ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εντοπισμός Επιθέσεων σε Ασύρματα Αυτοργανούμενα Δίκτυα με Μεθόδους Ανάλυσης Κύριων Συνιστωσών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Θεοδόσιου Λ. Τζούτζα

Επιβλέπων : Συμεών Παπαβασιλείου
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2009

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

Εντοπισμός Επιθέσεων σε Ασύρματα Αυτοργανούμενα Δίκτυα με Μεθόδους Ανάλυσης Κύριων Συνιστωσών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Θεοδόσιου Λ. Τζούτζα

Επιβλέπων : Συμεών Παπαβασιλείου
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7η Ιουλίου 2009.

.....	.....	.....
Συμεών Παπαβασιλείου	Ιάκωβος Βενιέρης	Δήμητρα–Θεοδώρα Κακλαμάνη
Επικ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγήτρια Ε.Μ.Π.

Αθήνα, Ιούλιος 2009

.....

Θεοδόσιος Λ. Τζούτζας

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

Περίληψη

Οι ασύρματες επικοινωνίες αποτελούν μια από τις πλέον αναπτυσσόμενες και πολλά υποσχόμενες περιοχές της σύγχρονης τεχνολογίας. Όμως η αυξανόμενη δραστηριότητα γύρω από την περιοχή των τηλεπικοινωνιών θα πρέπει να συνοδεύεται από την ανάλογη αξιοπιστία εφόσον η χρήση των συγκεκριμένων τεχνολογιών επεκτείνεται με ταχείς ρυθμούς σε όλο το φάσμα της ανθρώπινης δραστηριότητας. Συνεπώς, εξίσου κρίσιμος καθίσταται και ο τομέας της έρευνας για την ασφάλεια αυτών των τεχνολογιών ώστε να προστατευτούν ευπαθή και όχι μόνο τμήματα συστημάτων επικοινωνιών και κατ' επέκταση να διατηρείται η υψηλή πιστότητα, αποδοτικότητα και αξιοπιστία των συστημάτων αυτών.

Σκοπός της παρούσας διπλωματικής εργασίας υπήρξε η ανάπτυξη και μελέτη μεθόδου για τον εντοπισμό επιθέσεων σε ασύρματα αυτοργανούμενα (ad hoc) δίκτυα η οποία βασίζεται στην μέθοδο M^3L που χρησιμοποιεί την Ανάλυση Κύριων Συνιστωσών (*PCA*). Ο εντοπισμός κόμβων που είναι ευπαθείς σε επιθέσεις καθώς επίσης και ο ίδιος ο εντοπισμός της πηγής των επιθέσεων είναι καθοριστικής σημασίας για την εύρυθμη λειτουργία ενός ασύρματου συστήματος. Όπως έχει προκύψει από προηγούμενες μελέτες τα χαρακτηριστικά της επίθεσης διαδραματίζουν σημαίνοντα ρόλο στη λειτουργία του ασύρματου δικτύου. Στην παρούσα διπλωματική επιχειρήθηκε ο εντοπισμός επιθέσεων με βάση τα χαρακτηριστικά δικτύου και επίθεσης που είναι κρίσιμα.

Το πλαίσιο της παρούσας εργασίας διαφέρει θεμελιωδώς από εκείνο των μελετών με παραδοσιακές προσεγγίσεις αφού βασίζεται σε πιθανοτικά μοντέλα. Η μη ντετερμινιστική προσέγγιση που υιοθετείται και αναλύεται αρχικά αποτελεί ένα ρεαλιστικό μοντέλο τόσο για τα ασύρματα αυτοργανούμενα δίκτυα αισθητήρων όσο και για το μοντέλο επίθεσης πάνω στα οποία βασίστηκε το πειραματικό μέρος αυτής της εργασίας και τα τελικά πειραματικά αποτελέσματα.

Το μοντέλο του υποκείμενου δικτύου της παρούσας εργασίας λαμβάνει πιθανοτικές προσεγγίσεις για την περίπτωση επιθέσεων σε ασύρματα ad hoc δίκτυα από τυχαία κινούμενη απειλή. Επεκτείνει προηγούμενες μελέτες που αφορούν τη διάδοση μολύνσεων σε τέτοιου είδους δίκτυα προχωρώντας στον εντοπισμό επιθέσεων σε ασύρματα ad hoc και δίκτυα αισθητήρων με μια μέθοδο ικανή να ανιχνεύσει επιθέσεις συγκεκριμένων χαρακτηριστικών κατά τη διάρκεια της εξέλιξης αυτών των επιθέσεων. Εν κατακλείδι, ο αλγόριθμος που προτείνεται σε αυτήν την εργασία δύναται να ανιχνεύσει μολυσμένους κόμβους του δικτύου και κατ' επέκταση είναι δυνατόν να ιχνηθετήσει την πορεία του επιτιθέμενου κατά την εξέλιξη μιας επίθεσης.

Λέξεις κλειδιά: <<ασύρματα αυτοργανούμενα δίκτυα, δίκτυα αισθητήρων, ασφάλεια, μοντέλο επίθεσης και διάδοσης, ανίχνευση ανωμαλίας, Ανάλυση Κύριων Συνιστωσών>>

Abstract

Wireless communications have recently become one of the most rising areas of modern technology. Owing to this development, such systems are necessary to be accordingly reliable, on the grounds that these technological advances are being used in almost all kinds of human activity. Therefore, the area of security is considered to be equally crucial, and it is related with wireless technology in order to protect vulnerable communication systems and thus to ensure the quality, performance and reliability of these systems.

The aim of this thesis is to develop and study a methodology concerning the detection of potential attacks that might occur in ad hoc networks. This methodology is based upon the M^3L method that uses the *Principal Component Analysis* (*PCA*). The detection of vulnerable nodes in ad hoc networks as well as the detection of the attacker is of major significance for the functionality of any kind of wireless systems. Previous studies have shown that the characteristics of the attack play central role to the way the wireless system works. The main attempt of this thesis is to detect attacks that have the aforementioned characteristics which are essential for network performance.

The proposed framework has fundamental differences from the traditional approaches that were presented in the past basically because it relies on probabilistic models. The non–deterministic approach is adopted as the proper realistic model both for the underlying ad hoc/sensor network and the attack model. The latter was used as the basis for the experiments that took place during this thesis and subsequently for the final outputs of the corresponding experiments.

The underlying network model follows probabilistic approaches for potential attacks in wireless ad hoc networks by a randomly mobile threat. This thesis expands the existing studies that are related to the propagation of infections in such networks and moreover detects attacks in wireless sensor networks by applying a method which can detect those attacks that have specific characteristics during the attack process. To sum up, the proposed algorithm is capable to detect efficiently infected and thus malfunctioned nodes of the network and therefore, this algorithm enables us to track the attacker's path during the course of its action.

Keywords: <<wireless ad hoc networks, sensor networks, security, attack modeling and propagation, anomaly detection, Principal Component Analysis>>

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία συντάχθηκε το ακαδημαϊκό έτος 2008–2009 στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου υπό την επίβλεψη του επίκουρου Καθηγητή κ. Συμεών Παπαβασιλείου, προς τον οποίο θα ήθελα να εκφράσω τις ευχαριστίες μου για την εμπιστοσύνη που έδειξε στο πρόσωπό μου στην ανάθεση ενός τόσο επιστημονικά ενδιαφέροντος θέματος. Με αυτό τον τρόπο, μου έδωσε τη δυνατότητα να γνωρίσω ένα εντελώς άγνωστο σε εμένα επιστημονικό πεδίο, τα ασύρματα ad hoc δίκτυα. Δουλεύοντας πάνω σε αυτά ανακάλυψα το τεράστιο επιστημονικό ενδιαφέρον που παρουσιάζει το συγκεκριμένο αντικείμενο, πράγμα το οποίο θα επηρεάσει σημαντικά την επιλογή του αντικειμένου των μελλοντικών σπουδών μου.

Επίσης ένα μεγάλο ευχαριστώ οφείλω στους υπεύθυνους της διπλωματικής μου εργασίας κκ. Βασίλειο Καρυώτη και Βασίλειο Χατζηγιαννάκη για την άψογη συνεργασία που είχαμε. Χωρίς την υπομονή, τις συμβουλές, τις ιδέες, την επιστημονική τους πείρα και την αμέριστη βοήθειά τους δεν θα ήταν δυνατή η επιτυχής περάτωση του παρόντος έργου. Οι υποδείξεις και οι παρατηρήσεις που μου έγιναν κατά τη διάρκεια εκπόνησης της εργασίας αυτής εκτός από την καθοριστική συμβολή που είχαν για την ολοκλήρωσή της με ενθάρρυναν στο να εμβαθύνω ακόμα περισσότερο σε κάποια ζητήματα, ενώ παράλληλα με βοήθησαν στο να βελτιώσω τον τρόπο εργασίας μου.

Αφιερώνεται στους γονείς μου
Λεωνίδα και Φωτούλα

Περιεχόμενα

Πίνακας Σχημάτων

Κεφάλαιο 1^{ο} – Εισαγωγή

1.1 Γενικά στοιχεία

Είναι γεγονός ότι τα τελευταία χρόνια οι ασύρματες επικοινωνίες κερδίζουν συνεχώς μερίδιο στις προτιμήσεις των καταναλωτών. Κάτι τέτοιο επιβάλλεται, κατά κύριο λόγο, από τις απαιτήσεις του σύγχρονου τρόπου ζωής. Η τεχνολογία, δηλαδή, στρέφεται από τις παραδοσιακά ενσύρματες, στις σύγχρονες ασύρματες λύσεις επικοινωνίας. Πράγματι, οι σύγχρονοι χρήστες έχουν αφήσει κατά μέρος τις κλασικές μεθόδους επικοινωνίας όπως αυτές του ενσύρματου τηλεφώνου και της γραπτής αλληλογραφίας και στρέφονται προς νέες όπως η κινητή τηλεφωνία, το διαδίκτυο, κ.ά.

Γενικά, όπως είναι γνωστό, οι ασύρματες τεχνολογίες χρησιμοποιούν για τη μεταφορά δεδομένων το ηλεκτρομαγνητικό φάσμα (ραδιοκύματα, μικροκύματα και οπτικά κύματα). Το γεγονός αυτό καθιστά πιο ευέλικτη και πρόσφορη την ανάπτυξη των ασύρματων εφαρμογών έναντι των αντίστοιχων ενσύρματων. Η μεγάλη αύξηση του όγκου των χρηστών και των απαιτήσεων επικοινωνίας αυτών καθιστά αναγκαία την ανάπτυξη ασύρματων δικτύων. Στις μέρες μας έχουν αναπτυχθεί διάφορα είδη ασύρματων δικτύων για την παροχή διαφορετικών υπηρεσιών.

1.2 Ad hoc δίκτυα

Μεταξύ άλλων η κύρια τάση που συγκεντρώνει μεγάλο ενδιαφέρον σήμερα είναι η ανάπτυξη των λεγόμενων ασύρματων δικτύων συγκεκριμένου σκοπού η οποία είναι διεθνώς γνωστή με την ορολογία ad hoc ασύρματα δίκτυα. Τα ad hoc δίκτυα, που αποτελούν αντικείμενο μελέτης αυτής της εργασίας, διαφέρουν θεμελιωδώς από τα κλασικά ασύρματα δίκτυα διότι δεν εμφανίζουν, όπως τα τελευταία, συγκεκριμένη δομή. Ένα ad hoc δίκτυο αποτελείται από έναν αριθμό κινητών εν γένει χρηστών – κόμβων που επικοινωνούν μεταξύ τους μέσω ασύρματων μεταδόσεων. Αυτοί οι κόμβοι σχηματίζουν ένα δίκτυο το οποίο δεν παρουσιάζει μια σταθερή, κεντρική υποδομή ή οργάνωση.

Τα ad hoc δίκτυα παρουσιάζουν ευρεία χρήση λόγω της εύκολης, γρήγορης ανάπτυξής τους και της μειωμένης εξάρτησης τους από υποδομή. Βασικά χαρακτηριστικά αυτών των δικτύων είναι η αυτοργάνωση, η αυτοανάπτυξη, η αποκέντρωση και η δυναμική τοπολογία τους. Μερικά από τα ζητήματα που προσπαθούν να συμβιβάσουν τα εν λόγω δίκτυα είναι το περιορισμένο εύρος ζώνης, το πρόβλημα κατανάλωσης της ενέργειας τους, καθώς και διάφορα θέματα ασφάλειας που χρίζουν αντιμετώπισης.

1.3 Δίκτυα αισθητήρων

Μια κατηγορία ασύρματων δικτύων που εμφανίζει αλματώδη ανάπτυξη στις μέρες μας είναι τα λεγόμενα δίκτυα αισθητήρων. Τα δίκτυα αισθητήρων αποτελούν μια ειδική κατηγορία αυτοργανούμενων δικτύων η οποία δεν αποτελεί λύση επικοινωνίας γενικού σκοπού. Είναι σαφώς προσανατολισμένα στη μέτρηση παραμέτρων του περιβάλλοντος. Αποτελούνται από χωρικά κατανεμημένες αυτόνομες συσκευές αισθητήρων με σκοπό τη συλλογική παρακολούθηση διαφόρων φυσικών ή περιβαλλοντικών συνθηκών μιας περιοχής, όπως θερμοκρασία, ήχος, δόνηση, πίεση, υγρασία ή κίνηση.

Όμως, οι κόμβοι – αισθητήρες, εκτός της ασύρματης δυνατότητας επικοινωνίας, διαθέτουν κάποιο επίπεδο υπολογιστικής ισχύος για επεξεργασία των συλλεγόμενων

δεδομένων κάτι το οποίο δεν παρατηρείται γενικά στα ad hoc δίκτυα. Αυτά τα δίκτυα εξαρτώνται σε μεγαλύτερο βαθμό από τα χαρακτηριστικά της εφαρμογής για την οποία διατίθενται σε σχέση με τα ad hoc δίκτυα τα οποία εμφανίζουν εν γένει μια σχετική ομοιομορφία. Διαφορές μεταξύ των δύο τύπων δικτύων εμφανίζονται, επίσης, ως προς την επεκτασιμότητα (τα δίκτυα αισθητήρων είναι πιο επεκτάσιμα), την ενεργειακή κατανάλωση (τα δίκτυα αισθητήρων έχουν μεγαλύτερες ενεργειακές απαιτήσεις), την ποιότητα υπηρεσίας (*Quality of Service*, *QoS*) ανάλογα με την εφαρμογή, τα πρότυπα κίνησης των κόμβων και τη δρομολόγηση των δεδομένων μέσα στα δίκτυα.

Ένα βασικό πρόβλημα που εμφανίζεται και στα δύο προαναφερθέντα είδη ασύρματων δικτύων (ad hoc και δίκτυα αισθητήρων) είναι αυτό της ασφάλειας. Μάλιστα το θέμα της ασφάλειας γίνεται επιτακτικότερο λόγω της αυξανόμενης διάδοσης αυτών των δικτύων και της ευρείας χρήσης τους σε κρισιμότερες, για τον άνθρωπο, εφαρμογές.

Είναι γεγονός ότι τα θέματα ασφάλειας στα ασύρματα δίκτυα τίθενται με έναν ιδιαίτερο τρόπο. Η διαφοροποίηση των απαιτήσεων ασφάλειας στα ασύρματα δίκτυα γίνεται λόγω της φύσης αυτών των δικτύων και κυρίως λόγω των μεγάλων περιορισμών ενέργειας, υπολογιστικής επεξεργασίας και εύρους μετάδοσης που συνήθως υπάρχει σε αυτά τα δίκτυα. Η εμπιστευτικότητα, η πιστοποίηση, η ακεραιότητα, η διαθεσιμότητα και η εγκυρότητα (ακρίβεια) των δεδομένων είναι γενικά κάποιες από τις πιο βασικές απαιτήσεις και αντικείμενα ασφάλειας. Από την άλλη, μια από τις βασικότερες απειλές για ένα ασύρματο δίκτυο αισθητήρων είναι η εισβολή/εμφάνιση ενός κακόβουλου κόμβου στο δίκτυο ο οποίος μπορεί να παρεμποδίσει την αποστολή/λήψη δεδομένων από έναν κόμβο του δικτύου ή να αλλάξει τα δεδομένα ενός κόμβου.

1.4 Παρούσα εργασία

Στόχος της παρούσας διπλωματικής εργασίας είναι η ανίχνευση ανωμαλιών και ο εντοπισμός επιθέσεων σε ασύρματα αυτοργανούμενα δίκτυα. Επιχειρείται η μελέτη της παρουσίας ενός κακόβουλου κόμβου σε ένα ασύρματο δίκτυο αισθητήρων με σκοπό την ανάπτυξη ενός αλγόριθμου για την ανίχνευση και τον εντοπισμό μιας απειλής που κινείται εντός της περιοχής του δικτύου. Ο κακόβουλος κόμβος κινείται εντός των ορίων του δικτύου και μπορεί να επικοινωνήσει ασύρματα με όλους τους κόμβους αυτού μολύνοντας τους. Η μόλυνση των κόμβων έχει την έννοια της εισαγωγής κάποιου ποσοστιαίου σφάλματος επί των πραγματικών μετρούμενων μεγεθών. Δηλαδή, κάθε μολυσμένος κόμβος περιέχει λανθασμένη τιμή του μεγέθους το οποίο μετρά και εν δυνάμει, λόγω της επικοινωνίας του με άλλους κόμβους, μπορεί να τη μεταδώσει. Επειδή, όπως αναφέρθηκε παραπάνω, αυτά τα δίκτυα χρησιμοποιούνται σε σημαντικές εφαρμογές καθίσταται βασική απαίτηση η έγκαιρη ανίχνευση ενός σφάλματος – μόλυνσης.

Για τη διάδοση της κινητής απειλής μέσα στο δίκτυο χρησιμοποιούμε ένα πιθανοτικό μοντέλο που έχει αναλυθεί και μελετηθεί με βάσει ένα ad hoc δίκτυο **[1], [2]** και **[3]**. Από την άλλη, η ανίχνευση της κινητής απειλής και οι ανωμαλίες που αυτή με τυχαίο τρόπο μπορεί να μεταδώσει μέσα στο δίκτυο γίνεται με τη βοήθεια της τεχνικής M^3L η οποία βασίζεται στην Ανάλυση Κύριων Συνιστωσών (*Principal Component Analysis – PCA*) και έχει χρησιμοποιηθεί στην ανίχνευση της ακεραιότητας και ακρίβειας δεδομένων σε μεμονωμένους έκθετους ή δυσλειτουργούντες κόμβους **[4]**.

Με άλλα λόγια γίνεται σύγκλιση δύο τεχνικών, η μια από τη σκοπιά του επιτιθέμενου και η άλλη από τη σκοπιά του δικτύου που δέχεται την επίθεση με σκοπό την εύρεση ενός αλγόριθμου ανίχνευσης σφαλμάτων/ανωμαλιών στα πλαίσια ενός ασύρματου δικτύου αισθητήρων όταν υπάρχει μία ελεύθερα κινούμενη απειλή στο εν λόγω δίκτυο. Η πειραματική προσέγγιση που παρουσιάζουμε στα επόμενα κεφάλαια έγινε για πραγματικές

περιπτώσεις που χρησιμοποιούν μετεωρολογικά δεδομένα τα οποία συλλέγονται από κατανεμημένο σύνολο κόμβων αισθητήρων.

Το υπόλοιπο αυτής της εργασίας είναι οργανωμένο, ως ακολούθως: Στο Κεφάλαιο 2 περιγράφονται παλαιότερες εργασίες οι οποίες είναι σχετικές με το αντικείμενο της παρούσας διπλωματικής. Στις εργασίες που περιγράφονται στο Κεφάλαιο 2 στηρίζεται και το μοντέλο συστήματος της εργασίας αυτής το οποίο αναλύεται στο Κεφάλαιο 3. Κατόπιν, στο Κεφάλαιο 4 γίνεται αναφορά στο μοντέλο προσομοίωσης που υλοποιήθηκε στα πλαίσια αυτής της εργασίας, ενώ ακολουθεί το Κεφάλαιο 5 με τα πειραματικά αποτελέσματα των προσομοιώσεων που πραγματοποιήθηκαν. Τέλος, στο Κεφάλαιο 6 δίνονται περιληπτικά τα συμπεράσματα των πειραματικών εξαγομένων του Κεφαλαίου 5 και προτείνονται μελλοντικές κατευθύνσεις έρευνας.

Κεφάλαιο 2^{ο} – Σχετικές Εργασίες

2.1 Μοντέλα μόλυνσης και διάδοσης κακόβουλου λογισμικού

Οι μολύνσεις και οι διάφοροι τύποι αυτών αποτελούσαν από παλιά αντικείμενο ενδιαφέροντος για τις κοινότητες της βιολογίας. Πρόσφατα το ενδιαφέρον αυτό επεκτάθηκε στις περιοχές της πληροφορικής και των δικτύων επικοινωνιών. Όπως μπορεί εύκολα να διαπιστωθεί από πραγματικά σενάρια επίθεσης δικτύων υπολογιστών, η διάδοση ιών μεταξύ υπολογιστών μοιάζει με αυτήν των μολύνσεων στους βιολογικούς πληθυσμούς.

Η σύγκλιση που παρατηρείται στην πληροφορική και τη βιολογία, σε ότι αφορά το θέμα των μολύνσεων, οφείλεται στον κοινό μηχανισμό που υπάρχει στην εξάπλωσης μιας μόλυνσης. Γενικά, σε έναν κλειστό βιολογικό πληθυσμό, κάποια από τα μέλη του μπορεί να μολυνθούν. Διάφορα μέλη αυτού του πληθυσμού που αλληλεπιδρούν με ήδη μολυσμένα μέλη, μπορεί να μολυνθούν (*infection*) και αυτά και έτσι να αρχίσουν να διαδίδουν περαιτέρω τη μόλυνση. Από την άλλη, την ίδια στιγμή, υπάρχει πιθανότητα κάποιο μέλος να ανακάμψει (*recovery*) από τη μόλυνση. Ο ίδιος μηχανισμός μπορεί να υιοθετηθεί, αντίστοιχα, για κάποιον κλειστό πληθυσμό στο πεδίο της πληροφορικής και των δικτύων υπολογιστών ειδικότερα.

Γενικά, στη βιβλιογραφία καταγράφονται δύο κύριες κλάσεις μοντέλων μόλυνσης. Στο πρώτο μοντέλο, ένα μέλος πληθυσμού μπορεί να είναι ευπαθές ή μολυσμένο. Ευπαθές καλείται οποιοδήποτε μέλος του πληθυσμού το οποίο είναι υγιές αλλά επιρρεπές στο να γίνει μολυσμένο. Το συγκεκριμένο μοντέλο έχει πάρει την ονομασία του από τη συνήθη ακολουθία των καταστάσεων των κόμβων, ευπαθές – μολυσμένο – ευπαθές (***Susceptible – Infected – Susceptible, SIS***) **[5]**. Στο δεύτερο μοντέλο, που αναφέρεται ως ευπαθές – μολυσμένο – διαγραφόμενο (***Susceptible – Infected – Removed, SIR***), τα μέλη του πληθυσμού θεωρούνται αρχικά ευπαθή (υγιή αλλά πιθανά να μολυνθούν). Όταν μολυνθούν, δεν μπορούν να επανέλθουν στην υγιή κατάσταση και αφαιρούνται από τον εν λόγω πληθυσμό **[5]**. Το τελευταίο μοντέλο αναφέρεται σε περιπτώσεις στις οποίες η μόλυνση είναι είτε θανάσιμη (μη βιώσιμη) για τα μέλη του πληθυσμού, είτε η διαδικασία επαναφοράς οδηγεί σε μόνιμη ανοσία κατά συγκεκριμένου τύπου μολύνσεως. Σε οποιαδήποτε περίπτωση, τα μολυσμένα μέλη του πληθυσμού μπορούν να αγνοηθούν (μετά από ένα συγκεκριμένο χρονικό διάστημα) από την υπόλοιπη ανάλυση χωρίς να επηρεαστεί η ακεραιότητα της τελευταίας, αφού αυτά δε μπορούν να μολυνθούν ξανά και κατά συνέπεια να επηρεάσουν τη διάδοση μιας απειλής.

Στη βιολογία, το πρόβλημα της διάδοση ενός ιού, όσον αφορά τους ζωντανούς οργανισμούς, έχει μελετηθεί εκτενώς στο πεδίο της επιδημιολογίας **[5]** και **[6]**. Στην επιδημιολογία, ο πληθυσμός αντιστοιχίζεται σε ένα γράφο δικτύου, όπου οι κόμβοι παριστάνουν τα μέλη του εν λόγω πληθυσμού και οι ακμές του δικτύου τις αλληλεπιδράσεις ανάμεσα σε διάφορα ζεύγη του πληθυσμού. Ο αριθμός των μολυσμένων κόμβων του δικτύου είναι η άγνωστη μεταβλητή του προβλήματος και το ζητούμενο είναι η διατύπωση μιας διαφορικής εξίσωσης συναρτήσει του χρόνου. Από τη λύση αυτής της εξίσωσης προκύπτει μια αναλυτική συνάρτηση του αναμενόμενου αριθμού των μολυσμένων κόμβων του δικτύου σε σχέση με το χρόνο **[5]** και **[6]**. Η λύση αυτή δίνεται ως συνάρτηση του αριθμού των συνολικών κόμβων, του ρυθμού μόλυνσης l για ένα

ζευγάρι κόμβων, του ρυθμού επαναφοράς m για έναν κόμβο και της χρονικής συνάρτησης του αριθμού των διαγραφόμενων κόμβων (αν υπάρχει).

Η αναλυτική λύση οδηγεί σε ένα επιδημικό κατώφλι

$$\frac{l}{m} < \frac{1}{\langle k \rangle} \quad \acute{\eta} \quad \frac{l}{m} < \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{1}$$

όπου $\langle k \rangle$ δηλώνει το μέσο όρο του βαθμού ενός κόμβου και $\langle k^2 \rangle$ τη διασπορά του βαθμού ενός κόμβου. Το κατώφλι αυτό ορίζει πότε μια μόλυνση θα γίνει ενδημική ή θα εξασθενήσει (εκθετικά) στο τέλος. Οι συγκεκριμένες παράμετροι της αναλυτικής λύσης μεταβάλλονται ανάλογα με το θεωρούμενο μοντέλο (*SIS* ή *SIR*) και το επιλεγμένο επιδημιολογικό μοντέλο (π.χ. **[7], [8], [9]** ή **[10]**, **[11]**). Για παράδειγμα, θεωρώντας το μοντέλο *SIS* (**[7]**) για τη μοντελοποίηση μιας επίθεσης, η μη γραμμική διαφορική εξίσωση:

$$\frac{dn}{dt} = b\,n(1-n) - dn \tag{2}$$

περιγράφει τη δυναμική του συστήματος, όπου το $n(t)$ είναι το ποσοστό των μολυσμένων κόμβων, b είναι ο ρυθμός μόλυνσης και d ο ρυθμός ανάκαμψης **[12]**. Η λύση της εξίσωσης (2) είναι :

$$n(t) = \frac{n_0(1-r)}{n_0 + (1-r-n_0)e^{-(b-d)t}} \tag{3}$$

όπου $n_0 = n(t=0)$ και $p = d/b$. Όπου χρησιμοποιείται το μοντέλο *SIR* (**[10]**) η δυναμική του συστήματος ελέγχεται από το σύνολο των εξισώσεων:

$$\begin{cases} dI(t)/dt = b\,I(t)S(t) - g\,I(t) \\ dU(t)/dt = g\,I(t) \\ N = I(t) + U(t) + S(t) \end{cases} \quad \textbf{(4)}$$

όπου το N δηλώνει το συνολικό αριθμό των κόμβων, τα $I(t)$, $S(t)$, $U(t)$ δηλώνουν τον αριθμό των μολυσμένων, ευπαθών και διαγραφόμενων κόμβων αντίστοιχα, b είναι ο ρυθμός μόλυνσης των κόμβων και g είναι ο ρυθμός απομάκρυνσης των κόμβων **[11]**.

Ωστόσο, αυτή η προσέγγιση, αναδεικνύει μόνο το υποκείμενο πρότυπο του μοντέλου διάδοσης και όχι τους σχετικούς κρίσιμους παράγοντες. Επιπλέον, όλα τα μέλη θεωρούνται ικανά για απευθείας επικοινωνία μεταξύ τους σε οποιαδήποτε χρονική στιγμή, ώστε να ορίζουν έναν ομογενή πληθυσμό. Αυτό όμως διαφέρει θεμελιωδώς από τις αντίστοιχες αλληλεπιδράσεις και την ανάπτυξη των ασύρματων ad hoc δικτύων, όπου συνήθως εγκαθίσταται πολυδιαδρομικές (*multihop*) επικοινωνίες.

Από τη σκοπιά των δικτύων υπολογιστών, η μοντελοποίηση μολύνσεων σε τερματικούς (*end-host*) υπολογιστές βασίζεται στην αναγνώριση ειδικών λειτουργιών, που έχουν άμεση σχέση με επιθέσεις σε δίκτυα υπολογιστών, για παράδειγμα, *port scanning*, επιθέσεις *IP domain* ή εφαρμογή επιδημιολογικών εξισώσεων **[11]**, **[12]**, **[13]** και **[14]**. Η παρατήρηση κάποιας ασυνήθιστης συμπεριφοράς του δικτύου (κίνηση) δείχνει την παρουσία κάποιου είδους επίθεσης και η εξέλιξη της επίθεσης μοντελοποιείται επίσης, όπως στην περίπτωση των βιολογικών πληθυσμών, από επιδημιολογικά μοντέλα. Τα τελευταία προσαρμόζονται ώστε να απεικονίζουν τα διαφορετικά λειτουργικά χαρακτηριστικά των υπολογιστικών δικτύων – *mailing lists*, *web pages*, και *physical link interface connectivity* **[15], [11]**, **[12]** και **[13]**. Προσομοιώσεις και πειράματα που πραγματοποιήθηκαν έχουν μελετηθεί και χρησιμοποιηθεί σε μεγάλο βαθμό για την αξιολόγηση των αντίστοιχων επιδημιολογικών μοντέλων όπως αυτά που παρουσιάζονται στο **[14]**, όπου το μοντέλο *SIS* έχει επιλεγεί και μοντελοποιηθεί σύμφωνα με την

εξίσωση (2) και στο **[11]**, όπου το μοντέλο *SIR* έχει επιλεγεί και μοντελοποιηθεί μέσω της εξίσωσης (4).

Και στις δύο προαναφερθείσες μεθοδολογίες η συνήθης προσέγγιση για την αντιμετώπιση της διάδοσης είναι η προληπτική, ή κατά τη διάρκεια της ανάκαμψης, ανοσοποίηση πολλών κρίσιμων σημείων του δικτύου **[11]**, **[12]** και **[13]**. Αυτό το πρότυπο μεταφράζεται σε ρυθμούς εξαφανίσεων των κόμβων του δικτύου (απορρίπτοντας τους ανοσοποιημένους και άρα μη ευπαθείς κόμβους), οι οποίοι ενσωματώνονται στη διαφορική εξίσωση του μοντέλου.

Επιπλέον, στη βιβλιογραφία καταγράφονται αρκετές προσπάθειες για τη μελέτη και αντιμετώπιση των μολύνσεων που φέρουν συγκεκριμένα χαρακτηριστικά. Η αλγεβρική προσέγγιση με πίνακες **[9]**, λόγου χάρη, αποτελεί μια συστηματική και θεωρητική πρόταση για την αντιμετώπιση της διάδοσης ενός προγράμματος τύπου *worm* στα δίκτυα υπολογιστών. Η συγκεκριμένη μέθοδος συνδέει το ρυθμό μόλυνσης l και το ρυθμό ανάκαμψης m, με τον πίνακα γειτνίασης A του υποκείμενου δικτύου. Ορίζοντας την πιθανότητα μολύνσεως για έναν k–συνδεδεμένο κόμβο, λαμβάνεται η πιθανότητα αυτός ο κόμβος να είναι υγιής και κατόπιν η πιθανότητα αυτή συνδέεται με τις ιδιοτιμές του πίνακα γειτνίασης. Η μεγαλύτερη από αυτές, δηλαδή η φασματική ακτίνα του πίνακα γειτνίασης, χρησιμοποιήθηκε για την εξαγωγή ενός πιο γενικού επιδημιολογικού ορίου, που έχει αποδειχθεί ότι διατηρείται για αυθαίρετο τύπο δικτύων (χαμηλής ισχύος – ενέργειας, τυχαία, ομογενή, κ.τ.λ.). Το προτεινόμενο επιδημιολογικό όριο t δίνεται από τη σχέση:

$$t = \frac{1}{l_{1,\mathrm{A}}} \qquad \textbf{(5)}$$

όπου $l_{1,A}$ είναι η φασματική ακτίνα του πίνακα A. Πρέπει να σημειωθεί ότι η εξίσωση (5) περιγράφει τα όρια για όλους τους τύπους γράφων δικτύων. Όμως, το μοντέλο που

μελετήσαμε εμπλέκει διακριτό χρόνο σε αντίθεση με το συνεχή χρόνο που θεωρήθηκε σε όλα τα προαναφερθέντα μοντέλα.

Από την άλλη, στη βιβλιογραφία, απαντώνται στοχαστικές μέθοδοι οι οποίες βασίζονται στην πιθανότητα ενός κόμβου να μολυνθεί από ένα κανάλι επικοινωνίας και σχετίζονται με ένα ρυθμό ανάκαμψης. Αυτές οι μέθοδοι φαίνονται να παρέχουν ένα πολλά υποσχόμενο και πιο συστηματικό πλαίσιο μοντελοποίησης. Σύμφωνα με αυτό, σε κάθε κόμβο και για κάθε κανάλι επικοινωνίας αντιστοιχίζεται μια μη μηδενική πιθανότητα μόλυνσης και μία αντίστοιχη πιθανότητα ανάκαμψης. Τα εφαρμοζόμενα πρότυπα μόλυνσης και ανάκαμψης μπορούν να επιλεγούν ώστε να ακολουθούν διάφορα στοχαστικά μοντέλα **[16]** και **[17]**. Καθώς οι περισσότερες από τις προηγούμενες προσεγγίσεις αναφέρονται σε ενσύρματα δίκτυα, η τελευταία εμφανίζεται ως πιο ελκυστική και κατάλληλη για ασύρματα ad hoc δικτυακά περιβάλλοντα.

2.2 Αντιμετώπιση μολύνσεων στα δίκτυα αισθητήρων

Ανεξάρτητα με τα παραπάνω, οι μολύνσεις και η διάδοση τους έχουν μελετηθεί και σε επίπεδο δικτύων αισθητήρων. Ο έλεγχος τέτοιου είδους δικτύων γίνεται μέσω της δυναμικής εξέτασης των στοιχείων που προέρχονται από διάφορες ροές κόμβων αισθητήρων. Ο συνδυασμών αυτών των στοιχείων με κανονικά πρότυπα, κρίνεται αναγκαίος, προκειμένου να ανιχνευτούν πιθανές ανωμαλίες. Σε περιπτώσεις που υπάρχουν απαιτήσεις υποστήριξης στην αποστολή κρίσιμων εφαρμογών, οι αισθητήρες πρέπει να έχουν μηχανισμούς για τη διασφάλιση των επικοινωνιών και την εγκυρότητα των συλλεχθέντων στοιχείων. Μερικές ενδεικτικές εργασίες που αφορούν στην ασφάλεια των δικτύων αισθητήρων παρατίθεται, εν συντομία, ακολούθως.

Αρχικά, στα **[18]**, **[19]** και **[20]**, οι συγγραφείς παρουσιάζουν αρκετά σενάρια επίθεσης που αξιοποιούν τις αδυναμίες των Ασύρματων Δικτύων Αισθητήρων (***Wireless Sensor Networks – WSN***). Η επεκτασιμότητα όσο αφορά την ανάπτυξη των *WSN* απαιτεί

προσεκτικές αποφάσεις και συμβιβασμούς μεταξύ διάφορων και αντίθετων φιλοσοφιών σχεδίασης μέτρων ασφάλειας. Οι συγγραφείς πραγματεύονται αυτά τα θέματα και θεωρούν μηχανισμούς που πετυχαίνουν ένα υψηλότερο επίπεδο ασφάλειας και αξιοπιστίας στα εν λόγω δίκτυα.

