ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ανίχνευση ανωμαλιών σε δίκτυα μεγάλης κλίμακας με χρήση Ανάλυσης Κυρίων Συνιστωσών

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Βασίλης Χ. Χατζηγιαννάκης

Αθήνα, Μάιος 2009

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ανίχνευση ανωμαλιών σε δίκτυα μεγάλης κλίμακας με χρήση Ανάλυσης Κυρίων Συνιστωσών

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Βασίλης Χ. Χατζηγιαννάκης

Συμβουλευτική Επιτροπή : Συμεών Παπαβασιλείου, Επίκ. Καθηγητής ΕΜΠ
Βασίλης Μάγκλαρης, Καθηγητής ΕΜΠ
Ευστάθιος Συκάς, Καθηγητής ΕΜΠ

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την:

…… Συμεών Παπαβασιλείου Επίκ. Καθηγητής ΕΜΠ	…… Βασίλης Μάγκλαρης Καθηγητής ΕΜΠ	…… Ευστάθιος Συκάς Καθηγητής ΕΜΠ
…… Μιχαήλ Θεολόγου Καθηγητής ΕΜΠ	…… Νικόλαος Μήτρου Καθηγητής ΕΜΠ	…… Ιάκωβος Βενιέρης Καθηγητής ΕΜΠ

……
Χρήστος Δουληγέρης
Καθηγητής Παν. Πειραιώς

Αθήνα, Μάιος 2009

.....

Βασίλης Χ. Χατζηγιαννάκης

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Μια από τις σημαντικότερες προκλήσεις στην διαχείριση δικτύων μεγάλης κλίμακας είναι η ανίχνευση ανωμαλιών στη δικτυακή κίνηση εξαιτίας απειλών όπως οι Επιθέσεις Απάρνησης Υπηρεσίας ή οι αυτομεταδιδόμενοι ιοί. Συνήθως οι μεθοδολογίες ανίχνευσης ανωμαλιών στηρίζονται στην ανάλυση του δικτύου και το χαρακτηρισμό στατιστικών ιδιοτήτων που αντιστοιχούν σε κίνηση που δεν εμπεριέχει ανωμαλίες. Με βάση το πρότυπο ότι αποκλίσεις από την κανονική συμπεριφορά υποδηλώνουν την ύπαρξη ανωμαλιών, όπως λάθη και επιθέσεις, οι μεθοδολογίες ανίχνευσης ανωμαλιών μπορούν να εφαρμοστούν καθολικά ακόμα και για την ανίχνευση νέων επιθέσεων των οποίων η φύση είναι άγνωστη. Στην παρούσα διδακτορική διατριβή θεωρούμε ένα περιβάλλον πολλαπλών κατανεμημένων αισθητήρων που ανταλλάσσουν πληροφορίες μέσω μιας κοινής πλατφόρμας, και υποστηρίζουν την διαδικασία ανίχνευσης ανωμαλιών. Η προτεινόμενη προσέγγιση βασίζεται στην εφαρμογή της Ανάλυσης Κυρίων Συνιστωσών σε δεδομένα που προέρχονται από πολλαπλά μετρικά και πολλαπλές ζεύξεις/κόμβους του δικτύου και προσφέρει ένα αποτελεσματικό και ενιαίο τρόπο συνδυασμού συσχετισμένων δεδομένων για την ανίχνευση ανωμαλιών. Στην ουσία παρουσιάζεται μια γενικευμένη μεθοδολογία ανίχνευσης ανωμαλιών που δεν έχει μόνο δυνατότητα ανίχνευσης ανωμαλιών που επηρεάζουν τον όγκο της κίνησης, αλλά ανιχνεύει ένα μεγαλύτερο εύρος επιθέσεων όπως αυτές που επηρεάζουν τη σύσταση της κίνησης σε ένα ή περισσότερα δικτυακά μονοπάτια. Η προτεινόμενη μεθοδολογία συνδυάζει αποτελεσματικά συσχετισμένα δεδομένα με σκοπό την αποκάλυψη ανωμαλιών που διατρέχουν ένα ή περισσότερα γειτονικά στοιχεία του δικτύου, και επιτρέπει την αναγνώριση των δικτυακών μονοπατιών στο γράφο του δικτύου που περιέχουν την ανωμαλία. Στα πλαίσια της παρούσας διατριβής, η προτεινόμενη μεθοδολογία εφαρμόζεται σε διαφορετικά είδη δικτύων που αναμένεται να είναι αναπόσπαστα τμήματα των μελλοντικών υποδομών του Διαδικτύου, όπως είναι τα δίκτυα ευρείας ζώνης, τα Ασύρματα Δίκτυα Αισθητήρων και τα υβριδικά Δίκτυα Αυτοκίνησης. Σε κάθε περίπτωση, αναλύεται η εφαρμογή της προτεινόμενης μεθόδου λαμβάνοντας υπόψη τα ιδιαίτερα χαρακτηριστικά κάθε δικτύου, και μελετάται η αποτελεσματικότητα της μεθοδολογίας μέσω προσομοίωσης ή/και εξομοίωσης. Η προτεινόμενη μέθοδος μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα υπηρεσιών

ανίχνευσης ανωμαλιών, όπως η αναγνώριση ανωμαλιών σε μια σειρά μετρήσεων, η αποκάλυψη ύπαρξης ελαττωματικών αισθητήρων, η ανίχνευση πιθανών δικτυακών επιθέσεων και το φιλτράρισμα ύποπτων αναφορών κατά την διαδικασία εξαγωγής αποφάσεων.

Abstract

One of the main challenges in security management of large scale networks is the detection of suspicious anomalies in network traffic patterns due to threats such as Distributed Denial of Service (DDoS) attacks or worm propagation. Usually network anomaly detection methodologies rely on the analysis of network traffic and the characterization of the dynamic statistical properties of traffic normality. Anomaly detection is based on the concept that perturbations of normal behavior suggest the presence of anomalies such as faults and attacks and can be uniformly applied in order to detect network attacks, even in cases where novel attacks are present and the nature of the intrusion is unknown.

In this context, we envisage a variety of distributed monitors that exchange information through a common automated platform and assist the anomaly detection algorithms by sharing their views. The proposed approach is based on the application of Principal Component Analysis on multi-metric-multi-link data, and provides an efficient and unified way of taking into account the combined effect of the correlated observed data, for anomaly detection purposes. It actually introduces a generalized anomaly detection methodology, capable of detecting not only volume based anomalies, but a much wider range of classes of anomalies, such as the ones that may result in alterations in traffic composition or traffic paths. The proposed methodology effectively combines correlated data in order to reveal anomalies that span through a number of neighboring elements in the network graph and allows for the identification of the corresponding paths that contain the anomaly.

Throughout this thesis, the proposed anomaly detection methodology is applied in different kinds of large scale networks that present different chracteristics and application environments, and are quite representative of the networks that will compose the Internet of the Future, such as high speed networks, wireless sensor networks and hybrid vehicular networks. In each case, the application of the proposed methodology within the specific environemnt is discussed, and its effectiveness is evaluated via simulation or/and emulation.

Throughout this thesis it is demonstrated that such an approach can be used in principle for several anomaly detection applications and scenarios, including: identify an abnormal situation in a series of measurements (e.g. cases where the values of the

measured or monitored parameters may deviate significantly from the norm), discover the existence of faulty sensors in a sensor network, detect potential network attacks, and/or filter suspicious reports throughout the overall decision making process.

Ευχαριστίες

Κατά τη διάρκεια της εκπόνησης της διδακτορικής διατριβής μου ήταν πολλοί οι άνθρωποι που με στήριξαν και με βοήθησαν και προέρχονται τόσο από το ακαδημαϊκό και τον εργασιακό χώρο, όσο και από το φιλικό και συγγενικό μου περιβάλλον.

Θα ήθελα να ευχαριστήσω μέσα από την καρδιά μου τον επιβλέποντα καθηγητή μου κ. Συμεών Παπαβασιλείου για την ουσιαστική υποστήριξη που μου προσέφερε. Η συνεργασία μας όλα αυτά τα χρόνια ήταν άψογη και η επίβλεψη του διδακτορικού έγινε με μεράκι. Επίσης, θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου κ. Βασίλη Μάγκλαρη για την βοήθεια του αλλά και την εμπιστοσύνη που μου έδειξε όταν με δέχθηκε ως υποψήφιο διδάκτορα στο εργαστήριο Διαχείρισης και Βέλτιστου Σχεδιασμού Δικτύων (Netmode) αλλά και ως νεαρό εργαζόμενο στο Κέντρο Δικτύων του ΕΜΠ, καθώς και για τη διπλωματική μου εργασία που επέβλεψε η οποία ήταν προπομπός της διδακτορικής διατριβής. Η πορεία αυτή θα ήταν πολύ πιο δύσβατη χωρίς την άψογη συνεργασία και τις γόνιμες συζητήσεις με τα υπόλοιπα μέλη του Netmode και ιδιαίτερα τον Γιώργο Ανδρουλιδάκη και τη Μαίρη Γραμματικού. Σημαντική συμβολή και υποστήριξη μου προσέφεραν οι επιστημονικοί υπεύθυνοι, συνάδελφοι και φίλοι μου στο Κέντρο Δικτύων του ΕΜΠ τόσο για το ευχάριστο εργασιακό περιβάλλον όσο και για τη γνώση, την εμπειρία αλλά και την πρόσβαση σε δικτυακές υποδομές που ήταν σημαντικότατες για την διεξαγωγή της έρευνάς μου.

Στη συνέχεια, θα ήθελα να ευχαριστήσω τους Γιώργο Μπράβο, Γιώργο Δημητρακόπουλο, Αντώνη Ζήσιμο και Δημήτρη Αλεξόπουλο με τους οποίους με συνδέουν άρρηκτοι δεσμοί φιλίας, για τις αξέχαστες στιγμές και την κοινή πορεία που είχαμε από τα φοιτητικά μας χρόνια.

Τέλος, θα ήθελα να ευχαριστήσω την αδερφή μου Βάνια, τη Ροδούλα, και κυρίως τους γονείς μου Χρήστο και Άννα για την ανεκτίμητη υποστήριξη και φροντίδα που μου έχουν προσφέρει. Τους αγαπώ και τους ευχαριστώ μέσα από την καρδιά μου.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

1. Εισαγωγή

Το διαδίκτυο έχει αναπτυχθεί με ραγδαίους ρυθμούς τα τελευταία χρόνια. Σχεδόν καθημερινά επεκτείνεται σε μέγεθος ενώ αναπτύσσονται καινούριες εφαρμογές οι οποίες έχουν συνεχώς αυξανόμενες απαιτήσεις, τόσο σε εύρος ζώνης όσο και σε ποιότητα της προσφερόμενης υπηρεσίας. Με τη δημιουργία του Παγκόσμιου Ιστού προσφέρθηκε ελεύθερη πρόσβαση σε όλο και περισσότερους χρήστες με αποτέλεσμα την εμφάνιση νέων εφαρμογών όπως η τηλεδιάσκεψη, η τηλεφωνία μέσω δικτύου και το ηλεκτρονικό εμπόριο.

Καθώς λοιπόν το διαδίκτυο εισβάλει όλο και περισσότερο στη ζωή μας γίνεται επιτακτική η ανάγκη για συνεχή και ασφαλή λειτουργία των κρίσιμων κόμβων και υπηρεσιών. Παράλληλα με την αύξηση των χρηστών αυξάνονται και αυτοί που είναι διατεθειμένοι να κάνουν κακή χρήση του δικτύου για την εξυπηρέτηση οικονομικών, πολιτικών ή άλλου είδους συμφερόντων. Οι μεγαλύτερες απειλές στο διαδίκτυο είναι οι επιθέσεις απάρνησης υπηρεσίας (Denial of Service – DoS) και οι αυτό-μεταδιδόμενοι ιοί (worms). Ο σκοπός των επιθέσεων απάρνησης υπηρεσίας είναι είτε να πλημμυρίσουν το δίκτυο-στόχο με κίνηση είτε να καταναλώσουν όλους τους διαθέσιμους πόρους ενός εξυπηρετητή με «νόμιμα» αλλά συνεχή αιτήματα. Χρησιμοποιούν συνήθως ένα ευρύ φάσμα από «μολυσμένους» ηλεκτρονικούς υπολογιστές στους οποίους έχει νωρίτερα εγκατασταθεί κακόβουλος κώδικας, όπως ένας ιός. Όταν ο κώδικας αυτός ενεργοποιηθεί, οι μολυσμένοι υπολογιστές εξαπολύουν μια μαζική ροή από πακέτα επίθεσης. Οι ροές των πακέτων που απαρτίζουν την επίθεση είναι ιδιαίτερα δύσκολο να ανιχνευτούν στα αρχικά στάδια της. Αυτό συμβαίνει γιατί ποσοτικά κάθε ροή μπορεί να είναι πολύ μικρή σε σχέση με το συνολικό άθροισμα των πακέτων που περνούν από τα μεγάλα ενδιάμεσα δίκτυα. Φτάνοντας όμως στον στόχο τους προερχόμενες από διάφορα μέρη του διαδικτύου, οι ροές αυτές έχουν συχνά καταστρεπτικό αποτέλεσμα για τον τελικό στόχο. Ένα άλλο χαρακτηριστικό είναι ότι ακόμα και όταν γίνει αντιληπτή μια επίθεση (π.χ. τύπου DoS), είναι συχνά αδύνατο να εντοπιστούν οι υπολογιστές που τη δημιουργούν γιατί αφενός η διεύθυνση IP πηγής στα πακέτα της επίθεσης είναι συνήθως ψεύτικη και αφετέρου οι υπολογιστές αυτοί είναι διασκορπισμένοι στο διαδίκτυο.

Εξίσου επικίνδυνοι για ένα δίκτυο ευρείας ζώνης είναι και οι αυτό-μεταδιδόμενοι ιοί. Εκμεταλλευόμενοι κάποια «τρύπα ασφαλείας» τα προγράμματα αυτά μπορούν να

εξαπλωθούν από υπολογιστή σε υπολογιστή και μετά να χρησιμοποιηθούν από το δημιουργό τους για κακόβουλες ενέργειες, όπως μια κατανεμημένη επίθεση απάρνησης υπηρεσίας. Η άμυνα ενάντια σε μια κατανεμημένη επίθεση απάρνησης υπηρεσίας είναι μια διαδικασία πολλών βημάτων. Το πρόβλημα είναι ότι το δίκτυο που δέχεται την επίθεση δεν έχει τη δυνατότητα να ελέγξει την εισερχόμενη κυκλοφορία. Ο διαχειριστής πρέπει να καθορίσει τα χαρακτηριστικά της επίθεσης και να έρθει σε επαφή με τους διαχειριστές των γειτονικών δικτύων από τα οποία προέρχεται η επίθεση. Συνεπώς οι επιθέσεις αυτές είναι μια απειλή η οποία δεν μπορεί να αντιμετωπιστεί αποτελεσματικά από ένα μεμονωμένο δίκτυο. Η επίθεση πρέπει ιδανικά να ανιχνεύεται κοντά στις πηγές της, εκεί όμως που οι ροές των κακόβουλων πακέτων είναι ακόμα μικρές και δύσκολα ανιχνεύσιμες.

Στόχος της διατριβής αυτής είναι η ανάπτυξη μιας ολοκληρωμένης μεθοδολογίας ανίχνευσης ανωμαλιών η οποία θα μπορεί, με κατάλληλες ρυθμίσεις και παραμετροποιήσεις να εφαρμοστεί σε διαφορετικού τύπου δίκτυα μεγάλης κλίμακας, που αναμένεται να είναι αναπόσπαστα τμήματα των μελλοντικών υποδομών του Διαδικτύου (Internet of the Future). Τέτοια παραδείγματα, που αξίζει να σημειωθεί ότι μπορεί να παρουσιάζουν διαφορετικά μοντέλα λειτουργίας και να υποστηρίζουν διαφορετικές υπηρεσίες, αποτελούν και τα δίκτυα αισθητήρων.

Ένα Ασύρματο Δίκτυο Αισθητήρων - ΑΔΑ (Wireless Sensor Network - WSN) συνδυάζει την ανίχνευση, την επεξεργασία σήματος και την ασύρματη επικοινωνία, ώστε να παρέχει μια πλατφόρμα για την ιεραρχική και αποδοτική συλλογή και επεξεργασία πληροφοριών. Σε ένα ασύρματο δίκτυο αισθητήρων τα δεδομένα που συλλέγονται υποβάλλονται σε επεξεργασία σε διαφορετικά επίπεδα λεπτομέρειας, που κυμαίνεται από τη λεπτομερή μικροσκοπική εξέταση συγκεκριμένων στόχων ως μια μακροσκοπική άποψη της συνολικής συμπεριφοράς του περιβάλλοντος το οποίο ελέγχεται. Ένα κατανεμημένο δίκτυο αισθητήρων είναι συνήθως ένα αυτό-οργανωμένο σύστημα που αποτελείται από ένα μεγάλο αριθμό κόμβων-αισθητήρων που συνεργάζονται ο ένας με τον άλλον για τη μέτρηση πολλαπλών διαφορετικών παραμέτρων που διαφοροποιούνται στο χρόνο και το χώρο, και την αποστολή των αντίστοιχων στοιχείων σε ένα κεντρικό κόμβο για περαιτέρω επεξεργασία.

Σε αντίθεση με τα παραδοσιακά ασύρματα δίκτυα, στα οποία η επικοινωνία γίνεται από πρόσωπο σε πρόσωπο και τα περιεχόμενα διαφορετικών συνομιλιών είναι

ασυσχέτιστα, στα δίκτυα αισθητήρων, τα δεδομένα στους γειτονικούς κόμβους θεωρούνται ιδιαίτερα συσχετισμένα. Τυπικές εφαρμογές στα ΑΔΑ απαιτούν αρκετά μεγάλη πυκνότητα αισθητήρων στο χώρο προκειμένου να επιτευχθεί ικανοποιητική κάλυψη και κατά συνέπεια τα δεδομένα που παράγονται είναι ιδιαίτερα συσχετισμένα. Ένα παράδειγμα μιας τέτοιας εφαρμογής είναι η μετεωρολογία, όπου οι μετρήσεις σε στοιχεία όπως η υγρασία, η θερμοκρασία και η βαρομετρική πίεση σε γειτονικούς κόμβους φθάνουν σχεδόν σε συσχέτιση 100%. Στα δίκτυα που παρουσιάζουν αυτά τα χαρακτηριστικά χρησιμοποιείται συχνά η έννοια της συνάθροισης δεδομένων. Μετρήσεις από γειτονικούς κόμβους συναθροίζονται ώστε να χρειάζεται να αποσταλούν στο κεντρικό κόμβο μικρότερος αριθμός μηνυμάτων και να γίνεται οικονομία στους πόρους του δικτύου.

Λαμβάνοντας υπόψη τα παραπάνω λειτουργικά χαρακτηριστικά των δικτύων αισθητήρων καθώς επίσης και την κρισιμότητα των εφαρμογών που υποστηρίζουν (π.χ. έλεγχος και προστασία περιβαλλοντικών, στρατιωτικών και άλλων κρίσιμων υποδομών και πόρων), προβλήματα στην ακεραιότητα και την ακρίβεια των δεδομένων που μπορούν να προκληθούν από προβληματικούς κόμβους είναι εξαιρετικής σημασίας και ερευνητικού ενδιαφέροντος. Πιο συγκεκριμένα οι κόμβοι που δυσλειτουργούν επειδή κάποιος τρίτος έχει αποκτήσει τον έλεγχό τους μπορούν να επιτεθούν και να απορυθμίσουν επιμέρους λειτουργίες του δικτύου όπως η δρομολόγηση, η συνάθροιση δεδομένων και η δίκαιη κατανομή των πόρων.

Συμπερασματικά, για ένα ευρύ φάσμα δικτύων, βασικός στόχος ενός διαχειριστή είναι η εύρεση και αντιμετώπιση προβλημάτων και αδυναμιών στην ασφάλεια και πιθανών δυσλειτουργιών που μπορεί να έχουν τα συστήματα για τα οποία είναι υπεύθυνος. Η σωστή ρύθμιση, η τακτική ενημέρωση και αναβάθμιση όμως δυστυχώς δεν καθιστούν ένα σύστημα άτρωτο σε επιθέσεις. Για παράδειγμα, οι επιθέσεις απάρνησης υπηρεσίας δεν στοχεύουν σε κάποια συγκεκριμένη αδυναμία ενός συστήματος αλλά βομβαρδίζουν το στόχο με συνεχή αλλά φαινομενικά νόμιμα αιτήματα με σκοπό την δέσμευση πόρων του συστήματος αλλά και του δικτύου. Στις περιπτώσεις αυτές πρέπει τουλάχιστον να υπάρχει ένα εργαλείο ανίχνευσης επιθέσεων προς το διαχειριζόμενο σύστημα ή δίκτυο. Το εργαλείο αυτό πρέπει να είναι σε θέση να ειδοποιήσει τον διαχειριστή όταν γίνεται κάποια επίθεση προκειμένου αυτός να προσπαθήσει να την αντιμετωπίσει έγκαιρα και να προστατέψει το στόχο της επίθεσης.

Στην επόμενη ενότητα γίνεται εκτενής περιγραφή στις αρχές και τα πρότυπα λειτουργίας των εργαλείων αυτών, τα οποία ονομάζονται συστήματα ανίχνευσης εισβολής.

1.1. Συστήματα Ανίχνευσης Εισβολής

Η ανίχνευση εισβολής (Intrusion Detection) αποτελεί μια προσέγγιση για την παροχή ασφάλειας στους υπάρχοντες υπολογιστές και τα δίκτυα δεδομένων, επιτρέποντας παράλληλα σε αυτά να λειτουργήσουν χωρίς περιορισμούς όσον αφορά στην πρόσβαση στις υπηρεσίες που προσφέρουν.

Ο στόχος της ανίχνευσης εισβολής είναι να προσδιοριστεί, κατά προτίμηση σε πραγματικό χρόνο, η αναρμόδια χρήση, η κακή χρήση και η κατάχρηση των συστημάτων ηλεκτρονικών υπολογιστών – κόμβων του δικτύου- και από τα ίδια τα μέλη των συστημάτων αλλά και από εξωτερικούς διεισδύοντες. Το πρόβλημα αυτό γίνεται μια πρόκληση καθώς η συνεχώς αυξανόμενη διαδικτύωση των ηλεκτρονικών υπολογιστών δίνει μεγαλύτερη πρόσβαση στους εισβολείς και τους διευκολύνει να καλύπτουν τα ίχνη τους.

Υπάρχουν δύο κύριες ταξινομήσεις των συστημάτων ανίχνευσης εισβολής. Η πρώτη κατηγοριοποίηση διαιρεί τις τεχνικές ανίχνευσης εισβολής σε δύο κύριους τύπους: ανίχνευση ανωμαλίας (anomaly detection) και ανίχνευση κακής χρήσης (misuse detection). Το πρότυπο ανίχνευσης ανωμαλίας στηρίζεται στη συγκέντρωση ενός συνόλου στατιστικών μετρικών που χαρακτηρίζουν τη συμπεριφορά μιας οντότητας. Οντότητα μπορεί να είναι ένας χρήστης, μια ομάδα χρηστών ή ένας άλλος υπολογιστής. Το προφίλ μιας οντότητας χρηστών για παράδειγμα, μπορεί να περιλαμβάνει πληροφορίες όπως η μέση διάρκεια των συνόδων υπηρεσιών Telnet και FTP, το ποσό δεδομένων που μεταδίδονται και προς τις δύο κατευθύνσεις, τις ώρες της ημέρας ή τα τερματικά από τα οποία συνδέεται κλπ. Το προφίλ ενός υπολογιστή ή πιο συγκεκριμένα ενός προγράμματος – υπηρεσίας που τρέχει σε κάποιον υπολογιστή, μπορεί να περιλαμβάνει τη μέση χρησιμοποίηση της κεντρικής μονάδας επεξεργασίας ή της μνήμης, τον μέσο αριθμό συνδεδεμένων χρηστών, κλπ. Σε ένα δίκτυο αισθητήρων, το προφίλ ενός κόμβου του δικτύου θα μπορούσε να περιγράφεται από παραμέτρους όπως η ακτίνα, ο ρυθμός εκπομπής μηνυμάτων, η μέση κατανάλωση ενέργειας, κ.λ.π.

Ένα σύστημα ανίχνευσης εισβολή που λειτουργεί με βάση το πρότυπο ανίχνευσης ανωμαλίας μπορεί να ελέγχει τη λειτουργία ενός υπολογιστικού συστήματος, και να συγκρίνει συνεχώς το προφίλ της τρέχουσας συνόδου του χρήστη, με αυτήν που είναι αποθηκεμένη στη βάση δεδομένων του. Σε περίπτωση που ανιχνεύσει μια «μεγάλη» απόκλιση από την κανονική συμπεριφορά επισημαίνει έναν συναγερμό στον αντίστοιχο διαχειριστή ασφάλειας του συστήματος. Το μέγεθος της απόκλισης συγκρίνεται με κάποιο ανώτατο όριο που τίθεται αυτόματα από το σύστημα ή το διαχειριστή με βάση την εμπειρία του. Συνήθως τα αποθηκευμένα προφίλ ενημερώνονται συνεχώς προκειμένου να απεικονιστούν οι αλλαγές στη συμπεριφορά χρηστών ή του συστήματος. Δεδομένου ότι αυτό το πρότυπο λειτουργεί ελέγχοντας συμπεριφορές που αποκλίνουν από τις κανονικές, καλείται πρότυπο ανίχνευσης ανωμαλίας.

Το πρότυπο ανίχνευσης κακής χρήσης αφ' ετέρου λειτουργεί με βάση ένα σύνολο γνωστών επιθέσεων που έχουν αποθηκευτεί στη βάση δεδομένων του συστήματος. Η γνώση των επιθέσεων κωδικοποιείται ως ένα σύνολο από «υπογραφές» (signatures) επιθέσεων, οι οποίες είναι ουσιαστικά ακολουθίες, χνάρια που εμφανίζονται κάθε φορά που πραγματοποιείται μια επίθεση. Ο τρόπος που μια γνωστή επίθεση αντιπροσωπεύεται στο σύστημα είναι ένα σημαντικό χαρακτηριστικό της λειτουργίας του. Ο τρόπος που αυτό το πρότυπο λειτουργεί είναι παρόμοιος με αυτόν ενός προγράμματος anti-virus. Η εφαρμογή ενός τέτοιου συστήματος ασφαλείας περιλαμβάνει συνήθως ένα έμπειρο σύστημα (expert system) που εκτελεί τη σύγκριση της τρέχουσας κατάστασης του συστήματος με κανόνες αποθηκευμένους σε μια βάση. Μια προφανής δυσκολία σε αυτήν την αρχιτεκτονική είναι η ανάγκη για τη σταθερή ενημέρωση της βάσης όταν αποκαλύπτονται νέες μέθοδοι επίθεσης, καθώς επίσης και το γεγονός ότι αδυνατεί να ανιχνεύσει καινοτόμες επιθέσεις που δεν είναι γνωστές και δεν έχουν αποθηκευτεί οι υπογραφές τους. Δεδομένου ότι το πρότυπο λειτουργεί με την έρευνα για χνάρια που είναι αντιπροσωπευτικά διαφόρων επιθέσεων, αναφέρεται ως πρότυπο ανίχνευσης κακής χρήσης.

Η δεύτερη ταξινόμηση είναι βασισμένη στο εάν το σύστημα ανίχνευσης εισβολής ελέγχει τη δραστηριότητα σε έναν μοναδικό υπολογιστή ή στους διάφορους κόμβους και στοιχεία ενός δικτύου. Τα πρώτα συστήματα ανίχνευσης εισβολής συνήθιζαν να εξετάζουν στοιχεία σε μια μεμονωμένη μηχανή και να παράγουν τα συμπεράσματά τους βασισμένα απλώς σε εκείνες τις πληροφορίες. Συνεπώς, δεν θα μπορούσαν να

ανιχνεύσουν τις επιθέσεις που παρήχθησαν από πολλές πηγές, ή τις επιθέσεις που εκτείνονται σε πολλές μηχανές σε ένα δίκτυο. Επιπλέον, στηρίζονται σε μεγάλο ποσοστό στα αρχεία καταγραφής του συστήματος (log files) που παρέχονται από το λειτουργικό σύστημα του μηχανήματος. Το γεγονός αυτό τα καθιστά αρχιτεκτονικά εξαρτώμενα και πιο τρωτά σε επιθέσεις απάρνησης υπηρεσίας ενάντια στο ίδιο το σύστημα ανίχνευσης, δεδομένου ότι ένας εισβολέας μπορεί να κατορθώσει να καθυστερήσει το μηχανισμό καταγραφής, ή ακόμα και να τον απενεργοποιήσει τελείως. Μια αποδοτική λύση παρέχεται από τα συστήματα ανίχνευσης που ελέγχουν παθητικά το δίκτυο για ύποπτη δραστηριότητα. Δεδομένου ότι εξαρτώνται απλώς από το πρωτόκολλο TCP/IP, είναι ανεξάρτητα από την υποκείμενη αρχιτεκτονική και μπορούν να ελέγξουν τα ετερογενή δίκτυα αρκετά αποδοτικά. Λαμβάνοντας όμως υπόψη τη σύγχρονη τάση για σφαιρική σύνδεση μέσω δικτύων, σχεδόν κάθε επίθεση ασφάλειας περιλαμβάνει ένα ευρύτερο κομμάτι του δικτύου και όχι μεμονωμένα στοιχεία του.

Ένα άλλο ζήτημα που πρέπει να αντιμετωπίζεται αποδοτικά από τις σύγχρονες αρχιτεκτονικές είναι η προοπτική εξέλιξης και επέκτασης των συστημάτων ανίχνευσης και η αποτελεσματική λειτουργία σε διαφορετικές κλίμακες των δικτύων που ελέγχουν. Μονολιθικά κατασκευασμένα συστήματα ανίχνευσης που συλλέγουν δεδομένα ύστερα από λεπτομερή έλεγχο και τα διαβιβάζουν σε έναν κεντρικό υπολογιστή για την επεξεργασία είναι μη αποδοτικά σε ένα μεγάλο επιχειρηματικό δίκτυο, με μεγάλο και συνεχώς αυξανόμενο αριθμό κόμβων και ζεύξεων και με συνεχώς νέες δικτυακές απειλές ασφαλείας. Η λύση σε αυτό το πρόβλημα περιλαμβάνει την κατασκευή του συστήματος ανίχνευσης με τη χρήση διαστρωμάτωσης στον αρχιτεκτονικό σχεδιασμό του. Κάθε στρώμα σε αυτό το πρότυπο λειτουργεί με τη συνάθροιση των στοιχείων ελέγχου που λαμβάνει από τα χαμηλότερα στρώματα και τα διαβιβάζει συνοψίζοντας και συναθροίζοντας τα δεδομένα στο ανώτερο στρώμα. Στο σχεδιασμό τέτοιων συστημάτων έχουν βοηθήσει αλγόριθμοι που βασίζονται στη συνάθροιση και συγχώνευση δεδομένων (Data Fusion algorithms) και που θα παρουσιαστούν στη συνέχεια. Κατά συνέπεια, η πραγματική ανίχνευση μιας εισβολής μπορεί να πραγματοποιηθεί σε οποιοδήποτε στρώμα, με τις απλούστερες να εμφανίζονται στα χαμηλότερα στρώματα και τις πιο προηγμένες στα υψηλότερα στρώματα.

1.2. Ανίχνευση Ανωμαλιών

Μια από τις μεγαλύτερες προκλήσεις στη διαχείριση των δικτύων μεγάλης κλίμακας είναι η ανίχνευση ανωμαλιών στα μοτίβα της δικτυακής κίνησης λόγω επιθέσεων απάρνησης υπηρεσίας ή αυτομεταδιδόμενων ιών. Η ανίχνευση ανωμαλιών είναι μια από τις πλέον προτεινόμενες μεθόδους για την ανίχνευση της κακής χρήσης του δικτύου, καθώς μπορεί να αποδώσει ακόμα και όταν η φύση της επίθεσης είναι πρωτοεμφανιζόμενη και κατά συνέπεια άγνωστη. Αυτό επιτυγχάνεται συγκρίνοντας την τρέχουσα δικτυακή κίνηση με ένα μοντέλο ή γενικότερα ένα σύνολο παραμέτρων που θεωρείται πως περιγράφουν το δίκτυο σε κανονικές συνθήκες λειτουργίας. Αποκλίσεις από την προβλεπόμενη από το μοντέλο συμπεριφορά θεωρούνται ως ανωμαλίες: κακόβουλες επιθέσεις ή σφάλματα στο δίκτυο.

Σε αυτό το πλαίσιο, θεωρούμε ένα σύνολο από διασκορπισμένους ετερογενείς κόμβους οι οποίοι ανταλλάσσουν πληροφορίες μέσω μιας κοινής αυτοματοποιημένης πλατφόρμας και συνεισφέρουν στους αλγόριθμους ανίχνευσης ανωμαλιών μοιράζοντας τη γνώση τους για την κατάσταση του δικτύου. Συσχετίζοντας τα δεδομένα από αισθητήρες που συλλέγουν πληροφορία από διαφορετικά ενεργά στοιχεία του δικτύου, οι διαχειριστές θα μπορούσαν να εντοπίσουν το ή τα μονοπάτια της επίθεσης και να εφαρμόσουν κατάλληλα αντίμετρα κοντά στις πηγές της ανωμαλίας πειράζοντας τον αντίστοιχο δικτυακό εξοπλισμό, π.χ. firewalls. Ένα ακόμα πιο εξελιγμένο σενάριο θα μπορούσε να εμπεριέχει ανταλλαγή πληροφοριών και συνεργασία ανάμεσα σε διαφορετικές διαχειριστικές αρχές μέσω μιας κοινής αυτοματοποιημένης πλατφόρμας. Με σημερινά δεδομένα, η εναλλακτική λύση που βασίζεται στην επικοινωνία των διαχειριστών μέσω τηλεφώνου ή ηλεκτρονικών μηνυμάτων (π.χ τα Computer Security Incidence Response Teams - CSIRTs), είναι μη αυτοματοποιημένη και αποκλείει άμεσες αντιδράσεις.

1.3. Συγχώνευση Δεδομένων

Η συγχώνευση δεδομένων (Data Fusion) από πολλαπλούς αισθητήρες είναι μια σχετικά νέα πρακτική η οποία επιτρέπει το συνδυασμό δεδομένων από πολλαπλούς και ετερογενείς αισθητήρες και πηγές με σκοπό τη δημιουργία συσχετίσεων ανάμεσα σε διαφορετικά γεγονότα ή καταστάσεις [Hall92]. Τα συστήματα που χρησιμοποιούν αυτούς τους αλγορίθμους συχνά συγκρίνονται με τον τρόπο που ο ανθρώπινος εγκέφαλος συγκεντρώνει την πληροφορία από τα δικά του αισθητήρια όργανα, επεξεργάζεται τα δεδομένα, καταλήγει σε κάποια απόφαση και ενεργεί αντίστοιχα. Κάποια από τα πιο χαρακτηριστικά παραδείγματα στα οποία έχει χρησιμοποιηθεί η συγχώνευση δεδομένων είναι στρατιωτικά συστήματα και μετεωρολογικά συστήματα πρόγνωσης του καιρού. Γενικά, η συγχώνευση δεδομένων είναι μια διαδικασία που εκτελείται σε δεδομένα που προέρχονται από πολλές πηγές με σκοπό την ανίχνευση, συσχέτιση, υπολογισμό και συνδυασμό πολλαπλών ροών δεδομένων στοχεύοντας σε ένα ψηλότερο επίπεδο αφαιρετικότητας και σημασίας.

1.3.1. Φυσικά Μοντέλα

Οι αλγόριθμοι στη κατηγορία αυτή προσπαθούν να δημιουργήσουν ένα ακριβές μοντέλο του περιβάλλοντος το οποίο παρατηρούν και να κάνουν τις κατάλληλες εκτιμήσεις, συγκρίνοντας τις τρέχουσες πραγματικές μετρήσεις με αυτές που έχουν προβλεφθεί από το μοντέλο του συστήματος. Σε αυτήν την κατηγορία περιλαμβάνονται επίσης οι μέθοδοι που προσπαθούν να αποσυνθέτουν το παρατηρηθέν αντικείμενο (το δίκτυο ή ένα στοιχείο του δικτύου, όπως μια ζεύξη) στα περιγραφικά συστατικά του (τις κύριες συνιστώσες του). Η μέθοδος που προτείνεται και αναλύεται σε αυτή τη διατριβή ανήκει σε αυτή την κατηγορία, καθώς στηρίζεται στην Ανάλυση Κυρίων Συνιστωσών στην προσπάθεια να αποσυνθέσει τη τρέχουσα κατάσταση του δικτύου στις κύριες συνιστώσες του. Οι κύριες συνιστώσες συλλαμβάνουν τις σημαντικές συσχετίσεις και τα μοτίβα κυκλοφορίας μεταξύ των δικτυακών στοιχείων και επομένως δημιουργούν ένα μοντέλο του ελεγχόμενου δικτύου.

1.3.2. Ταξινόμηση με βάση τις παραμέτρους του συστήματος

Οι αλγόριθμοι που ανήκουν σε αυτήν την κατηγορία κάνουν μια άμεση αντιστοίχηση των παραμετρικών στοιχείων στο πεδίο που περιγράφει την κατάσταση του συστήματος. Μπορούν να διαιρεθούν περαιτέρω σε αλγορίθμους που βασίζονται στη στατιστική, όπως το Μπεϋζιανό Συμπέρασμα (Bayesian Inference) ή τη μεθοδολογία Dempster-Shafer (D-S), αλλά και σε θεωρητικές τεχνικές επεξεργασίας πληροφοριών όπως τα νευρωνικά δίκτυα και μέθοδοι βασισμένες στην εντροπία.

Το Μπεϋζιανό συμπέρασμα υπολογίζει την πιθανότητα να ισχύει μια παρατήρηση δεδομένης της ισχύος μιας a priori υπόθεσης. Η θεωρία αποδείξεων D-S είναι μια μαθηματική θεωρία [Shaf76] βασισμένη σε συναρτήσεις υπολογισμού της πεποίθησης και άγνοιας για την ισχύ μιας υπόθεσης, οι οποίες χρησιμοποιούνται συνδυαστικά για τον υπολογισμό της πιθανότητας να ισχύει η υπόθεση αυτή. Στη δημοσίευση [Siat05], η D-S δοκιμάστηκε για την ανίχνευση ανωμαλιών σε ένα λειτουργικό πανεπιστημιακό δίκτυο. Η θεωρία D-S δίνει τη δυνατότητα να υπολογιστεί ποσοτικά η έλλειψη βεβαιότητας για κάποια υπόθεση. Επίσης, κάποιες παρατηρήσεις μπορούν να οδηγήσουν στην υπόθεση ότι το σύστημα βρίσκεται σε κάποιο υποσύνολο καταστάσεων χωρίς όμως να μπορούν να προσδιορίσουν σε ποια ακριβώς κατάσταση. Για παράδειγμα μπορεί κάποιες αποδείξεις να οδηγούν με μεγάλη πιθανότητα στην υπόθεση $H=\{\theta_1,\theta_2)$, ενώ παράλληλα να μη δίνουν καμιά πληροφορία (πλήρης άγνοια) για τον αν το σύστημα είναι τελικά στην θ_1 ή στην θ_2. Η διαφοροποίηση λοιπόν ανάμεσα στην άγνοια και τη μη βεβαιότητα, καθώς και το γεγονός ότι δεν χρειάζεται τις κατανομές πιθανότητας των πιθανών καταστάσεων του δικτύου είναι τα μεγάλα πλεονεκτήματα της μεθόδου αυτής, ειδικά όταν το παρατηρηθέν δίκτυο είναι πολύπλοκο και η μοντελοποίηση του δύσκολη. Το κύριο μειονέκτημα της μεθόδου είναι ότι είναι εφαρμόσιμη μόνο σε ένα κεντρικό σημείο του δικτύου και όχι σε πολλαπλές ζεύξεις/σημεία του δικτύου, που είναι στην ουσία και ο κύριος στόχος της διατριβής αυτής. Ένα άλλο μειονέκτημα είναι ότι οι ενδείξεις από τους διάφορους αισθητήρες πρέπει να είναι στατιστικά ανεξάρτητες μεταξύ τους. Αυτό αποτελεί πρόβλημα γιατί πολλά από τα μετρικά που χρησιμοποιούνται συνήθως στην ανίχνευση ανωμαλιών όπως (πακέτα, ροές (flows) και Bytes) στη μονάδα του χρόνου είναι συνήθως συσχετισμένα όπως θα δούμε στη συνέχεια. Συνολικά όμως η μέθοδος αυτή

είναι ενδιαφέρουσα και αν χρησιμοποιηθεί κάτω από κατάλληλες συνθήκες μπορεί να είναι αποδοτική στην ανίχνευση ανωμαλιών σε κομβικά σημεία του δικτύου.

Τα προσαρμοστικά νευρωνικά δίκτυα παρέχουν μια ενδιαφέρουσα και γενική μέθοδο που δεν χρειάζεται την ύπαρξη ενός μοντέλου για το παρατηρηθέν σύστημα, αλλά βασίζει την εξαγωγή συμπερασμάτων στην εκπαίδευση των κόμβων της (νευρώνες) χρησιμοποιώντας δεδομένα κατάρτισης. Τα διαφορετικά είδη νευρωνικών δικτύων διαφέρουν στον αριθμό κόμβων και στρωμάτων που χρησιμοποιούν, καθώς επίσης και την επεξεργασία δεδομένων που εκτελείται σε κάθε κόμβο. Αυτές οι μέθοδοι έχουν χρησιμοποιηθεί στα πλαίσια των συστημάτων ανίχνευσης παρείσφρησης αλλά απαιτούν δεδομένα που είναι αντιπροσωπευτικά της κανονικής κυκλοφορίας και δεν περιέχουν ανωμαλίες, τα οποία είναι γενικά αρκετά δύσκολο να συλλεχθούν ή να παραχθούν. Στη δημοσίευση [Brot01] παρουσιάζεται μια μέθοδος ανίχνευσης ανωμαλιών με τη χρήση νευρωνικών δικτύων. Οι μετρήσεις από διάφορους αισθητήρες συγχωνεύονται και η τελική κατάσταση συγκρίνεται με το σύνολο των πιθανών καταστάσεων που έχουν προκύψει από τα δεδομένα κατάρτισης. Το κατώφλι ανίχνευσης ορίζεται σαν μια ακτίνα στο επίπεδο, η περιφέρεια της οποίας καλύπτει το σύνολο των μη ανώμαλων μετρήσεων.

Επιπλέον, οι μέθοδοι που είναι βασισμένες στην εντροπία χρησιμοποιούν την έννοια της εντροπίας πληροφοριών για να περιγράψουν την έμφυτη τυχαιότητα ενός συστήματος επικοινωνιών. Η εντροπία μετρά την τυχαιότητα ενός συνόλου στοιχείων. Υψηλές τιμές εντροπίας υποδηλώνουν ότι η κατανομή πιθανότητας των στοιχείων είναι διασκορπισμένη, ενώ χαμηλές τιμές εντροπίας υποδηλώνουν τη συγκέντρωση της κατανομής γύρω από συγκεκριμένα στοιχεία. Η εντροπία έχει χρησιμοποιηθεί εκτενώς στη βιβλιογραφία για την ανίχνευση αυτομεταδιδόμενων ιών και επιθέσεων απάρνησης υπηρεσίας [Lakh05]. Για το χαρακτηρισμό της δικτυακής κίνησης χρησιμοποιούνται συχνά οι κατανομές των πηγαίων διευθύνσεων (source IP address), διευθύνσεων προορισμού (destination IP address), πηγαίας πόρτας (source port) και πόρτας προορισμού (destination port) στα πακέτα IP. Για παράδειγμα, μια ανωμαλία που προέρχεται από ένα μολυσμένο υπολογιστή που προσπαθεί να μολύνει άλλους υπολογιστές στο διαδίκτυο (αυτομεταδιδόμενος ιός) οδηγεί στη μείωση της εντροπίας των πηγαίων διευθύνσεων στα πακέτα IP. Η μολυσμένη μηχανή παράγει έναν μεγάλο

αριθμό ροών αναγκάζοντας την ίδια πηγαία διεύθυνση να κυριαρχεί στη κατανομή των πηγαίων διευθύνσεων IP.

1.3.3. Γνωστικοί Αλγόριθμοι

Τα μέλη της τρίτης κατηγορίας, δηλαδή οι γνωστικοί αλγόριθμοι (Cognitive algorithms), προσπαθούν να μιμηθούν την γνωστική διαδικασία του ανθρώπινου εγκεφάλου για τον προσδιορισμό ενός αντικειμένου. Δύο αντιπροσωπευτικές προσεγγίσεις που ανήκουν σε αυτήν την κατηγορία είναι: τα έμπειρα συστήματα και οι τεχνικές που βασίζονται στη θεωρία ασαφούς λογικής (fuzzy logic). Τα έμπειρα συστήματα αποτελούνται από μια βάση γνώσεων που αντιπροσωπεύει τη γνώση κάποιου «εμπειρογνώμονα» συνήθως σε μορφή κανόνων. Αυτή η γνώση μπορεί να είναι γεγονότα, αλγόριθμοι, μετρικά κ.λπ. Τα έμπειρα συστήματα έχουν χρησιμοποιηθεί σε ευρεία κλίμακα για λόγους ανίχνευσης παρείσφρησης. Παραδείγματος χάριν, το λογισμικό NIDES [NIDES] έχει μια βάση δεδομένων αποτελούμενη από κανόνες που ορίζουν γνωστές δικτυακές επιθέσεις, όπως αυτές περιγράφονται στα αρχεία καταγραφής του συστήματος, και ενεργοποιεί συναγερμούς όταν τα μηνύματα στα αρχεία καταγραφής του συστήματος ταιριάζουν με κάποιον από τους αποθηκευμένους κανόνες. Η θεωρία ασαφούς λογικής χρησιμοποιείται ως εναλλακτική λύση του λογικού συλλογισμού. Στην ασαφή λογική, μια δήλωση δεν είναι απαραίτητα μόνο αληθινή ή ψευδής, αλλά είναι μια πρόταση με μια σχετική αξία μεταξύ 0, που αντιπροσωπεύει μια απολύτως ψευδή πρόταση, και 1, δηλαδή απολύτως αληθινή [Gom03].

1.4. Στόχοι και σύνοψη της Διατριβής

Στα πλαίσια της συγκεκριμένης διατριβής προτείνουμε και αναλύουμε μια προσέγγιση ανίχνευσης ανωμαλιών που συγχωνεύει δεδομένα που συλλέγονται από διάφορα μέρη του δικτύου έτσι ώστε να επιτυγχάνεται αποδοτική ανίχνευση. Η μέθοδος ονομάζεται μέθοδος πολλαπλών μετρικών πολλαπλών ζεύξεων, Multi-Metric-Multi-Link (M^3L) καθώς αναφέρεται στη χρήση πολλαπλών μετρικών που προέρχονται από ετερογενείς, διασκορπισμένους στο δίκτυο, αισθητήρες. Ένα από τα κυριότερα στοιχεία

της προτεινόμενης μεθόδου είναι ότι προσφέρει επεξεργασία δεδομένων τα οποία παρουσιάζουν μεγάλη συσχέτιση και τα αξιοποιεί σε δυο διαφορετικά επίπεδα:

Α. Επιτρέπει την ταυτόχρονη αξιοποίηση πολλαπλών μετρικών ανά ενεργό μέρος του δικτύου (όπως για παράδειγμα ένας δρομολογητής, ή ακόμα πιο ειδικά μια ζεύξη (link)). Η επεξεργασία πολλαπλών μετρικών όπως π.χ. η ρυθμαπόδοση σε πακέτα, η ρυθμαπόδοση σε bytes TCP ή UDP, και η διαφοροποίηση της συσχέτισής τους, μπορεί να αποκαλύψει ανωμαλίες που δεν επηρεάζουν τον όγκο των δεδομένων και θα ήταν αδύνατο να ανιχνευτούν διαφορετικά.

Β. Συσχετίζοντας μετρικά που έχουν συλλεχθεί από διαφορετικούς δρομολογητές προβάλει μια συνολική εικόνα του δικτύου επιτρέποντας την ανίχνευση ανωμαλιών που δεν είναι τοπικές και διατρέχουν ένα μεγαλύτερο μέρος του δικτύου. Επίσης σημαντικό στοιχείο της μεθόδου είναι ότι είναι ικανή να αναγνωρίσει το ή τα μονοπάτια τα οποία διατρέχει η ανωμαλία, αναγνωρίζοντας και συσχετίζοντας τις συνδέσεις στις οποίες η κίνηση παρουσιάζει μη προβλεπόμενη συμπεριφορά. Παράλληλα, σημαντικό στοιχείο είναι ότι μπορεί να εφαρμοστεί με κατάλληλες προσαρμογές σε ευρύ πεδίο διαφορετικών δικτύων, όπως τα δίκτυα ευρείας ζώνης και τα δίκτυα αισθητήρων.

Για να πετύχει τους στόχους αυτούς, η προτεινόμενη μέθοδος χρησιμοποιεί την Ανάλυση Κυρίων Συνιστωσών - ΑΚΣ (Principal Component Analysis - PCA). Η ΑΚΣ έχει ως στόχο την μείωση των διαστάσεων σε ένα σύνολο δεδομένων διατηρώντας όσο το δυνατόν περισσότερη από την υπάρχουσα διακύμανση (variance) στα δεδομένα.

Ανάμεσα στους στόχους της διατριβής είναι η εφαρμογή και η ανάλυση της επίδοσης της μεθόδου σε διαφορετικού τύπου δίκτυα μεγάλης κλίμακας: δίκτυα ευρείας ζώνης [Cha109] [Cha107], δίκτυα αισθητήρων [Cha107], αλλά και υβριδικά δίκτυα, διαδίκτυα δηλαδή που αποτελούνται από ετερογενή υποδίκτυα, όπως για παράδειγμα τα δίκτυα αυτοκίνησης [Cha307].

