Πολυτεχνείο Κρήτης

Τμήμα : Ηλεκτρονικών Μηχανικών
& Μηχανικών Υπολογιστών

Διπλωματική εργασία :

Σύγκριση αλγορίθμων για ανίχνευση εξέχουσων τιμών στα δίκτυα αισθητήρων

Σάββας Μπουρδάκης

Εξεταστική Επιτροπή:
Επ.Καθ. Αντώνιος Δεληγιαννάκης, επιβλέπων
Επ.Καθ. Μιχαήλ Γ. Λαγουδάκης
Καθ. Μίνως Γαροφαλάκης

Δεκέμβριος 2011

Ευχαριστίες

Πρώτα απ' όλα, θα επιθυμούσα να ευχαριστήσω, τον επιβλέπων καθηγητή μου Δρ.Αντώνιο Δεληγιαννάκη, για την αφιερωμένη καθοδήγηση, τις συμβουλές, την ενθάρρυνση, την γνώση και εμπειρία του, τον πολύτιμο χρόνο που αφιέρωνε συνεχώς για μένα και τη συνεχή υποστήριξή του σε όλη αυτήν την εργασία. Να ευχαριστήσω επίσης τους καθηγητές μου, κ. Λαγουδάκη και κ. Γαροφαλάκη για το χρόνο που αφιέρωσαν για την ανάγνωση και αξιολόγηση της παρούσας εργασίας. Επίσης ευχαριστώ τους Daniel Keren, Izchak Sharfman, Assaf Schuster, κ. Βασίλη Σαμολαδά, κ. Γαροφαλάκη και κ. Νίκο Γιατράκο για τις χρήσιμες συζητήσεις και τις προτάσεις τους.

Επίσης θα επιθυμούσα να ευχαριστήσω : την Εμμανουέλλα Μαρή, τους γονείς μου Γιώργο και Ευαγγελία Μπουρδάκη και την αδερφή μου Έυα, για την ατελείωτη αγάπη τους, την υποστήριξή τους, την πίστη και υπομονή που είχαν σε μένα όλα αυτά τα χρόνια. Επίσης σε όλους αυτούς που με ενθάρρυναν τα χρόνια σπουδών μου στο Πολυτεχνείο Κρήτης.

Πίνακας περιεχομένων

Κατάλογος σχημάτων..... 5
Κατάλογος πινάκων, Κατάλογος Γραφημάτων..... 6
Περίληψη..... 7
Κεφάλαιο I : Εισαγωγή..... 8
Κεφάλαιο II : Σχετική εργασία..... 12
Κεφάλαιο III : Βασικές γνώσεις στη γεωμετρική μέθοδο..... 21
Κεφάλαιο IV : Οργάνωση Προβλήματος..... 28
Κεφάλαιο V : Θεωρητικό υπόβαθρο των μετρικών ομοιότητας στη γεωμετρική μέθοδο..... 30
A. Έκφραση μετρικών ομοιότητας χρησιμοποιώντας τη γεωμετρική μέθοδο..... 30
B. Εκφράζοντας τους τοπικούς περιορισμούς στις σφαίρες ελέγχου..... 34
Κεφάλαιο VI : Λειτουργία κόμβων για έλεγχο ομοιότητας..... 41
A. Μοντέλο επικοινωνίας δύο κόμβων…..... 42
B. Γειτονική κόμβοι σε άμεση επικοινωνία.....49
C. Χειρισμός ερωτημάτων ελάχιστης υποστήριξης..... 50
D. Σημείωση για περίπτωση απώλειας μηνυμάτων..... 52
E. Χειρισμός προσθήκης και αφαίρεσης κόμβων..... 53
Κεφάλαιο VII : Πειράματα..... 54
A. Μεταβάλλοντας τις λειτουργίες ομοιότητας..... 56
B. Ερωτήματα ελάχιστης υποστήριξης 66
C. Μεταβάλλοντας τη διάσταση διανύσματος…..... 68
D. Μεταβάλλοντας το μέγεθος του Cluster…..... 70
Κεφάλαιο VIII : Συμπεράσματα.....72
Γλωσσάρι..... 73
Βιβλιογραφία..... 74

Κατάλογος σχημάτων

Όνομα *Σελίδα*
Εικόνα 1.1 : Τμήματα του κόμβου αισθητήρα mica2, πηγή : [37].....8
Εικόνα 1.2 : Sensor board του κόμβου mica, πηγή : [38].....8
Εικόνα 2 : Πηγές outlier και οι αντίστοιχες τεχνικές ανίχνευσης.....13
Εικόνα 3 : Αναπαράσταση outlier βασισμένο στην απόσταση, πηγή : [40].....15
Εικόνα 4 : Παράδειγμα προβλήματος της απόστασης βασισμένης τεχνικής και επίλυση του με την τεχνική βασισμένη στη πυκνότητα της γειτονίας, πηγή : [40].....16
Εικόνα 5 : Παράδειγμα πλεονεκτήματος της minimum Support προσέγγισης, πηγή :[40].....19
Εικόνα 6.1 : Τοπικοί περιορισμοί χρησιμοποιώντας γεωμετρική μέθοδος.....23
Εικόνα 6.2 : Διαδικασία balancing στη γεωμετρική μέθοδος.....25
Εικόνα 7 : Εισαγωγή περιθωρίου λάθους γύρο από το κατώτατο όριο, πηγή[41].....27
Εικόνα 8 : Η έκφραση λειτουργιών ομοιότητας χρησιμοποιώντας γεωμετρική μέθοδο.....29
Εικόνα 9 : Υπολογισμός μέγιστης / ελάχιστης L_2 *νόρμας, στο κύκλο ελέγχου*.....35
Εικόνα 10 : Υπολογισμός μέγιστης / ελάχιστης L_∞ *νόρμας, στο κύκλο ελέγχου*.....36
Εικόνα 11 : Υπολογισμός ελάχιστης τιμής όταν $\|C\|_2 < R$.....39
Εικόνα 12 : Ο Algorithm 1 που περιγράφει τον Simple τρόπο λειτουργίας του κόμβου.....44
Εικόνα 13 : Ο Algorithm 2 για την διαχείριση εισερχομένων πιθανόν μηνυμάτων.....45
Εικόνα 14 : Ο Algorithm 3 για την ανίχνευση παραβίασης τοπικού περιορισμού.....45
Εικόνα 15 : Ασφαλής ζώνες για L_∞, L_2 *και* L_1 *στις δυο διαστάσεις*.....48
Εικόνα 16 : Έλεγχοι τοπικών περιορισμών με χρήση ασφαλών ζωνών.....53
Εικόνα 17 : Τοπολογία των κόμβων για το Intel Lab Data.....54

Κατάλογος πινάκων

Όνομα *Σελίδα*
Πινάκας I : Ορισμοί συμβόλων που χρησιμοποιούνται στη γεωμετρική προσέγγιση 22
Πινάκας II : Τα σημεία μεγίστου ελαχίστου για τις L_∞, L_1 *και* L_2 39
Πίνακας III : Στοιχεία που κρατούνται, ανάλογα με το μοντέλο λειτουργίας.....51
Πίνακας IV : Τα περιεχόμενα του μηνύματος.....52
Πινάκας V : Παράμετροι και προεπιλεγμένες τιμές. 55

Κατάλογος Γραφημάτων

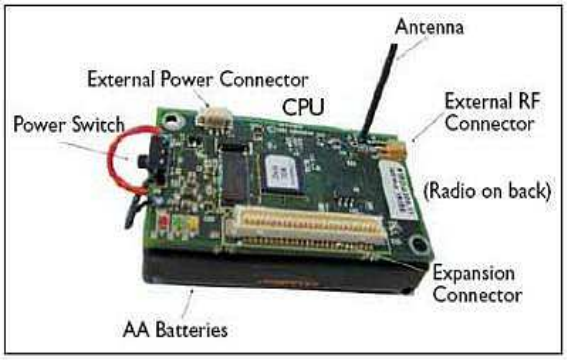
Όνομα *Σελίδα*
Γράφημα 1 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο L_1,..... 58
Γράφημα 2 : Intel Lab Data, ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1, ... 58
Γράφημα 3 : Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1,...59
Γράφημα 4:Weather Data, ηλιακή ακτινοβολία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1,..59
Γράφημα 5 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞.....60
Γράφημα 6 : Intel Lab Data, ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞.....60
Γράφημα 7 : Weather Data, σχετική υγρασία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞....61
Γράφημα 8:Weather Data, ηλιακή ακτινοβολία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞..61
Γράφημα 9 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2.....62
Γράφημα 10 : Intel Lab Data, ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2.....62
Γράφημα 11: Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2..63
Γράφημα 12:Weather Data, ηλιακή ακτινοβολία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2.62
Γράφημα 13 : Intel Lab Data, θερμοκρασία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, Cosine.64
Γράφημα 14 : Intel Lab Data, θερμοκρασία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, Jaccard.64
Γράφημα 15:Intel Lab Data, θερμοκρασία :#Μηνυμάτων σε σχέση με το κατώτατο όριο, Correlation……....65
Γράφημα 16 : Intel Lab Data, θερμοκρασία : #Ολικών παραβιάσεων σε σχέση με το κατώτατο όριο, μεταβάλλοντας την ελάχιστη υποστήριξη, L_2 *για την Proactive Broadcast Convex*……..……67
Γράφημα 17 : Weather Data, σχετική υγρασία : #Ολικών παραβιάσεων σε σχέση με το κατώτατο όριο, μεταβάλλοντας την ελάχιστη υποστήριξη, L_2 *για την Proactive Broadcast Convex*.....67
Γράφημα 18 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τη διάσταση του διανύσματος, L_2 *για την Proactive Broadcast Convex*.....68
Γράφημα 19 : Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τη διάσταση του διανύσματος, L_2 *για την Proactive Broadcast Convex*.....69
Γράφημα 20 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τον αριθμό των κόμβων στο cluster, L_2 *για την Convex Autobalance*.....70
Γράφημα 21 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τον αριθμό των κόμβων στο cluster, L_2 *για την Proactive Broadcast Convex*.....71

Περίληψη :

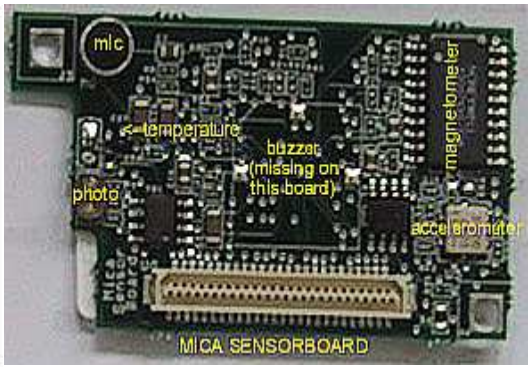
Σημαντικός αριθμός ερευνητικών εργασιών έχουν εστιάσει τα τελευταία χρόνια στην ανίχνευση εξέχουσων τιμών (outliers) σε δίκτυα αισθητήρων. Υπάρχουν πολλοί τρόποι να οριστεί κάποια τιμή/κόμβος ως outlier, αλλά γενικά outlier μπορεί να θεωρηθεί ένας αισθητήρας του οποίου οι μετρήσεις αποκλίνουν σημαντικά από τις μετρήσεις άλλων, γειτονικών, κόμβων. Η ανίχνευση μη φυσιολογικών μετρήσεων ενός κόμβου είναι ενδιαφέρουσα, επειδή αυτό το γεγονός μπορεί να βοηθήσει είτε στον εντοπισμό δυσλειτουργίας κάποιου κόμβου είτε για να ανιχνεύσει έναν κόμβο που παρατηρεί ένα τοπικά ενδιαφέρον φαινόμενο (π.χ. μια πυρκαγιά). Στην εργασία αυτή παρουσιάζουμε νέους αλγόριθμους για την ανίχνευση outlier στα δίκτυα αισθητήρων, βασισμένους στη γεωμετρική προσέγγιση. Σε αντίθεση με προηγούμενες εργασίες οι αλγόριθμοι μας εκτελούν έναν κατανεμημένο έλεγχο παρακολούθησης outlier κόμβων, παρουσιάζουν 100% ακρίβεια στον έλεγχο (υποθέτοντας ότι δεν υπάρχουν απώλειες μηνυμάτων) και η μετάδοση των μηνυμάτων μπορεί να περιοριστεί μόνο σε ένα μέρος των εποχών, επιτρέποντας έτσι στους κόμβους με ασφάλεια να απέχουν στη μετάδοση μηνυμάτων για αρκετές εποχές, περιορίζοντας έτσι την κατανάλωση ενέργειας. Η προσέγγισή μας στηρίζεται στη μετατροπή κοινών μετρικών ομοιότητας με τέτοιο τρόπο, ώστε να επιτρέπει την εφαρμογή της πρόσφατα προτεινόμενης γεωμετρικής προσέγγισης η οποία μας επιτρέπει να ελέγχουμε με ακρίβεια, κατά πόσον μια σύνθετη, ενδεχομένως μη γραμμική συνάρτηση είναι πάνω ή κάτω από ένα συγκεκριμένο όριο. Στη συνέχεια προτείνουμε ένα γενικό πλαίσιο ενώ παρουσιάζουμε πολλαπλούς τρόπους λειτουργίας, οι οποίοι επιτρέπουν σε κάθε κόμβο αισθητήρα να παρακολουθεί με ακρίβεια την ομοιότητά του με άλλους κόμβους. Τα πειράματα μας αποδεικνύουν ότι οι αλγόριθμοι μας μπορούν να ανιχνεύσουν με ακρίβεια outlier, σε ένα κλάσμα του κόστους επικοινωνίας που θα απαιτούσε μια συγκεντρωτική προσέγγιση (ακόμα και στην συγκεντρωτική περίπτωση όπου ο κεντρικός κόμβος βρίσκεται σε ακτίνα επικοινωνίας με όλους τους κόμβους αισθητήρων). Επιπλέον, δείχνουμε ότι αυτή η εξοικονόμηση εύρους ζώνης γίνεται ακόμα μεγαλύτερη, δεδομένου ότι ενσωματώνουμε περαιτέρω βελτιστοποιήσεις στον αλγόριθμό μας.

I. ΕΙΣΑΓΩΓΗ

Ένα δίκτυο αισθητήρων αποτελείται συνήθως από ένα μεγάλο αριθμό μικρών κόμβων αισθητήρων χαμηλού κόστους που κατανέμονται σε μια μεγάλη περιοχή, με έναν ή ενδεχομένως περισσότερους κόμβους βάσεων, συλλέγοντας τις αναγνώσεις των κόμβων αισθητήρων. Τα δίκτυα αισθητήρων αποτελούνται χαρακτηριστικά από τις μικρές συσκευές που εξοπλίζονται με μια πηγή ισχύος, μια μονάδα επεξεργασίας με περιορισμένη επεξεργαστική ισχύ και μνήμη, μία ή περισσότερες συσκευές αίσθησης (όπως για θερμοκρασία, υγρασία, ακτινοβολία, θερμότητα, πίεση, ήχο, δόνηση, τάση κ.α.) οι οποίες λαμβάνουν μετρήσεις και μια μονάδα επικοινωνίας για να αναμεταδώσουν αυτές τις αναγνώσεις ή το αποτέλεσμα της επεξεργασίας τους σε άλλους κόμβους αισθητήρων που βρίσκονται στην εμβέλεια τους.



Εικόνα 1.1: Τμήματα του κόμβου mica2



Εικόνα 1.2 : Το Sensorboard κόμβου mica

Οι πρόσφατες πρόοδοι στη μικροηλεκτρονική έχουν επιτρέψει την ανάπτυξη μεγάλης κλίμακας δικτύων αισθητήρων για ποικίλες εφαρμογές παρακολούθησης και ελέγχου. Εκτίνονται από οικολογικές εφαρμογές (παρακολούθηση της πανίδας ή ενός οικολογικού συστήματος), εφαρμογές στην υγειονομική περίθαλψη (την υγεία και την ιατρική παρακολούθηση), στρατιωτικές εφαρμογές (ανίχνευση και παρακολούθηση οχημάτων πίσω από τις εχθρικές γραμμές, την επιτήρηση πεδίων μαχών), τον έλεγχο κυκλοφορίας, τη γεωργία, τον έλεγχο παραγωγής και εφαρμογές ελέγχου καταστροφών (παρακολούθηση εκ των προτέρων ενός επικίνδυνου καιρικού φαινόμενου π.χ. τσουνάμι ή μιας πυρκαγιάς) είναι μόνο λίγες που μπορούμε να αναφέρουμε. Σε τέτοιες εφαρμογές που ανιχνεύουν ενδιαφέροντα γεγονότα, μπορούν να απαιτούν από τους αισθητήρες να συλλέγουν τις μετρήσεις που κρίνονται "παρόμοιες" ή να ανιχνευθεί εάν οι μετρήσεις των αισθητήρων είναι "παρόμοιες", με τις μετρήσεις των κοντινών αισθητήρων [6]. Ενδιαφέρον παρουσιάζει η διαδικασία ανίχνευσης της περίπτωσης όπου οι μετρήσεις των δύο κόμβων δεν είναι όμοιες, επειδή το γεγονός αυτό μπορεί να βοηθήσει στον εντοπισμό (i) είτε κάποιας δυσλειτουργίας του κόμβο είτε (ii) έναν κόμβο που ξεκινάει να παρατηρήσει ένα τοπικά ενδιαφέρον φαινόμενο (πχ, μια πυρκαγιά).

Η προαναφερόμενη διαδικασία ανίχνευσης, αναφέρεται συχνά ως ανίχνευση outlier. Ανάλογα με την εμβέλεια των δεδομένων που χρησιμοποιούνται και μέχρι ποιο επίπεδο

ελέγχω για την ανίχνευση, έχει γίνει κατηγοριοποίηση ως προς τον τύπο outlier[34]. Το local outlier αναφέρεται σε κάποια τοπική αλλαγή για μια περιορισμένη περιοχή ελέγχου και global outlier αναφέρεται για μια ποιο σφαιρική αλλαγή που συσχετίζεται συνήθως με όλο το δίκτυο. Υπάρχουν τεχνικές που προσδιορίζουν outlier βασιζόμενες μόνο σε χρονικές συσχετίσεις (μετρήσεις που παρατηρήθηκαν σε μια στιγμή του χρόνου, σχετίζονται με μετρήσεις που παρατηρήθηκαν σε προηγούμενες χρονικές στιγμές) ή χωρικές συσχετίσεις (μετρήσεις από τους κόμβους αισθητήρων που γεωγραφικά ο ένας βρίσκεται κοντά στον άλλο, αναμένεται να συσχετίζονται κατά ένα μεγάλο βαθμό). Πρέπει να σημειωθεί ότι ενώ υπάρχουν διάφοροι τρόποι για την ταξινόμηση και τον εντοπισμό των κόμβων με ακραίες τιμές, η εργασία μας στοχεύει στη σημαντική περίπτωση όπου η ομοιότητα των μετρήσεων μεταξύ διαφορετικών κόμβων απαιτείται στη διαδικασία ανίχνευσης outlier, εκμεταλλευόμενη και χωρικές συσχετίσεις. Επομένως δεν μπορεί να εφαρμοστεί σε άλλα σενάρια, παραδείγματος χάριν όταν μπορούν οι μετρήσεις ενός κόμβου να χαρακτηριστούν ως outlier, βασισμένες μόνο στις προηγούμενες μετρήσεις του κόμβου εκμεταλλευόμενη απλώς χρονικές συσχετίσεις.

Για μια εφαρμογή παρακολούθησης είναι συχνά κρίσιμο να είναι σε θέση να ανιχνεύσει τα ενδιαφέροντα γεγονότα με απόλυτη ακρίβεια. Παραδείγματος χάριν στις εφαρμογές όπου οι αισθητήρες έχουν αναπτυχθεί για την ανίχνευση φαινομένων της φύσης, όπως τα τσουνάμι, ή καθιζήσεις εδάφους ή χιονοστιβάδες, μια αποτυχία να ανιχνευθεί γρήγορα και να γίνει αναφορά αυτών των φαινομένων, μπορεί να οδηγήσουν σε αδυναμία ασφαλούς εκκένωσης των ανθρώπων από τις περιοχές που βρίσκονται σε κίνδυνο. Σε άλλες εφαρμογές όπως η κατάσβεση πυρκαγιών στα δάση είναι πολύ πιο εύκολο να γίνει όταν η φωτιά ανιχνεύεται νωρίς, αλλά πολύ πιο δύσκολο ενώ μια πυρκαγιά βρίσκεται σε εξέλιξη ή έχει είδη κλιμακωθεί. Στο πλαίσιο της παρακολούθησης εξοπλισμού η αποτυχία να ανιχνευθεί με ακρίβεια ένα μέρος που δυσλειτουργεί, μπορεί να οδηγήσει στην καταστροφή ενός πολύ μεγαλύτερου και ακριβότερου συστήματος. Σε όλες τις παραπάνω εφαρμογές είναι κρίσιμο να έχουν ένα σύστημα ανίχνευσης ακραίων τιμών που αναφέρει κόμβους outlier με 100% ακρίβεια. Αλλά και στις περισσότερες εφαρμογές η ακρίβεια της ανίχνευσης αποτελεί σημαντικό πλεονέκτημα. Φυσικά, για να εξασφαλιστεί η μακροζωία του δικτύου αισθητήρων, η απαίτηση για ακριβή ανίχνευση πρέπει να συνυπάρξει με τεχνικές επεξεργασίας δεδομένων που μειώνουν αποτελεσματικά τον αριθμό των μηνυμάτων που μεταδίδονται από τους κόμβους αισθητήρων.

Ένα σημαντικό ποσό της προσπάθειας [2], [4], [6], [8], [27], [29], [33] ,[34] έχει πρόσφατα αναπτυχθεί στην ανίχνευση outlier. Ωστόσο καμία από τις υπάρχουσες τεχνικές δεν έχει αντιμετωπίσει το συγκεκριμένο πρόβλημα, το να είναι σε θέση να ανιχνεύσει με 100% ακρίβεια την ομοιότητα μεταξύ οποιουδήποτε ζευγαριού των κόμβων αισθητήρων, ενώ παράλληλα να επιτρέπει σε ορισμένες περιπτώσεις τους κόμβους αισθητήρων να απέχουν στη διαβίβαση οποιωνδήποτε πληροφοριών που αφορούν στις μετρήσεις τους.

Οι υπάρχουσες τεχνικές ανίχνευσης outlier έχουν ένα ή περισσότερα από τα εξής μειονεκτήματα:

i. Δεν μπορούν να αναγνωρίσουν την ομοιότητα των δύο κόμβων με 100% ακρίβεια. (υποθέτοντας ότι δεν υπάρχουν απώλειες μηνυμάτων).
ii. Απαιτούν τη μετάδοση πληροφοριών που είναι συγκρίσιμη σε κόστος επικοινωνίας με την εκτίμηση μιας συγκεντρωτικής ερώτησης.
iii. Απαιτούν από τους κόμβους να εκτελούν μεταδόσεις σε κάθε εποχή.

iv. Αντιμετωπίζουν λίγες μετρικές ομοιότητας και συχνά με διαφορετική λειτουργία για κάθε μετρική ομοιότητας που αντιμετωπίζουν και / ή δεν μπορούν να χειριστούν τις λειτουργίες ομοιότητας στις μετρήσεις κατά τη διάρκεια που συλλέγονται από διαφορετικούς κόμβους (όπως correlation coefficient ή L_∞ νόρμα).

Σε αυτή την εργασία παρουσιάζουμε ένα πλαίσιο ανίχνευσης outlier που είναι βασισμένο στην πρόσφατα προτεινόμενη γεωμετρική προσέγγιση [21], [22], [23], [24]. Η γεωμετρική προσέγγιση μας επιτρέπει να ελέγχουμε με ακρίβεια (και αποτελεσματικότητα) κατά πόσον μια σύνθετη ενδεχομένως μη γραμμική συνάρτηση η οποία υπολογίζεται πάνω στο μέσο όρο των διανυσμάτων που διατηρούνται στους κόμβους, είναι πάνω ή κάτω από ένα συγκεκριμένο κατώτατο όριο. Με λίγα λόγια κάθε αισθητήρας εκχωρείται αυτόματα σε μια ζώνη ελέγχου, η οποία είναι ένα υποσύνολο του πεδίου ορισμού και εξετάζει αν η ελεγχόμενη συνάρτηση σε οποιοδήποτε σημείο μέσα στη ζώνη ελέγχου της, μπορεί να είχε ξεπεράσει το κατώτατο όριο. Κάθε αισθητήρας που ανιχνεύει μια πιθανή παραβίαση κατώτατων ορίων κάνει μια μετάδοση (π.χ., σε έναν συντονιστή). Ενώ μας δίνεται η δυνατότητα να επιλέξουμε ζώνες ελέγχου με διαφορετικές μορφές [24] (σφαίρα, τρίγωνα, ελλείψεις), χωρίς απώλεια γενικότητας. Σε αυτήν την εργασία υιοθετούμε την απλούστερη περίπτωση όπου η μορφή κάθε ζώνης ελέγχου είναι μια σφαίρα [22]. Αυτό που μας εγγυάται η γεωμετρική προσέγγιση είναι ότι οποιαδήποτε στιγμή ο μέσος όρος των διανυσμάτων μετρήσεων (που είναι άγνωστο στους κόμβους αισθητήρων), θα βρεθεί εντός της ένωσης των τοπικών ζωνών ελέγχου που εξετάζονται από τους μεμονωμένους κόμβους αισθητήρων. Κατά συνέπεια εάν κανένας αισθητήρας δεν παρατηρήσει τοπικά μια πιθανή παραβίαση κατώτατων ορίων με βάση τις μετρήσεις του, τότε ούτε η ελεγχόμενη συνάρτηση που βασίζεται στις μετρήσεις όλων των κόμβων δεν θα έχει διασχίσει το κατώτατο όριο, δεδομένου ότι η τιμή της ελεγχόμενης συνάρτησης για τον αληθινό μέσο όρο διανυσμάτων έχει εξεταστεί σίγουρα από έναν τουλάχιστον κόμβο των αισθητήρων.

Όπως θα δείξουμε σε αυτό το έγγραφο, διάφορες κοινές μετρικές ομοιότητας(που απεικονίζονται στην Εικόνα 8) μπορούν να μετασχηματιστούν με τέτοιο τρόπο ώστε να επιτρέπεται η εφαρμογή της γεωμετρικής προσέγγισης. Κατόπιν η ομοιότητα μεταξύ οποιουδήποτε ζεύγους κόμβων, μπορεί απλά να εκφραστεί ως μια συνάρτηση της οποίας η αξία επιθυμούμε να βρεθεί επάνω/κάτω από ένα προκαθορισμένο κατώτατο όριο. Για παράδειγμα μπορούμε να θεωρήσουμε δύο κόμβους παρόμοιους, αν τα διανύσματα μετρήσεων τους έχουν μια L_1 απόσταση που είναι κάτω από ένα όριο ή να έχουν μια cosine ομοιότητα πάνω από ένα συγκεκριμένο κατώτατο όριο. Φυσικά, ο αριθμός των λειτουργιών που πρέπει να ελέγχουμε στο σενάριο της εφαρμογή μας μπορεί να αυξηθεί τετραγωνικά (στη χειρότερη περίπτωση $O\left(\binom{n}{2}\right)$) σε σχέση με τον αριθμό n των κόμβων αισθητήρων. Οι μετασχηματισμοί μας, μας επιτρέπουν να χρησιμοποιούμε τη γεωμετρική προσέγγιση και να αναπτύξουμε ένα γενικό πλαίσιο το οποίο (i) υποστηρίζει μια ποικιλία μετρικών ομοιότητας, (ii) εκτελεί τον έλεγχο με 100% ακρίβεια και (iii) επιτρέπει έτσι στους κόμβους συχνά να απέχουν από οποιαδήποτε επικοινωνία, ενώ όπως φαίνεται στην πειραματική αξιολόγηση μας, καταναλώνοντας μόνο ένα μέρος του εύρους ζώνης που θα απαιτούσαν οι συγκεντρωτικές προσεγγίσεις. Αυτό έρχεται σε αντίθεση με τις πρόσφατα προτεινόμενες τεχνικές ανίχνευσης outlier [6], [8], που απαιτούν για να εκτελεστούν από κάθε κόμβο να γίνεται μετάδοση σε κάθε εποχή. Η επέκταση ώστε να επιτρέπει τον προσδιορισμό μιας

ελάχιστης υποστήριξης (δηλαδή, πόσοι κόμβοι πρέπει να είναι παρόμοιοι με τον κόμβο ελέγχου, έτσι ώστε να μην θεωρείτε outlier, minimum support[6]) ενσωματώνεται στην εργασία μας.
Οι συνεισφορές μας μπορούν να συνοψιστούν ως εξής :

- Θα αποδείξουμε ότι πολλές κοινές μετρικές ομοιότητας που χρησιμοποιούνται για την ανίχνευση outlier, μπορούν να μετατραπούν κατάλληλα με τέτοιο τρόπο ώστε να επιτρέπεται η εφαρμογή της γεωμετρικής προσέγγισης, που μας επιτρέπει να παρακολουθούμε με ακρίβεια αν η ελεγχόμενη συνάρτηση βρίσκεται πάνω/κάτω από ένα συγκεκριμένο κατώτατο όριο.
- Προτείνουμε ένα γενικό πλαίσιο για την ανίχνευση outlier στα δίκτυα αισθητήρων χρησιμοποιώντας τη γεωμετρική προσέγγιση και προτείνουμε διάφορους τρόπους λειτουργίας των κόμβων, ώστε να επιτυγχάνουμε τον επιθυμητό στόχο παρακολούθησης. Το πλαίσιό μας υπολογίζει με απόλυτη ακρίβεια (υποθέτοντας καμία απώλεια μηνυμάτων) την ομοιότητα δύο κόμβων, μια ιδιότητα που απορρέει άμεσα από τη γεωμετρική προσέγγιση μετασχηματισμού.
- Εξετάζουμε τις περιπτώσεις για την παρακολούθηση της ομοιότητας μεταξύ των κόμβων αισθητήρων, όταν η επικοινωνία μεταξύ αυτών των αισθητήρων είναι είτε άμεση, είτε όχι. Δείχνουμε επίσης πώς το πλαίσιο μας επιτρέπει την απάντηση σε minimum support queries.
- Το προτεινόμενο πλαίσιο μας μπορεί να χειριστεί ποικίλες μετρικές ομοιότητας, χρησιμοποιώντας το ίδιο μοντέλο λειτουργίας στο κόμβο, χωρίς να εκφράζει διαφορετικό αλγόριθμο για κάθε μετρική ομοιότητας.
- Εκτελούμε μια εκτενή πειραματική αξιολόγηση με τη χρήση πραγματικών δεδομένων. Η ανάλυσή μας αποδεικνύει ότι οι αλγόριθμοί μας μπορούν να ανιχνεύσουν με ακρίβεια outlier σε ένα κλάσμα του κόστους επικοινωνίας που θα απαιτούσε μια συγκεντρωτική προσέγγιση (ακόμα και στην περίπτωση, όπου ο κεντρικός κόμβος βρίσκεται μόλις ένα hop μακριά από όλους τους κόμβους αισθητήρων). Ενώ οι κόμβοι μπορούν να απέχουν στη μετάδοση σε κάποιες εποχές. Επιπλέον, αποδεικνύουμε ότι αυτή η εξοικονόμηση εύρους ζώνης γίνεται ακόμα μεγαλύτερη, δεδομένου ότι ενσωματώνουμε περαιτέρω βελτιστοποιήσεις στον προτεινόμενο τρόπο λειτουργίας μας.

Η εργασία συνεχίζει ως εξής. Στην παράγραφο II παρουσιάζουμε τη σχετική εργασία. Η παράγραφος III παρέχει το απαραίτητο υπόβαθρο της γεωμετρικής προσέγγισης. Το τμήμα IV περιγράφει με λεπτομέρειες την οργάνωση του προβλήματος που θεωρούμε στη παρούσα εργασία. Στο τμήμα V δείχνουμε πώς διάφορες κοινές μετρικές ομοιότητας μπορούν να μετατραπούν ώστε να καταστεί δυνατή την εφαρμογή της γεωμετρικής προσέγγισης. Τμήμα VI περιέχει το πλαίσιο λειτουργίας των κόμβων και τους αλγόριθμους μας, για την παρακολούθηση της ομοιότητας ενός κόμβου με οποιοδήποτε άλλο αισθητήρα, το οποίο είναι απαραίτητο στην εφαρμογή μας. Η πειραματική αξιολόγησή μας παρουσιάζεται στην παράγραφο VII, ενώ η παράγραφος VIII περιέχει τις τελικές παρατηρήσεις και τις μελλοντικές προτάσεις.

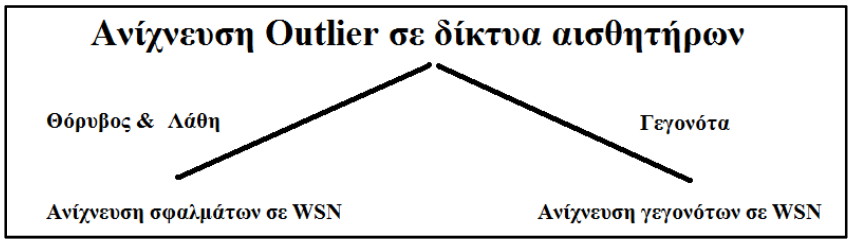
II. ΣΧΕΤΙΚΗ ΕΡΓΑΣΙΑ

Τα τελευταία χρόνια έχει γίνει σημαντική προσπάθεια σχετικά με τον καθορισμό και τον σχεδιασμό πρότυπων, για την ανάκτηση δεδομένων που βασίζονται σε δίκτυα αισθητήρων [17], [30]. Σε ένα τέτοιο περιβάλλον είναι σημαντικό να επεξεργάζονται τα περισσότερα στοιχεία, όσο είναι δυνατόν με αποκεντρωμένο τρόπο ώστε να αποφευχθούν οι περιττές επικοινωνίες και να μειώνετε το κόστος υπολογισμού. Επιτρέποντας εντός δικτύου επεξεργασία, μειώνετε δραματικά το κόστος της επικοινωνίας συχνά με τάξεις μεγέθους και συνεπώς οδηγεί σε παρατεταμένη διάρκεια ζωής του δικτύου. Ένας άλλος σημαντικός στόχος είναι ότι η όλη διαδικασία πρέπει να είναι αυτόνομη, έτσι ώστε να μπορεί να εργαστεί σε περιβάλλοντα χωρίς παρακολούθηση και να προσαρμόζεται αυτόματα στις αλλαγές στο περιβάλλον του δικτύου αισθητήρων. Πολλαπλοί τρόποι οργάνωσης του δικτύου έχουν προταθεί, συμπεριλαμβανομένων και ιεραρχικών (δηλ., με μορφή δέντρου) οργανώσεων όπως το δέντρο συνάθροισης [17], [26], [32], σχηματισμούς με cluster [3], [9], [20] , [31] ή ακόμα και εντελώς ad-hoc σχηματισμούς [1], [13], [16].

Η ανίχνευση outlier έχει ερευνηθεί ευρέως τα τελευταία χρόνια σε διάφορα γνωστικά αντικείμενα [39] (στατιστική, μηχανική μάθηση, θεωρία πληροφορίας και φασματική αποσύνθεση) και έχει χρησιμοποιηθεί σε πολυάριθμες περιοχές εφαρμογών. Στον τομέα των ασύρματων δικτύων αισθητήρων[34], οι μετρήσεις που ανιχνεύονται και αποκλίνουν σημαντικά από το φυσιολογικό πρότυπο, θεωρούνται outlier. Ο εντοπισμός κόμβων outlier παρέχει : (i) γεγονότα που πρέπει να αναφερθούν, (ii) αξιόπιστα δεδομένα και (iii) ασφαλής λειτουργία του δικτύου. Συγκεκριμένα η ανίχνευση outlier ελέγχει την ποιότητα των μετρημένων δεδομένων και βελτιώνει την ευρωστία της ανάλυσης δεδομένων, κάτω από την παρουσία θορύβου και ελαττωματικών αισθητήρων, έτσι ώστε να μειώνονται τα γενικά έξοδα επικοινωνίας. Παρέχει επίσης έναν ικανό τρόπο να ψάξει για τις τιμές που δεν ακολουθούν τη πρότυπο μορφή των δεδομένων, των υπολοίπων αισθητήρων στο δίκτυο. Οι ανιχνευμένες τιμές συνεπώς αντιμετωπίζονται όπως και τα γεγονότα, που δείχνουν την αλλαγή κάποιου φαινομένου που είναι ενδιαφέρον.