Ωστόσο, το πρόβλημα των εσφαλμένων δεδομένων λόγω της ύπαρξης είτε εκτεθειμένων είτε ελαττωματικών κόμβων στο δίκτυο γίνεται μεγαλύτερης σημασίας όταν εμφανίζεται συνάθροιση δεδομένων. Στο **[21]**, οι συγγραφείς παρουσιάζουν ένα στατιστικό ενδο–δρομολογιακό (*en–route*) μηχανισμό για ανίχνευση και εξαγωγή ψευδών αναφορών κατά τη διάρκεια της διαδικασίας προώθησης. Υποθέτοντας ότι ένα γεγονός μπορεί να ανιχνευτεί από πολλούς αισθητήρες, καθένας από τους οποίους παράγει ένα κλειδωμένο κώδικα επικύρωσης (επιβεβαίωσης – πιστοποίησης) μηνυμάτων (*MAC*) και πολλαπλά *MACs* επισυνάπτονται στην αναφορά του γεγονότος. Όσο η αναφορά προωθείται, μεταξύ των κόμβων, κάθε κόμβος επιβεβαιώνει την ακρίβεια του *MAC* με πιθανοτικό τρόπο και απορρίπτει εκείνους με άκυρα *MACs*.

Στο **[22]**, το πρόβλημα των ελαττωματικών ή κακόβουλων κόμβων τυποποιείται όπως κατά την κατασκευή ενός ανεξάρτητου δένδρου για να καλύψει όλους τους ύποπτους γείτονες και να δώσει το χαμηλότερο όριο της πολυπλοκότητας των μηνυμάτων. Η προηγούμενη σχετική εργασία που αναφέρεται στη βιβλιογραφία έχει εστιάσει στην ανίχνευση των αποκλίσεων στα πρότυπα των δεδομένων μεταξύ των αισθητήρων.

Στο **[23]**, οι συγγραφείς παρουσιάζουν ένα μοντέλο με μηχανισμό τυχαίας δειγματοληψίας ώστε να ελέγξουν αν οι τιμές που επιστρέφονται από τους συναθροιστές είναι προσεγγιστικά καλές σε σχέση με τις πραγματικές, ακόμα και όταν οι συναθροιστές και ένα μέρος των κόμβων των αισθητήρων είναι εσφαλμένοι. Ωστόσο, οι δοκιμές που περιγράφονται χρησιμεύουν μόνο ως απόδειξη της ιδέας που παρουσιάζεται, επειδή αποτελούνται από απλούς αλγόριθμους όπως η εύρεση των ελάχιστων, των μέγιστων και των ενδιάμεσων τιμών.

Στο **[24]**, προτείνεται μια χωροχρονική ανάλυση συσχέτισης αποκαλούμενη "ανώμαλη δοκιμή σχέσεων" ("*abnormal relationships test*" – *ART*), για την ανίχνευση πολύ μεγάλων διακυμάνσεων από τη μέση τιμή στα συλλεχθέντα δεδομένα. Αυτή η μέθοδος είναι βασισμένη στις δοκιμές των συντελεστών συσχέτισης μεταξύ των κόμβων. Κάθε κόμβος αποθηκεύει τις τιμές των δοκιμών σε ένα κινούμενο παράθυρο, ώστε να σημειώνεται κάθε γείτονας που αρχίζει να αναφέρει αποκλίνουσες τιμές, οι οποίες υπερβαίνουν ένα προκαθορισμένο όριο. Η ανίχνευση αυτή επιτυγχάνεται με συνεργασία μεταξύ των κόμβων ώστε να απομονώνονται οι προβληματικοί κόμβοι.

Στο **[25]**, οι συγγραφείς περιγράφουν μια τεχνική για την αναγνώριση σε πραγματικό χρόνο πολύ μεγάλων διακυμάνσεων από τη μέση τιμή στις ενδείξεις που συλλέγονται από μεμονωμένους ασύρματους αισθητήρες και προσπαθούν να επεκτείνουν αυτή την τεχνική σε ένα ολόκληρο δίκτυο αισθητήρων, που λαμβάνουν υπόψη την κατανεμημένη επεξεργασία των γεγονότων. Εντούτοις, η τεχνική αυτή απαιτεί την πλήρη γνώση της συνάρτησης κατανομής πυκνότητας των συλλεχθέντων δεδομένων.

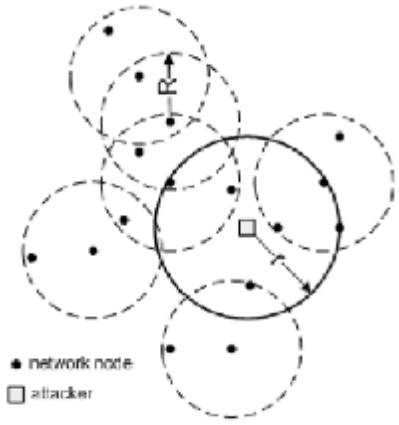
Τέλος, η προσέγγιση που ακολουθούμε εστιάζει στην αποδοτική ανίχνευση των πολύ μεγάλων διακυμάνσεων από τη μέση τιμή σε όλο το δίκτυο αισθητήρων με κατανεμημένο τρόπο και βασίζεται στη χρήση της τεχνικής *PCA* **[26]**. Η *PCA* έχει χρησιμοποιηθεί εκτενώς για να παρέχει πολύ αποδοτικούς τρόπους μοντελοποίησης των χωροχρονικών συσχετίσεων που εμφανίζουν τα δεδομένα και η βασική της αρχή έχει χρησιμοποιηθεί σε αρκετά πεδία, όπως η ανίχνευση γεγονότων σε πραγματικό χρόνο **[27]**, η ανίχνευση εισβολής **[28]**, και η ανίχνευση ανωμαλιών στη διαδικτυακή κίνηση **[29]**. Σημειώνουμε ότι υπάρχουν και άλλες τεχνικές που έχουν προέλθει από τη *PCA*, όπως η *Kernel PCA* (*KPCA*) τεχνική που παρέχει μη γραμμική μείωση της διάστασης των δεδομένων, σε αντίθεση με την αρχική *PCA* που είναι γραμμική. Στο **[30]**, οι συγγραφείς προτείνουν τη χρήση της *KPCA* για ανίχνευση ελαττωματικών αισθητήρων στη διαδικασία του χημικού ελέγχου.

Κεφάλαιο 3^{ο} – Μοντελοποίηση

3.1 Μοντελοποίηση ασύρματου δικτύου αισθητήρων

Ένα ασύρματο ad hoc δίκτυο μπορεί να αναπαρασταθεί από ένα γράφο δικτύου σε οποιαδήποτε χρονική στιγμή, ανάλογα με τη θέση των κόμβων, την ποιότητα του ασύρματου καναλιού, την ακτίνα μετάδοσης και το μοντέλο μετάδοσης/λήψης. Σε ένα ασύρματο δίκτυο, ένας κόμβος μπορεί να συνδέεται με πολλούς άλλους κόμβους μέσω ενός μοναδικού φυσικού ασύρματου καναλιού. Δύο κόμβοι θεωρούνται γειτονικοί αν υπάρχει κανάλι επικοινωνίας μεταξύ τους, δηλαδή αν μπορούν με αξιοπιστία να ανταλλάξουν μεταξύ τους πακέτα. Δεδομένου της φύσης των μεταδόσεων των ασύρματων ζεύξεων (*broadcast* μεταδόσεις), ένας κόμβος μπορεί να συνδέεται ταυτοχρόνως με περισσότερους από ένα γείτονες. Εντούτοις, σε αυτά τα περιβάλλοντα δικτύωσης είναι μάλλον απίθανο όλοι οι κόμβοι να μπορούν να επικοινωνήσουν απευθείας ο ένας με τον άλλον. Στις περισσότερες περιπτώσεις, οι κόμβοι πρέπει να κάνουν πολλαπλά ενδιάμεσα βήματα (*hops*) για να φτάσουν έναν τυχαία επιλεγμένο προορισμό μέσα στο δίκτυο. Αυτό είναι μια θεμελιώδης διαφορά από την ομοιογενή περίπτωση που έχει υποτεθεί στην επιδημιολογία. Επομένως, τα αντίστοιχα μοντέλα που αναπτύσσονται στην επιδημιολογία και αναφέρονται σε βιολογικούς πληθυσμούς, δεν μπορούν άμεσα να εφαρμοστούν για τη μοντελοποίηση της διάδοσης ιών τύπου *worm* στα ασύρματα ad hoc δίκτυα. Από την άλλη, οι πιθανοτικές προσεγγίσεις φαίνονται καταλληλότερες στη σύλληψη των ιδιαίτερων χαρακτηριστικών και γνωρισμάτων τέτοιων περιβαλλόντων δικτύωσης.

Στην ανάλυσή μας εξετάζεται ένα σύνολο ασυρμάτων κόμβων (με πληθυκότητα ίση με N), όπου ο καθένας έχει μια σταθερή ακτίνα μετάδοσης R. Η εικόνα που ακολουθεί παρουσιάζει κατά γενικό τρόπο ένα ασύρματο ad hoc δίκτυο.



Εικόνα 1 Τοπολογία ασύρματου ad hoc δικτύου

Οι κόμβοι είναι ομοιόμορφα και τυχαία κατανεμημένοι πάνω σε μία δισδιάστατη περιοχή συγκεκριμένου μεγέθους και σχήματος. Κατά συνέπεια, η πυκνότητα των κόμβων γίνεται ανάλογη προς τον αριθμό των κόμβων του δικτύου. Στο μοντέλο μας θεωρούμε ένα μοναδικό κακόβουλο, κινητό κόμβο, με δυνατότητα να ρυθμίζει την ακτίνα μετάδοσής του, *r*, έτσι ώστε να προσαρμόζει τη στρατηγική του, ενώ οι κόμβοι του υποκείμενου δικτύου θεωρούνται στατικοί. Επιπλέον, θεωρείται δυνατόν οι κόμβοι του δικτύου, μόλις μολυνθούν, να μολύνουν τους γείτονές τους, διαδίδοντας έτσι τις πιθανές μολύνσεις σε ολόκληρο το δίκτυο. Η συμπεριφορά αυτή μοντελοποιεί τις περιπτώσεις όπου ένας επιτιθέμενος παίρνει τον έλεγχο αξιοποιήσιμων μηχανημάτων και χρησιμοποιεί μερικά από τα τμήματα (*modules*) τους που τον ενδιαφέρουν **[13]**. Πρέπει να σημειωθεί, ότι το συγκεκριμένο μοντέλο αποτελεί το τμήμα του μοντέλου συστήματος της παρούσας εργασίας που αφορά τη σκοπιά του επιτιθέμενου και δίνει τη δυνατότητα να αντιληφθούμε τις δυνατότητες για πρόκληση βλάβης στο δίκτυο, οι οποίες στη συνέχεια μπορούν να ληφθούν υπόψη στη σχεδίαση αποδοτικών αντίμετρων. Αργότερα, σε αυτό το κεφάλαιο,

συμπληρώνεται το μοντέλο με το τμήμα που αφορά τη σκοπιά του δικτύου το οποίο δέχεται την επίθεση. Όσον αφορά το μοντέλο από τη σκοπιά του επιτιθέμενου, παρατηρούμε ότι η χρήση της υποκείμενης τοπολογίας του δικτύου, μας επιτρέπει να αντιληφθούμε καλύτερα την επίδραση των λειτουργικών χαρακτηριστικών του κινητού κακόβουλου κόμβου στη συνολική διαδικασία της διάδοσης. Το πρότυπο κινητικότητας του επιτιθέμενου, όπως και η δυνατότητά του να μεταβάλλει την ακτίνα μετάδοσής του, επηρεάζουν τα κανάλια επικοινωνίας ολόκληρου του δικτύου, παρά το γεγονός ότι το υποκείμενο δίκτυο παραμένει αμετάβλητο, δημιουργώντας έτσι ένα δυναμικό περιβάλλον.

Παρόλα αυτά, εκτός από την κινητικότητα, οι διακυμάνσεις του ασύρματου καναλιού συμβάλουν επίσης στην τυχαιότητα του δικτύου. Συγκεκριμένα, όσο το ασύρματο κανάλι ποικίλει σημαντικά με το χρόνο, οι διακυμάνσεις του μπορούν να περιγραφούν στατιστικά ως μεταβολές μικρής και μεγάλης κλίμακας. Ο πρώτος τύπος μοντελοποιείται μέσω του μοντέλου λογαριθμοκανονικών διαλείψεων (***lognormal shadowing***) και ο δεύτερος μέσω της κατανομής ***Rayleigh*** **[31]**. Στην παρούσα εργασία, αγνοούμε τελείως τις παραπάνω διακυμάνσεις μικρής κλίμακας, ενώ λαμβάνουμε το μέσο όρο όλων των διακυμάνσεων μεγάλης κλίμακας. Συνεπώς, θεωρούμε ότι το λαμβανόμενο σήμα έχει εξάρτηση αντιστρόφου δύναμης της απόστασης από τον κόμβο που μεταδίδει (δηλώνοντας τον αντίστοιχο εκθέτη με a , τη σταθερά απωλειών διαδρομής).

Υπό αυτές τις προϋποθέσεις, η γειτονιά ενός κόμβου καθορίζεται πλήρως από την ακτίνα μετάδοσής του. Επομένως, όλα τα μηχανήματα (***hosts***) μέσα σε δίσκο ακτίνας ίσης με το εύρος μετάδοσης, με κέντρο έναν κόμβο X , αποτελούν τη γειτονιά του X . Ο αποστολέας μπορεί να προσαρμόσει τη γειτονιά του αλλάζοντας την ισχύ μετάδοσης, **[32]** και **[33]**, αφού υψηλότερη ισχύς μετάδοσης οδηγεί σε αντίστοιχα υψηλότερο εύρος μετάδοσης και άρα σε μεγαλύτερη περιοχή γειτονίας.

3.2 Μοντέλο επίθεσης

Όπως σημειώθηκε προηγουμένως, εφόσον οι ασύρματοι κόμβοι ενός ad hoc δικτύου ή δικτύου αισθητήρων είναι τοπολογικά περιορισμένοι (ένας κόμβος μπορεί να επικοινωνήσει άμεσα μόνο με ένα μικρό υποσύνολο των υπολοίπων κόμβων του δικτύου), το ομοιογενές μοντέλο που θεωρήθηκε στην επιδημιολογία, **[5]** και **[6]**, δεν είναι κατάλληλο για αυτό το σύστημα. Όμοια, η διαδικασία της διάδοσης δεν μπορεί να μοντελοποιηθεί πλήρως ως αιτιοκρατικό (ντετερμινιστικό) φαινόμενο που περιγράφεται από διαφορικές εξισώσεις. Αντιθέτως, λόγω της πολυπλοκότητας του δικτύου, τα διάφορα γεγονότα πραγματοποιούνται με διακριτό, στοχαστικό τρόπο.

Οι προαναφερθείσες τοπολογικές ιδιότητες, μαζί με το δυναμικό χαρακτήρα του ασύρματου περιβάλλοντος και τη στοχαστική φύση της διάδοσης ιών τύπου *worm*, απαιτούν πιθανοτική ανάπτυξη ενός ακριβούς και ανιχνεύσιμου μοντέλου διάδοσης επίθεσης. Τα δίκτυα του συγκεκριμένου τύπου σε αντίθεση με τα κλασικά ασύρματα δίκτυα μονής διαδρομής (***single – hop networks***) βασίζονται στην απουσία κεντρικής σταθερής δομής και οργάνωσης (***lack of central infrastructure)***. Κάθε κόμβος τους δύναται να ενεργεί ως δρομολογητής επιτρέποντας με αυτόν τον τρόπο τη δημιουργία επικοινωνίας πολλαπλής διαδρομής μεταξύ των κόμβων του δικτύου (***multi – hop networks***). Επιπλέον, οι κόμβοι των ασύρματων ad hoc δικτύων είναι σε υψηλό βαθμό κινούμενοι (***highly mobile***) γεγονός που οδηγεί στην ταχύτατη και μη προβλέψιμη αλλαγή της τοπολογίας του δικτύου. Παράλληλα, η συνδεσιμότητα (***connectivity***) μεταξύ των κόμβων επίσης δεν είναι σταθερή αλλά μεταβάλλεται με το χρόνο **[34]**. Για όλους τους παραπάνω λόγους, υιοθετήθηκε ένα πρωτότυπο στοχαστικό πλαίσιο (***framework***) για τη μοντελοποίηση της διάδοσης επίθεσης σε ασύρματα δίκτυα τύπου ad hoc.

Από την άλλη, εισήχθησαν και οριστήκαν μεγέθη με σκοπό την αποτίμηση της συνολικής απόδοσης και το χαρακτηρισμό των διαφόρων στρατηγικών επίθεσης. Όσον αφορά τις στρατηγικές επίθεσης χρησιμοποιήθηκε το μετρικό (***metric***) ***Infection Efficiency***

(*βλ.* παρ. 3.4) ως μέτρο αποδοτικότητας της εκάστοτε στρατηγικής επίθεσης. Επιπρόσθετα, έγινε χρήση μιας σειράς μεγεθών όπως το SPE, $P_{contribution}$, P_d, P_f, P_{in}, P_{ingat} (*βλ.* παρ. 3.6.4, 3.10 και 3.11) ως μέτρο αποδοτικότητας από τη σκοπιά του δικτύου (μέτρηση ανίχνευσης επιθέσεων, μολυσμένων κόμβων).

Πριν από τη διαδικασία εισαγωγής και την περιγραφή του πιθανοτικού μοντέλου παρατίθενται κάποια σχόλια που αφορούν τη διαδικασία μόλυνσης των κόμβων σε ad hoc δίκτυα. Οι πιο κοινές απειλές, σήμερα, προέρχονται από προγράμματα *worm* που είτε επισυνάπτονται σε *emails* και εκτελούνται από το χρήστη που τα λαμβάνει **[13]**, είτε ενσωματώνονται σε άλλο λογισμικό. Η σάρωση δικτύου (*network scanning*), διαδικασία που γίνεται σε ένα δίκτυο για την ανίχνευση ενεργών *host* είτε για λόγους επίθεσης είτε για λόγους ασφαλείας, εκτελείται από ειδικά μηχανήματα *hosts* που ονομάζονται *bots*. Ο κύριος ρόλος των τελευταίων είναι να συλλέγουν και να αναλύουν πληροφορίες που αφορούν εξυπηρετητές του Διαδικτύου. Για την περίπτωση που μας αφορά, τα *bots* αναζητούν την *IP* διεύθυνση ευάλωτων μηχανημάτων. Αυτού του είδους οι επιθέσεις ακολουθούν δύο προσεγγίσεις: είτε ενεργοποιούνται όταν ο επιτιθέμενος θέλει να βρει ανοικτές πόρτες (*port scanning*), είτε είναι προενεργοποιημένοι, όταν ο ιός είναι ενσωματωμένος μέσα στο λογισμικό και ο χρήστης επιλέγει να εκτελέσει το αντίστοιχο πρόγραμμα **[35]**. Επιπλέον, ένας επιτιθέμενος μπορεί να επηρεάσει τα λειτουργικά χαρακτηριστικά ενός δικτυακού κόμβου, ειδικά στην περίπτωση των δικτύων αισθητήρων. Στέλνοντας είτε εσφαλμένα, είτε μη αναμενόμενου τύπου/ποσότητας δεδομένα, ο επιτιθέμενος μπορεί να μειώσει τη λειτουργική ικανότητα ενός δικτυακού κόμβου. Παρόμοια, υπάρχουν προενεργοποιημένες και ενεργοποιημένες μέθοδοι για την ανάκαμψη όταν συμβαίνει μια μόλυνση από ιό. Με τον όρο *antivirus* εννοούμε κάθε πρόγραμμα σχεδιασμένο για την προστασία του υπολογιστή από διαφόρων ειδών επιθέσεις **[36]**. Στην προενεργοποιημένη μέθοδο, ένα πρόγραμμα *antivirus* εκτελείται μόνιμα και όταν γίνεται αντιληπτή η λειτουργία ασυνήθιστου κώδικα, το πρόγραμμα συνδέεται με κεντρικό

εξυπηρετητή και κατεβάζει τα κατάλληλα *patches*. Με τον όρο *patch* αναφερόμαστε σε ένα πρόγραμμα το οποίο ενσωματώνεται και βελτιώνει ένα άλλο, συνήθως προγενέστερης έκδοσης, πρόγραμμα **[37]**. Στη διαδραστική μέθοδο, το λογισμικό *antivirus* ενημερώνει το χρήστη για συγκεκριμένη μόλυνση και ο χρήστης επιλέγει την κατάλληλη ενέργεια – αντίδραση, για την επαναφορά και/ή επαναλειτουργία του *patch* του αντίστοιχου πόρου.

Στην ενότητα που ακολουθεί, υιοθετούμε ένα πλαίσιο για τη μοντελοποίηση της διάδοσης μιας επίθεσης σε ασύρματα ad hoc δίκτυα.

3.3 Στοχαστικό πλαίσιο διάδοσης κακόβουλου λογισμικού

Χωρίς να επηρεάζουμε καθόλου την αξιοπιστία της ανάλυσης, εστιάζουμε στη μόλυνση ενός κόμβου και όχι στη διαδικασία ή στον τύπο της μόλυνσης. Σύμφωνα με αυτό, ένας κόμβος που ανακάμπτει (*recovered*), μπορεί να επαναμολυνθεί καθ' όλη τη διάρκεια λειτουργίας του δικτύου. Ειδικότερα, δεν ενδιαφερόμαστε για το μηχανισμό (έλλειψη ανοσοποίησης ή καινούρια απειλή) της μόλυνσης αυτής καθ' αυτής, αλλά για το γεγονός ότι ένας κόμβος μπορεί να επαναμολυνθεί κάποια στιγμή. Ενδιαφερόμαστε, με άλλα λόγια, για τη μακροσκοπική συμπεριφορά που παρουσιάζει ένας κόμβος, που περιγράφεται μόνο από την κατάσταση του κόμβου και τη διάδοση της απειλής σε όλο το δίκτυο. Αυτή η προσέγγιση ακολουθεί γενικότερες θεωρήσεις της πραγματικής συμπεριφοράς ενός ad hoc δικτύου, όπου συνδυασμός επιθέσεων (ευρέως γνωστών, νέων επιθέσεων και ποικίλων τύπων μολύνσεων) μπορούν να συνυπάρξουν και να διασπαρθούν στο δίκτυο **[14]**, **[16]** και **[12]**. Σε αυτό το πλαίσιο, θεωρούμε έναν κακόβουλο κόμβο, ο οποίος μπορεί να μολύνει κόμβους που έχουν επανακάμψει, είτε εφαρμόζοντας παραλλαγή μιας επίθεσης που έχει ήδη εφαρμοστεί, είτε εισάγοντας μια καινούρια επίθεση. (Επιπλέον, αυτό δικαιολογεί το γεγονός ότι ένας κακόβουλος κόμβος συνήθως έχει υψηλότερη πιθανότητα μόλυνσης άλλων κόμβων, σε σχέση με τη δυνατότητα που έχουν οι ήδη μολυσμένοι κόμβοι του δικτύου.) Έτσι, το *SIS* μοντέλο, που περιγράφηκε στο

προηγούμενο κεφάλαιο, ταιριάζει καλύτερα στα παραπάνω πλαίσια για τη διάδοση μιας απειλής.

Η πιθανότητα μόλυνσης ενός καναλιού είναι η βάση για την πιθανοτική διατύπωση του προβλήματος. Για έναν κόμβο που έχει ένα κανάλι επικοινωνίας με έναν μολυσμένο γείτονα, υπάρχει μια συγκεκριμένη πιθανότητα να μολυνθεί και ο ίδιος από το κανάλι αυτό. Ως γνωστό, ένας κόμβος σε ένα ασύρματο ad hoc δίκτυο έχει, συνήθως, περισσότερους από έναν κόμβους–γείτονες. Ο αριθμός των μολύνσεων που δέχεται ένας κόμβος από ένα κανάλι γίνεται επομένως τυχαία μεταβλητή. Υποθέτουμε ότι η μεταβλητή αυτή ακολουθεί την κατανομή *Poisson*. Συνεπώς, οι ενδιάμεσοι χρόνοι άφιξης μιας μόλυνσης σε ένα κανάλι επικοινωνίας κατανέμονται εκθετικά **[38]** και **[39]**. Επειδή διάφοροι χρήστες ακολουθούν διαφορετικές διαδικασίες ανάκαμψης όταν μολυνθούν, η διαδικασία ανάκαμψης είναι και αυτή στοχαστικής φύσης. Μάλιστα, θεωρούμε ότι και αυτή η διαδικασία ακολουθεί την κατανομή *Poisson*, οπότε τα χρονικά διαστήματα μεταξύ διαδοχικών ανακάμψεων είναι ανεξάρτητα και εκθετικά κατανεμημένα **[38]** και **[39]**. Η χρήση των υποθέσεων *Poisson* αναφορικά με τις διαδικασίες μόλυνσης και ανάκαμψης, απλοποιεί και κάνει πιο εφικτή την αντίστοιχη ανάλυση του συστήματος, επιτρέποντας μια συστηματική αξιολόγηση του πιθανοτικού πλαισίου που προτείνεται, ενώ παράλληλα έρχεται σε συμφωνία με τις σχετικές προσεγγίσεις και υποθέσεις που χρησιμοποιούνται στη βιβλιογραφία για την αντιμετώπιση παρόμοιων προβλημάτων όπως για παράδειγμα στα **[16]** και **[14]**. Εντούτοις, η γενίκευση των διαδικασιών μόλυνσης – ανάκαμψης και η ενσωμάτωση τυχαίων προτύπων, αποτελεί μέρος παρούσας και μελλοντικής έρευνας.

Επομένως, το σχεδιαζόμενο μοντέλο κόμβων μπορεί να αντιστοιχιστεί σε ένα μοντέλο ουράς *M/M/1* **[40]**, όπου μια μόλυνση αντιστοιχεί σε μια άφιξη που απαιτεί εξυπηρέτηση, ενώ η ανάκαμψη αντιστοιχεί σε μια εξυπηρέτηση. Τόσο οι αφίξεις, όσο και οι αναχωρήσεις κατανέμονται εκθετικά. Δεδομένου ότι ο επιτιθέμενος μπορεί να κινείται σε όλο το δίκτυο, μολύνοντας τους υπόλοιπους κόμβους και μειώνοντας την ενέργειά του, το δίκτυο θα αποκατασταθεί πλήρως μετά από μικρό χρονικό διάστημα αφότου ο κινητός

κόμβος εξαντλήσει την ενέργειά του και σβήσει οριστικά. Στο μοντέλο *SIS*, η κατάσταση ενός κόμβου είναι δυαδική. Δηλαδή, ένας κόμβος μπορεί να είναι είτε μολυσμένος είτε μη μολυσμένος. Έτσι, η κατάσταση όλου του συστήματος δίνεται από ένα δυαδικό διάνυσμα, τα στοιχεία του οποίου αντιστοιχούν στους κόμβους του δικτύου και η τιμή κάθε γραμμής δείχνει την τρέχουσα κατάσταση του κόμβου. Για ένα δίκτυο ad hoc με N κόμβους, το διάστημα καταστάσεων του συστήματος έχει πληθυκότητα 2^N.

Όπως έχει γίνει ήδη σαφές υπάρχουν δύο τύποι γεγονότων: μόλυνση ενός κόμβου (*infection*) και ανάκαμψη ενός μολυσμένου κόμβου (*recovery*). Από τα παραπάνω μπορούμε εύκολα να συμπεράνουμε ότι τα χρονικά διαστήματα μεταξύ δύο επιτυχημένων γεγονότων, σε ότι αφορά έναν κόμβο, είναι εκθετικά κατανεμημένα με ρυθμό το άθροισμα των επιμέρους ρυθμών των διαδικασιών μόλυνσης και ανάκαμψης.

Αν συμβολίσουμε με m το συνολικό αριθμό των κόμβων που δεν είναι μολυσμένοι σε ένα δίκτυο N κόμβων και ενός επιτιθέμενου, τότε $N-m$ κόμβοι του δικτύου είναι μολυσμένοι (εξαιρουμένου του επιτιθέμενου). Επιπλέον, έστω m' ο αριθμός των κόμβων που δεν είναι μολυσμένοι σε ένα δεδομένο στιγμιότυπο και έχουν τον κακόβουλο κόμβο ως γείτονα. Θεωρώντας ότι ένα γεγονός έχει μόλις πραγματοποιηθεί, με τις παραπάνω υποθέσεις για τις διαδικασίες μόλυνσης και ανάκαμψης, το χρονικό διάστημα για το επόμενο γεγονός είναι εκθετικά κατανεμημένο με ρυθμό:

$$\sum_{i=1}^{m'}[(k_i-1)\,l_i + l_{mal}] + \sum_{i=m'+1}^{m} k_i\,l_i + \sum_{i=m+1}^{N} m_i \qquad \textbf{(6)}$$

όπου k_i είναι ο αριθμός των μολυσμένων γειτόνων του κόμβου i (συμπεριλαμβανομένου του επιτιθέμενου όπου αυτό εφαρμόζεται), m_i δηλώνει το ρυθμό ανάκαμψης, l_i δηλώνει το ρυθμό μόλυνσης του καναλιού μεταξύ του κόμβου i και οποιουδήποτε άλλου γείτονα του συγκεκριμένου κόμβου και l_{mal} δηλώνει το ρυθμό μόλυνσης μεταξύ ενός κόμβου του δικτύου και του επιτιθέμενου. Ο δείκτης αθροίσματος διατρέχει όλη την ακολουθία των

κόμβων του δικτύου, όπου χωρίς βλάβη της γενικότητας, υποθέτουμε ότι οι μη μολυσμένοι κόμβοι που έχουν τον επιτιθέμενο ως γείτονα είναι οι πρώτοι m' στη σειρά, οι μη μολυσμένοι κόμβοι που δεν έχουν τον επιτιθέμενο ως γείτονα έπονται στην ακολουθία μέχρι τον αριθμό m και οι υπόλοιποι κόμβοι είναι οι μολυσμένοι για το παραπάνω θεωρούμενο χρονικό στιγμιότυπο. Οι διαδικασίες ανάκαμψης και μόλυνσης υποθέτουμε ότι είναι ανεξάρτητες μεταξύ τους.

Έχει δειχθεί ότι κατά τη διάρκεια της διαδικασίας της επικοινωνίας έχουμε τη μεγαλύτερη κατανάλωση ενέργειας για έναν ασύρματο κόμβο, ενώ η επεξεργασία και η άεργος ενεργειακή κατανάλωση είναι σχετικά μικρές συγκρινόμενες με τις ανάγκες του πομποδέκτη **[41]**. Συνεπώς, ο χρόνος ζωής του κακόβουλου κόμβου εξαρτάται κυρίως από τη διαθέσιμη ενέργεια και την ακτίνα μετάδοσής του. Δεδομένου ενός συγκεκριμένου χρονικού διαστήματος και δύο ασύρματων κόμβων με διαφορετικές ακτίνες μετάδοσης, η κατανάλωση ενέργειας θα είναι μεγαλύτερη για τον κόμβο με τη μεγαλύτερη ακτίνα, ενώ ταυτόχρονα ο ίδιος κόμβος θα έχει μειωμένη διάρκεια ζωής. Στην πραγματικότητα, όσο μεγαλύτερη είναι η διαφορά στην ακτίνα, τόσο μεγαλύτερη είναι και η διαφορά στην κατανάλωση. Σε αυτή την εργασία, υιοθετούμε ένα πρότυπο κατανάλωσης της ενέργειας μέσω της ακόλουθης επαναληπτικής διαδικασίας:

$$E_{avail} = E_{prev} - C(t_{sp} r^{a}) E_{step} \quad \textbf{(7)}$$

όπου E_{avail} είναι η νέα (ενημερωμένη) ενέργεια, E_{prev} είναι η ενέργεια πριν την ενημέρωση, E_{step} δηλώνει το βήμα καταναλώσεως της ενέργειας, r η ακτίνα μετάδοσης του επιτιθέμενου, t_{sp} δηλώνει το χρονικό διάστημα από την τελευταία ενημέρωση της ενέργειας, a είναι η σταθερά των απωλειών διάδοσης (*path loss*) και C είναι μια σταθερά κανονικοποίησης.

3.4 Μέτρα επίδοσης και παραμετροποίηση

Ένας κινητός κόμβος, όπως έχει ήδη ειπωθεί, μπορεί να ρυθμίσει μερικά από τα λειτουργικά του χαρακτηριστικά, προκειμένου να ενισχυθεί η ικανότητα του στο να βλάψει το δίκτυο αλλά και να αναπτυχθούν διαφορετικές και αποτελεσματικές στρατηγικές επίθεσης. Λόγω των συσχετίσεων των παραμέτρων του κάθε κόμβου και των αλληλεπιδράσεων μεταξύ των διαφόρων κόμβων στο δίκτυο, η επίδραση που προκαλείται από τις διαφορετικές στρατηγικές επίθεσης στο δίκτυο, εξαρτάται από διάφορους – ενδεχομένως αντίθετους – παράγοντες. Επομένως, προκειμένου να αξιολογηθούν καλύτερα αυτές οι στρατηγικές και για να αποκτηθεί βαθύτερη γνώση όσον αφορά την ευπάθεια του δικτύου, απαιτείται ένα γενικό μέτρο επίθεσης – αξιολόγησης, το οποίο να περιγράφει τη συνολική αντίδραση που προκαλείται από τον επιτιθέμενο στο δίκτυο. Για παράδειγμα, η μη διαθεσιμότητα ενός ποσοστού κόμβων του δικτύου θα προκαλέσει μείωση στη λειτουργία του δικτύου αυτού, καθ' όλη τη διάρκεια ζωής του επιτιθέμενου.

Όπως έχει ήδη επισημανθεί στην προηγούμενη υποενότητα, η διάρκεια ζωής του επιτιθέμενου μπορεί να επηρεάσει σημαντικά την εξέλιξη του δικτύου και να επιδράσει αρνητικά στην απόδοσή του. Εντούτοις, πρέπει να σημειωθεί ότι η διάρκεια ζωής ενός ενεργειακά περιορισμένου κόμβου είναι στενά συνδεδεμένη με το πρότυπο κατανάλωσης της ενέργειάς του, το οποίο με τη σειρά του εξαρτάται άμεσα από την ακτίνα μετάδοσής του. Παρόμοια, το ποσοστό (ή ο αριθμός) των μολυσμένων κόμβων σε κάθε χρονικό στιγμιότυπο είναι ενδεικτικό της ικανότητας μολύνσεως ενός κακόβουλου κόμβου. Σαφώς, όταν ο αριθμός των μολυσμένων κόμβων είναι υψηλός για το μεγαλύτερο μέρος της διάρκειας ζωής του επιτιθέμενου, το δίκτυο υπολειτουργεί για το μεγαλύτερο μέρος και ο επιτιθέμενος από τη δική του οπτική γωνία θεωρείται επιτυχημένος. Επιπλέον, ένας υψηλότερος αριθμός από μολυσμένους κόμβους εντείνει τη βλαπτική ικανότητα ενός επιτιθέμενου.

Λαμβάνοντας υπόψη τις παραπάνω εκτιμήσεις, έχει εισαχθεί ένα νέο μέγεθος απόδοσης, η αποδοτικότητα μολύνσεως (*Infection Efficiency*) I_E, ενός επιτιθέμενου, η οποία ορίζεται ως το ολοκλήρωμα της συνάρτησης χρόνου του αριθμού των μολυσμένων κόμβων του δικτύου. Διαισθητικά, αυτό μπορεί να ερμηνευτεί ως το γινόμενο του αριθμού των μολυσμένων κόμβων σε ένα χρονικό διάστημα επί τη διάρκεια του αντίστοιχου χρονικού διαστήματος και επομένως παρέχει ένα συνδυαστικό μέτρο της απόλυτης τιμής του αριθμού των μολυσμένων κόμβων (π.χ. μέτρο της στιγμιαίας βλάβης του επιτιθέμενου) και του αντίστοιχου χρονικού διαστήματος όπου η σχετική κατάσταση (βλάβη) συντελείται.