Η παρούσα διατριβή διαρθρώνεται ως εξής. Στο δεύτερο κεφάλαιο γίνεται η παρουσίαση του βασικού προτεινόμενου αλγόριθμου ανίχνευσης και η γενική μεθοδολογία που μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα δικτύων μεγάλης κλίμακας Παρουσιάζονται αναλυτικά τα διάφορα βήματα από τα οποία αποτελείται ο αλγόριθμος: η δημιουργία των δεδομένων κατάρτισης και η εξαγωγή των περιγραφικών συστατικών του (κύριες συνιστώσες) των δεδομένων κατάρτισης. Οι κύριες συνιστώσες

συλλαμβάνουν τις σημαντικές συσχετίσεις και τα μοτίβα κυκλοφορίας μεταξύ των δικτυακών στοιχείων και επομένως δημιουργούν ένα μοντέλο του ελεγχόμενου δικτύου. Στη συνέχεια, οι συνιστώσες αυτές συγκρίνονται κατά μια έννοια με τις συνιστώσες της τρέχουσας κίνησης και τυχόν μεγάλες αποκλίσεις θεωρούνται ανωμαλίες.

Στο τρίτο κεφάλαιο της διατριβής παρουσιάζεται η εφαρμογή της μεθοδολογίας στα δίκτυα ευρείας ζώνης. Αριθμητικά αποτελέσματα βασισμένα σε πειράματα εξομοίωσης δείχνουν ότι η προτεινόμενη μεθοδολογία επεξεργασίας πολλαπλών μετρικών μπορεί να αποκαλύψει ανωμαλίες που δεν επηρεάζουν τον όγκο των δεδομένων και θα ήταν αδύνατο να ανιχνευτούν διαφορετικά. Επίσης εξετάζεται ένα ακόμα σημαντικό στοιχείο της μεθόδου: η δυνατότητα αναγνώρισης του ή των μονοπατιών τα οποία διατρέχει η ανωμαλία Επιπροσθέτως, στο κεφάλαιο αυτό παρουσιάζεται η υλοποίηση της προτεινόμενης μεθοδολογίας στα δίκτυα ευρείας ζώνης, με τη βοήθεια της τεχνολογίας Πλέγματος. Το Πλέγμα (Grid) είναι ένας μηχανισμός διαμοίρασης ετερογενών πόρων μέσα σε ένα δίκτυο ευρείας ζώνης και γνωρίζει έντονη ανάπτυξη τα τελευταία χρόνια. Στο κεφάλαιο αυτό επίσης εξετάζεται η αποτελεσματικότητα της προτεινόμενης μεθόδου ανίχνευσης ανωμαλιών χρησιμοποιώντας πραγματικά δεδομένα από το δίκτυο του Εθνικού Δικτύου Έρευνας και Τεχνολογίας (ΕΔΕΤ), δημιουργώντας μια πραγματική ελεγχόμενη επίθεση απάρνησης υπηρεσίας.

Όπως αναφέρθηκε και προηγουμένως έντονο ενδιαφέρον παρουσιάζουν τα Ασύρματα Δίκτυα Αισθητήρων και κατά συνέπεια η μεταφορά της προτεινόμενης μεθόδου σε αυτά, όπως παρουσιάζεται στο τέταρτο κεφάλαιο, αποκτά ιδιαίτερη σημασία. Κλειδί στην εφαρμογή της προτεινόμενης μεθοδολογίας στα ΑΔΑ είναι ότι η συγχώνευση δεδομένων πρέπει να γίνει κατανεμημένα και ιεραρχικά. Συνεπώς, το δίκτυο θα πρέπει να χωριστεί σε υπό-ομάδες με γειτονικούς κόμβους, ώστε να εκμεταλλευτούμε τις υψηλές συσχετίσεις στις μετρήσεις γειτονικών κόμβων. Με τον συνδυασμό και την ανάλυση αποτελεσμάτων από γειτονικές περιοχές, είναι δυνατή η ανίχνευση συσχετισμένων επιθέσεων/ανωμαλιών που επηρεάζουν πολλαπλές ομάδες κόμβων. Η αποτελεσματικότητα της προτεινόμενης μεθόδου ανίχνευσης ανωμαλιών εξετάζεται με τη χρήση μετεωρολογικών δεδομένων που έχουν συλλεχθεί από ένα πραγματικό δίκτυο αισθητήρων.

Στο πέμπτο κεφάλαιο, η εφαρμογή της προτεινόμενης μεθόδου μελετάται και σε ένα παράδειγμα δικτύου με σταθερούς και κινούμενους κόμβους, όπως είναι τα Δίκτυα Αυτοκίνησης. Στα υβριδικά αυτά δίκτυα η εφαρμογή της προτεινόμενης μεθόδου παρουσιάζει αντικειμενικές δυσκολίες, καθώς όταν οι κόμβοι αλλάζουν συνεχώς θέση, αλλάζουν συνεχώς και οι μεταξύ τους συσχετίσεις με αποτέλεσμα η μοντελοποίηση του δικτύου να πρέπει να γίνει με διαφορετικό τρόπο από ότι στα σταθερά δίκτυα αισθητήρων. Η αξιολόγηση της απόδοσης και τα αντίστοιχα αριθμητικά αποτελέσματα αποδεικνύουν ότι η προτεινόμενη μεθοδολογία επιτυγχάνει υψηλή ακρίβεια στην ανίχνευση συμβάντων για διαφορετικά σενάρια συμβάντων και διάφορα μεγέθη ανωμαλίας.

Τέλος, στο έκτο κεφάλαιο παρουσιάζεται μια σύντομη επισκόπηση των βασικών συμπερασμάτων που παρουσιάστηκαν στα κυρίως κεφάλαια της διατριβής, και παρατίθενται θέματα που μπορούν να αποτελέσουν αντικείμενο μελλοντικών επεκτάσεων και περαιτέρω έρευνας και ανάλυσης, όπως αυτά προέκυψαν κατά την εκπόνηση της παρούσας διατριβής.

2. Περιγραφή του αλγόριθμου ανίχνευσης ανωμαλιών

Η συσχέτιση πολλαπλών μετρικών που συλλέγονται από διάφορα μέρη του δικτύου παρουσιάζει διάφορες σχεδιαστικές δυσκολίες. Αφενός, η προτεινόμενη μεθοδολογία στοχεύει στον εντοπισμό των στοιχείων που περιέχουν την ανωμαλία και αφετέρου προσπαθεί να εκμεταλλευτεί το συσχετισμό πολλαπλών μετρικών έτσι ώστε να μπορεί να ανακαλύπτει ανωμαλίες που μπορεί να μην επηρεάζουν σημαντικά κάποιο μετρικό, αλλά επηρεάζουν τις συσχετίσεις ανάμεσα σε μετρικά.

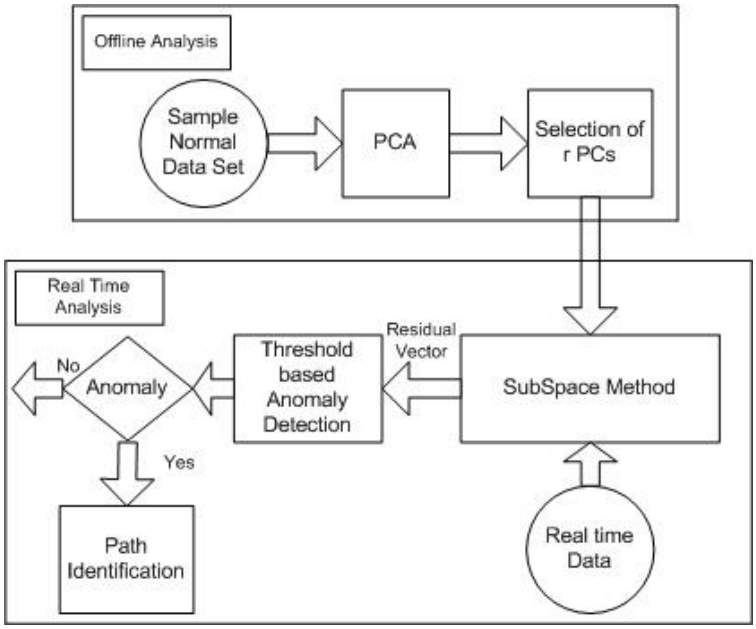
Στη συνέχεια, στην ενότητα 2.1 παρουσιάζεται σύνοψη της μεθοδολογίας ανίχνευσης ανωμαλιών που προτείνεται στα πλαίσια αυτής της διατριβής. Στην ενότητα 2.2 παρουσιάζονται οι βασικές αρχές της Ανάλυσης Κυρίων Συνιστωσών και στη συνέχεια τα επιμέρους βήματα της μεθοδολογίας: η Μέθοδος Υποχώρων στην ενότητα 2.3, ο τρόπος με τον οποίο επιλέγονται οι Κύριες Συνιστώσες στην 2.4, και μια διαδικασία για την κανονικοποίηση των δεδομένων στην ενότητα 2.5. Τέλος, στην ενότητα 2.6 παρουσιάζονται οι απαιτήσεις του αλγόριθμου ανίχνευσης ανωμαλιών σε υπολογιστικούς και δικτυακούς πόρους.

2.1. Σύνοψη μεθοδολογίας

Γενικά, για κάποιο στοιχείο του δικτύου (κόμβο ή ζεύξη για παράδειγμα) συλλέγονται ένα ή περισσότερα μετρικά που χαρακτηρίζουν τη κίνηση που περνά από το στοιχείο αυτό. Προκειμένου να μοντελοποιήσουμε αυτή τη θεώρηση, αντιστοιχούμε ένα σύνολο από εικονικά στοιχεία σε κάθε πραγματικό, με κάθε εικονικό στοιχείο σε αυτό το σύνολο να αντιστοιχεί σε διαφορετικό μετρικό. Αυτή η θεώρηση μας δίνει τη δυνατότητα να χρησιμοποιούμε αδιαφανώς για τον αλγόριθμό πολλαπλά μετρικά. Κάθε μετρικό θεωρείται ίσης σημασίας και συγκρίνεται με τα υπόλοιπα, είτε αφορά το ίδιο στοιχείο είτε κάποιο γειτονικό.

Στη συνέχεια περιγράφεται συνοπτικά και σχηματικά η προτεινόμενη μεθοδολογία. Η συνολική διαδικασία μπορεί να χωριστεί σε δυο βασικά μέρη, όπως φαίνεται στο Σχήμα 2.1: α) τη διαδικασία κατάρτισης (offline analysis), όπου δημιουργείται το μοντέλο της κίνησης από δεδομένα που θεωρούνται κανονικά (δεν περιέχουν ανωμαλίες), και β) τη διαδικασία ανίχνευσης που γίνεται σε πραγματικό χρόνο (real

time analysis) και η οποία συγκρίνει τη τρέχουσα κίνηση με τη μοντελοποιημένη με στόχο την ανίχνευση ανωμαλιών.



Σχήμα 2.1: Παρουσίαση της μεθοδολογίας

Κατά τη διαδικασία κατάρτισης, η ΑΚΣ εφαρμόζεται σε ένα δείγμα της κίνησης (Sample – Normal data set) που θεωρείται ότι δεν περιέχει ανωμαλίες με σκοπό να εξαχθούν οι Κύριες Συνιστώσες (ΚΣ) που επαρκούν για τη περιγραφή των σημαντικών συσχετίσεων στα δεδομένα. Θα πρέπει να τονιστεί ότι η επιλογή του αριθμού των ΚΣ επηρεάζει την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας και συνεπώς γίνεται ιδιαίτερος λόγος για τη διαδικασία αυτή στη συνέχεια. Η ανίχνευση επιτυγχάνεται με τη Μέθοδο Υποχώρων (Subspace Method) η οποία χρησιμοποιεί τις ΚΣ που εξήχθησαν κατά τη διαδικασία κατάρτισης για τον διαχωρισμό του διανύσματος που περιγράφει τη τρέχουσα κίνηση σε δύο μέρη. Το πρώτο μέρος ορίζεται σαν $\mathbf{y_{norm}}$ και περιλαμβάνει το κομμάτι της κίνησης που θεωρείται κανονικό και το δεύτερο ονομάζεται διάνυσμα-υπόλοιπο ($\mathbf{y_{res}}$) και περιλαμβάνει την υπόλοιπη κίνηση. Η ύπαρξη ανωμαλίας στη κίνηση θα έχει σαν αποτέλεσμα να παρατηρηθούν σημαντικές αλλαγές στο διάνυσμα-υπόλοιπο $\mathbf{y_{res}}$. Τελικά, αν ανιχνευτεί κάποια ανωμαλία, γίνεται περεταίρω ανάλυση των δεδομένων με τελικό στόχο την ανίχνευση των στοιχείων του δικτύου τα οποία διαπερνά η ανωμαλία.

2.2. Ανάλυση Κυρίων Συνιστωσών

Η βασική ιδέα πίσω από την ΑΚΣ είναι ο εντοπισμός των γραμμικών συνδυασμών των αρχικών μεταβλητών οι οποίοι είναι γραμμικά ανεξάρτητοι μεταξύ τους. Ο στόχος είναι να μειωθεί ο αριθμός των αρχικών μεταβλητών, διατηρώντας τόσες ώστε το σύστημα να εμπεριέχει όσο το δυνατόν μεγαλύτερο μέρος της διακύμανσης των αρχικών μεταβλητών [Joll02] [Jack03] [Jack79].

Ας υποθέσουμε ότι τα αρχικά δειγματοληπτικά δεδομένα είναι ένας πίνακας **x** με διαστάσεις n×p που αποτελείται από n παρατηρήσεις. Κάθε παρατήρηση είναι ένα σύνολο από p μεταβλητές (x_1, x_2, …, x_p). Αν k είναι ο αριθμός των δικτυακών στοιχείων και l είναι ο αριθμός των διαφορετικών μετρικών ανά στοιχείο, τότε ο αριθμός των συνολικών μεταβλητών είναι p = l × k. Παράλληλα, κάθε γραμμή του πίνακα **x** είναι το διάνυσμα – παρατήρηση, ορίζεται ως **y** και περιγράφει τη κατάσταση του δικτύου μια δεδομένη χρονική στιγμή. Κάθε στήλη του πίνακα **x** αντιστοιχεί σε κάποιο μετρικό και περιέχει όλες τις μετρήσεις της συγκεκριμένης τυχαίας μεταβλητής.

Ορίζουμε ως **S** τον δειγματοληπτικό πίνακα συσχετίσεων των μεταβλητών x_1, x_2, …,x_p με διάσταση p×p. Ο **S** είναι ένας διαγώνιος πίνακας που έχει ως στοιχεία τις $\frac{1}{2}p(p-1)$ συσχετίσεις των αρχικών τυχαίων μεταβλητών και μπορεί να προκύψει από ένα χαρακτηριστικό δείγμα τους. Αν (λ_1, $\mathbf{e_1}$), (λ_2, $\mathbf{e_2}$), …, (λ_p, $\mathbf{e_p}$) είναι τα p τον αριθμό ζεύγη ιδιοτιμών/ιδιοδιανυσμάτων του πίνακα **S**, τότε η ι-στή ΚΣ δίνεται από τον τύπο:

$$z_i = e_i^T y^* \qquad (1)$$

όπου $l_1 \geq l_2 \geq \ldots l_p \geq 0$ και e_i^T είναι το ι-στο ανάστροφο ιδιοδιάνυσμα, y^* είναι η κανονικοποιημένη μορφή του διανύσματος **y** και έχει ως j τον αριθμό στοιχείο το $(y_j - y_{mj})/\sigma_{jj}^{1/2}$, j=1,2…p, όπου y_{mj} και σ_{jj} είναι ο μέσος όρος και η τυπική απόκλιση του y_j αντίστοιχα.

Στην πράξη, ο πίνακας συνδιακύμανσης των κανονικοποιημένων μεταβλητών ισοδυναμεί με τον πίνακα συσχέτισης των αρχικών μεταβλητών. Παραδοσιακά, η ΑΚΣ βασίζεται στην εξαγωγή των ιδιοδιανυσμάτων από το δειγματοληπτικό πίνακα συνδιακύμανσης των τυχαίων μεταβλητών. Όμως ο συνδυασμός πολλών διαφορετικών μετρικών (όπως για παράδειγμα ο αριθμός πακέτων και ο αριθμός bytes σε κάποια ζεύξη), προκαλεί πρόβλημα στην εφαρμογή της μεθόδου ΑΚΣ λόγω της διαφορετικής

κλίμακας στα μεγέθη των μετρικών. Συνεχίζοντας το παράδειγμα σε ένα δίκτυο ευρείας ζώνης, η διακύμανση ενός μεγέθους που μετριέται σε δεκάδες χιλιάδες σε μια κεντρική ζεύξη όπως είναι ο αριθμός πακέτων/δευτερόλεπτο είναι σημαντικά μικρότερη από τη διακύμανση του αριθμού Bytes/δευτερόλεπτο που κινείται σε εκατοντάδες εκατομμύρια στην ίδια ζεύξη. Στην ΑΚΣ, τα στοιχεία με τη μεγαλύτερη διακύμανση τείνουν να κυριαρχούν στις πρώτες ΚΣ, συνεπώς μια ανωμαλία στα άλλα στοιχεία μπορεί να κρυφθεί λόγω του μικρού τους βάρους. Για το λόγο αυτό, γίνεται κανονικοποίηση όλων των μεταβλητών που παίρνουν μέρος στην ΑΚΣ και χρησιμοποιείται ο πίνακας συσχετίσεων αντί για το πίνακα συνδιακύμανσης.

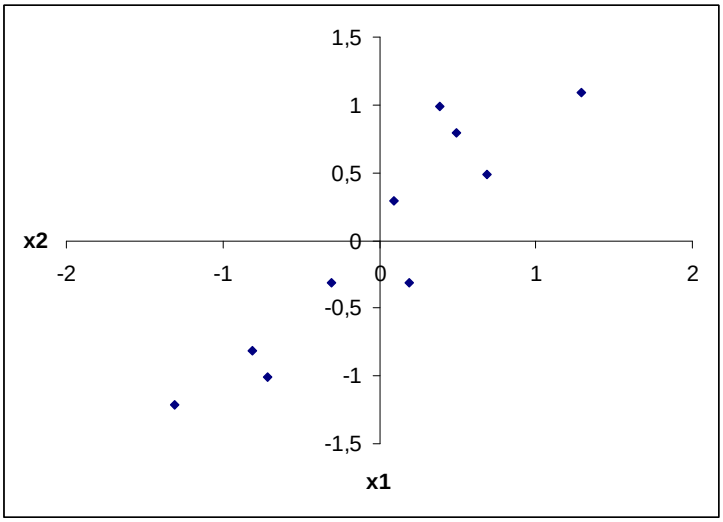
Η έξοδος της ΑΚΣ είναι ένας πίνακας p×p που περιέχει τις ΚΣ. Το επόμενο βήμα στη συνολική διαδικασία είναι η επιλογή των πρώτων r ΚΣ (όπου r<<p) που απαιτούνται προκειμένου να διατηρηθεί το ποσοστό διακύμανσης που χρειάζεται το σύστημα για την μοντελοποίηση του δικτύου. Μετά την παραγωγή των αξόνων από το δειγματοληπτικό πίνακα συσχετίσεων και την επιλογή του r, χρησιμοποιείται η Μέθοδος Υποχώρων. Πριν προχωρήσουμε στη ανάλυση και περιγραφή της μεθόδου αυτής, θα περιγράψουμε κάποια από τα χαρακτηριστικά και τις ιδιότητες της ΑΚΣ.

Ας υποθέσουμε ότι έχουμε ένα σύνολο από p μεταβλητές (x_1, x_2, …, x_p) που αποτελούν ένα διάνυσμα **x** και ότι μας ενδιαφέρουν οι διακυμάνσεις αυτών των p τυχαίων μεταβλητών και η δομή των συνδιακυμάνσεων ή των συσχετίσεών τους. Εκτός αν το p είναι πολύ μικρό, ή η δομή των συσχετίσεων είναι πολύ απλή, συνήθως είναι αδύνατο να συγκρίνει κανείς απλά όλες τις $\frac{1}{2}p(p-1)$ συσχετίσεις ή συνδιακυμάνσεις. Μια εναλλακτική μέθοδος είναι να εστιάσει κανείς σε λίγες (<<p) παράγωγες μεταβλητές που διατηρούν την περισσότερη από την αρχική πληροφορία που δίνεται από τις διακυμάνσεις και τις συσχετίσεις των αρχικών μεταβλητών. Η δημιουργία των ΚΣ μπορεί να περιγραφεί ως εξής:

Το πρώτο βήμα είναι η δημιουργία μιας γραμμικής συνάρτησης $\mathbf{a_1x}$ των στοιχείων του **x** που να έχει τη μέγιστη δυνατή διακύμανση , όπου $\mathbf{a_1}$ είναι ένα διάνυσμα που αποτελείται από p σταθερές τέτοιες ώστε

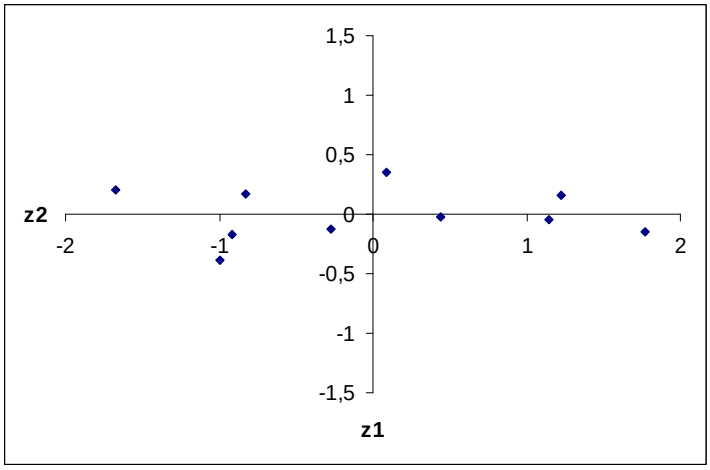
$$a_1 x = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p = \sum_{j=1}^{p} a_{1j}x_j \qquad (2)$$

Στη συνέχεια, στόχος είναι η δημιουργία της γραμμικής συνάρτησης $\mathbf{a_2x}$ με τη μέγιστη δυνατή διακύμανση, η οποία να είναι ασυσχέτιστη με την $\mathbf{a_1x}$ και ούτως καθεξής. Κάθε μια από αυτές τις νέες μεταβλητές είναι μια ΚΣ. Μπορούν να παραχθούν μέχρι p ΚΣ, αλλά συνήθως το μεγαλύτερο ποσοστό της συνολικής διακύμανσης υπάρχει στις πρώτες ΚΣ. Μια γεωμετρική αναπαράσταση της σημαντικής αυτής ιδιότητας που παρουσιάζουν οι ΚΣ μπορεί να φανεί με το απλοϊκό παράδειγμα όπου p=2 και τα αρχικά δεδομένα μπορούν να τυπωθούν σε 2 διαστάσεις. Το Σχήμα 2.2 δίνει ένα διάγραμμα από 10 συσχετισμένες σε υψηλό βαθμό μεταβλητές, x_1 και x_2. Υπάρχει σημαντική διακύμανση και στις 2 μεταβλητές, αν και περισσότερη στην x_1 από ότι στην x_2.



Σχήμα 2.2: Παράδειγμα δεδομένων στο επίπεδο

Με τη μετατροπή σε ΚΣ προκύπτουν οι μεταβλητές z_1 και z_2 που φαίνονται στο Σχήμα 2.3. Μετά τη μετατροπή αυτή, είναι προφανές ότι υπάρχει μεγαλύτερη διακύμανση στην κατεύθυνση του z_1 από τις αρχικές μεταβλητές και ελάχιστη διακύμανση στη z_2. Στη γενικότερη περίπτωση, για ένα σύνολο από μεταβλητές όπου p>2 παρουσιάζει σημαντικές συσχετίσεις ανάμεσα στις μεταβλητές, τότε οι πρώτες λίγες ΚΣ θα έχουν το μεγαλύτερο ποσοστό της συνολικής διακύμανσης των αρχικών μεταβλητών.



Σχήμα 2.3: Παράδειγμα μετασχηματισμένων δεδομένων στο επίπεδο

Κλειδί για τη καλή λειτουργία της μεθόδου είναι ο έλεγχος των συσχετίσεων μεταξύ των αρχικών μεταβλητών. Αν οι μεταβλητές είναι εντελώς ασυσχέτιστες τότε η προτεινόμενη μέθοδος δε θα έχει κανένα αποτέλεσμα. Επίσης, μεταβλητές οι οποίες εμφανίζονται ασυσχέτιστες με τις άλλες θα πρέπει να αφαιρούνται από την ανάλυση, καθώς αν παραμείνουν τότε κάποια από τις κύριες συνιστώσες που θα προκύψουν από την ΑΚΣ απλά θα ταυτίζονται με τις αρχικές μεταβλητές. Στον πίνακα συσχετίσεων που προκύπτει τελικά, απόλυτες τιμές κοντά στην μονάδα υποδεικνύουν υψηλή συσχέτιση, σε αντίθεση με τιμές κοντά στο μηδέν, που υποδεικνύουν χαμηλή συσχέτιση.

2.3. Μέθοδος Υποχώρων

Στη συνέχεια ακολουθεί η ανάλυση πραγματικού χρόνου. Στη φάση αυτή παίρνουμε δεδομένα από τα στοιχεία του δικτύου και προσπαθούμε να διαχωρίσουμε και να προβάλουμε το διάνυσμα παρατήρησης **y** (διαστάσεων 1 x p) σε δυο υποχώρους, τον S που περιέχει τη μοντελοποιημένη κίνηση και τον $\tilde{S}$ που περιέχει την υπόλοιπη [Duni98]. Με τη μέθοδο αυτή που ονομάζεται Μέθοδος Υποχώρων αναλύουμε περιοδικά το κάθε διάνυσμα παρατήρησης **y** σε δυο μέρη:

$$\mathbf{y} = \mathbf{y}_{norm} + \mathbf{y}_{res} \qquad (3)$$

Το διάνυσμα $\mathbf{y_{norm}}$ αντιστοιχίζεται ουσιαστικά στο «ομαλό» κομμάτι της κίνησης και το $\mathbf{y_{res}}$ στο υπόλοιπο μέρος της κίνησης που δε μπορεί να μοντελοποιηθεί. Το $\mathbf{y_{norm}}$ μπορούμε να το υπολογίσουμε προβάλλοντας το **y** στον ομαλό υποχώρο S και αντίστοιχα υπολογίζεται και το $\mathbf{y_{res}}$. Δηλαδή έχουμε:

$$\mathbf{y_{norm}} = \mathbf{PP}^{\mathrm{T}}\,\mathbf{y} = \mathbf{C}\mathbf{y} \text{ και } \mathbf{y_{res}} = (\mathbf{I} - \mathbf{PP}^{\mathrm{T}})\mathbf{y} = \tilde{C}\,\mathbf{y} \qquad (4)$$

Επειδή ο υποχώρος S και κατ' επέκταση το διάνυσμα $\mathbf{y_{norm}}$ αντιστοιχεί στις πρώτες ΚΣ περιέχει όλες τις ισχυρές συσχετίσεις και τις μεταβλητές με τις ισχυρότερες διακυμάνσεις που υπάρχουν στα αρχικά δειγματοληπτικά δεδομένα. Αντίστοιχα, το διάνυσμα-υπόλοιπο $\mathbf{y_{res}}$ περιέχει το υπόλοιπο μικρό μέρος της συνολικής διακύμανσης του αρχικού διανύσματος **y**. Σε περίπτωση όμως που εμφανιστεί μια ανωμαλία που διαταράσσει αυτή την ισορροπία και δημιουργήσει μια καινούρια ισχυρή συσχέτιση που δεν υπήρχε στα δεδομένα κατάρτισης, η διακύμανση που εισάγει η ανωμαλία αυτή στο αρχικό διάνυσμα θα μεταφερθεί στο μεγαλύτερο κομμάτι της στο διάνυσμα υπόλοιπο. Για το λόγο αυτό, οι ανωμαλίες στο δίκτυο έχουν σαν αποτέλεσμα μεγάλες αλλαγές στο $\mathbf{y_{res}}$. Ένα συνηθισμένο στατιστικό μέτρο που χρησιμοποιείται για την ανίχνευση ανωμαλίας είναι το τετραγωνικό σφάλμα πρόβλεψης – Squared Prediction Error (SPE) – το οποίο ορίζεται ως εξής :

$$SPE \equiv \|y_{res}\|^2 = \left\|\tilde{C}\,y\right\|^2 \qquad (5)$$

Αν λοιπόν το SPE ξεπεράσει κάποιο προκαθορισμένο κατώφλι τότε μπορούμε να θεωρήσουμε ότι έχουμε ανωμαλία στο δίκτυο. Στην περίπτωση που οι μεταβλητές που έχουμε χρησιμοποιήσει αναφέρονται σε πολλαπλές φυσικές ζεύξεις μπορούμε να βρούμε και το μονοπάτι της ανωμαλίας/επίθεσης, επιλέγοντας τις συνιστώσες του $\mathbf{y_{res}}$ που συνεισφέρουν στην αύξηση της νόρμας $\|\mathbf{y_{res}}\|$.

2.4. Επιλογή Κυρίων Συνιστωσών

Σημαντικό μέρος κάθε μεθόδου που χρησιμοποιεί την ΑΚΣ είναι η επιλογή του κατάλληλου αριθμού ΚΣ. Στη βιβλιογραφία υπάρχουν γενικοί κανόνες οι οποίοι όμως είναι διαισθητικοί και βασίζονται κυρίως στην εμπειρία και τον πειραματισμό αυτών που τους χρησιμοποιούν [Joll02].

Στην περίπτωση της μεθόδου που προτείνεται σε αυτή την διατριβή, πρέπει να επιλεγεί ο κατάλληλος αριθμός r των ΚΣ, με στόχο τον βέλτιστο διαχωρισμό του διανύσματος **y** στους δυο υποχώρους S και $\tilde{S}$. Ο διαχωρισμός γίνεται έτσι ώστε οι πρώτες r ΚΣ να αντιστοιχούν στον ομαλό υποχώρο S και οι υπόλοιπες (p-r) στον $\tilde{S}$.

Ένα από τα πιο κοινά κριτήρια για την επιλογή του r, είναι το αθροιστικό ποσοστό συνολικής διακύμανσης. Καθώς η ιδιοτιμή κάθε ΚΣ είναι στην ουσία μια ποσοτικοποίηση της διακύμανσης που αντιστοιχεί στην ΚΣ, το αθροιστικό ποσοστό συνολικής διακύμανσης t_r προκύπτει ως εξής: $t_r = \frac{100}{p}\sum_{k=1}^{r} l_k$, όπου με λ_κ εννοείται η k τον αριθμό ιδιοτιμή. Στην περίπτωση αυτή ο βέλτιστος αριθμός ΚΣ προκύπτει από το πλήθος των ΚΣ για το οποίο ξεπερνιέται κάποιο προκαθορισμένο κατώφλι (π.χ 90% ή 95%). Το πρόβλημα όσον αφορά στο κριτήριο αυτό είναι ότι δεν υπάρχει κάποιος αυστηρός τρόπος για την επιλογή του κατωφλίου. Ένας εναλλακτικός κανόνας, ο οποίος μπορεί να χρησιμοποιηθεί μόνο όταν η εξαγωγή των ΚΣ βασίζεται στον πίνακα συσχετίσεων, βασίζεται στο μέγεθος της διακύμανσης που έχει αιχμαλωτίσει κάθε ΚΣ. Η κύρια ιδέα πίσω από αυτό το κανόνα είναι ότι αν όλες οι μεταβλητές των αρχικών δεδομένων είναι ανεξάρτητες, τότε οι ΚΣ αντιστοιχούν στις αρχικές μεταβλητές και θα έπρεπε όλες να έχουν διακύμανση ίση με τη μονάδα (μετά την κανονικοποίηση). Συνεπώς, κάθε ΚΣ με διακύμανση/ιδιοτιμή μικρότερη της μονάδας περιέχει λιγότερη πληροφορία από οποιαδήποτε από τις αρχικές μεταβλητές και δεν αξίζει να διατηρηθεί. Για παράδειγμα, αν τα αρχικά δεδομένα περιέχουν μια ομάδα από μεταβλητές που έχουν υψηλή συσχέτιση μεταξύ τους, τότε θα υπάρχει μόνο μια ΚΣ που να συσχετίζεται με την ομάδα αυτή με διακύμανση μεγαλύτερη της μονάδας. Συνεπώς ο κανόνας θα διατηρήσει μόνο μια ΚΣ που σχετίζεται με τη συγκεκριμένη ομάδα.

Πέρα από τους παραπάνω κανόνες που μπορούν να εφαρμοστούν γενικά στην ΑΚΣ, στα πλαίσια της διατριβής αυτής παρουσιάζεται ένας ειδικός κανόνας που βασίζεται στη χρήση της Μεθόδου Υποχώρων. Πιο συγκεκριμένα, ας υποθέσουμε ότι το $\mathbf{y_s}$ είναι ένα διάνυσμα-παρατήρηση που προκύπτει δειγματοληπτικά από τα δεδομένα κατάρτισης και είναι χαρακτηριστικό της κανονικής κατάστασης του δικτύου. Το $\mathbf{y_s}$ περιγράφει τις πιο χαρακτηριστικές συσχετίσεις ανάμεσα στα γειτονικά στοιχεία του δικτύου και δεν περιέχει καθόλου ανωμαλίες. Αν r είναι ο αριθμός των ΚΣ που έχουν

επιλεγεί , το SPE κάτω από κανονικές συνθήκες (SPE_n) είναι η ευκλείδεια νόρμα του διανύσματος-υπόλοιπου που προκύπτει από το $\mathbf{y_s}$ και υπολογίζεται με βάση τη σχέση (3) ως εξής:

$$SPE_n = \left\| \tilde{C}\, y_s \right\|^2 \qquad (6)$$

Το SPE_n προκύπτει από την προβολή του διανύσματος **y** στον υποχώρο $\tilde{S}$ ο οποίος περιγράφεται από τις πρώτες r τον αριθμό ΚΣ. Οι ΚΣ ταξινομούνται από την ΑΚΣ με σειρά σημαντικότητας, με την σημαντικότητα να μετριέται με την ιδιοτιμή της ΚΣ, η οποία περιγράφει τη διακύμανση που έχει αιχμαλωτίσει η αντίστοιχη ΚΣ. Οι τελευταίες (p-r) ΚΣ ορίζουν κατευθύνσεις στις οποίες υπάρχει μικρή διακύμανση και περιγράφουν γραμμικές, σχεδόν σταθερές σχέσεις ανάμεσα στις αρχικές μεταβλητές. Για το λόγο αυτό, το διάνυσμα-υπόλοιπο $\mathbf{y_{res}}$ που είναι η προβολή του **y** στις τελευταίες (p-r) ΚΣ θα πρέπει να έχει μικρή διακύμανση. Συνεπώς, κάτω από ιδανικές συνθήκες, δηλαδή με αρχικές μεταβλητές με υψηλή συσχέτιση, βέλτιστη επιλογή του r και σωστή επιλογή του διανύσματος $\mathbf{y_s}$, η διακύμανση του SPE_n θα πρέπει να είναι σχεδόν μηδενική.

Στη συνέχεια, ορίζουμε ως $\mathbf{y_i}$ το διάνυσμα-παρατήρηση που προκύπτει όταν το δίκτυο παρουσιάζει κάποια ανωμαλία. Το διάνυσμα αυτό μπορεί να γραφεί ως:

$$\mathbf{y_i = y_s + y_{res} = y_s + \Xi_i f_i} \qquad (7)$$

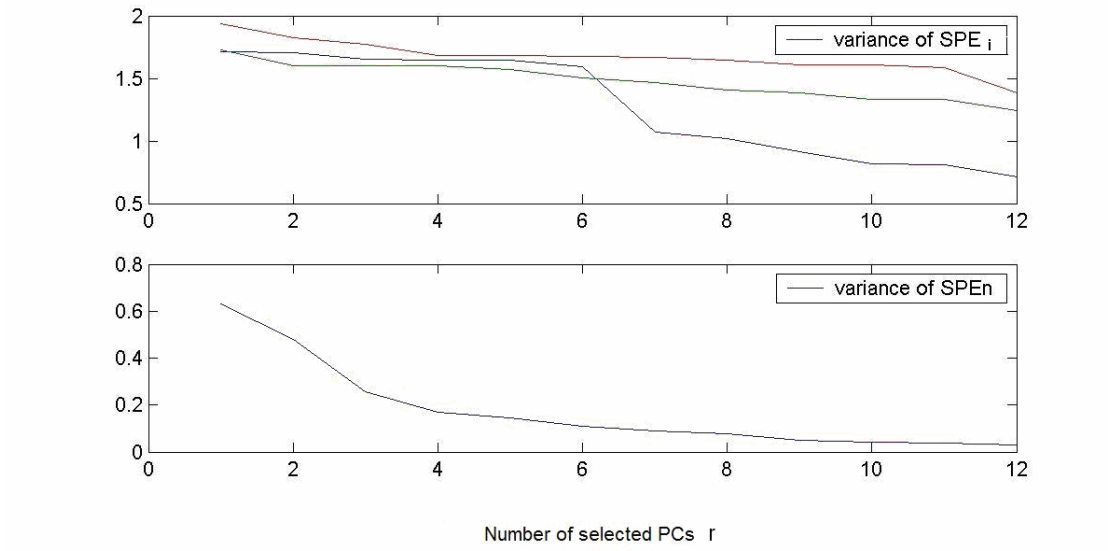
Ο παράγοντας $\mathbf{\Xi_i f_i}$ περιγράφει την αλλαγή που η ανωμαλία προκαλεί στο διάνυσμα παρατήρησης $\mathbf{y_s}$. Το $\mathbf{\Xi_i}$ είναι ένα διάνυσμα με μοναδιαία νόρμα, του οποίου οι διαστάσεις αντιστοιχούν στα στοιχεία του δικτύου που παρακολουθούνται από το σύστημα ενώ το $\mathbf{f_i}$ περιγράφει το μέγεθος της ανωμαλίας. Για παράδειγμα, στη περίπτωση ενός δικτύου όπου όλες οι μεταβλητές του συστήματος αντιστοιχούν σε ένα μετρικό, κάθε μια διάσταση του $\mathbf{\Xi_i}$ αντιστοιχεί σε μια ζεύξη του δικτύου και μπορεί να παίρνει μηδενικές ή μοναδιαίες τιμές, ανάλογα με το αν η ανωμαλία επηρεάζει το συγκεκριμένο στοιχείο ή όχι. Το σύνολο όλων των $\mathbf{\Xi_i}$ δίνει όλες τις πιθανές κατευθύνσεις λαθών. Για παράδειγμα σε ένα δίκτυο ευρείας ζώνης, το σύνολο όλων των $\mathbf{\Xi_i}$ δίνει όλα τα δικτυακά μονοπάτια στα οποία μπορεί να δρομολογηθεί κίνηση (για παράδειγμα πλήρη μονοπάτια εισόδων-εξόδων του δικτύου), ή εναλλακτικά σε ένα υποσύνολο αυτού στο οποίο ο διαχειριστής του δικτύου θέλει να εστιάσει. Στη

περίπτωση χρήσης πολλαπλών μετρικών, το σύνολο όλων των **Ξ** περιλαμβάνει και τα εικονικά στοιχεία. Συνεπώς, αν αγνοήσουμε το παράγοντα μεγέθους της ανωμαλίας, το πλήθος των πιθανών ανωμαλιών που μπορεί να επιβληθούν στο κανονικό δειγματοληπτικό διάνυσμα $\mathbf{y_s}$ δίνεται από το $\mathbf{\Xi_i}$.

Επιπλέον, το αντίστοιχο SPE (που στη περίπτωση της ανωμαλίας αναφέρεται ως SPE_i) υπολογίζεται ως εξής:

$$SPE_i = \left\| \tilde{C}\, y_i \right\|^2 = \left\| \tilde{C}\left(y_s + \Xi_i\, f_i\right) \right\|^2 \qquad (8)$$

Στο Σχήμα 2.4 το ανώτατο διάγραμμα παρουσιάζει τη διακύμανση του SPE_i ως συνάρτηση του αριθμού των επιλεγμένων ΚΣ r για τρεις διαφορετικές ανωμαλίες, ενώ το κατώτατο διάγραμμα παρουσιάζει τη διακύμανση του SPE_n ως συνάρτηση του αριθμού των επιλεγμένων r ΚΣ .



Σχήμα 2.4: Η διακύμανση του SPE ως συνάρτηση του r

Όπως φαίνεται από το σχήμα, οι διακυμάνσεις και των δυο SPE τείνουν να μειωθούν καθώς αυξάνεται ο αριθμός r. Ο στόχος μας είναι διπλός: Αρχικά, το r πρέπει να είναι τέτοιο ώστε η διακύμανση του SPE_n να είναι όσο το δυνατόν πιο κοντά στο μηδέν. Αυτό έχει ως αποτέλεσμα σχεδόν μηδενικές τιμές του SPE κάτω από κανονικές

συνθήκες και τη μείωση των λανθασμένων ανιχνεύσεων. Στη συνέχεια, καθώς το $\mathbf{y_i}$ αντιστοιχεί στο διάνυσμα παρατήρησης κατά στη διάρκεια κάποιας ανωμαλίας, θα πρέπει να αυξάνεται σημαντικά η διακύμανση στο SPE_i για να εξασφαλίζεται η ανίχνευση της συγκεκριμένης ανωμαλίας. Συνεπώς, προκύπτει το πρόβλημα βελτιστοποίησης, όπου στόχος είναι η μεγιστοποίηση της διακύμανσης του SPE_i για ένα σύνολο τεχνητών υποτιθέμενων ανωμαλιών $\mathbf{y_i}$ και παράλληλα η ελαχιστοποίηση της διακύμανσης του SPE_n, χρησιμοποιώντας τον αριθμό r ως παράμετρο. Ο παράγοντας $\mathbf{f_i}$ στο SPE_i μπορεί να χρησιμοποιηθεί σαν βάρος, ώστε να δοθεί μεγαλύτερη σημασία σε κάποιες από τις ανωμαλίες. Παράδειγμα για το πώς μπορεί να χρησιμοποιηθεί αυτή η προσέγγιση για την επιλογή των ΚΣ δίνεται στο κεφάλαιο 3.4.3.3.

2.5. Κανονικοποίηση Δεδομένων

Γενικά μια μέθοδος ανίχνευσης ανωμαλίας εξαρτάται από την κατοχή ενός καλού δείγματος δεδομένων που δεν περιέχει ανωμαλίες και το οποίο συχνά ονομάζεται στην βιβλιογραφία ως δεδομένα κατάρτισης. Στην περίπτωσή μας, ένα ακριβές δείγμα δεδομένων παράγει τις ΚΣ που αντιστοιχούν σε πραγματικές και σημαντικές διακυμάνσεις και συσχετίσεις μεταξύ των στοιχείων του δικτύου, και επομένως διαμορφώνουν το μοντέλο του δικτύου με ακρίβεια. Διάφορες μεθοδολογίες για τη δημιουργία ενός κανονικού συνόλου δεδομένων που στηρίζονται σε διαφορετικές τεχνικές έχουν προταθεί στη βιβλιογραφία [Lee01] [Cann98] [Barf01]. Εντούτοις, πριν χρησιμοποιηθούν για την κατάρτιση του αλγόριθμου ανίχνευσης, τα δεδομένα πρέπει να υποβληθούν σε επεξεργασία από κάποια τεχνική κανονικοποίησης για την ανεύρεση και εξαγωγή των ανώμαλων τιμών. Πρέπει να σημειωθεί ότι η προτεινόμενη μέθοδος ανίχνευσης ανωμαλιών δεν απαιτεί κάποια συγκεκριμένη μέθοδο για τη δημιουργία των δεδομένων κατάρτισης, και ο τρόπος με τον οποίο αποχτήθηκε το κανονικό σύνολο δεδομένων δεν έχει επιπτώσεις στο υπόλοιπο της διαδικασίας. Η διαδικασία κατάρτισης του αλγορίθμου ανίχνευσης θα πρέπει να εκτελείται περιοδικά, δημιουργώντας νέες ΚΣ που να ανταποκρίνονται στη τρέχουσα κατάσταση του δικτύου. Η περίοδος εξαρτάται από τα ιδιαίτερα χαρακτηριστικά κάθε δικτύου και θα πρέπει να επιλέγεται κάθε φορά από το διαχειριστή. Στη συνέχεια παρουσιάζεται μια μέθοδος κανονικοποίησης των δεδομένων που βασίζεται στις αρχές της μεθόδου ΑΚΣ.

Η διαδικασία κατάρτισης του αλγόριθμου ανίχνευσης μπορεί να βασιστεί στην ανίχνευση ανώμαλων τιμών στο σύνολο των δεδομένων. Οι ανωμαλίες αυτές μπορούν να ανιχνευθούν με τον έλεγχο των κατευθύνσεων που καθορίζονται είτε από τις πρώτες είτε τις τελευταίες ΚΣ. Οι πρώτες ΚΣ αντιστοιχούν στις ισχυρές διακυμάνσεις και συσχετίσεις στα δεδομένα και για το λόγο αυτό μπορεί να περιέχουν μόνο έντονες ανωμαλίες που προκάλεσαν σημαντικότατη διακύμανση και διαφοροποίηση στα δεδομένα. Αντίθετα, με τη χρησιμοποίηση των τελευταίων ΚΣ, μπορούν να ανιχνευθούν ανωμαλίες που δεν είναι προφανείς όσον αφορά τις αρχικές μεταβλητές. Ένας τρόπος να εκτελέσει κανείς αυτόν τον έλεγχο, είναι ο υπολογισμός των ΚΣ κατ' επανάληψη, αφήνοντας έξω κάθε φορά μια ή περισσότερες παρατηρήσεις και να επαναλάβει αυτήν την διαδικασία για κάθε παρατήρηση (ή σύνολο παρατηρήσεων). Αυτή η επαναληπτική μέθοδος, αν και είναι υπολογιστικά απαιτητική, εφαρμόζεται περιοδικά και προαιρετικά σε ένα δείγμα των διαθέσιμων δεδομένων κατάρτισης. Η σχέση για τον έλεγχο των πρώτων και των τελευταίων ΚΣ προκύπτει από τα αθροίσματα με βάρη τα τετράγωνα των τιμών των ΚΣ:

$$D_i^{\,2} = \sum_{k=1}^{r} \frac{z_{ik}^{\,2}}{l_k}, \; d_i^{\,2} = \sum_{k=p-r+1}^{p} \frac{z_{ik}^{\,2}}{l_k} \qquad (9)$$

όπου z_{ik} είναι η τιμή της κ-στης ΚΣ που προκύπτει από την ι-στη παρατήρηση και λ_κ είναι η διακύμανση (ιδιοτιμή) της κ-στης ΚΣ. Η χρήση της ιδιοτιμής λ_κ στην παραπάνω έκφραση στοχεύει στο να δώσει ίσο βάρος σε κάθε ΚΣ, καθώς οι πρώτες ΚΣ έχουν μεγαλύτερη διακύμανση. Αν μια παρατήρηση i παράγει μια ιδιότροπη τιμή (πολύ μικρή ή πολύ μεγάλη) στο D_i ή d_i τότε θεωρείται ανωμαλία. Αυτή η μέθοδος είναι αποτελεσματική στην ανίχνευση ανωμαλιών και μπορεί να χρησιμοποιηθεί περιοδικά για τη δημιουργία των δεδομένων κατάρτισης [Shyu03]. Δε μπορεί να χρησιμοποιηθεί όμως σαν αλγόριθμος ανίχνευσης ανωμαλιών γιατί δε προσφέρει καμία πληροφορία για τα επιμέρους στοιχεία του δικτύου που επηρεάστηκαν από την ανωμαλία ή για τη φύση αυτής.

2.6. Απαιτήσεις αλγορίθμου σε υπολογιστικούς και δικτυακούς πόρους

Το μέρος της προτεινόμενης μεθοδολογίας που είναι υπολογιστικά απαιτητικό είναι ο υπολογισμός των ΚΣ από το πίνακα συσχετίσεων. Η διαδικασία αυτή ισοδυναμεί με την εξαγωγή των ιδιοτιμών και ιδιοδιανυσμάτων του πίνακα συσχετίσεων. Η πολυπλοκότητα της διαδικασίας αυτής για ένα πίνακα t × m είναι $O(tm^2)$. Ο υπολογισμός των ΚΣ χρησιμοποιείται για την κατασκευή του μοντέλου της δικτυακής κίνησης, οπότε δε χρειάζεται να τρέχει συνεχώς και σε πραγματικό χρόνο, παρά σε αραιά χρονικά διαστήματα, ανάλογα με τα χαρακτηριστικά του δικτύου. Η Μέθοδος Υποχώρων η οποία τρέχει σε πραγματικό χρόνο, δεν είναι υπολογιστικά απαιτητική, γιατί ισοδυναμεί με ένα απλό πολλαπλασιασμό πινάκων (βλέπε σχέσεις (4) και (5)).