Ανάλογα με την πιθανή πηγή που τα παράγει, τα outlier έχουν διαφορετική ταυτότητα[34]. Πιθανές πηγές outlier στα δεδομένα που συλλέγονται από τα δίκτυα αισθητήρων περιλαμβάνουν τον θόρυβο και τα λάθη, πραγματικά γεγονότα(κάτι καλό) και κακόβουλες επιθέσεις. Outlier μπορούν να είναι είτε δεδομένα που φτάνουν από κακές μετρήσεις ή χαλασμένους κόμβους(errors), ή μπορεί να είναι και ένα ενδιαφέρον γεγονός που διαφέρει από τις υπόλοιπες μετρήσεις(events)(Εικόνα 2). (i)Errors είναι συνήθως μια αυθαίρετη αλλαγή που είναι πολύ διαφορετική από τα υπόλοιπα δεδομένα. Στόχος είναι αφού εντοπιστούν να διορθωθούν με σκοπό τη χρήση τους, ή να απορριφθούν ή να αναφερθούν. Λανθασμένοι αισθητήρες είναι πιθανό να παράγουν τέτοιες αυθαίρετες μετρήσεις οι οποίες δεν αντικατοπτρίζουν την πραγματική κατάσταση παρακολούθησης των περιβαλλοντικών φαινομένων. Outlier που προκαλούνται από άλλες πηγές πρέπει να εντοπιστούν, καθώς μπορεί να περιέχουν σημαντικές πληροφορίες όπως (ii)Events, γεγονότα που παρουσιάζουν μεγάλο ενδιαφέρον για τους ερευνητές. Αυτές οι event-based

εφαρμογές απαιτούν από τους κόμβους αισθητήρων, μόλις ανιχνευθεί το υπό παρακολούθηση φαινόμενο, να εκθέσουν το γεγονός στον κόμβο βάσης με έναν έγκυρο τρόπο. Event είναι ένα συγκεκριμένο φαινόμενο που αλλάζει την κατάσταση του πραγματικού κόσμου. Χαρακτηριστικό αυτού του είδους outlier είναι ότι διαρκεί φυσιολογικά για ένα σχετικά μεγάλο χρονικό διάστημα, αλλάζοντας το πρότυπο των δεδομένων των αισθητήρων που ακολουθούσαν μέχρι εκείνη τη στιγμή. Το πρόβλημα είναι ότι επίσης οι ελαττωματικοί αισθητήρες παραγάγουν παρόμοια μακρόχρονα outliers (Fail dirty[6]) όπως τα γεγονότα και άρα είναι δύσκολο να διακρίνουμε τις δύο διαφορετικές πηγές outlier, εξετάζοντας δεδομένα μόνο του ίδιου κόμβου, βασιζόμενη δηλαδή μόνο σε παρελθοντικές μετρήσεις(χρονικούς συσχετισμούς). Άρα πρέπει να γίνει χρήση της ομοιότητας των δεδομένων από τους γειτονικούς κόμβους και χωρικών δεδομένων των αισθητήρων, εκμεταλλευόμενη έτσι χωρικούς και χρονικούς συσχετισμούς. Το χαρακτηριστικό των τεχνικών ανίχνευσης outlier και εντοπισμού γεγονότος είναι ότι βασίζονται στους χώρο-χρονικούς συσχετισμούς μεταξύ των δεδομένων αισθητήρων των γειτονικών κόμβων, για να διακρίνουν μεταξύ των γεγονότων και των λαθών.

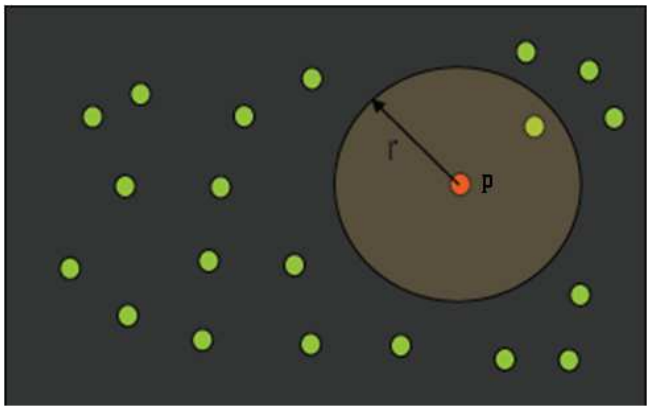


Εικόνα 2 : Δυο πηγές outlier και οι αντίστοιχες τεχνικές ανίχνευσης

Λόγω του χαμηλού κόστους του υλικού τους, οι κόμβοι αισθητήρων είναι επιρρεπής στο να παράγουν ακραίες αναγνώσεις. Έτσι πολλές τεχνικές έχουν προταθεί, που προσπαθούν να προσδιορίσουν τους κόμβους με "μη φυσιολογική" συμπεριφορά [34]. Στο [10], [11] προτείνεται ένας μηχανισμός καθαρισμού, πάνω σε ροές δεδομένων που παράγονται από τους αισθητήρες. Οι συγγραφείς παρουσιάζουν το ESP, ένα πλαίσιο καθαρισμού στοιχείων που επιτρέπει στους προγραμματιστές να προσδιορίσουν πέντε στάδια pipline καθαρισμού, με τη χρήση υψηλού επιπέδου declarative queries σε ροές δεδομένων που παράγονται από τους αισθητήρες. Υπάρχει υποδομή καθαρισμού στοιχείων μεταξύ των αισθητήριων συσκευών και των εφαρμογών. Όμοια η εργασία [7] εισάγει μια μονάδα καθαρισμού που έχει σχεδιαστεί για να συλλάβει τον θόρυβο σε δεδομένα ροής αισθητήρων, βασισμένο στην προκαθορισμένη κατανομή των δεδομένων με ένα δεδομένο μοντέλο λάθους $N(0, \delta^2)$. Συγκεκριμένα παρουσιάζει μια Bayesian προσέγγιση, για τη μείωση της αβεβαιότητας σχετικά με τα δεδομένα που προκύπτουν λόγω τυχαίου θορύβου. Στην εργασία[18] έχουν υιοθετηθεί Kalman φίλτρα κατά τη διάρκεια καθαρισμού των δεδομένων ή κατά τη διαδικασία ανίχνευσης outlier. Παρόλα αυτά, αν δεν γνωρίζουμε εκ των προτέρων την κατανομή των δεδομένων είναι δύσκολο να οριστούν οι παράμετροι και η συνδιακύμανση που χρησιμοποιούνται σε αυτά τα φίλτρα. Η τεχνική καθαρισμού που παρουσιάζεται στο [36], χρησιμοποιεί ένα μεταβαλλόμενο σταθμισμένο μέσο όρο που λαμβάνει υπόψη του τόσο τα πρόσφατα τοπικά δείγματα καθώς και τις αντίστοιχες τιμές από γειτονικούς κόμβους, για να εκτιμήσει τις πραγματικές μετρήσεις. Συγκεκριμένα συλλέγει τα βέβαια δεδομένα από τους αισθητήρες και υπολογίζει το σταθμισμένο

κινούμενο μέσο όρο(WMA). Δύο τεχνικές καθαρισμού outlier παρουσιάζονται στο [35]. Μια wavelet βασιζόμενη διαδικασία διόρθωσης των τιμών αναπτύσσεται για outliers, όπως ο θόρυβος ή τα περιστασιακά εμφανισμένα λάθη. Η άλλη τεχνική που χρησιμοποιεί Dynamic Time Warping (DTW) απόσταση μετρήσεων από γειτονικούς κόμβους. Αυτή βασίζεται στη σύγκριση της ομοιότητας ειδικά για outlier που είναι εσφαλμένα και θα διαρκέσουν για ένα ορισμένο χρονικό διάστημα. Σε αυτήν την τεχνική, κάθε κόμβος μετατρέπει ακατέργαστα δεδομένα στο wavelet πεδίο του χρόνου-συχνότητας και προσδιορίζει τις υψηλές συχνότητας εμφάνισης μετρήσεων ως outliers διορθώνοντας τις, χρησιμοποιώντας τους κατάλληλους συντελεστές wavelet. Η εργασία [28] προτείνει μια "fuzzy" προσέγγιση για να συμπεράνει το συσχετισμό μεταξύ των αναγνώσεων από διάφορους αισθητήρες, αντιστοιχεί μια τιμή εμπιστοσύνης σε καθένα από αυτούς και εκτελεί έπειτα ένα ασφαλές σταθμισμένο σχήμα μέσων όρων. Μια μέθοδος βασισμένη σε ιστογράμματα με μειωμένο κόστος επικοινωνίας παρουσιάζεται στο [25]. Ο σταθμός βάσης χρησιμοποιεί τις πληροφορίες των ιστογραμμάτων για να εξαγάγει τη κατανομή των δεδομένων του δικτύου και να φιλτράρει εκτός όσα δεν είναι outlier. Ο εντοπισμός των outlier επιτυγχάνεται με μια προκαθορισμένη απόσταση από το κατώτατο όριο ή το rank μεταξύ όλων των outlier. Μειονέκτημα αυτής της τεχνικής περιλαμβάνουν το γεγονός ότι η συλλογή περισσότερων πληροφοριών των ιστογραμμάτων από ολόκληρο το δίκτυο, θα προκαλέσει μεγάλη επιβάρυνση στο κόστος της επικοινωνίας.

Στο [15] οι συγγραφείς συζητούν ένα πλαίσιο για καθαρισμό εσφαλμένων δεδομένων εισόδου, χρησιμοποιώντας περιορισμούς ακεραιότητας, ενώ το [2], [33] είναι unsupervised τεχνικές ανίχνευσης outlier, που χρησιμοποιούνται για τον εντοπισμό των top-k τιμών που παρουσιάζουν την μεγαλύτερη απόκλιση από το global δείγμα του δικτύου. Συγκεκριμένα η εργασία [33] προτείνει μια απόσταση-βασισμένη τεχνική για να προσδιοριστούν k-top global outliers σε εφαρμογές snapshot και continuous query processing των δικτύων αισθητήρων. Κάθε κόμβος στο δέντρο διαβιβάζει μερικά χρήσιμα στοιχεία στο γονέα του, μετά τη συλλογή των δεδομένων που έχουν αποσταλεί από τα παιδιά του. Ο σταθμός βάσης έπειτα κατά προσέγγιση, υπολογίζει k-top global outliers και μεταδίδει αυτά τα outlier σε όλους τους κόμβους του δικτύου για επαλήθευση. Εάν οποιοσδήποτε κόμβος διαφωνεί στα global αποτελέσματα, αυτός θα στείλει τα πρόσθετα στοιχεία στο σταθμό βάσης πάλι, για την ανίχνευση outlier. Η διαδικασία αυτή επαναλαμβάνεται μέχρις ότου όλοι οι κόμβοι του δικτύου να συμφωνήσουν σχετικά με το συνολικό αποτέλεσμα που υπολογίζεται από τον σταθμό βάσης. Ωστόσο οι τεχνικές αυτές, δεν παρέχουν κανένα τρόπο, να ελέγχουν άμεσα την κατανάλωση του εύρους ζώνης, κατά συνέπεια απαιτούν εύρος ζώνης που είναι συχνά συγκρίσιμο με συγκεντρωτικές προσεγγίσεις[10] που χρησιμοποιούνται για την ανίχνευση outlier.

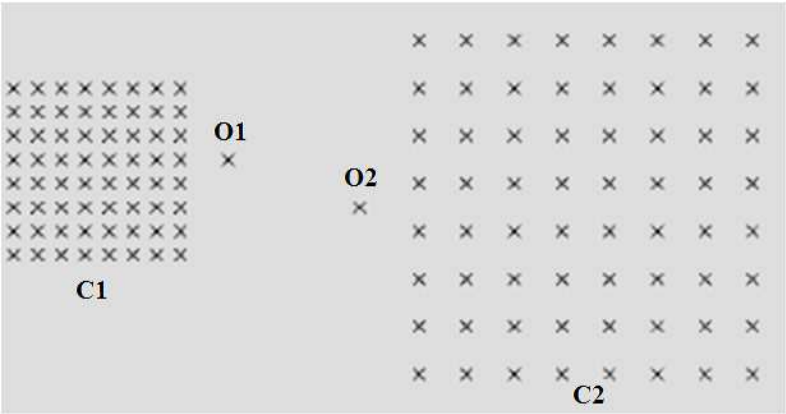


Εικόνα 3 : Σχηματική αναπαράσταση outlier βασισμένο στην απόσταση.
Μία τιμή p με λίγους γείτονες (σε απόσταση r έχει το πολύ K γείτονες)

Στη [12], παρουσιάζεται μια πιθανολογική τεχνική για τον καθαρισμό RFID ροών δεδομένων. Στο [27] οι συγγραφείς εισάγουν ένα νέο ορισμό της outlier τιμής, ως μια παρατήρηση που είναι αρκετά μακριά από τις περισσότερες άλλες παρατηρήσεις στο σύνολο δεδομένων. Ένας παρόμοιος ορισμός υιοθετείτε στην [19], όπου παρουσιάζεται μια διανεμημένη προσέγγιση ανίχνευσης outlier για δυναμικά σύνολα δεδομένων.

Στην εργασία[27] προτείνονται δύο διαφορετικοί μέθοδοι για outliers, η πρώτη χρησιμοποιεί έναν γρήγορο, βασισμένο στην απόσταση αλγόριθμο(distance-based algorithm) ενώ η δεύτερη υιοθετεί μια πιο ισχυρή προσέγγιση με βάση την πυκνότητα των μετρήσεων της κοντινής γειτονιάς του (local metrics based approach). Η πρώτη μέθοδος θεωρεί outlier μια τιμή που έχει λίγους γείτονες, δηλαδή σε απόσταση r έχει το πολύ K γείτονες. Εάν οι περισσότερες μετρήσεις του συνόλου δεδομένων, βρίσκονται σε απόσταση μεγαλύτερη από r από την μέτρηση, τότε η μέτρηση είναι outlier. Στη δεύτερη μέθοδος, outlier είναι τιμές που η πυκνότητα της κοντινής γειτονιάς τους είναι σημαντικά μικρότερη της ευρύτερης περιοχής. Ο εντοπισμός outlier στη συγκεκριμένη εργασία, γίνεται αναπτύσσοντας αποτελεσματικές τεχνικές για την εκτίμηση της κατανομής των δεδομένων του αισθητήρα, για ένα συρόμενο παράθυρο μετρήσεων. Αφού εκτιμηθεί η κατανομή των δεδομένων, μπορούμε να υπολογίσουμε την πυκνότητα του χώρου των δεδομένων γύρω από κάθε τιμή , ώστε να καθορίσουμε ποιες τιμές είναι outliers. Η εκτίμηση γίνεται με εκτιμητές πυκνότητας πυρήνων, αφού μπορούν επιτυχώς να προσεγγίσουν μια κατανομή που αλλάζει με την πάροδο του χρόνου, γιατί προσαρμόζουν τις παραμέτρους τους με κάθε νέα τιμή που έρχεται. Για τον αλγόριθμο βασισμένο στην απόσταση, κάθε φορά που εντοπίζεται ένα outlier, προωθείτε στον αντίστοιχο κόμβο γονέα όπου και ελέγχεται σε σχέση με το συνδυασμένο πρότυπο μοντέλο εκτίμησης εκείνου του κόμβου γονέα, προκειμένου να καθοριστεί εάν αυτή η αξία είναι επίσης outlier σε εκείνο το επίπεδο της ιεραρχίας. Συγκεκριμένα όταν οι κόμβοι φύλλα λαμβάνουν μια νέα παρατήρηση p, ο αισθητήρας μπορεί να χρησιμοποιήσει τη τρέχουσα κατανομή της συνάρτησης πυκνότητας, για να υπολογίσει τον αριθμό των τιμών που είναι στη γειτονιά αυτής της παρατήρησης p. Αν ο αριθμός αυτός είναι μικρότερος από ένα κατώτατο όριο t, το οποίο καθορίζετε από την εφαρμογή, τότε το p επισημάνετε ως outlier (Εικόνα 3). Οι ενδιάμεση κόμβοι αναζητούν outlier στην ένωση των αναγνώσεων που προέρχονται από αισθητήρες του υποδέντρου και γίνεται αν εξεταστούν μόνο οι τιμές που έχουν χαρακτηριστεί ως outliers από τα παιδιά τους. Ο καθορισμός outlier δεν γίνεται με χρήση κάποιας μετρικής ομοιότητας, αλλά

βασιζόμενος στο πλήθος των παρόμοιων τιμών που απέχουν λίγο από το τυχαίο δείγμα των πρόσφατων μετρήσεων. Δεν επηρεάζεται από το τι γίνεται στη γειτονία, αρκεί να βρεθούν k όμοιες τιμές είτε του κόμβου ή και από το υποδέντρο του. Outliers που βασίζονται στην απόσταση είναι δύσκολο να εντοπιστούν με ακρίβεια, αν ο χρήστης δεν είναι βέβαιος για κάποιο αξιόπιστο κατώτατο όριο. Επίσης με βάση την απόσταση υπάρχει κίνδυνος να υπάρξουν δυο ομάδες παρόμοιων τιμών (Εικόνα 4), η πρώτη C1 που η απόσταση μεταξύ των τιμών είναι μικρή και C2 που η απόσταση είναι λίγο μεγαλύτερη. Οι μετρήσεις στο C1 σύνολο είναι πιο πυκνές από το C2. Θα έπρεπε να προσδιοριστούν τα O1 και O2 ως outlier και οι ομάδες C1 και C2 ως ομάδες παρόμοιων τιμών. Η προσέγγιση με βάση την απόσταση δεν θα μπορούσε να ανιχνεύσει το O2, χωρίς να θεωρήσει και όλες τις παρατηρήσεις του C2 ως outlier. Το πρόβλημα αυτό επιλύει η τεχνική με βάση την πυκνότητα.



Εικόνα 4 : Παράδειγμα προβλήματος της απόστασης βασισμένης τεχνικής και επίλυση του με την τεχνική βασισμένη στη πυκνότητα της γειτονίας

Στη προσέγγιση με βάση την πυκνότητα των μετρήσεων της κοντινής γειτονιάς του, για να εκτιμηθεί αν μια παρατήρηση αποτελεί outlier, απαιτούνται δύο τιμές που μπορούν να υπολογιστούν από το μοντέλο πυκνότητας, (i) ο αριθμός των κοντινών γειτόνων και (ii) ο αριθμός παρατηρήσεων σε ένα ευρύτερο διάστημα της περιοχής. Οι κόμβοι φύλλα ελέγχουν αν μια μέτρηση είναι outlier στο δικό τους επίπεδο, βασιζόμενοι στο δικό τους μοντέλο εκτίμησης. Επίσης ελέγχουν και για τα υψηλότερα επίπεδα ιεραρχίας αφού χρησιμοποιούν αντίγραφο του μοντέλου εκτίμησης του πατέρα τους. Κάθε κόμβος φύλλο έχει ένα αντίγραφο της global συνάρτησης πυκνότητας πιθανότητας, εκτός από το δικό του μοντέλο εκτίμησης και έτσι μόνο οι αισθητήρες φύλλα ανιχνεύουν outliers. Για το λόγο αυτό οι ενδιάμεσοι κόμβοι στέλνουν τις μετρήσεις τους και το μοντέλο εκτίμησης τους στο πατέρα τους, ώστε να λάβουν αντίστοιχα ενημερώσεις του μοντέλου εκτίμησης του πάτερα τους. Ενώ επίσης μεταβιβάζουν αυτές τις πληροφορίες στα φύλλα ή το χαμηλότερο επίπεδο. Δεδομένου ότι αυτή η τεχνική για την ανίχνευση outlier είναι βασισμένη σε τοπικές συνθήκες είναι ακριβότερη από τη μέθοδο ανίχνευσης outlier βασισμένη στην απόσταση. Ενώ, η διατήρηση αντιγράφου της global συνάρτησης κατανομής σε όλους τους κόμβους αισθητήρων απαιτεί συχνές ενημερώσεις σε όλη τους αισθητήρες και τους γονείς τους.

Η εργασία [4],[29] αντιμετωπίζει το πρόβλημα ανίχνευσης outlier χρησιμοποιώντας ένα τοπικό πρωτόκολλο ψηφοφορίας. Στο έγγραφο αυτό[29] μελετάται το πρόβλημα του προσδιορισμού ελαττωματικών αναγνώσεων, χωρίς να ασχολείται με την ανίχνευση των

σημαντικών γεγονότων. Προγενέστερες εργασίες για το προσδιορισμό του συσχετισμού των αναγνώσεων από τους αισθητήρες, λάμβαναν υπόψη μόνο την απόσταση μεταξύ των κόμβων. Ενώ είναι δυνατό οι αναγνώσεις δύο κοντινών αισθητήρων να διαφέρουν αρκετά. Κατά συνέπεια είναι κρίσιμο να χρησιμοποιείτε ο συσχετισμός των αναγνώσεων και όχι η απόστασή τους. Προτείνεται ένα εντός δικτύου σχέδιο ψηφοφορίας για τον καθορισμό των ελαττωματικών αναγνώσεων, λαμβάνοντας υπόψη το συσχετισμό των αναγνώσεων μεταξύ των κόμβων αισθητήρων και της εμπιστοσύνης ενός κόμβου. Στον προτεινόμενο αλγόριθμο κάθε κόμβος συνδέεται με μια αξία εμπιστοσύνης (αποκαλούμενη sensor rank) που θα χρησιμοποιηθεί στην ψηφοφορία. Στην οποία ένας κόμβος που έχει μεγάλο αριθμό παρόμοιων γειτόνων ή έναν μεγάλο αριθμό "καλών αναφορών" θα έχει υψηλό sensor rank. Αρχικά για τον εντοπισμό outlier πρέπει να έχει κατασκευαστεί το δίκτυο συσχετισμού και να έχουν υπολογιστεί τα sensor rank. Στη συνέχεια στη πρώτη φάση(αυτοδιάγνωση)του αλγορίθμου, ο κόμβος ελέγχει εάν η τρέχουσα μέτρηση είναι ασυνήθιστη ή όχι, συγκρίνοντας το τρέχον διάνυσμα ανάγνωσης με το προηγούμενο χαρακτηρισμένο φυσιολογικό διάνυσμα μετρήσεων του. Αν οι μετρήσεις αυτές μοιάζουν συνεχίζω και δεν θεωρώ τη μέτρηση αυτή ως outlier, αν όμως δεν μοιάζουν προχωρεί στη δεύτερη φάση(φάση διαγνώσεων γειτόνων) όπου ο κόμβος συσκέπτεται με τους γειτονικούς κόμβους για την περαιτέρω επιβεβαίωση. Για μια ύποπτη μέτρηση πρέπει να συμβουλευτεί όλους τους πιθανούς γείτονες, από τους οποίους παίρνει αντίστοιχα θετική/αρνητική ψήφο με βάση τη συσχέτιση και το sensor rank τους(οι ψηφοφορίες από τους γείτονες με υψηλό sensor rank είναι πιο σημαντικές, ενώ οι ψηφοφορίες από τους γείτονες με χαμηλό πρέπει να μην επηρεάζουν τόσο πολύ την ψηφοφορία). Εάν μια ανάγνωση χαρακτηριστεί ως outlier θα πρέπει να βγει εκτός, χωρίς να λαμβάνεται υπόψη. Καθώς επίσης τέτοιοι κόμβοι δεν θα συμμετάσχουν στην ψηφοφορία, δεδομένου ότι αυτοί οι αισθητήρες είναι πιθανό να μολύνουν το αποτέλεσμα της ψηφοφορίας.

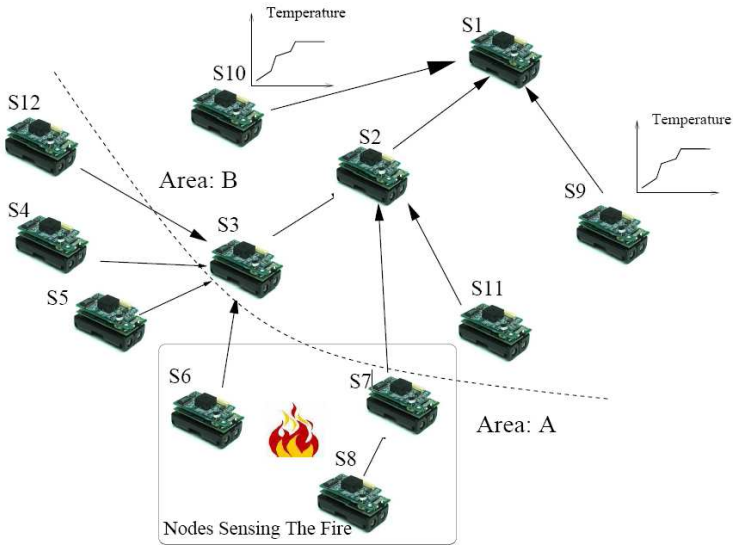
Ωστόσο συστήματα που χρησιμοποιούν τοπικό πρωτόκολλο ψηφοφορίας είναι επιρρεπείς σε λάθη(Εικόνα 5), όταν οι κόμβοι που παρατηρούν κάποιο ενδιαφέρον γεγονός με αποτέλεσμα να παράγουν outlier αναγνώσεις, δεν βρίσκονται σε άμεση επικοινωνία[6]. Επιπλέον, το πλαίσιο [29] απαιτεί να διατηρηθεί ένα δίκτυο συσχετισμού, ενώ ακολουθείτε μια πολύπλοκη διαδικασία από αιτήσεις, ψηφοφορίες καθώς και υπολογισμούς ομοιότητας και sensor rank. Ενώ επίσης το πλαίσιο αυτό είναι δύσκολο να είναι πλήρες ενημερωμένο σε κάθε εποχή, καθώς το μοντέλο υπολογισμού sensor rank αναφέρεται σε στατιστικά στοιχεία κάποιας παλιάς εποχής, ενώ θα έπρεπε να ενημερώνετε σε κάθε εποχή. Αυτό είναι πολύ δύσκολο να γίνει λόγο της μεγάλης πολυπλοκότητας, ενώ το ίδιο ισχύει και για το δέντρο συσχετισμού. Επίσης παρόλο που αναφέρονται κάποιες προτεινόμενες λύσεις για την σειρά εκτέλεσης των κόμβων στον αλγόριθμο, δεν παρέχεται αξιοπιστία επειδή η σειρά με την οποία θα εκτελέσουμε τον αλγόριθμο στους κόμβους επηρεάζει στην ανίχνευση ελαττωματικών αναγνώσεων. Ένα άλλο μειονέκτημα είναι ότι μόνο αν αποτύχει στην πρώτη φάση θα συμβουλεύεται τους γείτονες, αυτό σημαίνει ότι αν ένας αισθητήρας χαλάει σιγά-σιγά πιθανότατα δεν θα γίνει αντιληπτός. Με αποτέλεσμα να αλλοιωθεί το τελικό αποτέλεσμα και ακόμα χειρότερα αν ο κόμβος αυτός έχει μεγάλο Sensor Rank, θα επηρεάσει σύντομα και τις γειτονικές μετρήσεις. Απορρίπτοντας ύποπτες μετρήσεις μόλις εντοπιστούν αυξάνω την πιθανότητα να χάσω ενδιαφέροντα γεγονότα. Για παράδειγμα ένας κόμβος που θα εντοπίσει ένα ενδιαφέρον γεγονός και δεν μοιάζει με τους γείτονες του τον

απορρίπτω ή άμα παρατηρήσει κάτι ενδιαφέρον κάποιος με χαμηλό sensor rank ίσος δεν του δώσω την απαραίτητη σημασία.

Το [8] πλαίσιο TACO λειτουργεί πάνω σε cluster οργανώσεις δικτύων και προσπαθεί να προσδιορίσει outliers βασιζόμενο στις συμπιεσμένες αντιπροσωπεύσεις (LSH) των λαμβανόμενων μετρήσεων. Το πλαίσιο της [6], χρησιμοποιείτε για τον εντοπισμό και την αφαίρεση outlier κατά τη διάρκεια υπολογισμού aggregate και group-by queries που τίθενται σε ένα δέντρο συνάθροισης[5], [17]. Στην εργασία αυτή παρουσιάζετε ένα νέο πλαίσιο συνάθροισης, που ανιχνεύει ύποπτες μετρήσεις από κόμβους outlier και αποφεύγει να ενσωματώσει τέτοιες μετρήσεις στον υπολογισμό των aggregate values. Οι χρήστες είναι σε θέση να καθορίσουν το ελάχιστο ποσό υποστήριξης (minimum support), δηλαδή το απαραίτητο πλήθος από γειτονικούς κόμβους "μάρτυρες", έτσι ώστε μια μέτρηση να μην χαρακτηριστεί outlier. Κατά τη διάρκεια εκτέλεσης ενός aggregate query, το πρότυπο υπολογισμού είναι βασισμένο στην ταυτόχρονη επεξεργασία και ανίχνευση outlier, με αποτέλεσμα να ανακοινώνει outliers και μάρτυρες εκτός από τον υπολογισμό των aggregates. Μέσω αυτής της διαδικασίας σε κάθε εποχή, υπολογίζετε στη ρίζα του δέντρου το aggregate, αποκλείοντας τους κόμβους που καθορίζονται ως outliers, βασιζόμενη στη διευκρινισμένη ελάχιστη υποστήριξη minimum support. Η ρίζα καθορίζει επίσης το τελικό σύνολο outliers και μαρτύρων. Ο προτεινόμενος αλγόριθμος για τον υπολογισμό των aggregate και την ταυτόχρονη διαχείριση outliers στα δίκτυα αισθητήρων, λειτουργεί διαδίδοντας προς τα πάνω στο δέντρο συνάθροισης τα aggregate μαζί με τα σύνολα μαρτύρων και outliers που υπολογίζονται από τους κόμβους στα υποδέντρα του. Συγκεκριμένα κάθε κόμβος μόλις λάβει τα μηνύματα από τους κόμβους παιδιών του αρχικοποιεί πρώτα τους καταλόγους μαρτύρων και outlier, με την ένωση των αντίστοιχων καταλόγων που λαμβάνει από τους κόμβους παιδιών του. Στους καταλόγους αναφέρονται τα id των αντίστοιχων κόμβων και η μέτρηση. Έπειτα ελέγχετε κάθε outlier, εξετάζοντας εάν η μέτρηση του μπορεί να βεβαιωθεί υπό την παρουσία είτε άλλων ύποπτων μετρήσεων από κόμβους, που βρίσκονται στο κατάλογο outlier ή από μετρήσεις κόμβων που έχουν είδη τοποθετηθεί στο κατάλογο μαρτύρων. Για όλα τα πιθανά ζεύγη αισθητήρων συγκρίνονται οι τελευταίες n αναγνώσεις, για τις οποίες πήραν μέτρηση στην ίδια εποχή και στη παρούσα εποχή λάβανε και οι δύο μετρήσεις. Αν μια μέτρηση δεν βρει k = minimum support μάρτυρες από άλλους γειτονικούς κόμβους θεωρείτε προσωρινά outlier (την τοποθετώ στο κατάλογο outlier), στη περίπτωση αυτή δεν απορρίπτω την τιμή απλά την προωθώ σε κόμβους που βρίσκονται σε υψηλότερο επίπεδο. Στη συνέχεια υπολογίζεται το aggregate, χωρίς τους κόμβους outlier, ενώ ενημερώνεται και ο κατάλογος μαρτύρων. Το πλεονέκτημα είναι ότι μόνο αν μια μέτρηση δεν έχει αποκτήσει το απαραίτητο ποσό ελάχιστης υποστήριξης μέχρι τη ρίζα θα θεωρηθεί outlier.

Στο παράδειγμα της Εικόνας 5, οι μετρήσεις των κόμβων S6, S7 και S8 μεταβάλλονται απότομα, επειδή παρατηρούν μια πυρκαγιά. Στο παράδειγμα ορίζεται, minimum support ίσο με 2. Ο κόμβος S3 που λαμβάνει τις τιμές των παιδιών του, παρατηρεί ότι ο κόμβος S6 είναι ύποπτος, αφού κανένας άλλος κόμβος στο υποδέντρο δεν εντοπίζει την πυρκαγιά. Εάν είχε χρησιμοποιηθεί ένα πρωτόκολλο ψηφοφορίας και απέρριπτε την συγκεκριμένη μέτρηση, θα είχε χαθεί μια κρίσιμη παρατήρηση. Στο προτεινόμενο πλαίσιο ο S3 απλά τοποθετεί τη μέτρηση προσωρινά στο κατάλογο outlier και τον προωθεί μαζί με τις υπόλοιπες πληροφορίες στον S2. Οι αναγνώσεις των κόμβων S8, S7 είναι παρόμοιες, αλλά έχουν ελάχιστη υποστήριξη ίση με 1, άρα ο S7 τοποθετεί

προσωρινά τις αναγνώσεις των S7 και S8 στο κατάλογο outlier που διαβιβάζεται στον S2. Ο κόμβος S2 στη συνέχεια λαμβάνει τους καταλόγους outlier από τους κόμβους S7 και S3, όπου σε αυτό το σημείο, οι κόμβοι S6, S7 και S8 μπορούν να φθάσουν στην απαραίτητη ελάχιστη υποστήριξη ώστε να μπορέσουμε να αναγνωρίσουμε την πυρκαγιά.



Εικόνα 5 : Παράδειγμα της minimum Support προσέγγισης.

Ένα δεύτερο βασικό χαρακτηριστικό του πλαισίου είναι ότι δεν κατασκευάζει εκ των προτέρων ένα ενιαίο δέντρο συλλογής για να υπολογίσει τιμές ενός aggregate query. Συγκεκριμένα ο προτεινόμενος αλγόριθμος βασίζεται σε μια περιοδική αναδιοργάνωση του δέντρου συλλογής, χρησιμοποιώντας απλά στατιστικά στοιχεία, όπως το πόσο συχνά οι μετρήσεις δύο κόμβων αισθητήρων είναι παρόμοιες.

Οι αλγόριθμοι στις [6], [8], [27], έχουν σημαντικά μειονεκτήματα σε σχέση με τις προτεινόμενες τεχνικές μας. Ίσως το πιο σημαντικό είναι που δεν παρέχουν ισχυρές εγγυήσεις για την ανίχνευση κόμβων outlier. Ακολουθούν μια καλύτερη προσπάθεια προσέγγισης τα πλαίσια ([6], [27]), ή ακόμα καλύτερα στο (TACO) που προσφέρει μερικές πιθανολογικές εγγυήσεις, οι οποίες όμως είναι μικρές όταν η ομοιότητα των δύο αισθητήρων είναι κοντά στο κατώτατο όριο ομοιότητας. Επιπλέον η εργασία στο [27] χρησιμοποιεί διαφορετικό τρόπο λειτουργίας του κόμβου, για διαφορετική μετρική ομοιότητας. Οπότε δεν μπορεί να χειριστεί με ένα ενιαίο τρόπο λειτουργίας, περισσότερες μετρικές ομοιότητας όπως στοχεύουμε σε αυτό το έγγραφο. Η εξοικονόμηση του εύρους ζώνης, σε σύγκριση με συγκεντρωτικούς αλγόριθμους είναι περιορισμένη στην [6]. Ενώ το TACO δεν μπορεί να λειτουργεί σε κάθε εποχή (χωρίς να είναι απαγορευτικά ακριβό σε εύρος ζώνης), αλλά μάλλον λειτουργεί καλύτερα. Και τα δύο,[6] και Taco απαιτούν από κάθε κόμβο να γίνεται μια μετάδοση σε κάθε εποχή.

Αντίθετα, οι προτεινόμενοι αλγόριθμοι μας (i) μπορούν να εγγυηθούν ότι η ομοιότητα (ή όχι) των δύο κόμβων, μπορεί να ανιχνευθεί πάντα με απόλυτη βεβαιότητα (με προϋπόθεση την αξιόπιστη επικοινωνία), (ii) μπορούν να λειτουργήσουν συνεχώς, υπολογίζοντας κατά συνέπεια την ομοιότητα των κόμβων σε κάθε εποχή και (iii) επιτρέπουν στους κόμβους αισθητήρων να απέχουν με ασφάλεια στη μεταβίβαση μηνυμάτων για αρκετές εποχές επιτυγχάνοντας έτσι σημαντική εξοικονόμηση εύρους ζώνης.