Θεωρητικά, η αποδοτικότητα μολύνσεως, ως ολοκλήρωμα χρόνου, μπορεί να πάρει άπειρες τιμές. Όμως, σε όλες τις ρεαλιστικές περιπτώσεις, όπου τόσο ο συνολικός αριθμός των μολυσμένων κόμβων, όσο και η διάρκεια ζωής του επιτιθέμενου παρουσιάζουν κάποια όρια, τα οποία προέρχονται από το μέγεθος του δικτύου και τους ενεργειακούς περιορισμούς, αντίστοιχα, η αποδοτικότητα μολύνσεως μιας στρατηγικής επίθεσης αναμένεται να λαμβάνει πεπερασμένες τιμές. Με άλλα λόγια, $I_E \in \left[0, I_{E_{\max}}\right]$ όπου $I_{E_{\max}}$ είναι η μεγαλύτερη τιμή της αποδοτικότητας μολύνσεως που εξαρτάται από τα N, r, C και E_{init} (τα αρχικά ενεργειακά αποθέματα).

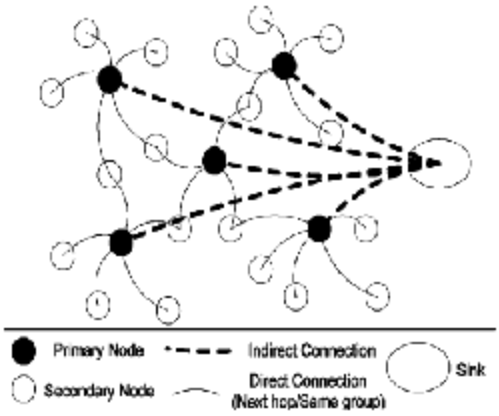
3.5 Μοντέλο συστήματος και αρχιτεκτονική δικτύου αισθητήρων

Στην παρούσα εργασία θεωρούμε ένα παράδειγμα ασύρματου δικτύου αισθητήρων με αρκετούς ετερογενείς χρήστες, όπου κάθε κόμβος μπορεί να έχει διαφορετικές δυνατότητες και να εκτελεί διαφορετικές συναρτήσεις. Για παράδειγμα, μερικοί κόμβοι μπορεί να έχουν μεγαλύτερη χωρητικότητα μπαταριών και ισχυρότερη ικανότητα επεξεργασίας, άλλοι μπορεί να αθροίζουν και να αναμεταδίδουν δεδομένα, ενώ μερικοί

μπορεί μόνο να εκτελούν τη λειτουργία επίβλεψης (*sensing*) χωρίς αναμετάδοση δεδομένων για λογαριασμό άλλων κόμβων.

Όπως έχει ήδη ειπωθεί, ένα ad hoc ή δίκτυο αισθητήρων συνήθως αναπαρίσταται με ένα γράφο δικτύου. Έστω ένας αλγόριθμος ο οποίος συσχετίζει τις μετρήσεις γειτονικών κόμβων – αισθητήρων, με σκοπό να ανιχνεύσει στον αντίστοιχο γράφο του δικτύου τους κόμβους που περιέχουν ανωμαλίες. Προκειμένου να αποκεντρωθεί ο αλγόριθμος ανίχνευσης, το δίκτυο αισθητήρων χωρίστηκε σε ομάδες κόμβων. Η διαίρεση μπορεί να γίνει είτε στατιστικά όταν το δίκτυο αναπτύσσεται αρχικά, είτε κατά περιόδους όταν το δίκτυο ρυθμίζεται δυναμικά. Σε κάθε περίπτωση, θεωρούμε ότι η διαίρεση του δικτύου σε υπο–ομάδες κόμβων βασίζεται σε δοκιμές συσχέτισης μεταξύ των μετρήσεων των κόμβων.

Η εικόνα 2 που ακολουθεί παρουσιάζει την τοπολογία του δικτύου αισθητήρων και την υπό μελέτη αρχιτεκτονική.



Εικόνα 2 Τοπολογία ασύρματου ad hoc δικτύου αισθητήρων και αρχιτεκτονική

Η δημιουργία των ομάδων βασίζεται στις συσχετίσεις των αντίστοιχων ενδείξεων των αισθητήρων.

Σε κάθε ομάδα υποθέτουμε ότι υπάρχει ένας πρωτεύον κόμβος, όπου είναι συνήθως εξοπλισμένος με περισσότερες δυνατότητες επεξεργασίας και ισχύος. Όπως

αναφέρθηκε προηγουμένως, στα δίκτυα αισθητήρων, τα δεδομένα των γειτονικών κόμβων θεωρούνται υψηλής συσχέτισης, δεδομένου ότι τα παρατηρούμενα αντικείμενα στην ίδια γεωγραφική θέση συνήθως εμφανίζουν ισχυρή συσχέτιση. Υπό αυτή την προϋπόθεση, η επεξεργασία και η άθροιση των δεδομένων μπορεί να εκμεταλλευτούν το χωρικό συσχετισμό των τοπικών κόμβων ώστε να πετύχουν εντυπωσιακή μείωση του ποσού των πληροφοριών που μεταδίδονται. Η εκτίμηση των ιδιοτήτων των συσχετισμένων δεδομένων, η οποία βασίζεται στα προηγούμενα και τα τρέχοντα δεδομένα που συλλέγονται, μπορεί να χρησιμοποιηθεί τοπικά για τη βελτιστοποίηση της συμπίεσης δεδομένων και της άθροισης των συνεπακόλουθων φάσεων συλλογής δεδομένων **[42]**.

Το αποτέλεσμα της διαδικασίας ομαδοποίησης είναι οι ομάδες να αποτελούνται από κόμβους με συσχετιζόμενες ενδείξεις. Πρέπει να σημειωθεί εδώ ότι οι διάφορες ομάδες δε χρειάζεται να έχουν αμοιβαίως αποκλειόμενα μέλη. Αντιθέτως, η ύπαρξη διαφόρων κοινών δευτερευόντων κόμβων είναι σε πολλές περιπτώσεις επιθυμητή με σκοπό τη βελτίωση της αποτελεσματικότητας ανίχνευσης.

Κάθε πρωτεύων κόμβος λαμβάνει τις ενδείξεις αισθητήρων από τους κόμβους της ομάδας του και μπορεί να εκτελέσει τοπική ανάλυση σε πραγματικό χρόνο. Γενικά, κάθε κόμβος δικτύου συλλέγει δεδομένα όσον αφορά ένα ή περισσότερα μετρικά (*metrics*) που περιγράφουν τις συγκεκριμένες παραμέτρους που ο κόμβος ελέγχει. Αν τα μετρικά είναι συσχετισμένα, τότε η διαδικασία μπορεί να επεκταθεί λαμβάνοντας υπόψη αυτή τη συσχέτιση. Με σκοπό την καλύτερη μοντελοποίηση και την αναπαράσταση αυτού, δημιουργήθηκε ένα σύνολο από εικονικούς κόμβους για κάθε πραγματικό κόμβο, με κάθε εικονικό κόμβο να αντιστοιχεί σε διαφορετικό μετρικό.

Οι αθροιζόμενες τιμές και πληροφορίες σχετικά με τις ανωμαλίες προωθούνται στην πηγή. Η πηγή γνωρίζοντας την τοπολογία του δικτύου μπορεί να αποφασίσει αν η ανωμαλία εξαπλώνεται ανάμεσα σε πολλαπλούς κόμβους και αν ακολουθεί ένα ή περισσότερα μονοπάτια του δικτύου. Στην περίπτωση μιας κινητής απειλής, η πηγή μπορεί να προσδιορίσει το μονοπάτι του δικτύου ελέγχοντας για ανωμαλίες μεταξύ γειτονικών

ομάδων στο πρόσφατο παρελθόν. Αυτή η διαδικασία μπορεί να παρέχει χρήσιμες πληροφορίες για μεγάλο αριθμό εφαρμογών. Για παράδειγμα, μπορεί να εμφανιστεί ένα ακραίο φυσικό φαινόμενο που να επηρεάζει τις ενδείξεις των αισθητήρων σε μια περιοχή ή μπορεί να υπάρχει ένας ιός ή ένας κακόβουλος κινητός κόμβος που δεσμεύει ή επηρεάζει μεγάλο αριθμό διασκορπισμένων κόμβων. Σε κάθε περίπτωση η ομάδα (–ες) που αναφέρουν μεγάλες αποκλίσεις, δείχνουν το μονοπάτι των ανιχνευμένων ανωμαλιών όπως και την τρέχουσα θέση της πηγής της ανωμαλίας.

3.6 Ανίχνευση ανωμαλιών με τη μέθοδο *M³L*

Το ζητούμενο του αλγόριθμου ανίχνευσης ανωμαλίας είναι να παρέχει μια αποδοτική και αποτελεσματική μέθοδο συγχώνευσης και συνδυασμού δεδομένων ετερογενών ελέγχων που καλύπτουν όλο το δίκτυο. Σκοπός είναι να παραχθεί ένα γενικό πλαίσιο, ικανό να ανιχνεύει ένα εκτεταμένο εύρος κλάσεων ανωμαλιών, όπως αυτά που δημιουργούνται τυχαία από ελαττωματικούς κόμβους ή άλλες που απορρέουν από συντονισμένους έκθετους κόμβους. Στην παρούσα εργασία αυτό επιτυγχάνεται με την εφαρμογή μιας προσέγγισης που βασίζεται στην τεχνική *PCA*.

3.6.1 Γενικά στοιχεία για την τεχνική *PCA*

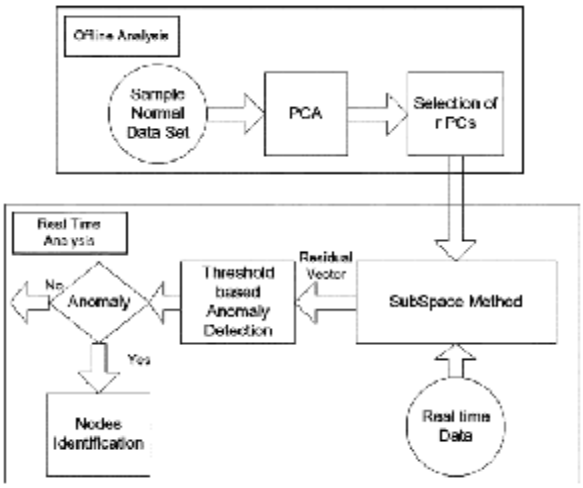
Η μέθοδος μέσω της οποίας πραγματοποιείται η ανίχνευση ανωμαλιών αποτελεί κρίσιμη παράμετρο και κατέχει κεντρικό ρόλο στην παρούσα εργασία. Η τεχνική που έχει επιλεχθεί για την ανίχνευση αυτή, όπως έχει ήδη ειπωθεί, είναι η *PCA*, μία γραμμική τεχνική μείωσης διαστάσεων. Αυτή η τεχνική περιλαμβάνει έναν μαθηματικό αλγόριθμο ο οποίος μετασχηματίζει διάφορες συσχετισμένες μεταβλητές σε έναν, εν γένει μικρότερο, αριθμό ασυσχέτιστων μεταβλητών. Οι τελευταίες μεταβλητές είναι γνωστές στη βιβλιογραφία ως κύριες συνιστώσες (*principal components*).

Η ιδέα πίσω από την *PCA* είναι να βρεθεί ένα διατεταγμένο σύνολο ορθογωνίων διανυσμάτων τα οποία να "συλλαμβάνουν" τις κατευθύνσεις στα δεδομένα με τη μεγαλύτερη απόκλιση (διασπορά). Η τεχνική *PCA* αποδεικνύεται αποτελεσματική μόνο εάν τα αρχικά δεδομένα, όπως στην περίπτωση που εξετάζουμε, παρουσιάζουν κάποιου είδους συσχέτιση. Η *PCA* μειώνει τη διαστατικότητα του αρχικού συνόλου των συσχετισμένων στοιχείων χωρίς να αφαιρεί σημαντικό ποσό πληροφοριών.

Το τελευταίο γίνεται εφικτό μετασχηματίζοντας τους άξονες του πολυδιάστατου διαστήματος των αρχικών δεδομένων με τέτοιον τρόπο ώστε μερικοί άξονες να περιέχουν το μεγαλύτερο μέρος της απόκλισης (διασποράς) και οι περισσότεροι να περιέχουν μικρές αποκλίσεις (διασπορές). Οι άξονες με τη μικρή απόκλιση (διασπορά) μπορούν να αφαιρεθούν χωρίς να επηρεάσουν σημαντικά τη γενική απόκλιση. Με αυτόν τον τρόπο επιτυγχάνεται μείωση των διαστάσεων των αρχικών δεδομένων χωρίς απώλεια σημαντικής πληροφορίας.

3.6.2 Διαδικασία ανίχνευσης ανωμαλίας

Η διαδικασία ανίχνευσης ανωμαλίας μπορεί να χωριστεί σε δύο διαφορετικά μέρη, όπως παρουσιάζεται στην εικόνα 3 που ακολουθεί: τη διαδικασία κατάρτισης του συστήματος (***offline analysis***), κατά την οποία δημιουργείται ένα μοντέλο συμπεριφοράς από δεδομένα που θεωρούνται κανονικά (δηλαδή δεν περιέχουν ανωμαλίες) και τη διαδικασία ανίχνευσης που γίνεται σε πραγματικό χρόνο (***real time analysis***) η οποία συγκρίνει την τρέχουσα συμπεριφορά με τη μοντελοποιημένη, οπότε πραγματοποιείται ανίχνευση τυχών ανωμαλιών.



Εικόνα 3 Αναπαράσταση μεθοδολογίας *PCA* σε υψηλό επίπεδο

Η είσοδος για τη διαδικασία κατάρτισης του συστήματος (***offline analysis***) είναι ο πίνακας συσχέτισης (ο διαγώνιος πίνακας που περιέχει τους συντελεστές συσχέτισης όλων των ελεγχόμενων μέτρων) ενός δείγματος του συνόλου των δεδομένων. Κατά τη διάρκεια αυτής της διαδικασίας, η τεχνική *PCA* μπορεί να εφαρμοστεί σε αυτό το σύνολο δεδομένων. Κατόπιν, επιλέγονται οι πρώτες πιο σημαντικές κύριες συνιστώσες – ***Principal Components*** (***PCs***). Η επιλογή του αριθμού των κύριων συνιστωσών είναι ένα σημαντικό μέρος της ανάλυσης αυτής και θα γίνει ιδιαίτερος λόγος για τη συγκεκριμένη διαδικασία στη συνέχεια. Ο αριθμός των επιλεγμένων *PCs* εξαρτάται από το δίκτυο αισθητήρων και τον αριθμό των εικονικών κόμβων και αναπαριστά τον αριθμό των *PCs* που απαιτούνται για τη μέτρηση του ποσοστού της διασποράς που το σύστημα χρειάζεται για τη μοντελοποίηση της κανονικής (χωρίς ανωμαλίες) κατάστασής του. Η έξοδος της διαδικασίας κατάρτισης, τα *PCs* δηλαδή, χρησιμοποιούνται στη λεγόμενη Μέθοδο Υποχώρων (***Subspace Method***) με την οποία επιτυγχάνεται η ανίχνευση **[43]**. Επειδή αυτή η διαδικασία είναι υπολογιστικά δαπανηρή, μπορεί να εκτελεστεί μόνο όταν υπάρχει μια αξιοσημείωτη αλλαγή σε έναν ή περισσότερους συντελεστές συσχέτισης. Μια εφικτή λύση είναι η χρήση παραθύρου ολίσθησης που περιέχει τις τελευταίες ενδείξεις και η επανεκτίμηση των *PCs* μόνο όταν η απόκλιση σε ένα ή περισσότερους συντελεστές συσχέτισης υπερβαίνει ένα όριο.

Ο στόχος της μεθόδου Υποχώρων (***Subspace Method***) είναι να διαιρέσει τα τρέχοντα δεδομένα σε δύο διαφορετικά διαστήματα: το ένα περιέχει τις ενδείξεις που θεωρούνται κανονικές και μοιάζουν στο μοντελοποιημένο πρότυπο των δεδομένων και το άλλο περιέχει το υπόλοιπο το οποίο δεν περιλαμβάνεται στο πρώτο διάστημα. Γενικά, οι ανωμαλίες τείνουν να καταλήγουν σε τεράστιες μεταβολές (αποκλίσεις) στο υπόλοιπο, επειδή παρουσιάζουν διαφορετικά χαρακτηριστικά. Στη διάρκεια της ανάλυσης πραγματικού χρόνου, το τρέχον διάνυσμα δεδομένων προβάλλεται σε δύο διαφορετικά υποδιαστήματα, με χρήση των *PCs* που εκτιμήθηκαν κατά την *offline* ανάλυση (***Subspace method***). Όταν συμβαίνει μια ανωμαλία, το διάνυσμα υπολοίπου παρουσιάζει μεγάλη απόκλιση (μεταβολή) σε μερικές από τις μεταβλητές του και το σύστημα ανιχνεύει το μονοπάτι που περιέχει την ανωμαλία επιλέγοντας αυτές τις μεταβλητές. Στις επόμενες υποενότητες παρέχουμε μια λεπτομερή περιγραφή για κάθε ένα από τα συστατικά (*components*) που εμπλέκονται στην όλη διαδικασία.

3.6.3 Συνδυασμός μέτρων

Στόχος της τεχνικής *PCA* είναι η μείωση της διάστασης του συνόλου των δεδομένων στα οποία υπάρχει μεγάλος αριθμός αλληλοσχετιζόμενων μεταβλητών και παράλληλα η διατήρηση όσο το δυνατόν περισσότερο της παρουσίας της μεταβολής (απόκλισης) στο σύνολο των δεδομένων **[26]** και **[44]**. Τα μη συσχετισμένα συστατικά (π.χ. *PCs*) εξάγονται από τα ιδιοδιανύσματα του πίνακα συνδιακύμανσης των αρχικών δεδομένων.

Ας υποθέσουμε ότι τα αρχικά δεδομένα είναι ένας πίνακας **x** με διαστάσεις $n \times p$, όπου n οι παρατηρήσεις για καθεμιά από τις p μεταβλητές ($x_1, x_2, \ldots, x_p$) και έστω **S** ένας πίνακας $p \times p$ δειγμάτων συνδιακύμανσης των $x_1, x_2, \ldots, x_p$. Αν (l_1,$\mathbf{e}_1$), (l_2,$\mathbf{e}_2$), …, (l_p,$\mathbf{e}_p$) είναι p (ιδιοτιμές, ιδιοδιανύσματα) ζεύγη του πίνακα **S**, τότε το i–οστο *PC* δίνεται από τη σχέση:

$$z_i = e_i^T (x - x_m) \quad \textbf{(8)}$$

όπου $l_1 \geq l_2 \geq ... \geq l_p \geq 0$, e_i^T είναι το i–οστο μετασχηματισμένο ιδιοδιάνυσμα και $\mathbf{x_m}$ είναι η μέση τιμή του **x**.

Αν k είναι ο αριθμός των κόμβων και l ο αριθμός των διαφορετικών μετρικών (*metrics*) που συλλέγονται σε κάθε κόμβο, τότε οι εικονικοί κόμβοι που αναπαριστούν τις μεταβλητές του συνόλου δεδομένων μας **x** είναι $p = 1 \times k$. Κάθε γραμμή του πίνακα είναι μια παρατήρηση του ελεγχόμενου δικτύου αισθητήρων σε ένα συγκεκριμένο χρονικό διάστημα (*bin*), ενώ κάθε στήλη περιέχει τη χρονική ακολουθία για κάθε εικονικό κόμβο. Η έξοδος της *PCA* είναι ένας $p \times p$ πίνακας. Το επόμενο βήμα στην όλη διαδικασία, είναι η επιλογή του αριθμού r ($r < p$) των *PCs* που απαιτούνται για να συλλάβουν το ποσοστό διασποράς που το σύστημα χρειάζεται για τη μοντελοποίηση κανονικών δεδομένων.

Παραδοσιακά η *PCA* βασίζεται στην εξαγωγή των ιδιοδιανυσμάτων από τον πίνακα συνδιακύμανσης ενός δείγματος συνόλου δεδομένων. Η μέθοδος πετυχαίνει να ξεπεράσει αυτό το πρόβλημα των διαφορετικών μέτρων χρησιμοποιώντας τον πίνακα συσχέτισης αντί του πίνακα συνδιακύμανσης. Το πρόβλημα της χρήσης της *PCA* η οποία βασίζεται σε πίνακες συνδιακύμανσης είναι ότι τα *PCs* είναι ευαίσθητα σε μεγάλες διαφορές μεταξύ μεταβολών ενός στοιχείου του **x**. Έτσι, με σκοπό την ελάττωση του προβλήματος της εξάρτησης από τη διάσταση της *PCA*, χρησιμοποιούμε κανονικοποίηση των εικονικών κόμβων. Κάθε *PC* μπορεί επίσης να οριστεί ως:

$$z_i = es_i^T x^* \quad \textbf{(9)}$$

όπου es_i^T είναι το i–οστο μετασχηματισμένο ιδιοδιάνυσμα του πίνακα συσχετίσεων και $\mathbf{x}^*$ αποτελείται από κανονικοποιημένες μεταβλητές. Ο στόχος της υιοθέτησης μιας τέτοιας προσέγγισης είναι η εύρεση των *PCs* της κανονικοποιημένης έκδοσης $\mathbf{x}^*$ του **x**, όπου $\mathbf{x}^*$

έχει j στοιχεία $x_j / s_{jj}^{1/2}$, $j = 1, 2, \mathbf{K}, p, x_j$ είναι το j–οστο στοιχείο του **x** και s_{jj} είναι η διακύμανση του x_j. Επομένως, ο πίνακας συνδιακύμανσης του $\mathbf{x}^*$ είναι ο πίνακας συσχέτισης του **x** και τα *PCs* του $\mathbf{x}^*$ μπορεί να εξαχθούν από την έκφραση (8).

Μια βασική και σημαντική εργασία στη προσέγγιση ανίχνευσης ανωμαλιών που βασίζεται στη *PCA* είναι η επιλογή του αριθμού των *PCs* που χρειάζονται για τη σύλληψη του επιθυμητού ποσοστού διασποράς. Στην περίπτωσή μας, χρειάζεται να προσδιορίσουμε την καταλληλότερη τιμή του αριθμού r των *PCs* που απαιτούνται για την εφαρμογή της μεθόδου Υποχώρων (***Subspace method***). Ένα από τα πιο κοινά κριτήρια για την επιλογή του r είναι το αθροιστικό (***cumulative***) ποσοστό της συνολικής μεταβολής **[26]**. Ένας εναλλακτικός κανόνας, που χρησιμοποιεί τους πίνακες συσχέτισης, όπως στην περίπτωσή μας, βασίζεται στο μέγεθος των διασπορών των *PCs*. Η βασική ιδέα πίσω από αυτόν τον κανόνα είναι ότι αν όλα τα στοιχεία του **x** είναι ανεξάρτητα, τότε τα *PCs* είναι τα ίδια με τα αρχικά δεδομένα και πρέπει όλα να έχουν διασπορές ίσες με 1. Έτσι, κάθε *PC* με διασπορά μικρότερη από 1 περιέχει λιγότερη πληροφορία από οποιοδήποτε αρχική μεταβλητή και ως αποτέλεσμα δεν αξίζει διατήρησης. Για παράδειγμα, αν το σύνολο δεδομένων περιέχει μια ομάδα μεταβλητών με μεγάλη ενδο–ομαδική συσχέτιση, τότε θα υπάρχει μόνο ένα *PC* που να σχετίζεται (συνδέεται) με αυτή την ομάδα. Αυτό το κριτήριο, που στην απλούστερη μορφή του καλείται κανόνας ***Kaizer*** **[26]**, έχει επίσης επιλεγεί για να χρησιμοποιηθεί στο περιβάλλον μας, ενώ εκτεταμένα πειράματα έχουν δείξει την καταλληλότητά και εφαρμοσιμότητά του.

3.6.4 Ανίχνευση ανωμαλιών μέσω υποχώρων

Αφότου πάρουμε τα *PCs* και καθοριστεί ο αριθμός αυτών που (τελικά) θα κρατηθούν, ένα κανονικοποιημένο διάνυσμα δείγμα μπορεί να αναλυθεί σε δύο τμήματα, ως ακολούθως:

$$y = y_{norm} + y_{res} \quad \textbf{(10)}$$

ώστε το y_{norm} να αντιστοιχεί στο μοντελοποιημένο (κανονικό) δεδομένο και το y_{res} στο υπόλοιπο. Σχηματίζουμε το y_{norm} προβάλλοντας το y στον κανονικό υποχώρο S και σχηματίζουμε το y_{res} προβάλλοντας το y στον ανώμαλο (μη κανονικό) υποχώρο $\tilde{S}$. Για να το πετύχουμε αυτό, διαρθρώνουμε το σύνολο των *PCs* που αντιστοιχεί στον κανονικό υποχώρο $(v_1, v_2, \mathbf{K}, v_r)$ σαν στήλη του πίνακα **P** μεγέθους $p \times r$ όπου r δηλώνει τον αριθμό των κανονικών αξόνων. Ακολουθώντας αυτή την προσέγγιση τα y_{norm} και y_{res} μπορούν να ξαναγραφτούν ως ακολούθως:

$$y_{norm} = PP^T y = Cy \quad \text{και} \quad y_{res} = (I - PP^T)y = \tilde{C}y \quad \textbf{(11)}$$

όπου ο πίνακας $C = PP^T$ αναπαριστά τον γραμμικό τελεστή που εκτελεί την προβολή στον κανονικό υποχώρο S και $\tilde{C}$ ομοίως προβάλει στον ανώμαλο (μη κανονικό) υποχώρο $\tilde{S}$. Έτσι, το y_{norm} περιέχει τα μοντελοποιημένα (κανονικά) δεδομένα ενώ το y_{res} περιέχει το υπόλοιπο. Γενικά, η εμφάνιση μια ανωμαλίας τείνει να επιφέρει μεγάλη αλλαγή στο y_{res}. Μια αλλαγή στη μεταβλητή συσχέτισης θα αυξήσει την προβολή του y στον υποχώρο $\tilde{S}$. Σε ένα τέτοιο πλαίσιο ένα στατιστικό μέγεθος συχνά χρησιμοποιούμενο για ανίχνευση ανώμαλων καταστάσεων είναι το τετραγωνικό σφάλμα πρόβλεψης (*Squared Prediction Error*, *SPE*) **[45]**:

$$SPE = \left\| y_{res} \right\|^2 = \left\| \tilde{C} \right\|^2 \quad \textbf{(12)}$$

Όταν εμφανίζεται μια ανωμαλία, το *SPE* υπερβαίνει τα προκαθορισμένα όρια και το σύστημα ανιχνεύει το σύνολο των κόμβων που περιέχουν την ανωμαλία, με την επιλογή των μεταβλητών που συνεισφέρουν περισσότερο στην αλλαγή του *SPE*. Αυτό μπορεί να

γίνει κατανοητό, αν στο διάνυσμα του υπόλοιπου επιλεγούν οι εικονικοί κόμβοι που η απόκλισή (μεταβολή) τους είναι σημαντικά μεγαλύτερη από την αντίστοιχη που εμφανίζουν υπό κανονικές συνθήκες.

3.7 Ομαδοποίηση

Όπως αναφέρθηκε νωρίτερα, η διαίρεση του δικτύου σε ομάδες βασίζεται στη συσχέτιση η οποία εμφανίζεται στις ενδείξεις των γειτονικών αισθητήρων. Αυτό μπορεί να οδηγήσει κόμβους που είναι πολύ κοντά ο ένας με τον άλλο να ανήκουν σε διαφορετικές ομάδες, εάν οι ενδείξεις τους είναι ασυσχέτιστες. Σε όλη την εργασία, θεωρούμε ως γείτονες τους αισθητήρες που απέχουν ένα βήμα μεταξύ τους (ο ένας από τον άλλον). Ο πρωτεύων κόμβος δημιουργεί μια ομάδα ρωτώντας όλους τους γείτονές του για τις πρόσφατες ενδείξεις τους. Όμως, ανάλογα με την πυκνότητα του δικτύου αισθητήρων, εάν το ποσοστό των πρωτευόντων κόμβων στο δίκτυο δεν είναι αρκετά μεγάλο για να καλύψει όλο το δίκτυο και μετά από τη διαδικασία αρχικοποίησης υπάρχουν κόμβοι που δεν ανήκουν σε κάποια ομάδα, το όριο εγγύτητας μπορεί να αυξηθεί σε δύο ή περισσότερα βήματα. Στη συνέχεια, για πειραματικούς κυρίως λόγους, υποθέτουμε μια απλή μέθοδο για τη σύνθεση των διαφόρων ομάδων του δικτύου αισθητήρων. Όμως, στη βιβλιογραφία, έχουν παρουσιαστεί και αξιολογηθεί αρκετές τεχνικές για αποδοτική ομαδοποίηση κόμβων σε ad hoc δίκτυα και δίκτυα αισθητήρων. Μια αναλυτική μέθοδος εκτίμησης της συσχέτισης μεταξύ δύο συνόλων δεδομένων είναι ο συντελεστής συσχέτισης $R_{X,Y}$ που δίνεται από τη σχέση:

$$R_{X,Y} = \frac{Cov(X,Y)}{S_X \cdot S_Y} \quad \textbf{(13)}$$

όπου $Cov(X,Y)$ δηλώνει τη συνδιακύμανση μεταξύ των συνόλων δεδομένων X και Y, ενώ S_X και S_Y είναι οι δειγματικές διασπορές των X και Y, αντίστοιχα. Μεταβλητές με

συντελεστή συσχέτισης κοντά στο 1 μεταβάλλονται προς την ίδια κατεύθυνση, ενώ μεταβλητές με συσχέτιση κοντά στο −1 μεταβάλλονται προς αντίθετες κατευθύνσεις. Κάθε πρωτεύων κόμβος μπορεί να χρησιμοποιήσει ένα προκαθορισμένο όριο για το $R_{X,Y}$ ώστε να επιλέξει τους κόμβους της ομάδας του.

3.7.1 Στατική και δυναμική ομαδοποίηση

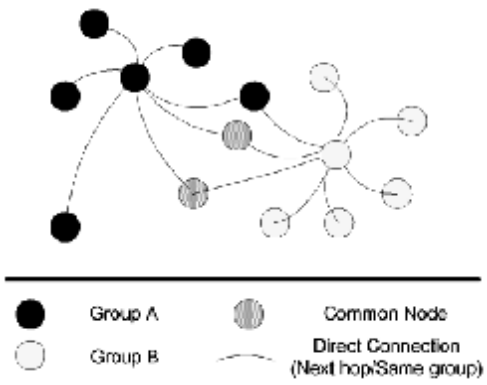
Γενικά, μια μέθοδος ανίχνευσης ανωμαλιών εξαρτάται από την ύπαρξη ενός καλού κανονικού δείγματος δεδομένων, δηλαδή ενός δείγματος δεδομένων χωρίς ανωμαλίες. Στην περίπτωσή μας επίσης, ένα ακριβές δείγμα δεδομένων θα παράγει *PCs* που αντιστοιχούν σε πραγματικά πρότυπα συσχέτισης μεταξύ των αισθητήρων και επομένως δημιουργούν ένα ακριβές μοντέλο. Θα πρέπει να σημειωθεί ότι η συνολικά προτεινόμενη προσέγγισή μας δεν απαιτεί κάποια ειδική μέθοδο κατάρτισης (*training*) δεδομένων. Χρησιμοποιεί τους συντελεστές συσχέτισης μεταξύ του πρωτεύοντος κόμβου και των γειτόνων του αρχικά για τη δημιουργία ομάδων και εν συνεχεία για τη μοντελοποίηση των κανονικών ενδείξεων των αισθητήρων μέσα σε κάθε ομάδα.

Εφόσον οι λειτουργίες της διάταξης των ομάδων και του υπολογισμού των *PCs* μπορεί να είναι υπολογιστικά δαπανηρές, μπορούν να εκτελεστούν μόνο κατά τη φάση αρχικοποίησης του δικτύου και επομένως, οι αντίστοιχες διατάξεις των ομάδων να παραμείνουν στατικές σε όλη την υπόλοιπη λειτουργία του δικτύου. Όμως, σε πολλές περιπτώσεις αφού τα δεδομένα που συλλέγονται εξαρτώνται σε μεγάλο βαθμό από το περιβάλλον, οι αλληλοσυσχετίσεις τους μπορεί να αλλάξουν αν υπάρχει κάποια αλλαγή στο περιβάλλον του αισθητήρα. Σε αυτή την περίπτωση, το σύστημα μπορεί περιοδικά να ελέγχει τους συντελεστές συσχέτισης μεταξύ κάθε αρχικού κόμβου και των άλλων μελών της ομάδας, προκειμένου να καθοριστεί εάν υπάρχουν σημαντικές αλλαγές στο περιβάλλον και κατά συνέπεια να απαιτείται η δυναμική επαναδιάταξη των δημιουργημένων ομάδων. Κάθε πρωτεύων κόμβος μπορεί να αποθηκεύσει ενδείξεις

αισθητήρων για χρονικά κινούμενο παράθυρο και να υπολογίσει το αντίστοιχο $R_{X,Y}$ ανά κόμβο. Αν η τιμή του $R_{X,Y}$ για έναν κόμβο είναι μικρότερη από ένα προκαθορισμένο όριο, τότε ο κόμβος πρέπει να αφαιρεθεί από την ομάδα και να ζητήσει να ενσωματωθεί σε άλλη, αν είναι διαθέσιμη. Επομένως, οι αλλαγές στο περιβάλλον που καταλήγουν σε τροποποιήσεις των συσχετίσεων μεταξύ των αισθητήρων μπορεί να οδηγήσουν και σε αλλαγές στη σύνθεση της ομάδας. Η αξιολόγηση της επίδρασης των τοπολογικών αλλαγών στην απόδοση της προτεινόμενης λύσης αποτελεί μέρος τρέχουσας και μελλοντικής έρευνας.

3.7.2 Επικαλυπτόμενοι κόμβοι

Όπως περιγράφηκε πριν, η δημιουργία μιας ομάδας βασίζεται στον έλεγχο των συντελεστών συσχέτισης μεταξύ του πρωτεύοντος κόμβου και των γειτονικών του αισθητήρων και κατά συνέπεια δεν υπάρχει ανάγκη για τις διαμορφούμενες ομάδες να έχουν αμοιβαίως αποκλειόμενα μέλη. Αντιθέτως, δύο ομάδες μπορεί να μοιράζονται έναν ή περισσότερους κόμβους αν η συσχέτιση αυτών των κόμβων με τον πρωτεύοντα κόμβο κάθε ομάδας είναι εξίσου υψηλή και εφόσον ο αριθμός των κοινών κόμβων δεν υπερβαίνει ένα προκαθορισμένο όριο. Η εικόνα 4 παρουσιάζει ένα παράδειγμα δύο διαφορετικών ομάδων που μοιράζονται δύο κοινούς κόμβους.



Εικόνα 4 Δίκτυο αισθητήρων με επικαλυπτόμενους κόμβους μεταξύ ομάδων

Επιτρέποντας την ύπαρξη επικαλυπτόμενων κόμβων μεταξύ γειτονικών ομάδων μας παρέχεται ένα πρόσθετο επίπεδο υπολογιστικής ευφυΐας στην αντίστοιχη διαδικασία λήψης αποφάσεων, συνδυάζοντας τις μεμονωμένες αποφάσεις των αντίστοιχων ομάδων. Ανάλογα με τη μεθοδολογία που χρησιμοποιείται για να συγχωνεύσει τα αποτελέσματα των ομάδων που αναφέρουν ανωμαλίες στους επικαλυπτόμενους κόμβους, μπορούμε να επιτύχουμε αύξηση του ρυθμού ή μείωση της εμφάνισης των εσφαλμένων ανιχνεύσεων (*false positives*). Πιο συγκεκριμένα, χρησιμοποιώντας την ένωση των αποτελεσμάτων που λαμβάνονται από δύο τέτοιες ομάδες, υπάρχει μεγαλύτερη πιθανότητα για ανίχνευση ανωμαλίας στις ενδείξεις των επικαλυπτόμενων κόμβων. Από την άλλη, χρησιμοποιώντας την τομή των αντίστοιχων δεδομένων μειώνουμε την πιθανότητα εμφάνισης ενός σφάλματος ανίχνευσης.