Πέρα από τις απαιτήσεις σε υπολογιστική ισχύ η προτεινόμενη μεθοδολογία έχει συγκεκριμένες απαιτήσεις σε δικτυακούς πόρους. Το ιεραρχικό μοντέλο συλλογής δεδομένων από διάσπαρτους κόμβους σε ένα δίκτυο επιφέρει φόρτο στο δίκτυο καθώς τα δεδομένα πρέπει να μεταφερθούν από τους αισθητήρες στον κεντρικό κόμβο όπου τρέχει ο αλγόριθμος ανίχνευσης. Το γεγονός αυτό μπορεί να προκαλεί προβλήματα στο δίκτυο ανάλογα με τα χαρακτηριστικά, το πλήθος των κόμβων και το ρυθμό αποστολής δεδομένων. Πιο συγκεκριμένα, για ένα δίκτυο στο οποίο υπάρχουν m ενεργά στοιχεία υπό έλεγχο, και σε κάθε στοιχείο συλλέγονται n μετρικά κατά μέσο όρο, το πλήθος των συνολικών μεταβλητών είναι p = m x n. Αν η περίοδος συλλογής είναι k δευτερόλεπτα, τότε στον κεντρικό κόμβο φτάνουν $\frac{m \times n}{k}$ μετρήσεις το δευτερόλεπτο. Για παράδειγμα σε ένα δίκτυο 1000 κόμβων με 6 μετρικά και περίοδο 30 sec έχουμε 200 μετρήσεις/δευτερόλεπτο με την κάθε μέτρηση να αντιστοιχεί σε μερικά bytes. Στα δίκτυα ευρείας ζώνης όπου οι ζεύξεις έχουν εύρος της τάξης μερικών Gbps ένας τέτοιος ρυθμός δεδομένων είναι αμελητέος, αλλά μπορεί να δημιουργήσει πρόβλημα σε άλλες εφαρμογές όπου οι πόροι του δικτύου είναι περιορισμένοι, όπως τα δίκτυα αισθητήρων.

Για να αντιμετωπιστεί το πρόβλημα των περιορισμένων δικτυακών πόρων στα δίκτυα που παρουσιάζουν αυτά τα χαρακτηριστικά χρησιμοποιείται συχνά η έννοια της ιεραρχικής επεξεργασίας και της δρομολόγησης χωρίζοντας τους κόμβους σε ομάδες με βάση κάποια χαρακτηριστικά και συναθροίζοντας τα δεδομένα σε διαφορετικά

επίπεδα. Στο κεφάλαιο που περιγράφεται η εφαρμογή της προτεινόμενης μεθοδολογίας στα δίκτυα αισθητήρων περιγράφεται αναλυτικά πως ο διαχωρισμός σε ομάδες όχι μόνο μειώνει τις απαιτήσεις σε δικτυακούς πόρους αλλά ταυτόχρονα αυξάνει και την ακρίβεια του αλγορίθμου ανίχνευσης.

Στη δημοσίευση [Huan07] παρουσιάζεται ένα σχέδιο προσέγγισης που μειώνει σημαντικά το φορτίο της χρησιμοποίησης ενός συγκεντρωτικού αλγορίθμου ανίχνευσης όπως αυτός που προτείνεται στα πλαίσια αυτής της διατριβής. Στόχος είναι η αποφυγή της συγκέντρωσης όλων των δεδομένων με την εκτέλεση ευφυούς φιλτραρίσματος στους αισθητήρες που συλλέγουν τις μετρήσεις από τα στοιχεία του δικτύου. Το φιλτράρισμα μειώνει το εύρος ζώνης που καταναλώνει το σύστημα ανίχνευσης ανωμαλιών για τα μηνύματα ελέγχου, αλλά ταυτόχρονα προκαλεί μείωση της ακρίβειας στην ανίχνευση ανωμαλιών καθώς η μειωμένη αποστολή δεδομένων διαστρεβλώνει την εικόνα που έχει ο κεντρικός κόμβος για τη κατάσταση του δικτύου. Το κύριο προτέρημα της προσέγγισης αυτής είναι ότι προσφέρει στο διαχειριστή τη δυνατότητα να ορίσει την κατάλληλη για αυτόν ισορροπία ανάμεσα στη μείωση κατανάλωσης δικτυακών πόρων και την ακρίβεια του αλγορίθμου ανίχνευσης.

3. Δίκτυα Ευρείας Ζώνης

Όπως προαναφέρθηκε, οι μεγαλύτερες απειλές στα δίκτυα ευρείας ζώνης είναι οι επιθέσεις απάρνησης υπηρεσίας και οι αυτό-μεταδιδόμενοι ιοί. Ο σκοπός των επιθέσεων απάρνησης υπηρεσίας είναι είτε να πλημμυρίσουν ένα δίκτυο θυμάτων με κίνηση είτε να καταναλώσουν όλους τους διαθέσιμους πόρους ενός εξυπηρετητή με νόμιμα αλλά συνεχή αιτήματα. Χρησιμοποιούν συνήθως ένα ευρύ φάσμα από «μολυσμένους» ηλεκτρονικούς υπολογιστές στους οποίους έχει νωρίτερα εγκατασταθεί κακόβουλος κώδικας, όπως ένας ιός. Όταν ο κώδικας αυτός ενεργοποιηθεί, οι μολυσμένοι υπολογιστές εξαπολύουν μια μαζική ροή από πακέτα επίθεσης. Οι ροές των πακέτων που απαρτίζουν την επίθεση είναι ιδιαίτερα δύσκολο να ανιχνευτούν στα αρχικά στάδια της, καθώς κάθε ροή μπορεί να είναι πολύ μικρή ποσοτικά σε σχέση με το συνολικό άθροισμα των πακέτων που περνούν από τα μεγάλα ενδιάμεσα δίκτυα. Σε ένα δρομολογητή με διεπαφές Gbps ροές μερικών εκατοντάδων kbps μπορεί να περάσουν απαρατήρητες. Φτάνοντας όμως στον στόχο τους, προερχόμενες από διάφορα μέρη του διαδικτύου έχουν συχνά καταστρεπτικό αποτέλεσμα για τον τελικό στόχο. Ένα άλλο χαρακτηριστικό των επιθέσεων αυτών είναι ότι ακόμα και όταν γίνει αντιληπτή μια επίθεση DoS, είναι συχνά αδύνατο να εντοπιστούν οι υπολογιστές που τη δημιουργούν γιατί η IP διεύθυνση πηγής στα πακέτα της επίθεσης είναι συνήθως ψεύτικη και οι υπολογιστές αυτοί είναι διασκορπισμένοι στο διαδίκτυο.

Εξίσου επικίνδυνοι για ένα δίκτυο ευρείας ζώνης είναι και οι αυτό-μεταδιδόμενοι ιοί. Εκμεταλλευόμενοι κάποια «τρύπα ασφαλείας» τα προγράμματα αυτά μπορούν να εξαπλωθούν από υπολογιστή σε υπολογιστή και μετά να χρησιμοποιηθούν από το δημιουργό τους για κακόβουλες ενέργειες, όπως μια κατανεμημένη επίθεση απάρνησης υπηρεσίας. Πολύ γνωστά παραδείγματα τέτοιων ιών είναι τα «SQL Slammer worm» [Slam09] και «Code Red worm» [Red09]. Χαρακτηριστικό παράδειγμα του πόσο επικίνδυνος μπορεί να είναι ένας τέτοιος ιός είναι η επίθεση απάρνησης υπηρεσίας προς διάφορους στόχους στο διαδίκτυο που εξαπολύθηκε στις 25 Ιανουαρίου του 2005 από υπολογιστές μολυσμένους με το «SQL Slammer worm». Η επίθεση ήταν τόσο μαζική ώστε δημιούργησε δραματικές καθυστερήσεις στο διαδίκτυο σε όλο τον κόσμο.

Έρευνες έδειξαν ότι το SQL Slammer worm διαδόθηκε εξαιρετικά γρήγορα, μολύνοντας τα περισσότερα από τα 75.000 θύματά του σε λιγότερο από δέκα λεπτά.

Η άμυνα ενάντια σε μια κατανεμημένη επίθεση απάρνησης υπηρεσίας είναι μια διαδικασία πολλών βημάτων. Το πρόβλημα είναι ότι το δίκτυο που δέχεται την επίθεση δεν έχει τη δυνατότητα να ελέγξει την εισερχόμενη κυκλοφορία. Ο διαχειριστής πρέπει να καθορίσει τα χαρακτηριστικά της επίθεσης και να έρθει σε επαφή με τους διαχειριστές των γειτονικών δικτύων από τα οποία προέρχεται η επίθεση. Το επόμενο βήμα είναι η εγκατάσταση ειδικών φίλτρων (access lists) στους συνοριακούς δρομολογητές τα οποία διαμορφώνονται κατάλληλα ώστε να μειώσουν την μη νόμιμη κίνηση. Εξάλλου, τα φίλτρα αυτά χρειάζονται προσοχή γιατί πολλές φορές σημειώνονται αλλαγές στα χαρακτηριστικά της επίθεσης με αποτέλεσμα να είναι απαραίτητες αλλαγές και στα αντίστοιχα χαρακτηριστικά των φίλτρων. Τελικά πρέπει να απενεργοποιηθούν όταν η επίθεση παύει. Η παραπάνω διαδικασία γίνεται μέχρι τώρα χειροκίνητα και είναι συχνά ιδιαίτερα χρονοβόρα ακριβώς επειδή βασίζεται ολοκληρωτικά στον ανθρώπινο παράγοντα.

Επίσης, οι επιθέσεις αυτές είναι μια απειλή η οποία δεν μπορεί να αντιμετωπιστεί αποτελεσματικά από ένα μεμονωμένο δίκτυο. Η επίθεση πρέπει ιδανικά να ανιχνεύεται κοντά στις πηγές της, εκεί όμως που οι ροές των κακόβουλων πακέτων είναι ακόμα μικρές και δύσκολα ανιχνεύσιμες.

Τα σύγχρονα συστήματα ανίχνευσης εισβολής στα δίκτυα υπολογιστών λειτουργούν είτε ελέγχοντας κάποιους κεντρικούς και συγκεκριμένους υπολογιστές είτε προσπαθούν να αναλύσουν την κίνηση σε ένα μικρό δίκτυο, για σημάδια κακόβουλης δραστηριότητας. Και στις δύο περιπτώσεις μια επίθεση ακόμα κι αν ανιχνευτεί, δεν είναι δυνατό να ενημερωθεί αυτόματα από το σύστημα κανείς άλλος πέρα από το διαχειριστή του συστήματος. Για το λόγο αυτό, η ομάδα εργασίας της IETF που ασχολείται με τα συστήματα ελέγχου εισβολής έχει προτείνει ένα πρωτόκολλο επικοινωνίας ανάμεσα σε τέτοια συστήματα το οποίο ονομάζεται Intrusion Detection Message Exchange Format [IDMEF]. Με το πρωτόκολλο αυτό είναι δυνατή η μετάδοση πληροφορίας που συλλέγει ένα σύστημα ανίχνευσης εισβολών σε μορφή XML. Παράλειψη του προτεινόμενου αυτού πρωτοκόλλου είναι ότι δεν αναφέρεται πως αυτή η πληροφορία θα μπορούσε να χρησιμοποιηθεί προκειμένου να αντιμετωπιστεί η επίθεση.

Το υπόλοιπο του κεφαλαίου αυτού οργανώνεται ως εξής. Στις ενότητες 3.1 και 3.2 γίνεται περιγραφή της εφαρμογής και των απαιτήσεων του αλγόριθμου ανίχνευσης στα δίκτυα ευρείας ζώνης. Στην ενότητα 3.3 πραγματοποιείται αξιολόγηση της προτεινόμενης μεθόδου, ενώ στην ενότητα 3.4 αναλύεται η επίδραση της δειγματοληπτικής συλλογής δεδομένων στην απόδοση της μεθόδου. Τέλος στην ενότητα 3.5 προτείνεται μια αρχιτεκτονική για την εφαρμογή της μεθόδου που βασίζεται στην Τεχνολογία Πλέγματος.

3.1. *Αλγόριθμοι ανίχνευσης επιθέσεων σε δίκτυα ευρείας ζώνης*

Η πλειοψηφία των σχετικών δημοσιεύσεων έχουν επικεντρωθεί στην ανίχνευση ανωμαλιών χρησιμοποιώντας μετρήσεις που αφορούν ένα συγκεκριμένο μέρος του δικτύου. Ο στόχος των δημοσιεύσεων αυτών είναι η χρησιμοποίηση αναλυτικών στατιστικών μεθόδων για την ανίχνευση των ανώμαλων στοιχείων στα δεδομένα.

Στην δημοσίευση [Barf02] παρουσιάζεται η ανάλυση των ροών της δικτυακής κίνησης με τη χρησιμοποίηση wavelets. Η ρύθμιση φίλτρων wavelet έτσι ώστε να αποκαλύπτουν μοναδικά χαρακτηριστικά κάθε τύπου ανωμαλίας παρέχει ένα αποτελεσματικό τρόπο ανίχνευσης καθώς μια ανωμαλία προκαλεί σημαντική αύξηση στη διακύμανση των αντίστοιχων στοιχείων που περνούν από το φίλτρο. Αρχικά οι ερευνητές ανέλυσαν μια ποικιλία σημάτων δικτυακής κίνησης εφαρμόζοντας γενικευμένα φίλτρα wavelet στα δεδομένα. Τα φίλτρα αυτά προσφέρουν ένα αποδοτικό τρόπο απομόνωσης ιδιαίτερων χαρακτηριστικών στα σήματα μέσω μιας συνδυασμένης χωρικής και χρονικής αντιπροσώπευσης του αρχικού σήματος. Χρησιμοποιώντας δεδομένα από διάφορες πηγές όπως για παράδειγμα το NETFLOW [Netf04], κατάφεραν να αναπτύξουν αλγόριθμους που αποκαλύπτουν αποτελεσματικά τα χαρακτηριστικά τόσο της φυσιολογικής κίνησης όσο και ανωμαλιών.

Στη δημοσίευση [Huss03] προτείνεται μια μέθοδος για την αναγνώριση επαναλαμβανόμενων σεναρίων από δικτυακές επιθέσεις. Πιο συγκεκριμένα, στη περίπτωση αυτή σα συγκεκριμένη επίθεση ορίζεται ο συνδυασμός από ένα συγκεκριμένο σύνολο κόμβων και εργαλείων επιθέσεων. Ο μετασχηματισμός Fourier χρησιμοποιήθηκε για την αναγνώριση ενός σεναρίου επίθεσης αναλύοντας το φάσμα

της κίνησης που περιέχει την επίθεση. Το χαρακτηριστικό φάσμα της επίθεσης προκύπτει από τα χαρακτηριστικά των μηχανών που χρησιμοποιούνται στην επίθεση. Στα πλεονεκτήματα της μεθόδου αυτής είναι τα υψηλά ποσοστά ανίχνευσης τα οποία πέφτουν μόνο αν μειωθεί αισθητά και το μέγεθος της επίθεσης, στοιχείο που αλλοιώνει το φάσμα της και δεν ανιχνεύεται από το αντίστοιχο φίλτρο.

Στη μελέτη [Emra02] εξερευνήθηκε μια διαδικασία χρησιμοποίησης πολυμεταβλητών δεδομένων για την ανίχνευση εισβολών σε ένα σύστημα (δίκτυο ή υπολογιστή). Η μέθοδος αυτή βασίζεται στο τεστ T2 του Hotelling το οποίο ανιχνεύει τόσο μεταβολές του μέσου όρου στις ελεγχόμενες μεταβλητές, όσο και μεταβολές συμπληρωματικές που δεν επηρεάζουν το μέσο όρο των μεταβλητών. Για την επιβεβαίωση της αποτελεσματικότητας της μεθόδου, χρησιμοποιήθηκαν δεδομένα από αρχεία καταγραφής ενός υπολογιστή και οι επιδόσεις της συγκρίθηκαν και με άλλες στατιστικές και πιθανοτικές μεθόδους.

Πρόσφατα, εμφανίστηκαν ένας αριθμός από ερευνητικές προσπάθειες που είχαν ως στόχο τη χρήση της ΑΚΣ για την ανίχνευση ανωμαλιών. Πιο συγκεκριμένα, στη δημοσίευση [Labi04] προτείνεται η χρήση ΑΚΣ για τον εντοπισμό ανωμάλων τιμών στα δεδομένα. Πιο συγκεκριμένα, χρησιμοποιώντας το γνωστό πακέτο δεδομένων «1998 DARPA Intrusion Detection» [1998] που περιέχει γνωστές επιθέσεις προσπάθησαν να επιτύχουν ανίχνευση επιθέσεων απάρνησης υπηρεσίας σε μια συγκεκριμένη ζεύξη δικτύου. Ουσιαστικά πρόκειται για την πρώτη προσπάθεια χρησιμοποίησης της ΑΚΣ στο πρόβλημα ανίχνευσης ανωμαλιών στο δίκτυο, η οποία όμως δε προσφέρει καμιά δυνατότητα κατηγοριοποίησης του είδους της επίθεσης και περιορίζεται στον έλεγχο μονάχα μιας ζεύξης του δικτύου.

Στη δημοσίευση [Oka04] οι ερευνητές προτείνουν μια μέθοδο που βασίζεται σε ΑΚΣ και ονομάζουν πίνακα επανεμφάνισης ιδιοτιμών (Eigen co-occurrence matrix - ECM). Η μέθοδος αυτή μοντελοποιεί ακολουθίες από συμβάντα όπως εντολές φλοιού σε ένα σύστημα UNIX και εξάγει τα κύρια χαρακτηριστικά τους με τη χρήση της ΑΚΣ. Ο στόχος είναι η ανίχνευση κακόβουλων χρηστών που έχουν υποκλέψει τη ταυτότητα ενός κανονικού χρήστη. Η ανίχνευση γίνεται συγκρίνοντας σε πραγματικό χρόνο τις ακολουθίες εντολών που εισάγονται στο σύστημα από κάθε χρήστη και το προφίλ του όπως έχει αυτό εξελιχθεί με βάση τη συμπεριφορά του χρήστη στο παρελθόν.

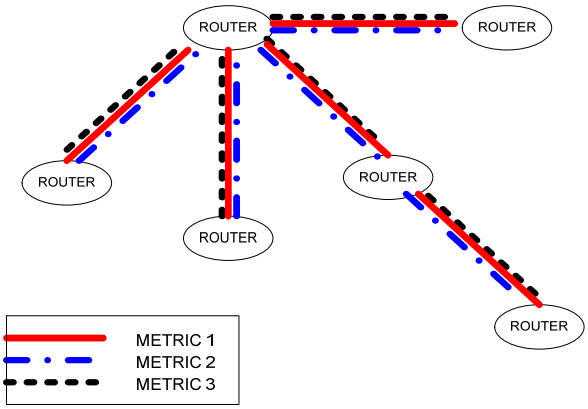
Στη δημοσίευση [Lakh04] η ΑΚΣ χρησιμοποιήθηκε για την ανίχνευση ανωμαλιών σε ολόκληρο το δίκτυο, συναθροίζοντας μετρήσεις ροών πακέτων ή bytes σε δικτυακά μονοπάτια που αποτελούν το ελεγχόμενο δίκτυο. Αυτή η μεθοδολογία, η οποία είναι η κοντινότερη με αυτή που προτείνεται στην παρούσα διατριβή για τα δίκτυα ευρείας ζώνης, εφαρμόζεται ως έχει μόνο σε δεδομένα που αποτελούνται μόνο από ένα μετρικό και έτσι περιορίζει τη μέθοδο σε αποτελεσματική ανίχνευση ανωμαλιών που επηρεάζουν το μέγεθος του μετρικού που ελέγχεται ανά ζεύξη του δικτύου. Για τη μερική αντιμετώπιση του προβλήματος αυτού στη δημοσίευση [Lakh05] γίνεται χρήση της έννοιας της εντροπίας. Πιο συγκεκριμένα, στη δημοσίευση αυτή υπολογίζεται η εντροπία της κατανομής χαρακτηριστικών μεγεθών στις επικεφαλίδες IP/TCP των πακέτων. Τα χαρακτηριστικά αυτά είναι οι διευθύνσεις IP και οι πόρτες προορισμού και πηγής των πακέτων. Οι ερευνητές δείχνουν με πειραματικά αποτελέσματα ότι οι κατανομές των χαρακτηριστικών αυτών επηρεάζονται από ανωμαλίες όπως οι επιθέσεις απάρνησης υπηρεσίας και οι αυτομεταδιδόμενοι ιοί και καταφέρνουν να τις ανιχνεύσουν με επιτυχία. Μειονέκτημα της μελέτης που παρουσιάζουν είναι ότι υποστηρίζουν τη χρησιμοποίηση μόνο της εντροπίας κατανομών ως τύπο μεταβλητής στην ΑΚΣ και δεν εξετάζουν τη βελτιστοποίηση παραμέτρων του αλγόριθμου ανίχνευσης, όπως για παράδειγμα το πλήθος των ΚΣ που θα χρησιμοποιηθούν στην ΑΚΣ.

3.2. Εφαρμογή του αλγορίθμου ανίχνευσης στα Δίκτυα Ευρείας Ζώνης

Η προτεινόμενη μεθοδολογία ανίχνευσης ανωμαλιών στοχεύει στον εντοπισμό του μονοπατιού που περιέχει την ανωμαλία και αφετέρου προσπαθεί να εκμεταλλευτεί το συσχετισμό πολλαπλών μετρικών έτσι ώστε να μπορεί να ανακαλύπτει ανωμαλίες που έχουν διαφορετική φύση και επηρεάζουν με διαφορετικό τρόπο τα αντίστοιχα μετρικά. Η συσχέτιση πολλαπλών μετρικών που συλλέγονται από διάφορα μέρη του δικτύου παρουσιάζει διάφορες σχεδιαστικές δυσκολίες. Στην ενότητα αυτή περιγράφεται η εφαρμογή του αλγόριθμου ανίχνευσης στα δίκτυα ευρείας ζώνης.

Γενικά, για κάθε ζεύξη δικτύου συλλέγονται ένα ή περισσότερα μετρικά που χαρακτηρίζουν τη κίνηση που περνά από τη ζεύξη αυτή. Προκειμένου να δημιουργήσουμε το μοντέλο ενός τέτοιου δικτύου, δημιουργούμε ένα σύνολο από

εικονικές ζεύξεις για κάθε πραγματική, με κάθε εικονική ζεύξη να αντιστοιχεί σε διαφορετικό μετρικό. Μια γραφική αναπαράσταση του μοντέλου αυτού παρουσιάζεται στο Σχήμα 3.1. Στο σχήμα αυτό παρουσιάζεται ένα μικρό δίκτυο όπου έχουμε θεωρήσει τρεις εικονικές ζεύξεις για κάθε πραγματική ζεύξη. Οι τρεις εικονικές ζεύξεις αντιστοιχούν σε τρία διαφορετικά μετρικά για κάθε πραγματική ζεύξη.



Σχήμα 3.1: Ανάλυση με χρήση πολλαπλών μετρικών

3.3. Περιγραφή Απαιτήσεων

Στην προσέγγισή μας η βασική προσπάθεια κατευθύνεται προς την οικοδόμηση ενός πλαισίου συνεργασίας μεταξύ διαφορετικών δικτύων. Για να το επιτύχουμε αυτό πρέπει να αντιμετωπίσουμε όλα τα ζητήματα ασφάλειας και διασύνδεσης ανάμεσα σε δίκτυα που ανήκουν σε διαφορετικές διαχειριστικές ομάδες, να αυτοματοποιήσουμε τις διαδικασίες ανταλλαγής δεδομένων και ανίχνευσης και σε δεύτερη φάση αν είναι δυνατό να αυτοματοποιήσουμε τις διαδικασίες αντίδρασης. Πρακτικά το διαδίκτυο απαρτίζεται από διαφορετικές διαχειριστικές οντότητες (Αυτόνομα Συστήματα) και για το λόγο αυτό απαιτείται η δημιουργία μιας κοινότητας ομότιμων συνεργατών που κοινό στόχο έχουν την ασφάλεια των συμμετεχόντων δικτύων. Η ομοσπονδία αυτή θα έχει ως σκοπό να προσφέρει αυτόματη επικοινωνία ανάμεσα στα δίκτυα που συμμετέχουν, την ανταλλαγή πληροφοριών για σημαντικά περιστατικά ασφαλείας και κυρίως επιθέσεις απάρνησης υπηρεσίας αλλά και αυτοματοποιημένους τρόπους άμεσης απάντησης στις επιθέσεις αυτές με κοινή αντιμετώπιση.

Εξάλλου, θα πρέπει σε κάθε δίκτυο να γίνεται συλλογή και συγχώνευση δεδομένων, ώστε αυτά να μπορούν να αξιοποιηθούν από το κεντρικό αλγόριθμο ανίχνευσης που περιγράψαμε στην προηγούμενη ενότητα. Για τη συλλογή, συγχώνευση και διάθεση των επεξεργασμένων δεδομένων απαιτείται ένας αυτοματοποιημένος τρόπος συλλογής και επεξεργασίας δεδομένων από επιμέρους συστήματα ανίχνευσης ανωμαλιών και αισθητήρες που είναι διασκορπισμένοι στο δίκτυο. Η επικοινωνία ανάμεσα στους κόμβους του συστήματος πρέπει να είναι αξιόπιστη και ασφαλής. Ακόμα σημαντικότερο όμως λόγω της ταχύτητας διάδοσης των επιθέσεων είναι το κατανεμημένο αυτό σύστημα ανίχνευσης ανωμαλιών να παρέχει εγγυήσεις για την έγκαιρη μεταβίβαση των δεδομένων ενώ παράλληλα για λόγους προσωπικού απορρήτου τα δεδομένα που ανταλλάσσονται μερικές φορές πρέπει να είναι ανώνυμα. Αυτό που αξίζει να αναφερθεί εδώ είναι ότι το σύστημα θα πρέπει να έχει εφεδρείες, δηλαδή αν ο κεντρικός κόμβος γίνει ο ίδιος στόχος επίθεσης και χάσει τη λειτουργικότητά του, θα πρέπει εύκολα να μπορεί να μεταφερθεί η λειτουργικότητα του σε κάποιο άλλο κόμβο. Επίσης βασική ιδέα στην σχεδίαση αυτού του κατανεμημένου συστήματος είναι η δυνατότητα που παρέχεται στο διαχειριστή της κάθε δικτυακής περιοχής να καθορίζει αυτός τα χαρακτηριστικά λειτουργίας του, να ελέγχει το βαθμό στον οποίο θα συμμετάσχει στην κοινότητα και να επιλέξει πόση πληροφορία θα δημοσιεύσει και θα μεταβιβάσει στον κεντρικό κόμβο.

Στη διατριβή [Κουτ04] παρουσιάζεται η ιδέα της συνεργατικής αυτόνομης οντότητας. Η οντότητα αυτή προσφέρει επικοινωνία και συνδεσιμότητα μεταξύ των συνεταιριστικών περιοχών, χωρίς να προσφέρει η ίδια κάποιο αλγόριθμο ανίχνευσης επιθέσεων. Η κύρια λειτουργία της είναι η παραλαβή και μεταβίβαση μηνυμάτων για την ασφάλεια και τη λειτουργική κατάσταση των μελών της κοινότητας και η μετάδοση πληροφοριών για τα τοπικά γεγονότα ασφάλειας. Η δεύτερη κύρια λειτουργία της οντότητας είναι να προσφέρει περιορισμένη δυνατότητα αντίδρασης σε κάποια επίθεση. Το στοιχείο αυτό ξεπερνά τις συνηθισμένες ενημερωτικές μόνο λειτουργίες ενός συστήματος ανίχνευσης. Ο στόχος στη σχεδίαση είναι οι αυτοματοποιημένες απαντήσεις σε επιθέσεις απάρνησης υπηρεσίας, τουλάχιστον σε μια περιορισμένη έκταση. Έτσι η οντότητα εγκαθίσταται με την ικανότητα, αφού κάνει αξιολόγηση του γεγονότος ασφάλειας, να αλληλεπιδρά με τα κατάλληλα τοπικά τμήματα των δικτύων με περιορισμένο και ελεγχόμενο τρόπο. Πιο συγκεκριμένα, η

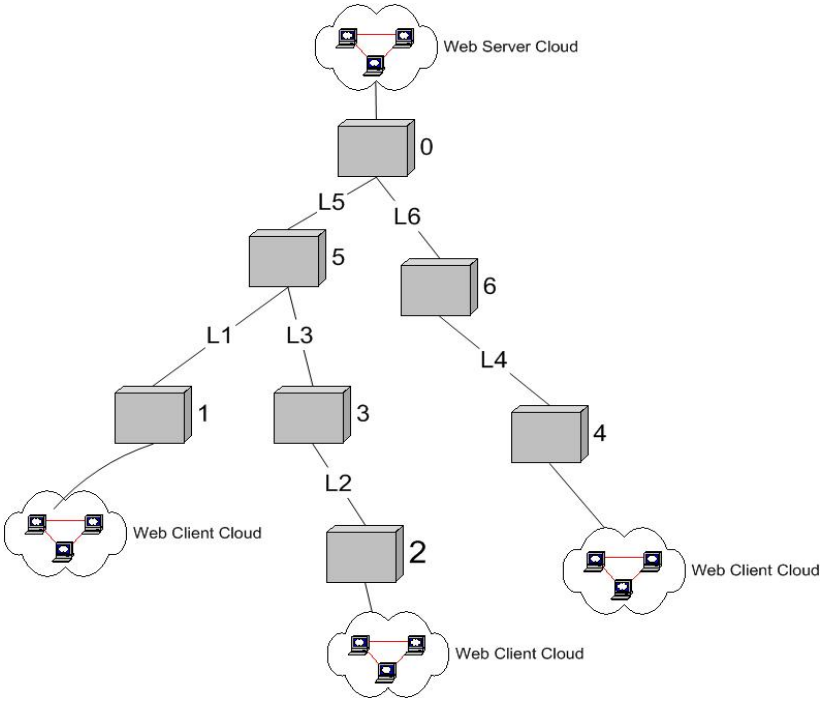
οντότητα εγκαθιστά στους δρομολογητές από τους οποίους περνούν τα πακέτα της επίθεσης κατάλληλα φίλτρα για την καταστολή της. Οι αλλαγές αυτές γίνονται προσωρινά και με την άμεση ενημέρωση του διαχειριστή. Αφετέρου, ο τελευταίος μπορεί να καθορίσει το επίπεδο αλληλεπίδρασης που η οντότητα θα έχει με το δίκτυο. Η σχεδίαση προβλέπει αρχείο το οποίο θα ορίζει την πολιτική που ακολουθείται στις αντιδράσεις της οντότητας στα περιστατικά ασφάλειας που λαμβάνει, όπου ο διαχειριστής μπορεί να καθορίσει ποια μηνύματα θα λαμβάνονται υπ' όψη και ποιες ενέργειες θα πραγματοποιεί η οντότητα.

3.4. Αξιολόγηση της προτεινόμενης μεθόδου στα δίκτυα ευρείας ζώνης με προσομοίωση

Για την αξιολόγηση της αποτελεσματικότητας της προτεινόμενης μεθόδου πραγματοποιήθηκαν εκτεταμένα πειράματα προσομοίωσης με τη βοήθεια του προσομοιωτή δικτύων NS-2 [NS08]. Επιπλέον συγκεκριμένα πειράματα πραγματοποιήθηκαν και με πραγματικά δεδομένα, ώστε να αξιολογηθεί η μέθοδος σε πραγματικές συνθήκες ενός δικτύου παραγωγής.

Στα πειράματα προσομοίωσης θεωρήσαμε μια τοπολογία – δέντρο με κίνηση αποτελούμενη από αιτήσεις και απαντήσεις HTTP. Για τη δημιουργία της δικτυακής κίνησης χρησιμοποιήθηκε το μοντέλο PackMime-http. Το μέγεθος της κίνησης ανάμεσα σε εξυπηρετητή και ένα πελάτη ορίζεται στο μοντέλο αυτό με το ρυθμό γεννήσεως νέων συνδέσεων σε κάθε δευτερόλεπτο. Το μοντέλο, του οποίου η υλοποίηση στο NS-2 μπορεί να παράγει ρεαλιστική συνθετική κίνηση HTTP 1.0 και HTTP 1.1, έχει επαληθευτεί τόσο πειραματικά όσο και θεωρητικά στη δημοσίευση [Cao04].

Η τοπολογία που χρησιμοποιήθηκε εμφανίζεται στο Σχήμα 3.2. Στην κορυφή του δέντρου υπάρχει ένα σύννεφο από εξυπηρετητές και σε κάθε φύλλο υπάρχει ένα σύννεφο από πελάτες. Κάθε σύννεφο από πελάτες επικοινωνεί με το σύννεφο από εξυπηρετητές, δημιουργώντας έτσι τρεις κεντρικές ροές κίνησης. Ο ρυθμός παραγωγής νέων συνδέσεων επιλέχθηκε τόσος ώστε σε κάθε ζεύξη να παράγεται συνολική κίνηση 100-200 Mbps.



Σχήμα 3.2: Τοπολογία Δικτύου

Τα δεδομένα που παράγει το NS-2 δίνονται ως είσοδο σε ένα άλλο πρόγραμμα το οποίο τα χωρίζει σε σειρές ανά μετρικό για κάθε ζεύξη που ελέγχεται. Έτσι, αν M είναι ο αριθμός μετρικών που ελέγχονται, L ο αριθμός των ζεύξεων και N ο αριθμός των χρονικών στιγμών που χωρίζουμε τα δεδομένα, τότε ο πίνακας που δίνεται ως είσοδος στην Ανάλυση Κυρίων Συνιστωσών είναι (P×N), όπου P=M×L. Κάθε γραμμή παρέχει τρέχουσες παρατηρήσεις για κάθε εικονική ζεύξη σε μια χρονική στιγμή, ενώ κάθε στήλη είναι μια χρονοσειρά μιας συγκεκριμένης εικονικής ζεύξης. Οι δικτυακές ανωμαλίες εξομοιωθήκαν εισάγοντας κίνηση TCP με τη χρησιμοποίηση των κλάσεων SimpleTCP and FullTCP του NS-2. Οι λεπτομέρειες και η φύση της μη ομαλής κίνησης περιγράφονται παρακάτω ξεχωριστά για κάθε πείραμα.

3.4.1. Μετρικά υπολογισμού της απόδοσης της μεθόδου

Προκειμένου να υπολογίσουμε την αποτελεσματικότητα του αλγορίθμου ανίχνευσης ανωμαλιών χρησιμοποιήθηκαν τρία ειδικά μετρικά: η πιθανότητα ανίχνευσης P_d , η πιθανότητα λανθασμένης ανίχνευσης (false alarm) P_f και η

πιθανότητα μη ανίχνευσης ανωμαλίας P_m. Η πιθανότητα ανίχνευσης ορίζεται ως η πιθανότητα μια ανωμαλία να αναγνωριστεί με επιτυχία, η πιθανότητα λανθασμένης ανίχνευσης ορίζεται ως η πιθανότητα κανονική κίνηση να καταχωρηθεί λανθασμένα ως ανώμαλη, και τέλος η πιθανότητα μη ανίχνευσης ορίζεται ως η πιθανότητα μια ανωμαλία στη δικτυακή κίνηση να μην ανιχνευτεί και να θεωρηθεί κανονική.

Ένας επιτυχής αλγόριθμος ανίχνευσης ανωμαλιών θα πρέπει να πετυχαίνει υψηλή P_d και χαμηλές P_f και P_m. Εφόσον $P_d+P_m=1$, συνήθως εξετάζουμε την αποτελεσματικότητα του αλγορίθμου ελέγχοντας μόνο τις πιθανότητες ανίχνευσης και λανθασμένης ανίχνευσης. Παράλληλα, χρησιμοποιούμε τις καμπύλες Receiver Operating Characteristic (ROC), προκειμένου να εκφράσουμε και να αναπαραστήσουμε τη σχέση ανάμεσα σε αυτές τις πιθανότητες.

3.4.2. Πειραματικά αποτελέσματα προσομοίωσης

Προκειμένου να εξετάσουμε την αποτελεσματικότητα του προτεινόμενου αλγορίθμου θα παρουσιάσουμε δυο σενάρια: ένα σενάριο στο οποίο η κίνηση αλλοιώνεται λόγω αλλαγής στη δρομολόγηση και ένα σενάριο στο οποίο η κίνηση αλλοιώνεται λόγω της ύπαρξης επίθεσης απάρνησης υπηρεσίας σε κάποιες από τις ζεύξεις του δικτύου. Επίσης, περιγράφεται ένα παράδειγμα ανίχνευσης και αναγνώρισης μιας ανωμαλίας που διέρχεται από διάφορες ζεύξεις του δικτύου.

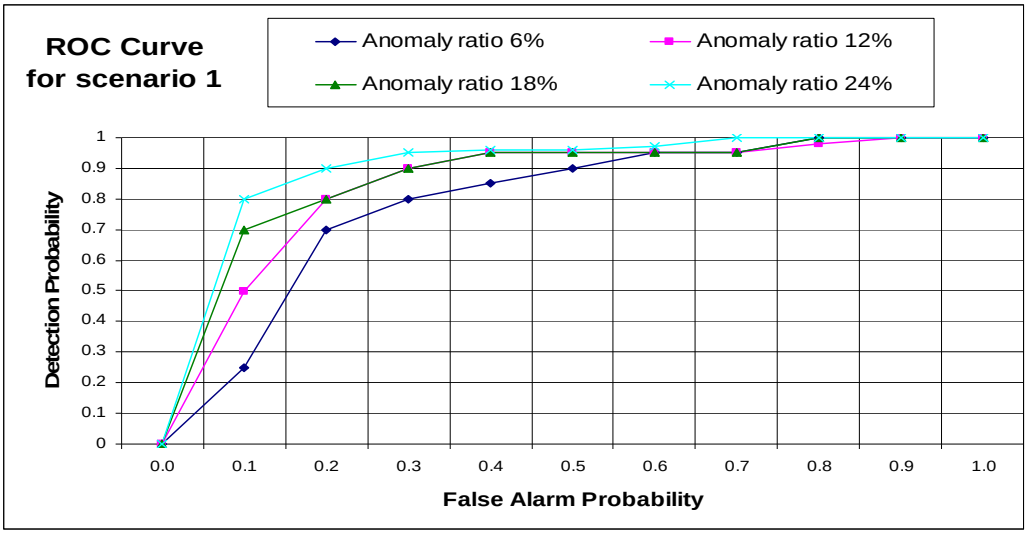
3.4.2.1. Σενάριο ανωμαλίας στη δρομολόγηση

Στο σενάριο αυτό η ανωμαλία έγκειται στην εισαγωγή επιπρόσθετης κίνησης στο δίκτυο με ίδια χαρακτηριστικά με την υπάρχουσα. Όπως περιγράφεται στη τοπολογία του δικτύου, η συνολική δικτυακή κίνηση αποτελείται από τρεις κύριες ροές. Στις ροές αυτές προστίθεται στο σενάριο αυτό ακόμα μια, ανάμεσα στους κόμβους 1 και 2 μέσω των κόμβων 5 και 3 (βλ. Σχήμα 3.2). Η επιπρόσθετη κίνηση αποτελεί ένα μικρό κλάσμα της υπάρχουσας. Στο σενάριο αυτό η αποτελεσματικότητα της μεθόδου συγκρίνεται με αυτής στη δημοσίευση [Lahk04], η οποία βασίζεται μόνο σε ένα μετρικό για τη ανίχνευση ανωμαλιών. Τα πειραματικά αποτελέσματα εμφανίζονται στα παρακάτω

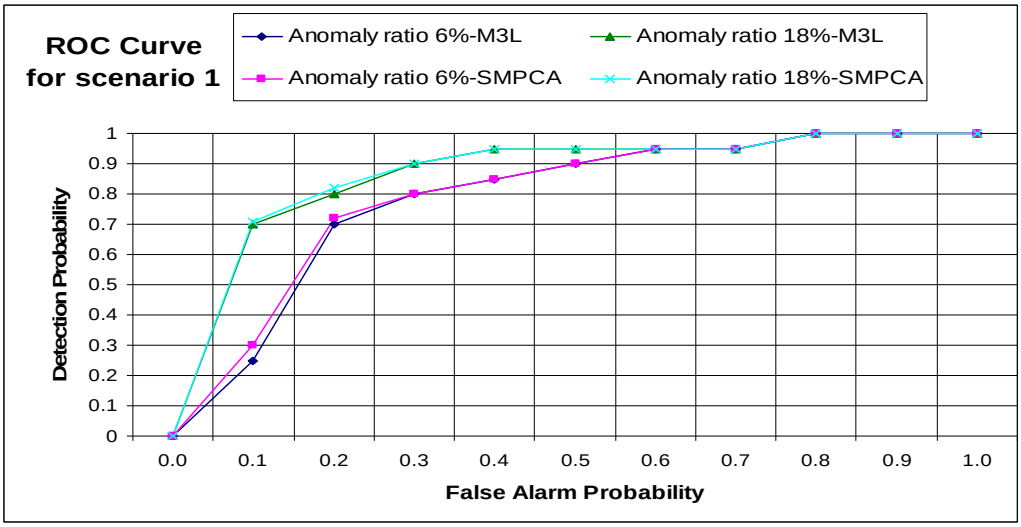
σχήματα, όπου η μέθοδος που χρησιμοποιεί μόνο ένα μετρικό ονομάζεται SMPCA (Single Metric PCA) και χρησιμοποιεί μονάχα πακέτα, ενώ η προτεινόμενη μέθοδος χρησιμοποιεί πακέτα και bytes και ονομάζεται M^3L (Multi-Metric-Multi-Link).

Πιο συγκεκριμένα, στο Σχήμα 3.3 εμφανίζεται η πιθανότητα ανίχνευσης σε συνάρτηση της πιθανότητας λανθασμένης ανίχνευσης, χρησιμοποιώντας τις καμπύλες ROC, όταν εφαρμόζεται η μέθοδος M^3L. Οι διαφορετικές καμπύλες αντιστοιχούν σε διαφορετικές περιπτώσεις όσον αφορά στο ποσοστό ανώμαλης κίνησης σε σχέση με την υπάρχουσα κανονική κίνηση. Ο οριζόντιος άξονας παρουσιάζει την πιθανότητα λανθασμένης ανίχνευσης και ο κατακόρυφος Y παρουσιάζει τη πιθανότητα επιτυχούς ανίχνευσης. Για κάθε καμπύλη, το σημείο στην πάνω και αριστερή γωνία του σχήματος αντιστοιχεί σε ιδανική λειτουργία, με μηδενική πιθανότητα λάθους και 100% επιτυχία ανίχνευσης. Με βάση τα αριθμητικά αποτελέσματα παρατηρούμε υψηλά ποσοστά επιτυχίας, τα οποία αυξάνονται όσο μεγαλώνει το ποσοστό της ανωμαλίας.

Στο Σχήμα 3.4, εμφανίζονται συγκριτικά αποτελέσματα ανάμεσα στις μεθόδους M^3L και SMPCA. Η παρατήρηση του σχήματος οδηγεί στο συμπέρασμα ότι οι δυο μέθοδοι έχουν σχεδόν την ίδια απόδοση στην ανίχνευση της ανωμαλίας. Ο κύριος λόγος είναι ότι η συγκεκριμένη ανωμαλία επηρεάζει μόνο τον όγκο της κίνησης και όχι τη σύνθεση της κίνησης. Στην περίπτωση αυτή η χρήση πολλαπλών μετρικών δεν επιφέρει ουσιαστικά βελτίωση στην ανίχνευση.



Σχήμα 3.3: Η απόδοση της M3L στο σενάριο 1

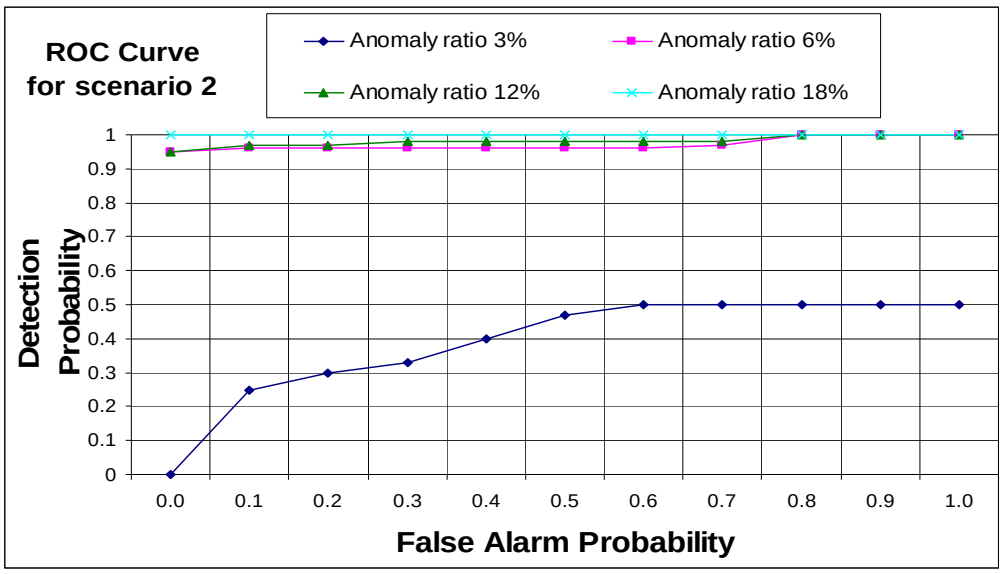


Σχήμα 3.4: Συγκριτικά Αποτελέσματα για το Σενάριο 1

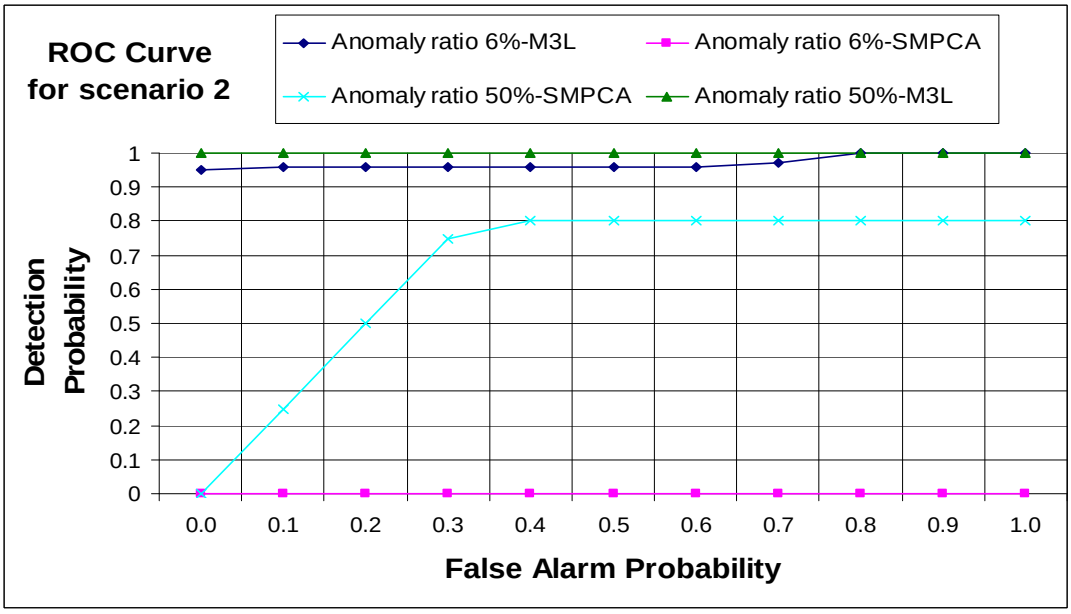
3.4.2.2. Σενάριο επίθεσης απάρνησης υπηρεσίας

Στο σενάριο αυτό η ανωμαλία έγκειται στην εισαγωγή κίνησης στο δίκτυο με διαφορετικά χαρακτηριστικά από την υπάρχουσα. Πιο συγκεκριμένα, με βάση το Σχήμα 3.2, ο κόμβος 4 στέλνει πακέτα SYN (δηλαδή πακέτα TCP με SYN flag που αντιστοιχούν σε αίτηση για αρχικοποίηση μιας συνόδου TCP) στον κόμβο 0 και ο κόμβος 0 απαντά με πακέτα ACK. Ο κόμβος 0 όμως δεν στέλνει άλλο πακέτο και η αίτηση για νέα σύνοδο μένει «κρεμασμένη». Αυτή η μη ολοκληρωμένη ακολουθία πακέτων εξομοιώνει μια επίθεση που προσπαθεί να πλημμυρίσει το δίκτυο με πακέτα SYN και να καταναλώσει όλους τους δικτυακούς και υπολογιστικούς πόρους του κόμβου 0 [Doul04] . Ο ρυθμός παραγωγής νέων ημιτελών συνδέσεων είναι ένα κλάσμα της παραγωγής κανονικών νέων συνδέσεων. Τα πειραματικά αποτελέσματα εμφανίζονται στα επόμενα δυο σχήματα σε καμπύλες ROC για διαφορετικά μεγέθη επίθεσης.

Πιο συγκεκριμένα, το Σχήμα 3.5 παρουσιάζει την πιθανότητα ανίχνευσης σε συνάρτηση της πιθανότητας λανθασμένης ανίχνευσης για διάφορα ποσοστά ανωμαλίας όταν εφαρμόζεται η προτεινόμενη μέθοδος. Στο Σχήμα 3.6 παρουσιάζονται συγκριτικά αποτελέσματα ανάμεσα στις μεθοδολογίες M^3L και SMPCA. Στην M^3L έχουν χρησιμοποιηθεί μετρήσεις πακέτων και bytes, ενώ στην SMPCA μετρήσεις πακέτων.