Η γεωμετρική προσέγγιση παρουσιάστηκε για πρώτη φορά στο [22]. Οι ζώνες παρακολούθησης που χρησιμοποιήθηκαν στο [22] ήταν σφαίρες, μια προσέγγιση που θα χρησιμοποιήσουμε και εμείς στην εργασία μας, λόγω των χαμηλότερων υπολογιστικών της απαιτήσεων. Η εργασία στο [14], [24] έδειξε ότι η χρήση ζωνών παρακολούθησης διαφορετικών σχημάτων δηλαδή, τρίγωνα ή ελλείψεις μπορεί να είναι πιο επωφελής σε σχέση με τον αριθμό των μεταδιδόμενων μηνυμάτων. Στην παράγραφο VI εκμεταλλευόμαστε επίσης (για μερικές μετρικές, όπως L_∞, L_1, L_2) την έννοια των ασφαλών ζωνών που εισάγονται στην [14], προκειμένου να μειωθεί περαιτέρω ο αριθμός διαβιβασθέντων μηνυμάτων. Η [21] παρουσιάζει ένα πλαίσιο για την αποδοτική αξιολόγηση aggregate threshold queries, γενικών συναρτήσεων σε κατανεμημένα δεδομένα (σε αντίθεση με ρυθμίσεις κατανεμημένων ροών δεδομένων που χρησιμοποιούνται στα [14], [22], [24]). Η εργασία στο [23] είναι η μόνη εργασία που γνωρίζουμε ότι εφαρμόζεται η γεωμετρική προσέγγιση, σε δίκτυα αισθητήρων (προκειμένου να απαντηθούν aggregate threshold queries). Στην [23] η aggregate function υπολογίζεται σε ολόκληρο το δίκτυο. Η εργασία μας δεν διαφέρει μόνο στο τύπο της ελεγχόμενης συνάρτησης, αλλά επίσης σε αυτή την εργασία ενδιαφερόμαστε να αξιολογήσουμε πολλαπλές ανά ζεύγη συναρτήσεις ομοιότητας, το οποίο δεν ήταν δυνατόν να γίνετε στο [23].

III. ΒΑΣΙΚΕΣ ΓΝΩΣΕΙΣ ΣΤΗ ΓΕΩΜΕΤΡΙΚΗ ΜΕΘΟΔΟ

Η γεωμετρική προσέγγιση μελετάει προσεγγίσεις της εξής μορφής : Υποθέτουμε ότι κάθε κόμβος κρατάει ένα d-διαστάσεων διάνυσμα μετρήσεων, ενώ μας έχει δοθεί μια αυθαίρετη συνάρτηση ελέγχου, η οποία είναι ορισμένη πάνω στο μέσο όρο των διανυσμάτων μέτρησης που συλλέγονται από τους κόμβους. Ενδιαφερόμαστε να ανιχνεύσουμε όταν η τιμή της συνάρτησης περνάει ένα προκαθορισμένο κατώτατο όριο. Η γεωμετρική προσέγγιση καθορίζει ένα τοπικό περιορισμό σε κάθε κόμβο, ενώ για όσο ικανοποιούνται αυτοί οι τοπικοί περιορισμοί με σιγουριά η τιμή της ελεγχόμενης συνάρτησης δεν έχει διασχίσει το κατώτατο όριο.

Στη συνέχεια θα περιγράψουμε βασικά σημεία της γεωμετρικής μεθόδου, η οποία προτείνει μια νέα προσέγγιση για την παρακολούθηση λειτουργιών ορισμένη πάνω στα δεδομένα n κατανεμημένων κόμβων. Ο Πίνακας I συνοψίζει τη χρήση γραφικών συμβόλων που χρησιμοποιείται στο παρόν έγγραφο. Οι αντίστοιχοι ορισμοί αναφέρονται σε αντίστοιχες περιοχές του κειμένου. Η Εικόνα 6.1 δείχνει τις βασικές ιδέες της γεωμετρικής προσέγγισης.

Κάθε κόμβος S_i διατηρεί ένα d - διαστάσεων τοπικό διάνυσμα μετρήσεων, το οποίο ορίζεται local statistics vector $\vec{v}_i$, όπου το j-οστό (j = 1…d) στοιχείο του local statistics vector του S_i, ορίζεται ως $\vec{v}_{j,i}$. Η διάσταση του διανύσματος είναι ίδια σε όλους τους κόμβους(δηλαδή, με το ίδιο πλήθος στοιχείων). Το global statistics vector $\vec{v}$ υπολογίζεται ως ο μέσος όρος όλων των local statistics vector (Το ίδιο πλαίσιο εφαρμόζεται επίσης όταν το global statistics vector διάνυσμα υπολογίζεται ως ο σταθμισμένος μέσος όρος των local statistics vectors). Κατά συνέπεια j-οστό (j=1…d) στοιχείο του global statistics vector, ορίζεται ως $\vec{v}_j$ και υπολογίζεται ως εξής :

$$\vec{v}_j = \frac{1}{n}\sum_{i=1}^{n} \vec{v}_{j,i} \ .$$

Για να ισχύει η γεωμετρική μέθοδος θα πρέπει, οποιαδήποτε υπό παρακολούθηση υποστηριζόμενη συνάρτηση $f : R^d \rightarrow R$ να ορίζεται πάνω στο global statistics vector $\vec{v}$, κατά συνέπεια να εκφράζεται πάνω στο μέσο όρο όλων των local statistics vectors. Ένα σημαντικό χαρακτηριστικό γνώρισμα είναι η δυνατότητα εφαρμογής της γεωμετρικής προσέγγισης ως λειτουργία κατώτατων ορίων που μπορεί γενικά να είναι και μη γραμμική. Λαμβάνοντας υπόψη ένα κατώτατο όριο T, το πλαίσιο στα [21], [22], [23], [24] μπορεί με ασφάλεια να προσδιορίσει αν $f(\vec{v}) > T$.

Η γεωμετρική προσέγγιση διαιρεί το στόχο της παρακολούθησης σε ένα σύνολο από τοπικούς περιορισμούς, έναν ανά κόμβο, τον οποίο κάθε κόμβος μπορεί να παρακολουθεί μεμονωμένα σε τοπικό επίπεδο. Για να το επιτύχει αυτό, κατά τη διάρκεια της λειτουργίας του αλγορίθμου, κάθε κόμβος S_i κρατάει (i) το estimate vector $\vec{e}$, που είναι ίσο με το global statistics vector $\vec{v}$, που υπολογίζεται από τα local statistics vectors που μεταδίδονται από τους κόμβους, σε κάποια χρονική περίοδο που έγινε κάποια ανταλλαγή

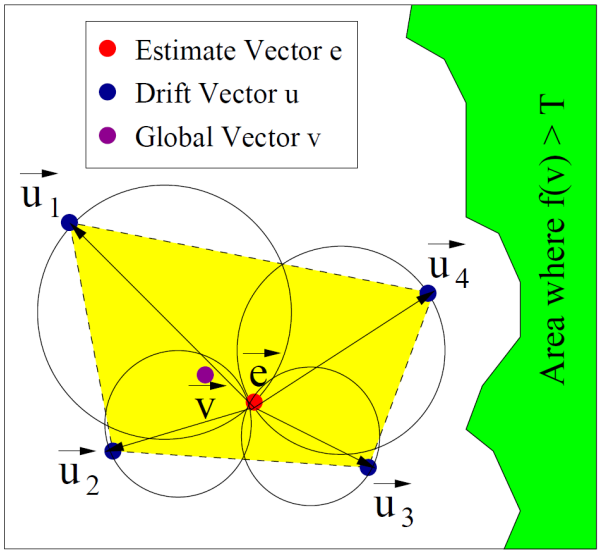
μηνυμάτων, και (ii) ένα delta vector $\overrightarrow{\Delta v_i}$ που ορίζεται ως η διαφορά του τρέχοντος local statistics vector με το τελευταίο local statistics vector που είχε μεταδοθεί από τον κόμβο S_i. Με βάση αυτές τις δύο ποσότητες, ο S_i υπολογίζει το drift vector του, $\vec{u}_i = \vec{e} + \overrightarrow{\Delta v_i}$. Επίσης πρόσθετες βελτιστοποιήσεις έχουν αναπτυχθεί στη γεωμετρική μέθοδο, όπως η δυνατότητα balance μόνο σε ένα μέρος του δικτύου σε περίπτωση violation. Σε αυτή την περίπτωση, ένα πρόσθετο διάνυσμα slack πρέπει να διατηρηθεί και να προστίθεται στον υπολογισμό του drift vector.

Symbol	Definition
S_i	The i-th sensor node
CN_i	The Comparison Neighborhood of S_i: With which nodes does S_i compute its similarity with?
W	Dimensionality of the measurements vector
d	Dimensionality of the local statistics vector
T	The similarity threshold
$\vec{e}$	The estimate vector. Its dimensionality is d
$\vec{v}_i$	The local statistics vector of S_i. Its dimensionality is d
$\vec{v}$	The true (not known by the sites) global statistics vector
$\Delta\vec{v}_i$	The delta vector of S_i. Calculated as the difference of the current local statistic vector from the last local statistic vector that S_i has transmitted.
$\vec{u}_i$	The drift vector of S_i. Equal to $\vec{e} + \Delta\vec{v}_i$
$B(\vec{e}, \vec{u}_i)$	The sphere having $\vec{e}$ and $\vec{u}_i$ as its diameter
$Conv(\vec{e}, \vec{u}_1, \ldots, \vec{u}_n)$	The convex hull determined by vectors $\vec{e}$, $\vec{u}_1, \ldots, \vec{u}_n$
$minSupp$	The specified minimum support

Πίνακας Ι : Ορισμοί συμβόλων που χρησιμοποιούνται στη γεωμετρική προσέγγιση.

Ο διανυσματικός χώρος R^d αντιπροσωπεύει τις πιθανές θέσεις που μπορεί να είναι το global statistics vector ανά πάσα στιγμή. Έστω όλα τα σημεία στον R^d για τα οποία $f(\vec{v}) \leq T$ χρωματίζονται με το ίδιο χρώμα (δηλαδή, το λευκό στην Εικόνα 6.1), ενώ τα υπόλοιπα σημεία χρωματίζονται με κάποιο διαφορετικό χρώμα (δηλαδή, το πράσινο Εικόνα 6.1). Ο στόχος είναι να μπορούμε να γνωρίζουμε το χρώμα του σημείου που αντιπροσωπεύει το global statistics vector $\vec{v}$. Επειδή οι κόμβοι δεν εκτελούν μεταδόσεις σε κάθε χρονική στιγμή, το τρέχων global statistics vector $\vec{v}$ δεν είναι γνωστό στους κόμβους. Ωστόσο αυτό που είναι εγγυημένο από τη γεωμετρική προσέγγιση είναι ότι το $\vec{v}$ θα βρίσκεται πάντα εντός του convex hull $Conv(\vec{u}_1, \ldots, \vec{u}_n)$ των drift vector και επομένως μέσα στο convex hull $Conv(\vec{e}, \vec{u}_1, \ldots, \vec{u}_n)$ των drift vector και του estimate vector. Κατά συνέπεια, εάν convex hull $Conv(\vec{e}, \vec{u}_1, \ldots, \vec{u}_n)$ είναι μονοχρωματικό (δηλ., είτε εξ ολοκλήρου ίσο ή κάτω από το κατώτατο όριο, είτε εξ ολοκλήρου πάνω από το κατώτατο

όριο), τότε όλοι οι κόμβοι είναι σίγουροι για το χρώμα της συνάρτησης f(), δεδομένου ότι αυτό θα συμπίπτει με το χρώμα της $f(\vec{e})$. Φυσικά κάθε κόμβος δεν μπορεί να υπολογίσει το $\mathrm{Conv}(\vec{e}, \vec{u}_1, \ldots, \vec{u}_n)$, δεδομένου ότι δεν γνωρίζει τα drift vector των άλλων κόμβων. Ωστόσο, μια σημαντική παρατήρηση στο[22] είναι ότι αν κάθε κόμβος S_i παρακολουθεί τη σφαίρα ελέγχου $B(\vec{e}, \vec{u}_i)$,που είναι κατασκευασμένη με διάμετρο από το estimate vector ως το drift vector του, τότε η ένωση αυτών των σφαιρών καλύπτει το convex hull. Κατά συνέπεια αρκεί κάθε κόμβος απλά να ελέγχει εάν η σφαίρα του είναι μονοχρωματική. Εάν όλες οι σφαίρες είναι μονοχρωματικές, τότε και το convex hull είναι σίγουρα μονοχρωματικό. Επομένως το ίδιο χρώμα θα έχουν και το $f(\vec{e})$ καθώς και το $f(\vec{v})$. Σε περίπτωση παραβίασης κάποιου τοπικού περιορισμού, οι κόμβοι μεταδίδουν τα local statistics vectors και ένα νέο estimate vector υπολογίζεται και γνωστοποιείται σε όλους τους κόμβους. Για να ελέγξουμε εάν μια σφαίρα είναι μονοχρωματική, υπολογίζουμε τη μέγιστη και ελάχιστη τιμή της f εντός της σφαίρας παρακολούθησης.



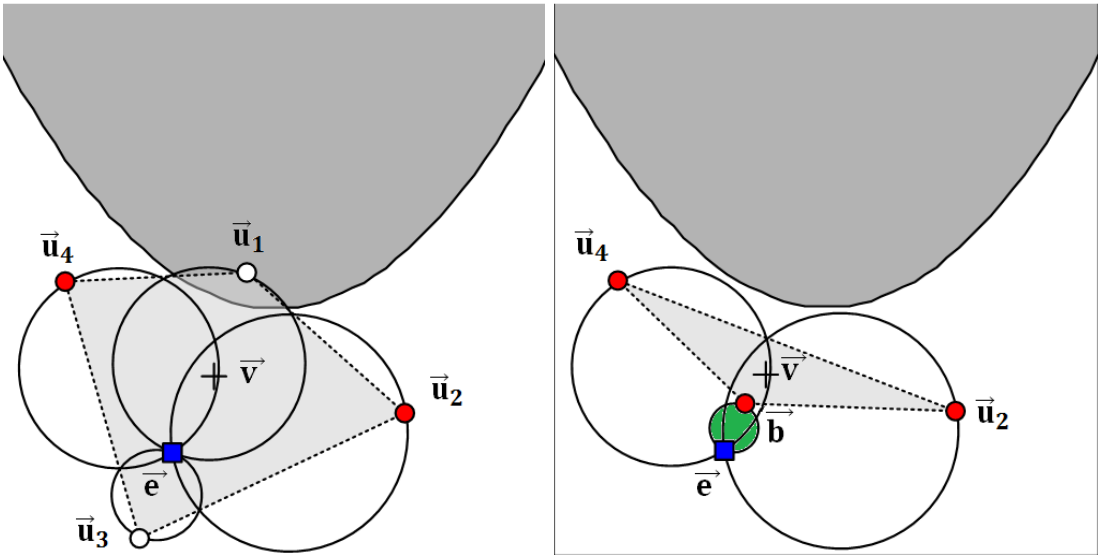
Εικόνα 6.1 : Τοπικοί περιορισμοί χρησιμοποιώντας τη γεωμετρική προσέγγιση. Κάθε κόμβος κατασκευάζει μια σφαίρα με διάμετρο το drift vector $\vec{u}$ του κόμβου και το estimate vector $\vec{e}$. Το global statistics vector $\vec{v}$ θα βρίσκεται εγγυημένα στο convex hull των $\vec{e}, \vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4$. Η ένωση των τοπικών σφαιρών καλύπτει το convex hull.

Στη συνέχεια θα περιγράψουμε τις πρόσφατες εργασίες [21], [22], [23], [24], που βασίζονται στη γεωμετρική προσέγγιση. Η εργασία [22] επιτρέπει τον έλεγχο των αυθαίρετων λειτουργιών κατώτατων ορίων πάνω σε κατανεμημένες ροές δεδομένων με αποδοτικό τρόπο. Ο προτεινόμενος αλγόριθμος είναι εξαιρετικά επεκτάσιμος και μειώνει σημαντικά το κόστος επικοινωνίας σε σύγκριση με τους συγκεντρωτικούς αλγορίθμους. Η επεκτασιμότητα και αποτελεσματικότητα στο περιορισμό της κατανάλωσης ενέργειας, σε προηγούμενες εργασίες παρακολούθησης ροών δεδομένων σε κατανεμημένα συστήματα, μπορούσαν να εφαρμοστούν μόνο σε απλές συναθροίσεις ή γραμμικές συναρτήσεις. Η εργασία αυτή πρότεινε μια λύση ενός γενικού πλαισίου για την παρακολούθηση αυθαίρετων ακόμα και μη γραμμικών συναρτήσεων κατώτατων ορίων. Στην εργασία παρουσιάζονται δυο αλγόριθμοι που βασίζονται στη γεωμετρική προσέγγιση και την διαίρεση του στόχου ελέγχου σε τοπικούς περιορισμούς για δύο διαφορετικά υπολογιστικά περιβάλλοντα. Ο

decentralized αλγόριθμος βασίζεται σε broadcast αλγόριθμο για τον έλεγχο της συνάρτησης κατώτατων ορίων. Υποθέτει ότι όλοι οι κόμβοι μοιράζονται μια περιοχή ραδιοφωνικής μετάδοσης, δηλαδή κάθε μήνυμα που αποστέλλεται από έναν από τους κόμβους παραλαμβάνεται από όλους τους υπόλοιπους κόμβους. Ο coordinator based αλγόριθμος προσπαθεί να μειώσει τα έξοδα επικοινωνίας, στέλνοντας μήνυμα μόνο σε όσους επηρεάζονται από τη τοπική παραβίαση και όχι σε όλους. Ο αλγόριθμος που είναι βασισμένος στο συντονιστή, ορίζει έναν από τους κόμβους ως συντονιστή και υποθέτει ότι όλοι οι κόμβοι επικοινωνούν απλώς με το συντονιστή. Επίσης στην εργασία, αναπτύσσονται και αποδεικνύονται οι βασικές έννοιες που αφορούν στη γεωμετρική προσέγγιση, όπως το θεώρημα που αποδεικνύει ότι το convex hull που σχηματίζεται με άκρα τα drift vector των κόμβων και εμπεριέχει το global vector είναι υποσύνολο της ένωσης των σφαιρών ελέγχου, που έχουν διάμετρο από το estimate vector ως το drift vector του κάθε κόμβου. Επίσης αποδεικνύετε ότι οι τοπική περιορισμοί είναι ασφαλής και ακριβής. Στους περιορισμούς αυτούς αρκεί να ελέγχουμε αν η σφαίρα είναι μονοχρωματική, δηλαδή να μην υπάρχει κάποια περιοχή τιμών στη σφαίρα για την οποία η ελεγχόμενη συνάρτηση αλλάζει συμπεριφορά (π.χ. περνάει το κατώτατο όριο). Καθορίζοντας την μέγιστη και την ελάχιστη τιμή στη σφαίρα, ελέγχω αν είναι μονοχρωματική. Συνεπώς, αν όλες οι σφαίρες είναι μονοχρωματικές, το convex hull είναι εγγυημένα μονοχρωματικό και ως εκ τούτου το global statistics vector είναι το ίδιο χρώμα με το convex hull. Δεδομένου ότι το estimate vector είναι μέρος του convex hull και γνωστό σε όλους τους κόμβους, το global measurement vector και το estimate vector έχουν το ίδιο χρώμα.

Κανένας από τους αλγορίθμους που παρουσιάζονται μέσα στο [22] δεν είναι κατάλληλος για τα δίκτυα αισθητήρων, δεδομένου ότι οι θεμελιώδεις περιορισμοί που υποθέτουμε σε πλαίσια εφαρμογών ενός δικτύου αισθητήρων (είτε μιας ενιαίας μετάδοσης ή παγκόσμιων επικοινωνιών με ένα μόνο κόμβο) απαιτεί αρκετή ενέργεια. Η εργασία [23] υιοθετεί τη γεωμετρική ερμηνεία που παρουσιαζόταν στο [22], προτείνει έναν αλγόριθμο που είναι σε θέση να επιλύσει σε τοπικό επίπεδο τις παραβιάσεις και ως εκ τούτου είναι ποιο προσεκτικός στη κατανάλωση της ενέργειας. Η κύρια συμβολή της εργασίας είναι ένας αποδοτικός αλγόριθμος επικοινωνίας, για να απαντήσει σε aggregate threshold queries. Ο στόχος μιας aggregate threshold query είναι να παρακολουθεί συνεχώς το δίκτυο και να δώσει μια ειδοποίηση κάθε φορά που μια συναθροιστική αξία ξεπερνάει ένα προκαθορισμένο κατώτατο όριο. Προτείνετε ένας αλγόριθμος βασισμένος στη γεωμετρική προσέγγιση από την οποία μια αυθαίρετη aggregate threshold query μπορεί να χωριστεί σε ένα σύνολο περιορισμών στις αναγνώσεις των επιμέρους αισθητήρων. Αυτοί οι περιορισμοί χρησιμοποιούνται από τους μεμονωμένους αισθητήρες για να ελέγξουν τις αναγνώσεις τους και είναι κατασκευασμένη έτσι ώστε, όσο κανένας από τους περιορισμούς δεν παραβιάζεται εγγυημένα το aggregate δεν έχει περάσει το κατώφλι. Στην εργασία περιγράφετε η αποδοτικότερη διαδικασία ως προς την κατανάλωση ενέργειας, balance αντί του τακτικού συγχρονισμού, για να επιλύονται παραβιάσεις των περιορισμών σε ένα ενιαίο cluster αλλά και σε πολλαπλά cluster με κόμβους. Το πλεονέκτημα της διαδικασίας balance είναι ότι σε αντίθεση με το συγχρονισμό δεν απαιτείτε το σύνολο των κόμβων για την επεξεργασία και είναι επομένως πιο αποτελεσματική. Στη περίπτωση ενός ενιαίου cluster η διαδικασία balance προσπαθεί να επιλύσει μια παραβίαση περιορισμού κατασκευάζοντας ένα σύνολο από κόμβους(balancing group), συμπεριλαμβανομένου του κόμβου που έχει παραβιαστεί ο τοπικός περιορισμός. Ο στόχος είναι ο μέσος όρος των drift vector των κόμβων που

ανήκουν στην ομάδα να δημιουργούν μονοχρωματική σφαίρα(Εικόνα 6.2). Όσο αυτό δεν ισχύει μπαίνει και άλλος κόμβος στο balancing group. Εάν στο balancing group είναι όλοι οι κόμβοι και η σφαίρα εξακολουθεί να είναι μην είναι μονοχρωματική, τότε απέτυχε η διαδικασία balance και γίνεται συγχρονισμός. Εάν ένα balancing group έχει κατασκευαστεί επιτυχώς, τότε όλοι οι κόμβοι που ανήκουν στην ομάδα τροποποιούν τα slack vectors τους, έτσι ώστε τα drift vector τους να είναι ίσα με το μέσο όρο των drift vectors των κόμβων που ανήκουν στην ομάδα. Σε περίπτωση πολλαπλών cluster γίνεται πρώτα intra cluster balancing, μια προσπάθεια balance του παραβιασθέντος κόμβου, χρησιμοποιώντας τα drift vector από μέλη το ίδιου cluster. Εάν η προσπάθεια αυτή είναι ανεπιτυχής, γίνεται extra cluster balancing, μια προσπάθεια balance της παραβίασης περιορισμού χρησιμοποιώντας τα drift vector από μέλη κόμβων άλλων cluster. Τέλος, εάν και extra cluster balancing έχει αποτύχει, τότε όλοι οι κόμβοι συγχρονίζονται. Επίσης στο πλαίσιο αναφέρεται περαιτέρω τρόπος βελτιστοποιήσεις στον προτεινόμενο τρόπο λειτουργίας εισάγοντας ένα καθορισμένο περιθώριο λάθους, ε γύρω από την τιμή κατώτατων ορίων, ώστε να περιορίζεται η κατανάλωση ενέργειας όταν το global statistics vector είναι κοντά στο κατώτατο όριο (Εικόνα7).

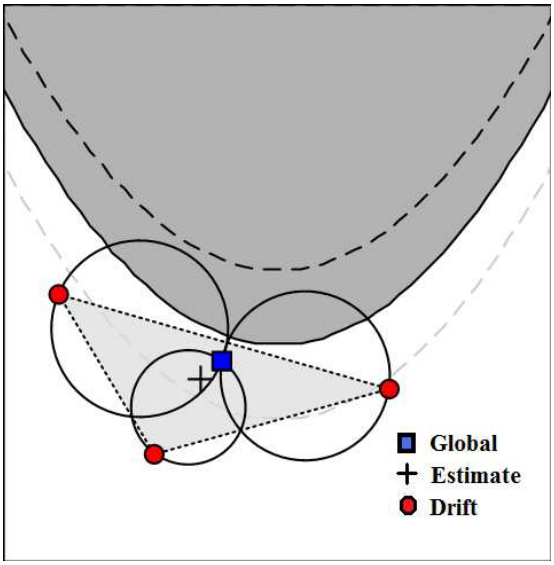


Εικόνα 6.2 : Διαδικασία balancing. Στη πρώτη απεικόνιση φαίνεται η τοπική παραβίαση της σφαίρας B(e,u_1), καθώς δεν είναι μονοχρωματική. Χρησιμοποιώντας το drift vector u_3 στο balancing group μαζί με το παραβιασμένο u_1, παρατηρώ στη δεύτερη απεικόνιση ότι ο μέσος όρος τους, δημιουργεί drift vector b το οποίο χτίζει μονοχρωματική σφαίρα B(e,b)(πράσινη). Άρα έχω επιτυχημένη διαδικασία balance. Τροποποιώ τα drift vector των u_1 και u_3 ώστε να είναι ίσα με το b, προσθέτοντας ένα διάνυσμα slack.

Στην εργασία[24], εισάγετε μια νέα μέθοδο για την παρακολούθηση κατανεμημένων δεδομένων, που ονομάζουμε shape sensitive geometric και είναι βασισμένη στη γεωμετρική προσέγγιση [22]. Γενικεύετε η προηγούμενη δουλειά στο γεωμετρικό έλεγχο και λύνονται δύο προβλήματα που επηρέαζαν αρκετά την απόδοση: σε αντίθεση με τους περιορισμούς που χρησιμοποιούνται μέχρι τώρα, που εξαρτώνται μόνο από τις τρέχουσες τιμές των τοπικών διανυσμάτων, εδώ ενσωματώνετε η χρονική συμπεριφορά τους στους περιορισμούς. Επίσης οι νέοι περιορισμοί προσαρμόζονται στις γεωμετρικές ιδιότητες της συγκεκριμένης λειτουργίας που ελέγχεται, ενώ οι προηγούμενοι περιορισμοί ήταν γενικοί. Τα πραγματικά δεδομένα εμφανίζουν συνήθως μια υποκειμενική κατανομή. Εγκαθιστάμε ένα πιθανολογικό

πρότυπο σε αυτήν την κατανομή και το χρησιμοποιούμε για να δημιουργήσουμε τους τοπικούς περιορισμούς που είναι ποιο βέλτιστοι για τα δεδομένα που λαμβάνονται για τους κόμβους, με την έννοια ότι ελαχιστοποιείται η πιθανότητα παραβίασης περιορισμού. Ένα άλλο μειονέκτημα των γεωμετρικών περιορισμών που προτεινόταν στη προηγούμενη εργασία είναι ότι ήταν γενικοί, επειδή οι ίδιοι περιορισμοί χρησιμοποιούνταν ανεξάρτητα από το αν με βάση τη λειτουργία ελέγχου θα μπορούσαν να εφαρμοστούν βελτιστοποιήσεις των περιορισμών. Παρουσιάζουμε μια νέα προσέγγιση η οποία επιτρέπει να κατασκευάσει μια σειρά από έγκυρους περιορισμούς, που είναι βέλτιστοι λαμβάνοντας υπόψη μια συγκεκριμένη συνάρτηση. Για κάθε περιορισμό ορίζουμε μια τυπική κατασκευή που ονομάζεται "ασφαλή ζώνη", η οποία επιτρέπει να μετρηθεί η αποτελεσματικότητα του περιορισμού σε σχέση με τη συγκεκριμένη λειτουργία. Αυτό το μέτρο στη συνέχεια χρησιμοποιείται για να επιλέξει έναν αποτελεσματικό περιορισμό. Στην εργασία παρουσιάζεται ένας θεωρητικά βέλτιστος περιορισμός(δηλ. όταν παραβιάζεται, οποιοσδήποτε άλλος πιθανός γεωμετρικός περιορισμός θα παραβιαστεί επίσης). Χρησιμοποιώντας αυτόν τον βέλτιστο περιορισμό αξιολογείτε, πόσο περαιτέρω βελτίωση των περιορισμών είναι δυνατόν να γίνει. Ενώ δίνεται η δυνατότητα να επιλεχτούν ελεγκτικές ζώνες με διαφορετικές μορφές [24](σφαίρα, τρίγωνα, ελλείψεις), χωρίς απώλεια γενικότητας.

Στην [21] παρουσιάζετε μια νέα προσέγγιση για την αποτελεσματική εκτέλεση γενικής κατανομής scoring ερωτημάτων κατώτατου ορίου. Παρουσιάζετε αρχικά μια λύση για τις μονότονες συναρτήσεις και εισάγετε έπειτα μια τεχνική, για να επιλυθούν άλλες λειτουργίες με την αντιπροσώπευση τους ως διαφορά μονότονων συναρτήσεων. Η προσέγγιση βασίζεται στην γνωστή από τη γεωμετρική μέθοδο διαίρεση του ερωτήματος σε τοπικούς περιορισμούς που για όσο ισχύουν, το global αποτέλεσμα δεν παραβιάζεται. Όσο τα περισσότερα αντικείμενα τους ικανοποιούν, ο όγκος επικοινωνίας μειώνεται δραστικά. Προτείνετε αρχικά ένας αλγόριθμο για τις μονότονες συναρτήσεις scoring , που ονομάζεται TB (αβέβαια συνδεδεμένος) - μονότονος. Ο αλγόριθμος αυτός συγκεντρώνει τοπικούς περιορισμούς που ελαχιστοποιούν τον όγκο της επικοινωνίας και των τοπικών παραβιάσεων. Η μονοτονία χρησιμοποιείται για την ανίχνευση αποτελεσματικών συνόλων και βασικούς περιορισμούς αντικειμένων. Εισάγεται επίσης μία αποτελεσματική τεχνική για να ελεγχθεί εάν ένα αντικείμενο παραβιάζει τους τοπικούς περιορισμούς της. Χρησιμοποιώντας TB-Monotonic ως δομική μονάδα, επεκτείνετε η τεχνική σε μια γενική μέθοδος, που ονομάζεται Tentative Bound (TB) αλγόριθμος για ερωτήματα που απασχολούν γενικά (όχι απαραιτήτως μονότονες) scoring λειτουργίες. Παρουσιάζεται μια νέα μέθοδος για τις scoring λειτουργίες ως διαφορά δύο μονοτονών λειτουργιών η οποία επιτρέπει να επιλυθούν οι μη μονότονες ερωτήσεις. Ο TB παρέχει μια αποτελεσματική μέθοδος για την εξακρίβωση, αν ένα αντικείμενο ικανοποιεί τους τοπικούς περιορισμούς, σε αντίθεση με ένα δύσκολο πρόβλημα βελτιστοποίησης που πρέπει να επιλυθεί στις [22, 24]. Επίσης μειώνει αποτελεσματικά τον αριθμό αντικειμένων που προσεγγίζονται σε κάθε κόμβο, σε αντίθεση με την πρόσβαση σε κάθε αντικείμενο σε κάθε κόμβο. Επίσης δεν περιορίζεται στον αριθμό των αντικειμένων. Ενώ το γεωμετρικό bounding θεώρημα από [22] είναι το μόνο συστατικό της προηγούμενης δουλειάς που χρησιμοποιείται εδώ και η εφαρμογή της περιορίζεται σε μια φάση του TB-Monotonic αλγορίθμου.



Εικόνα 7 : Εισαγωγή περιθωρίου λάθος ε γύρο από το κατώτατο όριο. Ενώ το estimate vector $\vec{e}$ και το global statistics vector $\vec{v}$ έχουν το ίδιο χρώμα, έχουμε παραβίαση τοπικών περιορισμών. Οι διακεκομμένες γραμμές δείχνουν, τη περιοχή που καλύπτει το κατώτατο όριο προσθέτοντας ή αφαιρώντας το περιθώριο λάθους *ε*. Με τον τρόπο αυτό ικανοποιούνται οι τοπικοί περιορισμοί.

IV. ΟΡΓΑΝΩΣΗ ΠΡΟΒΛΗΜΑΤΟΣ

Σε αυτή την διπλωματική εργασία ενδιαφερόμαστε για την οργάνωση του εξής γενικού προβλήματος. Ένας σταθμός βάσης, ελέγχει την ομοιότητα ανά ζεύγη των W-διάστασης διανυσμάτων μέτρησης που συλλέγονται από τους κόμβους αισθητήρων. Αν και οι τεχνικές μας δεν περιορίζονται στην πηγή αυτών των W τιμών, δύο πιθανές εναλλακτικές λύσεις είναι οι ακόλουθες: (i) το διάνυσμα περιέχει τις W πιο πρόσφατες μετρήσεις, που αφορούν την ίδια ποσότητα. Δηλαδή μονοδιάστατα χαρακτηριστικά ελέγχου (univariate attributes), που προέρχονται από έναν αισθητήρα του κόμβου σε τακτά χρονικά διαστήματα ή (ii) το διάνυσμα περιέχει την τρέχουσα ανάγνωση W διαφορετικών ποσοτήτων που ο αισθητήρας ελέγχει. Πολυδιάστατα χαρακτηριστικά (multivariate attributes) ελέγχου που προέρχονται για παράδειγμα από διαφορετικούς αισθητήρες του ίδιου κόμβου(π.χ. θερμοκρασία, πίεση, τάση κ.α.).

Κάθε αισθητήρας S_i καλείται να συγκρίνει συνεχώς το διάνυσμα μέτρησης του με ένα υποσύνολο(το οποίο θα ονομάσουμε comparison neighbourhood (CN, γειτονία σύγκρισης του S_i)) άλλων κόμβων στο δίκτυο. Εμείς δεν κάνουμε καμία υπόθεση σχετικά με το εάν ο S_i μπορεί να επικοινωνήσει άμεσα με όλους τους κόμβους της CN_i. Επομένως, περιγράφουμε αρχικά τη γενική περίπτωση, όπου κάθε αισθητήρας μπορεί να απαιτήσει μια επικοινωνία unicast (ενδεχομένως και multi-hop) με όλους τους κόμβους της CN και στη συνέχεια θα συζητήσουμε περαιτέρω πιθανές βελτιστοποιήσεις οι οποίες μπορούν να εφαρμοστούν όταν όλοι οι κόμβοι στη CN_i βρίσκονται σε άμεση επικοινωνία με τον S_i. Το in-between σενάριο όπου ένας κόμβος S_i βρίσκεται σε άμεση επικοινωνία με μόνο ένα μέρος από τους κόμβους στη CN_i, μπορεί να αντιμετωπιστεί με τον διαχωρισμό των κόμβων της CN_i σε δύο υποσύνολα. Το πρώτο σύνολο περιέχει εκείνους τους κόμβους που βρίσκονται σε άμεση επικοινωνία με τον S_i, ενώ το δεύτερο σύνολο περιέχει τους υπόλοιπους κόμβους. Για κάθε κόμβο που ανήκει σε ένα από αυτά τα υποσύνολα, εφαρμόζονται οι τεχνικές που αναπτύσσονται για το αντίστοιχο σενάριο.

Η απαίτηση της εφαρμογής μας είναι, ο σταθμός βάσης να γνωρίζει με απόλυτη βεβαιότητα (υποθέτοντας ότι δεν υπάρχουν απώλειες μήνυμα) αν κάθε κόμβος είναι όμοιος (ή όχι) με τους υπόλοιπους κόμβους στη γειτονία σύγκρισης του. Εάν ένας κόμβος S_i εντοπίσει ότι έχει αλλάξει η ομοιότητα του με έναν κόμβο $S_j \in CN_i$ (για παράδειγμα S_i και S_j ήταν ανόμοιοι / όμοιοι μέχρι και την προηγούμενη εποχή, αλλά τώρα κρίνονται όμοιοι / ανόμοιοι) τότε (ακριβώς), ένας από αυτούς τους κόμβους πρέπει να ενημερώσει το σταθμό βάσης. Προφανώς αυτές οι ενημερώσεις από όλους τους κόμβους αισθητήρων, μπορούν να χρησιμοποιούν για τη διάδοσή τους, ένα δέντρο συνάθροισης [17].