3.8 Σύγκλιση μοντέλων συστήματος

Στις προηγούμενες ενότητες, του παρόντος κεφαλαίου, περιγράφηκαν δύο ξεχωριστά μοντέλα συστήματος. Αρχικά καταγράφηκε το γενικό μοντέλο συστήματος ενός ad hoc δικτύου για μια πεπερασμένη περιοχή δύο διαστάσεων. Αναφέρθηκαν εκτενώς τα χαρακτηριστικά και οι ιδιότητες που πρέπει να φέρουν τα στοιχεία που δομούν ένα δίκτυο τέτοιων προδιαγραφών. Βασική επιδίωξη για την ανάπτυξη του μοντέλου αυτού ήταν η μελέτη της αντίδρασης του δικτύου σε μια επίθεση. Η επίθεση αυτή θεωρήθηκε συγκεκριμένου τύπου, ότι δηλαδή, προέρχεται από έναν ελεύθερα κινούμενο κόμβο στην περιοχή του δικτύου. Αυτός ο μοναδικός κακόβουλος εισβολέας, μπορούσε να μολύνει τους κόμβους που απαρτίζουν το δίκτυο. Όσοι από τους κόμβους του δικτύου μολύνονταν, βάσει του ίδιου μοντέλου, μπορούσαν με τη σειρά τους να μολύνουν γειτονικούς κόμβους. Έτσι, η προαναφερθείσα αντίδραση του δικτύου αναφέρεται από τη μία στην μόλυνση των κόμβων από τον επιτιθέμενο κόμβο και από την άλλη στη μόλυνση των κόμβων από γειτονικούς μολυσμένους κόμβους. Βάσει του εν λόγω μοντέλου εξετάστηκε διεξοδικά η

διαδικασία διάδοσης μιας επίθεσης σε ένα ad hoc δίκτυο. Τόσο η επίθεση όσο και το δίκτυο στο οποίο αυτή γίνεται έχουν τα χαρακτηριστικά που αναφέρθηκαν στις οικείες υποενότητες του κεφαλαίου αυτού.

Από την άλλη το δεύτερο μοντέλο συστήματος που αναπτύχθηκε στο παρόν κεφάλαιο έχει να κάνει με την ανίχνευση ανωμαλίας που εμφανίζεται σε δίκτυα αισθητήρων. Τα χαρακτηριστικά και η αρχιτεκτονική ενός δικτύου αισθητήρων καταγράφηκαν προηγούμενα σε αυτό το κεφάλαιο. Μάλιστα το μοντέλο του δικτύου αισθητήρων είναι απόλυτα συμβατό με αυτό του ad hoc δικτύου. Συνεπώς, τα δύο αυτά ξεχωριστά μοντέλα, μπορούν να συγχωνευτούν σε ένα με σκοπό την ανίχνευση ανωμαλίας η οποία θα φέρει τα χαρακτηριστικά της μόλυνσης που θεωρήθηκε στο μοντέλο του ad hoc δικτύου.

Ουσιαστικά θεωρείται από τη μία ένα καινούριο δίκτυο που κληρονομεί τα χαρακτηριστικά των δύο προαναφερθέντων δικτύων, του ad hoc και του δικτύου αισθητήρων και από την άλλη ένα μοντέλο επίθεσης που με τη σειρά του κληρονομεί τα χαρακτηριστικά του προαναφερθέντος μοντέλου επίθεσης. Η επίθεση, πλέον, γίνεται σε ένα αναπτυσσόμενο ad hoc δίκτυο αισθητήρων διαθέτοντας όλα τα χαρακτηριστικά μόλυνσης. Η επίθεση μεταφράζεται σε νέους όρους που τίθενται από το συγχωνευμένο ad hoc δίκτυο αισθητήρων. Αποτελεί τη μόλυνση μέσω της οποίας εισάγονται ανωμαλίες στις ενδείξεις των κόμβων αισθητήρων. Δηλαδή, μία μόλυνση, αποτέλεσμα της επίθεσης του κινούμενου κόμβου, έχει ως συνέπεια τη μεταβολή στις τιμές που πραγματικά μετρούν οι αισθητήρες. Αυτή η μόλυνση–ανωμαλία μπορεί να διαδοθεί μεταξύ των κόμβων στο καινούριο δίκτυο. Τα χαρακτηριστικά που ακολουθεί μια τέτοια διάδοση είναι ίδια με αυτά που ίσχυαν για το ad hoc δίκτυο αρχικά. Στόχος της παρούσας εργασίας είναι η ανάπτυξη ενός αλγόριθμου με τον οποίο θα ανιχνεύονται ανωμαλίες τέτοιων χαρακτηριστικών και θα γίνεται εντοπισμός του επιτιθέμενου. Ο αλγόριθμος αυτός υλοποιήθηκε βάσει των καινούριων δεδομένων που προέκυψαν μετά τη συγχώνευση των δύο προαναφερθέντων μοντέλων και αναπτύσσεται στη συνέχεια.

3.9 Κατηγοριοποίηση ανωμαλιών

Όπως ειπώθηκε προηγουμένως, οι κόμβοι αισθητήρες μπορεί να εμφανίσουν αποκλίσεις στις πραγματικά μετρούμενες τιμές τους λόγω ανωμαλιών. Υπάρχουν διάφορα είδη ανωμαλιών–μολύνσεων που μπορεί να εμφανίζονται σε κάθε χρονικό διάστημα σε μια ομάδα κόμβων. Η κατηγοριοποίηση αυτών των ανωμαλιών ποικίλει, ανάλογα με τα κριτήρια που λαμβάνονται υπόψη. Στην περίπτωση που ως κριτήριο λαμβάνεται ο αριθμός των ανωμαλιών μπορεί να εμφανιστεί μια απλή (μοναδική) ή πολλαπλή ανωμαλία (διπλή, τριπλή, κτλ.). Για παράδειγμα, σε ένα χρονικό διάστημα μπορεί σε κάποια ομάδα κόμβων να εμφανιστεί μόνο ένας κόμβος με ανωμαλία (απλή ανωμαλία) ή δύο κόμβοι με ανωμαλία (διπλή ανωμαλία). Απ’ την άλλη, όταν το κριτήριο που λαμβάνεται υπόψη για την ανωμαλία είναι το πλάτος της, δηλαδή, η τιμή της ανωμαλίας, έχουμε τις περιπτώσεις της ανιχνεύσιμης και της μη–ανιχνεύσιμης ανωμαλίας. Πέρα από τις ανωμαλίες, οι οποίες μπορούν να ανιχνευτούν και αποτελούν αντικείμενο της παρούσας εργασίας, που εξετάζονται εκτενώς στα επόμενα, υπάρχουν και ανωμαλίες μη ανιχνεύσιμες. Το γεγονός αυτό οφείλεται στο ότι η τιμή της ένδειξης ενός κόμβου δεν μεταβάλλεται – ή μεταβάλλεται ελάχιστα – από την εισερχόμενη, σε αυτόν, ανωμαλία.

3.10 Μέθοδος – Αλγόριθμος ανίχνευσης ανωμαλιών

Αρχικά πρέπει να επισημανθεί ότι η επιλογή του κατάλληλου αριθμού κύριων συνιστωσών αποτελεί πολύ σημαντικό μέρος σε κάθε μέθοδο που χρησιμοποιεί την τεχνική *PCA*. Στη βιβλιογραφία υπάρχουν γενικοί κανόνες οι οποίοι όμως είναι διαισθητικοί και βασίζονται κυρίως στην εμπειρία και τον πειραματισμό αυτών που τους χρησιμοποιούν. Περαιτέρω ανάλυση για την επιλογή της κατάλληλης τιμής των κύριων συνιστωσών γίνεται στο επόμενο κεφάλαιο, κατά την περιγραφή του μοντέλου προσομοίωσης.

Πριν το στάδιο εύρεσης των κύριων συνιστωσών εφαρμόζουμε κατάλληλη κανονικοποίηση τόσο στα δεδομένα κατάρτισης όσο και στα πραγματικά δεδομένα. Υπενθυμίζουμε ότι τα δεδομένα κατάρτισης δεν έχουν ανωμαλίες σε αντίθεση με τα πραγματικά δεδομένα στα οποία μπορούν να εμφανιστούν ανωμαλίες. Για την πραγματοποίηση της κανονικοποίησης αρχικά αφαιρούνται από τα πραγματικά δεδομένα οι τιμές που περιέχουν ανωμαλίες. Κατόπιν, από τις τιμές των πραγματικών δεδομένων που διατηρήθηκαν και δεν περιέχουν ανωμαλίες και τις τιμές των δεδομένων κατάρτισης αφαιρείται η μέση τιμή (m) και γίνεται διαίρεση με την τυπική απόκλιση (s)*. Τα κανονικοποιημένα δεδομένα που προκύπτουν έχουν μηδενική μέση τιμή και τυπική απόκλιση ίση με μονάδα, ενώ, ταυτόχρονα, οι τιμές τους απεξαρτόνται από μονάδες μέτρησης. Στη συνέχεια της μεθόδου χρησιμοποιούμε τις κανονικοποιημένες τιμές αυτών των δεδομένων. Στόχος μας είναι η ανάπτυξη ενός αλγόριθμου ο οποίος με χρήση της τεχνικής *PCA* να ανιχνεύει τις ανωμαλίες που τυχαία εμφανίζονται στα πραγματικά δεδομένα. Βέβαια τα αρχικά, μη κανονικοποιημένα, δεδομένα χρησιμοποιούνται και αποτελούν ένα επιπλέον κριτήριο στην ανίχνευση ανωμαλιών.

Η τεχνική *PCA*, όπως έχει ειπωθεί νωρίτερα, αποτελεί ισχυρό εργαλείο για την ανίχνευση ανωμαλιών. Πράγματι, όταν σε κάποιο χρονικό στιγμιότυπο προκύψει μια ανωμαλία, τότε το τετραγωνικό σφάλμα πρόβλεψης (*SPE*) υπερβαίνει ένα προκαθορισμένο κατώφλι. Οι συνιστώσες που συνεισφέρουν περισσότερο στην αλλαγή του *SPE* είναι αυτές που συνήθως περιέχουν ανωμαλία. Πρακτικά αυτό σημαίνει ότι στο διάνυσμα του υπόλοιπου, η απόκλιση (μεταβολή) αυτών των κόμβων είναι σημαντικά μεγαλύτερη από την αντίστοιχη που εμφανίζουν υπό κανονικές συνθήκες.

Το *SPE* είναι το κρίσιμο μέγεθος το οποίο μας βοηθά, αφενός να αποφασίσουμε αν σε κάποια χρονική στιγμή εμφανίζεται ανωμαλία και αφετέρου να ανιχνεύσουμε σε ποιους ακριβώς κόμβους εμφανίζεται ανωμαλία. Στις περισσότερες περιπτώσεις, οι κόμβοι που

* τα μεγέθη μέση τιμή (μ) και τυπική απόκλιση (σ), στο σημείο αυτό, αναφέρονται ξεχωριστά στις παρατηρούμενες τιμές κάθε μετρούμενου μεγέθους και δεν αφορούν τη μέση τιμή και την τυπική απόκλιση που εμφανίζει το σύνολο των μετρούμενων μεγεθών σε τυχαίο χρονικό στιγμιότυπο.

περιέχουν ανωμαλία συνεισφέρουν σε πολύ μεγάλο βαθμό στη μεταβολή του *SPE*. Αντίθετα, οι κόμβοι χωρίς ανωμαλία δεν παίζουν σημαντικό ρόλο στη διαμόρφωση της τιμής του *SPE*.

Η εξίσωση που δίνει το μέτρο του *SPE* έχει δοθεί στην υποενότητα 3.6.4 (*βλ*. εξ. 12). Χωρίς βλάβη της γενικότητας υποθέτουμε ότι η τιμή του *SPE* διαμορφώνεται από *n* αριθμό συνιστωσών, δηλαδή το διάνυσμα y_{res} αποτελείται από *n* διαφορετικές συνιστώσες. Τότε η εξίσωση για το *SPE* μπορεί γενικά να γραφεί ως εξής :

$$SPE = \|y_{res}\| = \sqrt{y_1^2 + y_2^2 + \mathbf{L} + y_n^2} \Leftrightarrow SPE^2 = y_1^2 + y_2^2 + \mathbf{L} + y_n^2 \quad \textbf{(14)}$$

Η ποσοστιαία συνεισφορά μιας τυχαίας συνιστώσας y_k στη διαμόρφωση του *SPE* θα γίνεται από τη σχέση :

$$P_{contribution} = \frac{y_k^2}{SPE^2} \cdot 100\%, \text{ όπου } 1 \le \mathrm{k} \le \mathrm{n} \quad \textbf{(15)}$$

Το ποσοστό συνεισφοράς των κόμβων με ανωμαλία στη διαμόρφωση της τιμής του *SPE* καθώς και η ίδια η τιμή του *SPE* σε μια ανωμαλία δεν είναι προκαθορισμένα. Αντιθέτως διαμορφώνονται διαισθητικά βάσει πειραματικής πρακτικής. Τιμές κατωφλίων για τα συγκεκριμένα μεγέθη δίνονται στο επόμενο κεφάλαιο.

3.10.1 Συγκριτική μέθοδος ανίχνευσης ανωμαλιών

Επικουρικά γίνεται χρήση των μη κανονικοποιημένων δεδομένων (κατάρτισης και πραγματικών) με σκοπό να επιτευχθεί ένα κατώτατο όριο στο συνολικό ποσοστό ανίχνευσης, άρα αποδοτικότητας, του αλγόριθμου που εφαρμόζουμε. Για κάθε κόμβο ξεχωριστά θεωρούμε ένα εύρος τιμών το οποίο οριοθετεί αν μια μέτρηση του κόμβου αυτού περιέχει ή όχι ανωμαλία. Αυτό το πεδίο τιμών εξάγεται από τα δεδομένα κατάρτισης του κάθε κόμβου. Μέσω των δεδομένων κατάρτισης του κάθε κόμβου λαμβάνονται η

μέση τιμή και η τυπική απόκλιση. Τελικά, το πεδίο τιμών διαμορφώνεται ως το σύνολο τιμών γύρω από τη μέση τιμή με εύρος ένα ποσοστό της τυπικής απόκλισης. Σημειώνεται ότι αυτή η μέθοδος συνεισφέρει μόνο στην πιθανότητα ανίχνευσης μιας ανωμαλίας και όχι στην πιθανότητα λανθασμένης ανίχνευσης.

3.11 Μεγέθη μέτρησης αποδοτικότητας αλγόριθμου ανίχνευσης ανωμαλιών

Για την αποτίμηση της αποτελεσματικότητας του αλγόριθμου ανίχνευσης ανωμαλιών της παρούσας εργασίας χρησιμοποιήθηκαν δύο μεγέθη απόδοσης: η πιθανότητα ανίχνευσης P_d και η πιθανότητα εσφαλμένης ανίχνευσης P_f. Εδώ, ως πιθανότητα ανίχνευσης, ορίζεται η πιθανότητα να ανιχνεύεται και να αναγνωρίζεται η ανωμαλία στα δεδομένα, ενώ ως πιθανότητα εσφαλμένης ανίχνευσης, ορίζεται η πιθανότητα να κατηγοριοποιούνται ως ανωμαλίες δεδομένα χωρίς ανωμαλία.

Επειδή μας ενδιαφέρει τα παραπάνω μεγέθη να χρησιμοποιηθούν για τη μέτρηση της συνολικής απόδοσης του αλγόριθμου, λαμβάνουμε το μέσο όρο αυτών των μεγεθών που αναφέρονται στο συνολικό διάστημα της προσομοίωσης και όχι τα αντίστοιχα στιγμιαία μεγέθη που αναφέρονται σε κάθε χρονικό στιγμιότυπο της προσομοίωσης.

Η πιθανότητα ανίχνευσης ανωμαλίας P_d καθώς και η πιθανότητα εσφαλμένης ανίχνευσης P_f αποτελούν τα κριτήρια για την αποτίμηση της αποτελεσματικότητας του αλγόριθμου που εφαρμόζουμε στην παρούσα εργασία. Όπως έχει ειπωθεί στο Κεφάλαιο 3, μας ενδιαφέρει η αποδοτικότητα του αλγόριθμου για το σύνολο των προσομοιώσεων. Για αυτό το λόγο στο πειραματικό μέρος λάβαμε τα μέσα μεγέθη των P_d και P_f τα οποία προέκυψαν ως ο μέσος όρος των αντίστοιχων μεγεθών που εμφανίζονται σε κάθε χρονικό στιγμιότυπο (στιγμιαία μεγέθη). Ο ορισμός των στιγμιαίων μεγεθών είναι αντιστοίχως:

$$P_d = \frac{\text{πλήθος ανιχνευθέντων ανωμαλιών}}{\text{πλήθος μολυσμένων κόμβων}}$$

$$P_f = \frac{\text{πλήθος λανθασμένων ανιχνεύσεων}}{\text{πλήθος μολυσμένων κόμβων}} \quad \textbf{(16)}$$

Ο αλγόριθμος που προτείνεται στην παρούσα εργασία θεωρείται αποτελεσματικός όταν επιτυγχάνει υψηλές τιμές μέσου P_d και ταυτόχρονα χαμηλές τιμές μέσου P_f. Αυτό σημαίνει ότι ο αλγόριθμος μπορεί να ανιχνεύει σωστά και με σχετικά μικρό σφάλμα ανωμαλίες που προκύπτουν κατά την εξέλιξη του πειράματος. Βέβαια, όπως θα διαφανεί και από τα πειραματικά αποτελέσματα του επόμενου κεφαλαίου, πρέπει να υπάρχει συμβιβασμός μεταξύ της τιμής των δύο αυτών μεγεθών. Μας ενδιαφέρει να έχουμε την υψηλότερη τιμή του P_d με παράλληλη διατήρηση σε όσο το δυνατόν πιο ανεκτά (χαμηλά) όρια της τιμής του P_f.

Επιπλέον, πέρα από την πιθανότητα ανίχνευσης ανωμαλίας P_d, εισήχθησαν δύο νέα μεγέθη, το μέσο ποσοστό μολυσμένων κόμβων P_{in} και το μέσο ποσοστό μολυσμένων κόμβων του *group* που βρίσκεται ο επιτιθέμενος P_{ingat}. Το P_{in} σχετίζεται με το σύνολο των κόμβων του δικτύου και δίνει μια γενική εικόνα της μόλυνσης του εκάστοτε πειράματος, ενώ το P_{ingat} αναφέρεται ειδικότερα στους κόμβους του *group* στο οποίο βρίσκεται ο επιτιθέμενος σε κάθε χρονικό στιγμιότυπο του πειράματος. Οι σχέσεις των στιγμιαίων μεγεθών P_{in} και P_{ingat} δίνονται ως εξής:

$$P_{in} = \frac{\text{πλήθος μολυσμένων κόμβων}}{\text{συνολικό πλήθος κόμβων}}$$

$$P_{ingat} = \frac{\text{πλήθος μολυσμένων κόμβων της ομάδας που βρίσκεται ο επιτιθέμενος}}{\text{συνολικό πλήθος κόμβων της ομάδας που βρίσκεται ο επιτιθέμενος}} \quad \textbf{(17)}$$

Υπενθυμίζουμε ότι παρόλο που τα μεγέθη ορίστηκαν εδώ ως στιγμιαία, δηλαδή αναφέρονται σε κάθε χρονικό στιγμιότυπο του πειράματος, εντούτοις στις γραφικές παραστάσεις που παρατίθενται παρακάτω λαμβάνονται τα αντίστοιχα μέσα μεγέθη τους, δηλαδή ο μέσος όρος τους για το συνολικό χρόνο του πειράματος.

Τέλος, ο αριθμός των ανωμαλιών που μπορούν να ανιχνευτούν σε κάθε χρονικό στιγμιότυπο του δικτύου είναι μείζονος σημασίας για τον προτεινόμενο αλγόριθμο. Πιο συγκεκριμένα, η τεχνική που χρησιμοποιήθηκε έχει τη δυνατότητα να ανιχνεύει μόνο ένα ποσοστό ανώμαλων κόμβων επί του συνόλου των κόμβων που συμμετέχουν στην τεχνική αυτή σε κάθε χρονικό στιγμιότυπο. Η ανίχνευση ανωμαλιών πέρα ενός ποσοστού επί των συνολικών συμμετεχόντων κόμβων είναι αδύνατη.

Κεφάλαιο 4^{o} – Περιγραφή Μοντέλου Προσομοίωσης

4.1 Τοπολογία δικτύου

Θεωρούμε ένα ad hoc ασύρματο δίκτυο αισθητήρων οι κόμβοι του οποίου είναι χωρικά ομοιόμορφα και τυχαία κατανεμημένοι σε μια τετραγωνική περιοχή $1500\,m \times 1500\,m$ και κάθε κόμβος έχει σταθερή ακτίνα μετάδοσης ίση με $R = 200\,m$. Υποθέτουμε ότι ο επιτιθέμενος μπορεί να ακολουθήσει διαφορετικές στρατηγικές επίθεσης. Επίσης, θεωρούμε, ότι η ακτίνα του κακόβουλου κόμβου μπορεί να μεταβάλλεται μεταξύ διαφορετικών στιγμιότυπων μιας προσομοίωσης, σύμφωνα με τη στρατηγική επίθεσης που ακολουθείται (π.χ. μικρή ή μεγάλη ακτίνα μετάδοσης). Επιπλέον, για λόγους απλοποίησης της παρουσίασης αποτελεσμάτων, υποθέτουμε ότι όλοι οι κόμβοι έχουν τον ίδιο ρυθμό μόλυνσης l (που παρακάτω θα αναφέρεται ως ρυθμός μόλυνσης κόμβου δικτύου), ενώ τα κανάλια στα οποία εμπλέκεται ο κακόβουλος κόμβος έχουν διαφορετικό ρυθμό μόλυνσης l_{mal} (που θα αναφέρεται ως κακόβουλος ρυθμός μόλυνσης). Ο επιτιθέμενος υποτίθεται ότι κινείται σε όλο το δίκτυο σύμφωνα με ένα τυχαίο μοντέλο κίνησης με μηδενικό χρόνο σταματήματος (τυχαίος περίπατος) **[46]**. Η ταχύτητα του επιτιθέμενου επιλέγεται τυχαία στο διάστημα $(0, u_{\max})$ και η κατεύθυνσή του είναι ομοιόμορφα κατανεμημένη στο διάστημα $(0, q_{\max})$. Σε αυτή την εργασία θεωρούμε

ένα τυχαίο πρότυπο κινητικότητας και τη $q_{\max}$ τη λαμβάνουμε ίση με $q_{\max} = 2p$

Επειδή η αναπτυσσόμενη, για τους σκοπούς της προσομοίωσης, περιοχή έχει πεπερασμένο εμβαδό, είναι πιθανό ο επιτιθέμενος, κατά την κίνησή του, να φτάσει, κάποια στιγμή, το όριο της περιοχής αυτής. Για να αντιμετωπιστεί αυτή η περίπτωση, υποθέτουμε ότι η περιοχή αναδιπλώνεται σαν τορός, ώστε αν ένας κόμβος διασχίσει ένα όριο, εισέρχεται ξανά στην περιοχή από το αντίθετο, σε σχέση με το κέντρο, όριο της περιοχής. Ένα τέτοιο μοντέλο κινητικότητας ονομάζεται τυχαίος περίπατος με αναδίπλωση **[46]**. Η σταθερά κανονικοποίησης C του μοντέλου κατανάλωσης ενέργειας που παρουσιάστηκε στο προηγούμενο κεφάλαιο (*βλ.* εξ. 7), έχει επιλεγεί έτσι ώστε ο ρυθμός κατανάλωσης να διασφαλίζει επαρκή χρόνο προσομοίωσης για το σύστημα, ώστε το τελευταίο να φτάσει στη σταθερή κατάσταση. Το αρχικό ενεργειακό απόθεμα έχει καθοριστεί για $E = 150$ μονάδες ενέργειας (***energy units***, ***eu***), η τιμή του βήματος κατανάλωσης είναι $E_{step} = 0.1$ ***eu*** , ενώ η σταθερά απωλειών χώρου *α* έχει θεωρηθεί $a = 2.5$.

Πρέπει να σημειώσουμε ότι το πρότυπο κατανάλωσης ενέργειας που χρησιμοποιείται στο μοντέλο μας βασίζεται κυρίως στην επίδραση της ακτίνας μετάδοσης. Ωστόσο, η χρήση μοντέλων που ενσωματώνουν επιπρόσθετο ενεργειακό κόστος λόγω κινητικότητας στον κανόνα κατανάλωσης της ενέργειας και η μελέτη της επίδρασής τους στα σχετικά αποτελέσματα, θα μπορούσε να θεωρηθεί εναλλακτικά.

4.2 Υποδιαίρεση δικτύου σε ομάδες

Η περιοχή του υπό μελέτη δικτύου χωρίστηκε σε εννέα μεγάλα τετράγωνα – ομάδες (***groups***). Κόμβοι μέσα στο ίδιο ***group*** θεωρείται ότι αποτελούν ένα σύνολο κόμβων. Τα χαρακτηριστικά που φέρει μια ομάδα κόμβων έχουν περιγραφεί στο Κεφάλαιο 3. Κάθε τέτοια ομάδα της περιοχής του δικτύου χωρίστηκε σε εννέα επιμέρους μικρότερα τετράγωνα – υπομάδες (***tiles***). Σαφέστερη εικόνα για τον τρόπο με τον οποίο χωρίστηκε η περιοχή του δικτύου δίνεται στο σχήμα που ακολουθεί.

<table>
<tr><td colspan="3">group 1</td><td colspan="3">group 2</td><td colspan="3">…</td></tr>
<tr><td>tile</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td>1</td><td>2</td><td>…</td></tr>
<tr><td></td><td></td><td></td><td>4</td><td>5</td><td>6</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td>7</td><td>8</td><td>9</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="3" rowspan="3">group 9</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Σχήμα 5 Χωρισμός περιοχής δικτύου σε ***groups*** και ***tiles***

Οι κόμβοι–αισθητήρες, στα πλαίσια των πειραμάτων της εργασίας, παίρνουν διάφορες μετρήσεις θερμοκρασίας του χώρου στον οποίο βρίσκονται. Στην περιοχή του δικτύου κατανέμονται οι τιμές των δεδομένων θερμοκρασίας, τις οποίες θεωρούμε ότι διαβάζουν οι κόμβοι–αισθητήρες ανάλογα με τη θέση τους στο δίκτυο. Η κατανομή των τιμών αυτών έγινε με τέτοιο τρόπο, ώστε κάθε ***group*** να εμφανίζει συσχέτιση στα δεδομένα του με το προηγούμενο και το επόμενό του ***group***. Για παράδειγμα, το ***group*** 2 εμφανίζει υψηλή συσχέτιση με το προηγούμενο του ***group*** 1 αλλά και με το επόμενό του ***group*** 3. Τονίζεται, ότι τα δεδομένα των ***tiles*** καθενός ***group*** ξεχωριστά εμφανίζουν πολύ υψηλή συσχέτιση μεταξύ τους.

Οι κόμβοι του δικτύου διαβάζουν τιμές από το ***tile*** του ***group*** στο οποίο βρίσκονται. Δηλαδή, σε κάθε ένα ***tile*** της περιοχής του δικτύου έχει αντιστοιχηθεί ένα σύνολο τιμών δεδομένων το οποίο μπορεί να διαβαστεί από κόμβους που βρίσκονται στο εν λόγω ***tile***. Η ανάγνωση τιμών γίνεται σε κάθε χρονικό στιγμιότυπο του δικτύου, το οποίο αντιστοιχεί σε κάθε νέο γεγονός, μόλυνσης ή επαναφοράς, που συμβαίνει στο δίκτυο. Στην περίπτωση που ένας κόμβος δεν είναι μολυσμένος διαβάζεται η τιμή του ***tile*** για το εν λόγω χρονικό

στιγμιότυπο. Όταν ο κόμβος αυτός είναι μολυσμένος εισάγεται τυχαίο σφάλμα, τυχαίου πρόσημου, στην τιμή του *tile* που διαβάζεται.

4.3 Παραμετροποίηση αλγόριθμου ανίχνευσης

Για την ανίχνευση ανωμαλιών που τυχαία μπορούν να εμφανιστούν ανά χρονικό στιγμιότυπο στο δίκτυο, εφαρμόζεται ο αλγόριθμος του προηγούμενου κεφαλαίου (*βλ.* παρ. 3.10). Αρχικά και πριν από τον αλγόριθμο, εφαρμόζεται η τεχνική *PCA* στις τιμές των δεδομένων των κόμβων που ανήκουν στο ίδιο *group*. Αυτή η τεχνική εφαρμόζεται σε κάθε χρονικό στιγμιότυπο και για κάθε ένα από τα εννέα *group* ξεχωριστά. Τα εξαγόμενα από την *PCA* κανονικοποιούνται κατάλληλα και αποτελούν είσοδο για τον αλγόριθμο ανίχνευσης ανωμαλιών. Τα κριτήρια και οι προϋποθέσεις για να ανιχνευτεί μία ανωμαλία αναφέρθηκαν στο προηγούμενο κεφάλαιο.

Στο προηγούμενο κεφάλαιο έγινε, επίσης, αναφορά στη σημαντικότητα της επιλογής κατάλληλου αριθμού κύριων συνιστωσών r. Για την περίπτωση της μεθόδου που προτείνεται σε αυτήν την εργασία επιλέγεται ο κατάλληλος αριθμός του r με στόχο το βέλτιστο διαχωρισμό του κανονικοποιημένου διανύσματος y (*βλ.* εξ. 10) στους δύο υποχώρους S και $\tilde{S}$. Ο διαχωρισμός γίνεται έτσι ώστε οι πρώτες r κύριες συνιστώσες να αντιστοιχούν στον ομαλό υποχώρο S και οι υπόλοιπες στον $\tilde{S}$.

Ένας κανόνας, που εμπλέκει τις κύριες συνιστώσες και αφορά την ανίχνευση ανωμαλιών, βασίζεται στο μέγεθος της διακύμανσης που έχει αιχμαλωτίσει κάθε κύρια συνιστώσα. Η κύρια ιδέα πίσω από αυτόν τον κανόνα είναι ότι αν όλες οι μεταβλητές των αρχικών δεδομένων είναι ανεξάρτητες, τότε οι κύριες συνιστώσες αντιστοιχούν στις αρχικές μεταβλητές και θα έπρεπε όλες να έχουν διακύμανση ίση με τη μονάδα (μετά την κανονικοποίηση). Συνεπώς, κάθε κύρια συνιστώσα με διακύμανση 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 κύρια συνιστώσα που να συσχετίζεται με την ομάδα αυτή με διακύμανση μεγαλύτερη της μονάδας και ο κανόνας αυτός θα διατηρήσει μονάχα μια κύρια συνιστώσα που σχετίζεται με την ομάδα αυτή.

Στα πλαίσια της διπλωματικής αυτής, λαμβάνοντας υπόψη όλες τις παραμέτρους για την κατάλληλη επιλογή του αριθμού των κύριων συνιστωσών, επελέγη ως το καταλληλότερο ποσοστό για τα πειράματα – προσομοιώσεις που έγιναν το κλάσμα $2/5$ επί του συνολικού αριθμού των κόμβων κάθε ***group*** για τα οποία εκτελείται ξεχωριστά η ***PCA***.

Στο προηγούμενο κεφάλαιο το ***SPE*** προκρίθηκε ως το πλέον κρίσιμο μέγεθος για την ανίχνευση ανωμαλιών, αφού κάθε φορά που εμφανίζεται μια ανωμαλία αυτό υπερβαίνει τα προκαθορισμένα όρια. Όμως, δεν υπάρχει μια συγκεκριμένη τιμή κατωφλίου για το ***SPE*** η οποία να καθορίζεται στατικά και η υπέρβασή της να επισημαίνει εμφάνιση ανωμαλίας. Αντιθέτως, θα πρέπει να σημειωθεί ότι εφόσον το ***SPE*** είναι το μέτρο–νόρμα του διανύσματος y_{res}, η τιμή της θα αυξάνεται με την αύξηση του αριθμού των προσθετέων–συνιστωσών του. Κατά συνέπεια, το όριο για την τιμή του ***SPE***, η υπέρβαση του οποίου θα σημαίνει και την εμφάνιση ανωμαλίας, θα πρέπει να μεταβάλλεται δυναμικά συναρτήσει του αριθμού των κόμβων που περιλαμβάνονται στην ***PCA***.

Ανάλογη λογική ισχύει και για την περίπτωση του $P_{contribution}$, μέγεθος εξαγόμενο από το ***SPE***, το οποίο επιλέχθηκε για την ανίχνευση του κόμβου στον οποίον εμφανίζεται ανωμαλία. Οι συνιστώσες του ***SPE*** που έχουν τη μεγαλύτερη συνεισφορά στην τιμή ***SPE***, και μόνο με την προϋπόθεση το ότι αυτό έχει ξεπεράσει το προκαθορισμένο κατώφλι, είναι αυτές που αντιστοιχούν στους κόμβους οι οποίοι εμφανίζουν ανωμαλία. Το ποσοστό συμμετοχής ενός μολυσμένου κόμβου στη διαμόρφωση του ***SPE*** είναι πολλαπλάσιο σε σύγκριση με το αντίστοιχο ποσοστό ενός μη μολυσμένου κόμβου. Βέβαια οι συνεισφορές

των προσθετέων του *SPE* – και κατ' επέκταση των κόμβων – διαφέρουν ανάλογα με τον αριθμό των κόμβων που συμμετέχουν κάθε φορά στην *PCA*. Επίσης, το κατώφλι του ποσοστού ενός κόμβου πάνω απ' το οποίο θεωρούμε ότι εμφανίζεται ανωμαλία μεταβάλλεται και αυτό δυναμικά συναρτήσει του αριθμού των κόμβων που παίρνουν μέρος στη *PCA*.

Τέλος, τα όρια για τη συγκριτική μελέτη των πραγματικών τιμών των δεδομένων με το μέσο όρο των τιμών των δεδομένων κατάρτισης ελήφθησαν σταθερά. Αποτελούν μια περιοχή τιμών γύρω από το μέσο όρο της κανονικά μετρούμενης τιμής του κάθε κόμβου με εύρος μία σταθερή απόσταση από το συγκεκριμένο μέσο όρο. Η απόσταση από το μέσο όρο λαμβάνεται σε σχέση με την τυπική απόκλιση που εμφανίζουν τα δεδομένα κατάρτισης του αντίστοιχου κόμβου.

4.3.1 Μειονεκτήματα της μεθόδου ανίχνευσης ανωμαλιών

Όπως αναφέρθηκε προηγουμένως στο υποκεφάλαιο 4.2, οι τιμές των κόμβων κάθε *tile* πάνω στα οποία εφαρμόζεται η τεχνική *PCA*, για την ανίχνευση ανωμαλιών, εμφανίζουν πολύ υψηλή συσχέτιση μεταξύ τους. Η συσχέτιση αυτή λαμβάνεται θετική και ίδιας τάξεως μεγέθους για όλες τις τιμές των *tile* ενός *group*. Όμως, όταν σε ένα σύνολο κόμβων, με υψηλή συσχέτιση, εμφανιστεί μια μεταβολή στην τιμή ενός κόμβου τότε αυτή επηρεάζει και τείνει να αλλάξει και τις τιμές των υπόλοιπων κόμβων του συνόλου αυτού. Μάλιστα οι δευτερεύουσες μεταβολές, οι μεταβολές που εμφανίζονται λόγω της αρχικής μεταβολής, έχουν αντίθετο πρόσημο από αυτό της αρχικής μεταβολής.

Δηλαδή, λόγω της συσχέτισης που υπάρχει μεταξύ των δεδομένων στους κόμβους μιας ομάδας έχουμε εικονική (ή επηρεαζόμενη) ανωμαλία. Η τελευταία δεν αποτελεί ανεξάρτητη ανωμαλία, όπως αυτές που αναφέρθηκαν στο υποκεφάλαιο 3.9, αλλά προκύπτει επειδή η τιμή του κόμβου με αυτού του είδους ανωμαλία επηρεάζεται, λόγω

ισχυρής συσχέτισης, από την τιμή κόμβου με πραγματική ανωμαλία που υπάρχει στην ίδια ομάδα. Με άλλα λόγια, η μεταβολή λόγω ανωμαλίας στην ένδειξη ενός κόμβου επηρεάζει και μεταβάλλει τις ενδείξεις κόμβων που συσχετίζονται ισχυρά με τον εν λόγω κόμβο.