Σχήμα 3.5: Η απόδοση της M3L στο σενάριο 2



Σχήμα 3.6: Συγκριτικά Αποτελέσματα για το Σενάριο 2

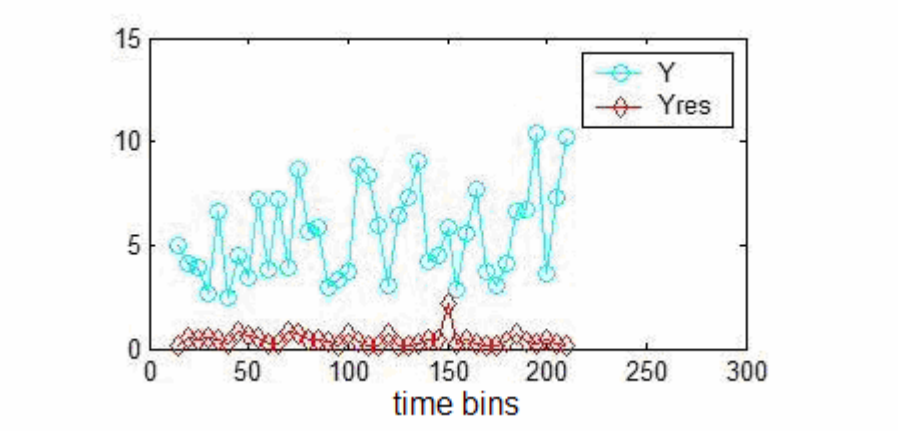
Όπως φαίνεται στο Σχήμα 3.5, η μέθοδος SMPCA αποτυγχάνει να ανιχνεύσει με επιτυχία την επίθεση, εκτός και αν η τελευταία γίνει πολύ μεγάλη. Η αποτελεσματικότητα της μεθόδου περιορίζεται σε περιπτώσεις μεγάλων σε όγκο ανωμαλιών, όπως προβλήματα δρομολόγησης. Στην επίθεση αυτή η αύξηση στον όγκο

της κίνησης είναι αμελητέα, επειδή η επιπρόσθετη κίνηση αποτελείται μόνο από μικρά πακέτα με λίγα bytes. Η εισαγωγή τέτοιων πακέτων επηρεάζει όμως σημαντικά το λόγο bytes/πακέτα στις ζεύξεις στις οποίες εμφανίζεται η ανωμαλία. Η αλλαγή στη συσχέτιση των μετρικών στις ζεύξεις αυτές, σε σχέση με τη μοντελοποιημένη, κάνει την επιτυχή ανίχνευση δυνατή μόνο με τη μέθοδο M^3L και τη χρήση πολλαπλών μετρικών γενικότερα.

3.4.2.3. Παράδειγμα αναγνώρισης του μονοπατιού της επίθεσης

Όπως περιγράψαμε σε προηγούμενο κεφάλαιο, η μέθοδος υποχώρων προβάλει τη μοντελοποιημένη κίνηση στο διάνυσμα $\mathbf{y}_{\mathbf{norm}}$, ενώ την υπόλοιπη κίνηση την προβάλει στο διάνυσμα-υπόλοιπο $\mathbf{y}_{\mathbf{res}}$. Αυτό έχει σαν αποτέλεσμα τη δραματική αύξηση του $\mathbf{y}_{\mathbf{res}}$ όταν υπάρχει ανωμαλία στη κίνηση. Επίσης, εξετάζοντας και αναλύοντας το $\mathbf{y}_{\mathbf{res}}$ μπορούμε να διακρίνουμε τις εικονικές ζεύξεις που επηρεάστηκαν από την ανωμαλία. Διαλέγοντας λοιπόν τις ζεύξεις με τις μεγαλύτερες αλλαγές στις τιμές που σχηματίζουν ένα ή περισσότερα μονοπάτια καταλήγουμε στο γράφο της επίθεσης/ανωμαλίας.

Στο ακόλουθο σχήμα 3.6, παρουσιάζεται στον κατακόρυφο άξονα η εξέλιξη της ευκλείδειας νόρμας των $\mathbf{y}$ και $\mathbf{y}_{\mathbf{res}}$ ως συνάρτηση το χρόνου. Στο χρονικό διάστημα 150 μια ανωμαλία με τα ίδια χαρακτηριστικά που περιγράψαμε στο κεφάλαιο 3.2.2.2 εισάγεται στο δίκτυο. Πιο συγκεκριμένα, με βάση τη τοπολογία δικτύου που περιγράφεται στο Σχήμα 3.2, η επίθεση περνά από τον κόμβο 4 στον 0 και πίσω, έχοντας διαφορετικά χαρακτηριστικά από την κανονική υπάρχουσα κίνηση. Όπως φαίνεται στο Σχήμα 3.7, η καμπύλη που αναφέρεται στη νόρμα του διανύσματος-υπόλοιπο παρουσιάζει μια μεγάλη διαφοροποίηση στο χρονικό διάστημα 150, δίνοντας έτσι μια προφανή ένδειξη για την ύπαρξη ανωμαλίας.



Σχήμα 3.7: Ανίχνευση ανωμαλιών

Ο Πίνακας 3.1 παρουσιάζει σε μεγαλύτερη λεπτομέρεια τα αριθμητικά αποτελέσματα που μας επιτρέπουν να εντοπίσουμε το μονοπάτι της επίθεσης. Πιο συγκεκριμένα, κάθε γραμμή του πίνακα είναι μια στιγμιαία απεικόνιση του διανύσματος $\mathbf{y_{res}}$ που αντιστοιχεί σε διαφορετικό κάθε φορά χρονικό διάστημα (από το 130 ως το 170) , ενώ κάθε στήλη αντιστοιχεί σε κάθε ξεχωριστή εικονική ζεύξη. Για ευκολία, στο πίνακα εμφανίζουμε σε τι μετρικό και πραγματική ζεύξη αντιστοιχεί κάθε εικονική ζεύξη. Όλες οι ζεύξεις θεωρούνται ότι έχουν φορά και οι μεταβλητές (στήλες) έχουν κανονικοποιηθεί με το μέσο όρο τους στο 0. Οι πρώτες 12 στήλες αντιστοιχούν στο μετρικό πακέτα/χρονικό διάστημα, ενώ οι υπόλοιπες 12 αντιστοιχούν σε bytes ανά χρονικό διάστημα.

Στο χρονικό διάστημα 150 η ανωμαλία περνά από τις ζεύξεις L6 και L4. Η ανώμαλη κίνηση πηγαίνει από τον κόμβο 4 στον 0 και πίσω, επηρεάζοντας τη συνδιακύμανση ανάμεσα σε πακέτα και bytes σε κάθε ζεύξη από την οποία περνά. Η διαφορά από το μέσο όρο σε κάθε εικονική ζεύξη από την οποία περνά είναι εμφανής, όπως φαίνεται στον πίνακα. Με τον τρόπο αυτό, αντιστοιχίζοντας τις εικονικές ζεύξεις στις πραγματικές, καταλήγουμε στο μονοπάτι από το οποίο περνά η επίθεση/ανωμαλία.

	1 packets: 0->5	2 packets: 0->6	3 packets: 1->5	4 packets: 2->3	5 packets: 3->2	6 packets: 3->5	7 packets: 4->6	8 packets: 5->0
130	-0.00772	-0.03931	0.001467	-0.01017	0.002484	-0.01215	0.066657	-0.0336
140	-0.0324	-0.02386	-0.02638	-0.00353	-0.00673	-0.01508	0.056583	-0.02386
150	-0.01209	**0.59109**	-0.06157	-0.01534	-0.01723	-0.03018	**-1.0411**	-0.04209
160	-0.0147	-0.01942	0.013112	-0.03496	0.030325	-0.03057	0.034571	-0.01895
170	-0.02498	-0.00553	-0.03443	0.00421	-0.01265	-0.01209	0.058096	-0.01142

	9 packets: 5->1	10 packets: 5->3	11 packets: 6->0	12 packets: 6->4	13 bytes: 0>5	14 bytes: 0->6	15 bytes: 1->5	16 bytes: 2->3
130	-0.02858	0.03499	-0.00785	0.07151	0.00018	-0.01722	0.02496	0.03203
140	-0.02281	-0.00714	-0.03104	0.07941	0.01774	-0.00933	0.02966	0.03620
150	-0.01072	-0.01298	**0.57351**	**-1.3872**	0.011028	**0.63242**	0.04153	0.05389
160	-0.0086	-0.00736	-0.00238	0.02416	-0.00131	-0.00085	0.01512	0.02043
170	-0.01701	-0.03137	-0.01754	0.07726	0.00960	-0.02328	0.02946	0.03543

	17 bytes: 3->2	18 bytes: 3->5	19 bytes: 4->6	20 bytes: 5->0	21 bytes: 5->1	22 bytes: 5->3	23 bytes: 6->0	24 bytes: 6->4
130	-0.00649	0.03201	-0.03409	0.03839	0.00573	-0.00651	-0.03433	-0.01722
140	0.01201	0.03618	-0.03716	0.04157	0.012115	0.01200	-0.03752	-0.00965
150	0.02429	0.05391	**0.67356**	0.06479	-0.00648	0.02468	**1.01127**	**0.63274**
160	0.02000	0.02043	-0.02048	0.02437	-0.01678	0.01948	-0.02123	-0.00092
170	0.01656	0.03544	-0.03882	0.04257	-0.00135	0.01656	-0.03861	-0.02325

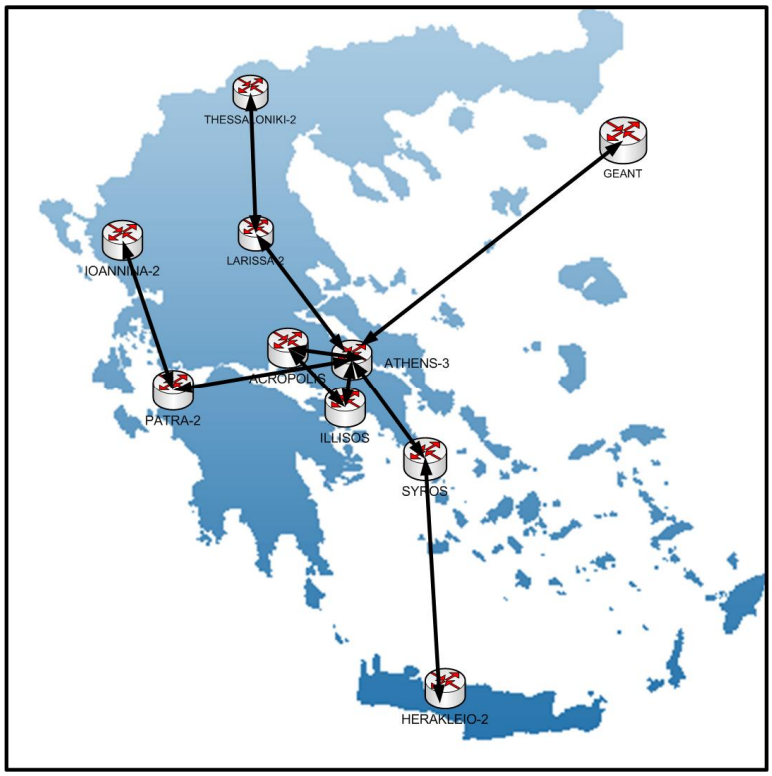
Πίνακας 3.1: Ανίχνευση Μονοπατιού Επίθεσης

3.4.3. Πειραματικά αποτελέσματα σε πραγματικό δίκτυο

Στο κεφάλαιο αυτό εξετάζουμε την αποτελεσματικότητα της προτεινόμενης μεθόδου ανίχνευσης ανωμαλιών χρησιμοποιώντας πραγματικά δεδομένα από το δίκτυο του Εθνικού Δικτύου Έρευνας και Τεχνολογίας (ΕΔΕΤ) [ΕΔΕΤ] και δημιουργώντας μια πραγματική ελεγχόμενη επίθεση απάρνησης υπηρεσίας. Παράλληλα, στο κεφάλαιο αυτό εξετάζουμε τη μέθοδο επιλογής του βέλτιστου αριθμού Κύριων Συνιστωσών. Χρησιμοποιήθηκαν μετρήσεις σε πακέτα και bytes από διάφορες διεπαφές των κεντρικών δρομολογητών του δικτύου τα οποία είναι διαθέσιμα στον δικτυακό τόπο του ΕΔΕΤ. Για την εξομοίωση της ανωμαλίας, δημιουργήθηκε ροή πακέτων η οποία είχε τα ίδια χαρακτηριστικά με αυτήν στο σενάριο που παρουσιάστηκε στο κεφάλαιο 3.4.2.2.

3.4.3.1. Τοπολογία και χαρακτηριστικά Δικτύου

Κατά τη διάρκεια του πειράματος, το κεντρικό δίκτυο του ΕΔΕΤ περιείχε κόμβους σε 6 κεντρικές πόλεις της Ελλάδας: Αθήνα, Θεσσαλονίκη, Πάτρα, Ιωάννινα, Λάρισα και Ξάνθη. Όλοι οι κόμβοι διασυνδέονταν με ζεύξεις των 2.5 Gbps. Στην Αθήνα υπήρχαν επίσης τρεις κεντρικοί δρομολογητές που αποτελούσαν το μητροπολιτικό δίκτυο της πόλης: οι Athens-3, Acropolis και Ilissos. Η τοπολογία του δικτύου κατά τη διάρκεια του πειράματος περιγράφεται στο Σχήμα 3.8:



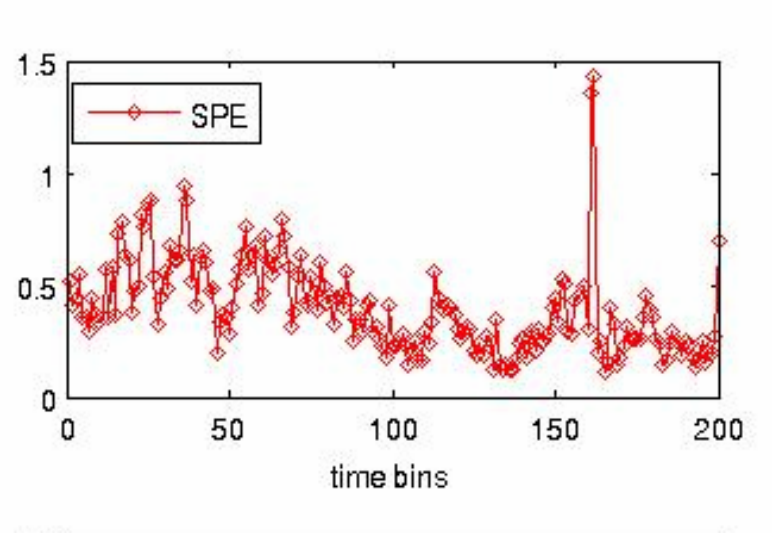
Σχήμα 3.8: Το δίκτυο κορμού του ΕΔΕΤ

Τα δεδομένα που συλλέξαμε αποτελούνται από 20 εικονικές ζεύξεις, που αντιστοιχούν σε μετρήσεις πακέτων και bytes από τις ζεύξεις που συνένωναν τους κεντρικούς δρομολογητές του ΕΔΕΤ. Εξάλλου, η συλλογή δεδομένων διήρκησε περίπου ένα μήνα, κατά τη διάρκεια του οποίου ανιχνεύθηκαν και άλλες πραγματικές ανωμαλίες, όπως ανωμαλίες στη δρομολόγηση λόγω προβλημάτων στον ενεργό εξοπλισμό του δικτύου.

3.4.3.2. Εξομοίωση επίθεσης

Η επίθεση εκτελέστηκε χρησιμοποιώντας ένα εργαλείο εξομοίωσης επίθεσης απάρνησης υπηρεσίας που δημιούργησε τα μεγάλα ποσά πακέτων τύπου TCP-SYN. Ο στόχος της επίθεσης ήταν ένας κόμβος που βρισκόταν στο δίκτυο του Εθνικού Μετσόβιου Πολυτεχνείου (ΕΜΠ) με σύνδεση 100 Mbps και υπήρξαν δύο επιτιθέμενοι που ήταν τοποθετημένοι σε τοπικά δίκτυα των δρομολογητών του ΕΔΕΤ patra-2 και thessaloniki-2. Κάθε επιτιθέμενος ήταν συνδεδεμένος με ζεύξη 100Mbps και έστελνε

τα πακέτα TCP-SYN προς το θύμα με ελεγχόμενο ρυθμό, χρησιμοποιώντας ψευδείς διευθύνσεις IP προερχόμενες από το τοπικό του δίκτυο. Η διαδικασία αυτή είχε ως στόχο την εξομοίωση μιας κατανεμημένης επίθεσης απάρνησης υπηρεσίας. Η επίθεση διήρκεσε 10 λεπτά, με κάθε επιτιθέμενο να παράγει σχεδόν 2000- 3000 πακέτα το δευτερόλεπτο. Η ήδη υπάρχουσα κανονική κίνηση περιείχε περίπου 100 χιλιάδες πακέτα το δευτερόλεπτο στις κεντρικές ζεύξεις του δικτύου κορμού. Η αντίστοιχη αύξηση στα bytes ήταν σχεδόν αμελητέα. Με τη χρησιμοποίηση μόνο των κεντρικών συνδέσεων του δικτύου κορμού, ήταν δυνατό να ανιχνεύσουμε επιτυχώς μια επίθεση μεγέθους μικρότερου από 5% με βάση των αριθμό πακέτων στις κεντρικές ζεύξεις υπό κανονικές συνθήκες, αξιοποιώντας την ύπαρξη υψηλών συσχετίσεων μεταξύ του όγκου πακέτων και bytes στις συνδέσεις αυτές. Όπως φαίνεται στο Σχήμα 3.9, η καμπύλη που αντιστοιχεί στο SPE (που αντιστοιχεί στην ευκλείδεια νόρμα του $\mathbf{y}_{\mathbf{res}}$) παρουσιάζει μια πολύ μεγάλη απόκλιση γύρω από τη χρονική περίοδο 160 που είναι η χρονική περίοδος κατά τη διάρκεια της οποίας έγινε η εξομοίωση της επίθεσης, παρέχοντας με τον τρόπο αυτό μια καθαρή ένδειξη για την ύπαρξη της επίθεσης.



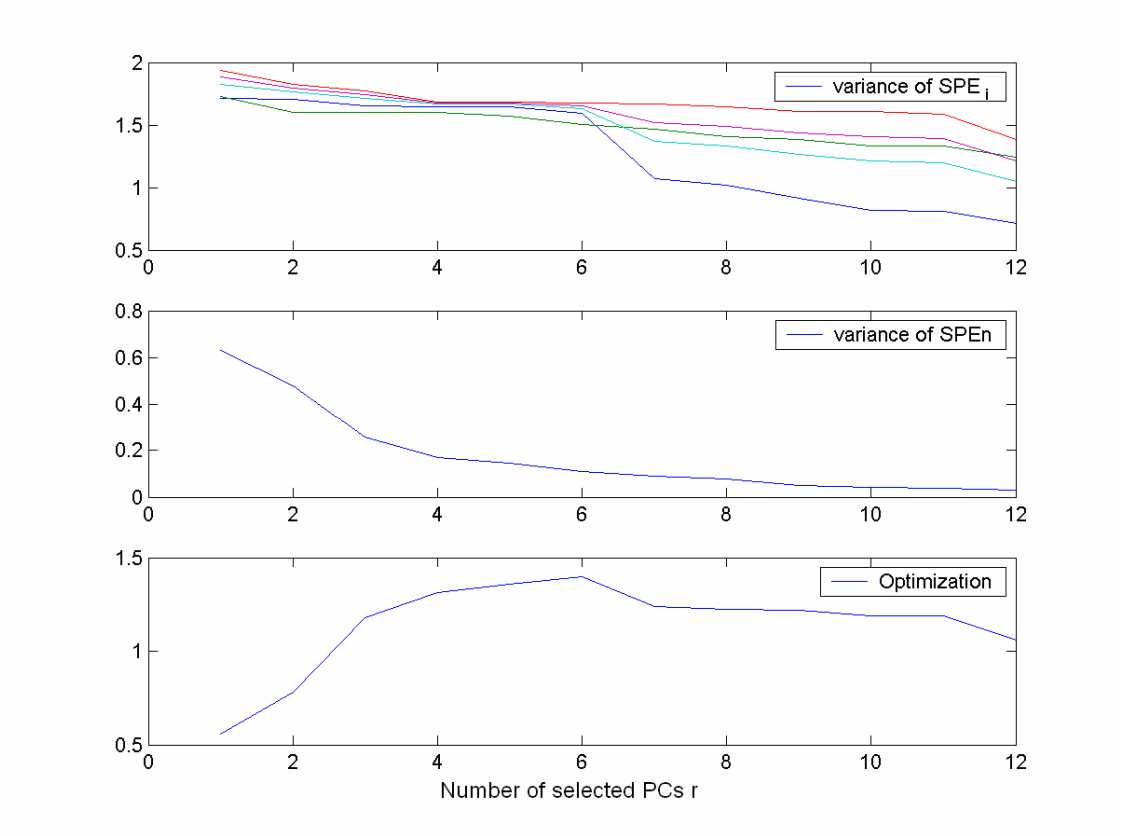
Σχήμα 3.9: Ανίχνευση Επίθεσης Απάρνησης Υπηρεσιών

3.4.3.3. Παράδειγμα επιλογής του βέλτιστου αριθμού ΚΣ

Σε αυτό το τμήμα, παρουσιάζεται ένα παράδειγμα που καταδεικνύει την εφαρμογή και την αποτελεσματικότητα της μεθοδολογίας για την επιλογή του βέλτιστου αριθμού των ΚΣ με τον προκαθορισμό ενός συνόλου πέντε διαδρομών για τις οποίες θέλουμε να βελτιστοποιήσουμε την ανίχνευση ανωμαλίας. Χρησιμοποιώντας το δίκτυο του ΕΔΕΤ, μιμούμαστε πέντε διαφορετικές ανωμαλίες που αντιστοιχούν στις διαδρομές με προορισμό τη διεπαφή του κεντρικού δρομολογητή που διασυνδέει και υποστηρίζει το δίκτυο του ΕΜΠ, και προέλευση πέντε διεπαφές από τους ακραίους δρομολογητές του ΕΔΕΤ. Αυτή η θεώρηση που χρησιμοποιείται στην περίπτωσή μας για λόγους επίδειξης, περιγράφει ένα σενάριο στο οποίο ο διαχειριστής του δικτύου ενδιαφέρεται πρωτίστως για ανωμαλίες/επιθέσεις με προορισμό το δίκτυο του ΕΜΠ.

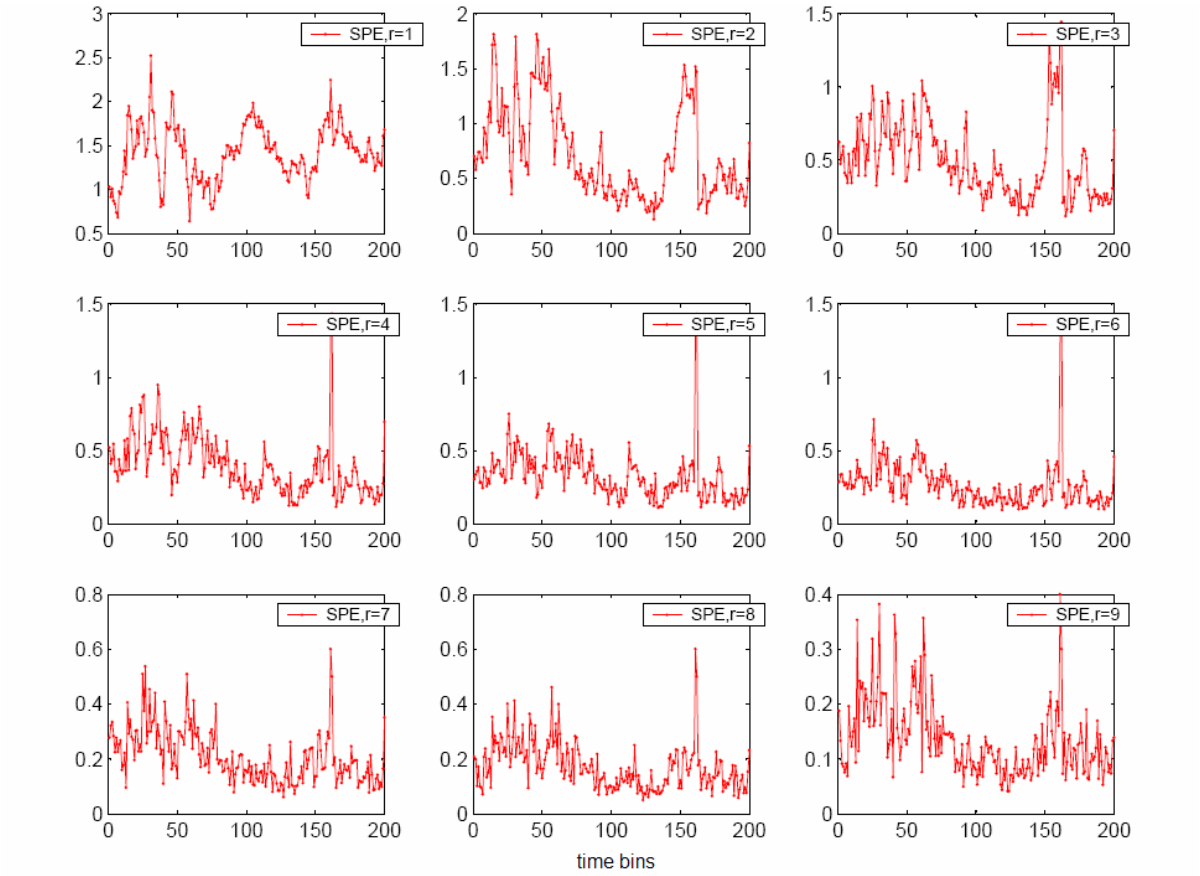
Στο Σχήμα 3.10 παρουσιάζουμε τα αποτελέσματα της τεχνικής βελτιστοποίησης επιλογής ΚΣ, σα συνάρτηση του αριθμού r των επιλεγμένων ΚΣ (οριζόντιος άξονας). Συγκεκριμένα, χρησιμοποιώντας τη σχέση (8), υπολογίζουμε την αντίστοιχη διακύμανση του SPEi για κάθε ανωμαλία. Όπως φαίνεται στο ανώτερο διάγραμμα του σχήματος 3.10 (όπου σχεδιάζουμε τη διακύμανση του SPEi για κάθε μια τις πέντε διαφορετικές υποτιθέμενες ανωμαλίες), η διακύμανση για κάθε SPEi μειώνεται καθώς ο αριθμός r των επιλεγμένων ΚΣ αυξάνεται. Με την επιλογή ενός δειγματοληπτικού διανύσματος $\mathbf{y}_s$ που περιγράφει τη δικτυακή κίνηση υπό κανονικές συνθήκες υπολογίζουμε και σχεδιάζουμε τη διακύμανση του SPEn στο δεύτερο (μέσο) διάγραμμα. Το διάνυσμα αυτό μπορεί να προκύψει με διάφορους τρόπους από τα δεδομένα κατάρτισης και στην περίπτωση μας επιλέχθηκε τυχαία από τα δεδομένα του πειράματος. Τέλος, με βάση ένα απλό (με ίσα βάρη) κριτήριο βελτιστοποίησης λαμβάνουμε και σχεδιάζουμε την αντίστοιχη καμπύλη «βελτιστοποίησης»: (VAR (SPE1) + VAR (SPE2) + VAR (SPE3) + VAR (SPE4) + VAR (SPE5))/5 - VAR (SPEn), για τις διαφορετικές τιμές του r, όπου με VAR() εννοείται η διακύμανση κάθε μεγέθους. Στο κριτήριο αυτό δεν υπάρχουν βάρη στις διακυμάνσεις που αντιστοιχούν στις διαφορετικές ανωμαλίες, γιατί όλες θεωρούνται ίσης σημασίας. Το κριτήριο αυτό μας δίνει τον αριθμό ΚΣ για τον οποίο έχουμε όσο το δυνατόν μεγαλύτερες τιμές διακύμανσης σε περίπτωση ανωμαλίας και όσο το δυνατόν μικρότερη τιμή

διακύμανσης υπό κανονικές συνθήκες. Η αντίστοιχη καμπύλη στο τρίτο διάγραμμα του σχήματος 3.10 υπαγορεύει ότι ο βέλτιστος αριθμός ΚΣ στην περίπτωση αυτή είναι ο αριθμός 6.



Σχήμα 3.10: Βελτιστοποίηση επιλογής ΚΣ

Τέλος, προκειμένου να αξιολογηθεί καλύτερα η προτεινόμενη τεχνική και να επικυρωθεί το προηγούμενο αποτέλεσμα, σχεδιάζουμε την καμπύλη που αντιστοιχεί στην εξέλιξη των τιμών SPE ως συνάρτηση του χρόνου για τις διαφορετικές τιμές του αριθμού ΚΣ. Οι αντίστοιχες τιμές SPE αυτή τη φορά υπολογίζονται από τα αποτελέσματα του πειράματος που περιγράφηκε στην αρχή του κεφαλαίου αυτού. Στην ουσία, κάθε καμπύλη από τις 9 που παρουσιάζονται στο Σχήμα 3.11 αντιστοιχεί σε εκτέλεση του αλγόριθμου ανίχνευσης με διαφορετικό αριθμό επιλεγμένων ΚΣ κάθε φορά. Η αποτελεσματικότητα της προτεινόμενης τεχνικής επιλογής ΚΣ επιβεβαιώνεται, δεδομένου ότι ο αλγόριθμος ανίχνευσης αποδίδει καλύτερα για r=5 και r=6. Για r=5 και r=6, η τιμή του SPE τη στιγμή της επίθεσης έχει τη μεγαλύτερη απόσταση από το μέσο όρο του SPE υπό κανονικές συνθήκες, δίνοντας έτσι τη δυνατότητα έγκυρης ανίχνευσης χωρίς λανθασμένες ανιχνεύσεις.



Σχήμα 3.11: Ανίχνευση ανωμαλιών για διαφορετικές τιμές του αριθμού ΚΣ

3.4.3.4. Τεχνικές και Μετρικά ανίχνευσης ανωμαλιών

Μια σημαντική προϋπόθεση για την αποτελεσματική εφαρμογή της προτεινομένης μεθόδου είναι η ύπαρξη υψηλών συσχετίσεων ανάμεσα στις εικονικές ζεύξεις που ελέγχονται στο δίκτυο. Ουσιαστικά, απαιτείται να γίνει μια αποδοτική επιλογή μετρικών, γιατί η χρησιμοποίηση μετρικών με μικρό βαθμό συσχέτισης στον αλγόριθμό ανίχνευσης εισάγει θόρυβο στο σύστημα (μεγαλύτερη διακύμανση στο διάνυσμα – υπόλοιπο), γιατί οι συσχετίσεις τους δε μπορούν να μοντελοποιηθούν από την ΑΚΣ. Συνεπώς, η χρήση μετρικών με μικρό βαθμό συσχέτισης οδηγεί στη μείωση της ακρίβειας στην ανίχνευση, είτε αυξάνοντας τη πιθανότητα λανθασμένης ανίχνευσης σε περίπτωση που το κατώφλι ανίχνευσης είναι χαμηλό, είτε χαμηλώνοντας τη πιθανότητα ανίχνευσης αν αυξηθεί το κατώφλι για την αποφυγή λανθασμένων ανιχνεύσεων.

Στην ενότητα αυτή μελετάμε τη συσχέτιση πραγματικών μετρήσεων από δυο κεντρικούς δρομολογητές του ΕΔΕΤ. Τα δεδομένα συγκεντρώθηκαν με τη βοήθεια του εργαλείου που παρουσιάστηκε στην προηγούμενη ενότητα. Το τεστ για τον έλεγχο συσχέτισης ανάμεσα σε δυο μεταβλητές είναι ο συντελεστής συσχέτισης (correlation coefficient). Μεταβλητές με συντελεστή συσχέτισης κοντά στη μονάδα μεταβάλλονται αντίστοιχα στην ίδια κατεύθυνση, ενώ μεταβλητές με συσχέτιση κοντά στο -1 μεταβάλλονται σε αντίθετες κατευθύνσεις.

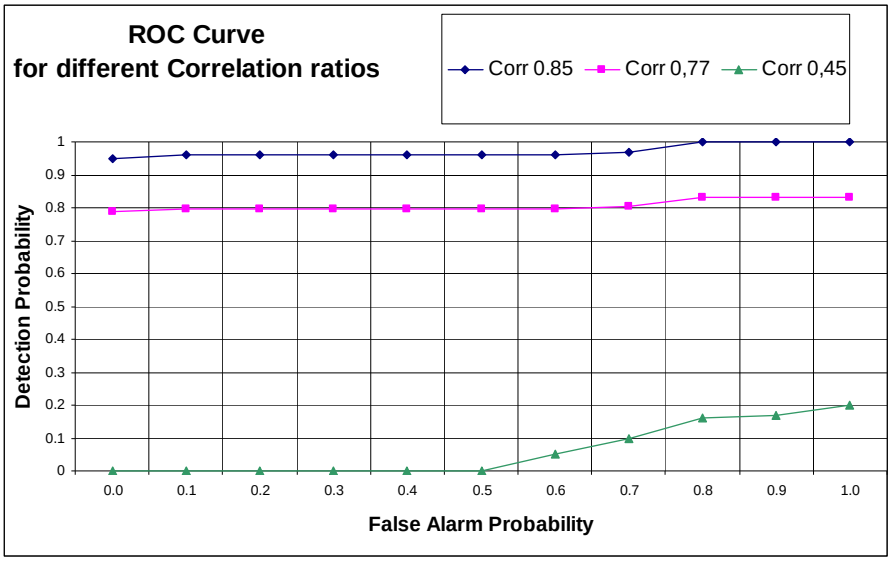
Στα πειράματά μας, με βάση τη συλλογή δεδομένων από το ΕΔΕΤ, διαπιστώσαμε ότι γειτονικές εικονικές ζεύξεις παρουσιάζουν υψηλή συσχέτιση. Ενδεικτικές τιμές εμφανίζει ο Πίνακας 3.2, και προέρχονται από το κεντρικό δρομολογητή του ΕΔΕΤ ilissos και αφορούν τη ζεύξη του με το δίκτυο του ΕΜΠ. Τα μεγέθη στα οποία αναφέρεται ο πίνακας περιγράφονται στη συνέχεια, στον Πίνακα 3.3.

	Np	**Nf**	**Nb**	**Np_tcp**	**Np_udp**	**Nf_tcp**	**Nf_udp**	**Nfs**	**Nft**
Np	1	0.9545	0.95568	0.99247	0.61995	0.95436	0.89097	0.93657	0.93136
Nf	0.9545	1	0.84984	0.97015	0.44234	0.99746	0.93906	0.9948	0.99039
Nb	0.95568	0.84984	1	0.93407	0.68717	0.85269	0.78128	0.82083	0.81382
Np_tcp	0.99247	0.97015	0.93407	1	0.51937	0.97325	0.88989	0.95304	0.94654
Np_udp	0.61995	0.44234	0.68717	0.51937	1	0.42335	0.51325	0.42571	0.43059
Nf_tcp	0.95436	0.99746	0.85269	0.97325	0.42335	1	0.9153	0.98816	0.98199
Nf_udp	0.89097	0.93906	0.78128	0.88989	0.51325	0.9153	1	0.95263	0.95656
Nfs	0.93657	0.9948	0.82083	0.95304	0.42571	0.98816	0.95263	1	0.99887
Nft	0.93136	0.99039	0.81382	0.94654	0.43059	0.98199	0.95656	0.99887	1

Πίνακας 3.2: Ανάλυση πολλαπλών μετρικών

Στο παρακάτω πείραμα εξετάζουμε τη σημασία της ύπαρξης υψηλών συσχετίσεων στα μετρικά. Πιο συγκεκριμένα, δοκιμάζουμε τον αλγόριθμο για τρία ζευγάρια από μετρικά με διαφορετικούς βαθμούς συσχέτισης για κάθε ζευγάρι (0.85, 0.77 και 0.45 αντίστοιχα). Τα αντίστοιχα πειραματικά αποτελέσματα που αφορούν τη πιθανότητα ανίχνευσης και λανθασμένης ανίχνευσης παρουσιάζονται στο Σχήμα 3.12. Λαμβάνοντας υπόψη τις ενδεικτικές τιμές συσχέτισης στα δεδομένα που λήφθηκαν από το δίκτυο του ΕΔΕΤ σε πραγματικές συνθήκες και την υψηλή απόδοση του

αλγόριθμου όταν υπάρχει ισχυρή συσχέτιση, περιμένουμε ότι η απόδοση του αλγορίθμου σε πραγματικές συνθήκες θα είναι εξίσου υψηλή.



Σχήμα 3.12: Η σημασία της συσχέτισης μετρικών στην αποτελεσματικότητα της μεθόδου

3.5. Δειγματοληπτική Συλλογή Δεδομένων

Όπως αναφέρθηκε προηγουμένως, ο προτεινόμενος αλγόριθμος ανίχνευσης ανωμαλιών βασίζεται στη συλλογή και συγχώνευση δεδομένων από διαφορετικά μέρη του δικτύου και συνεπώς η ακριβής και συνεπής εξαγωγή της κατάστασης του δικτύου βασίζεται στην εγκυρότητα των δεδομένων και την έγκαιρη μετάδοση τους. Στην πράξη όμως, τα συλλεχθέντα δεδομένα μπορεί είτε να περιέχουν ανακρίβειες λόγω κάποιου σφάλματος (π.χ. χαλασμένοι ή απορυθμισμένοι αισθητήρες, πρόβλημα στη μετάδοση των δεδομένων λόγω δικτυακών προβλημάτων), είτε να είναι ελλιπή επειδή έχει πραγματοποιηθεί δειγματοληψία στις μετρήσεις. Λόγω της μεγάλης κίνησης των σημερινών δικτύων και την αντίστοιχη δυσκολία στην αποθήκευση και επεξεργασία όλων των πληροφοριών που αφορούν τις ροές δεδομένων σε ένα δίκτυο, η δειγματοληψία (sampling) έχει επικεντρώσει το ενδιαφέρον της επιστημονικής κοινότητας ως ένας τρόπος συλλογής στατιστικών στοιχείων για τις ροές του δικτύου. Ως δειγματοληψία ορίζεται η διαδικασία με την οποία επιλέγουμε ένα μέρος από τα

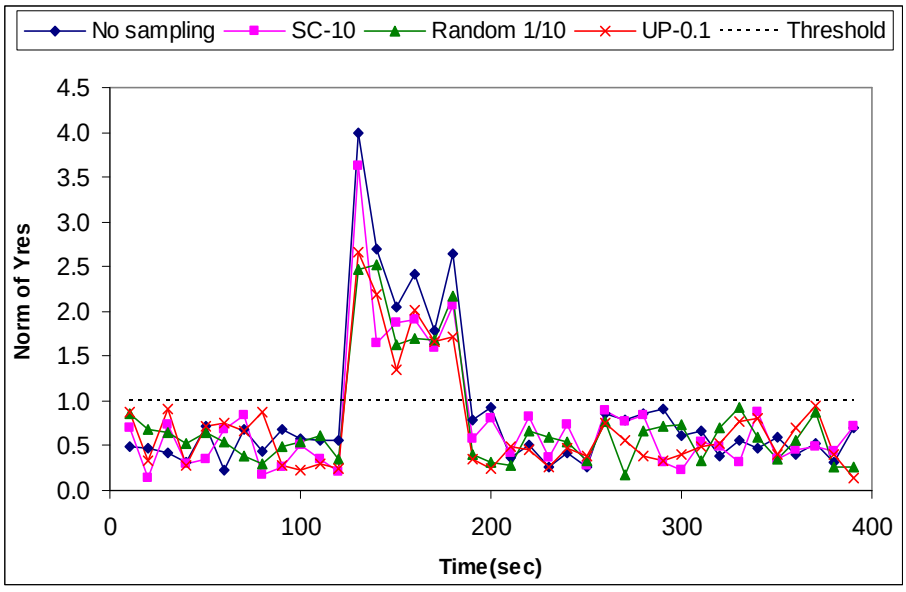
δεδομένα του συστήματος το οποίο μελετούμε, εξάγοντας στη συνέχεια συμπεράσματα για τη συνολική κατάσταση του συστήματος από τα συλλεχθέντα δεδομένα [Gand06] [Gand09]. Η πρακτική αυτή εφαρμόζεται συχνά στα δίκτυα ευρείας ζώνης με τη χρήση του πρωτοκόλλου NetFlow στους κεντρικούς δρομολογητές, όπου ο αριθμός των πακέτων που διέρχονται είναι τεράστιος. Οι κλασικότερες μορφές δειγματοληψίας, είναι οι ακόλουθες:

- ***Συστηματική δειγματοληψία (Systematic Sampling).*** Η συστηματική δειγματοληψία περιγράφει τη διαδικασία επιλογής των σημείων έναρξης και τη διάρκεια των διαστημάτων επιλογής σύμφωνα με μια ντετερμινιστική (deterministic) συνάρτηση. Πρακτικά αυτό σημαίνει ότι διαλέγουμε περιοδικά το κ-πακέτο κάθε φορά.
- ***Τυχαία δειγματοληψία ν-από-N (Random n-out-of-N sampling).*** Στην τυχαία δειγματοληψία ν-από-N, ο αρχικός πληθυσμός των στοιχείων διαιρείται σε τμήματα των N πακέτων το κάθε ένα, ενώ ν στο πλήθος πακέτα επιλέγονται τυχαία από κάθε τμήμα. Ένα παράδειγμα θα ήταν να παραχθούν οι διαφορετικοί τυχαίοι αριθμοί ν από το διάστημα [1,N] και να επιλεγούν όλα τα πακέτα που έχουν μια θέση πακέτων ίση με έναν από τους τυχαίους αριθμούς. Συνήθως αυτή η μεθοδολογία χρησιμοποιείται με ν=1.
- ***Ομοιόμορφη πιθανολογική τυχαία δειγματοληψία (Uniform Probabilistic Random Sampling).*** Στην πιθανολογική δειγματοληψία η απόφαση για το εάν ένα πακέτο επιλέγεται ή όχι γίνεται σύμφωνα με μια προκαθορισμένη πιθανότητα επιλογής. Για την ομοιόμορφη πιθανολογική τυχαία δειγματοληψία κάθε πακέτο επιλέγεται ανεξάρτητα με την ίδια πιθανότητα p.

Για την εξαγωγή συμπερασμάτων μελετήθηκε η σύνδεση μεταξύ του ΕΜΠ και του ΕΔΕΤ που συνδέει το ΕΜΠ με το Διαδίκτυο. Κατά τη διάρκεια διεξαγωγής του πειράματος, αυτή η σύνδεση είχε μια μέση κίνηση της τάξης των 70-80 Mbit/sec και περίπου 20000 πακέτα/sec, και περιείχε ένα πλούσιο μίγμα κίνησης αποτελούμενο από κίνηση web, ηλεκτρονικού ταχυδρομείου, FTP καθώς επίσης και κίνηση από εφαρμογές p2p. Εφαρμόστηκε μια κατανεμημένη επίθεση άρνησης υπηρεσίας (συγκεκριμένα επίθεση TCP SYN) ενάντια σε έναν υπολογιστή μέσα στο ΕΜΠ, η οποία παράχθηκε από ένα πραγματικό εργαλείο επίθεσης. Δημιουργήθηκαν τεχνητά

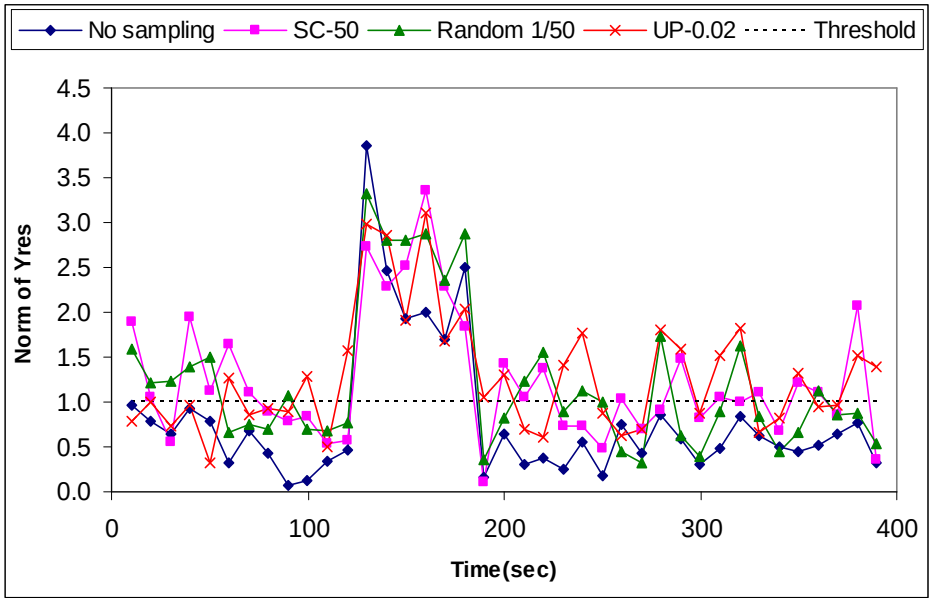
τρείς διαφορετικού μεγέθους επιθέσεις με αναλογία σε πακέτα ως προς την κανονική κίνηση 20%, 10% και 5% αντίστοιχα. Τα αποτελέσματα και των τριών ήταν όμοια και αναλογικά και για το λόγο αυτό θα αναλυθούν τα αποτελέσματα μονάχα από την επίθεση με λόγο 5% που παρουσιάζει περισσότερο ενδιαφέρον λόγω του σχετικά μικρού μεγέθους της. Τα μετρικά τα οποία χρησιμοποιήθηκαν ήταν ο αριθμός ροών, ο αριθμός πακέτων, ο αριθμός ροών TCP, ο αριθμός πακέτων TCP, ο αριθμός μικρών ροών (ροές με 2-3 πακέτα μόνο), ο αριθμός πακέτων SYN, και ο αριθμός πακέτων FIN στη μονάδα του χρόνου.

Το Σχήμα 3.13 αντιστοιχεί στο σενάριο όπου μια επίθεση με 1000 πακέτα/sec και αναλογία 5% σε σχέση με την κανονική κίνηση ανιχνεύεται με βάση δεδομένα τα οποία έχουν υποστεί δειγματοληψία με λόγο 1/10 ως προς τον αριθμό πακέτων. Με SC, Random και UP αναφερόμαστε αντίστοιχα στις τρεις προαναφερθείσες μεθόδους δειγματοληψίας: Systematic Sampling, Random n-out-of-N sampling και Uniform Probabilistic.



Σχήμα 3.13: Επίθεση με ποσοστό 5% - Δειγματοληψία 1/10

Το Σχήμα 3.14 παρουσιάζει το σενάριο όπου μια επίθεση με 1000 πακέτα/sec και αναλογία 5% σε σχέση με την κανονική κίνηση ανιχνεύεται με βάση δεδομένα τα οποία έχουν υποστεί δειγματοληψία με λόγο 1/50. Η ίδια συμπεριφορά εμφανίζεται και για δειγματοληψία με λόγο 1/100.



Σχήμα 3.14: Επίθεση με ποσοστό 5% - Δειγματοληψία 1/50

Όπως απεικονίζεται στο Σχήμα 3.13 και στο Σχήμα 3.14, η προτεινόμενη μεθοδολογία ανίχνευσης ανωμαλιών συλλαμβάνει την αντίστοιχη επίθεση που εμφανίστηκε κατά τη διάρκεια του διαστήματος 130-180sec, για όλες τις δειγματοληπτικές μεθόδους. Όπως ήταν αναμενόμενο, η απόδοση της μειώνεται όσο αυξάνεται ο λόγος δειγματοληψίας και η ακρίβεια των δεδομένων μειώνεται. Αυτό που όμως είναι άξιο αναφοράς είναι ότι η απόδοση της προτεινόμενης μεθοδολογίας είναι ανεξάρτητη από το είδος της δειγματοληψίας που εφαρμόζεται και επηρεάζεται μονάχα από το λόγο δειγματοληψίας. Αυτό είναι ένα σημαντικό πλεονέκτημα, καθώς η τυχαία και η ομοιόμορφη δειγματοληψία οι οποίες θεωρητικά αλλοιώνουν λιγότερο τα στατιστικά στοιχεία ενός δείγματος είναι δυσκολότερο να εφαρμοστούν σε κάποιο σύστημα (π.χ δρομολογητή) και απαιτούν περισσότερους υπολογιστικούς πόρους.

3.6. Προτεινόμενη αρχιτεκτονική βασισμένη σε Τεχνολογία Πλέγματος

3.6.1. Εισαγωγή στη Τεχνολογία Πλέγματος Συσκευών

Στο τμήμα αυτό θα παρουσιαστεί η εφαρμογή της προτεινόμενης μεθοδολογίας στα δίκτυα ευρείας ζώνης, με τη βοήθεια της τεχνολογίας Πλέγματος. Το Πλέγμα (Grid) είναι ένας μηχανισμός διαμοίρασης ετερογενών πόρων μέσα σε ένα δίκτυο ευρείας ζώνης και γνωρίζει έντονη ανάπτυξη τα τελευταία χρόνια. Προσφέρει τα εργαλεία για την ασφαλή και αποδοτική διαμοίραση πόρων όπως υπολογιστική ισχύς, μνήμη και χωρητικότητα για την αποθήκευση δεδομένων. Χρησιμοποιείται παραδοσιακά σε επιστημονικά πεδία των οποίων οι εφαρμογές χρειάζονται τεράστια ποσά υπολογιστικής ισχύος και χωρητικότητας, όπως για παράδειγμα η πυρηνική φυσική και η μετεωρολογία. Πρόσφατα οι τεχνολογίες Πλέγματος επεκτάθηκαν ώστε να επιτρέψουν απομακρυσμένο έλεγχο και συλλογή δεδομένων από συσκευές (Πλέγμα Συσκευών - Instrumentation Grid) [GRIDCC].

Το Πλέγμα Συσκευών εστιάζει στη δημιουργία μιας συνεκτικής συλλογής υπηρεσιών που επιτρέπουν την απομακρυσμένη ρύθμιση και λειτουργία μιας συσκευής. Για το λόγο αυτό εισάγει ένα νέο πόρο στο Πλέγμα: το στοιχείο «Συσκευή Πλέγματος» (Instrument Element). Η Συσκευή Πλέγματος είναι στην ουσία ένα σύνολο από Υπηρεσίες Ιστού (Web Services) που επιτρέπουν την απομακρυσμένη διαχείριση μιας πραγματικής συσκευής. Παράλληλα, το στοιχείο αυτό προσφέρει τη δυνατότητα για δια-δραστική συνεργασία με άλλους πόρους του Πλέγματος. Ο στόχος είναι η εξυπηρέτηση εφαρμογών που λειτουργούν σε πραγματικό χρόνο και χρειάζονται άμεσες αποκρίσεις από τις απομακρυσμένες συσκευές αλλά και από τους παραδοσιακούς πόρους του Πλέγματος, τα στοιχεία αποθήκευσης (Storage Elements) και υπολογισμών (Computer Elements). Εφαρμογές που αξιοποιούν το νέο αυτό στοιχείο και για τις οποίες έχουν αναπτυχθεί πειραματικές υλοποιήσεις είναι τα συστήματα ανίχνευσης φυσικών καταστροφών όπως πυρκαγιές, σεισμοί αλλά και συστήματα ελέγχου πολύπλοκων και πολυμερών συστημάτων όπως για παράδειγμα συστήματα ηλεκτροδότησης και επιταχυντές σωματιδίων μεγάλου μεγέθους.