Μια άλλη ενδιαφέρουσα εφαρμογή, περιλαμβάνει να παρακολουθείτε εάν οι μετρήσεις σε κάθε κόμβο S_i έχουν την απαραίτητη ελάχιστη υποστήριξη (minimum support) που απαιτείται. Η ελάχιστη υποστήριξη εκφράζεται ως ο αριθμός των κόμβων στη γειτονία σύγκρισης που θεωρούνται παρόμοιοι με τον S_i. Θα παρουσιάσουμε στο τμήμα VI-C ότι αυτή την εφαρμογή μπορεί επίσης να τη χειρίζεται εύκολα το πλαίσιό μας.

Η εφαρμογή μας, μπορεί να χρησιμοποιηθεί τόσο για την ανίχνευση outlier ως ένα ενδιαφέρον γεγονός καθώς και για ανίχνευση εσφαλμένου αισθητήρα. Καλύπτει δηλαδή και τις δυο ταυτότητες outlier (Event, error).

Similarity Metric	*Function based on vectors X, Y of two nodes*	*Transforming the Function Calculation into a Function over Average Quantities that each node may compute individually*	*Local Statistic Vectors X', Y' (may contain more elements than X, Y)*	*Value of function on any point Z*
$\|\mathbf{X} - \mathbf{Y}\|_\infty$	$dist(X,Y) = max_i \lvert X_i - Y_i \rvert$	$2\, max_i\left(\left\lvert \frac{X_i - Y_i}{2} \right\rvert\right)$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \end{bmatrix} \quad Y' = \begin{bmatrix} -Y_1 \\ \dots \\ -Y_W \end{bmatrix}$	$2\, max_i(\lvert Z_i \rvert)$
$\|\mathbf{X} - \mathbf{Y}\|_1$	$dist(X,Y) = \sum_{i=1}^{W} \lvert X_i - Y_i \rvert$	$2 \sum_{i=1}^{W} \left\lvert \frac{X_i - Y_i}{2} \right\rvert$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \end{bmatrix} \quad Y' = \begin{bmatrix} -Y_1 \\ \dots \\ -Y_W \end{bmatrix}$	$2 \sum_{i=1}^{W} \lvert Z_i \rvert$
$\|\mathbf{X} - \mathbf{Y}\|_2$	$dist(X,Y) = \sqrt{\sum_{i=1}^{W} (X_i - Y_i)^2}$	$\sqrt{4 \sum_{i=1}^{W} \left(\frac{X_i - Y_i}{2}\right)^2}$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \end{bmatrix} \quad Y' = \begin{bmatrix} -Y_1 \\ \dots \\ -Y_W \end{bmatrix}$	$2 \sqrt{\sum_{i=1}^{W} Z_i^2}$
$\|\mathbf{X} - \mathbf{Y}\|_k$	$dist(X,Y) = \sqrt[k]{\sum_{i=1}^{W} (\lvert X_i - Y_i \rvert)^k}$	$\sqrt[k]{2^k \sum_{i=1}^{W} \left(\left\lvert \frac{X_i - Y_i}{2} \right\rvert\right)^k}$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \end{bmatrix} \quad Y' = \begin{bmatrix} -Y_1 \\ \dots \\ -Y_W \end{bmatrix}$	$2 \sqrt[k]{\sum_{i=1}^{W} \lvert Z_i \rvert^k}$
Cosine Similarity	$cos(\theta(X,Y)) = \frac{XY}{\|X\|_2 \, \|Y\|_2}$	$\frac{2 \sum_{i=1}^{W} \left(\frac{X_i + Y_i}{2}\right)^2 - \frac{\|X\|_2^2 + \|Y\|_2^2}{2}}{2 \left(\frac{\|X\|_2 + \|Y\|_2}{2}\right)^2 - \frac{\|X\|_2^2 + \|Y\|_2^2}{2}}$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \\ \|X\|_2^2 \\ \|X\|_2 \end{bmatrix} \quad Y' = \begin{bmatrix} Y_1 \\ \dots \\ Y_W \\ \|Y\|_2^2 \\ \|Y\|_2 \end{bmatrix}$	$\frac{2 \sum_{i=1}^{W} Z_i^2 - Z_{W+1}}{2 Z_{W+2}^2 - Z_{W+1}}$
Extended Jaccard Coefficient	$J(X,Y) = \frac{XY}{\|X\|_2^2 + \|Y\|_2^2 - XY}$	$\frac{2 \sum_{i=1}^{W} \left(\frac{X_i + Y_i}{2}\right)^2 - \frac{\|X\|_2^2 + \|Y\|_2^2}{2}}{2 \frac{\|X\|_2^2 + \|Y\|_2^2}{2} - \left(2 \sum_{i=1}^{W} \left(\frac{X_i + Y_i}{2}\right)^2 - \frac{\|X\|_2^2 + \|Y\|_2^2}{2}\right)}$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \\ \|X\|_2^2 \end{bmatrix} \quad Y' = \begin{bmatrix} Y_1 \\ \dots \\ Y_W \\ \|Y\|_2^2 \end{bmatrix}$	$\frac{2 \sum_{i=1}^{W} Z_i^2 - Z_{W+1}}{2 Z_{W+1} - (2 \sum_{i=1}^{W} Z_i^2 - Z_{W+1})}$
Correlation Coefficient	$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$	$\frac{\left(\frac{2}{W} \sum_{i=1}^{W} \left(\frac{X_i + Y_i}{2}\right)^2 - \frac{1}{W} \frac{\|X\|_2^2 + \|Y\|_2^2}{2}\right) - \left(2 \left(\frac{E[X] + E[Y]}{2}\right)^2 - \frac{(E[X])^2 + (E[Y])^2}{2}\right)}{2 \left(\frac{\sigma_X + \sigma_Y}{2}\right)^2 - \frac{\sigma_X^2 + \sigma_Y^2}{2}}$	$X' = \begin{bmatrix} X_1 \\ \dots \\ X_W \\ \|X\|_2^2 \\ E[X] \\ (E[X])^2 \\ \sigma_X \\ \sigma_X^2 \end{bmatrix} \quad Y' = \begin{bmatrix} Y_1 \\ \dots \\ Y_W \\ \|Y\|_2^2 \\ E[Y] \\ (E[Y])^2 \\ \sigma_Y \\ \sigma_Y^2 \end{bmatrix}$	$\frac{\left(\frac{2}{W} \sum_{i=1}^{W} Z_i^2 - \frac{1}{W} Z_{W+1}\right) - (2 Z_{W+2}^2 - Z_{W+3})}{2 Z_{W+4}^2 - Z_{W+5}}$

Εικόνα 8 : Η έκφραση κοινών λειτουργιών ομοιότητας μεταξύ δύο κόμβων, με W - διάστασης διάνυσμα μέτρησης X και Y, χρησιμοποιώντας τη γεωμετρική προσέγγιση. Η λειτουργία πρέπει να εκφραστεί ως μια γενική λειτουργία πάνω στο μέσο όρο των local statistics vector X' και Y', όπου τα στοιχεία X' / Y' υπολογίζονται μόνο με βάση τα στοιχεία των X / Y και μπορούν να έχουν διαφορετική διάσταση από τα X / Y.

V. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΤΩΝ ΜΕΤΡΙΚΩΝ ΟΜΟΙΟΤΗΤΑΣ ΣΤΗ ΓΕΩΜΕΤΡΙΚΗ ΜΕΘΟΔΟ

Η παράγραφος αυτή, παρέχει το απαραίτητο θεωρητικό υπόβαθρο για την υλοποίηση του πλαισίου μας. Θα αποδείξουμε, πως αρκετές ενδιαφέρουσες μετρικές ομοιότητας (φαίνονται στην Εικόνα 8), μπορούν να μετασχηματιστούν με τέτοιο τρόπο, ώστε να μπορούν να χρησιμοποιηθούν με βάση τη γεωμετρική προσέγγιση. Ενώ επίσης θα καθορίσουμε με ποιο τρόπο ελέγχουμε τους τοπικούς περιορισμούς στις σφαίρες ελέγχου.

A. Έκφραση μετρικών ομοιότητας χρησιμοποιώντας τη γεωμετρική μέθοδο

Για να μπορέσουμε να χρησιμοποιήσουμε τη γεωμετρική μέθοδο στις μετρικές ομοιότητας που θέλουμε να ελέγχουμε, πρέπει να γίνουν οι απαραίτητη μετασχηματισμοί που είναι βασικές προϋποθέσεις της γεωμετρικής μεθόδου.

Συμβολισμός και σημαντικές σημειώσεις. Έστω ότι εξετάζουμε την ομοιότητα μεταξύ δυο κόμβων, με διανύσματα μέτρησης X και Y αντίστοιχα. Επειδή μερικοί από τους μετασχηματισμούς που πρέπει να εκτελέσουμε, δεν αντιμετωπίζουν το X και το Y με έναν συμμετρικό τρόπο, ορίζουμε πάντα X τον κόμβο (ανά ζεύγος δοκιμής) με το χαμηλότερο προσδιοριστικό (id), μεταξύ των δύο κόμβων. Για να χρησιμοποιήσουμε τη γεωμετρική προσέγγιση, κάθε λειτουργία ομοιότητας πρέπει να εκφράζεται ως γενική λειτουργία πάνω στο μέσο όρο των local statistics vectors X και Y. Κάποιες μετρικές ομοιότητας δεν φαίνονται να μπορούν να υπολογιστούν στο μέσο όρο των διανυσμάτων X/Y. Ο στόχος είναι να βρούμε τα μετασχηματισμένα διανύσματα X' και Y' τα οποία πρέπει να υπολογίζονται από τα διανύσματα μέτρησης X και Y, ενώ η ελεγχόμενη μετρική ομοιότητας πρέπει να εκφράζεται στο μέσο όρο των μετασχηματισμένων διανυσμάτων. Βασική ιδέα είναι :

$$\text{similarity Function}(X, Y) \xrightarrow{\textbf{μετασχηματισμός}} \text{similarity Function}\left(\frac{X' + Y'}{2}\right)$$

$$\textbf{έτσι ώστε}, \qquad \text{similarity Function}(X, Y) = \text{similarityFunction}\left(\frac{X' + Y'}{2}\right)$$

Πρέπει να δοθεί έμφαση στα ακόλουθα σημεία:

- Γενικά, τα X'/ Y' μπορεί να έχουν διαφορετική διάσταση από τα X/Y.
- Οι διαστάσεις των sub-space που κάθε κόμβος ελέγχει (δηλ., οι σφαίρες στην Εικόνα 6), αντιστοιχούν στη διάσταση X'/ Y', και όχι στη διάσταση X/Y.

- Κατά τη διάρκεια της παρακολούθησης κάθε κόμβος πρέπει να υπολογίσει την τιμή της συνάρτησης ομοιότητας πάνω σε οποιοδήποτε σημείο Z, που βρίσκεται στη ζώνη παρακολούθησης (π.χ., σφαίρα που κατασκευάζει). Οποιοδήποτε τέτοιο σημείο Z αντιπροσωπεύει μια πιθανή θέση του global statistics vector (που υπολογίζεται ως μέσος όρος των local statistics vector) και δεν πρέπει να μπερδεύεται με το πώς η λειτουργία υπολογίζεται σε καθένα $X'/Y'/X/Y$. Η τελευταία στήλη στην Εικόνα 8, δείχνει πώς να υπολογίσετε την τιμή της συνάρτησης σε οποιοδήποτε σημείο Z.

Εξετάζουμε πρώτα τις μετρικές ομοιότητας που είναι ευκολότερο να μετατραπούν κατάλληλα και στη συνέχεια θα προχωρήσουμε σε πιο σύνθετες μετρικές ομοιότητας.

Μετασχηματισμοί για L_∞, L_1, L_2 και L_k νόρμες. Μελετάμε αρχικά την απλούστερη περίπτωση της L_∞ νόρμας, που υπολογίζεται ως η μέγιστη διαφορά $\max_{i=1\ldots W}\{|X_i - Y_i|\}$. Τότε :

$$L_\infty(X,Y) = \max_{i=1\ldots W}\{|X_i - Y_i|\} = 2\max_{i=1\ldots W}\left\{\left|\frac{X_i - Y_i}{2}\right|\right\} = 2\max_{i=1\ldots W}\left\{\left|\frac{X_i + (-Y_i)}{2}\right|\right\}$$

Κατά συνέπεια, θέτοντας $X' = X$ και $Y' = -Y$, η $L_\infty(X,Y)$ μπορεί να υπολογιστεί ως μια συνάρτηση στο μέσο όρο των διανυσμάτων $\frac{1}{2}(X' + Y')$. Υπενθυμίζουμε ότι, όπως αναφέρθηκε νωρίτερα σε αυτό το τμήμα, το local statistics vector Y' αντιστοιχεί στο κόμβο του ζεύγους δοκιμής που έχει το υψηλότερο id.

Οι μετασχηματισμοί για της L_1, L_2 και L_k νόρμες είναι παρόμοιες και είναι οι ακόλουθοι:

$$L_1(X,Y) = \sum_{i=1}^{W}|X_i - Y_i| = 2\frac{1}{2}\sum_{i=1}^{W}|X_i - Y_i| = 2\sum_{i=1}^{W}\left|\frac{X_i + (-Y_i)}{2}\right|$$

$$L_2(X,Y) = \sqrt{\sum_{i=1}^{W}(X_i - Y_i)^2} = \sqrt{2^2\sum_{i=1}^{W}\left(\frac{X_i - Y_i}{2}\right)^2} = 2\sqrt{\sum_{i=1}^{W}\left(\frac{X_i + (-Y_i)}{2}\right)^2}$$

$$L_k(X,Y) = \sqrt[k]{\sum_{i=1}^{W}(X_i - Y_i)^k} = \sqrt[k]{2^k\sum_{i=1}^{W}\left(\left|\frac{X_i + (-Y_i)}{2}\right|\right)^k} = 2\sqrt[k]{\sum_{i=1}^{W}\left(\left|\frac{X_i + (-Y_i)}{2}\right|\right)^k}$$

Κατά συνέπεια όπως και στην περίπτωση L_∞, με τη χρήση $X' = X$ και $Y' = -Y$, ως local statistics vector των δυο κόμβων, οι νόρμες L_1, L_2 και L_k μπορούν να υπολογιστούν όπως φαίνεται στην Εικόνα 8.

Μετασχηματισμός για την ομοιότητα συνημίτονου (Cosine Similarity). Για να υπολογίσουμε την ομοιότητα συνημίτονου δύο διανυσμάτων, θα πρέπει αρχικά να εκφράσουμε το εσωτερικό γινόμενο (το οποίο είναι ίσο με το άθροισμα των γινομένων, των αντίστοιχων στοιχείων του διανύσματος) σε μια μορφή που αποδέχεται η γεωμετρική προσέγγιση.
Κατά συνέπεια :

$$X \cdot Y = \sum_{i=1}^{W} X_i Y_i = \frac{1}{2}\sum_{i=1}^{W} 2X_i Y_i = \frac{1}{2}\left(\sum_{i=1}^{W}(X_i + Y_i)^2 - \left(\sum_{i=1}^{W}(X_i)^2 + \sum_{i=1}^{W}(Y_i)^2\right)\right)$$
$$= 2\sum_{i=1}^{W}\left(\frac{X_i + Y_i}{2}\right)^2 - \frac{\|X\|_2^2 + \|Y\|_2^2}{2}$$

Χρησιμοποιώντας έναν ανάλογο μετασχηματισμό για την ποσότητα στο παρονομαστή της ομοιότητας συνημίτονου έχουμε:

$$\|X\|_2\|Y\|_2 = \frac{1}{2}2\|X\|_2\|Y\|_2 = \frac{1}{2}[\,(\|X\|_2 + \|Y\|_2)^2 - (\|X\|_2^2 + \|Y\|_2^2)] =$$
$$= 2\left(\frac{\|X\|_2 + \|Y\|_2}{2}\right)^2 - \frac{\|X\|_2^2 + \|Y\|_2^2}{2}$$

Οπότε η τελική μορφή του μετασχηματισμού για την ομοιότητα συνημίτονου (cosine similarity)

$$\cos\big(\theta(X,Y)\big) = \frac{X \cdot Y}{\|X\|_2\|Y\|_2}$$

μπορεί να εκφραστεί ως μια συνάρτηση στο μέσο όρο των διανυσμάτων. Έτσι ορίζοντας

$$X' = [X_1, \dots, X_W, \|X\|_2^2, \|X\|_2]^T$$

(για το κόμβο με το χαμηλότερο id στη δοκιμή σύγκρισης) και

$$Y' = [Y_1, \dots, Y_W, \|Y\|_2^2, \|Y\|_2]^T$$

(για το κόμβο με το υψηλότερο id στη δοκιμή σύγκρισης), η ομοιότητα συνημίτονου μπορεί να υπολογιστεί ως μια συνάρτηση στο μέσο όρο των local statistics vectors που έχουν οι κόμβοι και η τελική της μορφή απεικονίζεται στην Εικόνα 8. Σημειώστε ότι στην περίπτωση αυτή η διάσταση των local statistics vectors είναι W+2 και ως εκ τούτου, μεγαλύτερη από τη διάσταση του διανύσματος μέτρησης.

Μετασχηματισμός για Extended Jaccard Coefficient. Ο υπολογισμός του Extended Jaccard Coefficient : $J(X,Y) = \frac{X \cdot Y}{\|X\|_2^2 + \|Y\|_2^2 - X \cdot Y}$, περιέχει δύο ποσότητες που πρέπει να μετασχηματιστούν: (i) Το εσωτερικό γινόμενο $X \cdot Y$, το οποίο μετασχηματίζεται με τον ίδιο τρόπο όπως έχει δειχτεί στην ομοιότητα συνημίτονου και ο όρος $\|X\|_2^2 + \|Y\|_2^2$, που μπορεί να μετασχηματιστεί σε $2\frac{\|X\|_2^2 + \|Y\|_2^2}{2}$.

Όπως φαίνεται στην Εικόνα 8, ο Extended Jaccard Coefficient μπορεί να μετασχηματιστεί αρκεί να διατηρηθεί η συγκεκριμένη μορφή στα local statistics vectors :

$$X' = [X_1, \dots, X_W, \|X\|_2^2]^T$$

(για το κόμβο με το χαμηλότερο id στη δοκιμή σύγκρισης) και

$$Y' = [Y_1, \dots, Y_W, \|Y\|_2^2]^T$$

(για το κόμβο με το υψηλότερο id στη δοκιμή σύγκρισης).

Μετασχηματισμός για το συντελεστή συσχέτισης(Correlation Coefficient).

Το δυσκολότερο κομμάτι του μετασχηματισμού είναι ο υπολογισμός για το συντελεστή συσχέτισης : $\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$, υπολογίζεται με βάση τη συνδιακύμανση (covariance) των δύο διανυσμάτων μέτρησης, καθώς και την τυπική απόκλιση(standard deviations). Ο παρονομαστής μπορεί να μετασχηματιστεί βασισμένος σε τρόπο που έχουμε αποδείξει είδη:

$$\sigma_X \sigma_Y = 2\frac{1}{2}\sigma_X \sigma_Y = \frac{1}{2}[(\sigma_X + \sigma_Y)^2 - ({\sigma_X}^2 + {\sigma_Y}^2)] = 2\left(\frac{\sigma_X + \sigma_Y}{2}\right)^2 - \frac{{\sigma_X}^2 + {\sigma_Y}^2}{2}$$

Στη συνέχεια γνωρίζουμε ότι η συνδιακύμανση (covariance) $\mathrm{cov}(X,Y)$ μπορεί να υπολογιστεί ως εξής: $\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y]$. Το γινόμενο $E[X]E[Y]$ μπορεί να μετασχηματιστεί ως εξής :

$$E[X]E[Y] = 2\frac{1}{2}E[X]E[Y] = \frac{1}{2}[(E[X] + E[Y])^2 - ((E[X])^2 + (E[Y])^2)]$$
$$= 2\left(\frac{E[X] + E[Y]}{2}\right)^2 - \frac{(E[X])^2 + (E[Y])^2}{2}$$

Για τον όρο $E[XY]$ υπολογίζουμε όπως και στο εσωτερικό γινόμενο:

$$E[X,Y] = \frac{1}{W}\sum_{i=1}^{W} X_i Y_i = \frac{1}{2W}\sum_{i=1}^{W} 2X_i Y_i$$
$$= \frac{1}{2W}\left(\sum_{i=1}^{W}(X_i + Y_i)^2 - \left(\sum_{i=1}^{W}(X_i)^2 + \sum_{i=1}^{W}(Y_i)^2\right)\right)$$
$$= \frac{2}{W}\sum_{i=1}^{W}\left(\frac{X_i + Y_i}{2}\right)^2 - \frac{1}{W}\frac{\|X\|_2^2 + \|Y\|_2^2}{2}$$

Με βάση τους παραπάνω μετασχηματισμούς, όπως φαίνεται στην Εικόνα 8, ο συντελεστής συσχέτισης (Correlation Coefficient) μπορεί να μετασχηματιστεί, αρκεί να διατηρηθεί η συγκεκριμένη μορφή στα local statistics vectors.

$$X' = [X_1, \dots, X_W, \|X\|_2^2, E[X], (E[X])^2, \sigma_X, {\sigma_X}^2]^T$$

(για το κόμβο με το χαμηλότερο id στη δοκιμή σύγκρισης) και

$$Y' = [Y_1, \dots, Y_W, \|Y\|_2^2, E[Y], (E[Y])^2, \sigma_Y, {\sigma_Y}^2]^T$$

(για το κόμβο με το υψηλότερο id στη δοκιμή σύγκρισης).

B. Εκφράζοντας τους τοπικούς περιορισμούς στις σφαίρες ελέγχου

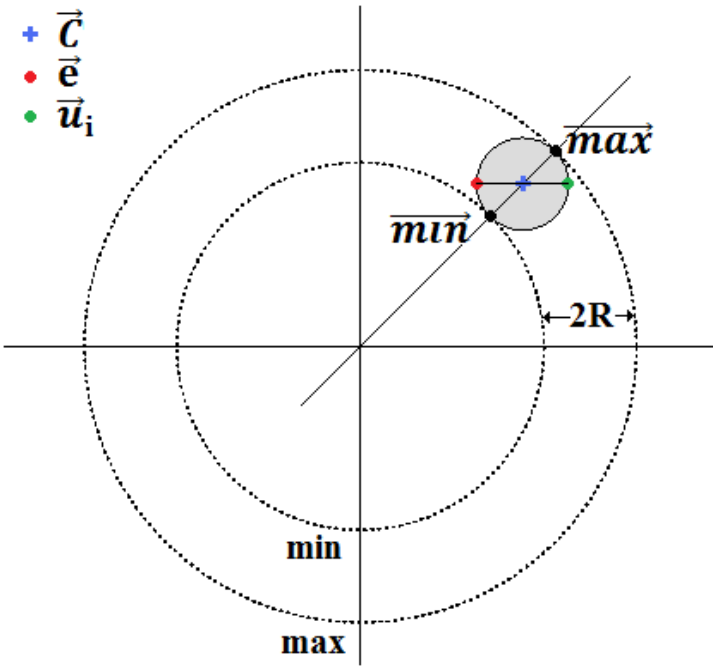
Γνωρίζουμε από τη γεωμετρική προσέγγιση ότι όσο η σφαίρα που ελέγχει κάθε κόμβος είναι μονοχρωματική, σίγουρα δεν έχει γίνει κάποια παραβίαση τοπικών περιορισμών. Στο κομμάτι αυτό θα καθορίσουμε με ποιο τρόπο υπολογίζουμε τη μέγιστη και ελάχιστη τιμή της ελεγχόμενης μετρικής ομοιότητας εντός της σφαίρας, με σκοπό να ελέγξουμε αν η σφαίρα είναι μονοχρωματική. Υπενθυμίζουμε ότι η σφαίρα ελέγχου $B(\vec{e}, \vec{u}_i)$ του κόμβου S_i είναι κατασκευασμένη με διάμετρο από το estimate vector ως το drift vector του, έχει κέντρο $\vec{C} = \frac{1}{2}(\vec{e} + \vec{u}_i)$ και ακτίνα $R = \left\|\frac{1}{2}(\vec{e} - \vec{u}_i)\right\|_2$.

Υπολογισμός σημείου μεγίστου / ελαχίστου, της σφαίρας ελέγχου για την $\|L\|_2$.
Ενδιαφερόμαστε να βρούμε το σημείο της σφαίρας ελέγχου, το οποίο αντιστοιχεί στη μέγιστη/ελάχιστη $\|L\|_2$. Τα διανύσματα που αντιστοιχούν στον L_2 χώρο είναι πάνω σε σφαίρες που έχουν κέντρο την αρχή των αξόνων. Στην Εικόνα 9 απεικονίζεται η διαδικασία (στις δυο διαστάσεις) που πρόκειται να περιγραφεί στη συνέχεια. Το διάνυσμα μεγίστου, έστω $\overrightarrow{max}$ θα είναι σίγουρα, το μοναδικό σημείο τομής της σφαίρας ελέγχου, με τη σφαίρα που έχει ακτίνα τη μέγιστη νόρμα και κέντρο την αρχή των αξόνων (0,0,…,0). Για να έχουμε ένα μοναδικό σημείο τομής και να εφάπτεται εσωτερικά η σφαίρα ελέγχου, θα πρέπει η διάκεντρος (απόσταση μεταξύ των κέντρων) να είναι ίση με την διαφορά της μέγιστης ακτίνας μείων την ακτίνα της σφαίρας ελέγχου. Αυτό ισχύει μόνο όταν το διάνυσμα $\overrightarrow{max}$, περνάει από το $\vec{C}$ της σφαίρας ελέγχου και από την αρχή των αξόνων. Όμοια για το διάνυσμα ελαχίστου $\overrightarrow{min}$ θα είναι σίγουρα, το μοναδικό σημείο τομής της σφαίρας ελέγχου, με τη σφαίρα που έχει ακτίνα την ελάχιστη νόρμα και κέντρο την αρχή των αξόνων (0,0,…,0). Για να έχουμε ένα μοναδικό σημείο τομής και να εφάπτεται εξωτερικά η σφαίρα ελέγχου, θα πρέπει η διάκεντρος να είναι ίση με το άθροισμα της ελάχιστης ακτίνας και την ακτίνα της σφαίρας ελέγχου. Αυτό ισχύει μόνο όταν το $\vec{C}$ της σφαίρας ελέγχου, περνάει από το σημείο $\overrightarrow{min}$ και από την αρχή των αξόνων. Οπότε το σημείο μεγίστου και ελαχίστου θα είναι προέκταση του διανύσματος του κέντρου της σφαίρας ελέγχου κατά την ακτίνα της, πάνω στη νοητή γραμμή που ενώνει το $\vec{C}$ και την αρχή των αξόνων. Οπότε αντίστοιχα

$$\|\min\|_2 = \|C\|_2 - \|R\|_2 \text{ και } \|\max\|_2 = \|C\|_2 + \|R\|_2$$

Οι τιμές αυτές , λόγο του μετασχηματισμού της μετρικής L_2, πρέπει να υπολογίζονται με βάση τη πέμπτη στήλη της Εικόνας 8. Οπότε προκύπτει ότι η μέγιστη και ελάχιστη $\|L\|_2$ στη σφαίρα ελέγχου θα είναι :

$$\|\min\|_2 = 2 \cdot (\|C\|_2 - \|R\|_2) \quad \text{και} \quad \|\max\|_2 = 2 \cdot (\|C\|_2 + \|R\|_2)$$



Εικόνα 9 : Υπολογισμός μέγιστης /ελάχιστης L_2 νόρμας, στο κύκλο ελέγχου. Ο κύκλος ελέγχου (γκρι) εφάπτεται εσωτερικά με το κύκλο με την μέγιστη ακτίνα max (μέγιστη $\|L\|_2$) και έχει μόνο ένα σημείο τομής το $\overrightarrow{max}$. Όμοια ο κύκλος ελέγχου (γκρι) εφάπτεται εξωτερικά με το κύκλο με την ελάχιστη ακτίνα min (ελάχιστη $\|L\|_2$) και έχει μόνο ένα σημείο τομής το $\overrightarrow{min}$. Η διαφορά της μέγιστης από την ελάχιστη $\|L\|_2$) είναι ίσο με τη διάμετρο της σφαίρας ελέγχου.
Οπότε $\|min\|_2 = \|C\|_2 - \|R\|_2$ και $\|max\|_2 = \|C\|_2 + \|R\|_2$.

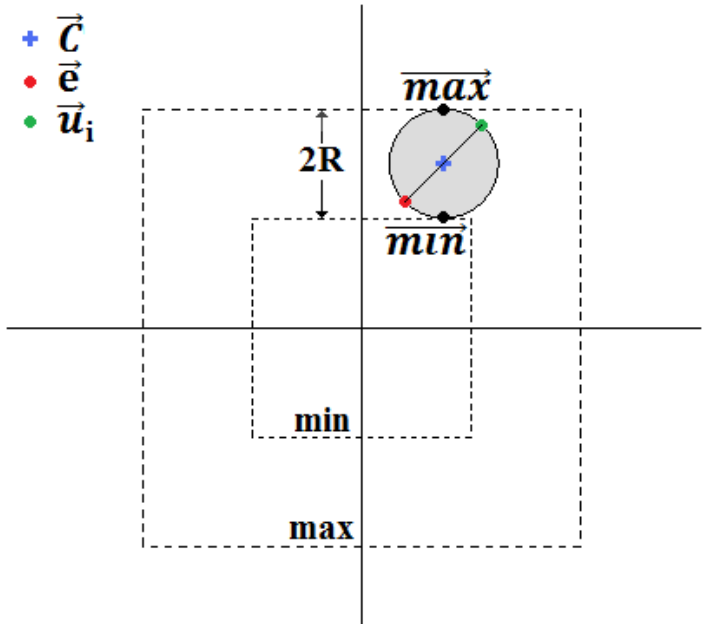
Υπολογισμός σημείου μεγίστου / ελαχίστου, της σφαίρας ελέγχου για την $\|L\|_\infty$.

Ενδιαφερόμαστε να βρούμε το σημείο της σφαίρας ελέγχου, το οποίο αντιστοιχεί στη μέγιστη/ελάχιστη $\|L\|_\infty$. Τα διανύσματα που αντιστοιχούν στον L_∞ χώρο είναι πάνω σε κύβους (υπερκύβους) που έχουν κέντρο την αρχή των αξόνων. Στην Εικόνα 10 απεικονίζεται η διαδικασία (στις δυο διαστάσεις) που πρόκειται να περιγραφεί στη συνέχεια. Η σφαίρα ελέγχου θα πρέπει να εφάπτεται σε ένα μοναδικό σημείο εσωτερικά με το τετράγωνο που αντιστοιχεί στη μέγιστη $\|L\|_\infty$ και εξωτερικά με το τετράγωνο που αντιστοιχεί στην ελάχιστη $\|L\|_\infty$. Τα αντίστοιχα σημεία είναι το $\overrightarrow{max}$ και $\overrightarrow{min}$. Η απόσταση μεταξύ αυτών των σημείων είναι όσο και η διάμετρος της σφαίρας ελέγχου. Οπότε το σημείο $\overrightarrow{max}$ δεν μπορεί να είναι άλλο από τη μέγιστη συντεταγμένη του κέντρου της σφαίρας ελέγχου $\|C\|_\infty$, αθροίζοντας επιπλέον την ακτίνα της σφαίρας, ενώ αντίστοιχα αφαιρώντας την για το σημείο $\overrightarrow{min}$.

Οπότε : $\|min\|_\infty = \max_{i=1\ldots W}|C_i| - \|R\|_2$ και $\|max\|_\infty = \max_{i=1\ldots W}|C_i| + \|R\|_2$

Οι τιμές αυτές , λόγο του μετασχηματισμού της μετρικής L_∞, πρέπει να υπολογίζονται με βάση τη πέμπτη στήλη της Εικόνας 8. Οπότε προκύπτει ότι η μέγιστη και ελάχιστη $\|L\|_\infty$ στη σφαίρα ελέγχου θα είναι :

$$\|min\|_\infty = 2 \cdot (\max_{i=1\ldots W}|C_i| - \|R\|_2) \text{ και } \|max\|_\infty = 2 \cdot (\max_{i=1\ldots W}|C_i| + \|R\|_2$$



Εικόνα 10 : Υπολογισμός μέγιστης /ελάχιστης L_∞ νόρμας, στο κύκλο ελέγχου. Ο κύκλος ελέγχου (γκρι) εφάπτεται εσωτερικά με το τετράγωνο με την μέγιστη ακμή 2max (μέγιστη $\|L\|_\infty$) και έχει μόνο ένα σημείο τομής το $\overrightarrow{max}$. Όμοια ο κύκλος ελέγχου (γκρι) εφάπτεται εξωτερικά με το τετράγωνο με την ελάχιστη ακμή 2min (ελάχιστη $\|L\|_\infty$) και έχει μόνο ένα σημείο τομής το $\overrightarrow{min}$. Η διαφορά της μέγιστης από την ελάχιστη $\|L\|_2$) είναι ίσο με τη διάμετρο της σφαίρας ελέγχου.
Οπότε $\|\min\|_\infty = \max_{i=1\ldots W}|C_i| - \|R\|_2$ και $\|\max\|_\infty = \max_{i=1\ldots W}|C_i| + \|R\|_2$

Υπολογισμός σημείου μεγίστου / ελαχίστου, της σφαίρας ελέγχου για την $\|L\|_1$.
Ενδιαφερόμαστε να βρούμε το σημείο της σφαίρας ελέγχου, το οποίο αντιστοιχεί στη μέγιστη/ελάχιστη $\|L\|_1$. Το σημεία ελαχίστου / μεγίστου πρέπει να είναι το μοναδικό σημείο τομής της σφαίρας ελέγχου με το επίπεδο της μορφής $y = \sum_{i=1}^{W} x_i + k$.