Πιο συγκεκριμένα, στις προσομοιώσεις που πραγματοποιήθηκαν, η αρχική μεταβολή στις τιμές των κόμβων ενός *group* οφείλεται στη μόλυνση των κόμβων αυτών. Στη συνέχεια, αυτή η μόλυνση, λόγω της υψηλής συσχέτισης τιμών, θα επηρεάσει και τις τιμές των υπολοίπων κόμβων του *group*. Με άλλα λόγια όταν σε κάποιο χρονικό στιγμιότυπο εμφανιστεί μία μόλυνση σε ένα *group* κόμβων τότε μεταβάλλονται όλες οι τιμές των κόμβων αυτού του *group*. Κατά γενικό κανόνα, επειδή η συσχέτιση μεταξύ κόμβων είναι ίδιας τάξεως μεγέθους, οι δευτερεύουσες μεταβολές κατανέμονται ομοιόμορφα στους υπόλοιπους κόμβους του *group*. Έτσι, στην περίπτωση μοναδικών μολύνσεων σε ένα *group* κόμβων, οι τιμές των υπόλοιπων μη μολυσμένων κόμβων δεν παρουσιάζουν σημαντική διακύμανση από το μέσο όρο της κανονικής μετρούμενης τιμής τους, δηλαδή σε όρους *PCA* και ανίχνευσης δεν παρουσιάζουν κάποια ανωμαλία τιμών.

Αντίθετα, στην περίπτωση που σε ένα *group* κόμβων εμφανίζονται πολλαπλές μολύνσεις, η διαδικασία ανίχνευσης περιπλέκεται. Υπενθυμίζουμε ότι η μόλυνση ενός κόμβου γίνεται με τελείως στοχαστικό τρόπο. Αυτό σημαίνει ότι στο μολυσμένο κόμβο εισάγεται ανωμαλία τυχαίου πρόσημου και τυχαίου ποσοστού επί της κανονικής μετρούμενης τιμής του συγκεκριμένου κόμβου. Συχνά εμφανίζονται φαινόμενα επικάλυψης ή απόκρυψης ανωμαλιών που αποτελούν τον κύριο παράγοντα της μη ανιχνευσιμότητας (δηλαδή της μη δυνατότητας ανίχνευσης) ανωμαλιών κατά την εξέλιξη της προσομοίωσης.

Για την πλήρη κατανόηση των παραπάνω παρατίθενται δύο αντιπροσωπευτικά παραδείγματα τα οποία αντλήθηκαν από την πειραματική διαδικασία.

1ο παράδειγμα

Έστω ένα ***group*** που αποτελείται από έξι κόμβους, οι τιμές των οποίων εμφανίζουν υψηλή συσχέτιση μεταξύ τους. Στους κόμβους που υπάρχουν μολύνσεις εμφανίζονται ενδεικτικά ποσοστά ανωμαλιών και το πρόσημό τους. Οι κόμβοι χωρίς μόλυνση εμφανίζονται με παύλες.

Αναγνωριστικό κόμβων	1	2	3	4	5	6
Ανωμαλίες ενός χρονικού στιγμιότυπου	-25%	–	-15%	-35%	–	-25%

Παρατηρείται ότι οι μολυσμένοι κόμβοι (1, 3, 4 και 6) έχουν ίδιου πρόσημου ανωμαλίες (αρνητικό). Ένα από τα αποτέλεσμα αυτού είναι ότι οι τιμές των κόμβων χωρίς ανωμαλία (2 και 5) θα μεταβληθούν επηρεαζόμενες συσσωρευτικά από τις εισερχόμενες ανωμαλίες των άλλων κόμβων. Έτσι οι κόμβοι 2 και 5 μπορεί να εμφανίζονται ότι έχουν ένα ποσοστό ανωμαλίας αντίθετου πρόσημου (θετικό). Από την άλλη, επειδή η ανωμαλία ενός κόμβου επηρεάζει την τιμή των υπολοίπων κόμβων του ***group***, υπάρχει ενδεχόμενο κάποιοι κόμβοι που έχουν αρχικά μολυνθεί, τελικά να εμφανίζονται χωρίς ανωμαλία (π.χ. το οριακά ανιχνεύσιμο ποσοστό ανωμαλίας 15% του κόμβου 3 ενδέχεται να μην ανιχνευτεί λόγω της μεταβολής του από τις ανωμαλίες των υπόλοιπων κόμβων). Μέσα από το συγκεκριμένο παράδειγμα γίνεται σαφές ότι ο αλγόριθμος που προτείνεται στην παρούσα διπλωματική αστοχεί στις περιπτώσεις εκτεταμένων μολύνσεων.

2ο παράδειγμα

Έστω το ***group*** κόμβων του προηγούμενου παραδείγματος με την παρακάτω κατανομή του ποσοστού ανωμαλιών:

Αναγνωριστικό κόμβων	1	2	3	4	5	6
Ανωμαλίες ενός χρονικού στιγμιότυπου	–	–	–	+35%	–	+20%

Στο παράδειγμα αυτό δεν έχουμε εκτεταμένη μόλυνση, εντούτοις εμφανίζεται ένα από τα πιο συχνά ενδεχόμενα, η επικάλυψη ομόσημων ανωμαλιών. Δηλαδή, η ανωμαλία του κόμβου 6 μπορεί να καταστεί μη ανιχνεύσιμη λόγω της ανωμαλίας ίδιου πρόσημου που εμφανίζεται στον κόμβο 4 (αφού λόγω αυτής και με δεδομένο ότι πολλές φορές κάποιες συσχετίσεις κόμβων είναι πιο ισχυρές από άλλες, η τιμή του κόμβου 6 τείνει τελικά στα επίπεδα τιμών που αυτός έχει χωρίς μόλυνση). ■

Μία εναλλακτική επιλογή αντιμετώπισης των προβλημάτων που μόλις παρουσιάστηκαν (αλλά και γενικά των προβλημάτων ανιχνευσιμότητας) είναι η κατάλληλη επεξεργασία των τιμών των κόμβων κάθε ***group*** βάσει των συσχετίσεων τους, αφού στην πραγματικότητα οι συσχετίσεις είναι πολύ υψηλές και εμφανίζουν διαφορές ανάμεσα στους κόμβους του ***group***. Όμως κάτι τέτοιο τελικά δεν υλοποιήθηκε από τη μία λόγω της αυξημένης πολυπλοκότητάς και από την άλλη λόγω του ότι βασική απαίτηση του αλγόριθμου ήταν η κατ' ελάχιστο συνεισφορά του στην πιθανότητα εσφαλμένης ανίχνευσης P_f. Δηλαδή, κύριο μέλημα ήταν η ανάπτυξη ενός αξιόπιστου αλγόριθμου ανίχνευσης που θα συνεισφέρει κατ' ελάχιστο στη δημιουργία ανωμαλιών χωρίς πραγματική μόλυνση (εικονικές ανωμαλίες).

Άλλωστε, αφού ο αλγόριθμος ενεργεί σε επίπεδο *group* για κάθε χρονικό στιγμιότυπο της προσομοίωσης ξεχωριστά, η πολυπλοκότητα του αλγόριθμου είναι $O(m)$ όπου $m = t \cdot N$. Με t συμβολίζουμε το σύνολο των χρονικών στιγμιότυπων της προσομοίωσης και με N το συνολικό αριθμό των κόμβων του δικτύου. Δηλαδή, η πολυπλοκότητα του προτεινόμενου αλγόριθμου εξαρτάται τόσο από τον αριθμό των χρονικών στιγμιότυπων του πειράματος όσο και από την πυκνότητα του δικτύου. Έτσι, κατά τη διάρκεια των πειραμάτων έγινε προσπάθεια συμβιβασμού μεταξύ του δείγματος στιγμιότυπων του πειράματος που λαμβάνεται, το οποίο θα πρέπει να είναι αντιπροσωπευτικό και να μην προδίδει τη στοχαστικότητα του μοντέλου επίθεσης και διάδοσης και της αποδοτικότητας του αλγόριθμου ανίχνευσης.

Κεφάλαιο 5^{ο} – Πειραματικά Αποτελέσματα

Στο κεφάλαιο αυτό επιχειρείται η αποτίμηση της απόδοσης του αλγόριθμου που αναπτύχθηκε προηγουμένως. Μελετάται, μέσω πειραμάτων προσομοίωσης, η επίδραση διαφόρων παραμέτρων του δικτύου αισθητήρων – που άλλοτε αφορούν τα λειτουργικά χαρακτηριστικά του κινητού επιτιθέμενου κόμβου και άλλοτε τα χαρακτηριστικά του ίδιου του δικτύου αισθητήρων – στην ικανότητα του αλγόριθμου για ανίχνευση ανωμαλιών. Συγκεκριμένα, η αξιολόγηση του αλγόριθμου πραγματοποιήθηκε για διάφορα σενάρια προσομοιώσεων και καταγράφηκε η αντίστοιχη συμπεριφορά του.

Η μελέτη της αποτελεσματικότητας του αλγόριθμου για διάφορες στρατηγικές επίθεσης γίνεται με κριτήριο το μέγεθος P_d που ορίστηκε στο προηγούμενο κεφάλαιο. Η αποτίμηση του αλγόριθμου μας παρέχει τη δυνατότητα να αξιολογούμε τόσο τη μολυσματική ικανότητα του επιτιθέμενου όσο και την ικανότητα του δικτύου που δέχεται την επίθεση να ανιχνεύσει έγκαιρα το σημείο που βρίσκεται ο επιτιθέμενος και κατ' επέκταση την περιοχή διάδοσης μιας μόλυνσης.

Τέλος, για την αξιολόγηση του αλγόριθμου αυτού πραγματοποιήθηκαν αρκετές σειρές πειραμάτων. Οι προσομοιώσεις αυτές πραγματοποιήθηκαν με διαφορετικές παραμέτρους κάθε φορά. Αρχικά, παρατίθενται πειράματα που αφορούν στην απόδοση του

αλγόριθμου από τη σκοπιά του δικτύου και εν συνεχεία πειράματα για διάφορες περιπτώσεις στρατηγικών επίθεσης που αναδεικνύουν την ανιχνευτική ικανότητα του αλγόριθμου από τη σκοπιά του επιτιθέμενου.

5.1 Μεταβάλλοντας τα χαρακτηριστικά του δικτύου

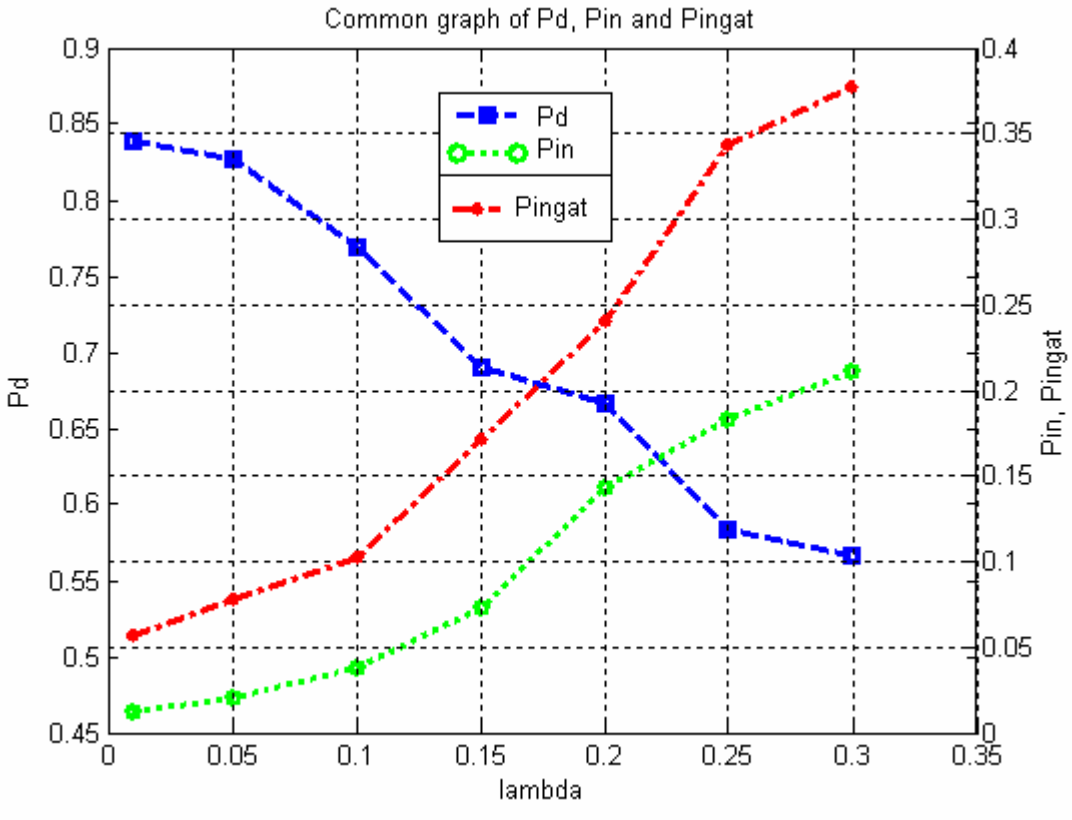
Σε αυτό το υποκεφάλαιο εξετάζεται η σχέση που υπάρχει ανάμεσα στην πιθανότητα ανίχνευσης ανωμαλιών P_d και στο ρυθμό μόλυνσης καναλιού μεταξύ των κόμβου του δικτύου, l . Για αυτό το σκοπό πραγματοποιήθηκαν τρεις σειρές πειραμάτων, κάθε μία για διαφορετική πυκνότητα δικτύου. Πρέπει να σημειωθεί ότι αύξηση της πυκνότητας του δικτύου επιτυγχάνεται με αύξηση του αριθμού των κόμβων του δικτύου N , διότι το εμβαδόν που αναπτύσσεται το δίκτυο παραμένει σταθερό.

Στα πειράματα αυτού του υποκεφαλαίου κρατήσαμε σταθερά όλα τα χαρακτηριστικά στοιχεία του επιτιθέμενου κόμβου και μεταβάλλαμε μόνο το ρυθμό μόλυνσης του καναλιού μεταξύ δύο κόμβων του δικτύου. Για λόγους απλότητας θεωρήσαμε το l σταθερό για όλους τους κόμβους του δικτύου δίνοντας του ένα εύρος τιμών από 0.01 έως 0.2. Επίσης, ο ρυθμός μόλυνσης του κακόβουλου κόμβου θεωρήθηκε ίσος με $l_{mal} = 0.1$.

α) Πυκνότητα $N = 100$

Στο Σχήμα 6 που ακολουθεί παρουσιάζονται οι διάφορες τιμές που λαμβάνει η πιθανότητα ανίχνευσης ανωμαλιών P_d και τα μεγέθη P_{in} και P_{ingat} συναρτήσει του l για μία αραιή τοπολογία δικτύου. Πιο συγκεκριμένα, ο αριθμός των κόμβων του δικτύου ελήφθη ίσος με $N = 100$. Επίσης, η ακτίνα του επιτιθέμενου λαμβάνεται ίση με $R_{mal} = 200\,m$ όπως και η ακτίνα μεταξύ των κόμβων του δικτύου θεωρήθηκε ίση με

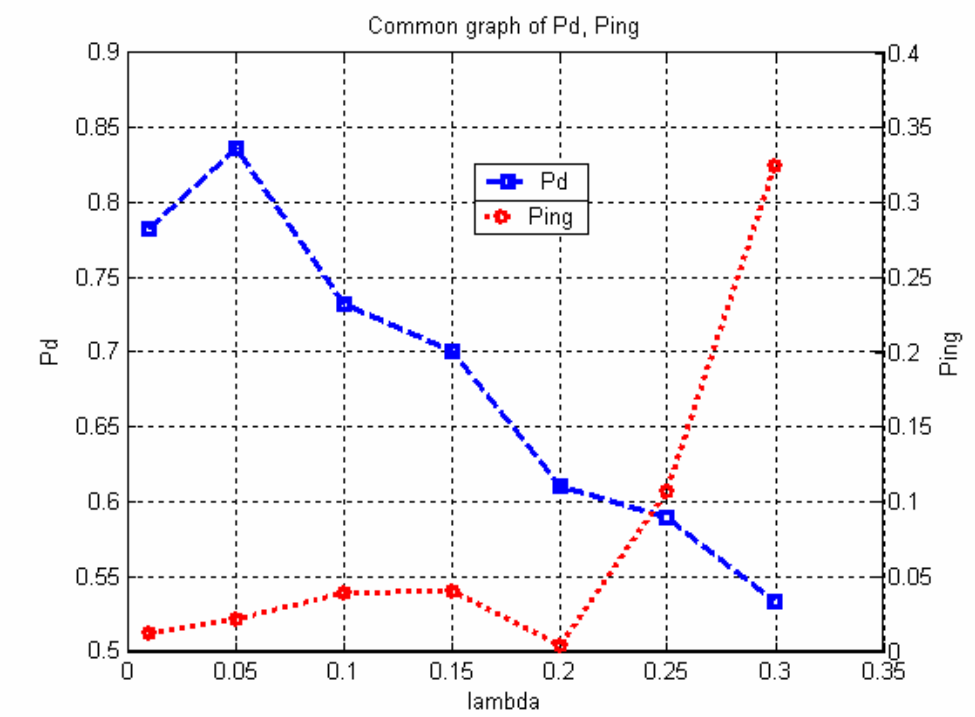
$R = 200\,m$ (τα μεγέθη αυτά μένουν σταθερά και για τα πειράματα που έπονται, εφόσον δεν αναφέρεται κάποια διαφοροποίηση). Τέλος, η μέγιστη ταχύτητα που κακόβουλου κόμβου θεωρήθηκε σταθερή και ίση με $u_{\max} = 25\,m/tu$.



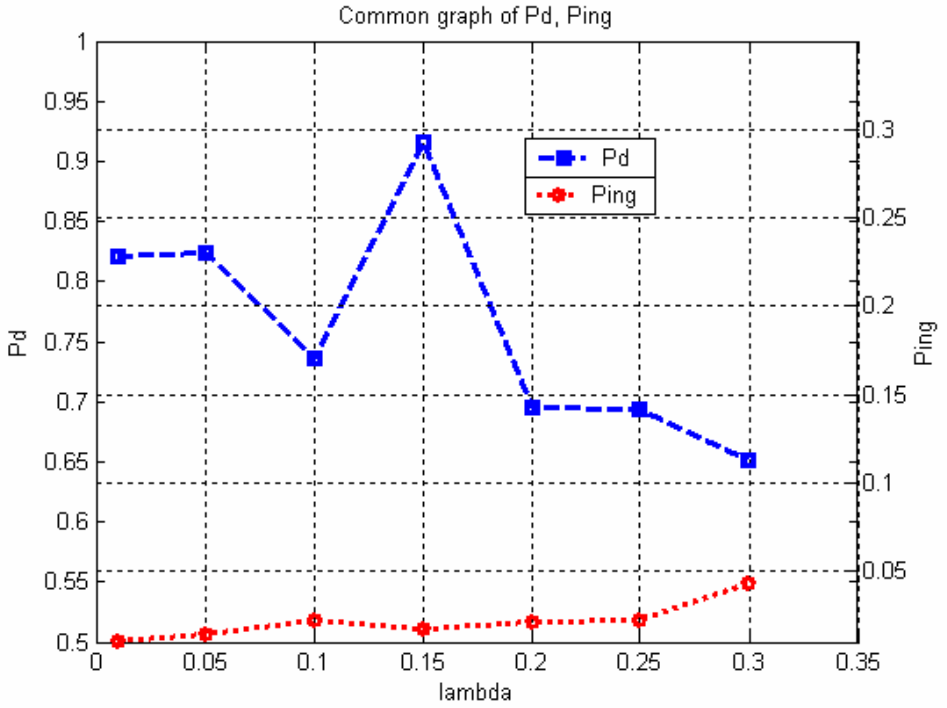
Σχήμα 6 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l

Ωστόσο, πέρα από το συγκεκριμένο διάγραμμα που αφορά στη συνολική εικόνα του δικτύου, κρίθηκε σκόπιμο να δοθούν και κάποιες χαρακτηριστικές γραφικές παραστάσεις που αφορούν τα διάφορα *group* του δικτύου αισθητήρων στο συγκεκριμένο πείραμα. Ο αλγόριθμος για κάθε χρονικό στιγμιότυπο του πειράματος εκτελείται σε κάθε *group* του δικτύου αισθητήρων ξεχωριστά. Κατά συνέπεια, τα παρακάτω σχήματα δίνουν μια εικόνα τοπικής απόδοσης του αλγόριθμου για κάθε *group* του δικτύου, σε αντίθεση με το προηγούμενο γράφημα που δίνει τη συνολική εικόνα απόδοσης του αλγόριθμου. Το

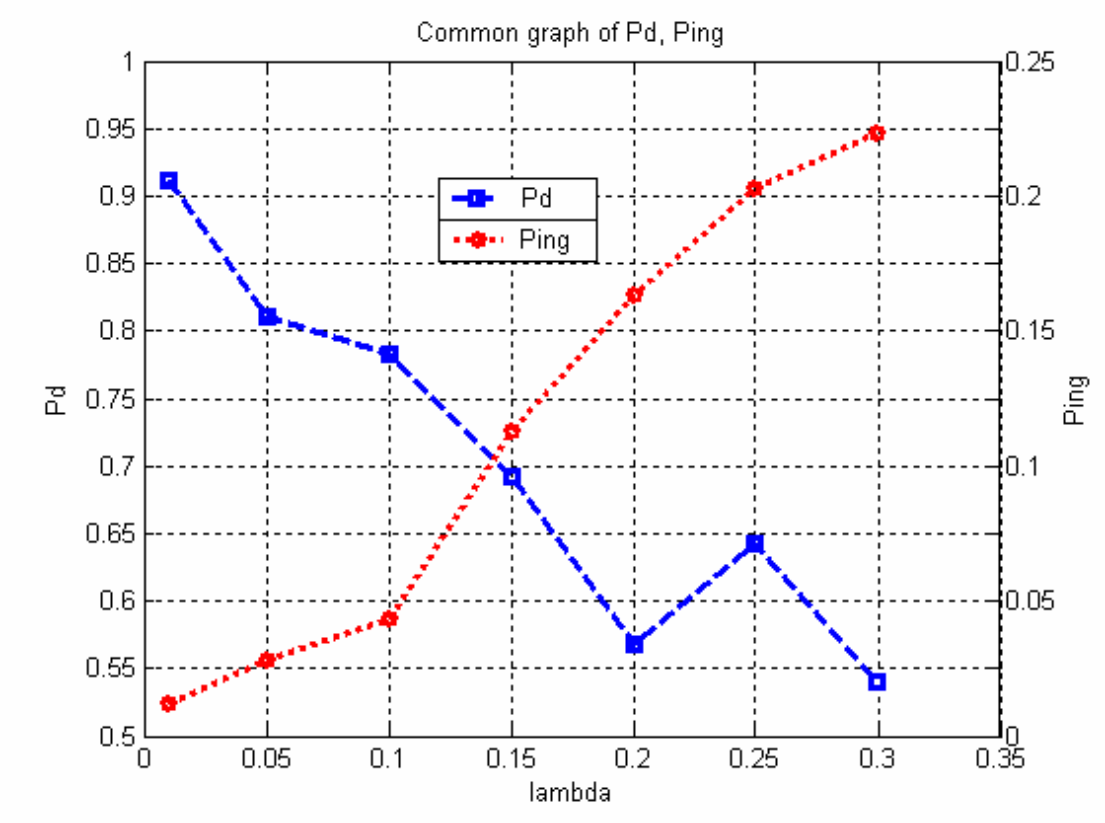
μέγεθος P_{ing} ορίζεται όπως το P_{in} με τη διαφορά ότι αναφέρεται τοπικά στο μέσο ποσοστό μολυσμένων κόμβων του κάθε *group*.



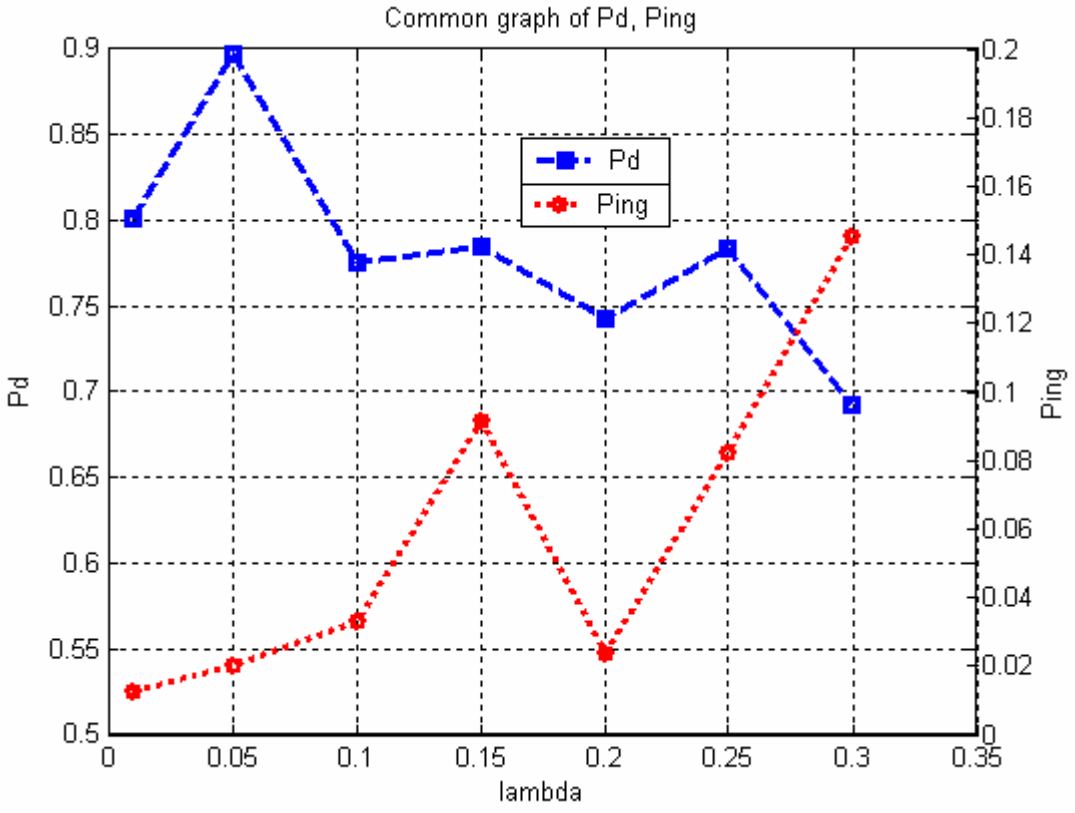
Σχήμα 7 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 1



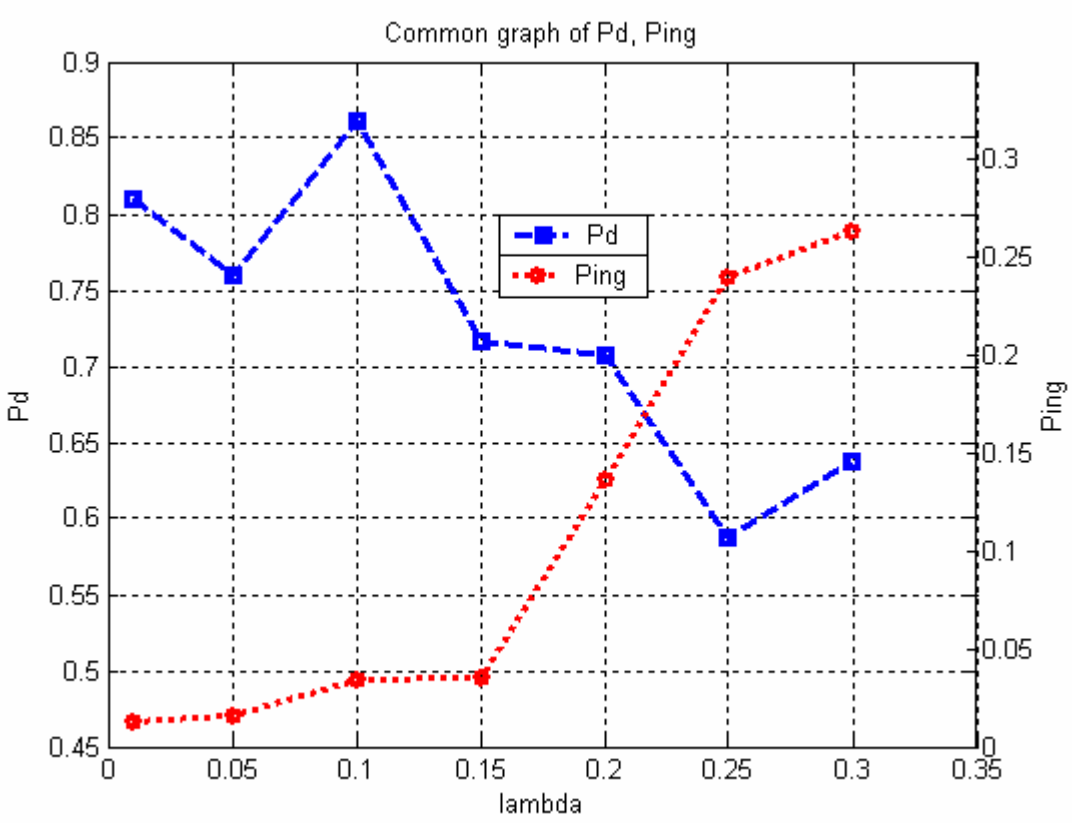
Σχήμα 8 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 3



Σχήμα 9 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 5



Σχήμα 10 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 6



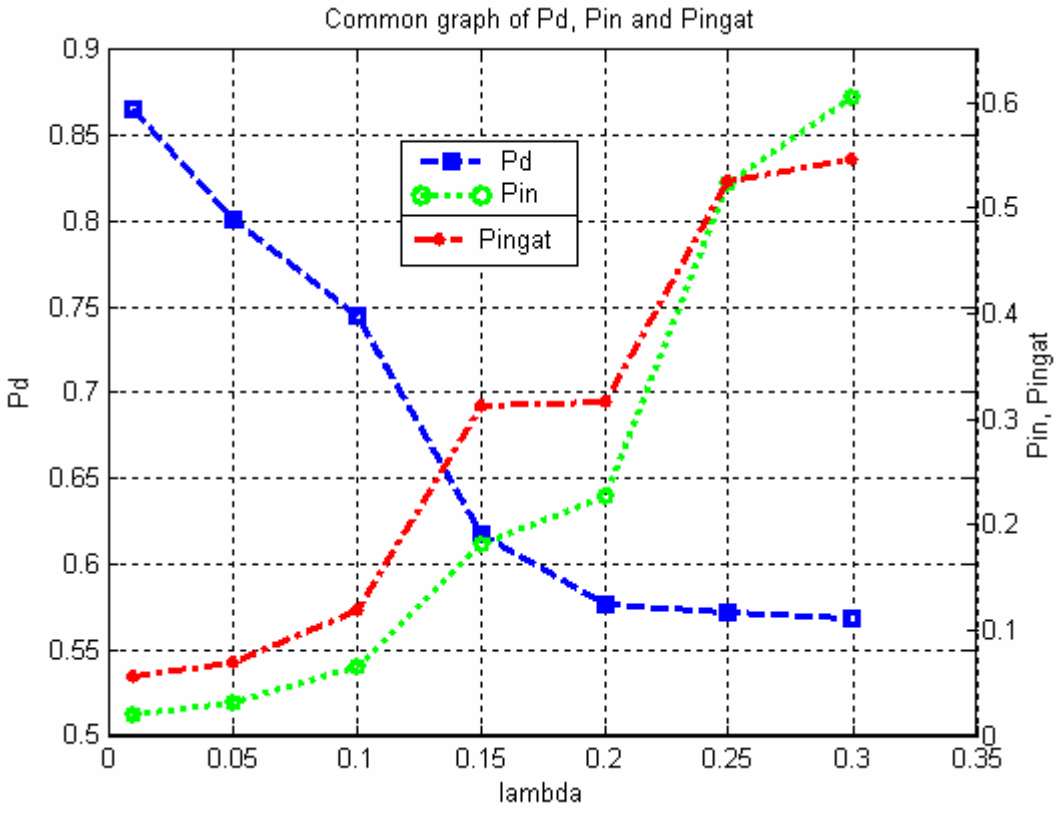
Σχήμα 11 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 7

Από τα διάγραμμα των *group* του δικτύου (*βλ.* σχ. 7 – 11) παρατηρούμε ότι η πιθανότητα ανίχνευσης ανωμαλίας P_d μειώνεται με την αύξηση του l. Στα *group* του δικτύου, χαμηλές τιμές του l έχουν ως αποτέλεσμα τις αντίστοιχα χαμηλές τιμές του P_{ing} οι οποίες, με τη σειρά τους, οδηγούν σε υψηλές τιμές του P_d και αντιστρόφως. Επιπλέον, αξίζει να σημειωθεί ότι το P_{ing} του *group* 3 (Σχήμα 8) παραμένει στα ίδια επίπεδα για όλες τις τιμές l που ελήφθησαν κατά τη διάρκεια των προσομοιώσεων. Συνέπεια αυτού, είναι το γεγονός ότι το P_d διατηρείται σε υψηλά επίπεδα συγκριτικά με τις τιμές των άλλων *group*. Τέλος, πρέπει να σημειωθεί ότι σε κάποια *group* του δικτύου το P_{ing} εμφανίζεται με υψηλότερες τιμές απ' ότι σε άλλα, λ.χ. στο *group* 3 οι τιμές που λαμβάνει το P_{ing} είναι της τάξης του 0.05 τη στιγμή που στο *group* 1 (Σχήμα 7) το ίδιο μέγεθος λαμβάνει τιμές κοντά στο 0.35. Αυτό βέβαια έχει ανάλογη επίπτωση στις τιμές του P_d, όπως ήδη έχουμε

αναφέρει και δείχνει ότι το κάθε *group* κόμβων επηρεάζεται διαφορετικά κατά την εξέλιξη των πειραμάτων λόγω της τυχαίας κίνησης του επιτιθέμενου.

β) Πυκνότητα $N = 125$

Σε πλήρη αντιστοιχία με το α), ακολουθεί διάγραμμα που απεικονίζει τα P_d, P_{in} και P_{ingat} συναρτήσει του l για μια μέση πυκνότητα δικτύου, δηλαδή για $N = 125$. Τα υπόλοιπα χαρακτηριστικά της προσομοίωσης έμειναν σταθερά και ίδια με αυτά της προηγούμενης περίπτωσης.