Στα πλαίσια του έργου GRIDCC [GRIDCC] αναπτύχθηκε μια εφαρμογή βασισμένη στη τεχνολογία Πλέγματος Συσκευών για την ανίχνευση ανωμαλιών στα δίκτυα ευρείας ζώνης. Η εφαρμογή αυτή αντιμετωπίζει τις απαιτήσεις ενός κατανεμημένου συστήματος ανίχνευσης ανωμαλιών, όπως αυτές περιγράφηκαν στην ενότητα 3.2. Τα βασικά στοιχεία της εφαρμογής είναι τα εξής:

- Οι Συσκευές Πλέγματος που αποτελούνται από διάφορους και ετερογενείς αισθητήρες ανωμαλίας και ελέγχουν διάφορα χαρακτηριστικά του δικτύου. Υπάρχει δυνατότητα απομακρυσμένης διαχείρισης των συσκευών μέσω ενός περιβάλλοντος που ονομάζεται «Εικονικό Δωμάτιο Ελέγχου-Virtual Control Room» με το οποίο ελέγχονται όλοι οι παράμετροι της Συσκευής και καταλήγουν τα μηνύματα ελέγχου.
- Η υπερκείμενη υπηρεσία Πλέγματος που επιτρέπει την πρόσβαση σε διασυνδεμένες απομακρυσμένες Συσκευές Πλέγματος χρησιμοποιώντας το δίκτυο και κατά συνέπεια καθιστά δυνατή τη συνεργασία διαφορετικών δικτυακών περιοχών.
- Η υπηρεσία υποστήριξης αποφάσεων (Decision Support), που παρέχει τους αλγορίθμους που στοχεύουν στη συγχώνευση της συλλεχθείσας γνώσης. Αυτή η υπηρεσία αναλύει τα δεδομένα που λαμβάνει από κάθε Συσκευή, τα οποία προέρχονται ενδεχομένως από ετερογενείς αισθητήρες, με σκοπό να συναγάγει μια ολοκληρωμένη άποψη για τη κατάσταση του συνολικού ελεγχόμενου δικτύου. Χρησιμοποιείται επίσης για να δημοσιεύσει τα συμπεράσματα για τις απειλές ασφάλειας που ανιχνεύει στους διαχειριστές του δικτύου (σε NOCs και CSIRTs), οι οποίοι μπορούν να αναλάβουν την εγκατάσταση απαραίτητων αντίμετρων ενάντια στην ανιχνευμένη ανωμαλία.

Ένας κρίσιμος περιορισμός ενός τέτοιου συστήματος είναι η ανάγκη η ανίχνευση να γίνεται σε σχεδόν πραγματικό χρόνο. Επομένως είναι απαραίτητο η υπερκείμενη υπηρεσία πλέγματος να παρέχει έναν αξιόπιστο μηχανισμό που μπορεί να παραδώσει τα μηνύματα ταχύτατα. Αυτός ο μηχανισμός έχει ως κύρια προτεραιότητα την έγκαιρη παράδοση μηνυμάτων από κάθε Συσκευή Πλέγματος και κατά δεύτερο λόγο την αξιόπιστη παράδοση (ένα καθυστερημένο μήνυμα δεν έχει αξία), και υπό αυτή την έννοια αυτός ο μηχανισμός διαφέρει από αυτούς που χρησιμοποιούνται στο

παραδοσιακό λογισμικό Πλέγματος. Τα μηνύματα ελέγχου και αντίδρασης εντούτοις πρέπει να παραδίδονται με υψηλή αξιοπιστία.

Μια άλλη λειτουργία του πρωτοτύπου αυτού είναι ο έλεγχος και η παρακολούθηση των ίδιων των αισθητήρων. Ένα από τα πιο κοινά προβλήματα στα δίκτυα μεγάλης κλίμακας είναι η ανεπαρκής διαχείριση. Με το λογισμικό Πλέγματος είναι δυνατή η κεντρική ρύθμιση και έλεγχος των δικτυακών στοιχείων, όπως δρομολογητές και μεταγωγοί μέσω της ανταλλαγής μηνυμάτων ελέγχου ανάμεσα στον κεντρικό κόμβο ελέγχου και τις Συσκευές Πλέγματος. Η τεχνολογία Πλέγματος Συσκευών προσφέρει μια διεπαφή στον διαχειριστή των συσκευών, η οποία μπορεί να παρομοιαστεί με ένα εικονικό δωμάτιο ελέγχου, στο οποίο είναι δυνατή η επιτήρηση, ο έλεγχος και ο συντονισμός των Συσκευών κεντρικά.

Η συμβολή της υπηρεσίας Πλέγματος στην ανίχνευση ανωμαλιών γίνεται όμως ακόμα πιο σημαντική στην περίπτωση δικτύων που αποτελούνται από διαφορετικά Αυτόνομα Συστήματα (Autonomous Systems). Ένα Κέντρο Δικτύων (NOC) που ενεργεί ως εικονικός οργανισμός μέσα στο Πλέγμα μπορεί να έχει περιορισμένη και ελεγχόμενη πρόσβαση στους αισθητήρες γειτονικών περιοχών. Τα ζητήματα κρυπτογράφησης των δεδομένων και εμπιστοσύνης ανάμεσα σε διαφορετικά συστήματα επιλύονται με τη βοήθεια της υπάρχουσας υποδομής της τεχνολογίας Πλέγματος. Πιο συγκεκριμένα, οι χρήστες ενός κατανεμημένου συστήματος που βασίζεται στην τεχνολογία Πλέγματος γενικά διαχωρίζονται σε Εικονικούς Οργανισμούς (Virtual Organizations). Οι Εικονικοί οργανισμοί είναι δυνατόν να διαφέρουν στο σκοπό, στην εμβέλεια, στο μέγεθος και στη δομή. Το Πλέγμα επιτρέπει στον ίδιο τον Εικονικό Οργανισμό να διαχειριστεί τους χρήστες του, τους ρόλους και τις δυνατότητές τους, αφήνοντας παράλληλα στους διαχειριστές των πόρων να έχουν τον απόλυτο έλεγχο πρόσβασης στους διάφορους ρόλους, χωρίς να συγκεκριμενοποιούν τους χρήστες. Ο χρήστης παρουσιάζει τα διαπιστευτήρια του σε μια υπηρεσία συνήθως μέσω ενός ψηφιακού πιστοποιητικού, συνεπώς ελέγχοντας το πιστοποιητικό μπορεί να ανακαλύψει κανείς σε ποιο Εικονικό Οργανισμό ανήκει ο χρήστης καθώς και τους ρόλους του. Με βάση αυτά και την τοπική πολιτική πρόσβασης, η υπηρεσία επιτρέπει ή αρνείται την πρόσβαση στον χρήστη.

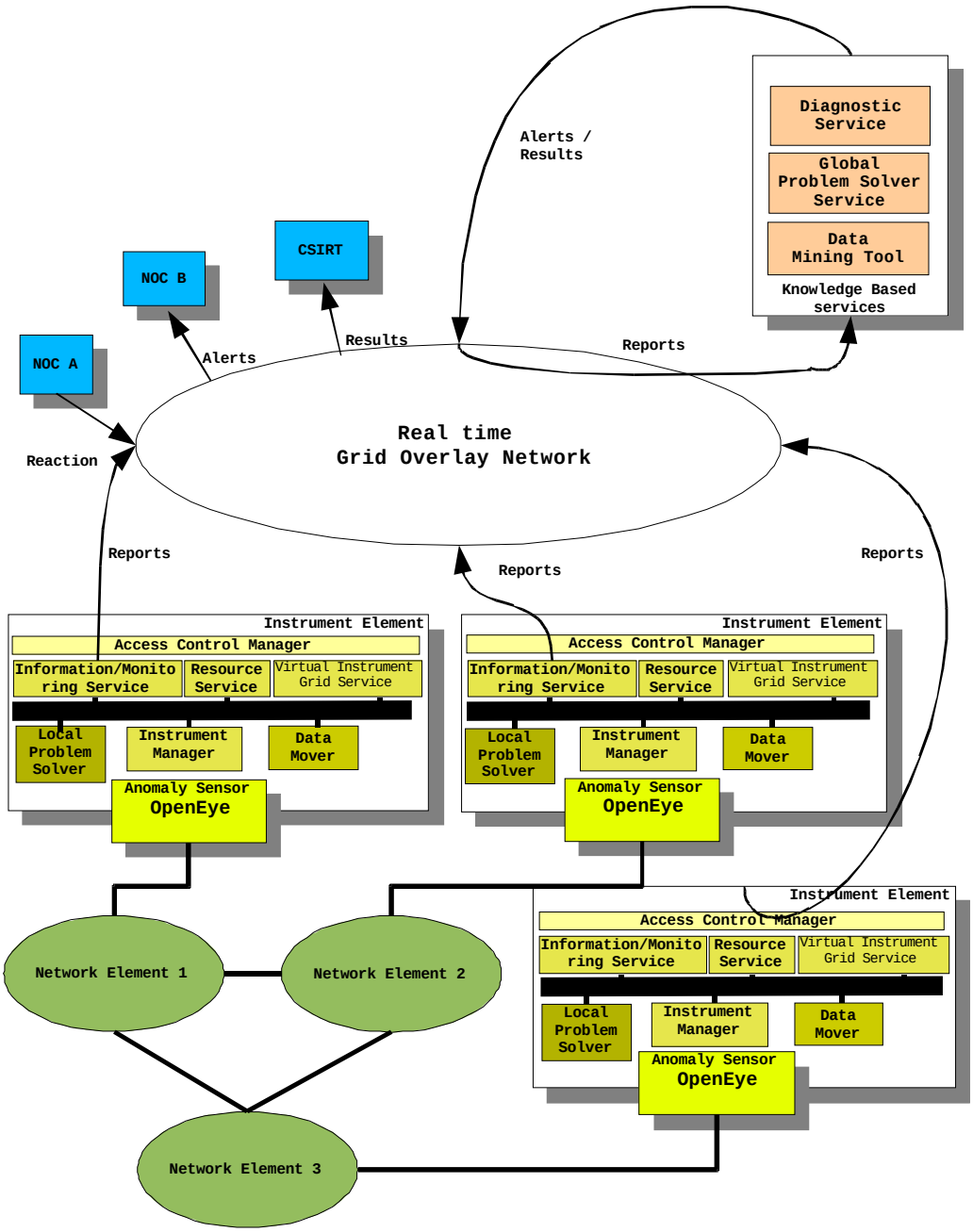
Η ασφαλής ανταλλαγή μηνυμάτων στο πρωτότυπο βασίζεται στο ανοιχτό πρότυπο Ασφάλειας Υπηρεσιών Ιστού (WS-Security). Στόχος του προτύπου αυτού είναι οι

ασφαλείς ανταλλαγές μηνυμάτων ανάμεσα σε Υπηρεσίες Ιστού. Οι κύριες προδιαγραφές ορίζουν ένα αφηρημένο μοντέλο ασφάλειας για την προστασία (εμπιστευτικότητα και ακεραιότητα) των μηνυμάτων. Η εμπιστευτικότητα του μηνύματος εξασφαλίζεται με την κρυπτογράφηση του σώματος του μηνύματος ή τμημάτων αυτού, ενώ η ακεραιότητα και η προέλευση εξακριβώνονται με χρήση ψηφιακών υπογραφών που μπορεί να χρησιμοποιηθούν στην επικεφαλίδα του μηνύματος, στο σώμα ή συνδυασμό αυτών. Για τη κρυπτογράφηση των μηνυμάτων στο Πλέγμα Συσκευών, όπου είναι σημαντικό τα μηνύματα να μεταδίδονται ταχύτατα, επιλέχθηκε η συμμετρική κρυπτογράφηση, η οποία προσφέρει ταχύτητα στη δημιουργία και επικύρωση κρυπτογραφημένων μηνυμάτων [Mora08].

3.6.2. Υλοποίηση του συστήματος ανίχνευσης ανωμαλιών βασισμένη στη Τεχνολογία Πλέγματος Συσκευών

Στην ενότητα αυτή θα παρουσιαστεί η αρχιτεκτονική και τα λειτουργικά μέρη του πρωτότυπου συστήματος ανίχνευσης ανωμαλιών. Πιο συγκεκριμένα, κάθε κόμβος του συστήματος αυτού είναι μια Συσκευή Πλέγματος και έχει σκοπό την ανάλυση της κίνησης που περνά από ένα δικτυακό στοιχείο, την εξαγωγή συγκεκριμένων μετρικών και την αποθήκευσή τους σε μια τοπική βάση. Η Συσκευή αυτή ονομάζεται Τοπικός Ανιχνευτής (Local Detector) και είναι σε θέση να πραγματοποιεί ανίχνευση ανωμαλιών στο τοπικό δίκτυο με βάση τα μετρικά που συλλέγει.

Ο κύριος σκοπός όμως του συστήματος αυτού είναι η εκτέλεση ανίχνευσης ανωμαλιών σε συνολικό επίπεδο, σε ένα μεγάλο δίκτυο ευρείας ζώνης, αποτελούμενο από πολλαπλά ξεχωριστά υποδίκτυα. Αυτό επιτυγχάνεται εφαρμόζοντας τη προτεινόμενη σε αυτή τη διατριβή μέθοδο ανίχνευσης ανωμαλιών, τη συγχώνευση δηλαδή πολλαπλών μετρικών από διάφορα δικτυακά στοιχεία. Για την εφαρμογή της μεθόδου δημιουργήθηκε μια νέα Συσκευή Πλέγματος που ονομάζεται Καθολικός Ανιχνευτής (Global Detector). Ο Καθολικός Ανιχνευτής συλλέγει δεδομένα από τους Τοπικούς Ανιχνευτές χρησιμοποιώντας την τεχνολογία Πλέγματος Συσκευών.



Σχήμα 3.15: Το σύστημα ανίχνευσης ανωμαλιών, βασισμένο στη τεχνολογία Πλέγματος Συσκευών

Κάθε Συσκευή (Instrument Element) χειρίζεται ένα αισθητήρα. Οι αισθητήρες αυτοί μπορεί να βασίζονται σε διαφορετικές τεχνολογίες, όπως συλλογή πακέτων, ροές Netflow [Netf04], ή δεδομένα SNMP, να μετρούν διαφορετικά μετρικά όπως αριθμοί

πακέτων, bytes και ροών στη μονάδα του χρόνου, για διάφορα είδη πακέτων όπως IP/TCP, UDP, ICMP. Άλλα είδη μετρικών είναι ο ρυθμός γεννήσεων νέων ροών, ο αριθμός μικρών ροών (π.χ. σε αριθμό πακέτων) στη μονάδα του χρόνου, λόγοι εισερχόμενων προς εξερχόμενων ροών πακέτων κτλ.

Για το πρωτότυπο ανίχνευσης ανωμαλιών χρησιμοποιήθηκαν ως πηγή δεδομένων ροές Netflow. Το Netflow είναι ένα ανοιχτό δικτυακό πρωτόκολλο το οποίο αναπτύχθηκε από την Cisco Systems για τη παραγωγή στατιστικών στοιχείων της δικτυακής κίνησης. Οι δρομολογητές που υποστηρίζουν το πρωτόκολλο, παράγουν εγγραφές Netflow τις οποίες στέλνουν σε ροές πακέτων UDP. Κάθε ροή θα πρέπει να έχει ως προορισμό ένα ειδικό Συλλέκτη (Collector) που μαζεύει και επεξεργάζεται την πληροφορία. Το πρωτόκολλο έχει διάφορες εκδόσεις με διαφορετική λειτουργικότητα, στην πιο συνηθισμένη του όμως μορφή κάθε εγγραφή έχει πληροφορία για μια ροή (flow) πακέτων. Η ροή στην περίπτωση αυτή ορίζεται από 7 χαρακτηριστικά: διεύθυνση IP πηγής και προορισμού, πηγαία και τελική πόρτα, το πρωτόκολλο, τη διεπαφή του δρομολογητή από την οποία στάλθηκε και τον τύπο της υπηρεσίας (Type of Service). Ο δρομολογητής παράγει μια νέα εγγραφή είτε όταν θεωρήσει ότι η αντίστοιχη ροή έχει λήξει, όταν δηλαδή δεν έχει σταλεί πακέτο που ανήκει σε αυτή τη ροή για κάποιο χρονικό διάστημα, είτε επειδή έχει ρυθμιστεί να αδειάζει τη λανθάνουσα μνήμη του περιοδικά, είτε τελικά αν έχει γεμίσει η λανθάνουσα μνήμη του. Το τελευταίο συμβαίνει σε ακραίες συνθήκες, όπως για παράδειγμα στη περίπτωση μιας ισχυρής επίθεσης απάρνησης υπηρεσίας.

Για τη συλλογή και επεξεργασία των ροών Netflow σε κάθε κόμβο χρησιμοποιήθηκε το σύστημα ανίχνευσης επιθέσεων OpenEye [OpE08]. Το πρωτότυπο αυτό εργαλείο αποτελείται από δυο κύρια μέρη: το Συλλέχτη δεδομένων και τον Ανιχνευτή. Ο Συλλέκτης είναι υπεύθυνος για την ασύγχρονη λήψη ροών δεδομένων από το δρομολογητή στον οποίο έχει ενεργοποιηθεί το Netflow. Ο Συλλέκτης μπορεί να συγκεντρώνει δεδομένα από πολλαπλές πηγές – δρομολογητές, ενώ τα δεδομένα μπορεί να προέρχονται από κίνηση τύπου IPv4 και IPv6 και τελικά διαχωρίζονται σε πολλά διαφορετικά μετρικά (π.χ. πακέτα ICMP, αριθμός σύντομων χρονικά ροών, αναλογίες πακέτων TCP τύπου SYN/FIN, αριθμός παραγωγής ροών και άλλα) που χαρακτηρίζουν την κατάσταση του δικτύου από το οποίο προέρχονται. Ο Πίνακας 3.3 που ακολουθεί παρουσιάζει ενδεικτικά μερικά από αυτά.

Τύπος	Περιγραφή
[Np]	Αριθμός πακέτων
[Nf]	Αριθμός ροών **(flows)**
[Nb]	Αριθμός οκτάδων **– bytes**
[Np_tcp]/[Np_udp]/[Np_icmp]	Αριθμός πακέτων **of TCP/UDP/ICMP**
[Nf_tcp]/[Nf_udp]/[Nf_icmp]	Αριθμός ροών **TCP/UDP/ICMP**
[Nb_tcp]/[Nb_udp]/[Nb_icmp]	Αριθμός οκτάδων**TCP/UDP/ICMP**
[Nft]	Αριθμός σύντομων χρονικά ροών
[Nfs]	Αριθμός μικρών ροών**(**σε αριθμό πακέτων**)**
[Nfsyn]	Αριθμός πακέτων με **flag TCP/SYN**

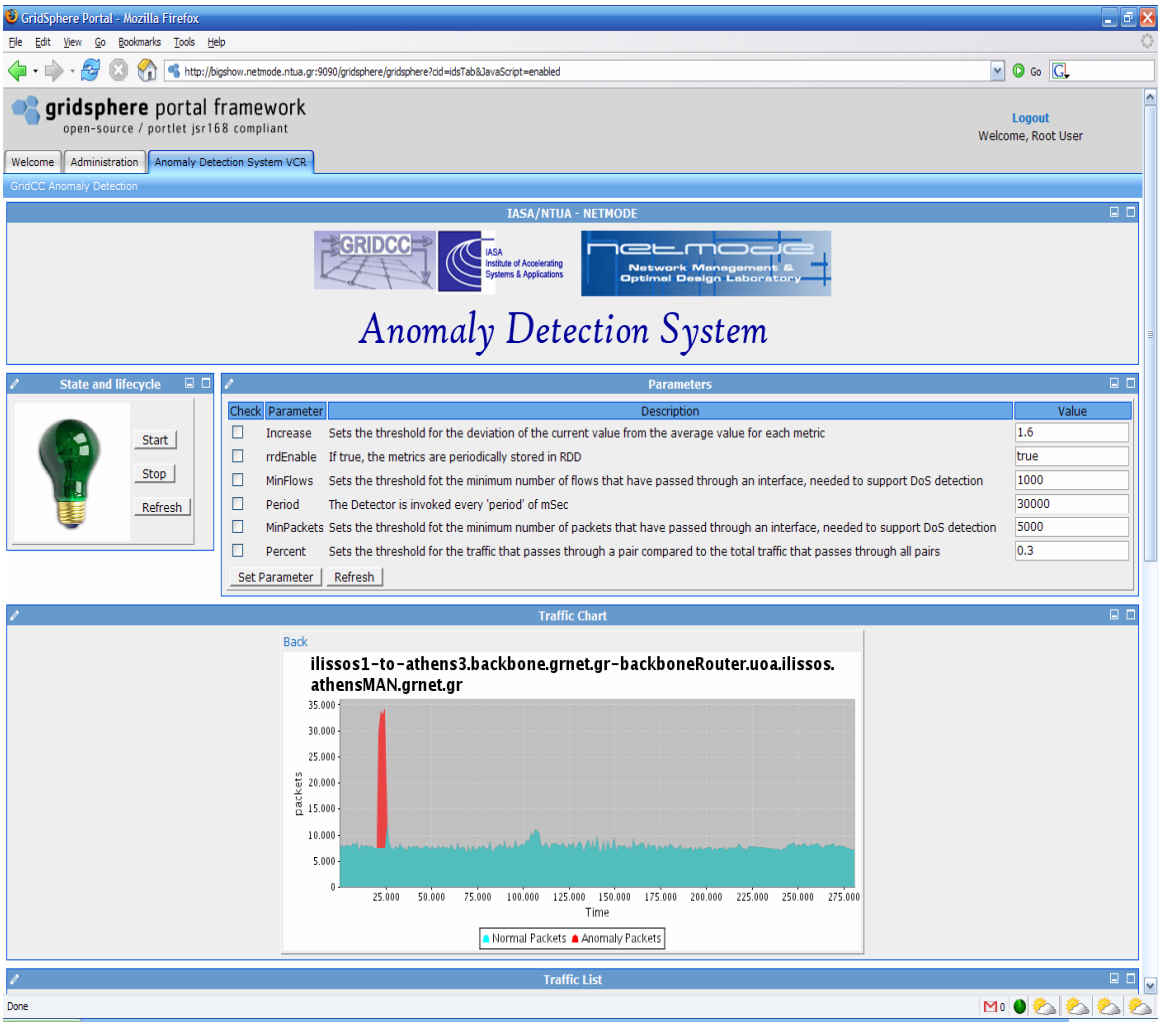
Πίνακας 3.3: Μετρικά, τύποι εικονικών ζεύξεων

Ο Ανιχνευτής καλείται περιοδικά από τον Συλλέκτη. Σε κάθε περίοδο ο Ανιχνευτής υπολογίζει και αποθηκεύει την απόκλιση ανάμεσα στην κανονική κίνηση και την τρέχουσα για κάθε ένα από τα μετρικά που παρακολουθούνται, ως ακολούθως:

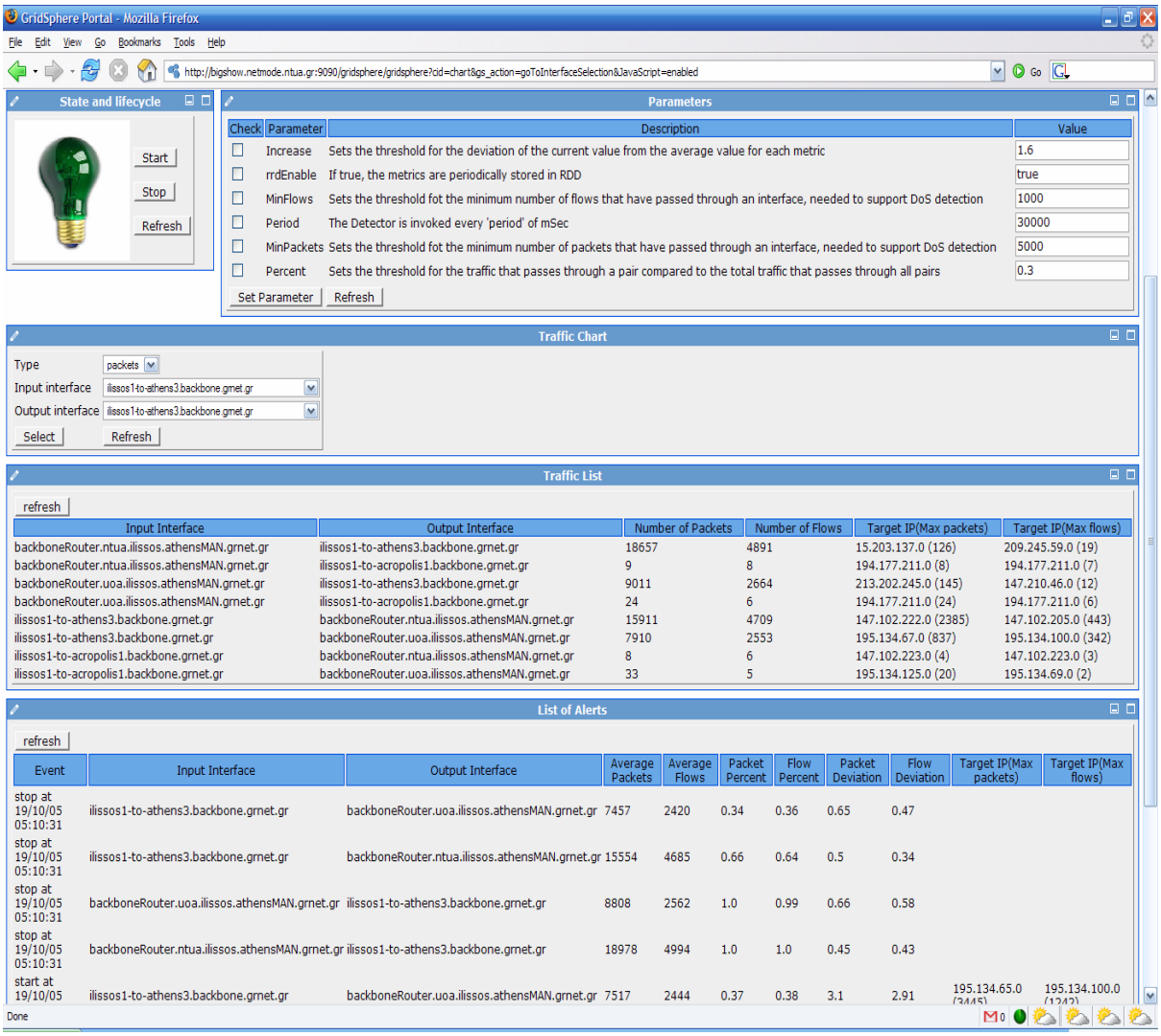
$$\frac{CP_{ij}}{\sum_j CP_{ij}} > k_2 \,,\ \frac{CP_{ij} - AP_{ij}}{AP_{ij}} > k_1 \qquad (10)$$

Με CPij συμβολίζουμε τα τρέχοντα πακέτα από τη διεπαφή i στη j και με APij τον μέσο όρο πακέτων από τη διεπαφή i στη j. Οι ίδιοι λόγοι υπολογίζονται αντίστοιχα και για τα υπόλοιπα μετρικά. Οι λόγοι αυτοί αποθηκεύονται προαιρετικά σε κυκλικές βάσεις για περαιτέρω ανάλυση. Ο πρώτος λόγος ελέγχει αν το ποσοστό της κίνησης που περνάει από το συγκεκριμένο ζεύγος προς την συνολική είναι κάτω από κάποιο κατώφλι. Ο λόγος αυτός είναι χρήσιμος σε κεντρικούς δρομολογητές με μοιρασμένη σχετικά κίνηση ανάμεσα στις διεπαφές. Ο δεύτερος λόγος ελέγχει τη τρέχουσα κίνηση που περνά από το ζεύγος σε σχέση με το μέσο όρο που έχει προκύψει από προηγούμενες περιόδους. Ο μέσος όρος αυτός ανανεώνεται συνεχώς με τη βοήθεια ενός κινούμενου παράθυρου. Με βάση τα στοιχεία αυτά γίνεται ένας απλός τοπικός έλεγχος σε κάθε κόμβο του κατανεμημένου συστήματος ανίχνευσης ανωμαλιών. Τα μετρικά και οι λόγοι που υπολογίζει κάθε Τοπικός Ανιχνευτής αποστέλλονται ασύγχρονα στον κεντρικό κόμβο ο οποίος συγχωνεύει τα εισερχόμενα δεδομένα και παράγει την συνολική εικόνα του δικτύου. Σε περίπτωση που ανιχνεύσει ανωμαλία, προχωρά στον εντοπισμό των ζεύξεων από τις οποίες πέρασε η ανωμαλία.

Η διαχείριση του Τοπικού και του Καθολικού Ανιχνευτή γίνεται κεντρικά από το εικονικό δωμάτιο ελέγχου. Ακολουθούν ενδεικτικές εικόνες του εργαλείου διαχείρισης ενός Τοπικού Ανιχνευτή ο οποίος συλλέγει ροές από ένα κεντρικό δρομολογητή του ΕΔΕΤ στο Σχήμα 3.16 και στο Σχήμα 3.17.



Σχήμα 3.16: Διάγραμμα κίνησης σε κάποια διεπαφή του δρομολογητή από τον οποίο συλλέγει δεδομένα ο Τοπικός Ανιχνευτής



Σχήμα 3.17: Το εικονικό δωμάτιο ελέγχου προσφέρει πληροφορίες για τη κατάσταση του απομακρυσμένου δικτύου

4. Δίκτυα Αισθητήρων

Τα Ασύρματα Δίκτυα Αισθητήρων (Wireless Sensor Networks) συνδυάζουν λειτουργίες όπως η ανίχνευση, η επεξεργασία σήματος και η επικοινωνία, για την παροχή μιας πλατφόρμας ιεραρχικής και αποδοτικής επεξεργασίας πληροφοριών. Σε ένα ΑΔΑ τα δεδομένα που συλλέγονται από τους αισθητήρες υποβάλλονται σε επεξεργασία σε διαδοχικά επίπεδα αφαίρεσης, που κυμαίνεται από τη λεπτομερή μικροσκοπική εξέταση των εκάστοτε στόχων ως τη μακροσκοπική άποψη της συνολικής κατάστασης του περιβάλλοντος στο οποίο βρίσκονται οι αισθητήρες. Ένα κατανεμημένο δίκτυο αισθητήρων είναι συνήθως ένα αυτο-οργανωμένο σύστημα που αποτελείται από ένα πολύ μεγάλο αριθμό κόμβων-αισθητήρων. Οι κόμβοι του δικτύου συνεργάζονται ο ένας με τον άλλον για τη μέτρηση διαφορετικών παραμέτρων που διαφέρουν στο πεδίο του χρόνου και του χώρου.

Τα στοιχείο που διαφοροποιεί ένα ΑΔΑ από τα παραδοσιακά ασύρματα δίκτυα είναι η δομική του μονάδα: ο πομποδέκτης – αισθητήρας. Σε γενικές γραμμές, ο κάθε κόμβος εμπεριέχει τη μονάδα παροχής ισχύος, τον πομποδέκτη, τον αισθητήρα, τον μικροεπεξεργαστή με την μνήμη, ενώ προαιρετικά είναι δυνατό να υπάρχουν μονάδες εντοπισμού θέσης και κίνησης. Το σημαντικό είναι ότι λόγου του μικρού μεγέθους και κόστους που έχουν οι κόμβοι, έχουν περιορισμένες υπολογιστικές δυνατότητες και περιορισμένη ισχύ. Όσον αφορά το στρώμα δικτύου, σε ένα ΑΔΑ πρέπει ο κάθε κόμβος να έχει τη δυνατότητα να εντοπίζει τη θέση του σε σχέση με τους γειτονικούς του κόμβους και να λαμβάνει απόφαση σχετικά με το ποιον θα διαλέξει για να του προωθήσει την προς μετάδοση πληροφορία. Οι αλγόριθμοι που αναφέρονται στις διαδικασίες δρομολόγησης δίνουν ιδιαίτερη έμφαση στην εξοικονόμηση ενέργειας.

Οι περιοχές εφαρμογής είναι διαφορετικές και μπορούν να καλύψουν ποικίλους τύπους δεδομένων συμπεριλαμβανομένου του ήχου, της εικόνας και διάφορων άλλων χημικών και φυσικών ιδιοτήτων. Η επικοινωνία μεταξύ των διαφορετικών κόμβων αισθητήρων υλοποιείται συνήθως μέσω δικτυακών αρχιτεκτονικών που πραγματοποιούν τη μεταφορά των δεδομένων προωθώντας μηνύματα από τον έναν κόμβο στον άλλο. Το μήνυμα μεταφέρεται στο δίκτυο μέσω πολλαπλών βημάτων (hops) και καταλήγει τελικά στον κεντρικό κόμβο (Sink) ο οποίος έχει μεγαλύτερη πολυπλοκότητα και διαθέτει συνήθως σύνδεση με κάποιο άλλο ενσύρματο ή ασύρματο

δίκτυο (π.χ το Διαδίκτυο). Πρόσφατες προσεγγίσεις έχουν εισαγάγει την ιεραρχική επεξεργασία και τη δρομολόγηση με την ομαδοποίηση των κόμβων (node grouping/clustering) και με τη συνάθροιση δεδομένων (data aggregation) σε διαφορετικά επίπεδα ιεραρχίας.

Πιο συγκεκριμένα, η χρήση της συνάθροισης δεδομένων χρησιμοποιείται για την μείωση της ενέργειας που χρειάζεται για την προώθηση μηνυμάτων. Με τη συνάθροιση δεδομένων, κάποιος κόμβος συλλέγει τα αποτελέσματα από διάφορους γειτονικούς αισθητήρες, τα επεξεργάζεται και δημιουργεί ένα μικρότερο μήνυμα το οποίο περιγράφει περιληπτικά την αρχική πληροφορία. Για παράδειγμα, ας υποθέσουμε ότι υπάρχει μια εφαρμογή η οποία ενδιαφέρεται για την μέση τιμή μιας μέτρησης σε όλο το δίκτυο. Ένας μη αποδοτικός τρόπος να βρεθεί η μέση τιμή είναι να στείλει κάθε αισθητήρας τη μέτρηση στον κεντρικό κόμβο (πιθανώς μέσω πολλαπλών βημάτων για κάθε κόμβο) και να γίνει ο υπολογισμός του μέσου όρου εκεί. Ένας περισσότερο αποδοτικός τρόπος είναι κάθε ενδιάμεσος κόμβος να προωθεί το μέσο όρο από τις μετρήσεις των απογόνων του στο δέντρο δρομολόγησης μαζί με το πλήθος των μετρήσεων που εκπροσωπεί. Η συνάθροιση δεδομένων μειώνει σημαντικά το κόστος επικοινωνίας σε ένα ΑΔΑ, όμως δημιουργεί προβλήματα ασφάλειας. Στο επίπεδο δικτύου, δεν είναι δυνατή η από άκρο σε άκρο κρυπτογράφηση των μηνυμάτων, γιατί οι ενδιάμεσοι κόμβοι πρέπει να επεξεργάζονται τα μηνύματα. Παράλληλα καθώς φτάνουμε στο κεντρικό κόμβο τα μηνύματα αποκτούν όλο και μεγαλύτερη αξία καθώς περιέχουν περισσότερη πληροφορία. Αν κάποιος ενδιάμεσος κόμβος αλλοιώσει ένα μήνυμα δημιουργεί σημαντικό πρόβλημα, ειδικά αν βρίσκεται κοντά στον κεντρικό κόμβο. Για την αντιμετώπιση των προβλημάτων αυτών έχουν προταθεί διάφορες λύσεις, μερικές από τις οποίες βασίζονται στην δημιουργία ομάδων (clustering) στο δίκτυο.

Αντίθετα με τα παραδοσιακά ασύρματα δίκτυα, στα οποία η επικοινωνία γίνεται ανάμεσα σε διαφορετικές οντότητες που δεν έχουν σχέση με το δίκτυο και το περιεχόμενο των συνομιλιών ανάμεσα σε δυο γειτονικούς κόμβους είναι συνήθως ασυσχέτιστο, στα ΑΔΑ τα δεδομένα στους γειτονικούς κόμβους είναι ιδιαίτερα συσχετισμένα δεδομένου ότι αφορούν κοντινές μετρήσεις του περιβάλλοντος. Οι περισσότερες εφαρμογές των ΑΔΑ απαιτούν πυκνή διασπορά αισθητήρων στο χώρο προκειμένου να επιτευχθεί ικανοποιητική κάλυψη του περιβάλλοντος αλλά και

προκειμένου το δίκτυο να είναι συνεκτικό. Κατά συνέπεια, συνήθως ένα γεγονός καταγράφεται από πολλαπλούς αισθητήρες, με τον καθένα να κάνει τη μέτρηση από τη δική του «οπτική γωνία». Όμως λόγω της υψηλής πυκνότητας στην τοπολογία του δικτύου, οι παρατηρήσεις που προέρχονται από γειτονικούς αισθητήρες είναι ιδιαίτερα συσχετισμένες [Mehm04]. Παραδείγματα τέτοιων ΑΔΑ είναι εφαρμογές που σχετίζονται με τη μετεωρολογία και την καλλιέργεια, όπου μετρήσεις στην υγρασία, τη θερμοκρασία, την ατμοσφαιρική πίεση σε γειτονικούς κόμβους παρουσιάζουν υψηλή συσχέτιση που μερικές φορές πλησιάζει και το 100% [Hung01]. Επιπρόσθετα παραδείγματα αφορούν το κυκλοφοριακό, όπου ΑΔΑ χρησιμοποιούνται για τη μέτρηση της ροής και της ταχύτητας των αυτοκινήτων για την ανίχνευση ατυχημάτων και τον υπολογισμό παραμέτρων όπως πιθανές καθυστερήσεις και εκτιμώμενη ώρα άφιξης σε καίρια σημεία του αυτοκινητιστικού δικτύου [Pfan05].

Λόγω της κρίσιμης φύσης μερικών εφαρμογών των ΑΔΑ, η ακεραιότητα των μετρήσεων και τα προβλήματα ακρίβειας που μπορούν να προκληθούν από κόμβους που δυσλειτουργούν ή των οποίων τον έλεγχο έχει αποκτήσει κάποιος τρίτος, έχουν πρακτική σπουδαιότητα και αποτελούν σημαντικό ερευνητικό στόχο. Πιο συγκεκριμένα, κόμβοι που δε λειτουργούν σύμφωνα με τις προδιαγραφές για τον ένα ή τον άλλο λόγο, μπορεί να επηρεάσουν διάφορες λειτουργίες του δικτύου όπως η δρομολόγηση, η συνάθροιση δεδομένων, η ψηφοφορία για την εξαγωγή μιας κοινής απόφασης και η δίκαιη κατανομή πόρων. Παράλληλα, επιτιθέμενοι μπορεί να επηρεάσουν την εύρυθμη λειτουργία του δικτύου με διάφορες επιθέσεις, όπως οι επιθέσεις απάρνησης υπηρεσιών και οι επιθέσεις ψευδοκόμβων για τις οποίες θα γίνει ιδιαίτερος λόγος στη συνέχεια. Σε ζωτικής σημασίας ή ευαίσθητες όσον αφορά την ασφάλεια εφαρμογές, είναι απαραίτητο το δίκτυο να διατηρείται λειτουργικό και διαθέσιμο για τη χρήση για την οποία προορίζεται. Η μειωμένη δυνατότητα που έχει κάθε κόμβος ατομικά για την ανίχνευση ανωμαλιών καθιστά αρκετά δύσκολη την εξασφάλιση της διαθεσιμότητας και ορθής λειτουργίας του δικτύου.

Λαμβάνοντας τα παραπάνω υπ’ όψη, στο κεφάλαιο αυτό εξετάζεται η εφαρμογή της προτεινόμενης μεθοδολογίας στα ΑΔΑ, η οποία βασίζεται στη συγχώνευση δεδομένων από τους ασύρματους αισθητήρες. Η προτεινόμενη μέθοδος παρέχει τη δυνατότητα λήψης πολλαπλών μετρήσεων από κάθε κόμβο του δικτύου, αρκεί αυτές να παρουσιάζουν υψηλό βαθμό συσχέτισης. Για την επιτυχή εφαρμογή της μεθοδολογίας

εξετάζεται ο διαχωρισμός του δικτύου σε ομάδες. Ένα από τα κύρια χαρακτηριστικά είναι ότι προσφέρεται η δυνατότητα του συνδυασμού συσχετισμένων μετρήσεων με κατανεμημένο τρόπο, με στόχο την αποκάλυψη ανωμαλιών που επηρεάζουν ένα η περισσότερους γειτονικούς αισθητήρες. Παράλληλα, ο συνδυασμός αποτελεσμάτων από γειτονικές περιοχές του δικτύου αποκαλύπτει συσχετισμένες ανωμαλίες/επιθέσεις που επηρεάζουν πολλαπλές ομάδες από κόμβους. Με τον τρόπο αυτό μπορεί να ανιχνευθεί μια μη κανονική κατάσταση στις μετρήσεις και να αντιμετωπιστούν προβλήματα όπως η ύπαρξη αισθητήρων που δυσλειτουργούν και παίρνουν λανθασμένες μετρήσεις, η ύπαρξη κάποιας τυχαίας ανωμαλίας από εξωτερικό παράγοντα, ή να ανιχνευθούν συντονισμένες επιθέσεις οι οποίες προέρχονται από κάποιον κοινό κακόβουλο παράγοντα.

Η απόδοση και η λειτουργική αποτελεσματικότητα της προτεινόμενης μεθοδολογίας στα ΑΔΑ αξιολογούνται με τη βοήθεια μετεωρολογικών δεδομένων που έχουν συλλεχθεί από ένα κατανεμημένο σύνολο αισθητήρων. Τα δεδομένα αυτά περιέχουν πραγματικές μετεωρολογικές μετρήσεις όπως η ταχύτητα ανέμου, η θερμοκρασία αέρα και η υγρασία από διάφορους γειτονικούς επίγειους σταθμούς στο νησί της Κρήτης. Εξετάστηκαν διάφορα σενάρια ανωμαλίας με διαφορετικό εύρος σε μέγεθος, αλλά και με διαφορετική φύση στο είδος της ανωμαλίας. Χαρακτηριστικό παράδειγμα στη διαφορά της φύσης μιας ανωμαλίας είναι οι τυχαίες και οι συσχετισμένες ανωμαλίες που μπορούν να αντιπροσωπεύσουν αντίστοιχα την ύπαρξη διάφορων ελαττωματικών αισθητήρων ή της επιτυχημένης προσπάθειας ενός επιτιθεμένου να κερδίσει τον έλεγχο κάποιων κόμβων προκειμένου να αλλοιωθούν οι τελικές παρατηρήσεις που φτάνουν στον κεντρικό κόμβο.

Στην ενότητα 4.1 θα παρουσιαστεί μια ανάλυση των κυριότερων απειλών στα ΑΔΑ και στη συνέχεια στην 4.2 θα παρουσιαστεί σχετική βιβλιογραφία που αναφέρει κάποιες υπάρχουσες λύσεις για την αντιμετώπιση των απειλών αυτών. Στην ενότητα 4.3 παρουσιάζεται η εφαρμογή της προτεινόμενης μεθοδολογίας ανίχνευσης ανωμαλιών στα ΑΔΑ και δίνεται έμφαση στη χρήση ομάδων κόμβων στο δίκτυο. Τέλος η απόδοση και η λειτουργική αποτελεσματικότητα της προτεινόμενης μεθοδολογίας παρουσιάζεται στην ενότητα 4.4 και το κεφάλαιο ολοκληρώνεται με κάποια συμπεράσματα για την εφαρμογή της μεθοδολογίας στα ΑΔΑ.

4.1. Απειλές στα ΑΔΑ

Μια από τις σημαντικότερες απειλές για τα ΑΔΑ που προέρχεται από τα παραδοσιακά δίκτυα επικοινωνιών είναι οι επιθέσεις απάρνησης υπηρεσίας. Οι επιθέσεις αυτές εκμεταλλεύονται αδυναμίες στη σχεδίαση κάποιου πρωτοκόλλου επικοινωνίας και δημιουργούν ένα μεγάλο πλήθος αιτήσεων προς το θύμα, καθιστώντας αδύνατη την παροχή υπηρεσιών στους υπόλοιπους κόμβους του δικτύου. Στα ΑΔΑ οι επιθέσεις DoS μπορούν να στοχεύσουν σε διάφορα επίπεδα επικοινωνίας, ξεκινώντας από το φυσικό επίπεδο και καταλήγοντας στο επίπεδο εφαρμογής, ανάλογα με τη χρήση για την οποία προορίζεται το δίκτυο [News04].

Μια γνωστή επίθεση στο φυσικό επίπεδο, είναι το μπλοκάρισμα των ραδιοσυχνοτήτων (frequency jamming) που χρησιμοποιούν οι κόμβοι του δικτύου. Ένας επιτιθέμενος μπορεί να αναστατώσει ολόκληρο το δίκτυο με τυχαία διανεμημένους κόμβους οι οποίοι να εκπέμπουν συνεχώς εμποδίζοντας την επικοινωνία των γειτονικών κόμβων. Για τα δίκτυα που χρησιμοποιούν μια συχνότητα εκπομπής, αυτή η επίθεση είναι απλή και αποτελεσματική. Ένας επιτιθέμενος μπορεί επίσης να αλλοιώσει το υλικό των κόμβων φυσικά και τελικά να αποκτήσει πρόσβαση και στο λογισμικό. Ρεαλιστικά, δεν είναι δυνατό να ελέγχεται η ασφάλεια σε δίκτυα που αποτελούνται από εκατοντάδες κόμβους που είναι εξαπλωμένοι σε αρκετά τετραγωνικά χιλιόμετρα. Σε τέτοια δίκτυα όπου η φυσική πρόσβαση στους κόμβους δεν μπορεί να εμποδιστεί, ο επιτιθέμενος μπορεί να βλάψει ή να αντικαταστήσει ολόκληρο τον αισθητήρα ή μέρος του υλικού του αποκτώντας πρόσβαση σε υψηλότερα επίπεδα επικοινωνιών.

Στο επίπεδο ζεύξης οι επιθέσεις DoS στοχεύουν στις συγκρούσεις πακέτων προκειμένου να δημιουργήσουν προβλήματα στο δίκτυο. Πιο συγκεκριμένα, αλλάζοντας μόνο ένα byte ενός πακέτου μπορεί να προκαλέσουν αναμεταδόσεις με ελάχιστη ενέργεια. Παράλληλα, αυτός ο τύπος της επίθεσης εκτός από καθυστερήσεις εξασθενίζει και τις μπαταρίες των κόμβων που αναμεταδίδουν τα πακέτα.

Στο επίπεδο δικτύου οι επιθέσεις έχουν να κάνουν κυρίως με τη δρομολόγηση. Η πιο απλή περίπτωση είναι κάποιος κακόβουλος κόμβος να μην προωθεί κάποια ή όλα από τα μηνύματα που διέρχονται από αυτόν. Εναλλακτικά, ο κακόβουλος κόμβος μπορεί να προωθεί τα μηνύματα σε λάθος κόμβους αλλοιώνοντας τη δρομολόγηση. Μια ακόμα πιο επικίνδυνη μορφή επίθεσης απάρνησης υπηρεσίας είναι αυτή που

εκμεταλλεύεται πρωτόκολλα Distance-Vector με χαρακτηριστικό παράδειγμα ένας κόμβος να διαφημίζει δρομολόγηση προς διάφορα μέρη του δικτύου με μηδενικό κόστος. Τέτοιες επιθέσεις όταν στοχεύουν συγκεκριμένα μέρη του δικτύου μπορούν να προκαλέσουν «τρύπες» στα σημεία αυτά, καθώς η έντονη αποστολή μηνυμάτων εξασθενίζει και τελικά νεκρώνει τους κόμβους που βρίσκονται εκεί.