Υπολογισμός γίνετε με βάση τους συντελεστές Lagrange, υπολογίζοντας που μεγιστοποιείτε το $\sum_{i=1}^{W} x_i$, τέτοιο ώστε $\sum_{i=1}^{W}(x_i - C_i)^2 = R^2$.

$$\sum_{i=1}^{W}(x_i - C_i)^2 = R^2 \Leftrightarrow \sum_{i=1}^{W} x_i^2 - 2\sum_{i=1}^{W} x_i C_i + \sum_{i=1}^{W} C_i = R^2$$

$$F = \sum_{i=1}^{W} x_i - \lambda\left(\sum_{i=1}^{W} x_i^2 - R^2\right), \qquad (1)$$

- Για $\frac{\partial F}{\partial x_i} = 0$

$$\frac{\partial F}{\partial x_i} = W - 2W\lambda x_i = 0 \Leftrightarrow 1 - 2\lambda x_i = 0 \Leftrightarrow$$

$$x_i = \frac{1}{2\lambda}\ , \qquad (2)$$

Η σχέση (1) αντικαθιστώντας την (2) γίνετε:

$$F = \sum_{i=1}^{W} \frac{1}{2\lambda} - \lambda \left(\sum_{i=1}^{W} \left(\frac{1}{2\lambda}\right)^2 - R^2 \right) = \frac{W}{2\lambda} - \lambda \left(\frac{W}{4\lambda^2} - R^2 \right) = \frac{W}{2\lambda} - \frac{W}{4\lambda} + \lambda R^2$$
$$= \frac{W}{4\lambda} + \lambda R^2$$

- Για $\frac{\partial F}{\partial \lambda} = 0$

$$\frac{\partial F}{\partial \lambda} = \frac{W}{4\lambda^2} + R^2 = 0 \Leftrightarrow -\frac{W}{4\lambda^2} = R^2 \Leftrightarrow -\frac{W}{4R^2} = \lambda^2 \Leftrightarrow$$

$$\lambda = \frac{\sqrt{W}}{2R}\ , \qquad (3)$$

Οπότε τώρα υπολογίζω το x_i για το λ που βρήκα. Άρα η (2) αντικαθιστώντας την (3) γίνετε:

$$x_i = \frac{1}{2\lambda} = \frac{1}{2\frac{\sqrt{W}}{2R}} = \frac{R}{\sqrt{W}}$$

$$x_i = \frac{R}{\sqrt{W}}$$

Και αντίστοιχα :

$$\sum_{i=1}^{W} x_i = \frac{W\,R}{\sqrt{W}}$$

Οπότε το $\frac{R}{\sqrt{W}}$ είναι ο παράγοντας που πρέπει να αθροίζουμε ή να αφαιρούμε σε κάθε συντεταγμένη του κέντρου της σφαίρας ελέγχου, για να βρούμε το μέγιστο ή το ελάχιστο σημείο. Συγκεκριμένα ανάλογα με τις συντεταγμένες του κέντρου, γνωρίζουμε ακριβώς σε

ποιο σημείο του χώρου βρίσκεται. Αν έστω η συντεταγμένη C_i είναι στο θετικό κομμάτι του τεταρτημορίου προσθέτουμε τον παράγοντα $\frac{R}{\sqrt{W}}$, αλλιώς τον αφαιρούμε. Οπότε βρίσκουμε με τον τρόπο αυτό το σημείο που ανήκει το μέγιστο. Όμοια για το σημείο ελαχίστου, αν η συντεταγμένη C_i είναι στο θετικό κομμάτι του τεταρτημορίου αφαιρούμε τον παράγοντα $\frac{R}{\sqrt{W}}$, αλλιώς τον προσθέτουμε. Οπότε το ζεύγος σημείων θα έχει την εξής μορφή :
Όπου $\mathrm{sign}C_i$ το πρόσημο της συγκεκριμένης συντεταγμένης του κέντρου.

$$\overrightarrow{\max} = \{C_1(\mathrm{sign}C_i)\frac{R}{\sqrt{W}}, \dots, C_W(\mathrm{sign}C_W)\frac{R}{\sqrt{W}}\}$$

$$\overrightarrow{\min} = \{C_1(-\mathrm{sign}C_i)\frac{R}{\sqrt{W}}, \dots, C_W(-\mathrm{sign}C_W)\frac{R}{\sqrt{W}}\}$$

Αντίστοιχα για τον υπολογισμό της μεγίστης και ελάχιστης $\|L\|_1$ θα είναι ίσο με το άθροισμα της $\|C\|_1 \pm W$ φορές τον αντίστοιχο όρο $\frac{R}{\sqrt{W}}$ (συν για το μέγιστο και πλην για το ελάχιστο). Οπότε :

$$\min = \sum_{i=1}^{W}|C_i| - \frac{W\|R\|_2}{\sqrt{W}} \quad \text{και} \quad \max = \sum_{i=1}^{W}|C_i| + \frac{W\|R\|_2}{\sqrt{W}}$$

Λόγο του μετασχηματισμού της μετρικής L_1 , οι τιμές των διανυσμάτων αυτών, υπολογίζονται με βάση τη πέμπτη στήλη της Εικόνας 8. Οπότε προκύπτει ότι η μέγιστη και ελάχιστη $\|L\|_1$ στη σφαίρα ελέγχου θα είναι :

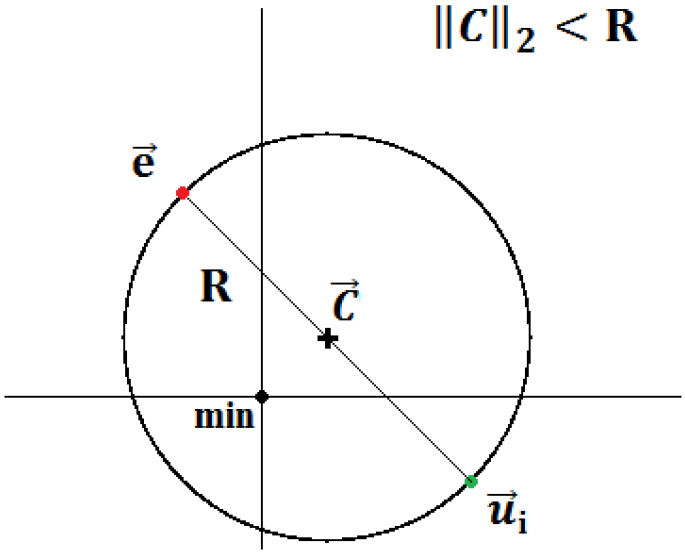
$$\min = 2 \cdot (\sum_{i=1}^{W}|C_i| - \frac{W\|R\|_2}{\sqrt{W}}) \quad \text{και} \quad \max = 2 \cdot (\sum_{i=1}^{W}|C_i| + \frac{W\|R\|_2}{\sqrt{W}})$$

Για όλες τις προαναφερόμενες περιπτώσεις θα πρέπει να κάνουμε μια διευκρίνηση σχετικά με τον υπολογισμό της ελάχιστης τιμής. Είναι σημαντικό στην σφαίρα ελέγχου, να συγκρίνουμε την απόσταση του κέντρου με την ακτίνα. Όταν η απόσταση του κέντρου από την αρχή των αξόνων είναι μικρότερη από την ακτίνα, η σφαίρα ελέγχου εμπεριέχει την αρχή των αξόνων και στη περίπτωση αυτή η τιμή ελαχίστου είναι μηδέν(Εικόνα 11). Συνοψίζοντας αν

$\|C\|_2 < R$ τότε min = 0 αλλιώς min υπολογίζεται με τον τρόπο που περιγράφεται σε κάθε μετρική ομοιότητας.

Οπότε ισχύει:

min = max {0, min(από αντίστοιχη μετρική ομοιότητας)}



Εικόνα 11 : Υπολογισμός ελάχιστης τιμής όταν $\|C\|_2 < R$. Όποτε η απόσταση C από την αρχή των αξόνων είναι μικρότερη από την R, η σφαίρα ελέγχου εμπεριέχει την αρχή των αξόνων και τότε min=0.

Για κάθε μια από τις L_∞, L_1 και L_2 μετρικές ομοιότητας, μπορέσαμε να καθορίσουμε με ακρίβεια τα σημεία μεγίστου και ελαχίστου στη σφαίρα ελέγχου, είτε γεωμετρικά είτε με Lagrange. Στο Πίνακα ΙΙ συνοψίζονται τα σημεία μεγίστου ελαχίστου, για αυτές τις μετρικές ομοιότητας. Όπου C το κέντρο της σφαίρας, R η ακτίνα της και W η διάσταση του διανύσματος.

Similarity Metric	Min value	Max value
L_∞	$2 \cdot (\max\{0, \max_{i=1\ldots W} \lvert C_i \rvert - \|R\|_2\})$	$2 \cdot (\max_{i=1\ldots W} \lvert C_i \rvert + \|R\|_2)$
L_1	$2 \cdot (\max\{0, \sum_{i=1}^{W} \lvert C_i \rvert - \frac{W\|R\|_2}{\sqrt{W}}\})$	$2 \cdot (\sum_{i=1}^{W} \lvert C_i \rvert + \frac{W\|R\|_2}{\sqrt{W}})$
L_2	$2 \cdot (\max\{0, \|C\|_2 - \|R\|_2\})$	$2 \cdot (\|C\|_2 + \|R\|_2)$

Πίνακας ΙΙ : Τα σημεία μεγίστου ελαχίστου για τις L_∞, L_1 και L_2

Υπολογισμός σημείου μεγίστου / ελαχίστου, της σφαίρας ελέγχου για τις υπόλοιπες μετρικές ομοιότητας, που έχουν επιπλέον τις παραγόμενες διαστάσεις (derived) .

Για τις υπόλοιπες μετρικές είναι δύσκολο για έναν κόμβο αισθητήρα να βρει με ακρίβεια, τη μέγιστη / ελάχιστη τιμή της μετρικής ομοιότητας εντός της σφαίρας ελέγχου που κατασκευάζει. Μια εφικτή εναλλακτική λύση είναι να εξετάζει την τιμή μέσα από ένα σύνολο σημείων που ανήκουν σε πλέγμα (το οποίο βρίσκεται εντός της σφαίρας). Τα σημεία αυτά έστω $\overrightarrow{\text{point}}$, κατασκευάζονται με βάση το κέντρο $\vec{C}$ και την ακτίνα R της σφαίρας ελέγχου. Το εύρος τιμών των συντεταγμένων του σημείου αυτού είναι από $\|C\|_2 - \|R\|_2$ ως $\|C\|_2 + \|R\|_2$. Αφού ολοκληρωθεί η διαδικασία, θα έχει υπολογίσει την μέγιστη και ελάχιστη τιμή των σημείων στο πλέγμα και κατά προσέγγιση τις τιμές min και max στη σφαίρα ελέγχου. Υπολογίζουμε την τιμή της μετρικής ομοιότητας στα σημεία αυτά, χρησιμοποιώντας την πέμπτη στήλη της Εικόνας 8. Η λύση αυτή όμως δεν μπορεί να

οδηγήσει σε απόλυτα ακριβή αποτελέσματα. Ένα δεύτερο πρόβλημα είναι ότι σε αυτές τις μετρικές ομοιότητας το ελεγχόμενο διάνυσμα τιμών έχει μεγαλύτερη διάσταση. Για τις πρώτες d-διαστάσεις (όσο και το local statistics vector)παράγουμε με τον προαναφερόμενο τρόπο το πιθανό σημείο $\overrightarrow{\text{point}}$. Ο κόμβος που θα κάνει την αναζήτηση της μέγιστης και ελάχιστης τιμής μπορεί με βάση το δικό του διάνυσμα μετρήσεων X και το σημείο $\overrightarrow{\text{point}}$ (πιθανό estimate vector), να υπολογίσει ποιο είναι το διάνυσμα Y ώστε να ισχύει $\text{point} = \frac{X+Y}{2}$:

$$\text{point} = \frac{X+Y}{2} \Leftrightarrow X + Y = 2\text{point}$$

Άρα, $Y = 2\text{point} - X$

Στη συνέχεια για τα διανύσματα X και Y, κατασκευάζω τα επεκταμένα διανύσματα με τις επιπλέον διαστάσεις τους. Ακολούθως γεμίζω τις επιπλέον διαστάσεις του $\overrightarrow{\text{point}}$ με βάση τα επεκταμένα διανύσματα X,Y και συγκεκριμένα στις επιπλέον θέσεις βάζω τα $\frac{X_i+Y_i}{2}$ των αντίστοιχων θέσεων των επεκταμένων X,Y. Επίσης ελέγχουμε για κάθε παραγόμενο σημείο $\overrightarrow{\text{point}}$, αν έχει απόσταση που είναι μεγαλύτερη από την ακτίνα της σφαίρας ελέγχου. Τέτοια σημεία τα απορρίπτουμε καθώς αντιστοιχούν σε σημεία εκτός σφαίρας.

VI. ΛΕΙΤΟΥΡΓΙΑ ΚΟΜΒΩΝ ΓΙΑ ΕΛΕΓΧΟ ΟΜΟΙΟΤΗΤΑΣ

Προτείνουμε τώρα διαφορετικές εναλλακτικές λύσεις για τη λειτουργία του κόμβου, έτσι ώστε κάθε κόμβος να μπορεί να παρακολουθεί με ακρίβεια την ομοιότητα του με οποιονδήποτε άλλο κόμβο στη γειτονιά σύγκρισης του. Για την ευκολία της παρουσίασης, αρχίζουμε την συζήτηση μας, με ένα πρότυπο μοντέλο δυο κόμβων που θέλουν να συγκρίνουν την ομοιότητα τους. Το μοντέλο αυτό από τη μια πλευρά δίνει την δυνατότητα να εξηγήσουμε τις διάφορες εναλλακτικές λύσεις και από την άλλη πλευρά, μπορεί να χρησιμοποιηθεί επίσης για να απαντάει σε ερωτήματα ομοιότητας :

(i) Μεταξύ των κόμβων που δεν επικοινωνούν άμεσα μεταξύ τους(Not Direct Communication). Στη περίπτωση αυτή κάθε αισθητήρας απαιτεί μια unicast/multi hop επικοινωνία με όλους τους κόμβους στη γειτονιά σύγκρισης του. Οπότε η αποστολή πληροφοριών γίνεται από έναν αποστολέα σε ένα και μόνο παραλήπτη.

(ii) Μεταξύ των κόμβων που επικοινωνούν άμεσα με όλους τους κόμβους στη γειτονιά σύγκρισης του (Direct, Broadcast) . Στη περίπτωση αυτή κάθε αισθητήρας απαιτεί μια broadcast επικοινωνία με όλους τους κόμβους στη γειτονιά σύγκρισης του. Οπότε η αποστολή πληροφοριών γίνεται από έναν αποστολέα σε όλους τους παραλήπτες.

Επιπρόσθετα εφαρμόζοντας σε όλους τους γείτονες ενός κόμβου τους ίδιους ανά ζεύγη αλγορίθμους που θα παρουσιάσουμε στη συνέχεια, μπορούμε να λύσουμε το πρόβλημα παρακολούθησης που αντιμετωπίζουμε σε αυτό το έγγραφο. Κάθε κόμβος, πρέπει απλά να διατηρήσει στατιστικά στοιχεία για κάθε κόμβο στη γειτονιά σύγκρισης του και να ακολουθεί για κάθε κόμβο την διαδικασία που περιγράφεται στο μοντέλο επικοινωνίας δυο κόμβων(VI-A). Τα απαραίτητα στατιστικά στοιχεία που κρατούνται από έναν κόμβο S_i, ανάλογα με το μοντέλο λειτουργίας του, περιγράφονται στο Πίνακα III. Τα στοιχεία αυτά είναι κυρίως πίνακες διαστάσεως όσο το πλήθος στη γειτονιά σύγκρισης του. Ενώ στο Πίνακα IV μπορούμε να δούμε τα περιεχόμενα του μηνύματος.

A. Μοντέλο επικοινωνίας δύο κόμβων

Στο πρότυπο μοντέλο μας, δεν μας απασχολεί αν οι δυο κόμβοι είναι σε άμεση επικοινωνία μεταξύ τους. Έτσι, όταν αναφερόμαστε σε ένα κόμβο S_i που μεταδίδει ένα μήνυμα στο κόμβο S_j, το μήνυμα αυτό μπορεί να περιλαμβάνει μια multi-hop επικοινωνία. Φυσικά είναι πιο λογικό να περιμένει κανείς ότι ο έλεγχος ομοιότητας μεταξύ των αναγνώσεων δύο κόμβων είναι ποιο χρήσιμος όταν οι αισθητήρες μπορούν να τοποθετηθούν κοντά, έτσι ώστε να επικοινωνούν απευθείας(εν αναμονή οποιουδήποτε εμποδίου που μπορεί να εμποδίσει την επικοινωνία τους) ή να είναι εφικτή η επικοινωνία με λίγα hop.

Όλοι εκτός από έναν τρόπο λειτουργίας (τον τελευταίο) που προτείνουμε στο παρόν έγγραφο, δεν απαιτούν μια προκαθορισμένη σειρά εκτέλεσης (μεταξύ των κόμβων των οποίων η ομοιότητα παρακολουθείται), σύμφωνα με την οποία οι κόμβοι αισθητήρων θα εξετάζουν κατά πόσον υπάρχει μια τοπική παραβίαση. Ωστόσο για την ευκολία της παρουσίασης, ας υποθέσουμε απλά ότι έχει καθοριστεί μια τέτοια προκαθορισμένη σειρά εκτέλεσης, η οποία δεν χρειάζεται να είναι ίδια σε όλες τις εποχές (δηλαδή, όπως στην περίπτωση κατά την οποία οι κόμβοι εναλλάσσουν τη σειρά τους).

Πρέπει να σημειώσουμε ότι μια τοπική παραβίαση (local violation) εμφανίζεται όταν ένας κόμβος κρίνει ότι ο τοπικός περιορισμός παραβιάζεται. Δηλαδή στη περίπτωση που η σφαίρα ελέγχου $B(\vec{e}, \vec{u}_i)$, που παρακολουθεί δεν είναι μονοχρωματική. Ο κόμβος που ανιχνεύει μια τοπική παραβίαση στέλνει το διάνυσμα μετρήσεών του, στον κόμβο για τον οποίο απέτυχε ο τοπικός έλεγχος ομοιότητας , με σκοπό να γίνει έλεγχος ολικής παραβίασης. Μια ολική παραβίαση (global violation) εμφανίζεται όταν ένας κόμβος είναι σίγουρος ότι η ομοιότητα του με έναν άλλο κόμβο έχει αλλάξει χρώμα. Δηλαδή η ομοιότητα άλλαξε συμπεριφορά σε σχέση με την προηγούμενη κατάσταση που ήταν (Το διάνυσμα εκτίμησης $\vec{e}$ ήταν πάνω/κάτω από το κατώτατο όριο και τώρα είναι κάτω/πάνω από το όριο). Ο κόμβος που ανιχνεύει μια ολική παραβίαση πρέπει να ενημερώσει το σταθμό βάσης. Στην περιγραφή που ακολουθεί όταν ο τοπικός περιορισμός παραβιάζεται, θα λέμε ότι ο έλεγχος τοπικού περιορισμού ήταν "ανεπιτυχής" ενώ όταν δεν παραβιάζεται "επιτυχής".

Simple τρόπος λειτουργίας. Στην Εικόνα 12 απεικονίζεται ο Algorithm 1, που παρουσιάζει τον Simple τρόπο λειτουργίας κάθε κόμβου. Αυτός ο τρόπος λειτουργίας δεν έχει καμία βελτιστοποίηση στις αποφάσεις του και κατά συνέπεια παρουσιάζει μόνο τη πρότυπη βασική λύση, για την σύγκριση της ομοιότητας μεταξύ των δύο κόμβων S_i και S_j . Μπορεί κανείς να θεωρήσει αυτόν τον Simple τρόπο λειτουργίας ως τον πιο φυσικό τρόπο εφαρμογής της γεωμετρικής προσέγγισης στο πρόβλημά μας. Ωστόσο, ενώ στην προηγούμενη εργασία συνήθως υπήρχε ένας κόμβος συντονιστής ως βελτιστοποίηση στον Simple τρόπο λειτουργίας, κάθε κόμβος σε ένα ζευγάρι ελέγχου μπορεί να ενεργήσει ως συντονιστής με σκοπό να μειώσει τις μεταδόσεις μηνυμάτων.

Σε κάθε εποχή και οι δυο κόμβοι ενημερώνουν τα διανύσματα μετρήσεων τους (Γραμμή 5), και υπολογίζουν τα delta ($\overrightarrow{\Delta v}$) και τα drift vectors τους (γραμμές 6-7) .

Υπάρχουν τρεις περιπτώσεις στις οποίες ο κόμβος S_i μπορεί να λάβει ένα μήνυμα από το κόμβο S_j :

- Γραμμές 8 - 17 : Εάν ο κόμβος S_i ενεργεί δεύτερος και λαμβάνει ένα μήνυμα από τον S_j , σημαίνει ότι ο έλεγχος τοπικού περιορισμού στον S_i ήταν επιτυχής αλλά έγινε τοπική παραβίαση στο κόμβο S_j . Στη περίπτωση αυτή ο κόμβος S_i ελέγχει αν έχει πραγματοποιηθεί μια ολική παραβίαση. Για να γίνει αυτό δεν είναι απαραίτητο να κατασκευαστεί οποιαδήποτε σφαίρα χρησιμοποιώντας τα drift vector των κόμβων. Αρκεί να αναφέρουμε ότι ο κόμβος S_i γνωρίζει το διάνυσμα μέτρησης του κόμβου S_j (από το μήνυμα που έλαβε) καθώς επίσης γνωρίζει και το δικό του διάνυσμα μετρήσεων. Οπότε ο S_i υπολογίζει ακριβός την τιμή του global statistics vector $\vec{e'} = v_j + v_i/2$ (Στον Αλγόριθμος 2, σειρά 5 που απεικονίζεται στη Εικόνα 13) και μπορεί απλά να ελέγξει αν η τιμή της ελεγχόμενης μετρικής ομοιότητας στο estimate vector $\vec{e}$ και $\vec{e'}$ έχουν διαφορετικό χρώμα. Αν είναι διχρωματικά (γραμμές 10-12), τότε μια παγκόσμια παραβίαση έχει συμβεί και ο κόμβο S_i θα ειδοποιήσει το σταθμός βάσης. Παράλληλα ο κόμβος S_i θα ενημερώσει το διάνυσμα εκτίμησης $\vec{e} = \vec{e'}$ και θα στείλει το διάνυσμα μετρήσεων του στον κόμβο S_j .
- Γραμμές 20-26: Εάν ο κόμβος S_i ενεργήσει πρώτος, σημαίνει ότι δεν έχει γίνει έλεγχος τοπικού περιορισμού στον S_j . Αν πάλι ο κόμβος S_i ενεργεί δεύτερος αλλά δεν λαμβάνει μήνυμα από τον κόμβο S_j σημαίνει ότι δεν υπάρχει τοπική παραβίαση περιορισμού στον κόμβο S_j . Στις περιπτώσεις αυτές ο κόμβος S_i κάνει τον έλεγχο τοπικού περιορισμού για να γνωρίζει αν υπάρχει τοπική παραβίαση (χρησιμοποιώντας τον Αλγόριθμο 3 που απεικονίζεται στην Εικόνα 14). Αν υπάρχει παραβίαση ο κόμβος S_i θα μεταδώσει το διάνυσμα μετρήσεων του στον S_j (με σκοπό να γίνει έλεγχος ολικής παραβίασης) και θα περιμένει να λάβει απάντηση. Η απάντηση θα περιέχει το διάνυσμα μετρήσεων του S_j , προκειμένου ο S_i να ενημερώσει το διάνυσμα εκτίμησης του $\vec{e} = \vec{e'}$ ενώ αν υπάρχει ολική παραβίαση ο σταθμός βάσης ενημερώνεται από τον S_j .
- Γραμμές 27-31: Εάν ο κόμβος S_i ενεργεί πρώτος και ο S_i δεν εντοπίσει κάποια παραβίαση τοπικού περιορισμού, υπάρχει περίπτωση να λάβει ένα μήνυμα από τον κόμβο S_j , Στη περίπτωση αυτή ο κόμβος S_j είχε μια τοπική παραβίαση περιορισμού και στέλνει μήνυμα με σκοπό να γίνει έλεγχος ολικής παραβίασης. Αυτή η περίπτωση είναι πανομοιότυπη με την πρώτη που περιγράψαμε.

Algorithm 1 Node i: Operation under Simple Mode

Require: Threshold T, Similarity Function F

1: Maintain $\vec{v'}_i$: last transmitted local statistics vector
2: Maintain $\vec{e}$: Estimate vector
3: Maintain $\vec{v'}_j$: last received local statistics vector from S_j
4: **while** Asked to Monitor Similarity to S_j **do**
5: Obtain new measurements and form local statistics vector $\vec{v}_i$
6: Compute delta vector $\Delta\vec{v}_i = \vec{v}_i - \vec{v'}_i$
7: Compute drift vector $\vec{u}_i = \vec{e} + \Delta\vec{v}_i$
8: **if** Acting Second in Pair **then**
9: **if** MessageWait(S_j, $\vec{v'}_j$, $\vec{e'}$) == true **then**
10: **if** $\vec{e'}$ and $\vec{e}$ are bi-chromatic **then**
11: Notify base station about global violation
12: **end if**
13: Send local measurements vector to S_j
14: $\vec{v'}_i = \vec{v}_i$, $\vec{e} = \vec{e'}$
15: Continue
16: **end if**
17: **end if**
18: {Acting first, or acting second but did not receive a message}
19: localViolation = checkIfViolation($\vec{e}$, $\vec{u}_i$, F, T)
20: **if** localViolation == true **then**
21: Send local measurements vector to S_j
22: $\vec{v'}_i = \vec{v}_i$
23: MessageWait(S_j, $\vec{v'}_j$, $\vec{e'}$) {Will definitely arrive}
24: Compute $\vec{e} = \frac{1}{2}(\vec{v'}_j + \vec{v}_i)$
25: Continue {If global violation, S_j will send notification}
26: **end if**
27: **if** Acting first **then**
28: **if** MessageWait(S_j, $\vec{v'}_j$, $\vec{e'}$) == true **then**
29: goto 10
30: **end if**
31: **end if**
32: **end while**

Εικόνα 12 : Ο Algorithm 1 που παρουσιάζει τον Simple τρόπο λειτουργίας κάθε κόμβου.

Algorithm 2 Node i: MessageWait Subroutine

Require: Paired node S_j, last received local statistics vector $\vec{v'}_j$, vector $\vec{e'}$

Wait for message from S_j
if Message Arrived, containing vector of measurements **then**
    Compute local statistics vector $\vec{v}_j$ of S_j
    $\vec{v'}_j = \vec{v}_j$
    Compute $\vec{e'} = \frac{1}{2}(\vec{v'}_j + \vec{v}_i)$
    RETURN true
end if
RETURN false

Εικόνα 13 : Ο Algorithm 2 για την διαχείριση εισερχομένων πιθανόν μηνυμάτων. Ο αλγόριθμος αυτός στον Simple τρόπο λειτουργίας κάθε κόμβου, ενημερώνει τον κόμβο που τον κάλεσε αν υπάρχει κάποιο λαμβανόμενο μήνυμα. Ενώ αν υπάρχει, υπολογίζει και το global statistics vector $\vec{e'}$.

Algorithm 3 Node i: checkIfViolation Subroutine

Require: estimate vector $\vec{e}$, drift vector $\vec{u}_i$, function F, threshold T

if $F(\vec{e}) > T$ **then**
    Find min value testVal in ball B($\vec{e}$, $\vec{u}_i$)
else
    Find max value testVal in ball B($\vec{e}$, $\vec{u}_i$)
end if
if $F(\vec{e})$ and testVal are monochromatic **then**
    Return false
end if
Return true

Εικόνα 14 : Ο Algorithm 3 για την ανίχνευση παραβίασης τοπικού περιορισμού. Ο αλγόριθμος αυτός στον Simple τρόπο λειτουργίας κάθε κόμβου, ελέγχει αν υπάρχει κάποια παραβίαση τοπικού περιορισμού, ελέγχοντας αν η σφαίρα είναι διχρωματική.

Η λειτουργία με Autobalance. Στον Simple τρόπο λειτουργίας είτε και οι δύο κόμβοι δεν θα ανιχνεύσουν κάποια τοπική παραβίαση και θα παραμείνουν αδρανείς ή και οι δυο κόμβοι θα μεταδώσουν τα διανύσματα μετρήσεων τους στον άλλο κόμβο, εάν ένας από τους δύο παρατηρήσει μια τοπική παραβίαση. Ο κόμβος που εντόπισε τη τοπική παραβίαση θα στείλει ένα μήνυμα ενώ θα περιμένει να λάβει απάντηση είτε ο άλλος κόμβος παρατηρήσει ολική παραβίαση είτε όχι. Οπότε αναγκάστηκα στέλνουν και οι δυο κόμβοι το αντίστοιχο διάνυσμα μετρήσεων, ακόμα και αν δεν παρατηρηθεί ολική παραβίαση. Αυτό είναι ένα σημαντικό μειονέκτημα του Simple τρόπου λειτουργίας, που στόχος μας είναι να βελτιώσουμε με τη λειτουργία Autobalance.

Προκειμένου να εξασφαλιστεί ακριβής έλεγχος, η γεωμετρική προσέγγιση απαιτεί ότι (i) οι δύο κόμβοι χρησιμοποιούν πάντα το ίδιο διάνυσμα εκτίμησης(estimate vector), και (ii) ο μέσος όρος των drift vectors είναι ίσος με το global statistics vector (δηλαδή, η πραγματική μέση τιμή των local statistics vectors). Το μοντέλο λειτουργίας Autobalance διαφέρει με το Simple στις επιπλέον ενέργειες που κάνει ο κόμβος που εντοπίζει πρώτος μια τοπική παραβίαση, αλλά και στους επιπλέον ελέγχους που κάνει ο κόμβος που λαμβάνει ένα μήνυμα για μια τέτοια παραβίαση.

Στην Autobalance λειτουργία, ο πρώτος κόμβος S_i (στο ζεύγος) που εντοπίζει μια τοπική παραβίαση κάνοντας ελέγχω ομοιότητας με τον κόμβο S_j :

- Τροποποιεί το estimate vector σε $\vec{e} + \frac{1}{2}\overrightarrow{\Delta v_i}$
- Μεταδίδει το διάνυσμα μετρήσεων του(μετά τη μετάδοση $\overrightarrow{\Delta v_i} = 0$).
- Περιμένει για ένα πιθανό μήνυμα από τον άλλο κόμβο S_j . Εάν παραλάβει το μήνυμα, ενημερώνει το διάνυσμα εκτίμησης του με την ενσωμάτωση μισού του delta vector $\frac{1}{2}\overrightarrow{\Delta v_j}$ που υπολογίζει για τον αντίστοιχο κόμβο, άρα το διάνυσμα εκτίμησης θα είναι ίσο με το πραγματικό μέσο όρο των μετρήσεων τους.

Στον προαναφερόμενο Αλγόριθμο 1 (Εικόνα 12), πρέπει πριν τη γραμμή 22 να υπολογίζει προσωρινά το $\vec{e} = \vec{e} + \frac{1}{2}\overrightarrow{\Delta v_i}$. Ενώ στη γραμμή 23 πλέον δεν είναι υποχρεωτικό να περιμένει απαντήσει από τον άλλο κόμβο, αλλά σε περίπτωση που λάβει θα πρέπει στη γραμμή 24 να ενημερώνει το διάνυσμα εκτίμησης με $\vec{e'} = \vec{e} + \frac{1}{2}\overrightarrow{\Delta v_j}$.

Από την άλλη πλευρά, ο κόμβος S_i που λαμβάνει πρώτα το μήνυμα από τον S_j , χρησιμοποιεί τα παρακάτω βήματα:

- Ελέγχει εάν η τιμή της μετρικής στο νέο ολικό διάνυσμα εκτίμησης $\vec{e'}$ (global statistics vector) έχει το ίδιο χρώμα με τη τιμή της μετρικής στο διάνυσμα εκτίμησης $\vec{e}$ που είχε υπολογιστεί πριν για τους κόμβους S_i και S_j .
- Αποθήκευση (σημείωση: η ενημέρωση πρέπει να έχει ολοκληρωθεί πριν τον επόμενο έλεγχο) ,του διανύσματος εκτίμησης με την ενσωμάτωση μισό του delta vector $\overrightarrow{\Delta v_j}$ που υπολογίζει για τον αντίστοιχο κόμβο.
- Αν το πρώτο βήμα σηματοδοτήσει ολική παραβίαση, ο κόμβος S_i ειδοποιεί το σταθμό βάσης, θέτει το διάνυσμα εκτίμησης ίσο με το global statistics vector που υπολόγισε στο πρώτο βήμα και μεταδίδει το διάνυσμα μέτρησης του στον

αντιστοίχιση κόμβο S_j . Διαφορετικά συνεχίζει χρησιμοποιώντας ως διάνυσμα εκτίμησης αυτό που υπολόγισε στο δεύτερο βήμα.

Στον προαναφερόμενο Αλγόριθμο 2(Εικόνα 13), πρέπει επιπλέον να υπολογίζει και να μας επιστρέφει το προσωρινά αποθηκευμένο διάνυσμα εκτίμησης (βήμα 2) έστω το συμβολίζω $\overrightarrow{e''}$. Ενώ στον Αλγόριθμο1 πριν τη γραμμή 13 θα πρέπει επιπλέον να ελέγχω αν $\overrightarrow{e''}, \overrightarrow{e'}$ και $\vec{e}$ έχουν διαφορετικό χρώμα. Σε περίπτωση που αυτό ισχύει ακολουθώ τα βήματα των γραμμών 13-14, αλλιώς αν δεν έχουν διαφορετικό χρώμα δεν στέλνεται μήνυμα και $\vec{e} = \overrightarrow{e''}$.

Η παραπάνω διαδικασία είναι παρόμοια σε γενικές γραμμές με τις τεχνικές που αναπτύσσονται στο [22], όπου ένα slack vector μπορεί να προστεθεί σε όλα τα υπολογιζόμενα drift vector, με την προϋπόθεση ότι το slack vector εξουδετερώνεται. Ο σκοπός τις διαδικασίας είναι ο μέσος όρος των drift vector να σχηματίζουν μονοχρωματική σφαίρα (απεικονίζεται στην Εικόνα 6.2). Είναι απλό να επαληθευθεί ότι το μοντέλο λειτουργίας Autobalance, μοιράζει το delta vector του κόμβου με την τοπική παραβίαση και στους δύο κόμβους. Αφού και οι δύο κόμβοι τροποποιούν το διάνυσμα εκτίμησης τους και τα drift vector τους κατά το ήμισυ του delta vector, ενώ ο κόμβος που μεταδίδει μηδενίζει το delta vector του. Κατά συνέπεια με μηδενική συνολική μεταβολή στον υπολογισμό. Για να καταλάβουμε γιατί συμβαίνει αυτό, ας υποθέσουμε ότι S_i ανίχνευσε μια τοπική παραβίαση στον έλεγχο ομοιότητας με το κόμβο S_j. Τότε, η συνολική αλλαγή στο drift vector του S_i είναι το $-\frac{1}{2}\overrightarrow{\Delta v_i}$. Καθώς ο κόμβος μηδενίζει το $\overrightarrow{\Delta v_i}$ του μετά από τη μετάδοση, αλλά το μισό από εκείνο το διάνυσμα, $\frac{1}{2}\overrightarrow{\Delta v_i}$ έχει προστεθεί στο estimate vector. Ενώ η συνολική αλλαγή στο drift vector του S_j είναι $\frac{1}{2}\overrightarrow{\Delta v_i}$, αφού τροποποίησε το διάνυσμα εκτίμησης του κατά $\frac{1}{2}\overrightarrow{\Delta v_i}$.

Έτσι, το άθροισμα των drift vector πριν και μετά τη λειτουργία autobalance παραμένει το ίδιο.

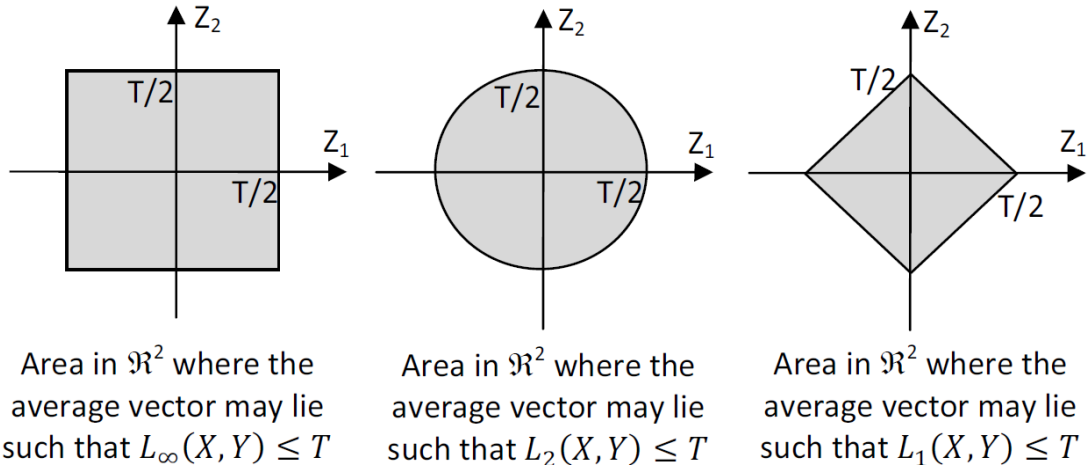
Στη λειτουργία Autobalance ο κόμβος που λαμβάνει πρώτος μήνυμα κατά τη διάρκεια μιας εποχής, δεν μεταδίδει κατ 'ανάγκη το διάνυσμα μέτρησής του, εκτός εάν εντοπίσει μια ολική παράβαση. Επομένως, αναμένουμε ότι η λειτουργία Autobalance θα οδηγήσει σε μια σημαντική μείωση του αριθμού των μεταδιδόμενων μηνυμάτων. Εντούτοις, δεδομένου ότι μόνο το δεύτερο μήνυμα (μήνυμα απάντησης) μπορεί να αποφευχθεί σε κάθε εποχή, αυτή η μείωση δεν μπορεί να υπερβεί το 50%, σε σύγκριση με τον Simple τρόπο λειτουργίας.

Εκμετάλλευση κυρτότητας / Convexity των ασφαλών ζωνών. Για τις L_∞, L_1 και L_2 λειτουργίες ομοιότητας, μπορούμε να μειώσουμε ακόμα περισσότερο τον αριθμό των μεταδιδόμενων μηνυμάτων με την εκμετάλλευση της έννοιας των ασφαλών ζωνών (Safe Zones, SZ) [14] και την κυρτότητα των ασφαλών ζωνών για τις προαναφερόμενες λειτουργίες ομοιότητας.