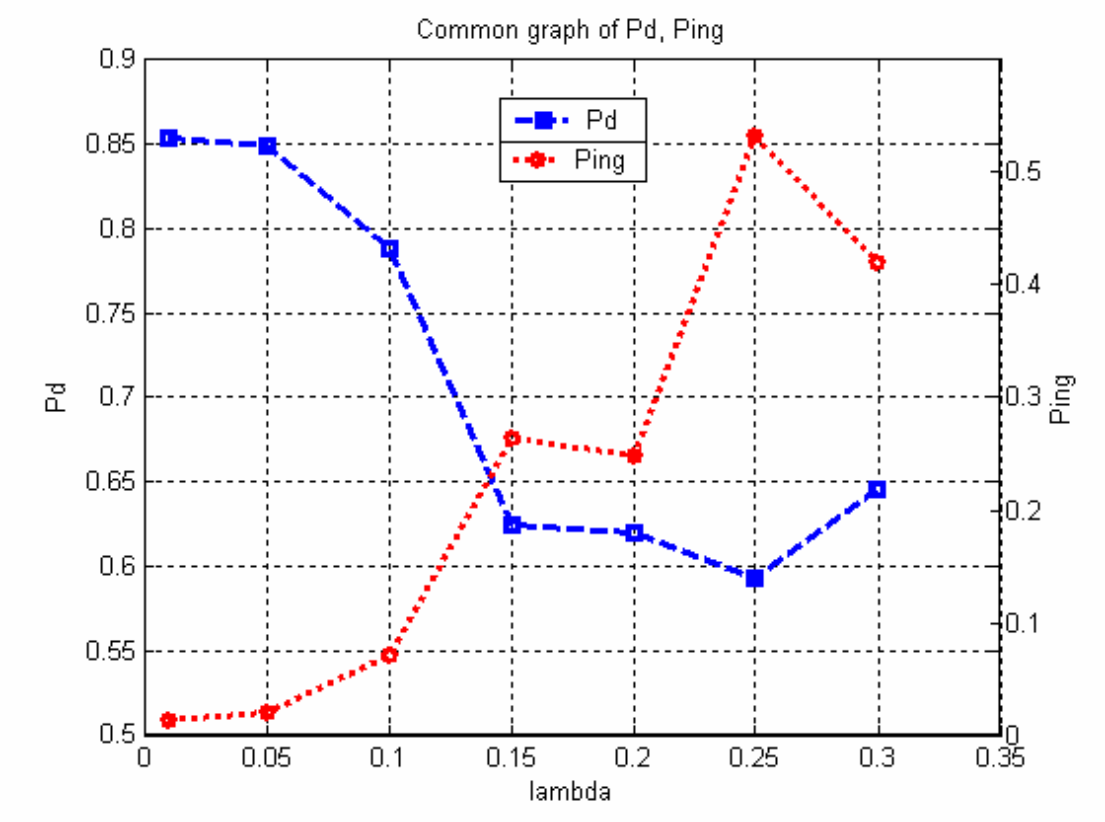


Σχήμα 12 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l

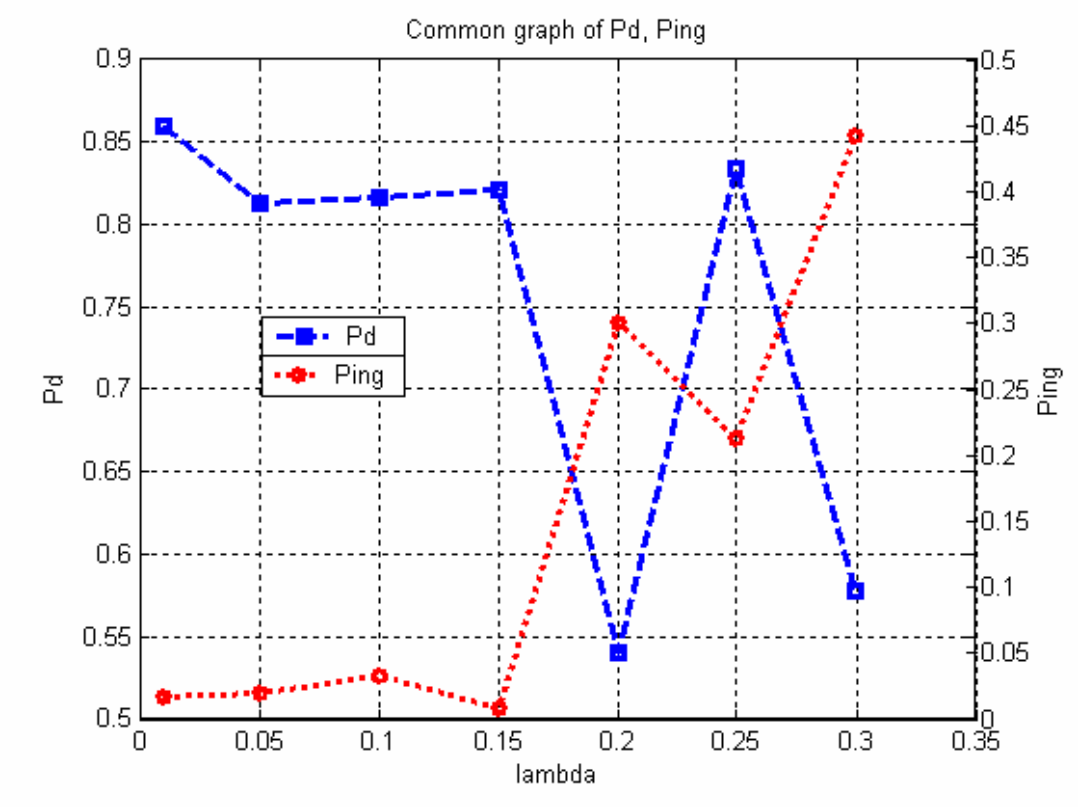
Παρατηρούμε ότι τα σχήματα 6 και 12 εμφανίζουν κοινά χαρακτηριστικά στις αντίστοιχες καμπύλες που απεικονίζουν. Αυτό γίνεται περισσότερο σαφές και εξετάζεται

πιο διεξοδικά παρακάτω, στην παράγραφο όπου παρατίθενται τα συγκεντρωτικά διαγράμματα της παρούσας σειράς προσομοιώσεων.

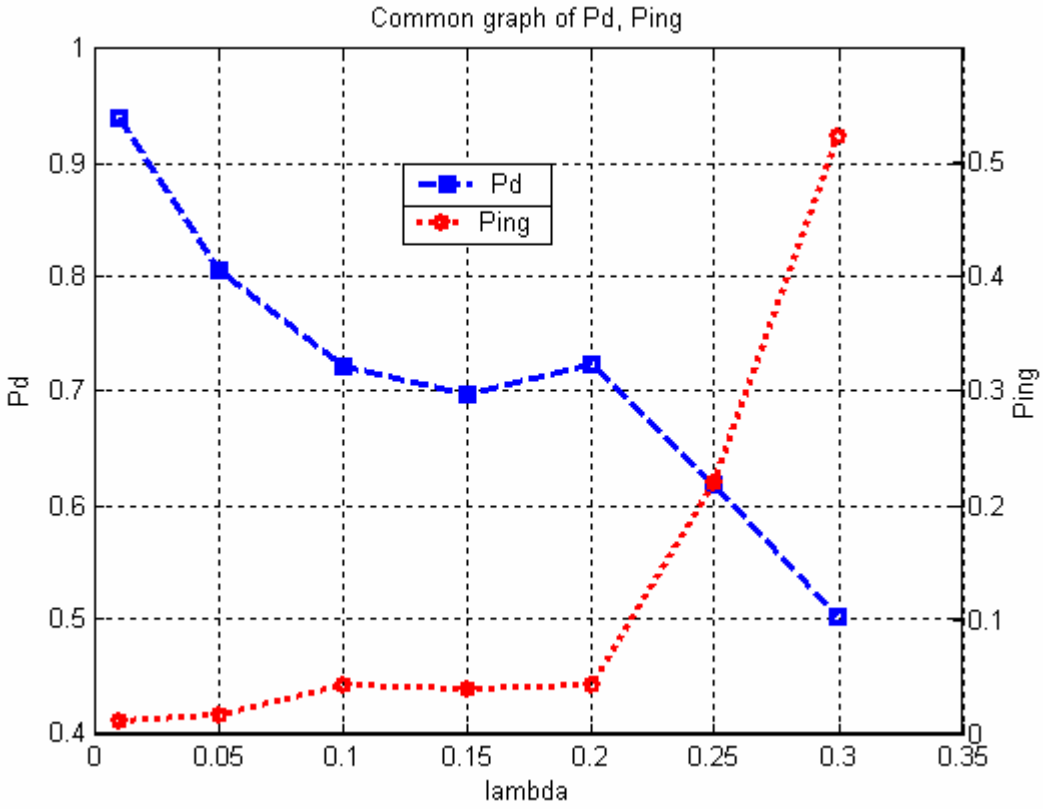
Ακολουθούν, όπως παραπάνω, αντιπροσωπευτικά γραφήματα του P_d και P_{ing} για διάφορα *group* του δικτύου.



Σχήμα 13 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 2



Σχήμα 14 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 6

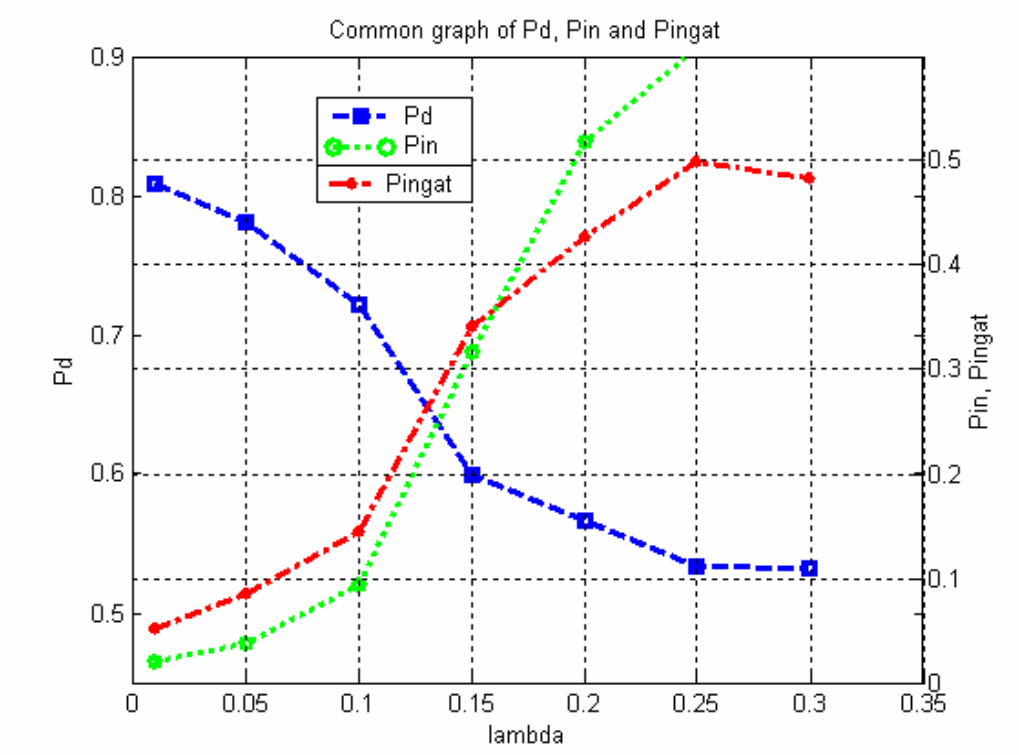


Σχήμα 15 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 7

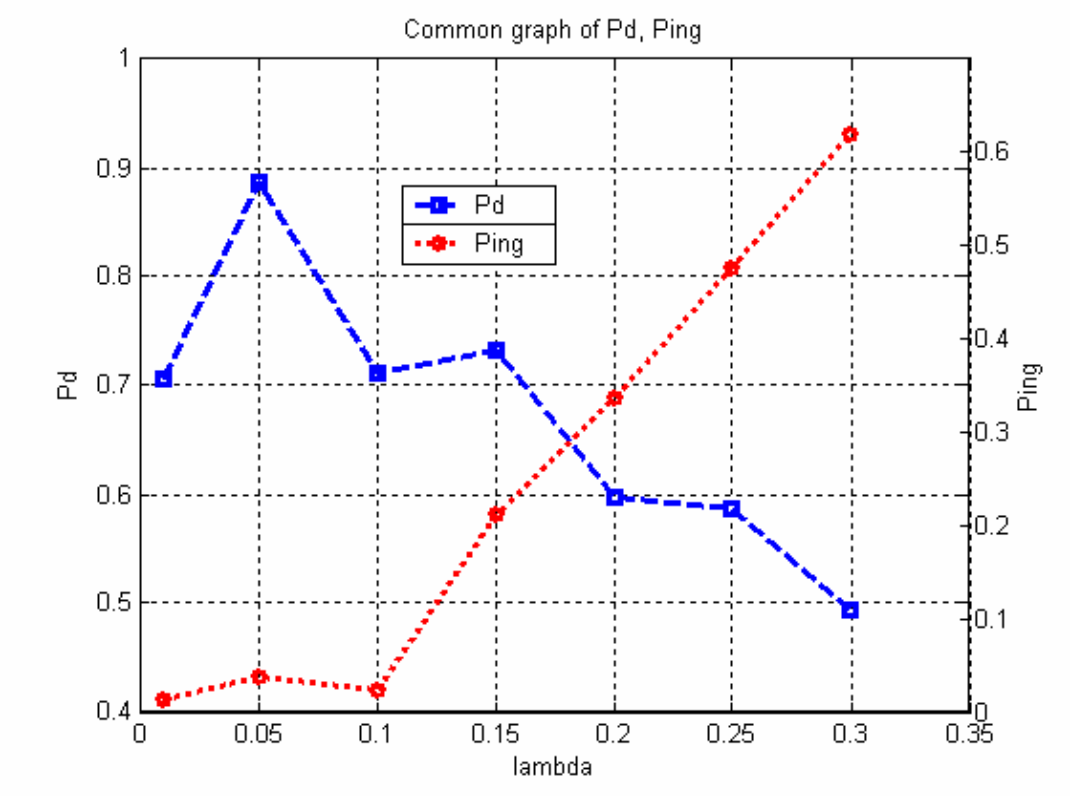
Τα σχήματα 13 – 15 έχουν κοινά γνωρίσματα με αυτά των σχημάτων 7 – 11 που δόθηκαν παραπάνω. Ξεχωριστό ενδιαφέρον παρουσιάζει το σχήμα 14 με το οποίο γίνεται εμφανές η αντίστροφη σχέση που υπάρχει μεταξύ των μεγεθών P_d και P_{ing}, αφού όταν αυξάνεται η τιμή του δεύτερου μειώνεται η τιμή του πρώτου και αντιστρόφως.

γ) Πυκνότητα $N = 150$

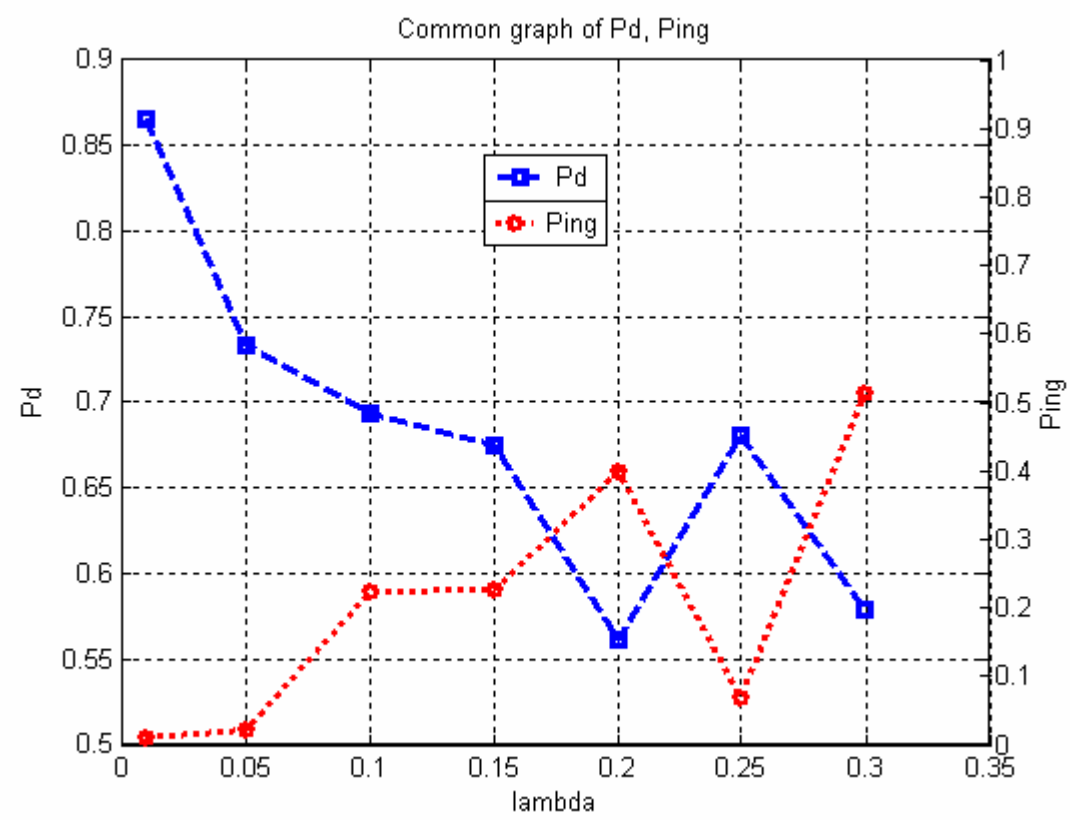
Τέλος, μετά την παράθεση του διαγράμματος των P_d, P_{in} και P_{ingat} συναρτήσει του l για την περίπτωση πυκνού δικτύου, δηλαδή για την περίπτωση που ο αριθμός των κόμβων του δικτύου είναι $N = 150$, ακολουθούν τα διαγράμματα των μεγεθών P_d και P_{ing} διαφόρων *group* του δικτύου τα οποία επιβεβαιώνουν αυτά που έχουν ειπωθεί παραπάνω και αφορούν στα αντίστοιχα διαγράμματα. Και σε αυτή τη σειρά πειραμάτων τα υπόλοιπα χαρακτηριστικά του δικτύου διατηρήθηκαν ίδια με παραπάνω.



Σχήμα 16 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l



Σχήμα 17 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 4



Σχήμα 18 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l του *group* 9

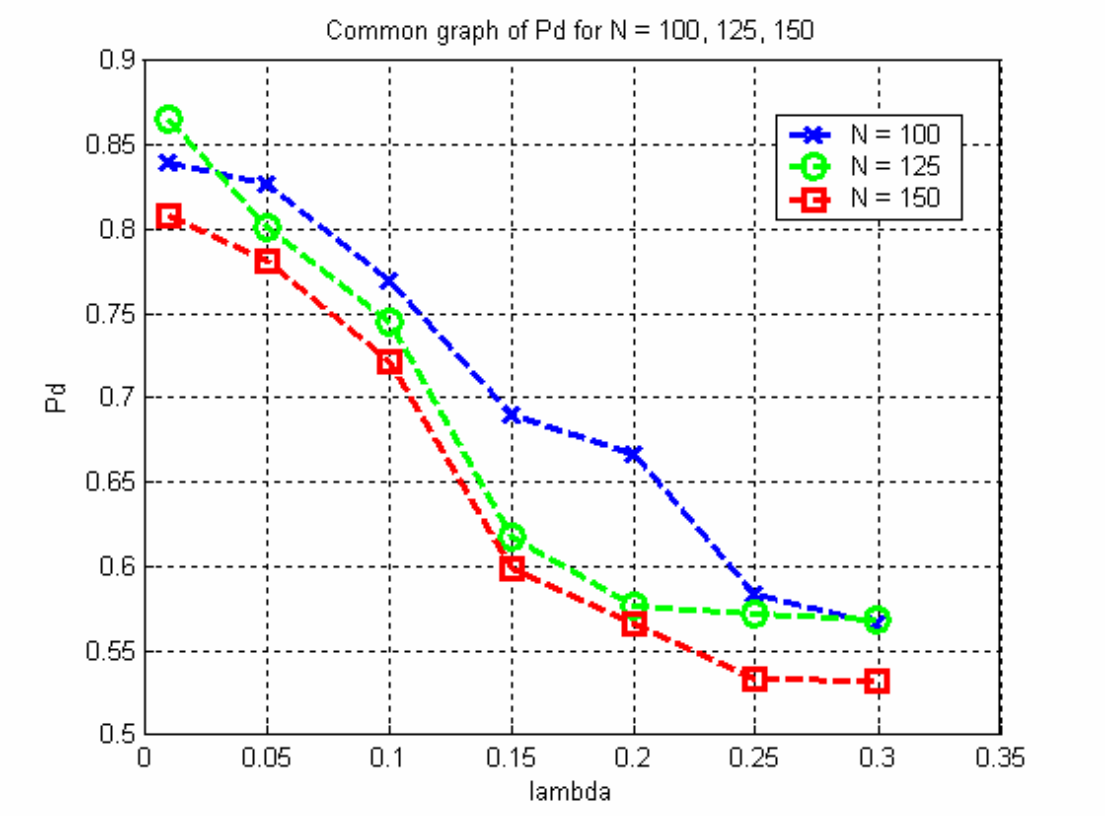
δ) Συγκεντρωτικά διαγράμματα

Στα σχήματα που ακολουθούν παρουσιάζεται το συγκριτικό διάγραμμα των τριών περιπτώσεων που εξετάστηκαν νωρίτερα. Από αυτό το διάγραμμα, γίνεται σαφές ότι η τιμή της πιθανότητας ανίχνευσης ανωμαλιών, P_d, μειώνεται με την αύξηση του ρυθμού μόλυνσης καναλιού μεταξύ των κόμβων του δικτύου, l. Με άλλα λόγια βλέπουμε ότι και για τις τρεις περιπτώσεις πυκνότητας δικτύου (αραιής, μέσης, πυκνής) η αποδοτικότητα του αλγόριθμου ακολουθεί γενικά πτωτική πορεία καθώς αυξάνεται το l, αφού η ανίχνευση των μολυσμένων κόμβων καθίσταται δυσκολότερη.

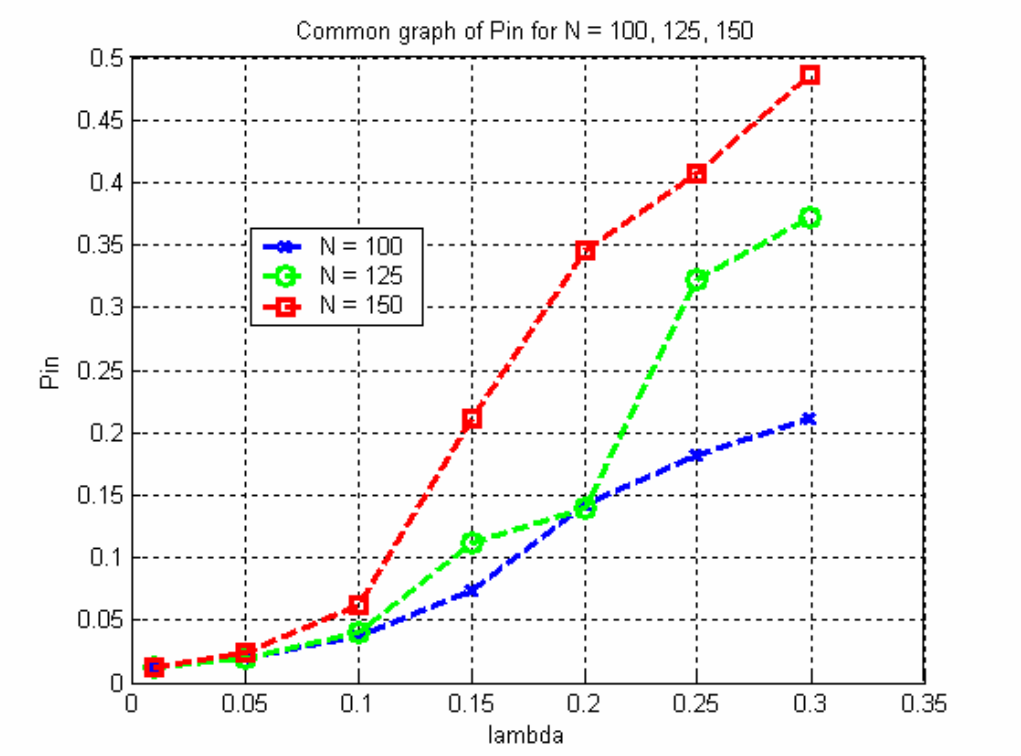
Το γεγονός αυτό μεταφράζεται σε κάτι το οποίο είναι το πλέον κρίσιμο για την αποδοτικότητα του αλγόριθμου. Βλέπουμε ότι η ικανότητα για ανίχνευση ανωμαλιών μειώνεται με τη γενική αυξητική τάση μολύνσεως, η οποία σημειώνεται με την αύξηση του P_{in}. Δηλαδή, όταν ο αριθμός των μολυσμένων κόμβων του δικτύου γίνεται μεγαλύτερος μειώνεται η δυνατότητα του αλγόριθμου για ανίχνευση ανωμαλιών. Το γεγονός αυτό οφείλεται στην εν γένει αδυναμία της τεχνικής, με την οποία πραγματοποιήθηκαν τα πειράματα, να ανιχνεύει ανωμαλίες πάνω από ένα αριθμητικό όριο. Το όριο αυτό, όπως σημειώσαμε και στο 4ο κεφάλαιο, συσχετίζεται άμεσα με το συνολικό αριθμό τον κόμβων που λαμβάνουν μέρος στον αλγόριθμο. Στα πλαίσια των προσομοιώσεων μας, ο αλγόριθμος εφαρμόστηκε για κάθε ένα *group* του χωρίου του δικτύου. Αυτό σημαίνει ότι η αποδοτικότητα του προτεινόμενου αλγόριθμου εξαρτάται τελικά από το συνολικό αριθμό των κόμβων κάθε *group* και από τον αριθμό των μολυσμένων κόμβων ανά χρονικό στιγμιότυπο σε αυτό το *group*. Το γεγονός της πτώσης της αποδοτικότητας του αλγόριθμου με την αύξηση του ποσοστού μόλυνσης των κόμβων βλέπουμε ότι ισχύει τόσο σε γενικό επίπεδο του δικτύου όσο και σε τοπικό επίπεδο για κάθε *group* κόμβων συνολικά. Μάλιστα, από τα τοπικά διαγράμματα των *group* του δικτύου βλέπουμε πιο καθαρά την αντίστροφη σχέση που υπάρχει μεταξύ των δύο αυτών μεγεθών, αφού αύξηση του P_{in} οδηγεί σε μείωση του P_d και αντιστρόφως. Επιπλέον σε

αρκετές περιπτώσεις στα τοπικά διαγράμματα των *group* παρατηρείται μία μεγάλη διακύμανση στις τιμές των μεγεθών που απεικονίζονται, η οποία οφείλεται στην στοχαστική φύση των πειραμάτων που πραγματοποιήθηκαν. Παράλληλα, από τα παρακάτω συγκεντρωτικά διαγράμματα, παρατηρούμε ότι η εν λόγω διακύμανση τιμών δεν εμφανίζεται καθώς λαμβάνονται αθροιστικά όλα τα *group* του δικτύου σε κάθε περίπτωση, παρέχοντας μας τη γενική τάση του συνόλου του δικτύου.

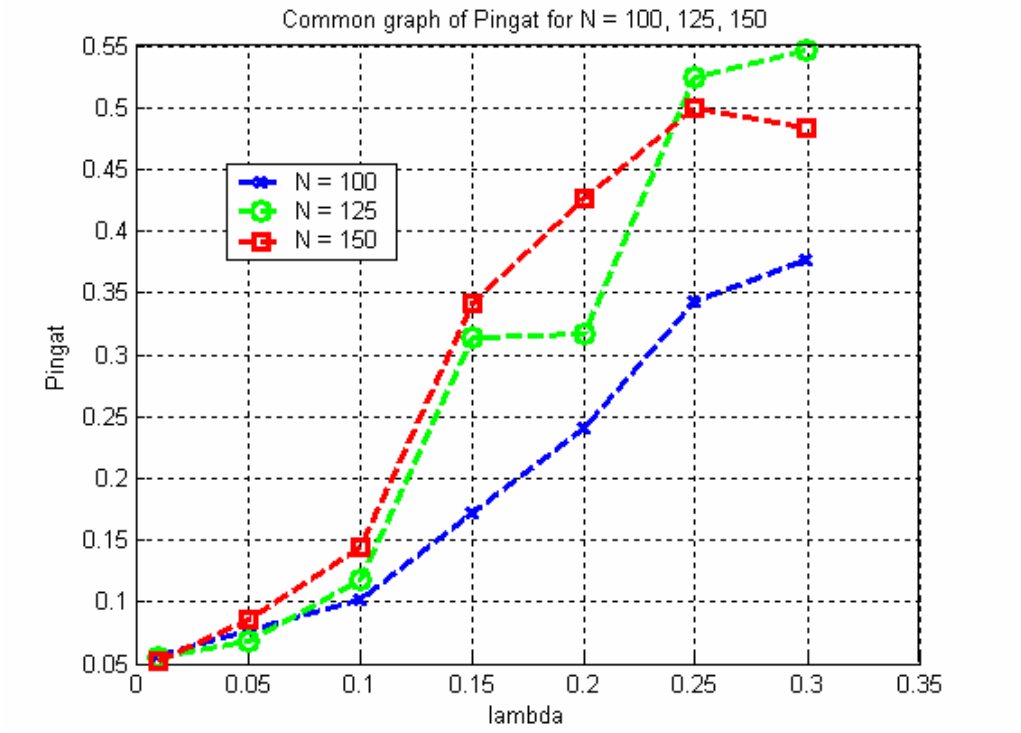
Για αυτή τη σειρά πειραμάτων, παρατηρείται γενικά, μια υψηλή απόδοση του εφαρμοζόμενου αλγόριθμου, αφού το P_d διατηρείται σε υψηλές τιμές. Η αδυναμία που ο αλγόριθμος παρουσιάζει έγκειται εκ κατασκευής στο γεγονός ότι αυτός μπορεί να ανιχνεύσει, με σχετικά μεγάλη ακρίβεια, μολύνσεις που λαμβάνουν χώρα σε ένα δίκτυο αισθητήρων οι οποίες είναι στα αρχικά στάδια ή δεν υπάρχει μεγάλος ρυθμός διάδοσης της μόλυνσης. Αντίθετα, η ικανότητα ανίχνευσης του αλγόριθμου μειώνεται σε περιπτώσεις εκτεταμένων μολύνσεων δικτύου.



Σχήμα 19 Συγκεντρωτικό διάγραμμα Πιθανότητας ανίχνευσης ανωμαλιών P_d συναρτήσει l



Σχήμα 20 Συγκεντρωτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων P_{in} συναρτήσει l



Σχήμα 21 Συγκεντρωτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων group επιτιθέμενου P_{ingat} συναρτήσει l

5.2 Μεταβάλλοντας τα χαρακτηριστικά του επιτιθέμενου

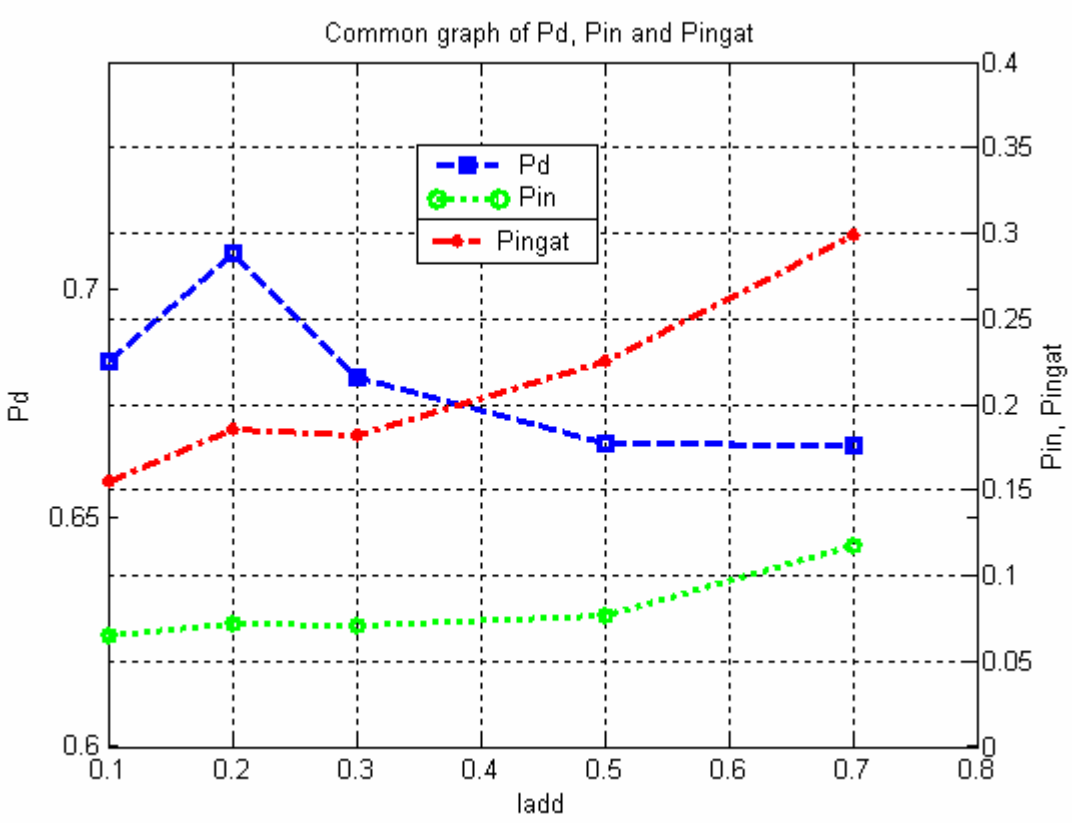
Σε αυτό το υποκεφάλαιο εξετάζουμε πως διαμορφώνεται η πιθανότητα ανίχνευσης ανωμαλιών P_d και των ποσοστών P_{in} και P_{ingat} για διάφορα σενάρια επίθεσης. Πραγματοποιήθηκαν διάφορες σειρές πειραμάτων στις οποίες μεταβάλλονται τα λειτουργικά χαρακτηριστικά του επιτιθέμενου σε αντίθεση με το προηγούμενο υποκεφάλαιο, στο οποίο αυτά παρέμειναν σταθερά.

5.2.1 Πιθανότητα ανίχνευσης ανωμαλιών συναρτήσει λ_{mal}

Εν συνεχεία παραθέτουμε πειράματα τα οποία πραγματοποιήθηκαν ώστε να εξεταστεί η μεταβολή του P_d συναρτήσει του ρυθμού μολύνσεως του κινούμενου επιτιθέμενου, l_{mal}.

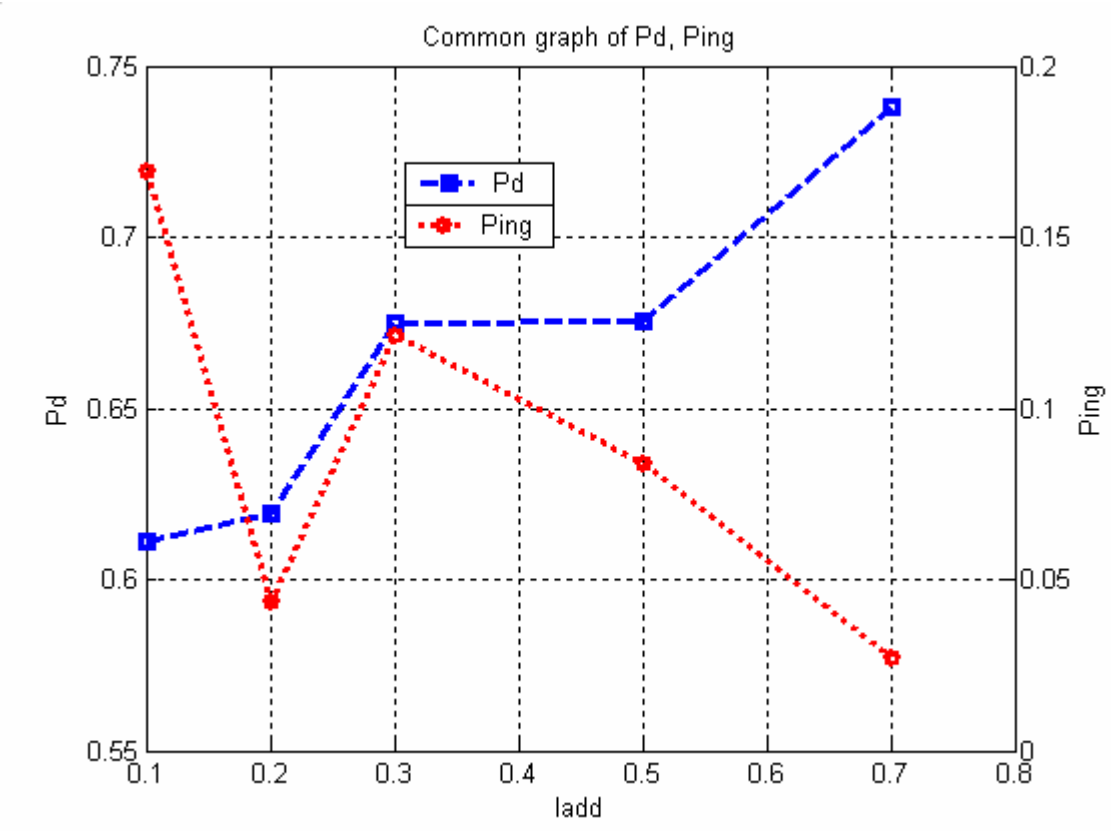
α) Πυκνότητα $N = 100$

Στο Σχήμα 22 που ακολουθεί, παρουσιάζονται οι διάφορες τιμές που λαμβάνει η πιθανότητα ανίχνευσης ανωμαλιών P_d, το P_{in} και το P_{ingat} συναρτήσει του l_{mal} για αραιό δίκτυο ($N = 100$). Υπενθυμίζεται ότι η ακτίνα του επιτιθέμενου και η ακτίνα μεταξύ των κόμβων διατηρήθηκαν σταθερές για τα πειράματα και ίσες με $R_{mal} = 200\,m$ και $R = 200\,m$ αντίστοιχα. Επιπλέον, η μέγιστη ταχύτητα του κακόβουλου κόμβου θεωρήθηκε σταθερή και ίση με $u_{\max} = 25\,m/tu$.