Μια σημαντική απειλή για τα ΑΔΑ είναι η επίθεση ψευδοκόμβων “Sybil attack” [Wood02]. Κατά την επίθεση αυτή, ένας κακόβουλος κόμβος συμπεριφέρεται σαν να ήταν ένα σύνολο από κόμβους, παρουσιάζοντας πολλαπλές ταυτότητες. Ο επιτιθέμενος στην χειρότερη περίπτωση δημιουργεί ένα αριθμό από νέες ταυτότητες κόμβων, χρησιμοποιώντας μονάχα μια συσκευή. Οι ψευδοκόμβοι μπορεί να έχουν δημιουργήσει καινούριες ταυτότητες, ή να χρησιμοποιούν ταυτότητες που έχουν κλέψει από κανονικούς κόμβους. Η δεύτερη περίπτωση είναι αναγκαία για τον επιτιθέμενο, αν το δίκτυο έχει κάποιο σύστημα αναγνώρισης ψεύτικων ταυτοτήτων. Παράλληλα, οι επιθέσεις αυτές μπορεί να διαφοροποιηθούν όσον αφορά στην παράλληλη ή μη παράλληλη χρήση των ψευδοκόμβων. Πιο συγκεκριμένα, στην πρώτη περίπτωση ο επιτιθέμενος ενεργοποιεί όλες τις πλαστές ταυτότητες ταυτόχρονα στο δίκτυο. Ακόμα και αν το υλικό του κόμβου δε του επιτρέπει να δρα ως πάνω από ένας ψευδοκόμβος κάθε φορά, μπορεί να αλλάζει κυκλικά την ταυτότητά του με αποτέλεσμα να φαίνονται όλοι οι ψευδοκόμβοι ενεργοί ταυτόχρονα. Εναλλακτικά, ο επιτιθέμενος μπορεί να παρουσιάζει στο δίκτυο ένα μεγάλο αριθμό ταυτοτήτων, χρησιμοποιώντας ταυτόχρονα ένα υποσύνολο αυτών κάθε φορά. Οι ψευδοκόμβοι μπορεί να υποκρίνονται ότι αποχωρούν από το δίκτυο και πιθανόν να συνδέονται ξανά αργότερα. Παράλληλα, υπάρχει η πιθανότητα να ανταλλάσσονται περιοδικά ψεύτικες ταυτότητες ανάμεσα σε διαφορετικούς πραγματικούς κόμβους, με αποτέλεσμα οι ψευδοκόμβοι να αλλάζουν θέση και να προκαλούν μεγαλύτερη σύγχυση στο δίκτυο. Οι επιθέσεις που προκαλούν οι ψευδοκόμβοι μπορεί να στοχεύουν διάφορες λειτουργίες των ΑΔΑ, όπως η δρομολόγηση, η συνάθροιση δεδομένων και η ψηφοφορία και να δεσμεύουν τους πόρους του δικτύου, προκαλώντας απάρνηση υπηρεσιών. Στην περίπτωση της δρομολόγησης, οι ψευδοκόμβοι μπορεί να δυσκολεύουν την αποστολή μηνυμάτων, καθώς η προσθήκη ψευδοκόμβων μπορεί να προκαλέσει βρόγχους στη δρομολόγηση, τεχνητές καθυστερήσεις επειδή οι ψευδοκόμβοι υποκρίνονται ότι μεταφέρουν μηνύματα από τον ένα στον άλλο, ή ακόμα και την απόρριψη μηνυμάτων εξαντλώντας

το χρόνο ζωής του πακέτου (χρήση TTL). Σε περίπτωση που το ΑΔΑ χρησιμοποιεί τεχνικές όπως η συνάθροιση δεδομένων για οικονομία σε πόρους ή η ψηφοφορία για την εξαγωγή μιας απόφασης από μια ομάδα, τότε είναι προφανές πως ο επιτιθέμενος με την εισαγωγή ψευδοκόμβων στο δίκτυο, μπορεί να αλλοιώσει το αποτέλεσμα στο βαθμό που του επιτρέπουν το πλήθος των ταυτοτήτων που έχει εισαγάγει.

4.2. Ανίχνευση εισβολών στα δίκτυα αισθητήρων

Στα δίκτυα αισθητήρων, τα δεδομένα που προέρχονται από τους διάφορους κόμβους πρέπει να εξεταστούν δυναμικά και να συνδυαστούν με πρότυπα προκειμένου να ανιχνευθούν οι πιθανές ανωμαλίες. Λόγω της απαίτησης για την υποστήριξη εφαρμογών υψηλής κρισιμότητας, οι αισθητήρες πρέπει να κατέχουν μηχανισμούς για τη διασφάλιση των επικοινωνιών και για την επικύρωση των δεδομένων που συλλέγουν. Σε αντίθεση με τα δίκτυα ευρείας ζώνης στα οποία η ανίχνευση εισβολών έχει μελετηθεί αρκετά, στα δίκτυα αισθητήρων είναι ακόμα σε πρώιμα στάδια, καθώς δεν έχουν αποκρυσταλλωθεί οι διαδικασίες και τα πρωτόκολλα επικοινωνίας. Στις εργασίες [News04] [Wood02] [Perr04] παρουσιάζονται διάφορα σενάρια επίθεσης που εκμεταλλεύονται τις αδυναμίες των ΑΔΑ. Η κλίμακα κάθε εφαρμογής απαιτεί προσεκτικές αποφάσεις για την ισορροπία ανάμεσα στην αποδοτικότητα του δικτύου και την ασφάλεια και αξιοπιστία του. Αναλύοντας αυτά τα ζητήματα οι ερευνητές προσπαθούν να σχεδιάσουν τους μηχανισμούς για να επιτύχουν υψηλότερου επιπέδου ασφάλεια και αξιοπιστία σε αυτά τα δίκτυα.

Το πρόβλημα των λανθασμένων δεδομένων λόγω της ύπαρξης ελαττωματικών κόμβων είτε κόμβων των οποίων τον έλεγχο έχει αποκτήσει κάποιος τρίτος κακόβουλος, έχει υψηλότερη σπουδαιότητα όταν το δίκτυο βασίζεται στη συνάθροιση δεδομένων για την οικονομία σε ενέργεια και δικτυακούς πόρους.. Στη δημοσίευση [Luo05] οι ερευνητές παρουσιάζουν έναν στατιστικό μηχανισμό φιλτραρίσματος δεδομένων σε πραγματικό χρόνο για την ανίχνευση και απόρριψη λανθασμένων μετρήσεων κατά τη διάρκεια της διαδικασίας συνάθροισης των δεδομένων. Υποθέτοντας ότι ένα γεγονός μπορεί να ανιχνευθεί από πολλαπλούς αισθητήρες, κάθε ένας από τους αισθητήρες ανίχνευσης παράγει έναν κώδικα επικύρωσης μηνυμάτων με βάση κάποιο κλειδί που του έχει δοθεί (MAC), και πολλαπλά MAC επισυνάπτονται σε

κάθε μήνυμα που αφορά ένα γεγονός. Δεδομένου ότι το μήνυμα διαβιβάζεται, κάθε κόμβος κατά μήκος της διαδρομής του μηνύματος ελέγχει πιθανοτικά την ακρίβεια των MAC και απορρίπτει τα μηνύματα που κατά τη γνώμη του περιέχουν άκυρα MAC.

Πιο συναφείς με την προτεινόμενη μέθοδο σε αυτή τη διατριβή, είναι οι παρακάτω μέθοδοι ανίχνευσης ανωμαλιών στα ΑΔΑ. Συγκεκριμένα, στη δημοσίευση [Wang03] οι ερευνητές παρουσιάζουν ένα υπόβαθρο δειγματοληπτικών μηχανισμών για τον έλεγχο της ορθότητας των δεδομένων πριν τη συνάθροιση και τη μεταφορά τους σε ψηλότερο επίπεδο. Ο στόχος είναι η απόρριψη ανώμαλων τιμών στα δεδομένα, ώστε το δίκτυο να επιστρέφει μετρήσεις με αποδεκτή ακρίβεια ακόμα και όταν μέρος των αισθητήρων του δυσλειτουργούν. Το μειονέκτημα της μελέτης αυτής είναι ότι οι έλεγχοι που παρουσιάζονται εξυπηρετούν μόνο ως απόδειξη της ορθότητας της προτεινόμενης αρχιτεκτονικής, καθώς βασίζονται σε πολύ απλοϊκούς αλγόριθμους όπως η εύρεση ελάχιστων, μέγιστων τιμών και διάμεσων στις μετρήσεις των αισθητήρων.

Στη δημοσίευση [Tana05] προτείνεται μια ανάλυση που βασίζεται σε χωρικές και χρονικές συσχετίσεις και ονομάζεται “Abnormal Relationships Test (ART)” . Στόχος είναι η ανίχνευση ανώμαλων τιμών στις μετρήσεις των αισθητήρων. Η μέθοδος βασίζεται σε ελέγχους της συσχέτισης των μετρήσεων ανάμεσα σε γειτονικούς κόμβους. Κάθε κόμβος αποθηκεύει τις τιμές συσχέτισης με τους γείτονές του σε ένα κινούμενο χρονικό παράθυρο, έτσι αν κάποιος γείτονας αρχίσει να δηλώνει μετρήσεις που έχουν χαμηλότερο βαθμό συσχέτισης με αυτές στο παρελθόν σημαδεύεται ως ύποπτος. Η ανίχνευση των προβληματικών κόμβων επιτυγχάνεται συνεργατικά και οι κόμβοι που θεωρούνται ύποπτοι από κάποιον αριθμό γειτόνων απομονώνονται από το δίκτυο.

Στη δημοσίευση [Palm03] οι ερευνητές περιγράφουν μια τεχνική για ανίχνευση ανωμαλιών στις μετρήσεις αισθητήρων σε πραγματικό χρόνο. Η τεχνική αυτή εφαρμόζεται σε ένα δίκτυο μεγάλης κλίμακας, κατανέμοντας την επεξεργασία μετρήσεων σε ομάδες γειτονικών κόμβων. Ένα από τα μειονεκτήματα της μεθόδου είναι ότι προϋποθέτει την γνώση κατανομών που αφορούν τη φύση των δεδομένων, όπως για παράδειγμα η συνάρτηση πυκνότητας κατανομής, που συνήθως δεν είναι γνωστές ή διαθέσιμες.

4.3. Εφαρμογή της προτεινόμενης μεθοδολογίας στα ΑΔΑ

Το ΑΔΑ στο οποίο στοχεύουμε την εφαρμογή της προτεινόμενης μεθοδολογίας είναι ένα δίκτυο με διάφορους ετερογενείς κόμβους αισθητήρων, όπου κάθε κόμβος μπορεί να έχει διαφορετικές δυνατότητες και να εκτελεί διαφορετικές λειτουργίες. Παραδείγματος χάριν, μερικοί κόμβοι μπορούν να έχουν περισσότερη διαθέσιμη ενέργεια (στοιχείο το οποίο μεταφράζεται σε κατοχή μεγαλύτερης μπαταρίας) και ισχυρότερη ικανότητα επεξεργασίας δεδομένων, άλλοι μπορούν να συναθροίσουν και να αναμεταδίδουν δεδομένα, ενώ άλλοι παίρνουν μόνο μετρήσεις από το περιβάλλον χωρίς να αναμεταδίδουν μηνύματα για λογαριασμό των άλλων κόμβων του δικτύου.

Ο αλγόριθμος ανίχνευσης ανωμαλιών στις μετρήσεις των αισθητήρων συσχετίζει μετρικά από γειτονικούς κόμβους. Για την αποκέντρωση του αλγόριθμου ανίχνευσης το ΑΔΑ χωρίζεται σε ομάδες από αισθητήρες. Ο διαχωρισμός μπορεί να γίνει είτε στατικά κατά την αρχική δημιουργία του δικτύου, είτε δυναμικά αν το δίκτυο αναδιοργανώνεται περιοδικά ή αλλάζουν σημαντικά οι παράμετροι του περιβάλλοντος που ελέγχουν οι αισθητήρες. Σε κάθε περίπτωση, υποθέτουμε ότι ο διαχωρισμός του δικτύου σε ομάδες βασίζεται στο βαθμό της συσχέτισης ανάμεσα στις μετρήσεις των κόμβων, όπως θα περιγράψουμε αναλυτικά στη συνέχεια και πραγματοποιείται από ειδικούς κόμβους που λέγονται κύριοι κόμβοι και έχουν περισσότερη διαθέσιμη ενέργεια και ισχυρότερη ικανότητα επεξεργασίας δεδομένων .

Η διαδικασία ανίχνευσης ανά ομάδα μπορεί να χωριστεί σε δυο βασικά μέρη, με βάση τη περιγραφή της μεθοδολογίας στο κεφάλαιο 2.1: τη διαδικασία κατάρτισης του συστήματος (offline analysis), όπου δημιουργείται το μοντέλο του περιβάλλοντος από μετρήσεις που θεωρούνται κανονικές (δεν περιέχουν ανωμαλίες) και τη διαδικασία ανίχνευσης που γίνεται σε πραγματικό χρόνο (real time analysis). Η διαδικασία ανίχνευσης εκτελείται κεντρικά ανά ομάδα και συγκρίνοντας τη τρέχουσα κατάσταση του περιβάλλοντος με τη μοντελοποιημένη, στοχεύει στην ανίχνευση ανωμαλιών.

Κατά τη διαδικασία κατάρτισης του συστήματος, κάθε κύριος κόμβος συλλέγει και προωθεί στον κεντρικό κόμβο του δικτύου χωρίς περαιτέρω επεξεργασία και συνάθροιση δεδομένων ένα δείγμα μετρήσεων από τους κόμβους που ανήκουν στην ομάδα του. Ο κεντρικός κόμβος επεξεργάζεται τα δεδομένα κάθε ομάδας, τα κανονικοποιεί και αφαιρεί τυχόν ανωμαλίες. Η διαδικασία αυτή που αντιστοιχεί στην

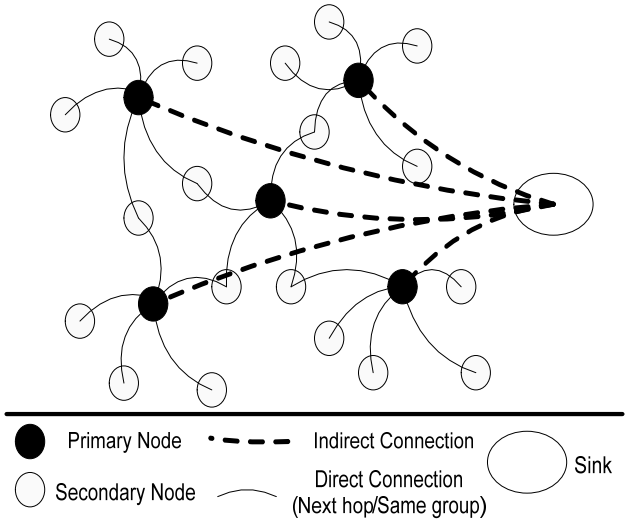
προετοιμασία των δεδομένων κατάρτισης μπορεί να γίνει με τη μέθοδο που περιγράφεται στο κεφάλαιο 2.5. Από τα δεδομένα κατάρτισης προκύπτουν οι ΚΣ, οι οποίες επαρκούν για τη περιγραφή των σημαντικών συσχετίσεων και διακυμάνσεων στις μετρήσεις. Οι ΚΣ επιστρέφονται με μηνύματα ελέγχου σε κάθε κύριο κόμβο. Η διαδικασία αυτή προτείνεται καθώς η δημιουργία των δεδομένων κατάρτισης και η εξαγωγή των ΚΣ είναι απαιτητική σε υπολογιστικούς πόρους.

Η ανίχνευση επιτυγχάνεται με τη Μέθοδο Υποχώρων (Subspace Method) η οποία χρησιμοποιεί τις ΚΣ που εξήχθησαν κατά τη διαδικασία κατάρτισης, για τον διαχωρισμό του διανύσματος που περιγράφει τη τρέχουσα κίνηση σε δύο μέρη: ένα το οποίο ορίζεται ως $\mathbf{y_{norm}}$ και περιλαμβάνει το κομμάτι της κίνησης που θεωρείται ότι δεν περιέχει ανωμαλίες και το διάνυσμα-υπόλοιπο ($\mathbf{y_{res}}$) το οποίο περιλαμβάνει την υπόλοιπη κίνηση. Η ύπαρξη ανωμαλίας στη κίνηση θα έχει σαν αποτέλεσμα να παρατηρηθούν σημαντικές αλλαγές στο διάνυσμα-υπόλοιπο $\mathbf{y_{res}}$, όπως έχει παρουσιαστεί αναλυτικά στην παρουσίαση του αλγόριθμου ανίχνευσης.

Συνεπώς, η υπολογιστικά απαιτητική διαδικασία της παραγωγής των δεδομένων κατάρτισης και της εξαγωγής των ΚΣ γίνεται στον κεντρικό κόμβο του δικτύου που δεν έχει περιορισμούς σε υπολογιστικούς πόρους. Το κόστος του αλγόριθμου ανίχνευσης είναι ότι πρέπει περιοδικά να επαναλαμβάνεται η διαδικασία μεταφοράς μετρήσεων χωρίς συνάθροιση στον κεντρικό κόμβο. Μια λύση στο πρόβλημα αυτό, που θα παρουσιαστεί αναλυτικά στη συνέχεια, είναι να απαιτείται νέος υπολογισμός των ΚΣ μόνο αν η απόκλιση του βαθμού συσχέτισης σε ένα η παραπάνω κόμβους της ομάδας έχει υπερβεί κάποιο όριο.

4.3.1. Ομαδοποίηση κόμβων στα ΑΔΑ

Στο Σχήμα 4.1 παρουσιάζεται η τοπολογία ενός ΑΔΑ με τα παραπάνω χαρακτηριστικά. Η δημιουργία ομάδων βασίζεται στη συσχέτιση των μετρήσεων γειτονικών κόμβων και με τον τρόπο αυτό συνδυάζεται η αποδοτικότερη συνάθροιση και δρομολόγηση των μηνυμάτων με την αποτελεσματικότερη ανίχνευση ανωμαλιών.



Σχήμα 4.1: Αρχιτεκτονική ΑΔΑ και Ομαδοποίηση κόμβων

Με τη λογική αυτή θα μπορούσαν δυο γειτονικοί κόμβοι να ανήκουν σε διαφορετικές ομάδες, αν οι μετρήσεις τους είναι ασυσχέτιστες. Γειτονικοί κόμβοι στα ΑΔΑ θεωρούνται αυτοί που είναι αρκετά κοντά ο ένας στον άλλο ώστε να είναι ο ένας μέσα στην εμβέλεια του άλλου και να μπορούν να επικοινωνήσουν άμεσα. Πιο συγκεκριμένα, σε κάθε ομάδα υποθέτουμε ότι υπάρχει ένας κύριος κόμβος ο οποίος διαθέτει μεγαλύτερη υπολογιστική ισχύ, ενέργεια και πολυπλοκότητα, από τους υπόλοιπους δευτερεύοντες κόμβους. Ο κύριος κόμβος δημιουργεί μια ομάδα ρωτώντας όλους τους γειτονικούς του κόμβους για τις τελευταίες μετρήσεις τους. Ανάλογα με την πυκνότητα του δικτύου και το ποσοστό των κύριων κόμβων προς τους συνολικούς, πρέπει να οριστεί ένα κατώτερο όριο στη συσχέτιση μετρήσεων με βάση το οποίο μπορεί να εισέρθει κάποιος κόμβος σε μια συγκεκριμένη ομάδα. Το κατώφλι πρέπει να οριστεί προσεκτικά ώστε ο γράφος του δικτύου να είναι συνεκτικός. Για τον υπολογισμό της συσχέτισης χρησιμοποιείται ο βαθμός συσχέτισης που δίνεται από τον παρακάτω τύπο:

$$R_{X,Y} = \frac{Cov(X,Y)}{S_X \cdot S_Y} \tag{11}$$

όπου Cov(X,Y) είναι η συνδιακύμανση μεταξύ των τυχαίων μεταβλητών X and Y, ενώ με S_X και S_Y ορίζεται η τυπική απόκλιση που προκύπτει από το δείγμα μετρήσεων για τις μεταβλητές X και Y αντίστοιχα. Μεταβλητές με βαθμό συσχέτισης κοντά στη μονάδα αυξάνονται ή μειώνονται αντίστοιχα προς την ίδια κατεύθυνση, ενώ

μεταβλητές με βαθμό συσχέτισης κοντά στο -1 έχουν αντίθετες αλλαγές. Όταν η απόλυτη τιμή του βαθμού συσχέτισης δεν είναι κοντά στη μονάδα οι μεταβλητές δεν συσχετίζονται και οι αλλαγές στα μεγέθη τους δεν έχουν αντιστοιχία.

Το αποτέλεσμα της διαδικασίας διαχωρισμού είναι ένα σύνολο από ομάδες αισθητήρων των οποίων οι μετρήσεις είναι ιδιαίτερα συσχετισμένες. Αξίζει να σημειωθεί στο σημείο αυτό ότι οι γειτονικές ομάδες μπορεί να έχουν κοινά μέλη καθώς η ύπαρξη κοινών δευτερευόντων κόμβων μπορεί να είναι επιθυμητή σε πολλές περιπτώσεις για την βελτίωση της δυνατότητας ανίχνευσης ανωμαλιών.

Κάθε κύριος κόμβος συγκεντρώνει τις μετρήσεις από όλους τους κόμβους της ομάδας και εφαρμόζει περιοδικά ανίχνευση ανωμαλιών. Αν κάθε κόμβος έχει περισσότερα από ένα αισθητήρια όργανα και τα μετρικά για τα οποία συλλέγει μετρήσεις είναι συσχετισμένα μεταξύ τους, η διαδικασία ανίχνευσης μπορεί να τα χρησιμοποιεί όλα ταυτόχρονα στη ίδια ΑΚΣ. Για τη μοντελοποίηση των πολλαπλών μετρικών ανά κόμβο, μπορεί να δημιουργηθεί ένας εικονικός κόμβος για κάθε μετρικό. Αν οι μετρήσεις δεν περιέχουν ανωμαλίες μπορούν να συναθροιστούν στον κύριο κόμβο κάθε ομάδας και να προωθηθούν στο κεντρικό κόμβο του δικτύου. Αντίθετα, αν ανιχνευτεί ανωμαλία και αναγνωριστεί ο κόμβος ή οι κόμβοι που έχουν το πρόβλημα, οι μετρήσεις τους μπορούν να φιλτραριστούν και τα στοιχεία για τους κόμβους αυτούς να μεταβιβαστούν με μηνύματα ελέγχου στο κεντρικό κόμβο του δικτύου. Ο κεντρικός κόμβος γνωρίζοντας την τοπολογία του δικτύου και έχοντας μια συνολική εικόνα από τις αναφορές από όλες τις ομάδες, είναι σε θέση να συμπεράνει αν κάποια ανωμαλία επηρεάζει πολλαπλές ομάδες και αν ακολουθεί κάποιο μονοπάτι. Στην περίπτωση ενός κινούμενου επιτιθέμενου κόμβου μπορεί να αναγνωρίσει το μονοπάτι που ακολουθεί ελέγχοντας για ανωμαλίες σε γειτονικές ομάδες σε ένα κινούμενο χρονικό παράθυρο. Η διαδικασία αυτή είναι ιδιαίτερα ενδιαφέρουσα και μπορεί να έχει διάφορες χρήσεις ανάλογα με την εφαρμογή στην οποία χρησιμοποιείται το ΑΔΑ. Για παράδειγμα, κάποιο ισχυρό φυσικό φαινόμενο μπορεί να επηρεάζει τους αισθητήρες σε μια περιοχή όπως είναι το μέτωπο μιας πυρκαγιάς, σε περίπτωση που το ΑΔΑ χρησιμοποιείται για τον έλεγχο και την προστασία ενός δάσους. Σε άλλη περίπτωση, θα μπορούσε να υπάρχει ένας ιός ο οποίος μεταδίδεται μολύνοντας γειτονικούς κόμβους. Γενικότερα, οι ομάδες οι οποίες ανέφεραν τη μεγαλύτερη σε μέγεθος ανωμαλία υποδεικνύουν το μονοπάτι και την τρέχουσα θέση της πηγής της ανωμαλίας.

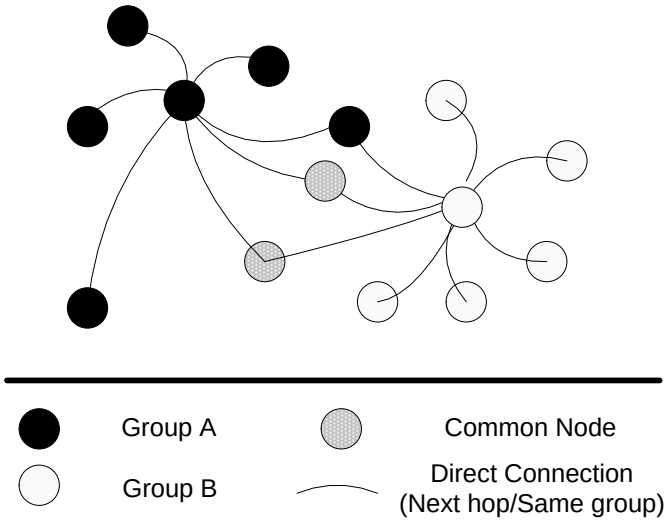
4.3.2. Στατική και Δυναμική ομαδοποίηση κόμβων στα ΑΔΑ

Γενικά μια μέθοδος ανίχνευσης ανωμαλίας εξαρτάται από την κατοχή ενός καλού κανονικού δείγματος δεδομένων. Στην περίπτωσή μας επίσης, ένα ακριβές δείγμα δεδομένων παράγει τις ΚΣ που αντιστοιχούν στις πραγματικές διακυμάνσεις και συσχετίσεις μεταξύ των μετρήσεων των γειτονικών αισθητήρων, και επομένως δημιουργούν ένα ακριβές μοντέλο του περιβάλλοντος. Πρέπει να σημειωθεί ότι η εφαρμογή της προτεινόμενης μεθοδολογίας στα ΑΔΑ δεν απαιτεί κάποια συγκεκριμένη μέθοδο κατάρτισης του αλγορίθμου ανίχνευσης, χρησιμοποιεί το βαθμό συσχέτισης μεταξύ του αρχικού κόμβου και των γειτόνων του ώστε αρχικά να χωρίσει το δίκτυο σε ομάδες και στη συνέχεια να εξάγει τις ΚΣ ανά ομάδα.

Δεδομένου ότι η διαδικασία διαχωρισμού του δικτύου σε ομάδες και η διαδικασία εξαγωγής των ΚΣ είναι υπολογιστικά και δικτυακά δαπανηρές, μπορούν να εκτελεστούν μόνο κατά τη διάρκεια της φάσης δημιουργίας του δικτύου, και κατά συνέπεια η αρχική ομαδοποίηση μπορεί να παραμείνει στατική κατά τη λειτουργία του δικτύου. Όμως σε πολλές περιπτώσεις τα δεδομένα που συλλέγονται εξαρτώνται σημαντικά από το περιβάλλον και οι συσχετίσεις τους μπορεί να αλλάξουν αν υπάρξει κάποια αλλαγή και στο περιβάλλον των αισθητήρων. Για το λόγο αυτό θα πρέπει να γίνεται περιοδικός έλεγχος των βαθμών συσχετίσεων ανάμεσα στον κύριο κόμβο και τους δευτερεύοντες κόμβους κάθε ομάδας. Κάθε κύριος κόμβος θα πρέπει να αποθηκεύει τις μετρήσεις των αισθητήρων σε ένα χρονικά κινούμενο παράθυρο και να υπολογίζει περιοδικά το βαθμό συσχέτισης με κάθε ένα αισθητήρα που ανήκει στην ομάδα του. Αν η απόλυτη τιμή του βαθμού συσχέτισης πέσει κάτω από ένα προκαθορισμένο όριο, τότε ο κόμβος θα πρέπει να απορρίπτεται από την ομάδα και να ζητήσει την ένταξη του σε κάποια γειτονική. Επίσης, ο κεντρικός κόμβος μπορεί να ζητήσει την εκ νέου αναδιοργάνωση του δικτύου αν παρατηρηθούν εκτεταμένες και συνεχείς αναφορές ανωμαλιών σε μια η περισσότερες ομάδες, στοιχείο το οποίο μπορεί να θεωρηθεί ως ένδειξη ότι έχει γίνει κάποια ακραία αλλαγή στο περιβάλλον και οι ΚΣ ανά ομάδα δεν αντιστοιχούν πλέον στη τρέχουσα κανονική κατάσταση. Συνεπώς, αλλαγές στο περιβάλλον μπορεί να προκαλέσουν αλλαγές στις συσχετίσεις των μετρήσεων και τελικά αλλαγές στην ομαδοποίηση των κόμβων του δικτύου.

4.3.3. Η ύπαρξη κοινών κόμβων σε γειτονικές ομάδες αισθητήρων

Όπως περιγράψαμε νωρίτερα, η δημιουργία ομάδων βασίζεται στο βαθμό συσχέτισης ανάμεσα σε κάθε κύριο κόμβο και τους γείτονές του και για το λόγο αυτό δεν απαιτείται οι δευτερεύοντες κόμβοι να ανήκουν αποκλειστικά σε μια ομάδα. Δυο γειτονικές ομάδες μπορεί να μοιράζονται έναν ή περισσότερους κοινούς κόμβους αν ο βαθμός συσχέτισής τους με το κύριο κόμβο κάθε ομάδας είναι αρκετά υψηλός, και εφόσον ο αριθμός κοινών κόμβων ανά ζεύγος γειτονικών ομάδων δεν περνάει κάποιο προκαθορισμένο όριο. Στο Σχήμα 4.2 παρουσιάζεται ένα παράδειγμα ύπαρξης κοινών κόμβων σε γειτονικές ομάδες.



Σχήμα 4.2: ΑΔΑ με κοινούς κόμβους σε γειτονικές ομάδες

Επιτρέποντας την ύπαρξη κοινών κόμβων ανάμεσα σε γειτονικές ομάδες παρέχεται ένα επιπρόσθετο πλεονέκτημα στη διαδικασία εξαγωγής της απόφασης, καθώς είναι δυνατό να συνδυαστούν αποφάσεις από κάθε ομάδα. Ανάλογα με την μέθοδο συγκερασμού των αναφορών από τις γειτονικές ομάδες για την κατάσταση ενός κοινού κόμβου, επιτυγχάνεται η αύξηση του ποσοστού ανίχνευσης ή η μείωση της πιθανότητας λανθασμένης ανίχνευσης. Πιο συγκεκριμένα, χρησιμοποιώντας την ένωση των αναφορών που προέρχονται από δυο τέτοιες ομάδες, υπάρχει μεγαλύτερη πιθανότητα ανίχνευσης ανωμαλίας στις μετρήσεις που προέρχονται από κοινούς κόμβους. Από την

άλλη, χρησιμοποιώντας την τομή των αντίστοιχων αναφορών, μειώνεται η πιθανότητα λανθασμένης ανίχνευσης.

4.4. Πειραματικά αποτελέσματα

Στην ενότητα αυτή εξετάζεται η απόδοση της προτεινόμενης μεθοδολογίας με τη χρησιμοποίηση μετεωρολογικών δεδομένων που έχουν συλλεχθεί από ένα σύνολο κατανεμημένων μετεωρολογικών σταθμών στο νησί της Κρήτης. Τα δεδομένα αποτελούνται από μετρήσεις όπως η ταχύτητα του αέρα, η θερμοκρασία και η υγρασία. Επίσης, παρουσιάζονται συγκριτικά αποτελέσματα ανάμεσα στη προτεινόμενη μέθοδο και μια αντίστοιχη υπάρχουσα μέθοδο που χρησιμοποιεί τις συσχετίσεις ανάμεσα σε μετρήσεις γειτονικών κόμβων και ονομάζεται «Δοκιμή Ανώμαλων Σχέσεων » - “Abnormal Relationships Test (ART)” [Tana05]. Προκειμένου να υπολογίσουμε την αποτελεσματικότητα του αλγορίθμου ανίχνευσης ανωμαλιών χρησιμοποιήθηκαν τα ίδια μετρικά με αυτά στο κεφάλαιο 3.3.1 για τα δίκτυα ευρείας ζώνης: η πιθανότητα ανίχνευσης P_d και η πιθανότητα λανθασμένης ανίχνευσης (false alarm) P_f. Παράλληλα, χρησιμοποιούμε τις καμπύλες Receiver Operating Characteristic (ROC), προκειμένου να εκφράσουμε και να αναπαραστήσουμε τη σχέση ανάμεσα σε αυτές τις πιθανότητες.

Η ανάλυση που παρουσιάζεται σε αυτό το τμήμα είναι βασισμένη σε ένα εκτενές σύνολο πραγματικών μετρήσεων θερμοκρασίας, με τις ανωμαλίες να έχουν προστεθεί μετά τη συλλογή των δεδομένων. Οι αρχικές μετρήσεις που για την κανονική κατάσταση (χωρίς ανωμαλίες) χρησιμοποιήθηκαν αυτούσιες, είναι περιοδικές με χρονική απόσταση μιας ώρας. Το δίκτυο, το οποίο αποτελείται από σαράντα κόμβους, χωρίστηκε σε ομάδες από αισθητήρες με βαθμό συσχέτισης ανά ομάδα μεγαλύτερο του 90%. Ο διαχωρισμός του δικτύου με βάση το κριτήριο αυτό οδήγησε σε τέσσερεις ομάδες με τουλάχιστον δέκα κόμβους ανά ομάδα, με μερικούς κόμβους να ανήκουν σε περισσότερες από μια ομάδες. Οι ανωμαλίες προστέθηκαν σε ένα ποσοστό των κόμβων ανά ομάδα με κυμαινόμενο μέγεθος κατά απόλυτη τιμή από 2% ως 15% σε σχέση με την αρχική τιμή της μέτρησης. Οι κύριοι κόμβοι κάθε ομάδας θεωρούνται απρόσβλητοι από εξωτερικούς παράγοντες και μόνο οι δευτερεύοντες κόμβοι μπορεί να παρουσιάσουν ανωμαλίες στις μετρήσεις. Οι καμπύλες ROC που παρουσιάζουν τις πιθανότητες ανίχνευσης και λανθασμένης ανίχνευσης παράχθηκαν χρησιμοποιώντας

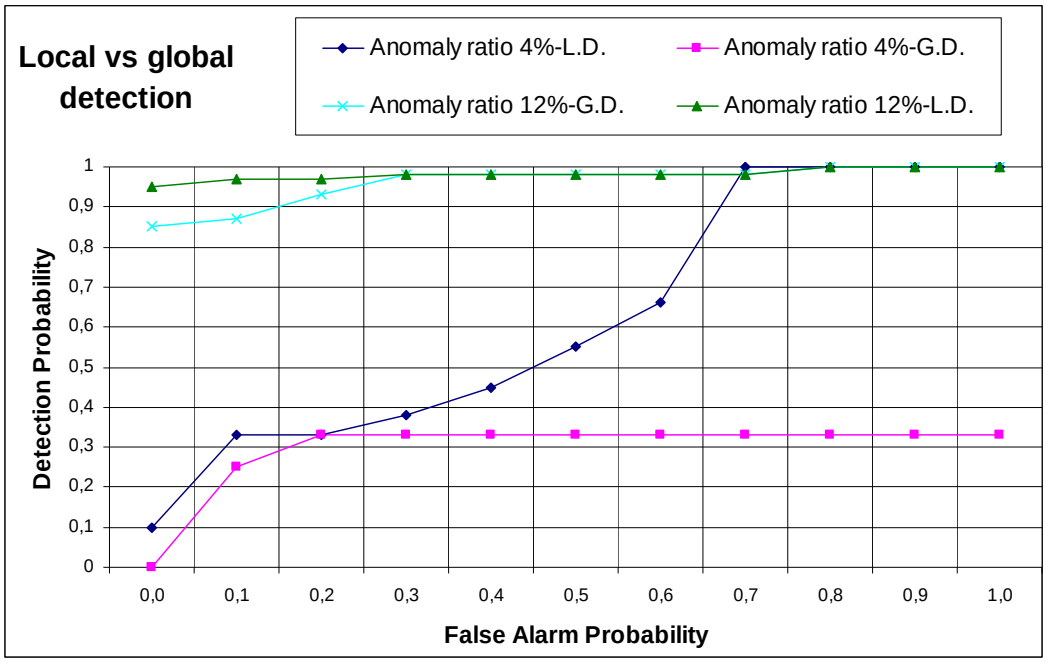
διαδοχικές τιμές στο κατώφλι ανίχνευσης (το όριο στη τιμή του SPE πάνω από το οποίο θεωρείται ότι τα δεδομένα περιέχουν ανωμαλία). Το κατώφλι ανίχνευσης (Detection Threshold – DT) με τιμή κοντά στο μέσο όρο του SPE οδηγεί σε μεγαλύτερες τιμές των πιθανοτήτων P_f και P_d ενώ μεγαλύτερες τιμές έχουν το αντίθετο αποτέλεσμα. Πιο συγκεκριμένα, καθώς όλα τα πειράματα και τα αποτελέσματα που παρουσιάζονται στο τμήμα αυτό έχουν προκύψει από τα ίδια αρχικά δεδομένα με μόνο τις ανωμαλίες να αλλάζουν σε μέγεθος και είδος, οι τιμές στις οποίες κυμάνθηκε το κατώφλι ανίχνευσης κυμάνθηκαν στο παρακάτω εύρος:

$$DT = [2 \times mean(SPE), 3 \times mean(SPE)] \quad (12)$$

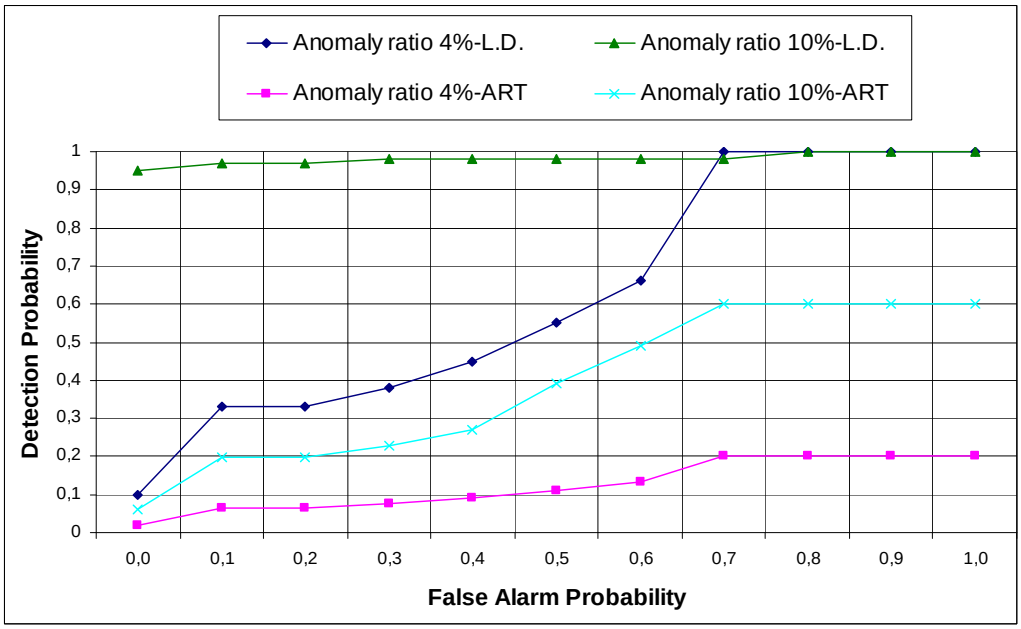
Στο Σχήμα 4.3 παρουσιάζονται οι καμπύλες ROC για δυο διαφορετικές περιπτώσεις: κατανεμημένη/τοπική ανίχνευση (Local Detection – LD) σε σύγκριση με την κεντρική/ολική ανίχνευση (Global Detection – GD) μιας ανωμαλίας που συμβαίνει στις μετρήσεις μιας συγκεκριμένης ομάδας αισθητήρων. Πιο αναλυτικά, οι δυο περιπτώσεις αναφέρονται στο τρόπο με τον οποίο εφαρμόζεται η προτεινόμενη μεθοδολογία στο ΑΔΑ. Η ολική ανίχνευση αναφέρεται στην κεντρική εφαρμογή της μεθόδου άμεσα σε όλους τους κόμβους του δικτύου, χωρίς το διαχωρισμό του δικτύου σε ομάδες. Αντίθετα, η τοπική ανίχνευση αναφέρεται στην κατανεμημένη εφαρμογή του αλγόριθμου ανίχνευσης ξεχωριστά σε κάθε μία ομάδα από κόμβους. Σε κάθε περίπτωση υπολογίζεται η αποτελεσματικότητα του αλγορίθμου για διαφορετικά μεγέθη ανωμαλίας από τα οποία προκύπτουν οι διάφορες καμπύλες ROC. Η ανωμαλία παράγεται τυχαία σε ένα τυχαίο κόμβο κάθε φορά, με το εύρος της να κυμαίνεται στις τιμές 4% και 12%. Ο οριζόντιος άξονας αντιστοιχεί στην πιθανότητα λανθασμένης ανίχνευσης P_f ενώ ο κατακόρυφος άξονας αντιστοιχεί στην πιθανότητα ανίχνευσης P_d. Για κάθε καμπύλη τα σημεία στην άνω-αριστερή πλευρά του διαγράμματος αντιστοιχούν σε ιδανική λειτουργία του αλγορίθμου, με υψηλή πιθανότητα ανίχνευσης και μηδενική πιθανότητα λάθους. Το διάγραμμα αυτό επιβεβαιώνει ότι όσο αυξάνει το μέγεθος της ανωμαλίας αυξάνεται και η πιθανότητα ανίχνευσης.

Επιπρόσθετα, παρατηρούμε ότι η απόδοση του αλγορίθμου ανίχνευσης βελτιώνεται σημαντικά αν εφαρμοστεί κατανεμημένα για κάθε ομάδα. Η συμπεριφορά αυτή βασίζεται στο γεγονός ότι οι μετρήσεις των κόμβων που ανήκουν στην ίδια ομάδα

παρουσιάζουν μεγάλο βαθμό συσχέτισης μεταξύ τους, στοιχείο που επιτρέπει την δημιουργία ενός καλύτερου και ακριβέστερου μοντέλου συσχετίσεων. Η σημασία της υψηλής συσχέτισης ανάμεσα στις μετρήσεις των αισθητήρων θα φανεί και στη συνέχεια με την ανάλυση ξεχωριστού διαγράμματος.



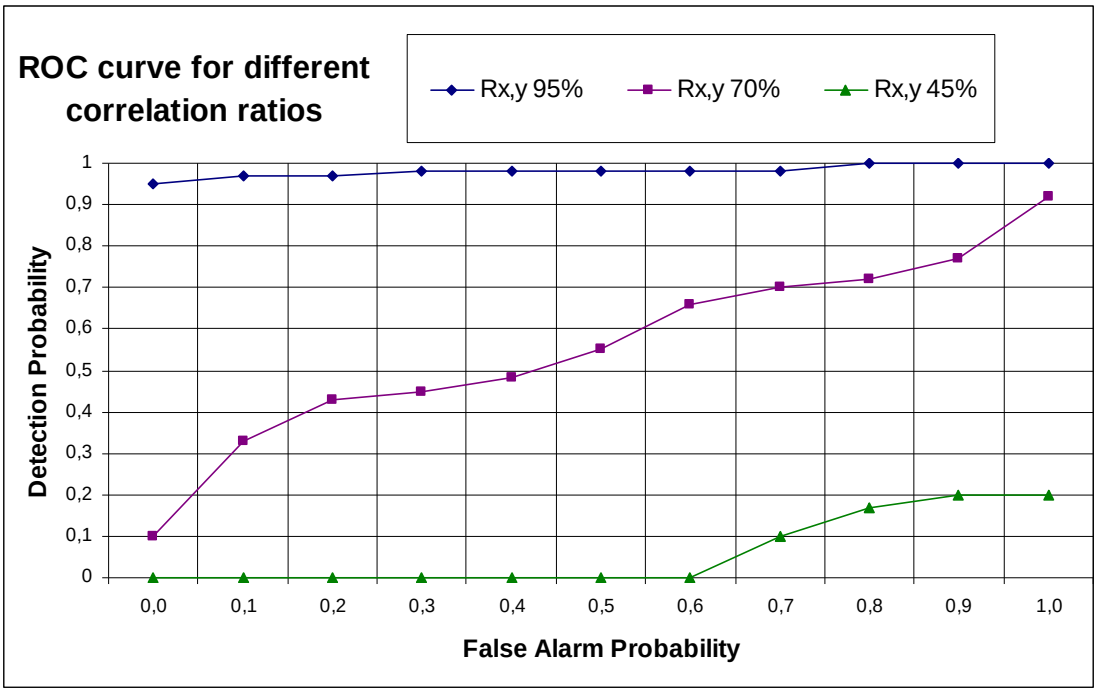
Σχήμα 4.3: Καμπύλες ROC για την κατανεμημένη ανίχνευση σε σύγκριση με την κεντρική ανίχνευση



Σχήμα 4.4: Συγκριτικές καμπύλες ROC για την προτεινόμενη μέθοδο και τη μέθοδο ART

Το Σχήμα 4.4 παρουσιάζει συγκριτικά αριθμητικά αποτελέσματα ανάμεσα στη προτεινόμενη μεθοδολογία και την μέθοδο ART, σε μια ομάδα κόμβων με υψηλή συσχέτιση στις μετρήσεις. Η ανωμαλία παράγεται τυχαία σε ένα κόμβο κάθε φορά, με το εύρος της να κυμαίνεται στις τιμές 4% και 10%. Όπως φαίνεται από το διάγραμμα, η μέθοδος ART αποτυγχάνει να ανιχνεύσει την ανωμαλία εκτός και αν αυτή γίνει αρκετά μεγάλη. Αυτό οφείλεται στο γεγονός ότι με τη μέθοδο ART κάθε αισθητήρας χρησιμοποιεί το βαθμό συσχέτισης μονάχα μεταξύ του ίδιου και κάποιου γειτονικού προκειμένου να αποφασίσει αν είναι ύποπτος ή όχι. Από την άλλη μεριά, η προτεινόμενη μέθοδος επιτυγχάνει καλύτερη απόδοση από την ART, διότι χρησιμοποιώντας την ΑΚΣ, λαμβάνει υπόψη περισσότερες από δυο συσχετίσεις ταυτόχρονα και παράγει λιγότερες λανθασμένες ανιχνεύσεις (false positives).

Στο Σχήμα 4.5 παρουσιάζεται η σημασία της ύπαρξης υψηλής συσχέτισης ανάμεσα στους κόμβους μιας ομάδας. Για το λόγο αυτό εξετάζεται η απόδοση της προτεινόμενης μεθοδολογίας σε ένα σύνολο από αισθητήρες των οποίων οι μετρήσεις παρουσιάζουν διαφορετικούς βαθμούς συσχέτισης $R_{X,Y}$. Πιο συγκεκριμένα, εφαρμόστηκε ανωμαλία με μέγεθος 12% σε σχέση με την αρχική τιμή της μέτρησης και υπολογίστηκαν οι ΚΣ για βαθμούς συσχέτισης με διαφορετική μέγιστη τιμή κάθε φορά: 45%, 70% και 95%. Όπως αναμενόταν, φαίνεται από το σχήμα ότι η αποτελεσματικότητα της προτεινόμενης μεθοδολογίας βελτιώνεται σημαντικά όσο αυξάνεται ο βαθμός συσχέτισης ανάμεσα στις μετρήσεις των αισθητήρων που συμμετέχουν στην ομάδα.



Σχήμα 4.5: Η σημασία της υψηλής συσχέτισης στις μετρήσεις των αισθητήρων μιας ομάδας

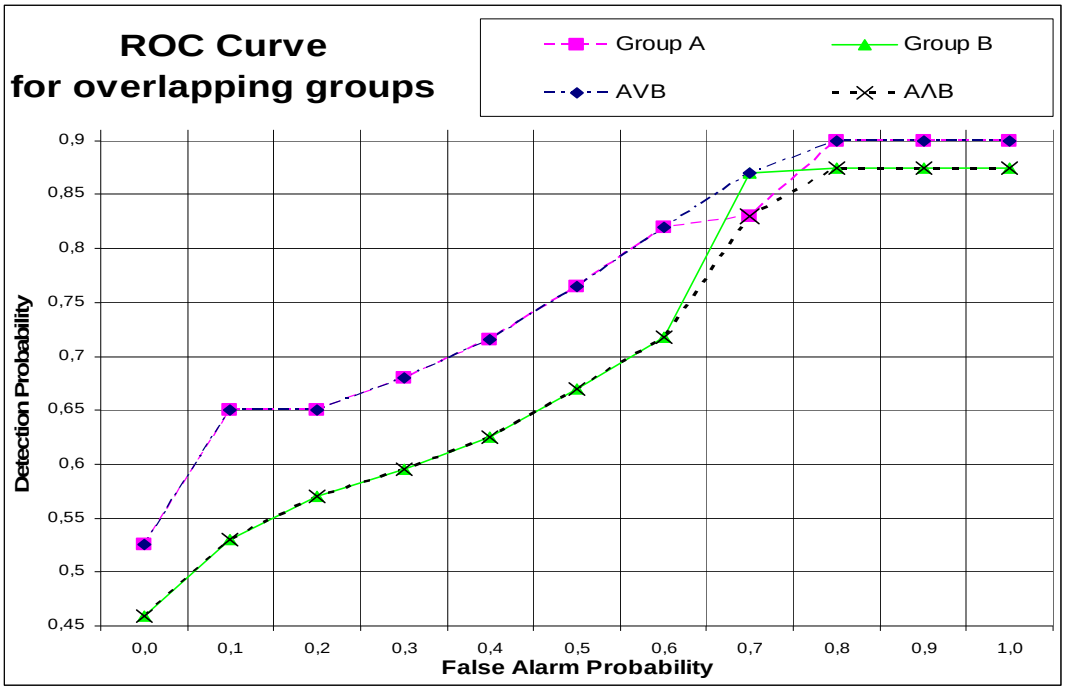
Για να εξασφαλιστεί μεγαλύτερη συσχέτιση ανάμεσα στις μετρήσεις των αισθητήρων σε μια ομάδα μπορεί να χρειαστεί να δημιουργηθούν περισσότερες και μικρότερες ομάδες και κατά συνέπεια χρειάζονται περισσότεροι κύριοι κόμβοι. Ανάλογα με τη κρισιμότητα της εφαρμογής θα πρέπει να αποφασιστεί η ισορροπία ανάμεσα στο αυξημένο κόστος και την καλύτερη απόδοση που προσφέρει η επιλογή αυτή. Σύμφωνα με το πείραμα αυτό, ο βαθμός συσχέτισης θα πρέπει να είναι κοντά στο 90% με 95% προκειμένου να εξαλειφθεί πρακτικά η πιθανότητα λανθασμένης ανίχνευσης.