Σύμφωνα με το [14], "SZ ενός κόμβου αποτελείται από το σύνολο των διανυσμάτων που ικανοποιούν τους τοπικούς περιορισμούς και εφ' όσον παραμένουν τα διανύσματα στη SZ καμία επικοινωνία δεν απαιτείται". Με τον παραπάνω ορισμό για τις

L_∞, L_1 και L_2 μετρικές ομοιότητας, μπορούμε να καθορίσουμε τις γκρίζες ασφαλείς ζώνες που απεικονίζονται για W = 2, στην Εικόνα 15. Σε υψηλότερες διαστάσεις οι αντίστοιχες ασφαλείς ζώνες είναι υπερκύβος (για L_∞), μια σφαίρα (για L_2) και η τομή από υπερεπίπεδα (για L_1). Σημειώνουμε ότι οι ασφαλείς ζώνες που απεικονίζουμε είναι κυρτές. Κατά συνέπεια, όσο το διάνυσμα εκτίμησης και το drift vector ενός κόμβου παραμένουν εντός της ασφαλής ζώνης, τότε το global statistics vector (που είναι ο μέσος όρος των drift vector), θα βρίσκεται επίσης μέσα στην ασφαλή ζώνη. Εκμεταλλευόμενοι αυτό το γεγονός, αρκεί να ελέγχουμε την τιμή της μετρικής ομοιότητας στο διάνυσμα εκτίμησης $\vec{e}$. Οσο αυτό βρίσκεται εντός ασφαλής ζώνης, μπορώ να ελέγχω τοπικούς περιορισμούς με βάση την τιμή της μετρικής στο drift vector (Εικόνα 16). Αν το διάνυσμα εκτίμησης είναι εκτός ασφαλής ζώνης, τότε ελέγχω τοπικούς περιορισμούς με έλεγχο μονοχρωματικής σφαίρας. Οπότε μπορούμε να τροποποιήσουμε επιπλέον τον Algorithm 3(Εικόνα 14) με τον ακόλουθο τρόπο:

Αν $f(\vec{e}) \leq T$ τότε απλά ελέγχω αν $f(\vec{u}) \leq T$(δηλαδή, δεν χρειάζεται να βρούμε τη μέγιστη τιμή σε μια σφαίρα), προκειμένου να εξεταστεί εάν υπάρχει μια τοπική παραβίαση. Στη περίπτωση όπου $f(\vec{e}) > T$, ακολουθείται η υπάρχουσα τεχνική που αναφέρεται στις γραμμές 1-2 του Algorithm 3. Συνοψίζοντας όσο το διάνυσμα εκτίμησης $\vec{e}$ είναι εντός ασφαλής ζώνης, ελέγχω τοπικούς περιορισμούς με βάση τη θέση του drift vector(εκτός SZ σηματοδοτεί local violation, εντός SZ καθορίζει μια ασφαλής μέτρηση). Αν το διάνυσμα εκτίμησης $\vec{e}$ είναι εκτός SZ τότε ελέγχω τοπικούς περιορισμούς με σφαίρες. Την παραλλαγή του Simple τρόπου λειτουργίας που εκμεταλλεύεται την κυρτότητα των ασφαλών ζωνών για L_∞, L_1 και L_2 τον ορίζουμε ως Convex τρόπο λειτουργίας. Ορίζουμε ως Convex Autobalance την αντίστοιχη παραλλαγή στο τρόπο λειτουργίας Autobalance.



Εικόνα 15 : Ασφαλής ζώνες για L_∞, L_2 και L_1 στις δυο διαστάσεις. Απεικόνιση των σκιασμένων περιοχών, όπου το global statistics vector μπορεί να βρεθεί ώστε L_∞, L_2 και L_1 απόσταση μεταξύ των δυο κόμβων αισθητήρων να μην είναι πάνω από το κατώτατο όριο T. Οι σκιασμένες περιοχές είναι επίσης κυρτές σε οποιαδήποτε διάσταση για αυτά τα μέτρα ομοιότητας.

B. Γειτονική κόμβοι σε άμεση επικοινωνία

Στην περίπτωση που όλοι οι γείτονες του S_i είναι σε άμεση επικοινωνία με το συγκεκριμένο κόμβο S_i, τότε οποιαδήποτε μετάδοση μηνυμάτων από τον S_i που αφορά στο διάνυσμα μετρήσεων του, μπορεί να πραγματοποιηθεί χρησιμοποιώντας broadcast επικοινωνία. Στέλνοντας λοιπόν μόνο ένα μήνυμα με το διάνυσμα μετρήσεων του, μεταδίδεται σε όλους τους γείτονες. Ο κόμβος θα πρέπει να κρατάει ένα bitmap που θα μαρκάρει σε ποιους αναφέρεται αυτό το μήνυμα, ώστε να μπορεί να χειριστεί καλυτέρα τα λαμβανόμενα μηνύματα. Για το λόγο ότι ένας κόμβος μπορεί να στείλει μόνο ένα μήνυμα(broadcast) ανά εποχή, πρέπει οι κόμβοι να είναι προσεκτικοί όταν λαμβάνουν κάποιο μήνυμα. Ακόμα και αν δεν αναφέρεται σε αυτούς, δεν πρέπει να αγνοηθεί αλλά πρέπει να γίνουν οι απαραίτητοι έλεγχοι. Καθώς επίσης για τις περιπτώσεις με autobalance πρέπει ο κόμβος να ενημερώνει προληπτικά τα διανύσματα εκτίμησης για τους κόμβους που πρόκειται να ενεργήσουν μετά από αυτόν, ακόμα και αν δεν παρατηρεί κάποια τοπική παραβίαση.

Λόγο της ομοιότητας που έχουν οι λειτουργίες μεταξύ τους ως προς τον τρόπο διαχείρισης broadcast μηνυμάτων, δεν θα απαριθμήσουμε όλους τους πιθανούς συνδυασμούς για broadcast επικοινωνία με Simple, Convex, Autobalance, Convex & Autobalance, αλλά επικεντρωνόμαστε στην πιο αποτελεσματική, που ορίζουμε Proactive Broadcast Convex λειτουργία κόμβου. Που βασίζεται στην λειτουργία Convex και Autobalance.

Σύμφωνα με αυτό τον τρόπο λειτουργίας, οι κόμβοι στη γειτονιά σύγκρισης θα πρέπει να ταξινομηθούν, έτσι ώστε κάθε κόμβος να γνωρίζει ποιος ενεργεί πριν από αυτόν και ποιος μετά. Ας εξετάσουμε τη λειτουργία του κόμβου S_i :

- Ο S_i περιμένει για πιθανά μηνύματα από τους κόμβους που ενεργούν πριν από αυτόν.
- Για κάθε λαμβανόμενο μήνυμα, ο S_i ενημερώνει το διάνυσμα εκτίμησης για τον αντίστοιχο γειτονικό κόμβο και ελέγχει αν μια ολική παραβίαση έχει εμφανιστεί. Στη περίπτωση ολικής παραβίασης ο κόμβος S_i θα αποθηκεύσει το global statistics vector ως διάνυσμα εκτίμησης και θα μεταδώσει αργότερα την μέτρηση του με ένα broadcast μήνυμα(μαρκάρει τον αντίστοιχο κόμβο στο bitmap) , έτσι παραλείπουμε το επόμενο βήμα.
- Ο κόμβος S_i εξετάζει εάν υπάρχει μια τοπική παραβίαση περιορισμών με βάση τον εαυτό του και τους κόμβους που ενεργούν μετά από τον S_i ή τους κόμβους που ενεργούν πριν από αυτόν αλλά δεν έστειλαν μήνυμα. Στη συνέχεια ο κόμβος S_i μαρκάρει όλους αυτούς τους κόμβους που παρατηρεί ένα local violation, με σκοπό να μεταδώσει αργότερα την μέτρηση του με ένα broadcast μήνυμα.
- Εάν ένα από τα βήματα 2,3 απαιτούν μια μετάδοση (υπάρχει δηλαδή κάποιος κόμβος μαρκαρισμένος στο bitmap), ο S_i μεταδίδει με ένα broadcast μήνυμα το διάνυσμα μετρήσεων του. Στη συνέχεια ο S_i θα ενημερώσει τα διανύσματα εκτίμησης για όλους τους κόμβους (i) οι οποίοι ενεργούν μετά από αυτόν ή είχε local violation με αυτούς (αυτό γίνεται λόγω autobalancing), (ii) για τους οποίους μια ολική παραβίαση ανιχνεύθηκε στο πρώτο βήμα. Αξίζει να σημειωθεί ότι σε αυτή

την περίπτωση ο κόμβος S_i τροποποιεί προληπτικά το διάνυσμα εκτίμησης του για τους κόμβους που ενεργούν μετά από αυτόν, ακόμα και όταν δεν ανιχνεύσει μια τοπική παραβίαση στον έλεγχο ομοιότητάς τους.

- Ο κόμβος S_i περιμένει στη συνέχεια, τα μηνύματα σχετικά με πιθανές παραβιάσεις. Αγνοεί τα λαμβανόμενα μηνύματα από οποιοδήποτε κόμβο S_j, που ενεργεί μετά από τον S_i και που δεν υπάρχει μια ολική παραβίαση με τον S_i (δηλ., το λαμβανόμενο μήνυμα οφειλόταν σε μια παραβίαση που είχε ο S_j με κάποιον άλλο κόμβο). Η τελευταία ενέργεια είναι εύκολο να προσδιοριστεί από τον S_i, αφού μπορεί να κάνει έλεγχο ολικής παραβίασης με βάση την λαμβανόμενη μέτρηση από τον S_j και την δική του. Αν υπάρχει ολική παραβίαση πρέπει να αποθηκεύσει το διάνυσμα εκτίμησης, ώστε να έχει την ίδια τιμή με το global statistics vector.

C. Χειρισμός ερωτημάτων ελάχιστης υποστήριξης (minimum support)

Στην περίπτωση των ερωτημάτων που περιλαμβάνουν μια απαραίτητη ελάχιστη υποστήριξη (minimum support , ο αριθμός των κόμβων που έχουν παρόμοιες μετρήσεις) ο σταθμός βάσης θέλει να λαμβάνει ενημερώσεις από οποιοδήποτε κόμβο αισθητήρων στις εξής περιπτώσεις (i) Ο κόμβος είχε την απαραίτητη ελάχιστη υποστήριξη στην προηγούμενη εποχή, αλλά πέφτει κάτω η απαραίτητη ελάχιστη υποστήριξη στην τρέχουσα εποχή και (ii)όμοια όταν δεν είχε την απαραίτητη ελάχιστη υποστήριξη στην προηγούμενη εποχή , αλλά φθάνει στην απαραίτητη ελάχιστη υποστήριξη στην τρέχουσα εποχή. Ωστόσο, αυτό το πρόβλημα είναι εύκολο να το χειριστή το πλαίσιο μας, δεδομένου ότι κάθε κόμβος S_i θα μεταδώσει μόνο ένα μήνυμα στο τέλος της εποχής(δηλ., αφότου έχει ολοκληρωθεί η διαδικασία ελέγχου για όλους τους γείτονες για αυτήν την εποχή) εάν η τρέχουσα ελάχιστη υποστήριξη δεν ακολουθεί την προηγούμενη ελάχιστη υποστήριξη. Να υπενθυμίσουμε ότι κάθε κόμβος ξέρει με απόλυτη βεβαιότητα εάν η επιθυμητή ομοιότητα του με οποιοδήποτε κόμβο στη γειτονιά σύγκρισής του είναι επάνω από ή κάτω από το απαιτούμενο όριο.

Αναμένουμε στην πραγματικότητα οι minimum support queries να είναι αποτελεσματικότερες στην εξοικονόμηση εύρους ζώνης, δεδομένου ότι ο αρχικός αλγόριθμος ενημερώνει το σταθμό βάσης κάθε φορά που η κατάσταση ομοιότητας τροποποιείται με κάποιον κόμβο στη γειτονιά σύγκρισης του. Ωστόσο, όπως δείχνουμε στην πειραματική αξιολόγηση μας, η συνολική εξοικονόμηση εύρους ζώνης περιορίζεται από το γεγονός ότι η μεγάλη πλειοψηφία των μεταδιδόμενων μηνύματα μεταξύ των ζευγαριών είναι λόγο των δοκιμών ομοιότητας. Έτσι, αν και μειώνονται οι εκπομπές στο σταθμό βάσης, αυτές οι μεταδόσεις ήταν σχετικά λίγες.

Σύμβολα που διατηρεί ο S_i	**Ερμηνεία συμβόλων** (Στη θέση j του αντίστοιχου πίνακα [ ])	**Εμβέλεια συμβόλων**		
Estimate [ ]	Estimate vector που υπολογίστηκε σε σχέση με τον κόμβο S_j, σε κάποια χρονική περίοδο.	**Simple , Convex**	**Autobalance**	**Broadcast**
Reference [ ]	Local statistics vector που έστειλε ο κόμβος S_i στον κόμβο S_j, την τελευταία φορά.			
Drift [ ]	Παρόν drift vector του S_i σε σχέση με κόμβο S_j. $\vec{u}_{i,j} = \vec{e}_{i,j} + \Delta\vec{v}_{i,j}$			
Failed map [ ]	'1'/ '0' : ο κόμβος S_i έχει/δεν έχει τοπική παραβίαση με τον κόμβο S_j, στη συγκεκριμένη εποχή. (για να ξέρει ο S_i αν ένα λαμβανόμενο μήνυμα οφείλετε σε δικιά του παραβίαση, οπότε είναι απάντηση από τον S_j ή οφείλετε σε τοπική παραβίαση του S_j.			
previous minimum Support	Το minimum support που είχα στη προηγούμενη εποχή.			
Receive [ ]	Local statistics vector που έλαβε ο κόμβος S_i από τον κόμβο S_j, την τελευταία φορά. Για να μπορεί να υπολογίζει το $\overrightarrow{\Delta v_j}$, ώστε να αποθηκεύσει πιθανός το $\overrightarrow{e'}$			
Flag map [ ]	'1'/ '0' : ο κόμβος S_i έχει/δεν έχει τοπική παραβίαση με τον κόμβο S_j, στη συγκεκριμένη εποχή. Πρέπει/δεν πρέπει να απαντήσει σε μήνυμα από S_j(Global violation με S_j) .			
Receive map [ ]	'1'/ '0' : ο κόμβος S_i έλαβε/δεν έλαβε σε αυτή την εποχή από τον κόμβο S_j κάποιο μήνυμα Για να ξέρω αν αποθηκεύσω προσωρινά $\overrightarrow{e'}$ άμα δεν μου απαντήσει.			

Πίνακας ΙΙΙ: Τα στατιστικά στοιχεία που κρατούνται από έναν κόμβο S_i, η ερμηνεία τους και εμβέλεια αποθήκευσης τους ανάλογα με το μοντέλο λειτουργίας. Στη δεύτερη στήλη εξηγούμε τη αποθηκεύει στη θέση j του αντίστοιχου πίνακα.

D. Σημείωση για περίπτωση απώλειας μηνυμάτων

Όλοι οι αλγόριθμοι που παρουσιάζονται υποθέτουν, όπως και στην προηγούμενη εργασία που χρησιμοποιεί τη γεωμετρική προσέγγιση, την αξιόπιστη παράδοση των μηνυμάτων. Εντούτοις στις εφαρμογές δικτύων αισθητήρων συχνά τα μηνύματα χάνονται λόγω των συγκρούσεων. Συζητάμε τώρα τις συνέπειες της απώλειας μηνυμάτων χρησιμοποιώντας την γεωμετρική προσέγγιση.

Σε περίπτωση που χάνεται ένα μήνυμα που αφορά στην ενημέρωση προς το σταθμό βάσης, ο σταθμός βάσης δεν γνωρίζει εάν οι δύο κόμβοι είναι παρόμοιοι ή όχι. Στην περίπτωση των ερωτημάτων ελάχιστης υποστήριξης, ο σταθμός βάσης δεν ξέρει εάν οι μετρήσεις ενός κόμβου έχουν φθάσει στην απαραίτητη ελάχιστη υποστήριξη ή όχι. Αυτό το ζήτημα θα επιλυθεί, στο επόμενο μήνυμα που αποστέλλεται από τον ίδιο κόμβο στο σταθμό βάσης.

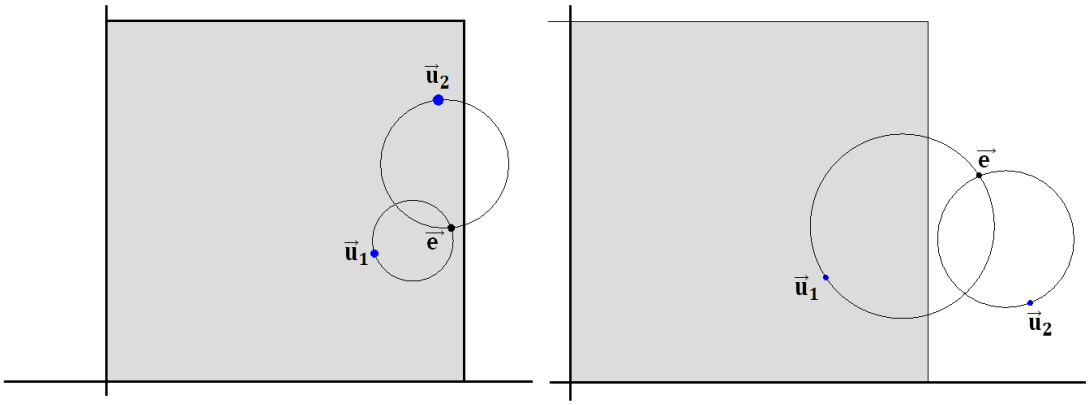
Στις εκδόσεις των αλγορίθμων μας που δεν περιλαμβάνουν τον τρόπο Autobalance, κάθε κόμβος που κάνει αρχικά μια μετάδοση λόγω μιας τοπικής παραβίασης αναμένει μια απάντηση. Κατά συνέπεια εάν ένα μήνυμα χαθεί ο κόμβος δεν θα λάβει μια απάντηση, οπότε θα αναγνωρίσει αυτήν την απώλεια μηνύματος και μπορεί να στείλει εκ νέου το μήνυμά του επιλύοντας κατά συνέπεια αυτό το ζήτημα. Από την άλλη πλευρά στις εκδόσεις των αλγορίθμων μας που περιλαμβάνουν τη λειτουργία Autobalance, υπάρχει περίπτωση ένας κόμβος που παρατηρεί μια τοπική παραβίαση και διαβιβάζει ένα μήνυμα να μην λάβει μια απάντηση. Σε αυτή την περίπτωση ο κόμβος δεν ξέρει εάν αυτό εμφανίστηκε λόγω μιας απώλειας μηνύματος ή λόγω μιας επιτυχούς λειτουργίας autobalance στον αντίστοιχο κόμβο ζεύγους (εκτός και αν κάνουμε χρήση ACK). Σε αυτήν την περίπτωση, οι 100% εγγυήσεις των αλγορίθμων μας δεν ισχύουν πλέον, δεδομένου ότι οι δύο αντίστοιχοι κόμβοι ελέγχου θα καταλήξουν να έχουν διαφορετικά διανύσματα εκτίμησης. Ας εξετάσουμε τώρα, πότε αυτό το πρόβλημα θα επιλυθεί. Έστω X χαρακτηρίζεται ο κόμβος του ζεύγος, του οποίου το μεταδιδόμενο μήνυμα δεν ελήφθη. Το πρόβλημα μπορεί να αντιμετωπιστεί μόνο είτε (i) αν στην επόμενη εποχή ο X εντοπίζει μια τοπική παράβαση (κατά τον έλεγχο ομοιότητας του ίδιου ζεύγους κόμβων), αν το μήνυμα του X δεν χαθεί και πάλι ή (ii) αν στην επόμενη εποχή ο αντίστοιχος κόμβος του ζεύγους ομοιότητας ανιχνεύσει για πρώτη φορά μια τοπική παραβίαση και ο X εντοπίσει παραβίαση κατώτατου ορίου. Αφού τότε ο X θα διαβιβάσει το διάνυσμα μετρήσεών του στον κόμβο του ζεύγους ομοιότητας.

Συμβολισμός	Ερμηνεία
Data	Η παρούσα μέτρηση του κόμβου (local statistics vector)είτε επειδή προκάλεσε τοπική παράβαση ή απάντηση σε κάποια τοπική παραβίαση του αποστολέα.
Sender	Το αναγνωριστικό id του κόμβου που στέλνει το μήνυμα

Πίνακας IV: Τα περιεχόμενα του μηνύματος

Ε. Χειρισμός προσθήκης και αφαίρεσης κόμβων

Όταν οι κόμβοι προστίθενται ή αφαιρούνται στο δίκτυο (παραδείγματος χάριν, λόγω μια αποτυχίας κόμβου), τότε κάθε αισθητήρας πρέπει απλά να ελέγξει την ομοιότητά του με περισσότερους ή αντίστοιχα λιγότερους κόμβους αισθητήρων. Στις περισσότερες περιπτώσεις, αυτό δεν απαιτεί οποιαδήποτε πρόσθετη εργασία από τους υπόλοιπους κόμβους αισθητήρων. Στην περίπτωση των ερωτημάτων ελάχιστης υποστήριξης, η αφαίρεση κόμβων αισθητήρων μπορεί να αναγκάσει μερικούς κόμβους να ενημερώσουν το σταθμό βάσης, δεδομένου ότι η ελάχιστη υποστήριξη τους μπορεί να πέσει κάτω από το απαιτούμενο όριο. Η μόνη ασήμαντη επιπλοκή που προκύπτει περιλαμβάνει την περίπτωση broadcast μετάδοσης που περιγράφεται στο VI-B: Κάθε νέος κόμβος που προστίθεται στο δίκτυο μετά τον καθορισμό της γειτονιάς σύγκρισης CN_i, θα πρέπει να καθοριστεί η σειρά του (πότε θα ενεργήσει) εντός CN_i, έτσι ώστε να γνωρίζει ποιοι άλλοι αισθητήρες ενεργούν πριν από/μετά από αυτόν.

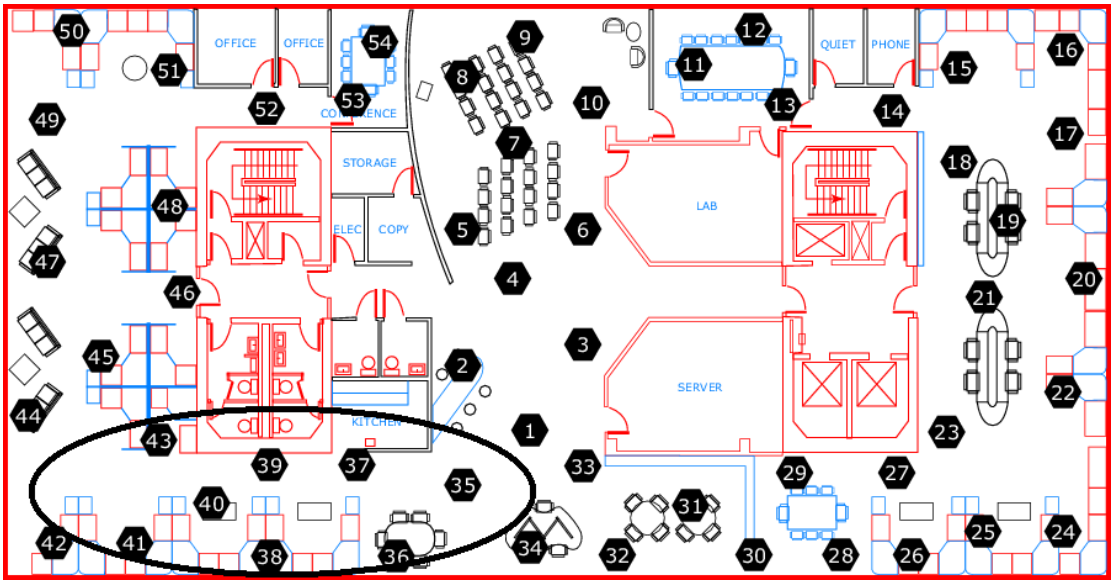


Εικόνα 16 : Έλεγχοι τοπικών περιορισμών με χρήση ασφαλών ζωνών. Η γκρι επιφάνεια είναι η ασφαλής ζώνη για την L_∞ μετρική. Στη πρώτη εικόνα βλέπουμε ότι το διάνυσμα εκτίμησης $\vec{e}$ είναι εντός ασφαλής ζώνης οπότε ελέγχω τοπικούς περιορισμούς με βάση τη θέση των drift vector($\vec{u}_1$ εντός SZ μονοχρωματική σφαίρα, $\vec{u}_2$ εντός SZ local violation). Στη δεύτερη εικόνα το διάνυσμα εκτίμησης $\vec{e}$ είναι εκτός SZ οπότε ελέγχω τοπικούς περιορισμούς με σφαίρες($\vec{u}_1$ local violation, $\vec{u}_2$ έχει μονοχρωματική σφαίρα).

VII. ΠΕΙΡΑΜΑΤΑ

Στα πειράματά μας χρησιμοποιήσαμε δύο σύνολα δεδομένων πραγματικού κόσμου. Το πρώτο σύνολο δεδομένων ορίζεται ως Intel Lab Data, είναι δεδομένα που συλλέγονται από Mica2Dot κόμβους αισθητήρες με weather boards για την υγρασία, θερμοκρασία και φως στο Intel Research εργαστήριο του Berkeley[42]. Τα δεδομένα συλλέχθηκαν με χρήση του TinyDB σε εντός δικτύου σύστημα επεξεργασίας ερωτήματος, φτιαγμένο σε πλατφόρμα TinyOS. Οι αισθητήρες τοποθετήθηκαν στο εργαστήριο σύμφωνα με το διάγραμμα που παρουσιάζετε στην Εικόνα 17. Επιλέξαμε τις μετρήσεις των ακόλουθων κόμβων (με την καθοριζόμενη σειρά) που βρίσκονται σε κοντινές θέσεις του εργαστηρίου : 38, 39, 40, 41, 43, 37, 35, 36. Στα πειράματα όπου θα μεταβάλλουμε τον αριθμό των χρησιμοποιούμενων κόμβων, κάθε πείραμα που περιέχει k κόμβους , χρησιμοποιεί τις μετρήσεις από τους k πρώτους κόμβους στην προαναφερόμενη λίστα για 30000 εποχές. Το δεύτερο σύνολο δεδομένων ορίζεται ως Weather Data, περιλαμβάνει μετρήσεις για τη θερμοκρασία του αέρα, σχετική υγρασία και την ηλιακή ακτινοβολία στο Πανεπιστήμιο της Ουάσιγκτον για το έτος 2002[43]. Χρησιμοποιήσαμε αυτές τις μετρήσεις για να παράγουμε τις αναγνώσεις μέχρι και 9 κόμβων για μια χρονική περίοδο 2000 εποχών. Για να αποφύγουμε την παρουσίαση πειραμάτων για την ίδια ποσότητα και στα δύο σύνολα δεδομένων (και αφού και για τις υπόλοιπες ποσότητες η συμπεριφορά ήταν παρόμοια), στο σύνολο δεδομένων Intel Lab Data παρουσιάζουμε τα αποτελέσματα για τη θερμοκρασία και το φως των δεδομένων, ενώ όσον αφορά τα Weather Data παρουσιάζουμε τα αποτελέσματα δεδομένων για την υγρασία και την ηλιακή ακτινοβολία .

Ο προσομοιωτής μας γράφτηκε στην Java. Όλα τα πειράματα εκτελέστηκαν σε έναν φορητό υπολογιστή που τρέχει Ubuntu, με έναν επεξεργαστή που χρονίζεται στα 2,4 GHz και 2GB μνήμη RAM.



Εικόνα 17 : Τοπολογία των κόμβων για το Intel Lab Data. Οι αισθητήρες τοποθετήθηκαν στο εργαστήριο σύμφωνα με το διάγραμμα που παρουσιάζετε. Στα πειράματα μας χρησιμοποιήθηκαν οι κυκλωμένοι αισθητήρες (38, 39, 40, 41, 43, 37, 35, 36)που βρίσκονται σε κοντινές θέσεις του εργαστηρίου.

Τεχνικές και ρυθμίσεις παραμέτρων. Συγκρίνουμε την απόδοση των τεχνικών Simple, Convex, Autobalance, Convex Autobalance και Proactive Broadcast Convex στα διαφορετικά σύνολα δεδομένων. Παράλληλα μεταβάλουμε διάφορες παραμέτρους, όπως το χρησιμοποιούμενο κατώτατο όριο, τη λειτουργία ομοιότητας, την ελάχιστη υποστήριξη, τη διάσταση του διανύσματος μετρήσεων ή των αριθμό των κόμβων αισθητήρων. Προκειμένου να είμαστε σε θέση να εξετάσουμε τις Convex εναλλακτικές λύσεις, εστιάσαμε στις L_∞, L_1 και L_2 λειτουργίες ομοιότητας. Επιπλέον αυτές οι λειτουργίες έχουν μια κλειστή μορφή λύσης που μας βοηθά να καθορίσουμε με ακρίβεια και έναν απλό τρόπο τις ελάχιστες και μέγιστες τιμές τους μέσα σε οποιαδήποτε σφαίρα.

Θεωρούμε επίσης στην αξιολόγηση μας (αλλά δεν απεικονίζεται στα διαγράμματα μας, λόγω της μεγάλης κατανάλωσης εύρους ζώνης της σε σχέση με τις τεχνικές μας), μια μέθοδο σύγκρισης που ονομάζεται NAIVE. Στην οποία οι κόμβοι αισθητήρων μεταδίδουν το διάνυσμα μέτρησης τους στο σταθμό βάσης σε κάθε εποχή. Για να ευνοήσουμε αυτόν τον NAIVE αλγόριθμο, παράγουμε τα cluster των k κόμβων αισθητήρων (η προκαθορισμένη τιμή του k ήταν 5), οι οποίοι είναι όλοι σε άμεση επικοινωνία μεταξύ τους και με το σταθμό βάσης. Ακόμα κι αν ο ανταγωνιστικός αλγόριθμος καλείται ως NAIVE, η ρύθμιση που χρησιμοποιήθηκε τον ευνοούσε εξαιρετικά. Αφού σε κάθε εποχή κάθε κόμβος μεταδίδει το διάνυσμα μετρήσεων του σε ένα μόνο hop. Να σημειωθεί επίσης ότι, όπως αναφέρεται στη Παράγραφο ΙΙ, οι πρόσφατες προσεγγίσεις [6], [8] απαιτούν από κάθε κόμβο να εκτελέσουν μια μετάδοση σε κάθε εποχή και επομένως δεν μπορούν να έχουν καλύτερες επιδόσεις από το βασικό μέτρο σύγκρισης του NAIVE με τον αλγόριθμο μας. Θα πρέπει να σημειωθεί επίσης, ότι σε αντίθεση με τον NAIVE αλγόριθμο, οι αλγόριθμοι μας μεταδίδουν προς το σταθμό βάσης μόνο όταν πραγματοποιείτε μια ολική παραβίαση. Έτσι, οι αλγόριθμοι μας σε αντίθεση με τον NAIVE, δεν θα έχουν σοβαρές επιβαρύνσεις σε περίπτωση που ο σταθμός βάσης είναι αρκετά hop μακριά από τους κόμβους αισθητήρων. Οι προεπιλεγμένες τιμές που χρησιμοποιούνται για τις παραμέτρους μας απεικονίζονται στον Πινάκας V.

Παράμετροι	**Προεπιλεγμένες τιμές**
k = μέγεθος του Cluster	5
W = διάσταση του διανύσματος	3
minSupp = ελάχιστη υποστήριξη	δεν καθορίζεται
#epochs = αριθμός των εποχών	Intel Lab Data: 30000,Weather Data: 2000

Πινάκας V : Παράμετροι και προεπιλεγμένες τιμές που χρησιμοποιήθηκαν στα πειράματα.

A. Μεταβάλλοντας τις λειτουργίες ομοιότητας

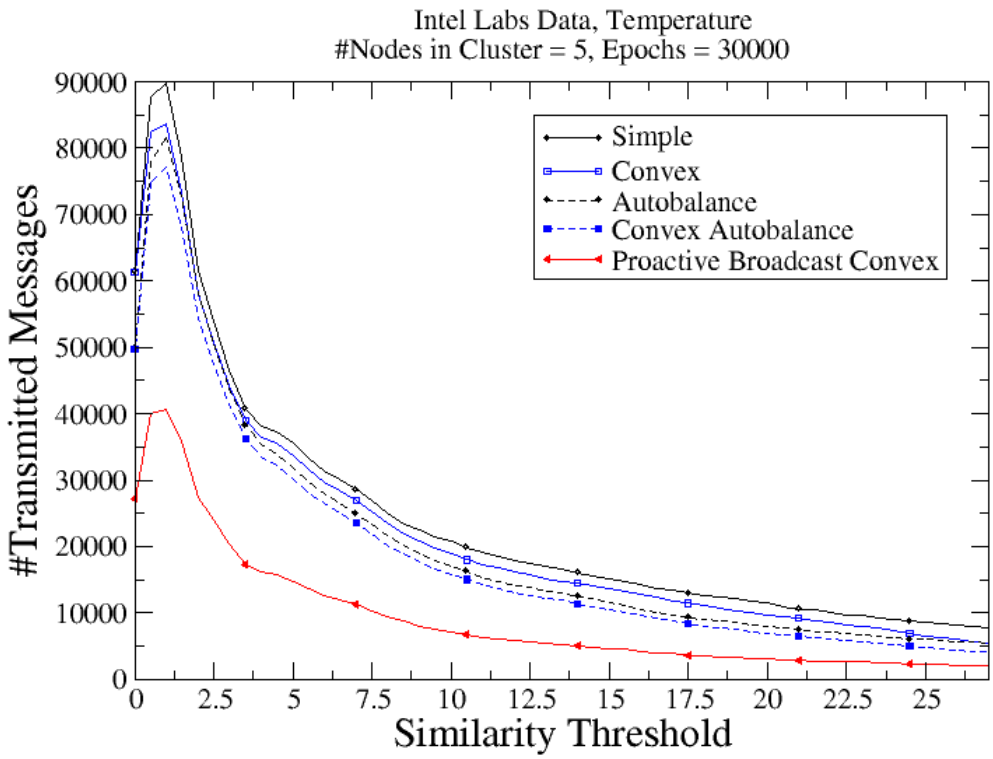
Στα γραφήματα 1 ως 4 δείχνουμε τον συνολικό αριθμό των μεταδιδόμενων μηνυμάτων των τεχνικών μας για τις διάφορες λειτουργίες κόμβων, για την L_1 λειτουργία ομοιότητας. Στα δεδομένα μεταβάλουμε ποικίλα το χρησιμοποιούμενο κατώτατα όριο. Το cluster περιέχει πέντε κόμβους, που σημαίνει ότι ο συνολικός αριθμός των διαβιβασθέντων μηνυμάτων για την μέθοδο αναφοράς NAIVE (δεν απεικονίζεται) ήταν 150.000 στα δεδομένα Intel Lab Data και 10.000 για τα Weather Data. Κάνουμε τρεις ενδιαφέρουσες παρατηρήσεις:

- Οι τεχνικές μας μπορούν να παρέχουν σημαντική εξοικονόμηση εύρους ζώνης σε σχέση με τη NAIVE προσέγγιση. Παραδείγματος χάριν με τη χρησιμοποίηση ενός μέτριου κατώτατου ορίου 5 βαθμών για την L_1 λειτουργία ομοιότητας στο γράφημα 1, για την λειτουργία κόμβου Simple, μειώνει τον αριθμό των μηνυμάτων κατά έναν παράγοντα 3.3 (σε σύγκριση με το NAIVE). Ενώ όσες περισσότερες βελτιστοποιήσεις χρησιμοποιούμε(Autobalance, Convex, Convex Autobalance) τόσο ο παράγοντας γίνεται καλύτερος, ενώ ο προτεινόμενος Proactive Broadcast Convex τρόπος λειτουργίας επιτυγχάνει περίπου 10 φορές μείωση στον αριθμό των μηνυμάτων. Παρατηρώντας αυτές τις βελτιώσεις, να θυμίσουμε ότι έχουμε επιλέξει ένα σύστημα που ευνοεί σημαντικά τον NAIVE αλγόριθμο.
- Στο ελεγχόμενο περιβάλλον του εργαστηρίου της Intel, ο μέγιστος αριθμός των μεταδιδόμενων μηνυμάτων παρατηρείτε σε πολύ μικρές τιμές κατωφλίου, το οποίο υπονοεί τις μικρές διαφορές στα διανύσματα μέτρησης των αισθητήρων. Το γεγονός αυτό φαίνεται να είναι ενθαρρυντικό, αφού δεν είναι και πολύ πιθανό κάποιος να επέλεγε τέτοιες χαμηλές τιμές κατώτατων ορίων προκειμένου να εξετάσει εάν οι αναγνώσεις μερικών αισθητήρων είναι "μη φυσιολογικές". Από την άλλη πλευρά το σύνολο δεδομένων Weather Data περιλαμβάνει μετρήσεις από εξωτερικούς αισθητήρες, με αποτέλεσμα να υπάρχουν σημαντικές μεταβολές μεταξύ των μετρήσεων τους. Αυτό εξηγεί γιατί ο μέγιστος αριθμός διαβιβασθέντων μηνυμάτων στα δεδομένα καιρού είναι και σε υψηλότερες τιμές κατωφλίου.
- Παρακαλώ να σημειωθεί ότι χρησιμοποιώντας τη γεωμετρική προσέγγιση, ένας κόμβος είναι λιγότερο πιθανό να έχει μια τοπική παραβίαση αν το estimate vector είναι πολύ μακριά από την επιφάνεια του κατώτατου ορίου, ανεξάρτητα από την πλευρά της επιφάνειας (πάνω από ή κάτω από το όριο) στην οποία ανήκει. Έτσι με πολύ χαμηλές (υψηλές) τιμές κατώτατων ορίων οι κόμβοι κατασκευάζουν συχνά τις μονοχρωματικές σφαίρες με τις τιμές εξ ολοκλήρου επάνω από (κάτω από) το επιθυμητό κατώτατο όριο, το οποίο στη συνέχεια οδηγεί σε λιγότερες τοπικές παραβιάσεις(local violation).