Σχήμα 22 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal}

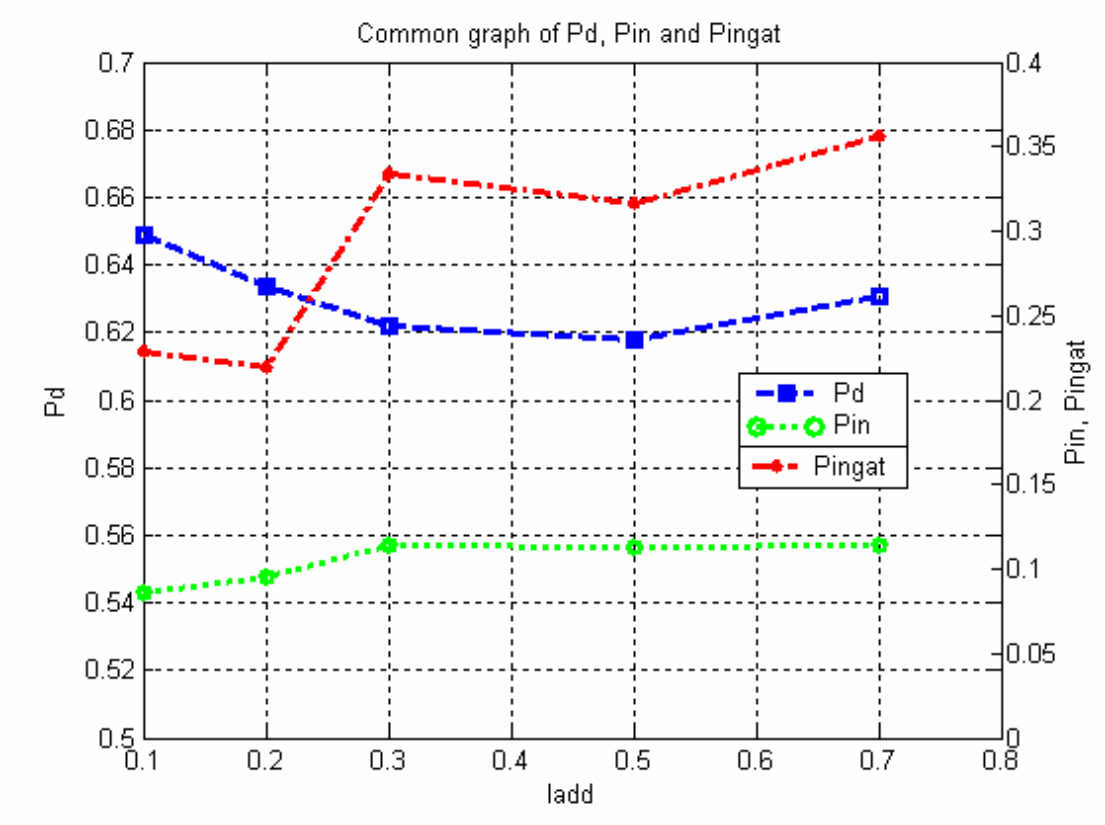
Ακολουθεί επίσης το χαρακτηριστικό διάγραμμα του P_d για το πέμπτο *group* του δικτύου.



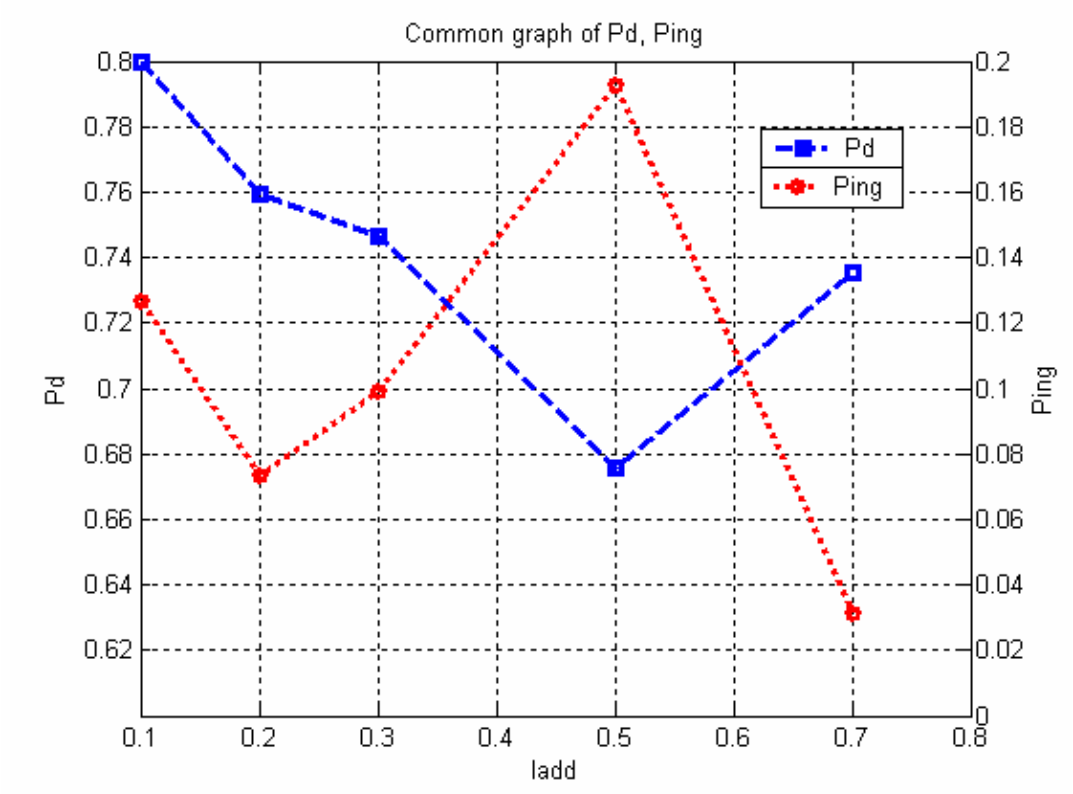
Σχήμα 23 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal} του *group* 5

β) Πυκνότητα $N = 125$

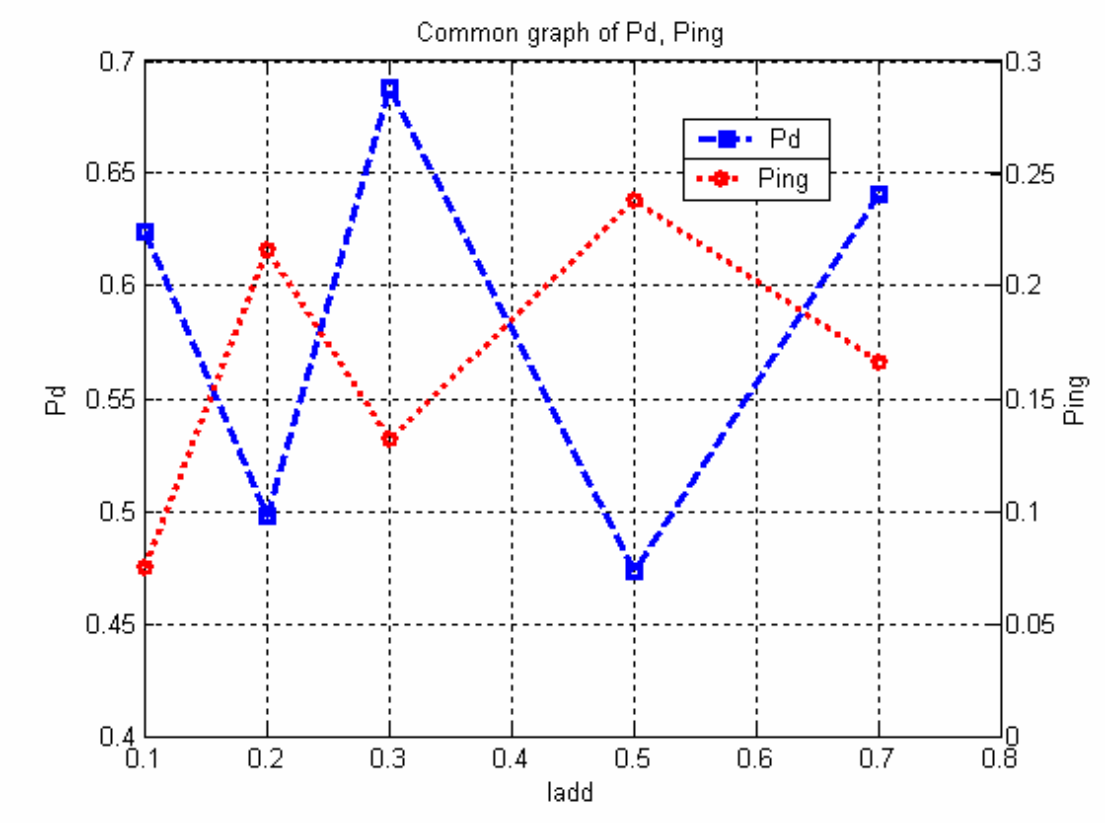
Στη συνέχεια παρουσιάζεται διάγραμμα που δείχνει τα P_d, P_{in}, P_{ingat} συναρτήσει του l_{mal} που αφορά στην περίπτωση της μέσης πυκνότητας δικτύου ($N = 125$). Τα υπόλοιπα χαρακτηριστικά της προσομοίωσης διατηρήθηκαν ίδια με την προηγούμενη περίπτωση. Επίσης, όμοια με παραπάνω δίνονται και τα διαγράμματα που αφορούν στα αντίστοιχα P_d και P_{in} κάποιων *group* του δικτύου.



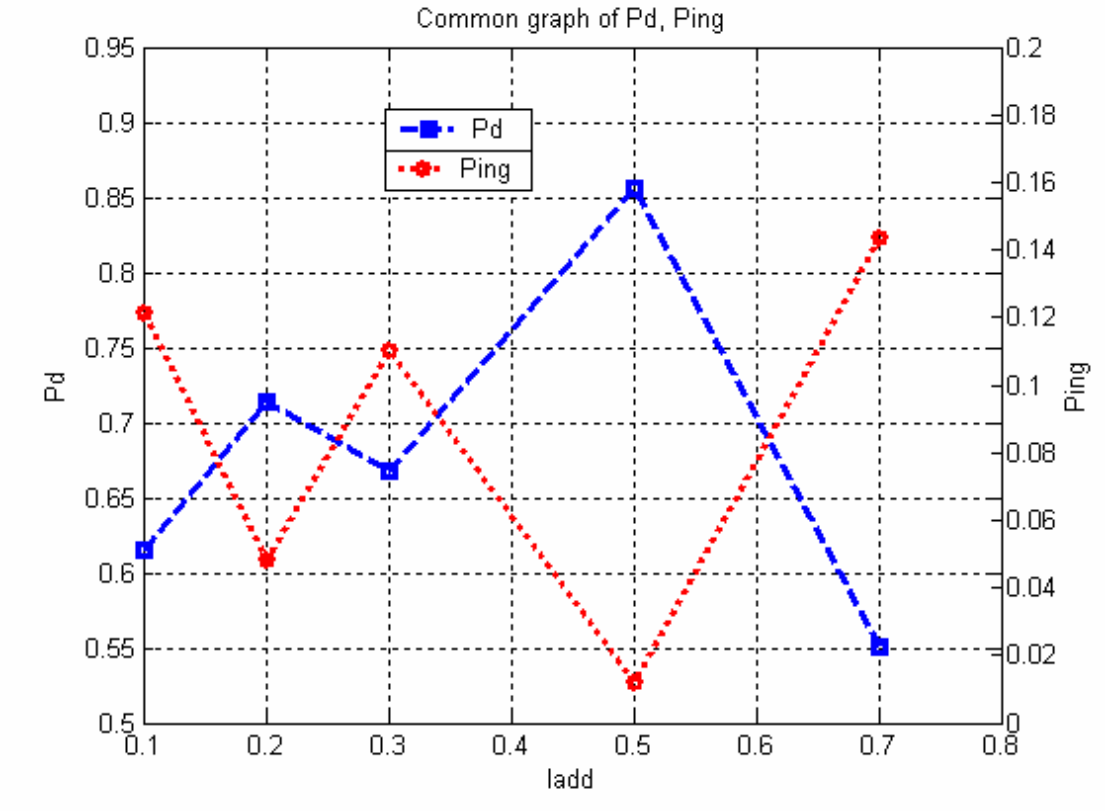
Σχήμα 24 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal}



Σχήμα 25 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal} του *group* 2



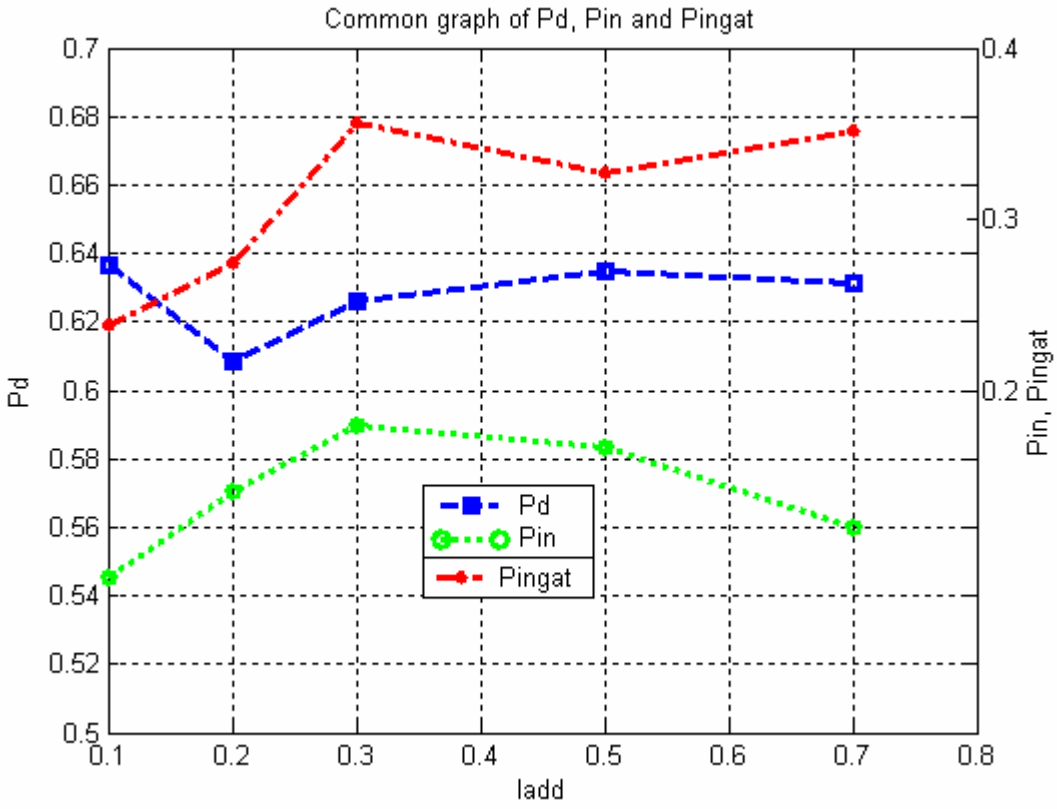
Σχήμα 26 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal} του *group* 4



Σχήμα 27 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal} του *group* 7

γ) Πυκνότητα $N = 150$

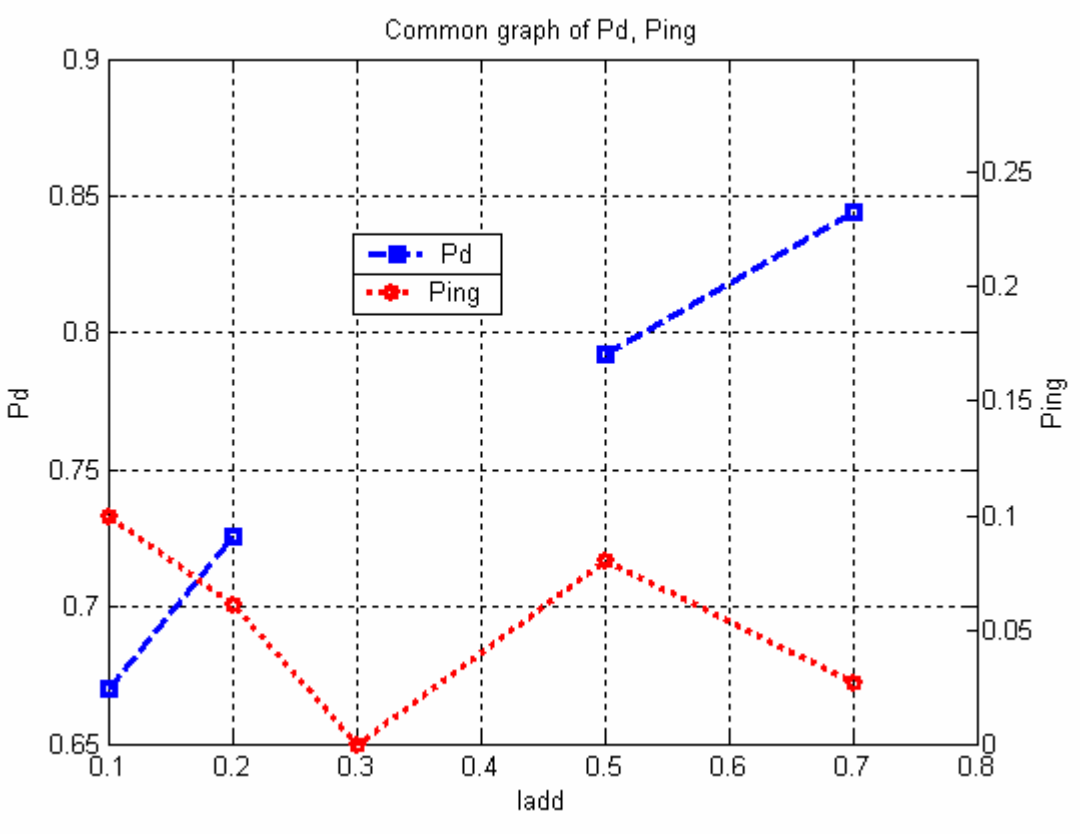
Το Σχήμα 28 δείχνει τη μεταβολή των P_d, P_{in} και P_{ingat} συναρτήσει του l_{mal} για πυκνό δίκτυο ($N = 150$). Επαναλαμβάνουμε ότι και στην παρούσα σειρά πειραμάτων τα υπόλοιπα χαρακτηριστικά του δικτύου παρέμειναν σταθερά και ίδια με παραπάνω.



Σχήμα 28 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal}

Τέλος, πριν προχωρήσουμε στην απεικόνιση και το σχολιασμό των συγκεντρωτικών διαγραμμάτων της παρούσας σειράς πειραμάτων, παραθέτουμε το διάγραμμα του P_d και P_{ing} του ένατου *group* όπου διαπιστώνουμε ότι για $l_{mal} = 0.3$ το P_d δεν εμφανίζει τιμή. Αυτό γίνεται διότι στην περίπτωση αυτή δεν υπάρχει κάποια μόλυνση κόμβου στο *group* 9 και συνεπώς ο παρανομαστής του P_d ισούται με μηδέν (*βλ.* εξ. 16) Άρα, το P_d κινείται στο άπειρο και εν τέλει δεν αντιπροσωπεύεται από πραγματικό αριθμό σε αυτήν την περίπτωση του πειράματος. Ωστόσο, παρατηρούμε, όπως σε όλα τα

διαγράμματα που αφορούν τοπικά τα *group* του δικτύου, ότι διατηρείται η αντίστροφη σχέση που υπάρχει μεταξύ των μεγεθών P_d και P_{ing}. Στο παρακάτω, αλλά και στα προηγούμενα διαγράμματα, κυριαρχεί το γεγονός ότι μείωση της τιμής του ενός μεγέθους οδηγεί σε αύξηση της τιμής του άλλου και αντίστροφα.

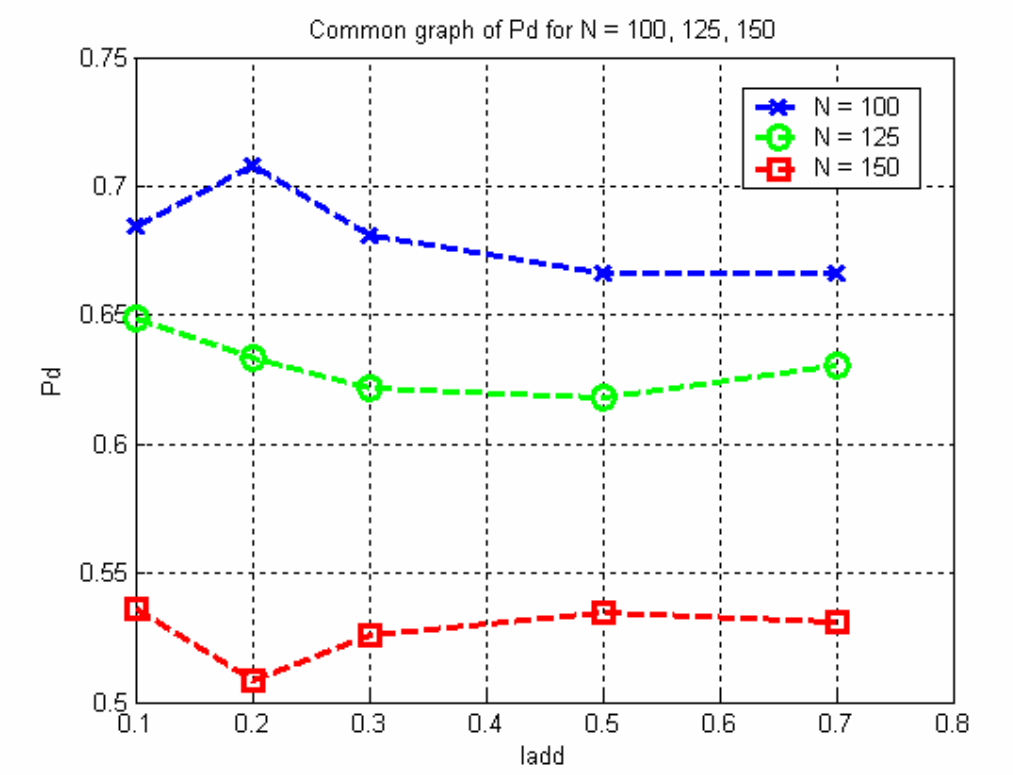


Σχήμα 29 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal} του *group* 9

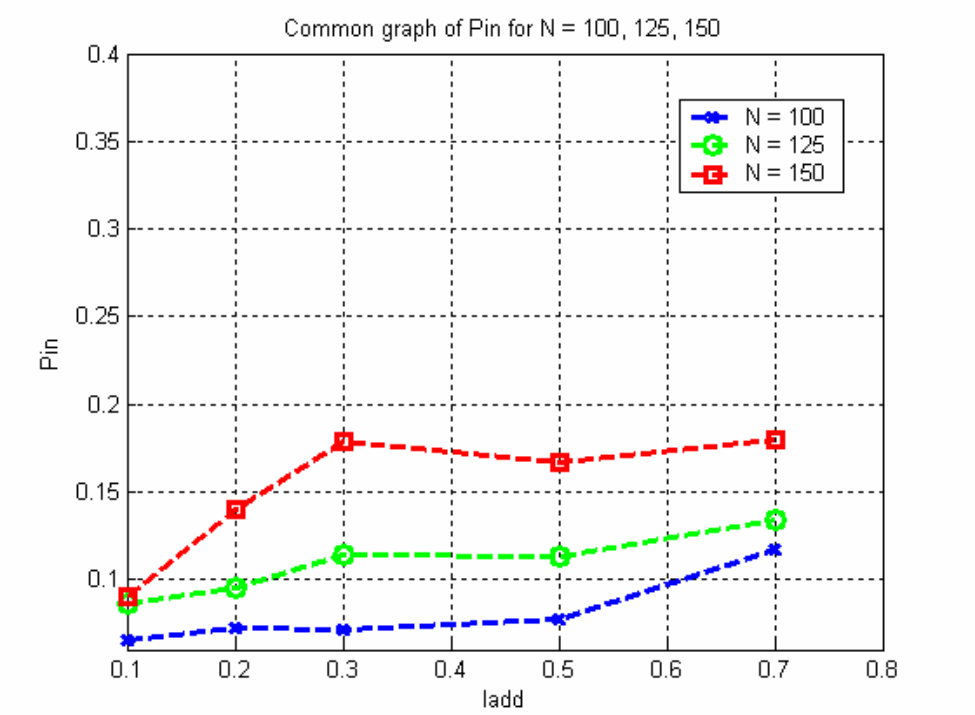
δ) Συγκεντρωτικά διαγράμματα

Τα σχήματα που ακολουθούν απεικονίζουν σε κοινό διάγραμμα τα μεγέθη P_d, P_{in} και P_{ing} για τις τρεις περιπτώσεις πυκνότητας δικτύου $N = 100$, 125 και 150. Με το συνδυασμό των διαγραμμάτων αυτών βλέπουμε και εδώ, όπως έγινε και με την προηγούμενη σειρά πειραμάτων, ότι η ανιχνευτική δυνατότητα του αλγόριθμου, που αντιπροσωπεύεται από υψηλό P_d, εξαρτάται με άμεσο τρόπο από την μολυσματική ικανότητα του επιτιθέμενου. Πράγματι, στα παρακάτω διαγράμματα παρατηρούμε ότι για

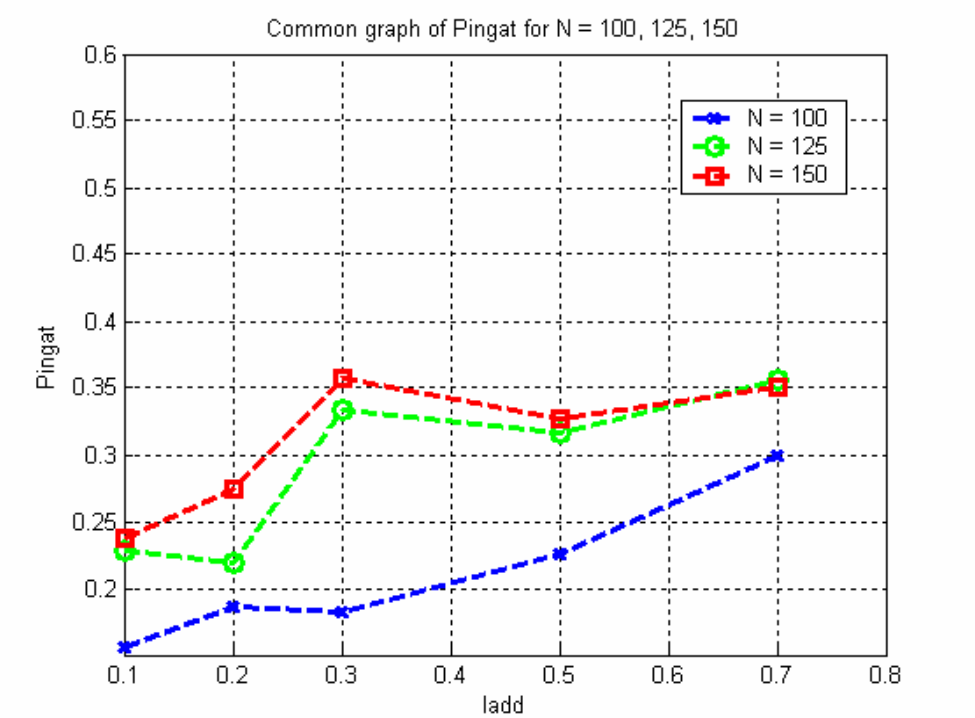
μεγαλύτερο P_{in} λαμβάνουμε μικρότερο P_d. Το γεγονός αυτό μάλιστα ισχύει για όλες τις τιμές του l_{mal} για μία συγκεκριμένη πυκνότητα κόμβων δικτύου. Δηλαδή παρατηρούμε ότι το P_d για $N=100$ είναι για όλες τις τιμές του l_{mal} μικρότερο με το αντίστοιχο P_d που λαμβάνουμε για $N=125$. Το ίδιο συμβαίνει για το συνδυασμό $N=125$ και 150. Η διαμόρφωση των διαφορών αυτών επιβάλλεται από το διαφορετικό μολυσματικό επίπεδο του δικτύου το οποίο διαπιστώνεται από το συγκεντρωτικό διάγραμμα του P_{in}. Από το τελευταίο βλέπουμε ότι π.χ. για $N=150$ και για όλες τις τιμές l_{mal} το P_{in} κυμαίνεται σε υψηλότερες τιμές από τα αντίστοιχα P_{in} για τις δύο άλλες περιπτώσεις πυκνοτήτων. Αντίστροφα τώρα, το P_d για $N=150$ κυμαίνεται, πάλι για όλες τις τιμές του l_{mal} σε χαμηλότερες τιμές από ότι για $N=100$ ή 125. Ανάλογα πράγματα ισχύουν και για τους υπόλοιπους συνδυασμούς πυκνοτήτων.



Σχήμα 30 Συγκεντρωτικό διάγραμμα πιθανότητας ανίχνευσης ανωμαλιών P_d συναρτήσει l_{mal}



Σχήμα 31 Συγκεντρωτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων P_{in} συναρτήσει l_{mal}



Σχήμα 32 Συγκεντρωτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων *group* επιτιθέμενου P_{ingat} συναρτήσει l_{mal}

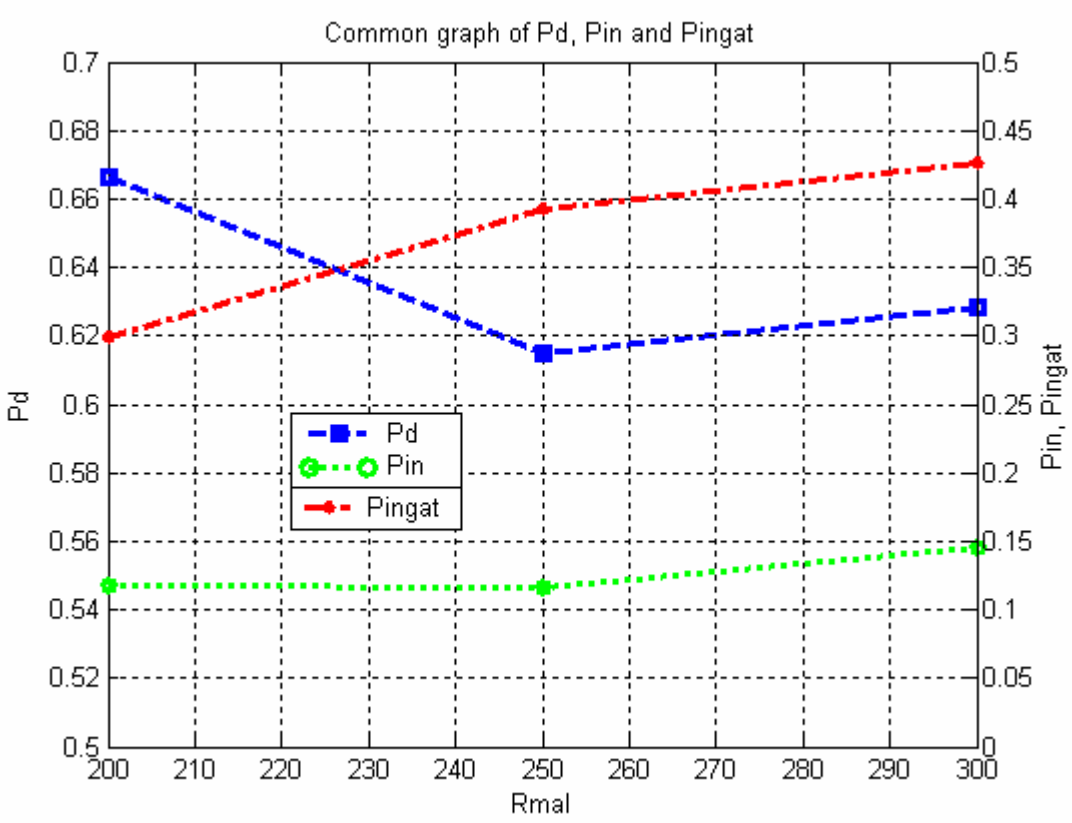
5.2.2 Πιθανότητα ανίχνευσης ανωμαλιών συναρτήσει R_{mal}

Όμοια με προηγουμένως, στο παρόν υποκεφάλαιο εξετάζουμε τη μεταβολή του P_d για διάφορα σενάρια μολύνσεως όταν μεταβάλλεται η ακτίνα του επιτιθέμενου R_{mal}. Με την τρέχουσα σειρά πειραμάτων θέλουμε να διερευνήσουμε τον τρόπο με τον οποίο επηρεάζεται το P_d μέσω του R_{mal} για διάφορες περιπτώσεις πυκνοτήτων δικτύου.

α) Πυκνότητα $N = 100$

Στο Σχήμα 33 που ακολουθεί απεικονίζονται οι τιμές της πιθανότητας ανίχνευσης ανωμαλιών P_d μαζί με τα P_{in} και P_{ingat} για διάφορες τιμές της ακτίνας του επιτιθέμενου R_{mal} για $N = 100$ (αραιό δίκτυο). Η ακτίνα μεταξύ των κόμβων του δικτύου ελήφθη ίση με $R = 200\,m$ και η μέγιστη ταχύτητα του κακόβουλου κόμβου σταθερή και ίση με $u_{\max} = 25\,m/tu$.

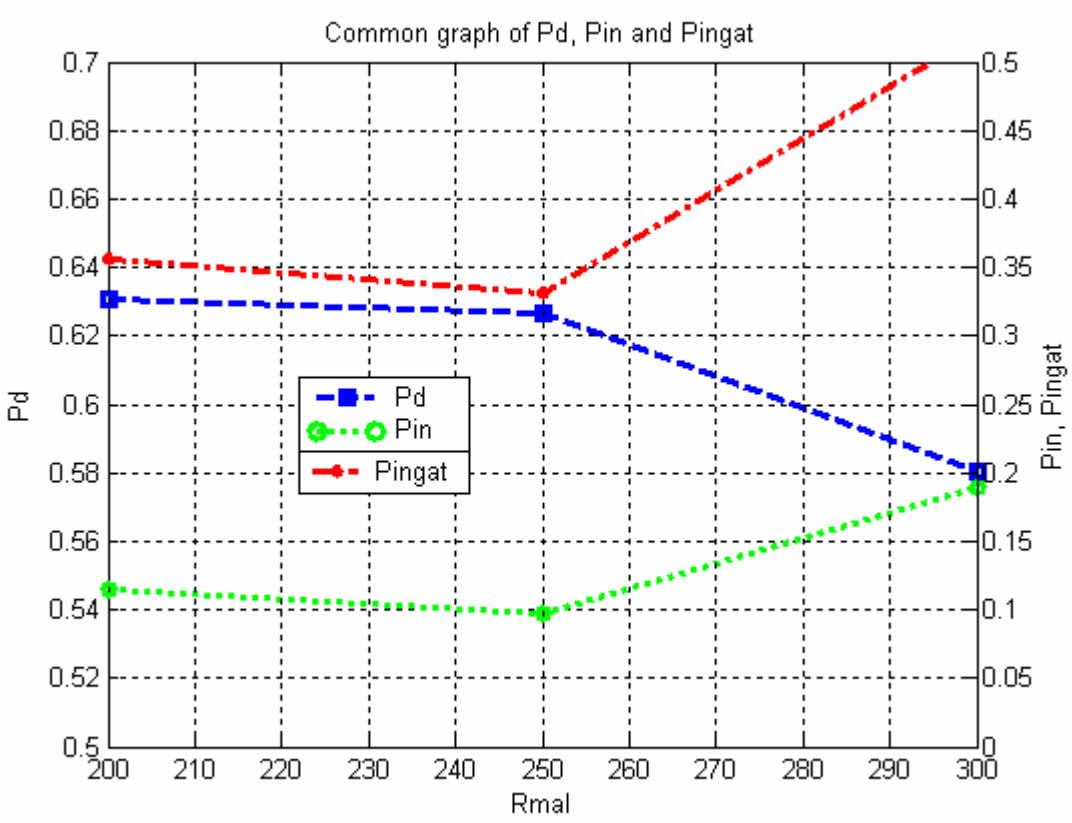
Οι τιμές των P_d, P_{in} και P_{ingat} βρέθηκαν πραγματοποιώντας τρία πειράματα για $R_{mal} = 200\,m$, $R_{mal} = 250\,m$ και $R_{mal} = 300\,m$. Στο σχήμα που ακολουθεί βλέπουμε ότι το P_{ingat} λαμβάνει και για τα τρία πειράματα υψηλότερες τιμές απ' ότι το P_{in}. Με άλλα λόγια η περιοχή που κινείται ο κακόβουλος κόμβος μολύνεται σε μεγαλύτερο βαθμό συγκριτικά με ολόκληρη την περιοχή του δικτύου. Επιπλέον, για μεγαλύτερη ακτίνα R_{mal} του κινούμενου κόμβου μολύνεται μεγαλύτερος αριθμός κόμβων του δικτύου και κατά συνέπεια παρατηρείται μείωση του P_d.