4.4.1. Χρησιμοποίηση κοινών κόμβων σε γειτονικές ομάδες

Όπως αναφέρθηκε νωρίτερα στο κεφάλαιο αυτό, η ύπαρξη κοινών κόμβων ανάμεσα σε γειτονικές ομάδες παρέχει επιπρόσθετες δυνατότητες στην διαδικασία εξαγωγής της απόφασης για την κατάσταση του δικτύου καθώς συνδυάζει αποφάσεις από διαφορετικές ομάδες. Στο Σχήμα 4.6 παρουσιάζουμε τα αποτελέσματα μιας σειράς

πειραμάτων που επιδεικνύουν και ποσοτικοποιούν τα πλεονεκτήματα της χρήσης κοινών κόμβων ανάμεσα σε γειτονικές ομάδες. Σε κάθε πείραμα, χρησιμοποιήθηκαν δυο ομάδες αισθητήρων με ποσοστό κοινών κόμβων 30%. Οι ανωμαλίες εισάχθηκαν μόνο στις μετρήσεις των κοινών αισθητήρων, είχαν διαφορετικά μεγέθη και εύρος από 2% μέχρι και 15%. Το ποσοστό κόμβων που παρουσίασαν ανωμαλίες κυμάνθηκε ανάμεσα στο 10% και το 30% των συνολικών κόμβων. Ο συνδυασμός των αποτελεσμάτων ανάμεσα στις δυο ομάδες Α και Β έγινε με δυο διαφορετικούς τρόπους: χρησιμοποιώντας τη τομή (A∩B) και την ένωση (AUB) των αποτελεσμάτων προκειμένου να εξαχθεί το τελικό αποτέλεσμα. Η ένωση προκύπτει υπολογίζοντας όλες τις ανιχνεύσεις, σωστές και λανθασμένες, από κάθε ομάδα, ενώ η τομή προκύπτει παίρνοντας μονάχα τις κοινές σωστές και λανθασμένες ανιχνεύσεις.

Γενικά, με τη χρησιμοποίηση της τομής μειώνουμε και το ποσοστό ανίχνευσης σε μια προσπάθεια να μειωθούν οι λανθασμένες ανιχνεύσεις ενώ η χρήση της ένωσης των αποτελεσμάτων έχει το ανάποδο αποτέλεσμα στις πιθανότητες ανίχνευσης και λανθασμένης ανίχνευσης. Όπως φαίνεται στο διάγραμμα στο Σχήμα 4.6, κρατώντας χαμηλό το ποσοστό λάθους (αυτό γίνεται αυξάνοντας το κατώφλι ανίχνευσης), το ξεχωριστό ποσοστό επιτυχίας στην ανίχνευση για την ομάδα Α είναι μεγαλύτερο από ότι αυτό της Β. Συνεπώς, κατά την χρησιμοποίηση της ένωσης AUB, η ομάδα Α κυριαρχεί στα αντίστοιχα συνολικά αποτελέσματα και η καμπύλη ROC της ένωσης είναι πολύ κοντά στην αντίστοιχη του Α. Όσο όμως η πιθανότητα λανθασμένης ανίχνευσης αυξάνεται (μειώνοντας το κατώφλι ανίχνευσης), η ομάδα Β παρουσιάζει μεγαλύτερη πιθανότητα ανίχνευσης και για P_f=0,7 ξεπερνά το αντίστοιχο ποσοστό επιτυχίας στην ανίχνευση της ομάδας Α. Στα πειράματα αυτά οι δυο ομάδες παρουσιάζουν πολλές κοινές λανθασμένες ανιχνεύσεις, έχοντας παράλληλα διαφορετικά ποσοστά επιτυχούς ανίχνευσης. Για το λόγο αυτό, συνολικά η ένωση των αποτελεσμάτων των δυο ομάδων παρουσιάζει καλύτερο συνολικό αποτέλεσμα από ότι η τομή. Αν οι δυο ομάδες ανίχνευαν τις ίδιες ανωμαλίες και παρουσίαζαν διαφορετικές πιθανότητες λανθασμένης ανίχνευσης, η τομή θα ήταν προτιμότερη.



Σχήμα 4.6: Ο αντίκτυπος της χρήσης κοινών κόμβων σε γειτονικές ομάδες

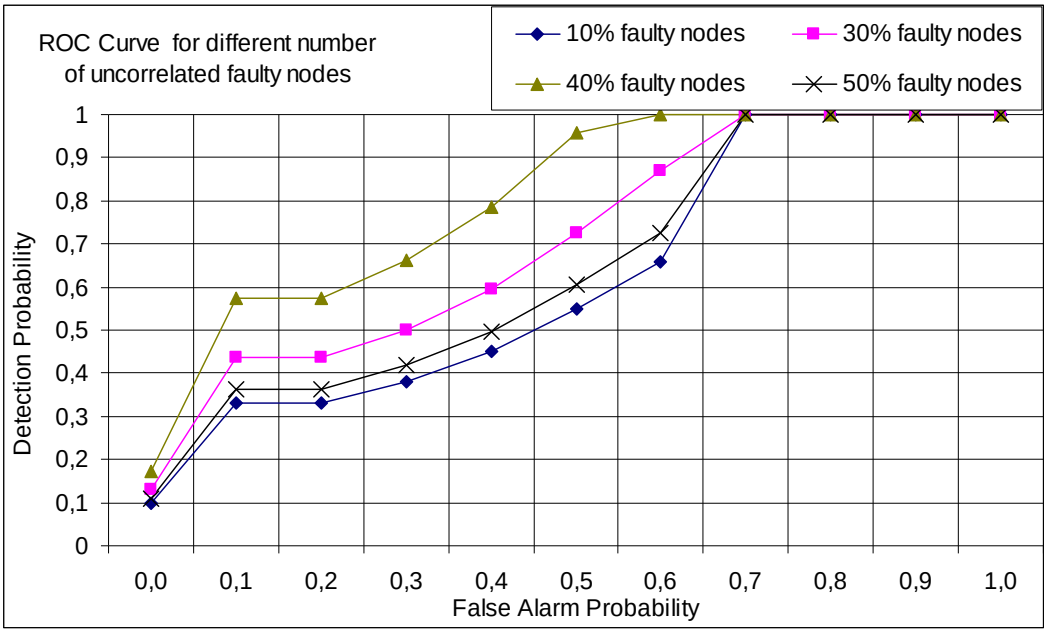
Ανάλογα με τη φύση και την κρισιμότητα της εφαρμογής για την οποία χρησιμοποιείται το ΑΔΑ ο διαχειριστής μπορεί να επιλέξει μια από τις δυο προσεγγίσεις με σκοπό να αυξήσει είτε την πιθανότητα ανίχνευσης είτε να μειώσει την πιθανότητα λανθασμένης ανίχνευσης. Θα πρέπει να σημειωθεί ότι υπάρχουν πιο πολύπλοκες συναρτήσεις οι οποίες θα μπορούσαν να ισορροπήσουν ανάμεσα στους δυο στόχους και θα μπορούσαν να χρησιμοποιηθούν για τη συγχώνευση των αντίστοιχων αποτελεσμάτων.

4.4.2. Σύγκριση τυχαίων και συσχετισμένων ανωμαλιών

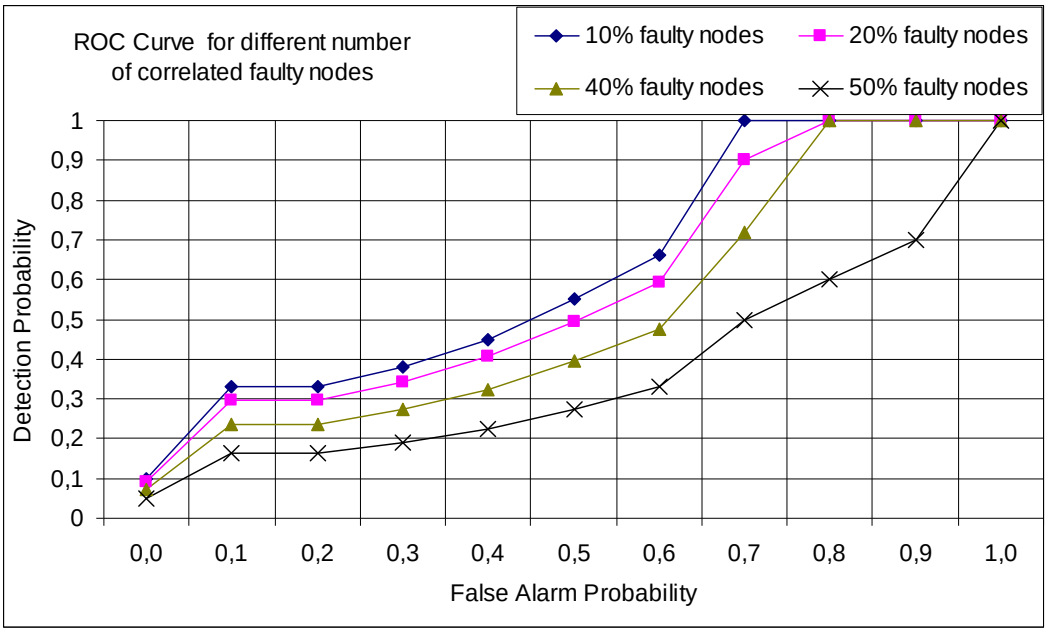
Ανάλογα με την εφαρμογή στην οποία χρησιμοποιείται το ΑΔΑ, η εισαγωγή λαθών στις μετρήσεις μπορεί να συμβεί είτε τυχαία, λόγω αστοχίας υλικού σε κάποιους αισθητήρες, είτε με την θελημένη και ενορχηστρωμένη αλλοίωση μετρήσεων σε γειτονικούς κόμβους με τέτοιο τρόπο ώστε να μην επηρεάζεται σημαντικά η συσχέτιση τους. Στη συνέχεια αναφερόμαστε στο πρώτο είδος ανωμαλίας ως τυχαία ανωμαλία και στο δεύτερο ως συσχετισμένη ανωμαλία. Στη περίπτωση της συσχετισμένης ανωμαλίας, που είναι και η πιο ενδιαφέρουσα, επηρεάζονται οι μετρήσεις γειτονικών

κόμβων κατ' αντιστοιχία με τη συσχέτισή τους, και η αύξηση ή η μείωση της τιμής είναι ίδια σε ποσοστό σε κάθε κόμβο. Όταν ένας αισθητήρας δίνει αλλοιωμένη μέτρηση, η τιμή του σε σχέση με το μέσο όρο και τη διασπορά των μετρήσεων στην ίδια ομάδα θα διαφέρει αισθητά. Όμως, εισάγοντας συσχετισμένες ανωμαλίες, ο μέσος όρος και η διασπορά αλλοιώνονται και η διαδικασία ανίχνευσης δυσχεραίνεται.

Τα αποτελέσματα για τα δυο διαφορετικά είδη ανωμαλιών παρουσιάζονται στο Σχήμα 4.7 και στο Σχήμα 4.8 αντίστοιχα. Και στην περίπτωση αυτή εξετάζουμε και συγκρίνουμε την πιθανότητα ανίχνευσης με την πιθανότητα λανθασμένης ανίχνευσης με χρήση καμπύλων ROC, έχοντας εισαγάγει διαφορετικά μεγέθη ανωμαλίας, με το εύρος τους να κυμαίνεται από 2% ως 15% της κανονικής αρχικής τιμής. Οι διαφορετικές καμπύλες αντιστοιχούν σε διαφορετικά σενάρια όσον αφορά στο ποσοστό των προβληματικών κόμβων, το οποίο κυμαίνεται από 10% ως 50% του συνολικού αριθμού κόμβων της ομάδας. Συνεπώς, στην περίπτωση των τυχαίων ανωμαλιών, κάθε καμπύλη είναι το αποτέλεσμα μιας μίξης ανωμαλιών που συμβαίνουν ταυτόχρονα, επηρεάζουν ένα συγκεκριμένο αριθμό κόμβων και έχουν διαφορετικό ποσοστιαίο μέγεθος ανά κόμβο. Στη περίπτωση των συσχετισμένων ανωμαλιών, η ανωμαλία είναι η ίδια σε ποσοστό σε κάθε κόμβο της ομάδας.



Σχήμα 4.7: Τυχαίες ανωμαλίες



Σχήμα 4.8: Συσχετισμένες ανωμαλίες

Όπως φαίνεται από το Σχήμα 4.7, καθώς ο αριθμός των προβληματικών κόμβων αυξάνει, αυξάνεται και το ποσοστό επιτυχούς ανίχνευσης. Συνεπώς, εκτός και αν το ποσοστό προβληματικών κόμβων υπερβεί το 50%, τυχαίες ανωμαλίες σε παραπάνω από ένα κόμβο αυξάνουν την ακρίβεια του αλγόριθμου ανίχνευσης. Αν όμως το ποσοστό προβληματικών κόμβων αυξηθεί σημαντικά (π.χ. το 50% σε σχέση με το 30%) το σύνολο των μετρήσεων αλλοιώνεται σημαντικά, με αποτέλεσμα η ΑΚΣ να μη μπορεί να διακρίνει ανάμεσα στις κανονικές και τις ανώμαλες τιμές.

Στην περίπτωση των συσχετισμένων ανωμαλιών όμως, όπως φαίνεται στο Σχήμα 4.8, η αύξηση του ποσοστού των προβληματικών κόμβων καταλήγει στη μείωση της πιθανότητας ανίχνευσης. Ακόμα και σε αυτή τη περίπτωση όμως, απαιτείται ένα αρκετά μεγάλο ποσοστό κόμβων να παρουσιάσουν συσχετισμένες ανωμαλίες προκειμένου να επηρεαστεί σημαντικά η δυνατότητα ανίχνευσης της προτεινόμενης μεθοδολογίας. Για να προσδιοριστεί με ακρίβεια μια τόσο μαζική επίθεση, θα μπορούσε να γίνεται σύγκριση με προηγούμενες χρονικά τιμές των μετρικών, δεδομένου ότι οι μετρήσεις παρουσιάζουν χρονική συσχέτιση. Το θέμα αυτό θα μπορούσε να αποτελέσει μέρος μελλοντικής έρευνας.

5. Δίκτυα Αυτοκίνησης

Η βιομηχανία αυτοκινήτων έχει εργαστεί με συνέπεια κατά τη διάρκεια των τελευταίων ετών προς τον εξοπλισμό των αυτοκινήτων με σύνθετα ηλεκτρονικά συστήματα και αισθητήρες που συγκεντρώνουν συνεχώς παραμέτρους που αφορούν την οδική συμπεριφορά και ασφάλεια του οχήματος. Οι αισθητήρες αυτοί συλλέγουν στοιχεία όπως η πίεση και η πρόσφυση των ελαστικών, η ανίχνευση απόστασης από τα γειτονικά οχήματα, η γεωγραφική θέση του οχήματος κ.λπ. Ενώ τα σημερινά συστήματα είναι σε θέση να ελέγχουν το περιβάλλον τους, δεν ανταλλάσσουν ενεργά πληροφορίες μεταξύ γειτονικών οχημάτων καθώς επίσης και μεταξύ οχημάτων και σταθερών σταθμών στην άκρη του δρόμου. Εντούτοις, η δυνατότητα επικοινωνίας γειτονικών οχημάτων και η δυνατότητα επικοινωνίας ανάμεσα σε οχήματα και σταθερούς σταθμούς (όπως για παράδειγμα σταθμούς της τροχαίας τοποθετημένους σε κεντρικά σημεία του αυτοκινητιστικού δικτύου), θα έδινε στους οδηγούς την ευκαιρία να αποφύγουν επικίνδυνες και δυσάρεστες καταστάσεις.

Τα δίκτυα αυτοκίνησης – ΔΑ (Vehicular Networks) έχουν καταστεί ακρογωνιαίος λίθος για την δημιουργία ευφυών συστημάτων μεταφορών (Intelligent Transportation Systems (ITS). Με την επικοινωνία μεταξύ οχημάτων (Vehicle to Vehicle - V2V) καθώς επίσης μεταξύ αυτοκινήτων και σταθμών βάσης στην άκρη του δρόμου (Roadside-to-Vehicle - R2V), τα ΔΑ μπορούν να συμβάλουν στη δημιουργία ασφαλέστερων και αποδοτικότερων δρόμων με την παροχή έγκαιρων πληροφοριών στους οδηγούς και τις αρχές. Στην κατεύθυνση αυτή, περιεκτικές παρατηρήσεις για τις κυκλοφοριακές ροές και τις ταχύτητες των οχημάτων μπορούν να χρησιμοποιηθούν για να υπολογίσουν διάφορες παραμέτρους του οδικού δικτύου, όπως πιθανές καθυστερήσεις και αναμενόμενους χρόνους άφιξης σε διάφορα σημεία κλειδιά [Pfan05].

Ένας αποφασιστικός παράγοντας για την επιτυχία και την αποτελεσματικότητα ενός ευφυούς συστήματος μεταφορών είναι τα συστήματα συλλογής πληροφοριών από τους αισθητήρες των οχημάτων και των σταθερών σταθμών να μπορούν να εξετάσουν διαφορετικά επίπεδα αβεβαιότητας στην πληροφορία που συλλέγουν, να ερμηνεύουν κατάλληλα τα στοιχεία και να συνάγουν ακριβή και χρήσιμα για τον οδηγό συμπεράσματα. Για να παραγάγει μια αξιόπιστη εικόνα της κατάστασης του οδικού δικτύου, το κάθε σύστημα πρέπει να συλλέξει και επεξεργαστεί δεδομένα από ποικίλα

διαφορετικά σημεία, από πολλά οχήματα και διαφορετικούς αισθητήρες και να τα ενσωματώσει σε μια περιεκτική μορφή. Κατά συνέπεια απαιτείται η αποδοτική συγχώνευση δεδομένων από πολλαπλές πηγές.

Για την επίτευξη των προαναφερθέντων στόχων, εφαρμόζεται η ΑΚΣ ταυτόχρονα σε πολλαπλά μετρικά που συλλέγονται από διάφορους αισθητήρες αξιοποιώντας τη δυνατότητα της ΑΚΣ στο εντοπισμό και προσδιορισμό χωρικών συσχετίσεων. Η συγχώνευση των δεδομένων από διάφορους απομακρυσμένους ετερογενείς αισθητήρες παρέχει μια γενικευμένη μεθοδολογία ανίχνευσης ατυχημάτων και γενικότερα ανωμαλιών στο οδικό δίκτυο. Η αξιολόγηση της απόδοσης της προτεινόμενης μεθοδολογίας καταδεικνύει τη λειτουργική αποτελεσματικότητά της στα ευφυή συστήματα μεταφορών. Το υπόλοιπο του κεφαλαίου αυτού οργανώνεται ως εξής. Στην ενότητα 5.1 παρουσιάζουμε σχετικές προσεγγίσεις που υπάρχουν στην βιβλιογραφία, ενώ στις ενότητες 5.2 και 5.3 παρουσιάζεται η μεταφορά και οι λεπτομέρειες εφαρμογής του αλγορίθμου ανίχνευσης ανωμαλιών στα ΔΑ. Η απόδοση και η λειτουργική αποτελεσματικότητα της προτεινόμενης μεθοδολογίας αξιολογούνται στις ενότητες 5.4 και 5.5 μέσω προσομοίωσης.

5.1. Ανίχνευση Ανωμαλιών στα ΔΑ- Βιβλιογραφία

Στη βιβλιογραφία το μεγαλύτερος μέρος των ερευνητικών προσπαθειών που σχετίζονται με την ανίχνευση ανεπιθύμητων καταστάσεων στα οδικά δίκτυα έχει εστιάσει στην ανάπτυξη αλγορίθμων ανίχνευσης ανωμαλίας. Υπάρχουν πολλοί διαφορετικοί τύποι αλγορίθμων που έχουν αναπτυχθεί κατά τη διάρκεια των τελευταίων τριάντα ετών με σκοπό την ανίχνευση ανεπιθύμητων συμβάντων στα οδικά δίκτυα, βασισμένοι σε διαφορετικές θεωρίες όπως η αναγνώριση προτύπων, η στατιστική ανάλυση, και η τεχνητή νοημοσύνη [Mart00].

Για παράδειγμα, στη δημοσίευση [Anto96] οι ερευνητές χρησιμοποίησαν το στατιστικό εργαλείο T-test για να αναλύσουν διαφορές στη πυκνότητα μιας οδικής αρτηρίας με τη βοήθεια ενός μοναδικού σταθμού ανιχνευτών. Κρατώντας ένα κινητό παράθυρο από 10 χρονικά διαστήματα, ο αλγόριθμος υπολογίζει την τυπική απόκλιση της πυκνότητας σε οχήματα. Το T-test χρησιμοποιείται προκειμένου να υπολογιστεί το διάστημα εμπιστοσύνης γύρω από το μέσο όρο. Συνεπώς, η τρέχουσα τιμή συγκρίνεται

με την τυπική απόκλιση και το διάστημα εμπιστοσύνης. Αν η απόκλιση είναι μεγαλύτερη από κάποιο προκαθορισμένο κατώφλι, το σύστημα αναφέρει ότι ανίχνευσε ανωμαλία. Εντούτοις η εφαρμογή αυτής της μεθόδου περιορίζεται σε ένα μοναδικό μέρος (κάποιο κομμάτι δρόμου) του οδικού δικτύου και δεν μπορεί να εφαρμοστεί σε ένα μεγάλο δίκτυο. Αυτό οφείλεται στο γεγονός ότι δεν συσχετίζει πληροφορίες από διαφορετικά μέρη, ώστε να παρασχεθεί μια πλήρης και ολοκληρωμένη εικόνα για την κατάσταση του ελεγχόμενου οδικού δικτύου.

Στη δημοσίευση [Pfan05] οι συγγραφείς παρουσίασαν μια προσέγγιση που χρησιμοποιεί μεθόδους αναγνώρισης προτύπων και συσχετισμού για τον προσδιορισμό ακολουθιών από οχήματα με τρόπο παρόμοιο με αυτόν που χρησιμοποιούν οι μηχανικοί επικοινωνιών για να προσδιορίσουν ακολουθίες από bit για την αποκωδικοποίηση μηνυμάτων. Τα πειραματικά αποτελέσματα που παρουσίασαν συνδυάζουν δεδομένα από έναν μικροκυματικό και έναν υπερηχητικό αισθητήρα.

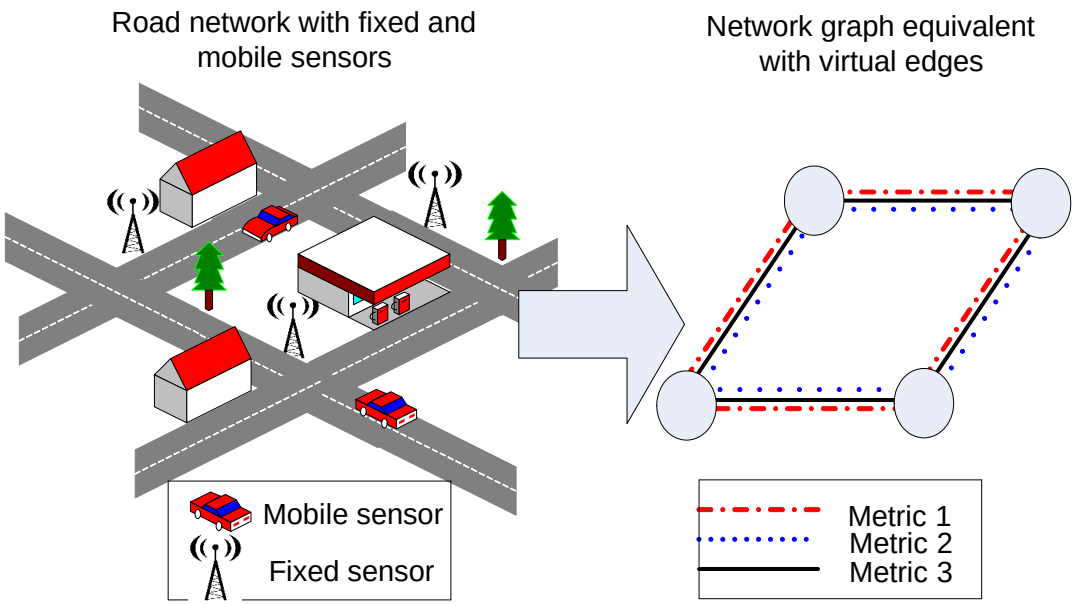
Πρόσφατα, η ανάπτυξη των ασύρματων επικοινωνιών έχει οδηγήσει τους ερευνητές στην υιοθέτηση της χρησιμοποίησης αυτό-οργανούμενων δικτύων για τη διαχείριση κυκλοφορίας και την ανίχνευση ατυχημάτων. Στη δημοσίευση [Chen06], οι συγγραφείς εξέτασαν το πρόβλημα εξομάλυνσης των ροών της κυκλοφορίας με την ανταλλαγή πληροφορίας που αφορά τη θέση και τη ταχύτητα των οχημάτων. Η εξομάλυνση επιτυγχάνεται με την αποστολή μηνυμάτων από το σημείο στο οποίο υπάρχει το πρόβλημα και το δυναμικό υπολογισμό και επιβολή μεταβλητών ορίων ταχύτητας βασισμένων στην τοπική πυκνότητα των οχημάτων. Στη δημοσίευση [Yang04] προτείνεται ένα πρωτόκολλο επικοινωνίας από όχημα σε όχημα για την παροχή συνεργατικής προειδοποίησης σύγκρουσης. Η τεχνική πρόκληση που προσπαθούν να αντιμετωπίσουν είναι η επίτευξη όσο το δυνατόν μικρότερης καθυστέρησης στην παράδοση των προειδοποιήσεων έκτακτης ανάγκης κάτω από διάφορες συνθήκες στο οδικό δίκτυο. Με βάση μια προσεκτική ανάλυση των απαιτήσεων της εφαρμογής, οι ερευνητές σχεδίασαν ένα πρωτόκολλο περιλαμβάνοντας πολιτικές ελέγχου συμφόρησης των πακέτων στο ασύρματο δίκτυο, ώστε να είναι πάντα δυνατή η έγκυρη διάδοση μηνυμάτων έκτακτης ανάγκης.

Επιπλέον, στη δημοσίευση [Chen97] οι ερευνητές παρουσιάζουν μια προσέγγιση που υπολογίζει τη πιθανότητα τα δεδομένα να μην περιέχουν ανωμαλίες με τη βοήθεια της ΑΚΣ. Το τρέχον διάνυσμα που περιγράφει τη κατάσταση του οδικού δικτύου

φέρεται να περιέχει ή όχι ανωμαλίες ανάλογα με την απόστασή του από το «κέντρο βάρους» των κανονικών (χωρίς ανωμαλίες) δεδομένων. Για τη μέτρηση της αξιοπιστίας της μεθοδολογίας χρησιμοποιήθηκαν πραγματικά και προσομοιωμένα δεδομένα από διάφορες πηγές. Αν και αυτή η προσέγγιση παρουσίαζε ελπιδοφόρα αποτελέσματα προς την κατεύθυνση της ανίχνευσης ανωμαλίας με τη χρησιμοποίηση της ΑΚΣ, το μεγάλο της μειονέκτημα ήταν ότι δεν παρείχε κάποιο μέσο για την απομόνωση της πηγής της ανωμαλίας και τον εντοπισμό των ακμών του οδικού δικτύου που επηρεάστηκαν. Αντίθετα, η προσέγγισή μας, που είναι επίσης βασισμένη στις αρχές και τα κύρια χαρακτηριστικά γνωρίσματα της ΑΚΣ, παρέχει μια ενσωματωμένη μεθοδολογία που όχι μόνο συνδυάζει αποτελεσματικά τα συσχετισμένα δεδομένα προκειμένου να αποκαλυφθούν ανωμαλίες που εκτείνονται σε διάφορες γειτονικές ακμές του γράφου του οδικού δικτύου, αλλά επιτρέπει τον προσδιορισμό των αντίστοιχων μονοπατιών που περιέχουν την ανωμαλία.

5.2. Εφαρμογή της μεθόδου ανίχνευσης ανωμαλιών σε δίκτυα αυτοκίνησης

Για τους σκοπούς των ευφυών συστημάτων μεταφορών και της ανίχνευσης ατυχημάτων, ο δρόμος και το ασύρματο δίκτυο οχημάτων υπό εξέταση μπορούν να χαρτογραφηθούν σε ένα υβριδικό παράδειγμα δικτύου με διάφορους ετερογενείς αισθητήρες, που αποτελείται από κινητούς και σταθερούς κόμβους. Πιο συγκεκριμένα, μερικοί κόμβοι αντιπροσωπεύουν οχήματα που φέρουν αισθητήρες και το απαραίτητο λογισμικό και υλισμικό για τη διασύνδεση στο ασύρματο δίκτυο. Άλλοι κόμβοι του δικτύου μπορούν να συναθροίσουν και να αναμεταδώσουν τα δεδομένα, ενώ κάποιοι άλλοι σταθεροί κόμβοι φέρουν αισθητήρες που ελέγχουν ένα συγκεκριμένο μέρος του δρόμου ή της κυκλοφοριακής ροής (δηλ. μια συγκεκριμένη ακμή του αντίστοιχου γράφου του οδικού δικτύου). Γενικά για κάθε ακμή μπορεί να συλλέγονται διάφορα μετρικά που περιγράφουν την κυκλοφορία που περνά από αυτή. Κατά αντιστοιχία με τα δίκτυα επικοινωνιών, δημιουργείται ένα σύνολο από εικονικές ακμές, με την κάθε μια να αντιστοιχεί σε διαφορετικό μετρικό της αρχικής πραγματικής ακμής. Η τοπολογία και η αρχιτεκτονική δικτύων απεικονίζονται στο Σχήμα 5.1.



Σχήμα 5.1: Τοπολογία Δικτύου και Αρχιτεκτονική

5.3. Περιβάλλον και ανάλυση της εφαρμογής

Στα ad hoc δίκτυα που αποτελούνται από οχήματα, η γεωγραφική θέση και οι σχετικές με την διαδρομή πληροφορίες θεωρούνται ως κρίσιμα στοιχεία. Ο λόγος είναι ότι τέτοιες πληροφορίες που είναι διαθέσιμες μόνο αν τα οχήματα έχουν δυνατότητα επικοινωνίας με σταθμούς βάσης, απαλείφουν μερικούς από τους περιορισμούς των υπαρχόντων πρωτοκόλλων που βασίζονται στη γνώση της τοπολογίας. Παράλληλα, τα ad hoc δίκτυα επιτρέπουν την εφαρμογή ευφυών και έγκαιρων αντιδράσεων σε ένα ανεπιθύμητο συμβάν στο οδικό δίκτυο.

Για την εφαρμογή της προτεινόμενης μεθοδολογίας εξετάζονται δυο διαφορετικοί τρόποι λειτουργίας. Η βασική λειτουργία βασίζεται μόνο στα στοιχεία που συλλέγονται από σταθερούς αισθητήρες και αναφέρεται ως ευφυής ανίχνευση συμβάντων (Intelligent Incident Detection - IID). Ο δεύτερος τρόπος λειτουργίας ενισχύεται σε σχέση με τον πρώτο με πρόσθετες μετρήσεις ή/και πληροφορίες που προέρχονται από τα οχήματα και αναφέρεται ως ευφυής ανίχνευση συμβάντων υποβοηθούμενη από τα οχήματα (Vehicle-assisted Intelligent Incident Detection - VIID). Αυτός ο τρόπος λειτουργίας προϋποθέτει τη δυνατότητα επικοινωνίας μεταξύ των οχημάτων και του ενσύρματου δικτύου και επιτρέπει τη συλλογή πιο εξειδικευμένων μετρικών ανά όχημα, όπως ο χρόνος ταξιδιού του οχήματος στο οδικό δίκτυο, η κατάσταση του οχήματος,

πληροφορίες για τη διαδρομή και τον τελικό προορισμό του, κ.λπ. Στην ενότητα 5.5.4 γίνεται μια προσπάθεια εκτίμησης και ποσοτικοποίησης της βελτίωσης στην αντίδραση μετά την ανίχνευση ενός συμβάντος που προσφέρει η χρήση του VIID. Στο εξής οι όροι ανίχνευση ανωμαλιών και ανίχνευση συμβάντων θα χρησιμοποιούνται εναλλακτικά έχοντας την ίδια έννοια.

Προκειμένου να αποκεντρωθεί η λειτουργία του αλγορίθμου ανίχνευσης συμβάντων σε ένα οδικό δίκτυο μεγάλης κλίμακας, το δίκτυο μπορεί να διαιρεθεί σε ομάδες, με κάθε ομάδα να αποτελείται από ένα υποσύνολο των ακμών του γράφου του συνολικού δικτύου. Σε κάθε ομάδα μπορούμε να υποθέσουμε ότι υπάρχει ένας κύριος κόμβος ο οποίος είναι σταθερός και εξοπλισμένος με περισσότερους υπολογιστικούς και δικτυακούς πόρους από τους υπόλοιπους. Ο διαχωρισμός σε ομάδες μπορεί να είναι είτε στατικός είτε δυναμικός (δηλ. να ρυθμίζεται εκ νέου περιοδικά) αν υπάρχουν δραστικές αλλαγές στο μέγεθος της κίνησης (π.χ. ώρα αιχμής) και τη δρομολόγηση. Η διαδικασία ομαδοποίησης των ακμών μπορεί σε γενικές γραμμές να βασιστεί είτε στα «φυσικά» χαρακτηριστικά (π.χ. εγγύτητα των ακμών) είτε σε άλλες «λογικές» ιδιότητες (π.χ. συσχέτιση των μετρικών). Στα οδικά δίκτυα τα δεδομένα που συλλέγονται σε γειτονικές ακμές του γράφου είναι συνήθως ιδιαίτερα συσχετισμένα. Για το λόγο αυτό, μετά τη διαδικασία ομαδοποίησης δημιουργούνται ομάδες που αποτελούνται από γειτονικές ακμές τω οποίων οι μετρήσεις είναι ιδιαίτερα συσχετισμένες.

Επομένως, με τη διαίρεση ενός δικτύου μεγάλης κλίμακας σε ομάδες, επιτυγχάνει κανείς περισσότερο έγκυρη επεξεργασία των μετρικών χωρίς να υποστεί επιβάρυνση στην πολυπλοκότητα και τη διαχείριση που προκαλούνται από τα μεγάλης κλίμακας οδικά και τηλεπικοινωνιακά δίκτυα. Κάθε κύριος κόμβος λαμβάνει μετρήσεις από τους κόμβους που ανήκουν στην ομάδα του και εκτελεί τοπικά τον αλγόριθμο ανίχνευσης σε πραγματικό χρόνο. Κάθε κινούμενος κόμβος συλλέγει και αποστέλλει δεδομένα σε σχέση με ένα ή περισσότερα μετρικά που περιγράφουν συγκεκριμένες παραμέτρους του ίδιου και της ακμής του οδικού δικτύου που διασχίζει την προκειμένη στιγμή. Πρέπει επίσης να σημειωθεί ότι οι διάφορες ομάδες μπορεί να έχουν κοινά μέλη, κατά αντιστοιχία με τα ΑΔΑ που περιγράφονται στο κεφάλαιο 4.3.3, καθώς η ύπαρξη διάφορων κοινών ακμών θα μπορούσε να επιδιωχτεί σε πολλές περιπτώσεις προκειμένου να βελτιωθεί η αποτελεσματικότητα ανίχνευσης. Στην ενότητα 5.5.2 αξιολογείται ο αντίκτυπος της διαδικασίας ομαδοποίησης στην ικανότητα ανίχνευσης,

εξετάζοντας την εφαρμογή της προτεινόμενης προσέγγισης ανίχνευσης συμβάντων σε δύο διαφορετικούς τρόπους λειτουργίας (τοπικός σε σύγκριση με τον κεντρικό). Πιο συγκεκριμένα, η κεντρική ανίχνευση (Global Detection - GD) αναφέρεται στη συγκεντρωτική εφαρμογή του προτεινόμενου αλγορίθμου ανίχνευσης συμβάντων άμεσα στο συνολικό αριθμό των ακμών (δηλ. χωρίς ομαδοποίηση), ενώ η τοπική ανίχνευση (Local Detection - LD) αναφέρεται στην αποκεντρωμένη εφαρμογή της προσέγγισής μας, χωριστά για κάθε συγκεκριμένη ομάδα ακμών.

Μια άλλη πιθανή επίδραση της διαίρεσης ενός μεγάλου δικτύου σε μικρότερες ομάδες ακμών είναι η απομόνωση των τοπικών λαθών και κατά αντιστοιχία η μείωση του αντίκτυπού τους στην ακρίβεια ανίχνευσης της προτεινόμενης μεθοδολογίας. Στην πράξη, η συλλογή δεδομένων μπορεί να περιέχει ανακριβείς μετρήσεις λόγω διάφορων ετερογενών παραγόντων όπως η ύπαρξη ελαττωματικών αισθητήρων, ανεπαρκής ενδοεπικοινωνία λόγω της φύσης των αυτό-οργανούμενων δικτύων ή ακόμα εξαιτίας της άρνησης ή αδυναμίας μερικών κόμβων να συνδεθούν στον κεντρικό κόμβο και να εκθέσουν τις μετρήσεις τους. Εάν αυτά τα λάθη παράγονται αραιά και τυχαία στο δίκτυο δεν θέτουν κάποια σοβαρή απειλή στην αποτελεσματικότητα ανίχνευσης συμβάντων. Εντούτοις, εάν αυτά τα λάθη εμφανίζονται έντονα σε μια συγκεκριμένη περιοχή είναι δυνατό να αλλοιώσουν σημαντικά μερικές από τις ελεγχόμενες παραμέτρους του οδικού δικτύου, προκαλώντας τελικά λανθασμένες ανιχνεύσεις σε όλο το δίκτυο. Με τη διαίρεση του δικτύου σε ομάδες, μπορούμε να απομονώσουμε σε τοπικό επίπεδο προβλήματα περιορισμένης κλίμακας.

Σε κάθε περίπτωση, οι συναθροισμένες μετρήσεις και οι πληροφορίες για τις τοπικές ανωμαλίες διαβιβάζονται στον κεντρικό κόμβο. Ο κεντρικός κόμβος που έχει γνώση για την τοπολογία του δικτύου μπορεί να καθορίσει εάν η ανωμαλία διαδίδεται σε πολλαπλές ομάδες και εάν ακολουθεί ένα η περισσότερα δικτυακά μονοπάτια. Αυτή η διαδικασία θα μπορούσε να παρέχει βελτιωμένη ανίχνευση συμβάντων στην κυκλοφορία. Οι ακμές που παρουσίασαν τις μεγαλύτερες αποκλίσεις αποκαλύπτουν το μονοπάτι που ακολουθεί το συμβάν καθώς επίσης και την τρέχουσα θέση της κύριας πηγής της ανωμαλίας.

5.4. Αξιολόγηση και εξέταση απόδοσης

Σε αυτή την ενότητα αξιολογούμε την απόδοση και τη λειτουργική αποτελεσματικότητα της προτεινόμενης προσέγγισης ανίχνευσης ανωμαλιών μας μέσω προσομοίωσης, χρησιμοποιώντας τον προσομοιωτή της αστικής κυκλοφορίας (Simulator of Urban Mobility - SUMO) [SUMO08] [Chow00]. Το SUMO είναι ένα ιδιαίτερα φορητό λογισμικό ανοιχτού κώδικα, ένα πακέτο μικροσκοπικής προσομοίωσης της οδικής κυκλοφορίας, σχεδιασμένο να χειρίζεται μεγάλα οδικά δίκτυα.

Στη συνέχεια θα παρουσιαστούν πέντε διαφορετικά σενάρια που έχουν προσομοιωθεί με τη βοήθεια του λογισμικού SUMO. Η έμφαση στα δυο πρώτα σενάρια δίνεται κυρίως στην ικανότητα ανίχνευσης ανεπιθύμητων συμβάντων στην κυκλοφορία και για το λόγο αυτό περιλαμβάνουν διαφορετικά πρότυπα ανωμαλίας που επαναλαμβάνονται για διάφορα ποσοστά ανάμεσα στις κανονικές και ανώμαλες ροές. Το τρίτο σενάριο επιδεικνύει πως η προτεινόμενη μεθοδολογία μπορεί να εφαρμοστεί με δυο διαφορετικούς τρόπους: με τη συγχώνευση δεδομένων από ολόκληρο το οδικό δίκτυο, είτε με την εφαρμογή της ΑΚΣ χωριστά σε τμήματα του οδικού δικτύου. Στο τέταρτο σενάριο εξετάζουμε τον τρόπο με τον οποίο δεδομένα που περιέχουν ανακριβείς μετρήσεις επηρεάζουν αρνητικά το ποσοστό επιτυχούς ανίχνευσης. Ο στόχος του τελευταίου σεναρίου είναι η μελέτη της αποδοτικότητας πιθανής αντίδρασης σε κάποιο συμβάν κάτω από δυο διαφορετικές περιπτώσεις: με ή χωρίς την εισαγωγή πληροφοριών στο σύστημα που προέρχονται από τα οχήματα (για την ακρίβεια το προορισμό κάθε οχήματος). Τα αποτελέσματα αυτά αποκαλύπτουν στην πραγματικότητα τις λειτουργικές διαφορές ανάμεσα στους δυο διαφορετικούς τρόπους λειτουργίας των ευφυών συστημάτων μεταφορών όπως αυτά παρουσιάστηκαν στην προηγούμενη ενότητα (σύγκριση IID με VIID). Τέλος παρουσιάζεται ένα παράδειγμα προσδιορισμού του μονοπατιού που ακολουθεί η ανωμαλία στο γράφο του οδικού δικτύου καθώς και ένα παράδειγμα που επιδεικνύει και ποσοτικοποιεί τη σημασία χρησιμοποίησης μετρικών με υψηλό βαθμό συσχέτισης.

Το υπόλοιπο μέρος της ενότητας αυτής είναι οργανωμένο ως εξής: στην ενότητα 5.4.1 παρουσιάζουμε τα μετρικά ποσοτικοποίησης της απόδοσης της μεθοδολογίας, ενώ στην ενότητα 5.4.2 περιγράφεται η τοπολογία του οδικού δικτύου και τα μοντέλα της

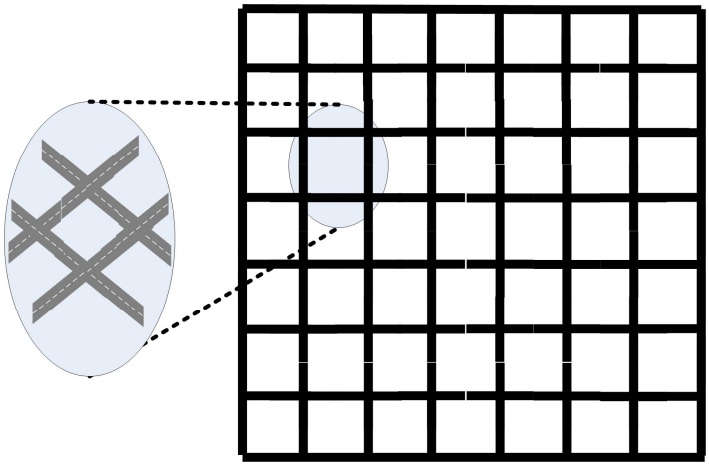
κίνησης που χρησιμοποιούνται στις προσομοιώσεις. Η ενότητα 5.5 περιέχει τα αριθμητικά αποτελέσματα και σχολιασμό των πειραμάτων και αποτελεσμάτων.

5.4.1. Μετρικά αξιολόγησης της απόδοσης

Προκειμένου να υπολογίσουμε την αποτελεσματικότητα του αλγορίθμου ανίχνευσης ανωμαλιών υιοθετήθηκαν τα ίδια μετρικά που χρησιμοποιήθηκαν και στα προηγούμενα κεφάλαια: η πιθανότητα ανίχνευσης P_d και η πιθανότητα λανθασμένης ανίχνευσης (false alarm) P_f. Παράλληλα, χρησιμοποιούμε τις καμπύλες Receiver Operating Characteristic (ROC), προκειμένου να εκφράσουμε και να αναπαραστήσουμε τη σχέση ανάμεσα σε αυτές τις πιθανότητες. Επιπλέον προκειμένου να αξιολογηθεί η αποτελεσματικότητα αντίδρασης του συστήματος, χρησιμοποιείται ο μέσος χρόνος ταξιδιού. Αυτό το μετρικό αντιστοιχεί στον υπολογισμό του μέσου χρόνου ταξιδιού όλων των αυτοκινήτων στο οδικό δίκτυο. Με βάση τις δυνατότητες ανίχνευσης της προτεινόμενης προσέγγισης και των γενικών διαθέσιμων πληροφοριών που παρέχονται από όλους τους κινητούς και τους σταθερούς κόμβους, εκτελείται αναδρομολόγηση της κίνησης με διαφορετικές μεθόδους προκειμένου να ανακουφιστεί το συγκεκριμένο τμήμα του δικτύου στο οποίο έχει ανιχνευθεί κάποιο συμβάν. Εντούτοις, η αναδρομολόγηση πρέπει να γίνει κατά τέτοιο τρόπο ώστε τα τμήματα με κυκλοφοριακό πρόβλημα να ανακουφίζονται και παράλληλα να μη δημιουργείται πρόβλημα στις γειτονικές περιοχές από τη νέα δρομολόγηση των οχημάτων.

5.4.2. Μοντέλο οδικού Δικτύου

Στα πειράματα που παρουσιάζονται στη συνέχεια έχει χρησιμοποιηθεί μια τοπολογία πλέγματος που αποτελείται από 64 διασταυρώσεις. Κάθε ακμή του γράφου μοντελοποιεί ένα δρόμο διπλής κατεύθυνσης με μια λωρίδα για κάθε κατεύθυνση και μήκους 400 μέτρων. Η μέγιστη ταχύτητα σε όλες τις λωρίδες είναι 60 km/hour. Δεν υπάρχουν φανάρια ελέγχου της κυκλοφορίας με αποτέλεσμα να ισχύει ο βασικός κανόνας προτεραιότητας (προτεραιότητα έχουν τα οχήματα που έρχονται από δεξιά σε μια διασταύρωση). Η τοπολογία του οδικού δικτύου φαίνεται στο παρακάτω σχήμα:



Σχήμα 5.2: Τοπολογία Οδικού Δικτύου

Σε κάθε πείραμα εισάγονται διάφορες κυκλοφοριακές ροές που αποτελούνται από τρία διαφορετικά είδη οχημάτων. Ο ακόλουθος Πίνακας 5.5.1 εμφανίζει τους διάφορους τύπους οχημάτων που ορίστηκαν και χρησιμοποιήθηκαν κατά την προσομοίωση. Οι τύποι “Car1” και “Car2” θεωρούνται μέρος της κανονικής κίνησης, ενώ ροή με οχήματα του τύπου “Car3” εισάγεται σε ειδικές περιπτώσεις για την προσομοίωση μιας ανωμαλίας που αποτελείται από μια ροή αργά κινούμενων αυτοκινήτων. Ο ρυθμός παραγωγής οχημάτων ανά ροή ορίζεται σε τέτοιες τιμές ώστε να μην προκαλούνται κυκλοφοριακά προβλήματα κάτω από κανονικές συνθήκες.

Type	Acceleration (m/s^2)	Deceleration (m/s^2)	Length (m)
Car1	4.6	4.5	4
Car2	2.6	4.5	4
Car3	1.6	4.5	8

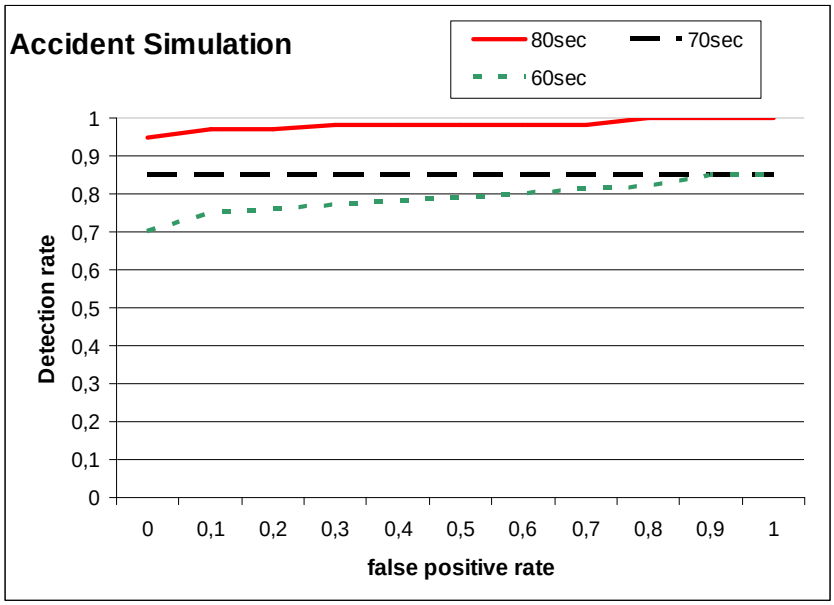
Πίνακας 5.5.1: Τύποι Οχημάτων

5.5. Αριθμητικά αποτελέσματα

5.5.1. Ανίχνευση συμβάντων

Σενάριο 1 – Προσομοίωση Ατυχήματος

Στο πρώτο σενάριο η ανωμαλία δημιουργείται από ένα σταματημένο αυτοκίνητο σε κάποια ακμή του δικτύου το οποίο προκαλεί κυκλοφοριακό πρόβλημα, καθώς τα οχήματα που ακολουθούν δε μπορούν να το προσπεράσουν και σταματούν και αυτά. Το αυτοκίνητο παραμένει ακίνητο για ένα σύντομο χρονικό διάστημα, προσομοιώνοντας ένα ατύχημα. Το Σχήμα 5.3 παρουσιάζει τη πιθανότητα ανίχνευσης σε σχέση με τη πιθανότητα λανθασμένης ανίχνευσης σε μια καμπύλη ROC. Το μετρικό που χρησιμοποιείται σαν είσοδο στον αλγόριθμο ανίχνευσης είναι η πυκνότητα των οχημάτων σε κάθε ακμή του γράφου του δικτύου.