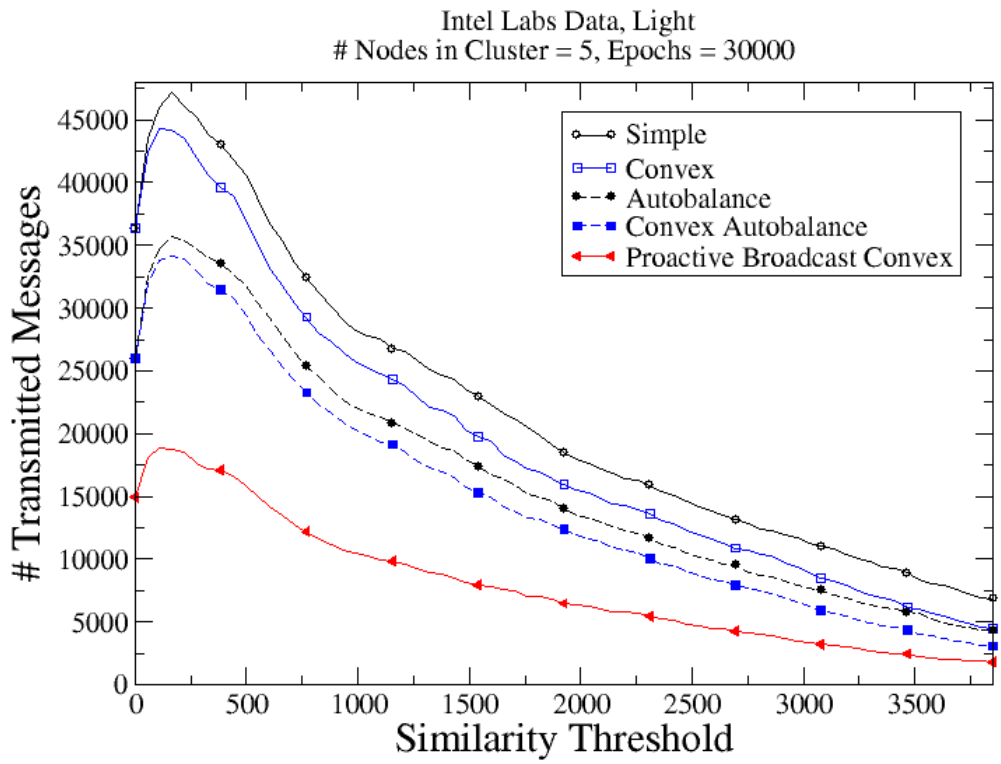
Στα γραφήματα 5 ως 8 που απεικονίζετε η απόδοση (για τα ίδια σύνολα δεδομένων), όταν χρησιμοποιείται η L_∞ λειτουργία ομοιότητας. Η απόδοση της περίπτωσης L_∞ είναι παρόμοια με την L_1, όπου από τη μια πλευρά δημιουργεί περισσότερα μηνύματα σε πολύ μικρές τιμές κατώτατων ορίων, αλλά από την άλλη πλευρά παρουσιάζεται μια πολύ γρήγορη μείωση στα μεταδιδόμενα μηνύματα όσο το κατώτατο όριο αυξάνεται. Στα γραφήματα 9 ως 12 που απεικονίζετε η απόδοση (για τα ίδια σύνολα δεδομένων), όταν

χρησιμοποιείται η L_2 λειτουργία ομοιότητας. Η απόδοση της περίπτωσης L_2 είναι επίσης παρόμοια με την L_1. Στη περίπτωση αυτή ο αριθμός των διαβιβασθέντων μηνυμάτων είναι μικρότερος, ενώ παρουσιάζεται μια πιο αργή μείωση στα μεταδιδόμενα μηνύματα όσο το κατώτατο όριο αυξάνεται. Οι τελευταίες παρατηρήσεις που προκύπτουν συγκρίνοντας L_1, L_∞και L_2 είναι αναμενόμενες, καθώς συγκρίνοντας τις μετρικές ομοιότητας για ένα συγκεκριμένο κατώτατο όριο T (π.χ. T=5), η έννοια ομοιότητας T=5 διαφέρει από μετρική σε μετρική. Συνοψίζοντας με τις συγκεκριμένες μετρικές ομοιότητας και χρησιμοποιώντας τις προτεινόμενες λειτουργίες των κόμβων, παρατηρούμε ότι σε όλες τις περιπτώσεις καταφέραμε να μειώσουμε αισθητά τον αριθμό των μεταδιδόμενων μηνυμάτων και να επιτρέψουμε έτσι στους κόμβους να απέχουν στη διαβίβαση μηνυμάτων των διανυσμάτων μέτρησης για αρκετές εποχές.

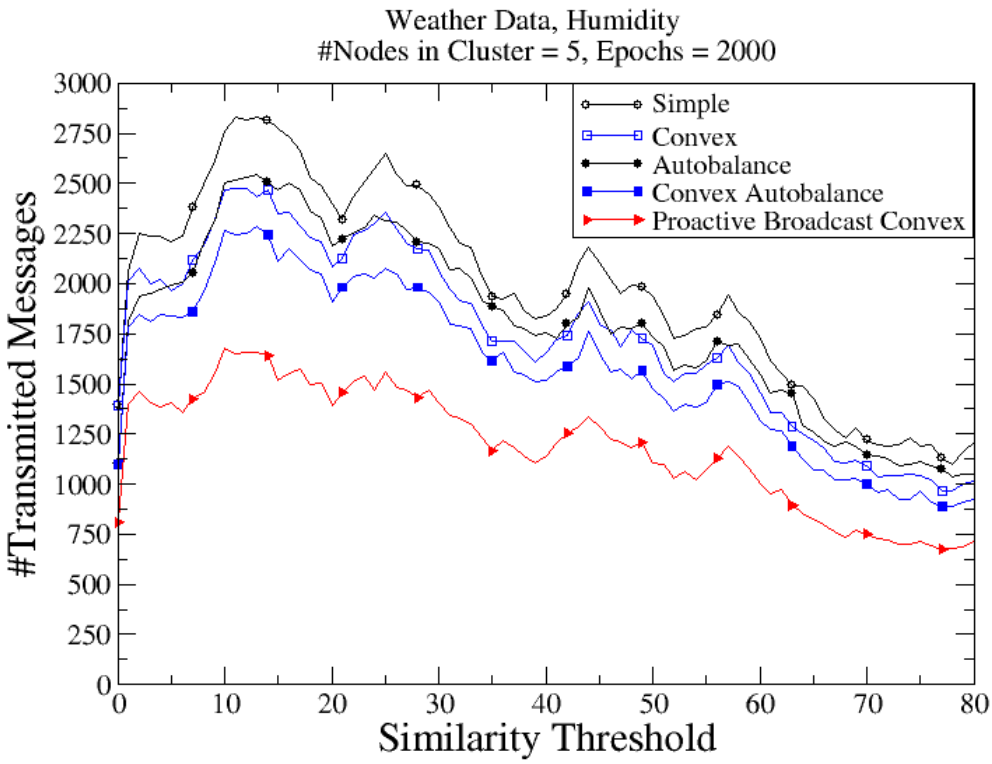
Επίσης στα γραφήματα 13 ως 15 απεικονίζονται αποτελέσματα για τις υπόλοιπες μετρικές ομοιότητας Cosine similarity, Extended Jaccard και Correlation coefficient. Όπως έχουμε αναφέρει λόγο της πολυπλοκότητας να γίνει έλεγχος μονοχρωματικής σφαίρας για τις συγκεκριμένες μετρικές ομοιότητας, ενώ και το χρησιμοποιούμενο πλέγμα δεν μας παρέχει ακρίβεια εμφανίζουμε λιγότερα γραφήματα. Επίσης στις περιπτώσεις αυτές, τις πρώτες W εποχές(όσο το μέγεθος του διανύσματος) στέλνονται μηνύματα από όλους τους κόμβους, τα οποία μηνύματα τα απαριθμούμε στις γραφικές. Αν δεν γίνει αυτό θα προκύψουν απροσδιόριστες μορφές λόγο των μηδενικών διανυσμάτων. Παρατηρούμε ότι τα αποτελέσματα στις cosine similarity και extended Jaccard είναι αρκετά αισιόδοξα, καθώς ο αριθμός των μηνυμάτων περιορίζεται αρκετά, με αποτέλεσμα οι κόμβοι να μην στέλνουν μηνύματα για αρκετές εποχές. Οι περισσότερες μεταδόσεις είναι σε κατώτατα όρια κοντά σε περιοχές ομοιότητας(κοντά στο 0.5 ως 1). Ενώ η μετρική για correlation coefficient δεν δείχνει να κερδίζει στη μετάδοση μηνυμάτων, παρά μόνο όταν οι μετρήσεις των κόμβων δεν μεταβάλλονται.



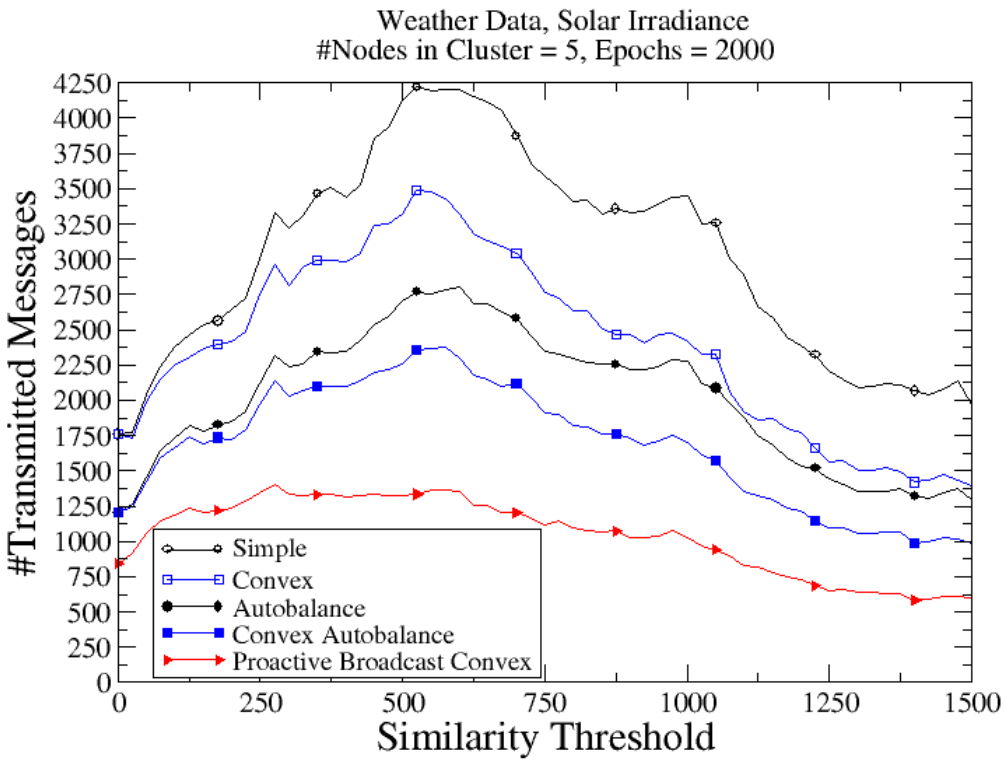
Γράφημα 1 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1



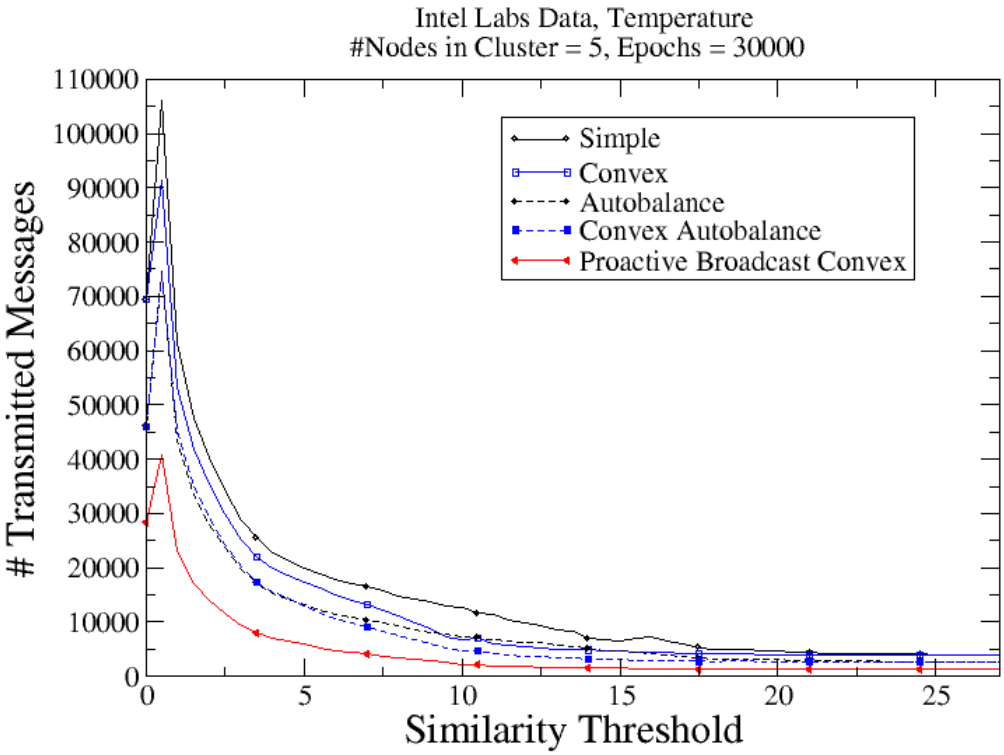
Γράφημα 2 : Intel Lab Data, ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1



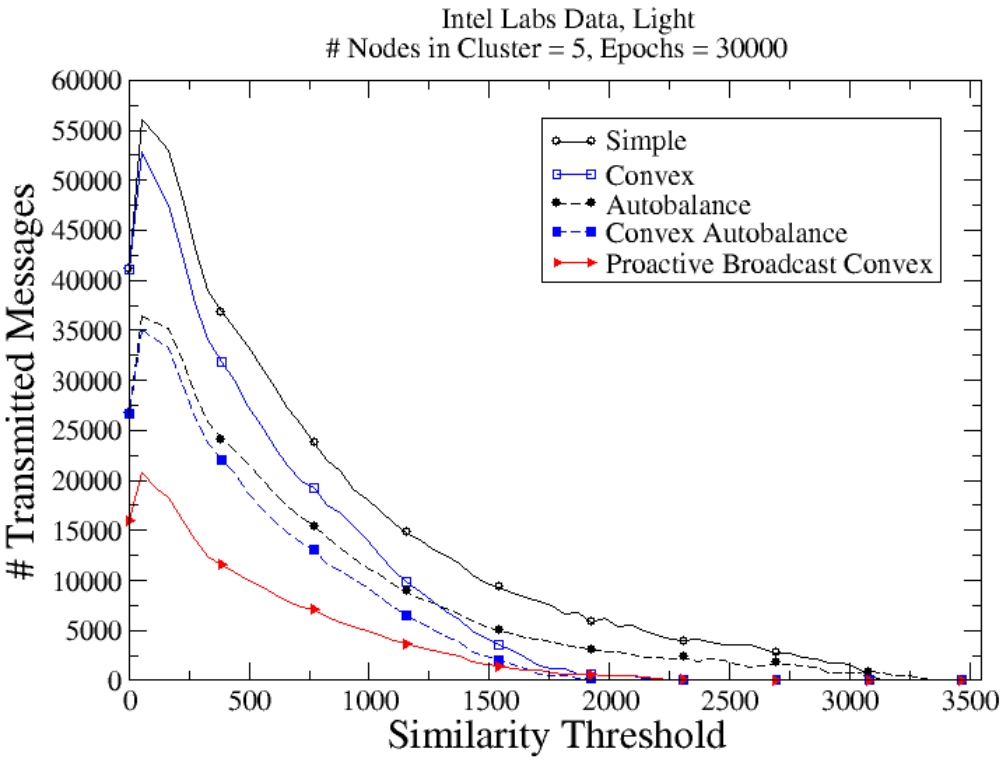
Γράφημα 3 : Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1



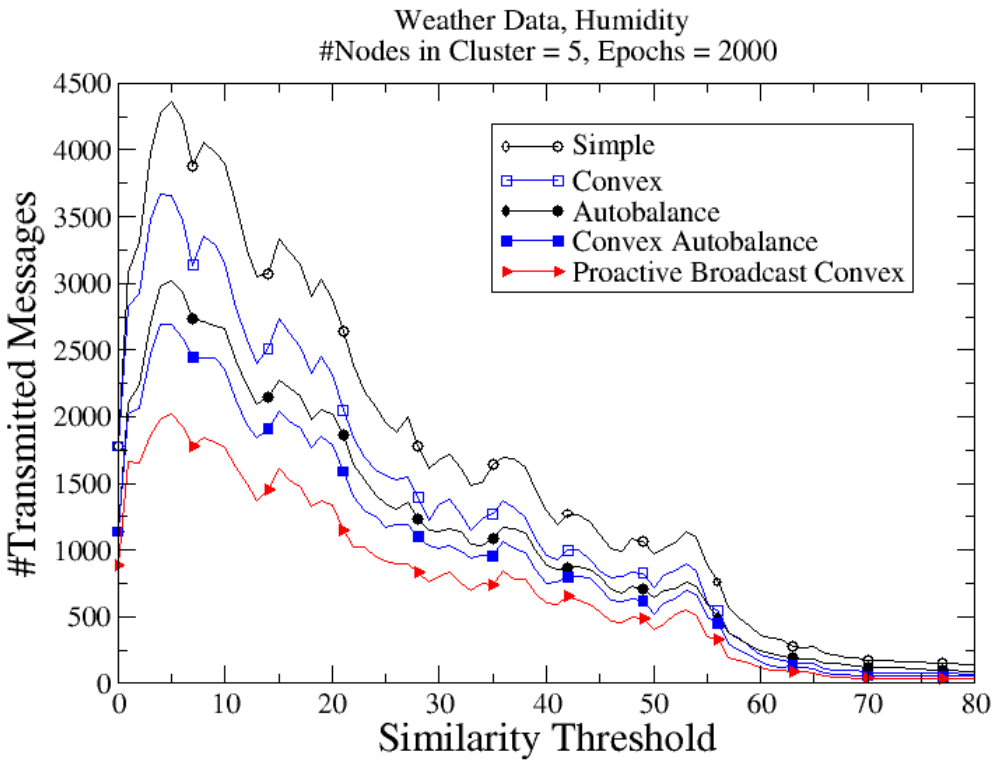
Γράφημα 4 : Weather Data, ηλιακή ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_1



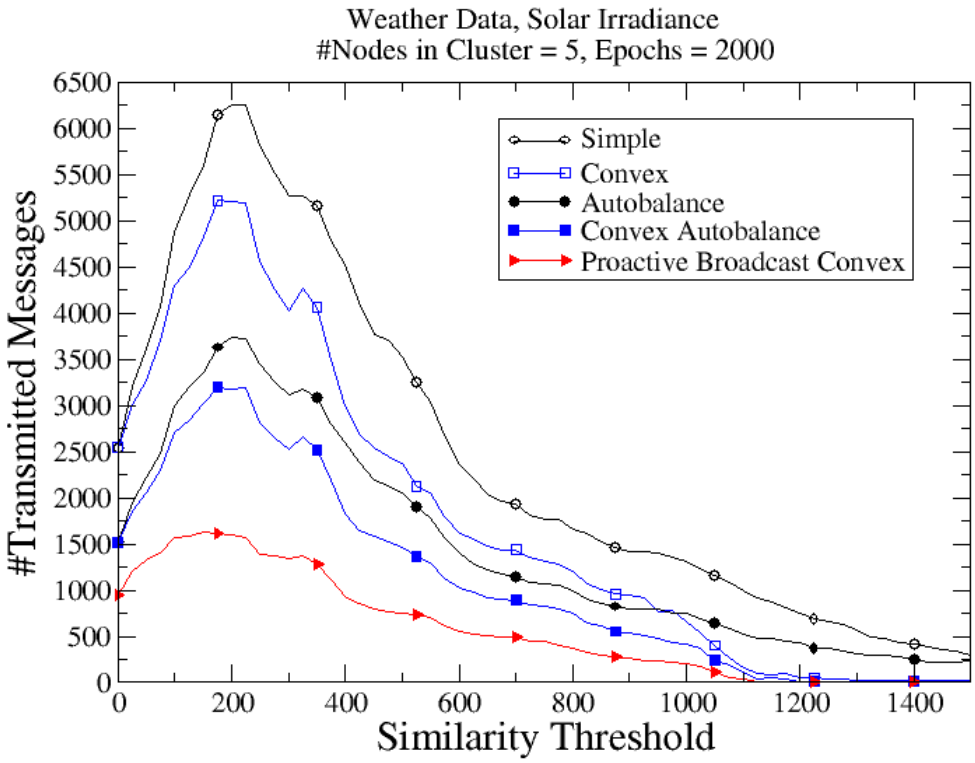
Γράφημα 5 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞



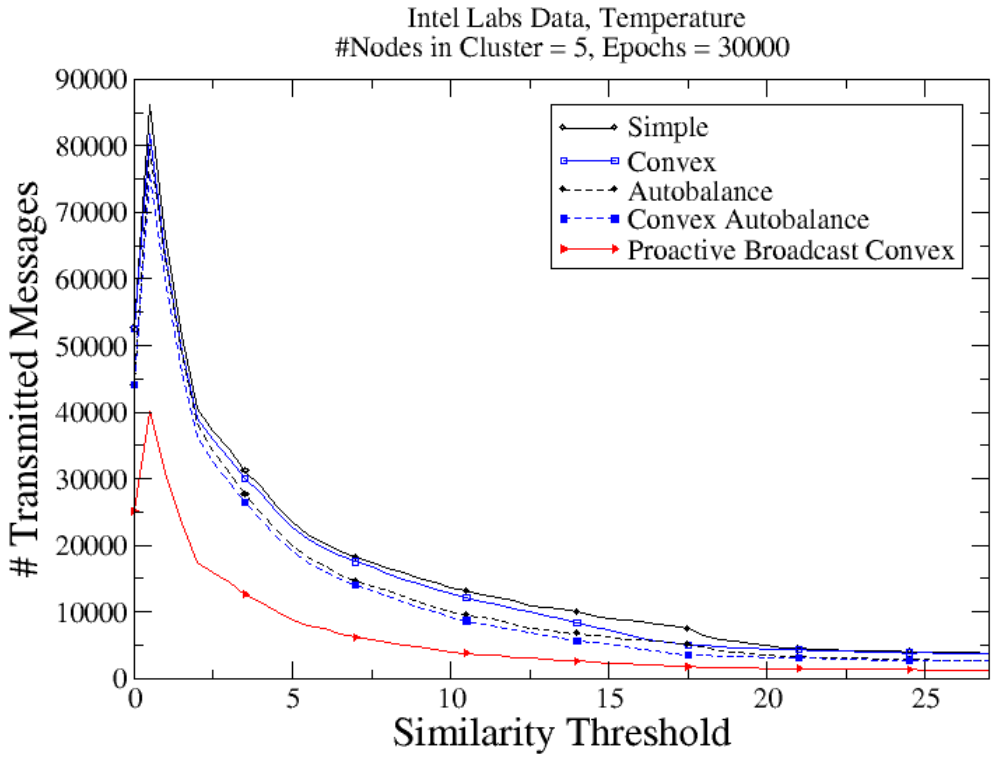
Γράφημα 6 : Intel Lab Data, ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞



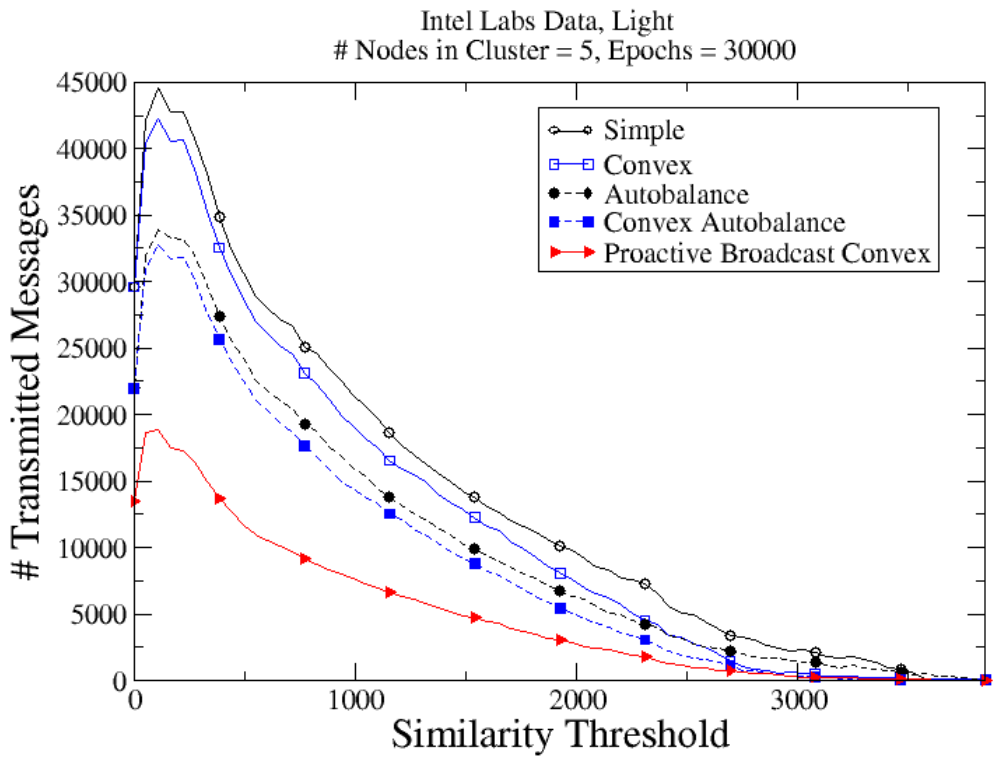
Γράφημα 7 : Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞



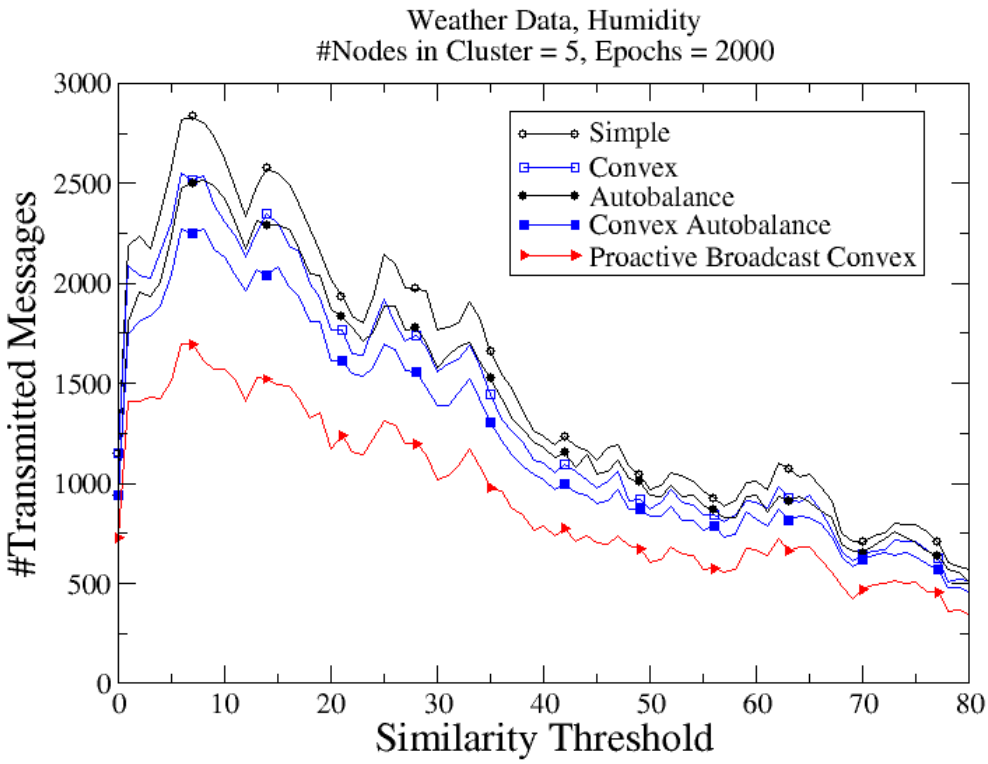
Γράφημα 8 : Weather Data, ηλιακή ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_∞



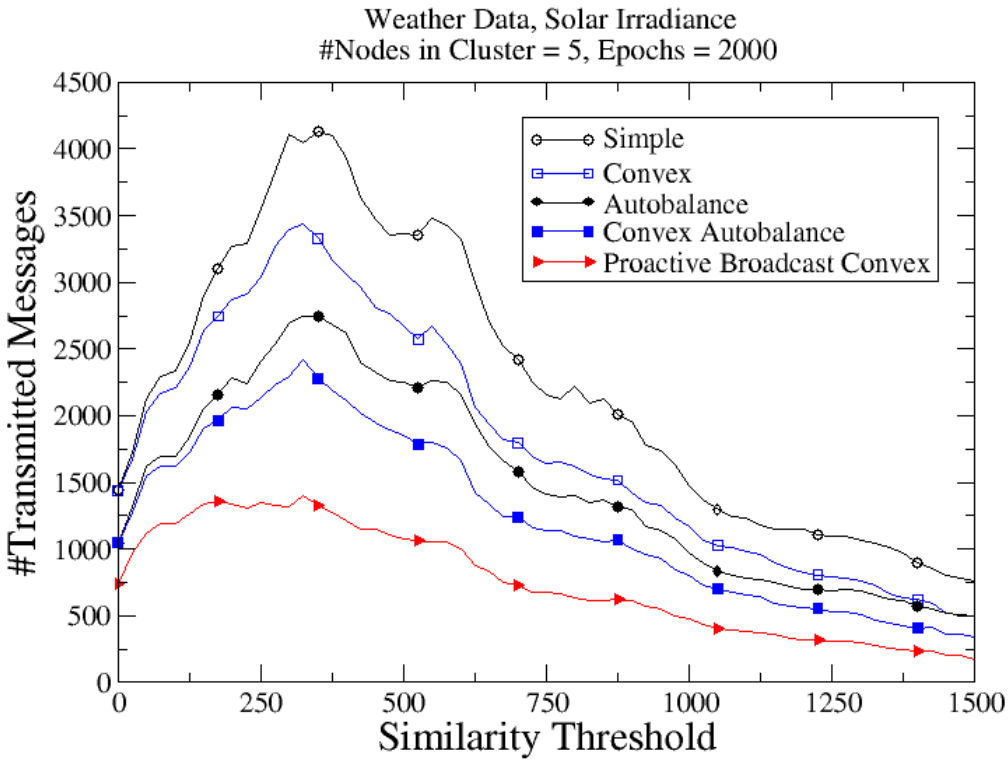
Γράφημα 9 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2



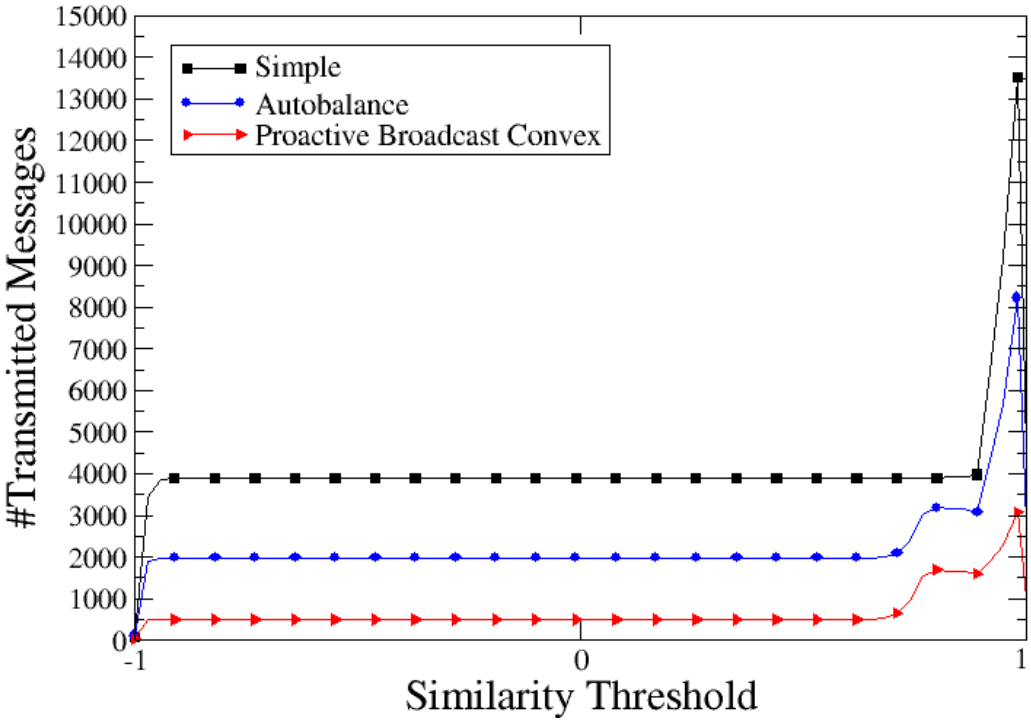
Γράφημα 10 : Intel Lab Data, ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2



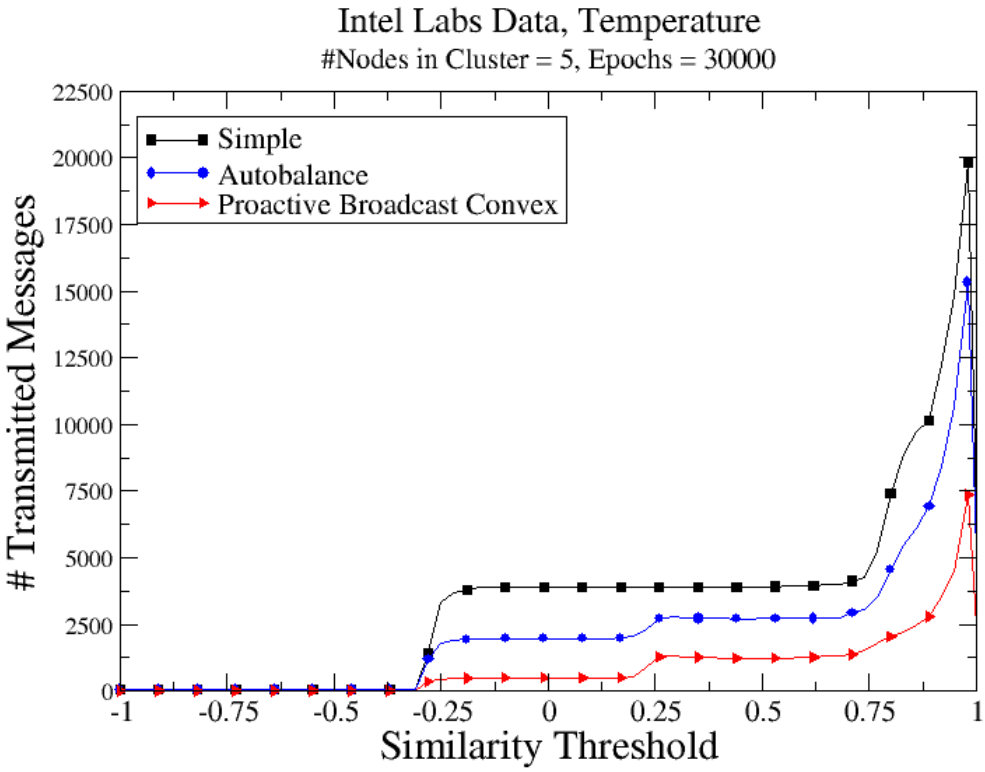
Γράφημα 11 : Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2



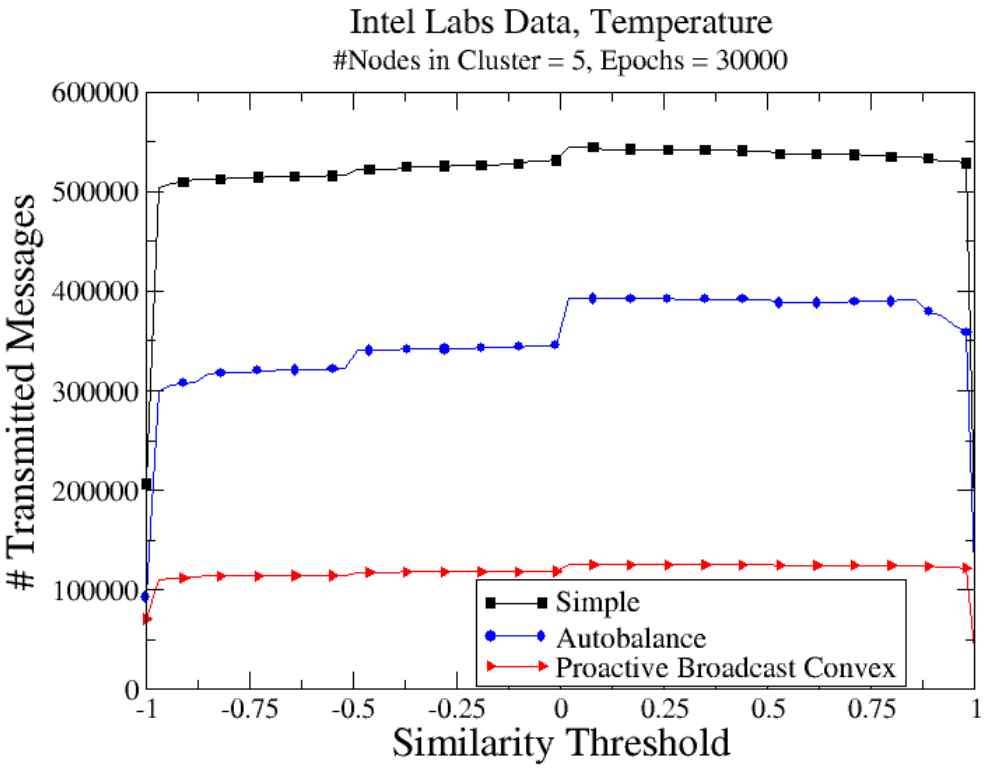
Γράφημα 12 : Weather Data, ηλιακή ακτινοβολία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, L_2



Γράφημα 13 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, Cosine



Γράφημα 14 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, Jaccard

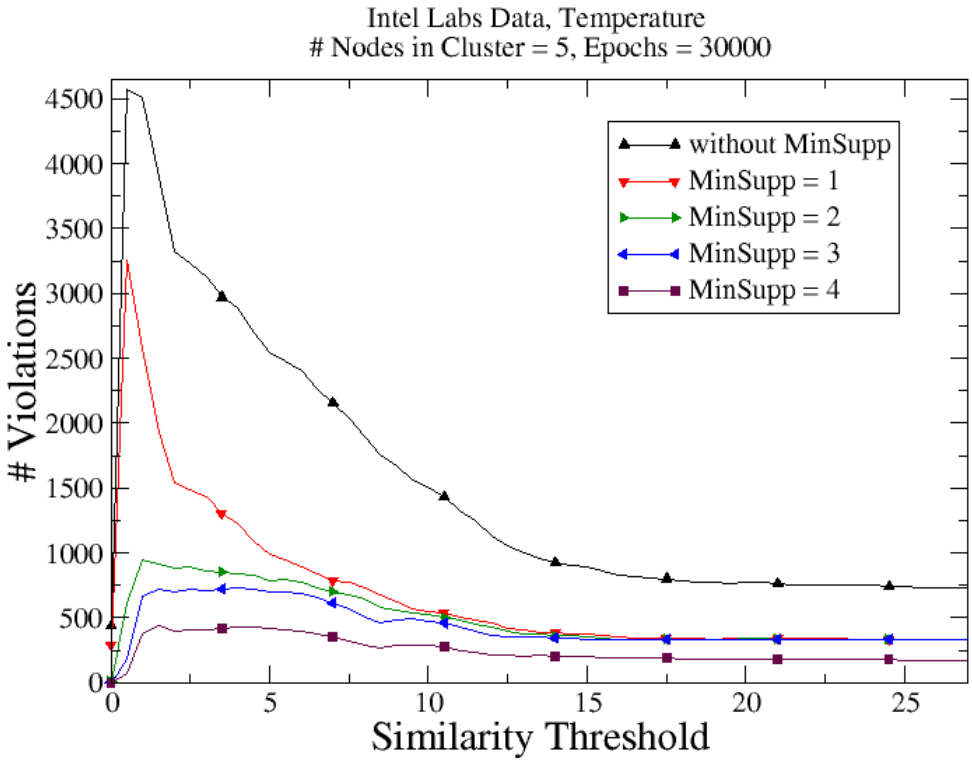


Γράφημα 15 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, Correlation

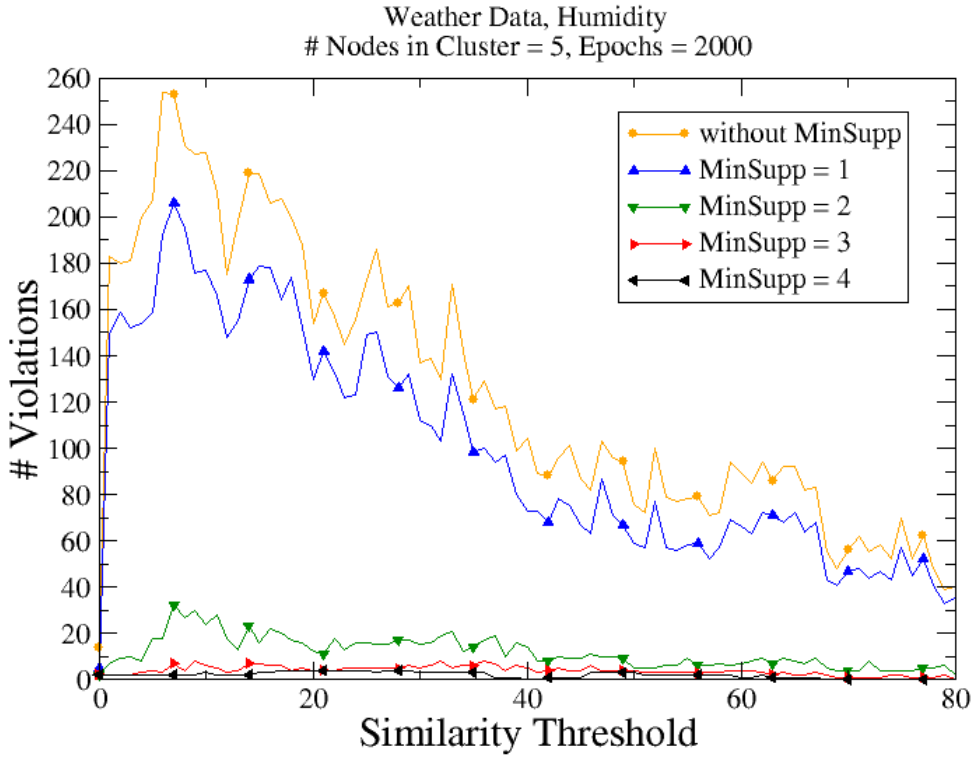
B. Ερωτήματα ελάχιστης υποστήριξης

Σε αυτό το κομμάτι θα εξετάσουμε κατά πόσο μπορούν να βελτιωθούν ακόμα περισσότερο τα οφέλη, όταν ορίσουμε ένα ελάχιστο ποσό υποστήριξης (minimum support). Στο τμήμα VI-C συζητήσαμε γιατί αυτό μπορεί να συμβεί. Στα γραφήματα 16 και 17 απεικονίζουμε τον αριθμό ολικών παραβιάσεων, (παρακαλώ να σημειωθεί η διαφορά σε σύγκριση με όλα τα άλλα γραφήματα που απεικονίζονται) για το τρόπο λειτουργίας Proactive Broadcast Convex και για τα δεδομένα θερμοκρασίας από το Intel Lab Data και τα δεδομένα υγρασίας από το Weather Data αντίστοιχα. Κάνουμε τρεις σημαντικές παρατηρήσεις:

- Ο αριθμός ολικών παραβιάσεων, στους τρόπους λειτουργίας που δεν απεικονίζονται είναι παρόμοιος για τις ίδιες τιμές κατώτατου ορίου και minimum support . Δεδομένου ότι αυτοί οι εναλλακτικοί τρόποι λειτουργίας τροποποιούν το διάνυσμα εκτίμησης τους με διαφορετικούς τρόπους(δηλαδή, περισσότερο ή λιγότερο επιθετικά), ο αριθμός των ολικών παραβιάσεων δεν μπορεί να είναι ακριβώς ο ίδιος. Έτσι δεδομένου ότι ο αριθμός των ολικών παραβιάσεων είναι περίπου ο ίδιος σε όλους τους τρόπους λειτουργίας, τα οφέλη των τεχνικών (συγκρίνοντας τα μεταξύ τους) προέρχονται από τη μείωση της επικοινωνίας μέσα σε κάθε ζεύγος κόμβων.
- Ο αριθμός των ολικών παραβιάσεων, είναι μόνο ένα μικρό μέρος του συνολικού αριθμού των μεταδιδόμενων μηνυμάτων. Κατά συνέπεια, οι τεχνικές μας είναι λιγότερο ευαίσθητες από το NAIVE αλγόριθμο, όταν ο σταθμός βάσης τοποθετείται πιο μακριά από το cluster.
- Ο καθορισμός ενός minimum support μειώνει τον αριθμό των ολικών παραβιάσεων σε όλες τις περιπτώσεις. Προφανώς όσο μεγαλύτερο ποσό στήριξης επιλέξουμε τόσο ο αριθμός των μηνυμάτων μειώνετε. Πρέπει να σημειώσουμε ότι η εντός ζεύγους επικοινωνία δεν επηρεάζεται (μόνο την ανακοίνωση στον σταθμό βάσης μπορεί να επηρεάσει), πράγμα που σημαίνει ότι η συνολική μείωση των μεταδιδόμενων μηνυμάτων είναι μέτρια.



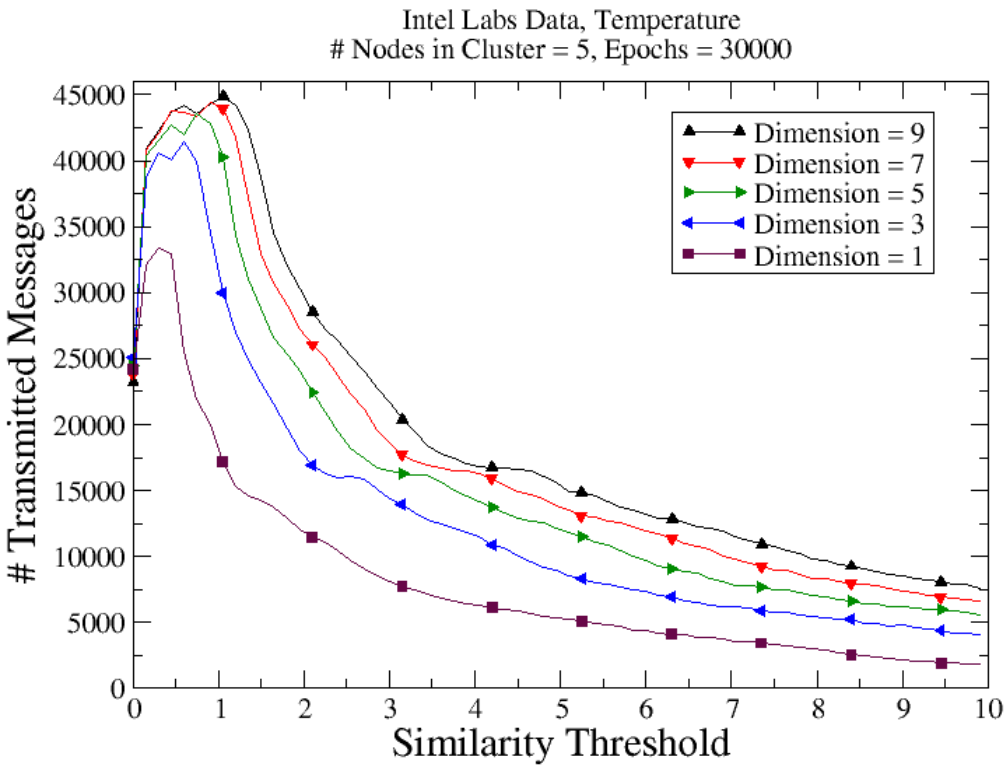
Γράφημα 16 : Intel Lab Data, θερμοκρασία : #Ολικών παραβιάσεων σε σχέση με το κατώτατο όριο, μεταβάλλοντας την ελάχιστη υποστήριξη, L_2 για την Proactive Broadcast Convex



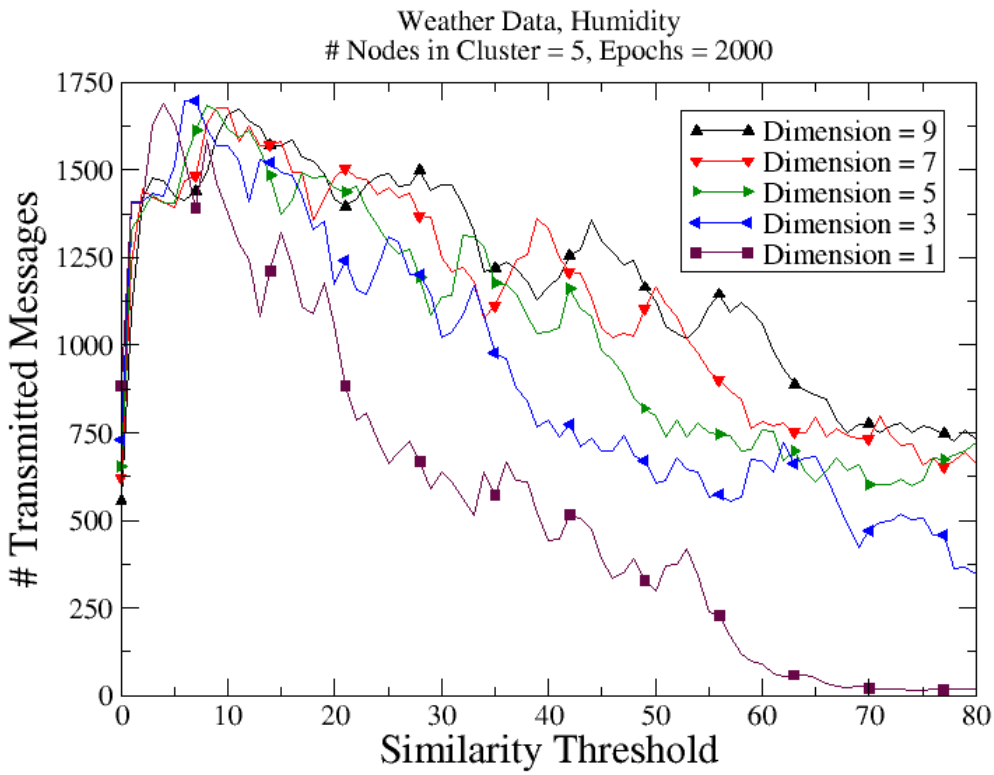
Γράφημα 17 : Weather Data, σχετική υγρασία : #Ολικών παραβιάσεων σε σχέση με το κατώτατο όριο, μεταβάλλοντας την ελάχιστη υποστήριξη, L_2 για την Proactive Broadcast Convex

C. Μεταβάλλοντας τη διάσταση διανύσματος W

Τα γραφήματα 18 και 19 απεικονίζουν τον αριθμό των μεταδιδόμενων μηνυμάτων για την L_2 μετρική ομοιότητας χρησιμοποιώντας Proactive Broadcast Convex λειτουργία κόμβου, καθώς μεταβάλουμε την διάσταση W του διανύσματος μετρήσεων, για δεδομένα θερμοκρασίας από το από το Intel Lab Data και υγρασίας από το Weather Data αντίστοιχα. Σε αυτήν την περίπτωση, η L_2 λειτουργία συσσωρεύει το άθροισμα περισσότερων όρων. Επομένως, αναμένεται ότι η μορφή των γραφικών παραστάσεων θα μετατοπιστεί βαθμιαία προς τις υψηλότερες τιμές κατώτατων ορίων(παραδείγματος χάριν, αν υποθέσουμε τα ίδια κατ' απόλυτη τιμή διανυσματικά στοιχεία σε όλες τις διαστάσεις, η μετρική L_2 αυξάνεται με τον αριθμό των διαστάσεων).



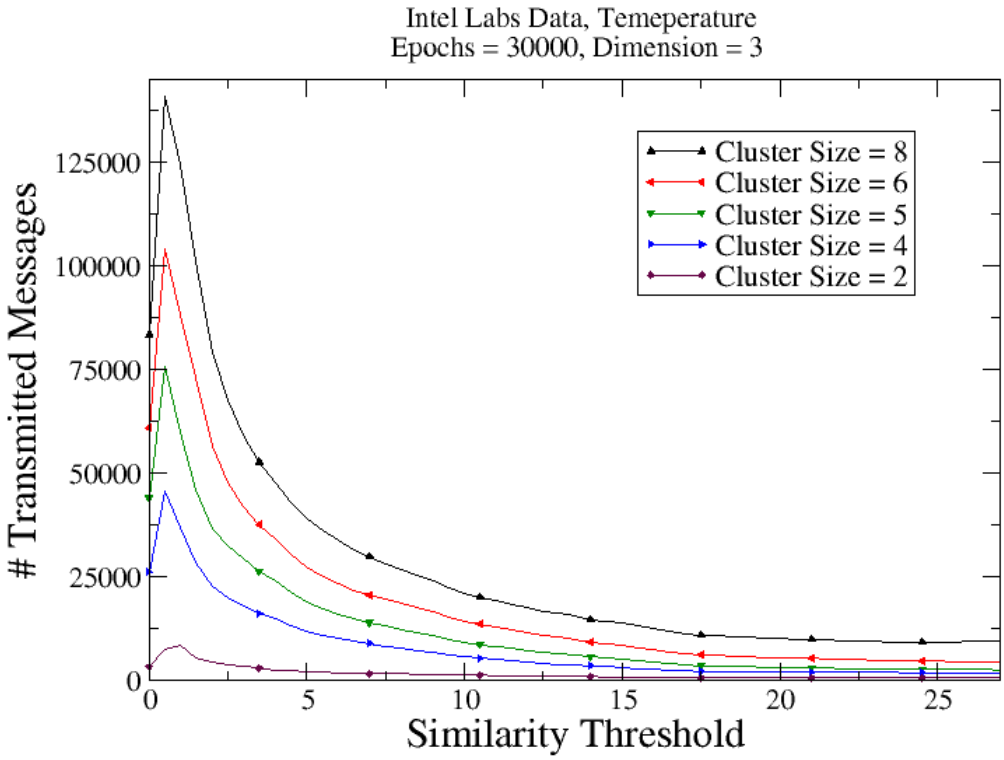
Γράφημα 18 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τη διάσταση του διανύσματος, L_2 για την Proactive Broadcast Convex



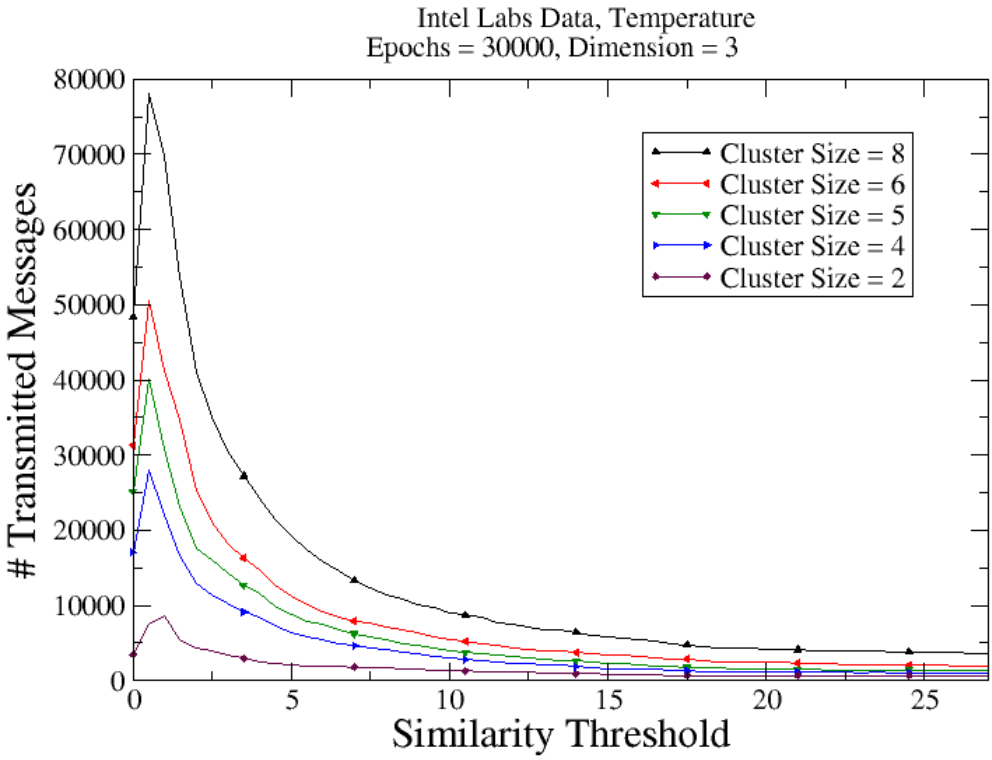
Γράφημα 19 : Weather Data, σχετική υγρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τη διάσταση του διανύσματος, L_2 για την Proactive Broadcast Convex

D. Μεταβάλλοντας τον αριθμό των κόμβων στο Cluster

Μελετάμε τώρα πώς οι τεχνικές μας επηρεάζονται κατά την αύξηση του μεγέθους του cluster. Για ένα cluster με k κόμβους, πρέπει να παρακολουθηθούν συνολικά οι $\binom{k}{2}$ σχέσεις ομοιότητας. Κατά συνέπεια, αναμένουμε μια αύξηση στον αριθμό διαβιβασθέντων μηνυμάτων όσο το μέγεθος του cluster αυξάνεται. Να σημειώσουμε εδώ ότι στη NAIVE τεχνική κλιμακώνεται γραμμικά με το μέγεθος του cluster (δηλαδή, για ένα cluster που περιέχει 8 κόμβους, για τα δεδομένα της Intel Lab Data, απαιτούνται 240000 μηνύματα) . Τα γραφήματα 20 και 21 απεικονίζουν την κλιμάκωση για τα δεδομένα θερμοκρασίας στο Intel Lab Data για τις λειτουργίες κόμβων Convex Autobalance και Proactive Broadcast Convex. Ακόμα και όταν αυξάνεται το μέγεθος του cluster, οι τεχνικές μας διατηρούν σημαντική εξοικονόμηση σε σχέση με την NAIVE προσέγγιση, με την Proactive Broadcast Convex λειτουργία κόμβου να είναι σε θέση να κλιμακώνεται καλύτερα.



Γράφημα 20 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τον αριθμό των κόμβων στο cluster, L_2 για την Convex Autobalance



Γράφημα 21 : Intel Lab Data, θερμοκρασία : #Μηνυμάτων σε σχέση με το κατώτατο όριο, μεταβάλλοντας τον αριθμό των κόμβων στο cluster, L_2 για την Proactive Broadcast Convex

VIII. ΣΥΜΠΕΡΑΣΜΑΤΑ

Σε αυτή την διπλωματική εργασία προτείναμε ένα νέο πλαίσιο για την ανίχνευση outlier στα δίκτυα αισθητήρων, που βασίζεται στη γεωμετρική προσέγγιση. Αποδείξαμε ότι πολλές κοινές μετρικές ομοιότητας μπορούν να μετασχηματιστούν και να εκφραστούν με τέτοιο τρόπο ώστε να επιτρέπεται η εφαρμογή της γεωμετρικής προσέγγισης. Παρουσιάσαμε έπειτα ένα γενικό πλαίσιο για την ανίχνευση outlier και προτείναμε πολλαπλούς τρόπους λειτουργίας για τους κόμβους αισθητήρων. Συζητήθηκαν επίσης και αξιολογήθηκαν κατάλληλες βελτιστοποιήσεις, όπως η λειτουργία Autobalance, η εκμετάλλευση της κυρτότητας ασφαλών ζωνών και τη δυνατότητα broadcast επικοινωνίας. Επιπλέον το πλαίσιο μας μπορεί εύκολα να χρησιμοποιηθεί για ερωτήματα ελάχιστης υποστήριξης. Ίσως το πιο σημαντικό στο πλαίσιο μας είναι ότι επιτρέπει σε κάθε κόμβο αισθητήρα να παρακολουθεί με ακρίβεια την ομοιότητά του με άλλους κόμβους, επιτρέποντας του να απέχει στη διαβίβαση μηνυμάτων σε κάποιες εποχές, ενώ η κατανάλωση εύρους ζώνης να είναι μόνο ένα μέρος του κόστους επικοινωνίας που απαιτεί μια συγκεντρωτική προσέγγιση.

Για τη μελλοντική εργασία, θα ερευνήσουμε πώς οι πρόσθετες λειτουργίες ομοιότητας μπορούν να ενσωματωθούν στο πλαίσιό μας. Επιπλέον για μερικές λειτουργίες ομοιότητας μπορεί να είναι δύσκολο για τον κόμβο αισθητήρα να βρει με ακρίβεια τη μέγιστη / ελάχιστη τιμή σε μια τοπική σφαίρα ελέγχου που κατασκευάζεται. Στις περιπτώσεις αυτές η εξέταση της ελάχιστης/μέγιστης τιμής της λειτουργίας, σε ένα σύνολο σημείων που ανήκουν σε ένα πλέγμα (που βρίσκεται μέσα στη σφαίρα) είναι μια εφικτή εναλλακτική λύση, αλλά δεν μπορεί πάντα να οδηγήσει σε τέλεια ακρίβεια. Ίσως, μια πιο ελπιδοφόρα προσπάθεια θα ήταν να ζητήσει από το σταθμό βάσης (που μπορεί να έχει περισσότερους υπολογιστικούς πόρους) όταν αυτό είναι δυνατόν, να υπολογίσει μια κατάλληλη κυρτή ασφαλή ζώνη την οποία οι αισθητήρες θα χρησιμοποιήσουν στη συνέχεια. Σχεδιάζουμε να ερευνήσουμε αυτές τις κατευθύνσεις στο μέλλον.

Γλωσσάρι

aggregate : Συνάθροιση αποτελεσμάτων
aggregate threshold queries : Συναθροίστηκα ερωτήματα κατώτατων ορίων
balance : διαδικασία εξισσορόπισης
broadcast : Όταν ο κόμβος είναι άμεσα συνδεδεμένος με όλους τους γείτονες
cluster : Ένα σύνολο από κόμβους
comparison neighbourhood : Γειτονία σύγκρισης
continuous queries : ενημερώνει συνεχώς το υπολογιζόμενο aggregate ώστε να αντικατοπτρίζει το σύνολο των τρεχουσών τιμών που έχουν οι αισθητήρες
convex hull : Κυρτή επιφάνεια που υπάρχει, ενώνοντας τις κορυφές διανυσμάτων(σημείων)
declarative queries : δηλωτικές ερωτήσεις
derived : επιπλέον διαστάσεις στο διάνυσμα μετρήσεων
direct : Άμεση επικοινωνία
error : Σφάλμα στον αισθητήρα
event : Κάποιο φαινόμενο που μας ενδιαφέρει να εντοπίσουμε
event-based : εφαρμογές που απαιτούν από τους κόμβους αισθητήρων να ανιχνεύουν ενδιαφέροντα γεγονότα
expanded : Τα κατασκευαζόμενα διανύσματα μαζί με τις επιπλέον διαστάσεις
global outlier : σφαιρική εξέχουσα τιμή
global statistics vector : μέσος όρος μεταξύ όλων των local statistics vector
global violation : ολική παραβίαση, εμφανίζεται όταν ένας κόμβος είναι σίγουρος ότι η ομοιότητα του με έναν άλλο κόμβο έχει αλλάξει χρώμα.
group-by queries : Ερωτήματα που ομαδοποιούνται με βάση κάποιο κοινό χαρακτηριστικό
local outlier: τοπική εξέχουσα τιμή
local violation : τοπική παραβίαση, εμφανίζεται όταν ένας κόμβος κρίνει ότι ο τοπικός περιορισμός παραβιάζεται. Στη περίπτωση που η σφαίρα ελέγχου $B(\vec{e}, \vec{u}_i)$, που παρακολουθεί δεν είναι μονοχρωματική.
minimum support : ελάχιστη υποστήριξη
multi-hop : Όταν σε ένας κόμβος σε ένα δίκτυο αισθητήρων είναι συνδεδεμένος πε περισσότερους κόμβους. Οι πληροφορίες μεταβιβάζονται από κόμβο σε κόμβο μέχρι να φτάσουν στο προορισμό τους.
outlier : Εξέχουσα τιμή
pipeline : διοχέτευσης
unicast : η αποστολή πληροφοριών από ένα αποστολέα σε ένα και μόνο παραλήπτη δικτύου.
unsupervised : χωρίς επίβλεψη
violation : Παραβίαση

Βιβλιογραφία :

[1] M. Bawa, H. Garcia-Molina, A. Gionis, and R. Motwani. Estimating Aggregates on a Peer-to-Peer Network. Technical report, Stanford, 2003.

[2] J. Branch, B. Szymanski, C. Giannella, R. Wolff, and H. Kargupta. Innetwork outlier detection in wireless sensor networks. In ICDCS , 2006.

[3] M. Chatterjee, S. Das, and D. Turgut. WCA: A Weighted Clustering Algorithm for Mobile Ad hoc Networks. Journal of Cluster Computing(Special Issue on Mobile Ad hoc Networks), 5, 2002.

[4] J. Chen, S. Kher, and A. Somani. Distributed Fault Detection of Wireless Sensor Networks. In DIWANS, 2006.

[5] J. Considine, M. Hadjieleftheriou, F. Li, J. Byers, and G. Kollios. Robust approximate aggregation in sensor data management systems. ACM Trans. Database Syst., 34(1):1–35, 2009.

[6] A. Deligiannakis, Y. Kotidis, V. Vassalos, V. Stoumpos, and A. Delis. Another Outlier Bites the Dust: Computing Meaningful Aggregates in Sensor Networks. In ICDE, 2009.

[7] E. Elnahrawy and B.Nath. Cleaning and querying noisy sensors. In WSNA, 2003.

[8] N. Giatrakos, Y. Kotidis, A. Deligiannakis, V. Vassalos, and Y. Theodoridis. TACO: Tunable Approximate Computation of Outliers in Wireless Sensor Networks. In SIGMOD, 2010.

[9] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-Efficient Communication Protocol for Wireless Microsensor Networks.In HICSS, 2000.

[10] S. Jeffery, G. Alonso, M. J. Franklin, W. Hong, and J. Widom. Declarative Support for Sensor Data Cleaning. In Pervasive, 2006.

[11] S. Jeffery, G. Alonso, M. J. Franklin, W. Hong, and J. Widom. A pipelined framework for online cleaning of sensor data streams. In ICDE, 2006.

[12] S. Jeffery, M. Garofalakis, and M. Franklin. Adaptive Cleaning for RFID Data Streams. In VLDB, 2006.

[13] D. Kempe, A. Dobra, and J. Gehrke. Gossip-Based Computation of Aggregate Information. In FOCS, 2003.

[14] D. Keren, I. Sharfman, A. Schuster, and A. Livne. Shape sensitive geometric monitoring. IEEE Transactions on Knowledge and Data Engineering (to appear).

[15] N. Khoussainova, M. Balazinska, and D. Suciu. Towards Correcting Input Data Errors Probabilistically using Integrity Constraints. In MobiDE, 2006.

[16] Y. Kotidis. Snapshot Queries: Towards Data-Centric Sensor Networks. In ICDE, 2005.

[17] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TAG: A Tiny Aggregation Service for ad hoc Sensor Networks. In OSDI, 2002.

[18] S. Meng, K. Xie, G. Chen, X. Ma, and G. Song. A kalman filter based approach for outlier detection in sensor networks. In CSSE , 2008.

[19] M. Otey, A. Ghoting, and S. Parthasarathy. Fast distributed outlier detection in mixed-attribute data sets. Data Min. Knowl. Discov., 12(2-3), 2006.

[20] M. Qin and R. Zimmermann. VCA: An Energy-Efficient Voting-Based Clustering Algorithm for Sensor Networks. J.UCS, 13(1), 2007.

[21] G. Sagy, D. Keren, I. Sharfman, and A. Schuster. Distributed threshold querying of general functions by a difference of monotonic representation.Proc. VLDB Endow., 4, November 2010.
[22] I. Sharfman, A. Schuster, and D. Keren. A geometric approach to monitoring threshold functions over distributed data streams. In SIGMOD, 2006.
[23] I. Sharfman, A. Schuster, and D. Keren. Aggregate threshold queries in sensor networks. In IPDPS, 2007.
[24] I. Sharfman, A. Schuster, and D. Keren. Shape sensitive geometric monitoring. In PODS, 2008.
[25] B. Sheng, Q. Li, W. Mao, and W. Jin. Outlier detection in sensor networks. In MobiHoc, 2007.
[26] S. Singh, M. Woo, and C. S. Raghavendra. Power-aware Routing in Mobile Ad Hoc Networks. In MobiCom, 1998.
[27] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos. Online Outlier Detection in Sensor Data Using Non- Parametric Models. In VLDB, 2006.
[28] Y.-J. Wen, A. M. Agogino, and K.Goebel. Fuzzy Validation and Fusion for Wireless Sensor Networks. In ASME, 2004.
[29] X. Xiao, W. Peng, C. Hung, and W. Lee. Using SensorRanks for In-Network Detection of Faulty Readings in Wireless Sensor Networks. In MobiDE, 2007.
[30] Y. Yao and J. Gehrke. The Cougar Approach to In-Network Query Processing in Sensor Networks. SIGMOD Record, 31(3), 2002.
[31] O. Younis and S. Fahmy. Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach. In INFOCOM, 2004.
[32] D. Zeinalipour, P. Andreou, P. Chrysanthis, G. Samaras, and A. Pitsillides. The Micropulse Framework for Adaptive Waking Windows in Sensor Networks. In MDM, 2007.
[33] K. Zhang, S. Shi, H. Gao, and J. Li. Unsupervised outlier detection in sensor networks using aggregation tree. In ADMA , 2007.
[34] Y. Zhang, N. Meratnia, and P. Havinga. Outlier detection techniques for wireless sensor networks: A survey. International Journal of IEEE Communications Surveys and Tutorials, 12(2), 2010.
[35] Y. Zhuang and L. Chen. In-network Outlier Cleaning for Data Collection in Sensor Networks. In Clean DB Workshop,2006.
[36] Y. Zhuang, L. Chen, S.Wang, and J. Lian. A Weighted Moving Averagebased Approach for Cleaning Sensor Data. In ICDCS, 2007
[37] http://images.devshed.com/wh/stories/Sensor%20Networks%20Intro/mica2_image.jpg
[38] http://www.tinyos.net/hardware/mica-sensorboard.jpg
[39] V. Chandola, A. Banerjee, and V. Kumar, Anomaly Detection: A Survey, Technical Report, University of Minnesota, 2007.
[40] http://courses.ece.tuc.gr/ , Σημειώσεις μαθήματος, Επεξεργασία και Διαχείριση Δεδομένων σε Δίκτυα Αισθητήρων
[41] http://www.cs.technion.ac.il/~assaf/talks/SEBD-Monitoring06.ppt
[42] http://db.csail.mit.edu/labdata/labdata.html
[43] http://www-k12.atmos.washington.edu/k12/grayskies/