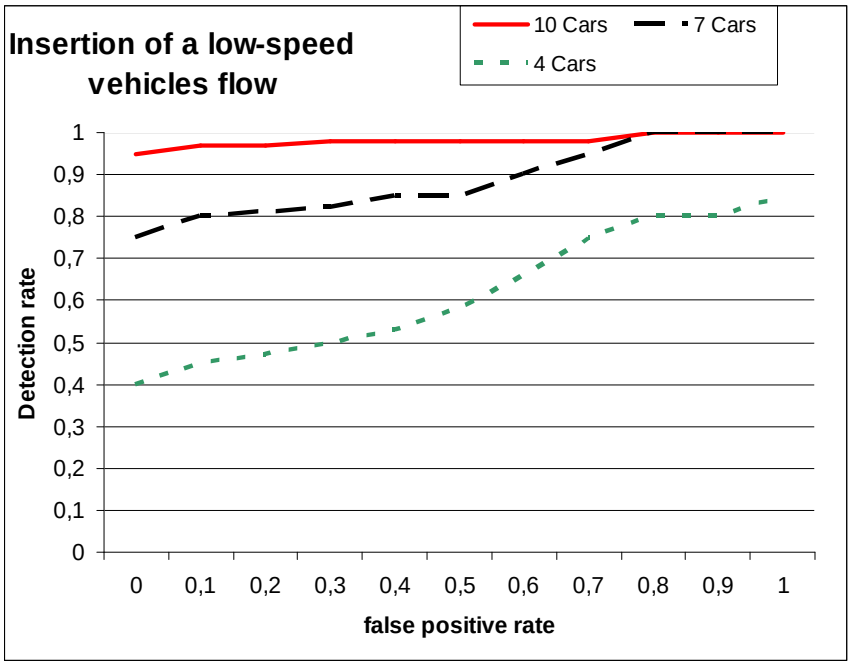
Σχήμα 5.3: Προσομοίωση Ατυχήματος

Οι διαφορετικές καμπύλες αντιστοιχούν σε διαφορετικό χρόνο για τον οποίο έμεινε ακινητοποιημένο το όχημα. Πιο συγκεκριμένα, η ακινησία του οχήματος διαρκεί για 60,70 και 80 δευτερόλεπτα και το αντίκτυπο στην κίνηση είναι η αύξηση της πυκνότητας της αντίστοιχης ακμής κατά 5%, 8% και 12% αντίστοιχα. Για κάθε

καμπύλη το σημείο στο άνω αριστερό σημείο αντιστοιχεί σε ιδανική λειτουργία του αλγορίθμου, με μεγάλη πιθανότητα ανίχνευσης και χαμηλότατη πιθανότητα λανθασμένης ανίχνευσης. Από το σχήμα φαίνεται ότι καθώς η διάρκεια ακινησίας αυξάνεται η πιθανότητα ανίχνευσης αυξάνεται. Φαίνεται επίσης ότι η προτεινόμενη προσέγγιση επιτυγχάνει υψηλά ποσοστά επιτυχούς ανίχνευσης ακόμα και για μικρές τιμές της ανωμαλίας, γεγονός που σημαίνει ότι μπορεί να χρησιμοποιηθεί σε πραγματικό χρόνο για την εφαρμογή αντίμετρων και τη βελτίωση της κυκλοφορίας.

Σενάριο 2 – Εισαγωγή ροής αργών οχημάτων

Στο δεύτερο σενάριο, η ανωμαλία που εισάγεται αντιστοιχεί στην εισαγωγή μιας ροής από αργά οχήματα. Η κανονική κίνηση αποτελείται από τα οχήματα τύπου “Car1” και “Car2” ενώ τα αργά οχήματα είναι του τύπου “Car3”. Τα αντίστοιχα αριθμητικά αποτελέσματα εμφανίζονται στις καμπύλες στο Σχήμα 5.4 Οι διαφορετικές καμπύλες αντιστοιχούν σε διαφορετικά μεγέθη της ανωμαλίας που εισάγεται στο δίκτυο. Πιο συγκεκριμένα, οι διαφορετικές ροές που εισάγονται αποτελούνται από 4,7 και 10 οχήματα και προκαλούν ποσοστό ανωμαλίας στη πυκνότητα κάθε ακμής 5%, 8% και 12% αντίστοιχα. Και στην περίπτωση αυτή επιβεβαιώνεται ότι η απόδοση του αλγορίθμου βελτιώνεται καθώς το μέγεθος της ανωμαλίας αυξάνεται.

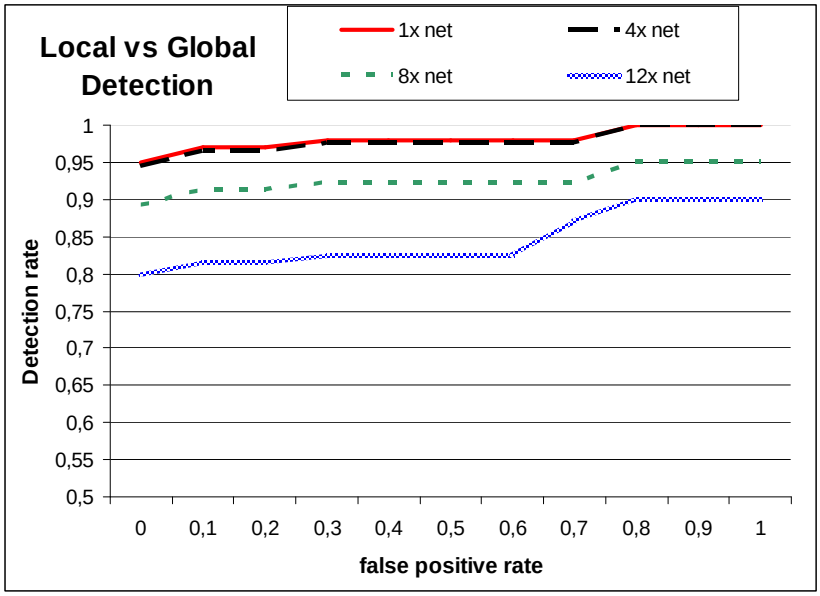


Σχήμα 5.4: Εισαγωγή Οχημάτων Χαμηλής Ταχύτητας

5.5.2. Αποκεντρωμένη και Κεντρική ανίχνευση

Σενάριο 3 – Μέγεθος του δικτύου και ομαδοποίηση ακμών

Σε αυτό το πείραμα εξετάζεται η συμπεριφορά του αλγόριθμου ανίχνευσης καθώς το μέγεθος του δικτύου αυξάνεται. Ο στόχος στη περίπτωση αυτή είναι διπλός: αρχικά μας ενδιαφέρει να αποκτήσουμε κάποια γνώση για το αντίκτυπο του μεγέθους του δικτύου στην αποτελεσματικότητα του αλγορίθμου και στη συνέχεια να μελετήσουμε πως ο διαχωρισμός των ακμών σε ομάδες μπορεί να χρησιμοποιηθεί για αντιμετωπιστούν προβλήματα κλίμακας. Για το λόγο αυτό, συγκρίνουμε την αποκεντρωμένη λειτουργία του αλγόριθμου (LD) με τη λειτουργία σε συνολικό επίπεδο (GD). Χρησιμοποιείται η ανωμαλία που παρουσιάστηκε στο σενάριο 1, με ποσοστό σε σχέση με την κανονική κίνηση 12%. Το οδικό δίκτυο αποτελείται από 64 διασταυρώσεις (Σχήμα 5.2) και στη συνέχεια θα αναφερόμαστε σε αυτό το βασικό δίκτυο ως «δίκτυο 1x». Στη συνέχεια η ίδια ανωμαλία εισάγεται σε μεγαλύτερα δίκτυα. Πιο συγκεκριμένα, δημιουργούνται δίκτυα με τέσσερις, οχτώ και δώδεκα φορές μεγαλύτερο μέγεθος, που αποτελούνται από 256, 512 και 768 διασταυρώσεις και ονομάζονται «δίκτυο 4x» «δίκτυο 8x» και «δίκτυο 12x» αντίστοιχα. Ας σημειωθεί ότι το αρχικό δίκτυο των 64 διασταυρώσεων είναι μέρος των 3 μεγαλύτερων δικτύων και η ανωμαλία εισάγεται πάντα στην ίδια ακμή του δικτύου αυτού.



Σχήμα 5.5: Σύγκριση Κεντρικής με Αποκεντρωμένη Ανίχνευση

Συγκρίνοντας το «δίκτυο 8x» με το «δίκτυο 1x» παρατηρούμε την αντίστοιχη μείωση στο ποσοστό επιτυχούς ανίχνευσης κατά τη χρησιμοποίηση κεντρικής ανίχνευσης ενός μεγάλου δικτύου (8x64) 512 διασταυρώσεων, αντί για τη χρησιμοποίηση 8 ομάδων με 64 διασταυρώσεις η καθεμία. Με βάση τα αποτελέσματα αυτά, παρατηρούμε ότι η απόδοση του αλγόριθμου βελτιώνεται αισθητά με τη χρήση ομάδων σε τοπικό επίπεδο. Η συμπεριφορά αυτή οφείλεται κυρίως στο γεγονός ότι οι ακμές του γράφου σε μια ομάδα παρουσιάζουν συνήθως μεγαλύτερη συσχέτιση από ότι συνολικά με όλες τις ακμές του δικτύου, με αποτέλεσμα το μοντέλο του δικτύου να είναι πιο ακριβές.

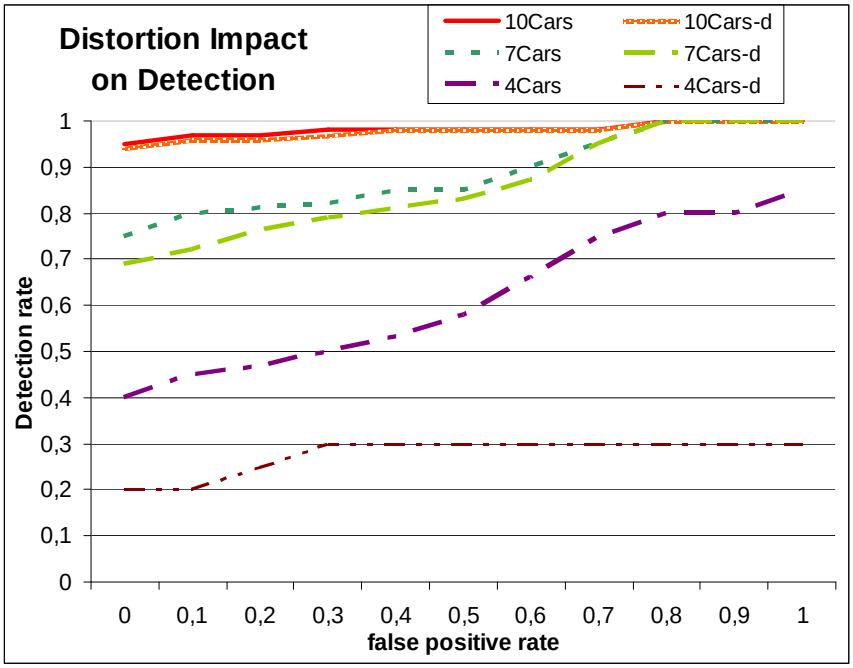
5.5.3. Το αντίκτυπο της διαστρέβλωσης των δεδομένων στην απόδοση του αλγορίθμου

Σενάριο 4 – Διαστρέβλωση μετρήσεων

Όπως αναφέρθηκε προηγουμένως, ο προτεινόμενος αλγόριθμος ανίχνευσης συμβάντων βασίζεται στη συλλογή και συγχώνευση δεδομένων από διάφορους αισθητήρες και συνεπώς η απόδοση του βασίζεται στην εγκυρότητα των δεδομένων. Στην πράξη όμως, τα συλλεχθέντα δεδομένα μπορεί να περιέχουν ανακρίβειες για διάφορους λόγους (π.χ. χαλασμένοι αισθητήρες ή πρόβλημα στη μετάδοση των δεδομένων λόγω της φύσης του δικτύου. Προκειμένου να μελετηθεί το συγκεκριμένο πρόβλημα στο σενάριο 4 εκτελούνται κάποια πειράματα όπου εισάγονται λάθη στις μετρήσεις κατά τη συλλογή των δεδομένων.

Χρησιμοποιώντας ως βάση το σενάριο 2, τα αρχικά δεδομένα αλλοιώθηκαν με τη προσθήκη τυχαίων λαθών σε κάποια μετρικά. Τα λάθη κυμαίνονταν από -6% ως 6% της κανονικής τιμής. Στο Σχήμα 5.6 συγκρίνουμε τα αποτελέσματα που δίνουν οι αλλοιωμένες μετρήσεις σε σχέση με τις αρχικές στο σενάριο 2. Οι καμπύλες “4Cars”, “7Cars” και “10Cars” αντιστοιχούν σε τρεις – διαφορετικές σε μέγεθος – ροές που αποτελούνται από 4,7 και 10 αργά οχήματα και προκαλούν 5%, 8% και 12% αύξηση στην χρησιμοποίηση των ακμών αντίστοιχα. Παράλληλα, οι “4Cars-d”, “7Cars-d” και “10Cars-d” περιέχουν αλλοιωμένες μετρήσεις για τις προαναφερθείσες καμπύλες. Παρατηρείται ότι οι καμπύλες “10Cars” και “10Cars-d” είναι σχεδόν πανομοιότυπες, συνεπώς όταν η ανωμαλία είναι μεγαλύτερη από το τυχαίο λάθος τα λάθη στις

μετρήσεις δεν έχουν πρακτικά αντίκτυπο στην απόδοση του αλγορίθμου. Συγκρίνοντας τις καμπύλες “7Cars” και “7Cars-d” παρατηρούμε ότι όταν το μέγεθος της ανωμαλίας είναι σχεδόν το ίδιο με το μέγεθος του λάθους που προστίθεται στις μετρήσεις, η ανωμαλία είναι ακόμα ανιχνεύσιμη αν και ο αλγόριθμος αρχίζει να εμφανίζει αύξηση στον αριθμό των λανθασμένων ανιχνεύσεων. Τέλος, στη περίπτωση που η ανωμαλία γίνεται μικρότερη από το επίπεδο του τυχαίου λάθους η απόδοση του αλγορίθμου χειροτερεύει αισθητά, όπως φαίνεται από τη σύγκριση των “4Cars” και “4Cars-d”.



Σχήμα 5.6: Επίδραση Αλλοιωμένων Μετρήσεων στην Ανίχνευση Ανωμαλιών

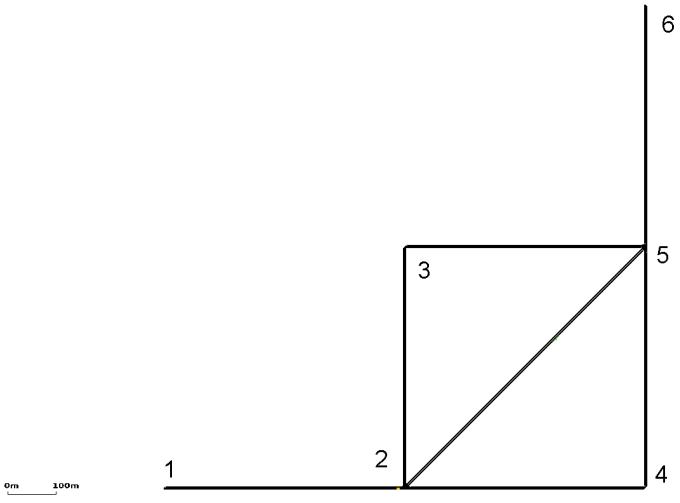
5.5.4. Αποτελεσματικότητα αντίδρασης

Σενάριο 5 – Στοχευμένη Αντίδραση

Στο πέμπτο σενάριο, εξετάζεται η αποτελεσματικότητα αντίδρασης σε κάποιο συμβάν κάτω από δυο διαφορετικές περιπτώσεις λειτουργίας: την ευφυή ανίχνευση συμβάντων (Intelligent Incident Detection - IID) σε σύγκριση με την ευφυή ανίχνευση συμβάντων υποβοηθούμενη από τα οχήματα (Vehicle-assisted Intelligent Incident Detection - VIID). Πιο συγκεκριμένα, το σύστημα συλλέγει τη χρησιμοποίηση κάθε ακμής του γράφου του οδικού δικτύου καθώς και την προγραμματισμένη πορεία των

οχημάτων που συνδέονται στο δίκτυο. Στη περίπτωση ανίχνευσης κάποιου συμβάντος, το σύστημα μπορεί να προτείνει συγκεκριμένη εναλλακτική οδό σε κάθε οδηγό που σκοπεύει να περάσει από το μέρος του δικτύου που παρουσίασε το πρόβλημα. Στο πείραμα που παρουσιάζεται, η διαφορά ανάμεσα στις λειτουργίες IID και VIID είναι ότι η τελευταία μπορεί να προτείνει εναλλακτική διαδρομή σε κάθε οδηγό, προσφέροντας έτσι μια περισσότερο εκλεπτυσμένη διαχείριση της κίνησης.

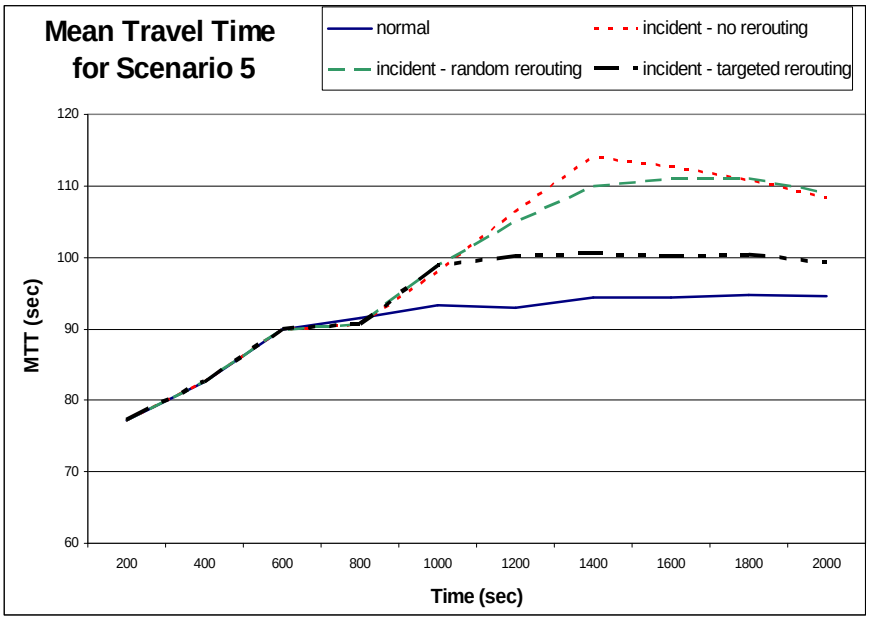
Η δικτυακή τοπολογία που χρησιμοποιείται στο συγκεκριμένο πείραμα φαίνεται στο Σχήμα 5.7. Το πείραμα διήρκησε 7000 δευτερόλεπτα. Στο χρονικό σημείο 700sec, ένα όχημα ακινητοποιείται για 300 sec στην ακμή που συνδέει τη διασταύρωση 2 με την διασταύρωση 5 προσομοιώνοντας ένα ατύχημα. Παράλληλα, οι ακμές που συνδέουν την διασταύρωση 2 με την 4 έχουν μεγάλη κίνηση και συνεπώς η χρήση τους σε πιθανή αναδρομολόγηση της κίνησης δεν συνίσταται. Εφαρμόζοντας τη λειτουργία IID, το σύστημα αναφέρει σε όλους τους οδηγούς που περνούν από τη διασταύρωση 2 ότι θα πρέπει να αποφύγουν την ακμή με το ατύχημα και συνεπώς κάθε οδηγός διαλέγει τυχαία κάποια από τις γειτονικές ακμές. Με τη χρησιμοποίηση όμως της λειτουργίας VIID, το σύστημα αναφέρει το πρόβλημα σε κάθε οδηγό και τους προτείνει να κινηθούν στην ακμή που συνδέει τις διασταυρώσεις 2 και 3.



Σχήμα 5.7: Η Τοπολογία του Δικτύου στο Σενάριο 5

Για λόγους σύγκρισης, στο Σχήμα 5.8 παρουσιάζουμε το μέσο χρόνο ταξιδιού (Mean Travel Time – MTT) όλων των οχημάτων που διέρχονται από το δίκτυο για τέσσερις διαφορετικές περιπτώσεις. Η πρώτη περίπτωση παρουσιάζει το MTT κάτω από κανονικές συνθήκες χωρίς το συμβάν. Η δεύτερη παρουσιάζει το MTT στη

περίπτωση του ατυχήματος χωρίς να υπάρχει η δυνατότητα ενημέρωσης των οχημάτων, η τρίτη παρουσιάζει το MTT με τη λειτουργία IID, ενώ η τελευταία με τη λειτουργία VIID. Όπως ήταν αναμενόμενο, η λειτουργία VIID έχει ως αποτέλεσμα το μικρότερο MTT (εξαιρουμένης της κανονικής κατάστασης φυσικά), προσφέροντας καλύτερη και εγκυρότερη αντίδραση στο συμβάν.



Σχήμα 5.8: Μέσος Χρόνος Ταξιδιού για το Σενάριο 5

6. Συμπεράσματα και Μελλοντικές Τάσεις

Η συνεχής αύξηση της χρηστικότητας και σημαντικότητας των δικτύων μεγάλης κλίμακας, έχει καταστήσει την έγκαιρη και έγκυρη ανίχνευση ανωμαλιών, όχι μια απλώς επιθυμητή προηγμένη υπηρεσία, αλλά μια άκρως βασική και απαραίτητη λειτουργία. Δεδομένου του γεγονότος ότι το Διαδίκτυο του μέλλοντος (Internet of the Future) θα έχει πιθανώς ως αναπόσπαστα τμήματά του ετερογενή δίκτυα μεγάλης κλίμακας, που μπορεί να παρουσιάζουν διαφορετικά μοντέλα λειτουργίας και να υποστηρίζουν διαφορετικές υπηρεσίες, είναι ιδιαίτερα κρίσιμο να αναπτυχθεί μια γενική, ευέλικτη και αποτελεσματική μεθοδολογία ανίχνευσης ανωμαλιών, που να προσφέρει ένα γενικευμένο πλαίσιο για την εφαρμογή της σε διαφορετικού τύπου δίκτυα. Προς την κατεύθυνση αυτή, σκοπός της διατριβής αυτής ήταν η μελέτη του προβλήματος της ανίχνευσης ανωμαλιών σε διαφορετικής φύσης δίκτυα μεγάλης κλίμακας, βασιζόμενη σε έναν αλγόριθμο συγχώνευσης δεδομένων από ετερογενείς αισθητήρες διασκορπισμένους στο δίκτυο. Για να πετύχει τους στόχους αυτούς, η προτεινόμενη μεθοδολογία στηρίζεται στην Ανάλυση Κυρίων Συνιστωσών.

Συγκεκριμένα, στα πλαίσια της συγκεκριμένης διατριβής προτάθηκε και αναλύθηκε μια μεθοδολογία ανίχνευσης ανωμαλιών που συγχωνεύει δεδομένα που συλλέγονται από διάφορα μέρη του δικτύου έτσι ώστε να επιτυγχάνεται αποδοτική ανίχνευση. Η μέθοδος ονομάζεται μέθοδος πολλαπλών μετρικών πολλαπλών ζεύξεων, Multi-Metric-Multi-Link (M^3L) με κύρια χαρακτηριστικά τα εξής:

Α. Η εφαρμογή της προτεινόμενης προσέγγισης επιτρέπει τη ταυτόχρονη αξιοποίηση πολλαπλών μετρικών ανά ενεργό μέρος του δικτύου (όπως για παράδειγμα ένας δρομολογητής, μια ζεύξη στα δίκτυα ευρείας ζώνης, ένας ασύρματος αισθητήρας στα Ασύρματα Δίκτυα Αισθητήρων, μια οδική αρτηρία στην εφαρμογή των Δικτύων Αυτοκίνησης). Η επεξεργασία πολλαπλών μετρικών μπορεί να αποκαλύψει ανωμαλίες που δεν επηρεάζουν τον όγκο των δεδομένων και θα ήταν αδύνατο να ανιχνευτούν διαφορετικά, αλλά επηρεάζουν σημαντικά τις συσχετίσεις ανάμεσα σε διάφορα μετρικά.

Β. Συσχετίζοντας μετρικά που έχουν συλλεχθεί από διαφορετικούς δρομολογητές προβάλει μια συνολική εικόνα του δικτύου επιτρέποντας την ανίχνευση ανωμαλιών που δεν είναι τοπικές και διατρέχουν ένα μεγαλύτερο μέρος του δικτύου.

Επίσης σημαντικό στοιχείο της μεθόδου είναι ότι είναι ικανή να αναγνωρίσει το ή τα μονοπάτια τα οποία διατρέχει η ανωμαλία, αναγνωρίζοντας και συσχετίζοντας τις συνδέσεις στις οποίες η κίνηση παρουσιάζει μη προβλεπόμενη συμπεριφορά. Ένα ακόμα κύριο καινοτόμο στοιχείο, είναι η διαδικασία εύρεσης του βέλτιστου αριθμού Κυρίων Συνιστωσών. Η διαδικασία αυτή εκμεταλλεύεται τη γνώση πιθανών ανωμαλιών, ώστε να υπολογίσει με μεγαλύτερη ακρίβεια τον βέλτιστο αριθμό Κυρίων Συνιστωσών ή εναλλακτικά να ορίσει ένα μοντέλο του δικτύου το οποίο να είναι προσαρμοσμένο στις εκάστοτε ανάγκες ή προτεραιότητες του διαχειριστή.

Αρχικά, παρουσιάστηκε η εφαρμογή της προτεινόμενης μεθοδολογίας στα δίκτυα ευρείας ζώνης. Πειραματικά αποτελέσματα βασισμένα σε πειράματα εξομοίωσης έδειξαν ότι η προτεινόμενη μεθοδολογία επεξεργασίας πολλαπλών μετρικών όπως π.χ. η ρυθμαπόδοση σε πακέτα, TCP ή UDP bytes, μπορεί να αποκαλύψει ανωμαλίες που δεν επηρεάζουν τον όγκο των δεδομένων και θα ήταν αδύνατο να ανιχνευτούν διαφορετικά.

Πραγματικές μετρήσεις σε διαφαινόμενες συνθήκες κανονικής λειτουργίας στο δίκτυο του ΕΔΕΤ έδειξαν ότι ο συντελεστής συσχέτισης μετρικών είναι ιδιαίτερα υψηλός, στοιχείο που είναι ιδιαίτερα ελπιδοφόρο για την αποτελεσματική λειτουργία του αλγορίθμου και σε πραγματικές συνθήκες. Παράλληλα, μελετήθηκε η επιρροή της συλλογής δεδομένων που υπόκεινται σε δειγματοληψία πριν την αποστολή τους στον κεντρικό κόμβο ανίχνευσης ανωμαλιών. Η πρακτική αυτή εφαρμόζεται συχνά στα δίκτυα ευρείας ζώνης με τη χρήση του πρωτοκόλλου NetFlow στους κεντρικούς δρομολογητές, όπου ο αριθμός των πακέτων που διέρχονται είναι τεράστιος. Στα πλαίσια επέκτασης της λειτουργίας της προτεινόμενης μεθοδολογίας, καθώς επίσης και της αξιολόγησης της επίδοσής της, ένα ενδιαφέρον στοιχείο το οποίο μπορεί να μελετηθεί μελλοντικά και που έχει ιδιαίτερη πρακτική αξία για την αποδοτική εφαρμογή της μεθοδολογίας σε ένα λειτουργικό πραγματικό δίκτυο, είναι η ανάλυση της δυνατότητας λειτουργίας με μειωμένο αριθμό κόμβων. Πιο συγκεκριμένα, μπορεί να αναλυθεί η αντοχή του αλγορίθμου ανίχνευσης σε τυχόν αδυναμία συλλογής δεδομένων από κάποια σημεία του δικτύου λόγω δικτυακών ή άλλων προβλημάτων.

Μεγάλο ενδιαφέρον παρουσιάζουν τα Ασύρματα Δίκτυα Αισθητήρων και η μεταφορά της προτεινόμενης μεθόδου σε αυτά. Κλειδί στη μεταφορά της προτεινόμενης μεθοδολογίας στα ΑΔΑ είναι ότι η συγχώνευση δεδομένων πρέπει να

γίνεται κατανεμημένα και ιεραρχικά. Προκειμένου να εφαρμοστεί με επιτυχία στα ΑΔΑ, το δίκτυο θα πρέπει να χωριστεί σε ομάδες με γειτονικούς κόμβους, ώστε να εκμεταλλευτούμε τις υψηλές συσχετίσεις στις μετρήσεις γειτονικών κόμβων. Με τον συνδυασμό και την ανάλυση αποτελεσμάτων από γειτονικές περιοχές, είναι δυνατή η ανίχνευση συσχετισμένων επιθέσεων/ανωμαλιών που επηρεάζουν πολλαπλές ομάδες κόμβων. Με τον τρόπο αυτό μπορούμε να εφαρμόσουμε τον αλγόριθμο ανίχνευσης για την αποκάλυψη ελαττωματικών κόμβων, για την ανίχνευση δικτυακών επιθέσεων και το φιλτράρισμα λανθασμένων μετρήσεων σε οποιοδήποτε βήμα της διαδικασίας συνάθροισης δεδομένων στα ΑΔΑ. Η αποτελεσματικότητα της προτεινόμενης μεθόδου ανίχνευσης ανωμαλιών εξετάστηκε με τη χρήση μετεωρολογικών δεδομένων που έχουν συλλεχθεί από ένα πραγματικό δίκτυο αισθητήρων. Επίσης, εξετάστηκε πως παράμετροι όπως ο συντελεστής συσχέτισης γειτονικών κόμβων, το ποσοστό κοινών κόμβων σε γειτονικές ομάδες, αλλά και διαφορετικά είδη επιθέσεων (τυχαίες ή συντονισμένες με στόχο κόμβους σε πολλαπλές γειτονικές ομάδες) επηρεάζουν την αποτελεσματικότητα ανίχνευσης. Αντικείμενο μελλοντικής έρευνας στην κατεύθυνση αυτή αποτελεί η ανάπτυξη δυναμικών μεθοδολογιών ομαδοποίησης των κόμβων, ώστε να προσαρμόζονται στις πιθανώς μεταβαλλόμενες συνθήκες του δικτύου και στις απαιτήσεις των μετρήσεων.

Τέλος, η εφαρμογή της προτεινόμενης μεθόδου μελετήθηκε και σε δίκτυα με σταθερούς και κινούμενους κόμβους, όπως είναι για παράδειγμα τα Δίκτυα Αυτοκίνησης, με σκοπό την υποστήριξη διαφόρων λειτουργιών και υπηρεσιών (π.χ. ανάπτυξη έξυπνων συστημάτων κυκλοφορίας και μεταφοράς). Στα υβριδικά αυτά δίκτυα η εφαρμογή της προτεινόμενης μεθόδου παρουσιάζει αντικειμενικές δυσκολίες, καθώς όταν οι κόμβοι αλλάζουν συνεχώς θέση, αλλάζουν συνεχώς και οι μεταξύ τους συσχετίσεις με αποτέλεσμα η μοντελοποίηση του δικτύου να πρέπει να γίνει με διαφορετικό τρόπο από ότι στα σταθερά δίκτυα αισθητήρων. Η αξιολόγηση της απόδοσης και τα αντίστοιχα αριθμητικά αποτελέσματα κατέδειξαν ότι η μεθοδολογία μας επιτυγχάνει υψηλή ακρίβεια στην ανίχνευση συμβάντων για διαφορετικά σενάρια συμβάντων και διάφορα μεγέθη ανωμαλίας. Παράλληλα, δίνοντας ένα απλό παράδειγμα ευφυούς ανίχνευσης συμβάντων ή οποία υποστηρίζεται από τα οχήματα, τονίστηκε η αναγκαιότητα ανάπτυξης αποδοτικών συστημάτων μεταφορών που μπορούν να εξετάσουν τα διαφορετικά επίπεδα αβεβαιότητας, να συλλέξουν δεδομένα

από τα οχήματα, να συνάγουν τα κατάλληλα συμπεράσματα, και τελικά να παράσχουν στον οδηγό χρήσιμες πληροφορίες με σκοπό την ασφαλή και γρήγορη μετακίνηση.

Ενδιαφέρον στοιχείο από τα ΔΑ το οποίο χρήζει περαιτέρω έρευνας και μπορεί να μελετηθεί γενικότερα στα αυτο-οργανούμενα δίκτυα είναι η κινητικότητα των κόμβων. Στο τέταρτο κεφάλαιο, στα ΑΔΑ θεωρήσαμε ότι οι αισθητήρες παραμένουν σταθεροί και ο διαχωρισμός σε νέες ομάδες γίνεται μόνο όταν αλλάζει συμπεριφορά το περιβάλλον. Στα ΔΑ αντίθετα, μελετήθηκε ένα υβριδικό μοντέλο με σταθερούς και κινούμενους κόμβους και η επιτυχής εφαρμογή της ΑΚΣ στηρίχθηκε στην εξαγωγή ΚΣ από τους σταθερούς κόμβους. Σε ένα περιβάλλον όμως όπου δεν υπάρχουν σταθεροί κόμβοι, η εφαρμογή της ΑΚΣ απαιτεί τη συνεχή δημιουργία νέων ομάδων με δυναμικό τρόπο. Η εύρεση μιας αποδοτικής μεθόδου για τη δημιουργία ομάδων αποτελεί ιδιαίτερα ενδιαφέρον πεδίο και αντικείμενο μελλοντικής έρευνας.

Τέλος, ενδιαφέρον στοιχείο είναι ότι κάποιες από τις διαδικασίες που παρουσιάστηκαν και εφαρμόστηκαν σε συγκεκριμένο είδος δικτύου, όπως ο διαχωρισμός του δικτύου σε ομάδες και η ύπαρξη κοινών κόμβων/αισθητήρων σε γειτονικές ομάδες, έχουν γενικότερη ισχύ και μπορούν να εφαρμοστούν καθολικά με σκοπό την αποτελεσματικότερη ανίχνευση ανωμαλιών στο δίκτυο. Παράλληλα, η προτεινόμενη μεθοδολογία έχει σχεδιαστεί με γενικευμένο τρόπο ώστε οι βασικές της αρχές να μην επηρεάζονται από τα ιδιαίτερα στοιχεία ή μετρικά του κάθε είδους δικτύου, με μοναδική προϋπόθεση αποδοτικής λειτουργίας οι μετρήσεις γειτονικών αισθητήρων να παρουσιάζουν υψηλό βαθμό συσχέτισης.

Κατάλογος Δημοσιεύσεων

Διεθνή περιοδικά με Κρίση

1. V. Chatzigiannakis, G. Androulidakis, S. Papavassiliou, “Improving Network Anomaly Detection Effectiveness via an Integrated Multi-Metric-Multi-Link (M3L) PCA-based Approach”, Security and Communication Networks (Wiley), (on-line Oct. 2008), 2008.
2. V. Chatzigiannakis, S. Papavassiliou, “Diagnosing Anomalies and Identifying Faulty Nodes in Sensor Networks”, IEEE Sensors Journal , Vol. 7, Issue 5, pp. 637-645, 2007
3. V. Chatzigiannakis, M. Grammatikou, S. Papavassiliou, “Extending Driver's Horizon Through Comprehensive Incident Detection in Vehicular Networks”, IEEE Transactions on Vehicular Technology, Vol. 56, Issue 6, pp.3256-3265, 2007
4. G. Androulidakis , V. Chatzigiannakis, S. Papavassiliou, “Network anomaly detection and classification via opportunistic sampling”, IEEE Network, Vol. 23, Issue 1, pp. 6-12, 2009

Πρακτικά Διεθνών Επιστημονικών Συνεδρίων με Κρίση

1. V. Chatzigiannakis, M. Grammatikou, S. Papavassiliou, “Improving Incident Detection in Vehicular Networks: Methodology and Evaluation”, in Proc. of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2007, pp. 1-5, Athens, Greece, 2007
2. V. Chatzigiannakis, G. Androulidakis, K. Pelechrinis, S. Papavassiliou, V. Maglaris, “Data fusion algorithms for network anomaly detection: classification and evaluation”, in Proc. of the Third International Conference on Networking and Services (ICNS 2007), pp. 50-57, Athens, Greece, 2007.

3. G. Androulidakis, V. Chatzigiannakis and S. Papavassiliou, “Using Selective Sampling for the Support of Scalable and Efficient Network Anomaly Detection”, in Proc. of the 2nd Distributed Autonomous Network Management Systems Workshop (IEEE DANMS 2007), Washington D.C., USA, 2007.

4. A. Lenis, V. Chatzigiannakis, M. Grammatikou, S. Papavassiliou, “An Architectural Framework for the Support of Intelligent Vehicular Network Monitoring”, in Proc. of International Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS 2007) (ACM Digital Library), Vancouver, Canada, 2007.

5. V. Chatzigiannakis, S. Papavassiliou, G. Androulidakis, V. Maglaris, “On the realization of a generalized data fusion and network anomaly detection framework”, in Proc. of the Fifth International Symposium in Communication Systems, Networks and Digital Signal Processing (CSNDSP 2006), pp. 251-255, Patra, Greece, 2006

6. V. Chatzigiannakis, S. Papavassiliou, M. Grammatikou, B. Maglaris, “ Hierarchical Anomaly Detection in Distributed Large- scale Sensor Networks”, in Proc. of IEEE Symposium on Computers and Communications (IEEE ISCC, 2006), pp. 761-767, Pula-Cagliari, Sardinia, Italy, 2006.

7. G.Androulidakis, V.Chatzigiannakis, S.Papavassiliou, M.Grammatikou, V.Maglaris, “Understanding and Evaluating the Impact of Sampling on Anomaly Detection Techniques”, in Proc. IEEE Military Communications Conference 2006 (IEEE MILCOM2006), pp. 1-7, Washington D.C., USA, 2006

8. V. Chatzigiannakis, A. Lenis, C. Siaterlis, M. Grammatikou, D. Kalogeras, S. Papavassiliou and V. Maglaris, “Distributed Network Monitoring and Anomaly Detection as a Grid Application”, In Proc. of 12th Workshop of HP OpenView University Association (HP-OVUA), Porto, Portugal, 2005.

Βιβλιογραφία

[Anto96] C. N. Antoniades, Y. J. Stephanedes, "Single-Station Incident Detection Algorithm (SSID) for Sparsely Instrumented Freeway Sites", in Proc. of the 4th International Conference on Applications of Advanced Technologies in Transportation Engineering, pp. 218-221, 1996.

[Band03] S. Bandyopadhyay, E.J. Coyle, "An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks", in Proc. of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE INFOCOM, Vol. 3, pp. 1713-1723, 2003.

[Barf01] P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies", In Proc. of the First ACM SIGCOMM Internet Measurement Workshop, pp. 69-74, 2001

[Barf02] P. Barford, J. Kline, D. Plonka, and A. Ron, “A Signal Analysis of Network Traffic Anomalies”, in Proc of the Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement, pp. 71-82, 2002.

[Brot01] T. Brotherton, T. Johnson, "Anomaly detection for advanced military aircraft using neural networks", in Proc of the IEEE Aerospace Conference, Vol.6, pp.3113-3123, 2001

[Cann98] J. Cannady, "Artificial neural networks for misuse detection", In Proc. of the National Information Systems Security Conference (NISSC'98), pp. 443-456, 1998.

[Cao04] J. Cao, W.S. Cleveland, Y. Gao, K. Jeffay, F.D. Smith, and M.C. Weigle. "Stochastic Models for Generating Synthetic HTTP Source Traffic", in Proc. of IEEE INFOCOM, Vol. 3, pp.1546 - 1557, 2004.

[Cha107] V. Chatzigiannakis, G. Androulidakis, K. Pelechrinis, S. Papavassiliou, V. Maglaris, "Data fusion algorithms for network anomaly detection: classification and evaluation", in Proc. of Third International Conference on Networking and Services (ICNS 2007), Athens, Greece, June 2007

[Cha109] V. Chatzigiannakis, G. Androulidakis, S. Papavassiliou, “Improving Network Anomaly Detection Effectiveness via an Integrated Multi-Metric-Multi-Link (M3L) PCA-based Approach”, Security and Communication Networks (Wiley), (on-line Oct. 2008), 2009.

[Cha207] V. Chatzigiannakis, S. Papavassiliou, "Diagnosing Anomalies and Identifying Faulty Nodes in Sensor Networks", IEEE Sensors Journal , Vol. 7, Issue 5, pp. 637-645, 2007

[Cha307] V. Chatzigiannakis, M. Grammatikou, S. Papavassiliou, "Extending Driver's Horizon Through Comprehensive Incident Detection in Vehicular Networks", IEEE Transactions on Vehicular Technology, Vol 56, Issue 6, pp.3256-3265, 2007

[Chen03] Z. Chen, L. Gao, K. Kwiat, "Modeling the spread of active worms", Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Vol. 3, pp. 1890 - 1900, 2003.

[Chen06] A. Chen, B. Khorashadi, C. Chuah, D. Ghosal, M. Zhang , "Smoothing Vehicular Traffic Flow using Vehicular-based Ad Hoc Networking & Computing Grid (VGrid)", in Proc. of the IEEE Intelligent Transportation Systems Conference (ITSC), pp. 349-354, 2006.

[Chen97] H. Chen, R. Boyle, F. Montgomery, H. Kirby, M. Dougherty, "Motorway Incident Detection using Principal Component Analysis", in Proc. IEE Colloquium on Incident Detection and Management, pp. 1/1-1/5, 1997.

[Chow00] D. Chowdhury, L. Santen, A. Schadschneider , "Statistical Physics of Vehicular Traffic and Some Related Systems", Physics Reports, Vol. 329, Nos. 4-6, pp. 199-329, 2000.

[Doul04] C. Douligeris, A. Mitrokotsa, "DDoS attacks and defense mechanisms: classification and state-of-the-art", Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 44, Issue 5 , pp: 643 - 666, 2004

[Duni98] R. Dunia and S. J. Qin. "A Subspace Approach to Multidimensional Fault Identification and Reconstruction", American Institute of Chemical

Engineers (AIChE) Journal, pp. 1813-1831, 1998.

[Emra02] N. Ye, S. Emran, Q. Chen, S. Vilbert, “Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection”, IEEE Transactions on Computers, Vol. 51, No. 7, July 2002.

[Gand06] G.Androulidakis, V.Chatzigiannakis, S.Papavassiliou, M.Grammatikou, V.Maglaris, "Understanding and Evaluating the Impact of Sampling on Anomaly Detection Techniques", IEEE MILCOM 2006, Washington D.C., USA, 2006

[Gand09] G. Androulidakis , V. Chatzigiannakis, S. Papavassiliou, "Network anomaly detection and classification via opportunistic sampling", IEEE Network, Vol. 23, Issue 1, pp. 6-12, 2009

[Gome03] J. Gomez, F. Gonzalez, D. Dasgupta, "An immuno-fuzzy approach to anomaly detection", In Proc. of The 12th IEEE International Conference on Fuzzy Systems, Vol. 2, pp. 1219- 1224, 2003.

[GRIDCC] GRIDCC: http://www.gridcc.org

[GRNET] GRNET: http://www.grnet.gr/

[Hall92] D. Hall. Mathematical Techniques in Multisensor Data Fusion. Artech House, Norwood, Massachussets, 1992.

[Huan07] L. Huang, X. Nguyen, M. Garofalakis, J. M. Hellerstein, M. I Jordan, A. D. Joseph, N. Taft, "Communication-Efficient Online Detection of Network-Wide Anomalies", In Proc. of IEEE INFOCOM , pp. 134-142, 2007

[Hung01] H. L. Huang, P. Antonelli, "Application of Principal Component Analysis to High-Resolution Infrared Measurement Compression and Retrieval", Journal of Applied Meteorology, Vol. 40, Issue 10, pp. 365-388, 2001

[Huss03] A. Hussain, J. Heidemann, and C. Papadopoulos, “A Framework for Classifying Denial of Service Attacks”, In Proc. of the ACM SIGCOMM Conference, pp. 99-110, 2003.

[IDMEF] Intrusion Detection Message Exchange Format, RFC 4765

[Jack03] J. Edward Jackson. "A user's Guide to Principal Components", Wiley, 2003

[Jack79] J. E. Jackson and G. S. Mudholkar, "Control Procedures for Residuals Associated with Principal Component Analysis", Technometrics, pp. 341-349, 1979.

[Joll02] I. T. Jollife. "Principal Component Analysis", Second Edition, Springer, 2002

[Labi04] K. Labib and V. R. Vemuri, "Detecting and Visualizing Denial of Service And Network Probe Attacks Using Principal Component Analysis", in Proc. of SAR'04 the 3rd Conference on Security and Network Architectures, 2004.

[Lakh04] A.Lakhina, M.Crovella, C.Diot. "Diagnosing network-wide traffic anomalies", in Proc. of the conference on applications, technologies, architectures, and protocols for computer communications (SIGCOMM), pp. 219 - 230, 2004.

[Lakh05] A. Lakhina, M. Crovella, C. Diot, "Mining Anomalies Using Traffic Feature Distributions",in Proc. of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM), pp. 217 – 228 , 2005

[Lee01] W. Lee and D. Xiang, "Information-Theoretic Measures for Anomaly Detection", In Proc. of the IEEE Symposium on Security and Privacy (S&P 2001), pp. 130 -143, 2001.

[Luo05] F. Y. Luo, H. S. Lu L. Zhang, "Statistical en-route filtering of injected false data in sensor networks", IEEE Journal on Selected Areas in Communications, Vol. 23, Issue 4, pp. 839- 850, April 2005.

[Mart00] P. T. Martin, H.J. Perrin, B. G. Hansen, "Incident Detection Algorithm Evaluation", Technical Report UTL-0700-31, Utah Traffic Laboratory, 2000.

[Mehm04] C. Mehmet Vuran, B. Akan, Ian F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks", Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 45, Issue 3, pp. 245-259, 2004.

[Mora08] A. Moralis, V. Pouli, S. Papavassiliou, V. Maglaris, "A Kerberos security architecture for web services based instrumentation grids", Future Generation Computer Systems, Elsevier, Available online, 2008

[Netf04] RFC 3954 (rfc3954) – Cisco Systems NetFlow Services Export Version 9

[News04] J. Newsome, E. Shi, D. Song, A. Perrig, "The Sybil Attack in Sensor Networks: Analysis& Defenses", in Proceedings of the Third International Symposium on Information Processing in Sensor Networks, pp. 259 - 268, 2004.

[Nides] NIDES: http://www.sdl.sri.com/projects/nides/

[NS08] Network Simulator: http://nsnam.isi.edu/nsnam/index.php/User_Information

[Oka04] M. Oka, Y. Oyama, H. Abe, and K. Kato, “Anomaly Detection Using Layered Networks Based on Eigen Co-occurrence Matrix”, in Proc. of the Seventh International Symposium on Recent Advances in Intrusion Detection (RAID),pp. 223-237, 2004.

[OpE08] OpenEye: http://sheep.netmode.ntua.gr/openeye/

[Palm03] T. Palpamas, et al, "Distributed Deviation Detection in Sensor Networks," in Proceedings of ACM SIGMOD, Vol. 32, Issue 4, pp. 77 - 82, 2003.

[Perr04] E. Shi, A. Perrig, "Designing secure sensor networks", Wireless Communications, Vol.11, Issue 6, pp. 38- 43, Dec. 2004.

[Pfan05] E. Pfannerstill, "Object Recognition and Correlation Methods for Traffic Flow Analysis", in Proc. of the 8th International IEEE Conference on Intelligent Transportation Systems, pp. 290-295, 2005.

[Red09] Code Red Worm: http://en.wikipedia.org/wiki/Code_Red_worm

[Shaf76] Glenn Shafer. A Mathematical Theory of Evidence. Princeton University Press, Princeton, 1976.

[Shyu03] M. Shyu, S. Chen, K. Sarinnapakorn, L. Chang, "A Novel Anomaly Detection Scheme Based on Principal Component Classifier", in Proc. of IEEE Foundations and New Directions of Data Mining Workshop (in conjunction with the Third IEEE International Conference on Data Mining), pp. 172-179, 2003.

[Siat05] C.Siaterlis, B. Maglaris, "One step ahead to multisensor data fusion for DDoS detection", Journal of Computer Security, pp. 779 - 806 , Vol. 13, Issue 5, 2005.

[Slam09] SQL Slammer: http://en.wikipedia.org/wiki/SQL_slammer_(computer_worm)

[Song03] B. Przydatek, D. Song, A. Perrig, "SIA: secure information aggregation in sensor networks", in Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, pp. 255 - 265, 2003.

[SUMO08] SUMO: http://sumo.sourceforge.net

[Tana05] S. Tanachaiwiwat, A. Helmy, "Correlation analysis for alleviating effects of inserted data in wireless sensor networks", in Proceedings of Mobile and Ubiquitous Systems: Networking and Services, pp. 97- 108, 2005.

[Wang03] G. Wang, W. Zhang, G. Cao, T. La Porta, "On supporting distributed collaboration in sensor networks", In Proceedings of the IEEE Military Communications Conference, Vol. 2, pp. 752- 757, Oct. 2003.

[Wood02] A. D. Wood, J. A. Stankovic, "Denial of service in sensor networks", IEEE Computer, Vol. 35, Issue 10,pp. 54- 62, Oct. 2002.

[Yang04] X. Yang, L. Liu, N.H. Vaidya, F. Zhao, "A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning", In Proc. of the Mobile and Ubiquitous Systems: Networking and Services, pp. 114-123 , 2004.

[Youn04] O. Younis, S. Fahmy, "A Hybrid, Energy-efficient, Distributed Clustering

Approach for Ad Hoc Sensor Networks", IEEE Transactions on Mobile Computing, Vol. 3, Issue 4, pp. 366-379, 2004.

[Κουτ04] Γεώργιος Κουτέπας, "Συνεργατικό Σύστημα Αντιμετώπισης Κατανεμημένων Επιθέσεων Άρνησης Υπηρεσίας σε Περιβάλλον Πολλαπλών Διαχειριστικών Περιοχών", Διδακτορική Διατριβή ΕΜΠ, 2004

[1998] 1998 DARPA Intrusion Detection:

http://www.ll.mit.edu/IST/ideval/data/data_index.html