Σχήμα 33 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}

β) Πυκνότητα $N=125$

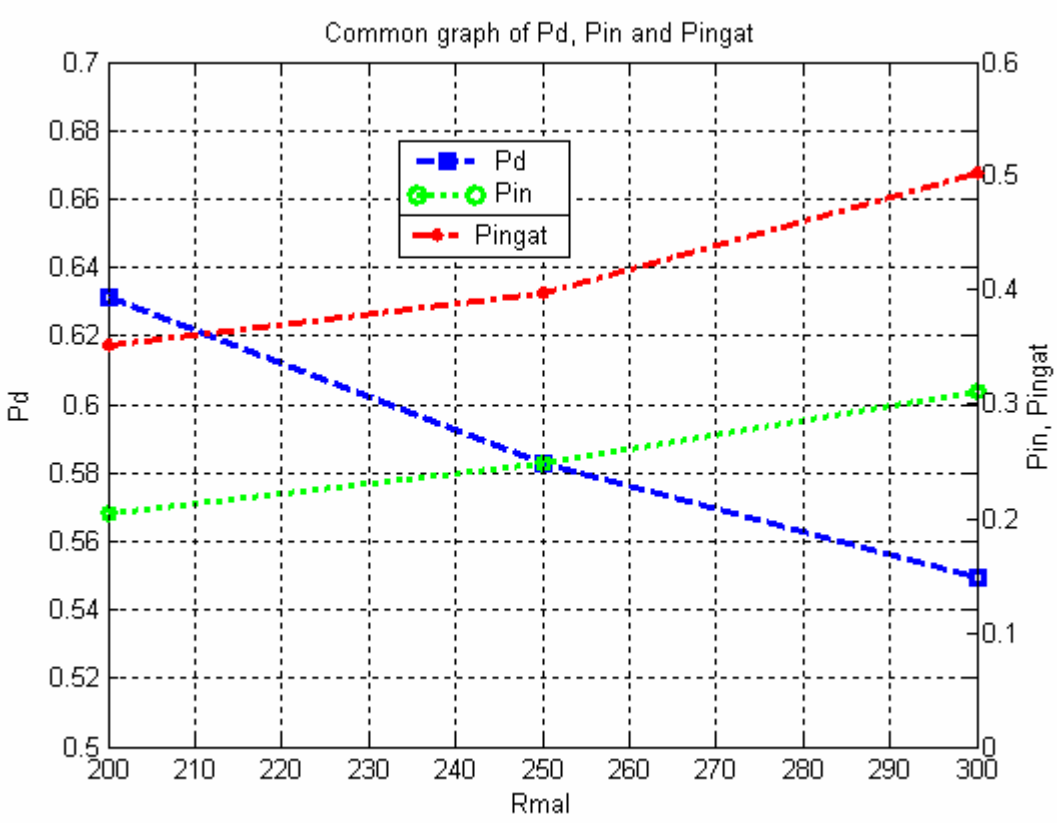
Όμοια με την προηγούμενη περίπτωση στο επόμενο σχήμα παρουσιάζεται η πιθανότητα ανίχνευσης ανωμαλιών P_d, P_{in} και P_{ingat} για διάφορες τιμές της ακτίνας του επιτιθέμενου R_{mal} για $N=125$ (μεσαίας πυκνότητας δίκτυο). Επισημαίνεται ότι το σχήμα που ακολουθεί παρουσιάζει ίδια ποιοτικά χαρακτηριστικά με αυτά του σχήματος 33. Η διαφορά αυτών των δύο σχημάτων έγκειται στο γεγονός ότι η αύξηση του R_{mal} για μεγαλύτερες πυκνότητες επιφέρει μεγαλύτερες αυξήσεις στα μεγέθη P_{in} και P_{ingat} και παρεπόμενα οδηγεί σε μεγαλύτερες μειώσεις του P_d.



Σχήμα 34 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}

γ) Πυκνότητα $N = 150$

Τέλος, το Σχήμα 35 που ακολουθεί, δίνει την πιθανότητα ανίχνευσης ανωμαλιών P_d, P_{in} και P_{ingat} συναρτήσει του R_{mal} για $N = 150$ (πυκνό δίκτυο). Όπως ειπώθηκε για τις προηγούμενες περιπτώσεις πυκνότητας δικτύων, η αύξηση του R_{mal} σε συνδυασμό με μεγαλύτερο N οδηγεί σε μεγαλύτερη αύξηση των P_{in} και P_{ingat} και επομένως μεγαλύτερη μείωση του P_d. Το γεγονός αυτό επιβεβαιώνεται και σε αυτήν την περίπτωση με το σχήμα που ακολουθεί. Παρατηρούμε ότι η μείωση της τιμής του P_d για αυτήν την περίπτωση πυκνότητας δικτύου είναι μεγαλύτερη σε σχέση με τις δύο προηγούμενες.



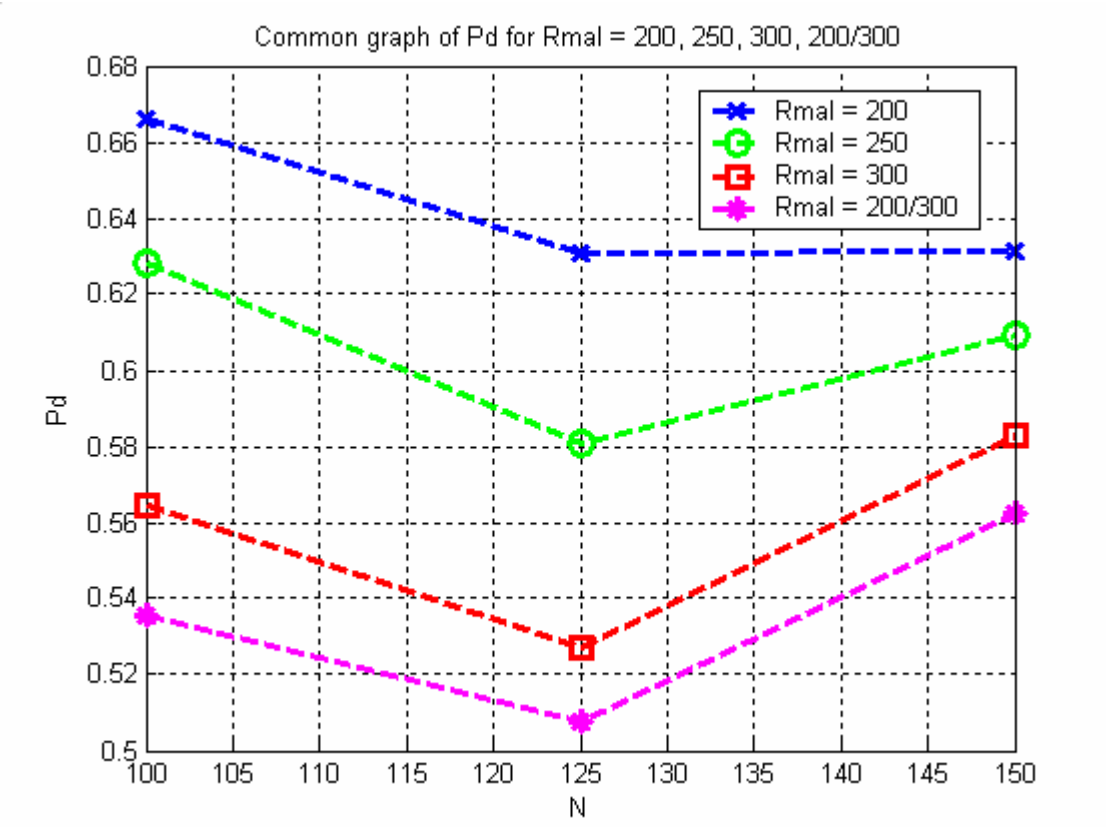
Σχήμα 35 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}

Και σε αυτή τη σειρά πειραμάτων διαφαίνεται η σχέση που υπάρχει μεταξύ των P_d και P_{in} γεγονός που έχει διαπιστωθεί και από τα πειράματα που προηγήθηκαν. Βέβαια σε αυτόν τον κύκλο πειραμάτων αυτή η σχέση δεν γίνεται απόλυτα εμφανής γιατί όπως παρατηρούμε δεν υπάρχει σημαντική διαφοροποίηση της τιμής του P_{in}. Παρόλα αυτά παραμένει αξιοσημείωτο το γεγονός ότι η τάση στις τιμές του P_d καθορίζεται από τις αντίστοιχες τάσεις των τιμών του P_{in}. Επιπλέον, αξιοσημείωτο είναι και το ίδιο το γεγονός ότι για τις τρεις διαφορετικές τιμές R_{mal} (όλα τα υπόλοιπα στοιχεία δικτύου και επιτιθέμενου δεν μεταβλήθηκαν) και ξεχωριστά για το κάθε πείραμα, οι μολύνσεις του δικτύου διακυμάνθηκαν στο ίδιο περίπου επίπεδο. Περισσότερο αντιπροσωπευτικά στοιχεία δίνονται με την επόμενη σειρά πειραμάτων.

5.2.3 Πιθανότητα ανίχνευσης ανωμαλιών συναρτήσει N για διαφορετικούς συνδυασμούς R_{mal}

Με αυτή τη σειρά πειραμάτων, εξετάζεται η μεταβολή του P_d στις στρατηγικές επίθεσης για διάφορες ακτίνες του επιτιθέμενου R_{mal} συναρτήσει της πυκνότητας του δικτύου. Το Σχήμα 36 που ακολουθεί δίνει συγκεντρωτικά τις τιμές που λαμβάνει το P_d για R_{mal} ίση με $200\,m$, $250\,m$, $300\,m$ και $200/300\,m$. Με τον τελευταίο συμβολισμό εννοούμε ότι ενώ αρχικά η ακτίνα του επιτιθέμενου είναι $300\,m$, όταν κατά την κίνηση του τελευταίου μέσα στο δίκτυο τα αποθέματα της ενέργειάς του πέσουν κάτω από το μισό των αρχικών τους τιμών τότε η ακτίνα του μειώνεται σε $200\,m$.

Στο παρακάτω διάγραμμα (Σχήμα 36), επαναβεβαιώνεται η διαπίστωση των προηγούμενων πειραμάτων, το γεγονός δηλαδή ότι τα επίπεδα μόλυνσης του δικτύου παίζουν κυρίαρχο ρόλο στην ανιχνευτική ικανότητα του προτεινόμενου αλγόριθμου. Από το διάγραμμα αυτό βλέπουμε ότι οι τιμές του P_d διατηρούνται σε υψηλότερα επίπεδα για την περίπτωση ακτίνας επιτιθέμενου ίση με $R_{mal} = 200\,m$ για κάθε πυκνότητα δικτύου. Μπορούμε, ακόμα, να παρατηρήσουμε ότι όσο αυξάνεται η ακτίνα του επιτιθέμενου τόσο περισσότερο μειώνονται οι τιμές του P_d, δηλαδή η δυνατότητα ανίχνευσης του αλγόριθμου. Τέλος, παρατηρούμε ότι για το συνδυασμό ακτίνας $200/300\,m$ το P_d κυμαίνεται στα χαμηλότερα επίπεδα τιμών.



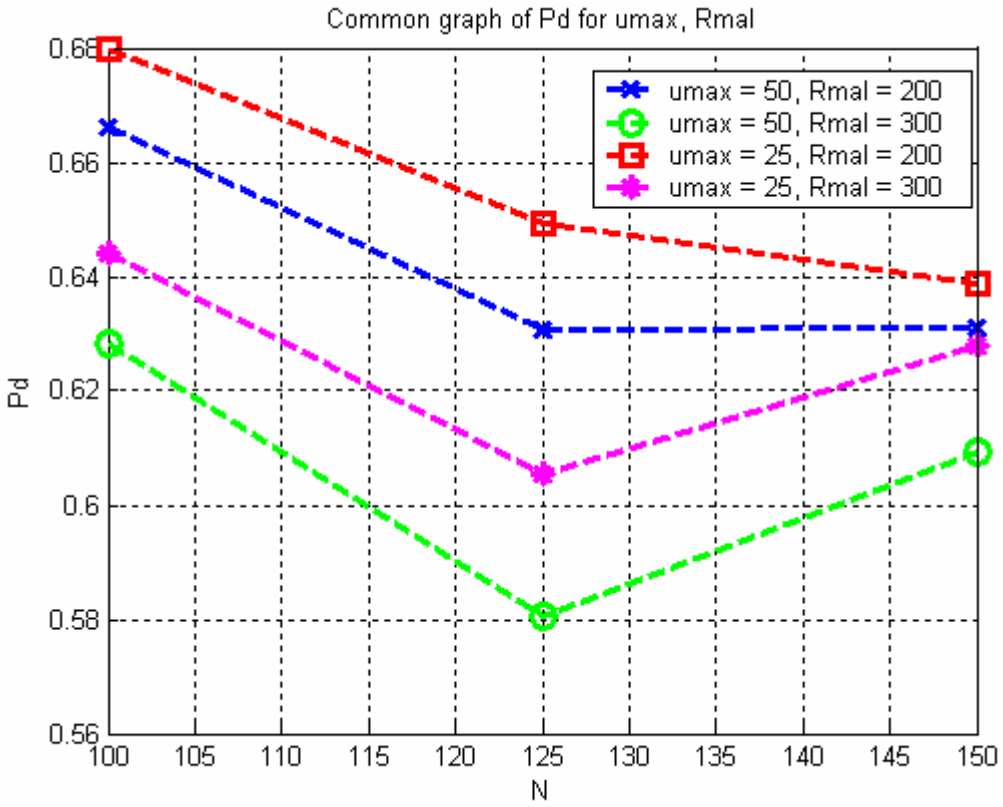
Σχήμα 36 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}

5.2.4 Πιθανότητα ανίχνευσης ανωμαλιών συναρτήσει *N*

Σε αυτό το υποκεφάλαιο εξετάζεται η εξάρτηση που υπάρχει ανάμεσα στην πιθανότητα ανίχνευσης ανωμαλιών P_d με δύο λειτουργικά χαρακτηριστικά μεγέθη του κακόβουλου κόμβου, της ακτίνας R_{mal} και της μέγιστης ταχύτητας του $u_{\max}$.

Πραγματοποιήθηκαν δύο ομάδες πειραμάτων. Η πρώτη περιλαμβάνει πειράματα για διάφορες τιμές πυκνότητας δικτύου ($N = 100, 125$ και 150) για $R_{mal} = 200\,m$ και $R_{mal} = 300\,m$. Στη δεύτερη σειρά πειραμάτων διατηρήθηκε ο ίδιος συνδυασμός πυκνότητας δικτύου N και ακτίνας R_{mal} ενώ μεταβλήθηκε μόνο η μέγιστη ταχύτητα του επιτιθέμενου $u_{\max}$, η οποία ελήφθη ίση με το μισό της ταχύτητας που θεωρήσαμε στην πρώτη ομάδα πειραμάτων.

Στο επόμενο διάγραμμα (Σχήμα 37) απεικονίζονται οι τιμές της πιθανότητας ανίχνευσης ανωμαλιών P_d για διάφορους συνδυασμούς της ακτίνας του επιτιθέμενου και της μέγιστης ταχύτητάς του για $N = 100, 125, 150$. Επίσης, για λόγους πληρότητας, δίνεται παρακάτω και το αντίστοιχο γράφημα του P_{in} (Σχήμα 38).



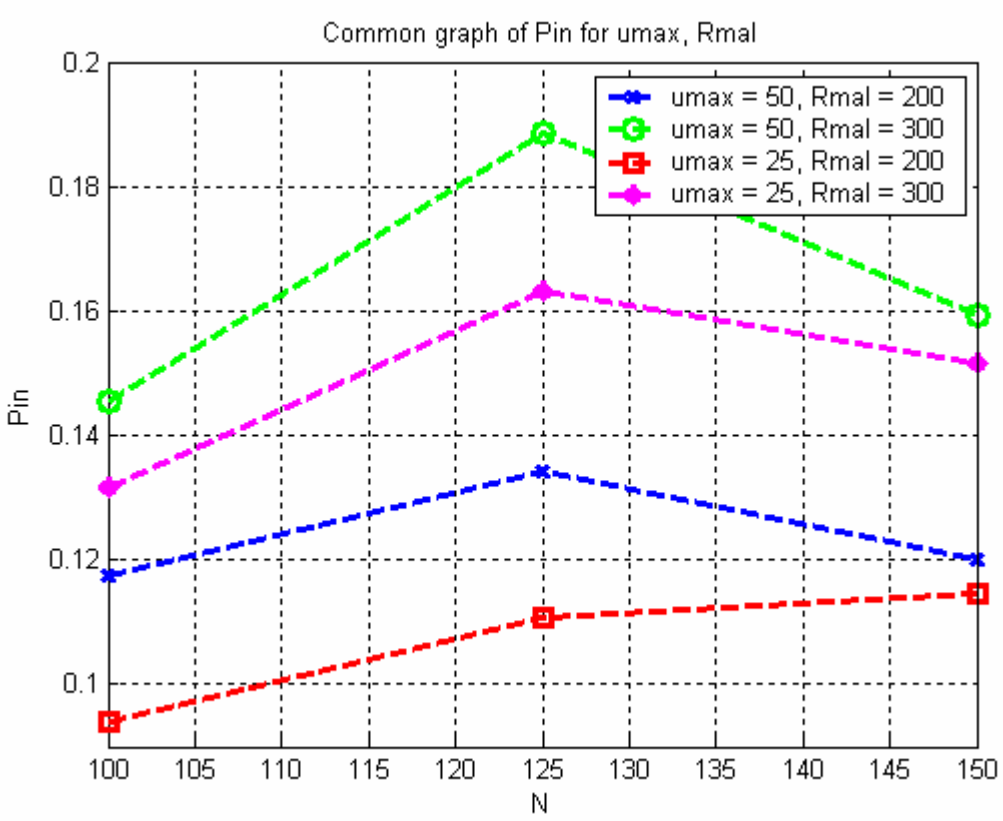
Σχήμα 37 Συγκεντρωτικό διάγραμμα πιθανότητας ανίχνευσης ανωμαλιών P_d συναρτήσει N

Από το παραπάνω (Σχήμα 37) αλλά και το επόμενο (Σχήμα 38) συγκριτικό διάγραμμα γίνεται σαφές ότι η πιθανότητα ανίχνευσης ανωμαλιών P_d είναι μέγεθος αντιστρόφως ανάλογο της μέγιστης ταχύτητας του κινούμενου επιτιθέμενου κόμβου. Πράγματι, από τα διαγράμματα αυτά βλέπουμε ότι το P_d λαμβάνει μικρότερες τιμές για μεγαλύτερη μέγιστη ταχύτητα επιτιθέμενου $u_{\max}$ σε σχέση με τις τιμές που αυτό λαμβάνει όταν το $u_{\max}$ έχει το μισό μέτρο.

Και για τις δύο περιπτώσεις ακτίνας του επιτιθέμενου R_{mal} (για $R_{mal} = 200\,m$ και $R_{mal} = 300\,m$) βλέπουμε ότι το P_d είναι μικρότερο για μεγαλύτερες τιμές ταχύτητας επιτιθέμενου $u_{\max}$. Αυτό σημαίνει ότι η ικανότητα που έχει ο αλγόριθμος για ανίχνευση ανωμαλιών μειώνεται με την αύξηση της ταχύτητας του επιτιθέμενου. Αυτό οφείλεται στο γεγονός ότι όταν ο επιτιθέμενος έχει μεγαλύτερες ταχύτητες μολύνει περισσότερους κόμβους δικτύου και άρα δυσχεραίνει τη διαδικασία ανίχνευσης ανωμαλιών. Η ταχύτερη διάδοση μόλυνσης στους κόμβους του δικτύου είναι ένας ακόμα παράγοντας που περιορίζει την ικανότητα ανίχνευσης του προτεινόμενου αλγόριθμου.

Διευκρινίζουμε ότι ως μέγιστη ταχύτητα του κακόβουλου κόμβου $u_{\max}$ εννοούμε το μέγιστο μέτρο της ταχύτητας που δυνητικά μπορεί να λάβει ο επιτιθέμενος σε κάθε χρονικό στιγμιότυπο της προσομοίωσης. Δηλαδή, το $u_{\max}$ δεν έχει την έννοια μιας σταθερής ταχύτητας του επιτιθέμενου καθ' όλη τη διάρκεια των πειραμάτων. Μέσω του $u_{\max}$ ορίζεται το εύρος τιμών της ταχύτητας του επιτιθέμενου που τυχαία εκλέγεται στην αρχή κάθε χρονικού στιγμιότυπου της προσομοίωσης.

Σημαντικό, βέβαια, είναι και το γεγονός ότι η τιμές του P_d για $R_{mal} = 200\,m$, ανεξαρτήτως ταχύτητας, κυμαίνονται σταθερά σε μεγαλύτερα επίπεδα από τις αντίστοιχες τιμές που λαμβάνει για $R_{mal} = 300\,m$, πράγμα τελείως συμβατό με το αντίστοιχο συγκεντρωτικό διάγραμμα που παρουσιάστηκε στο προηγούμενο υποκεφάλαιο.



Σχήμα 38 Συγκεντρωτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων P_{in} συναρτήσει N

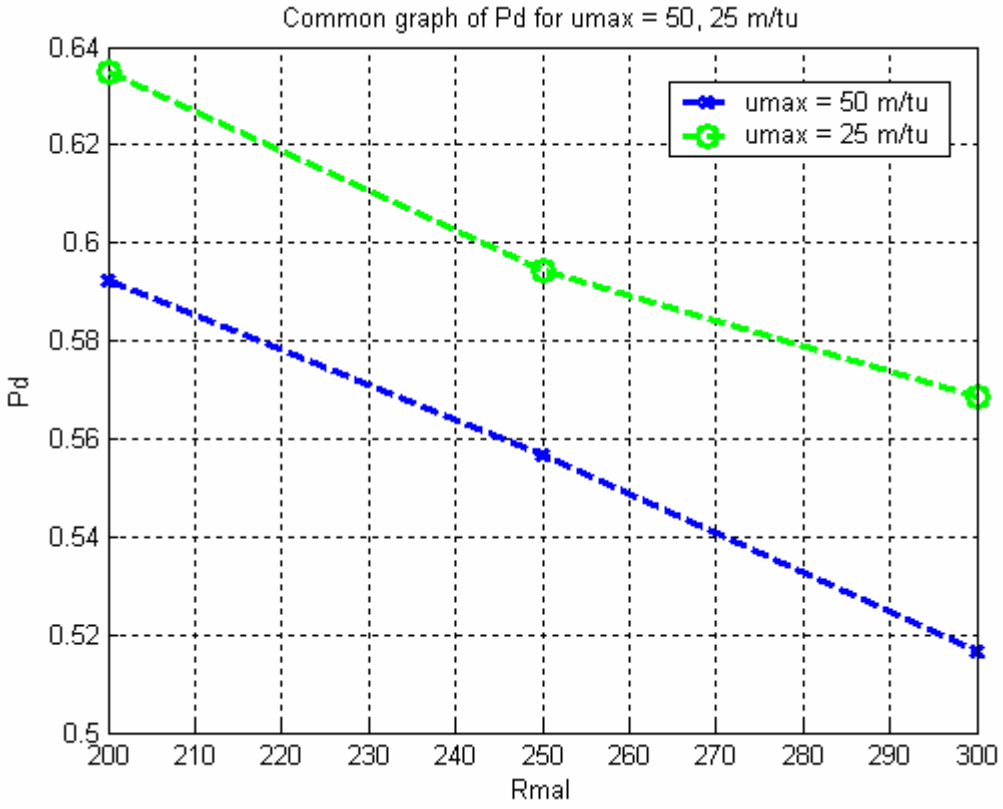
5.2.5 Πιθανότητα ανίχνευσης ανωμαλιών συναρτήσει R_{mal} για διαφορετικούς συνδυασμούς u_{max}

Στο υποκεφάλαιο αυτό παρατίθενται περιπτώσεις πειραμάτων για διαφορετικές τιμές ρυθμού μόλυνσης δικτύου και επιτιθέμενου, l και l_{mal} αντιστοίχως. Πραγματοποιήθηκαν δύο ομάδες πειραμάτων, στην πρώτη ομάδα ελήφθη ρυθμός μόλυνσης δικτύου ίσος με $l = 0.15$ ενώ στη δεύτερη ίσος με το $l = 0.075$, δηλαδή ίσος με το μισό της τιμής της πρώτης ομάδας πειραμάτων. Τα αποτελέσματα των πειραμάτων ελήφθησαν για διάφορες τιμές ακτίνας επιτιθέμενου R_{mal} (για $R_{mal} = 200\,m$, $R_{mal} = 250\,m$ και $R_{mal} = 300\,m$).

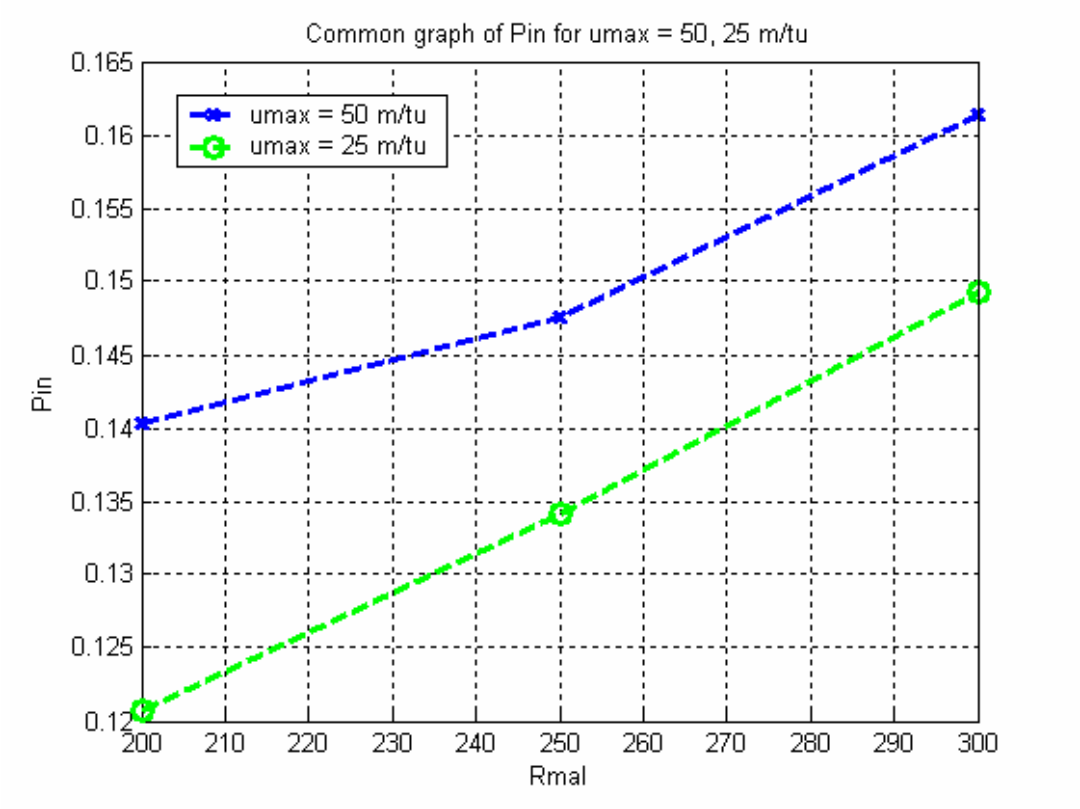
5.2.5.1 Διαγράμματα λ_{mal} και u_{max} συναρτήσει R_{mal}

Ακολούθως, δίνονται τα διαγράμματα για την πρώτη σειρά πειραμάτων που πραγματοποιήθηκαν. Τα πρώτα διαγράμματα είναι για $l = 0.15$ και $l_{mal} = 0.5$ ενώ τα δεύτερα για ίδιο l και $l_{mal} = 0.8$ για διάφορες τιμές ακτίνας R_{mal}. Αυτά τα πειράματα έγιναν για δύο ταχύτητες, μία με μέτρο $u_{\max} = 50\, m/tu$ και μία με $u_{\max} = 25\, m/tu$.

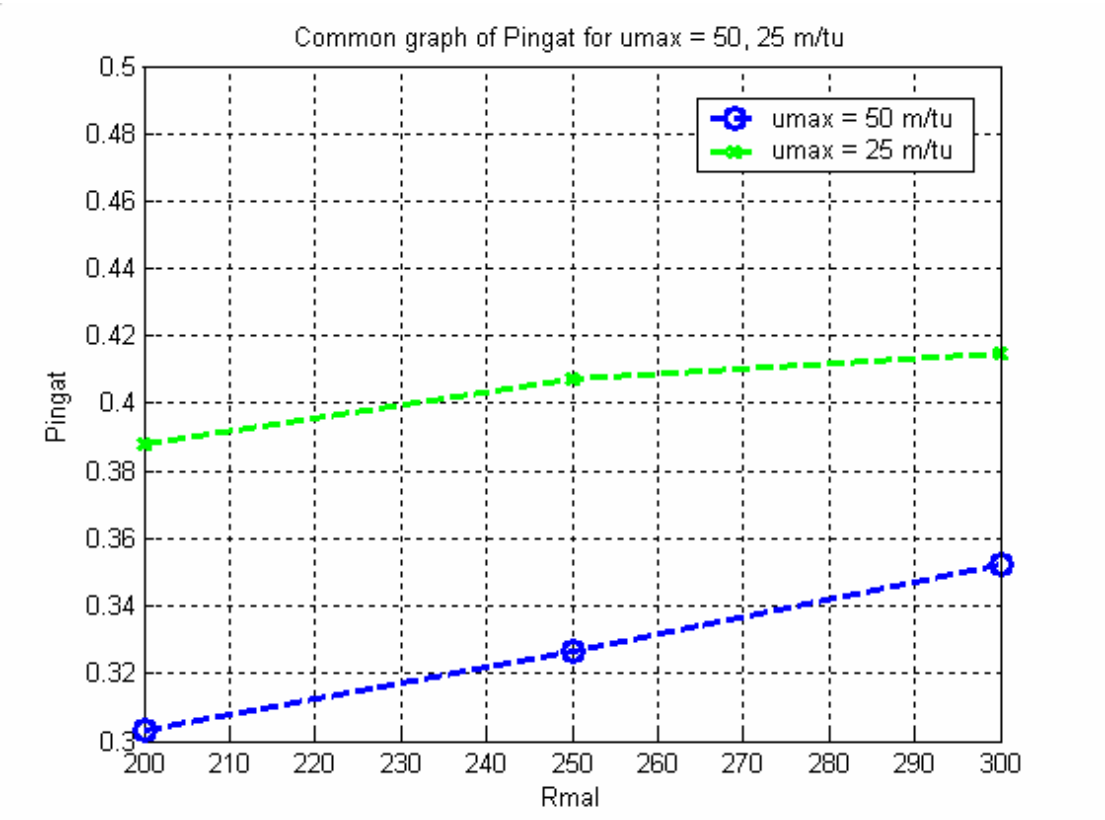
α) $l = 0.15,\ l_{mal} = 0.5$



Σχήμα 39 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}



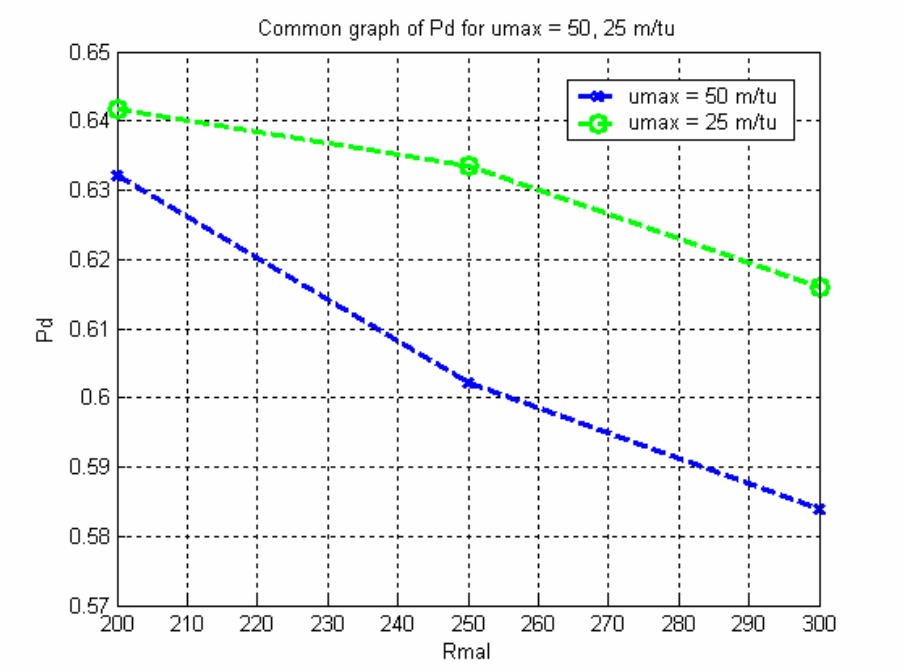
Σχήμα 40 Μέσο ποσοστό μολυσμένων κόμβων P_{in} συναρτήσει R_{mal}



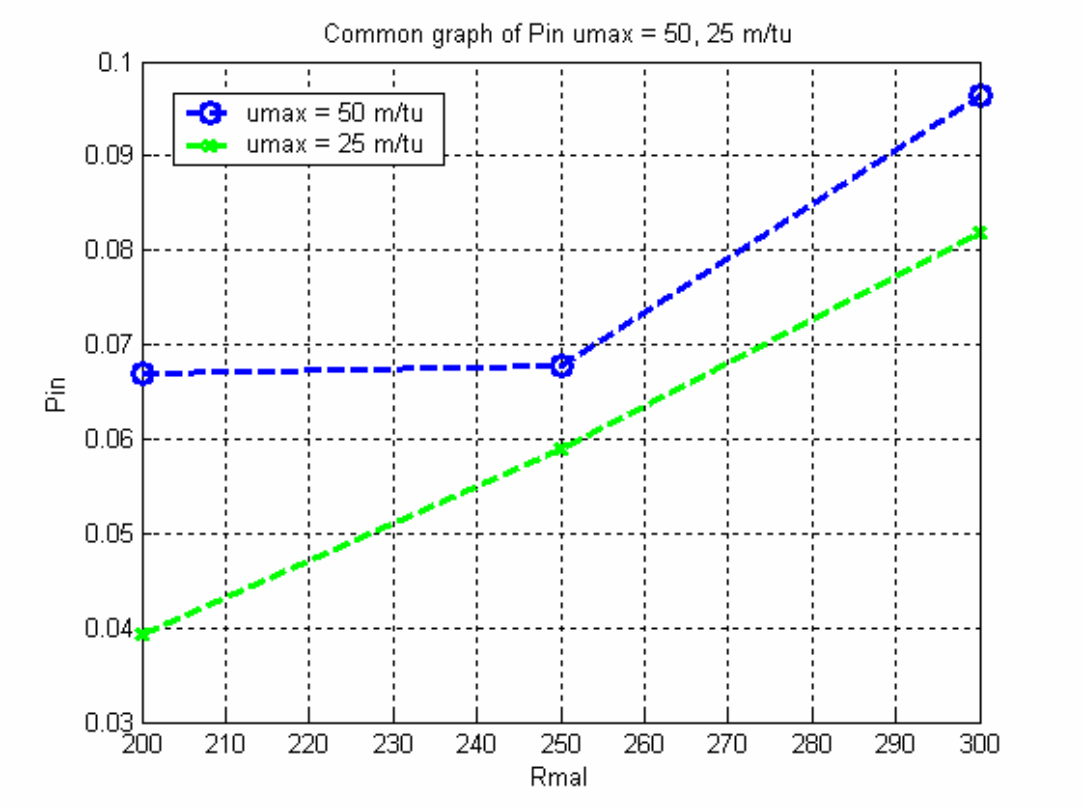
Σχήμα 41 Μέσο ποσοστό μολυσμένων κόμβων *group* επιτιθέμενου P_{ingat} συναρτήσει R_{mal}

β) $l = 0.15,\ l_{mal} = 0.8$

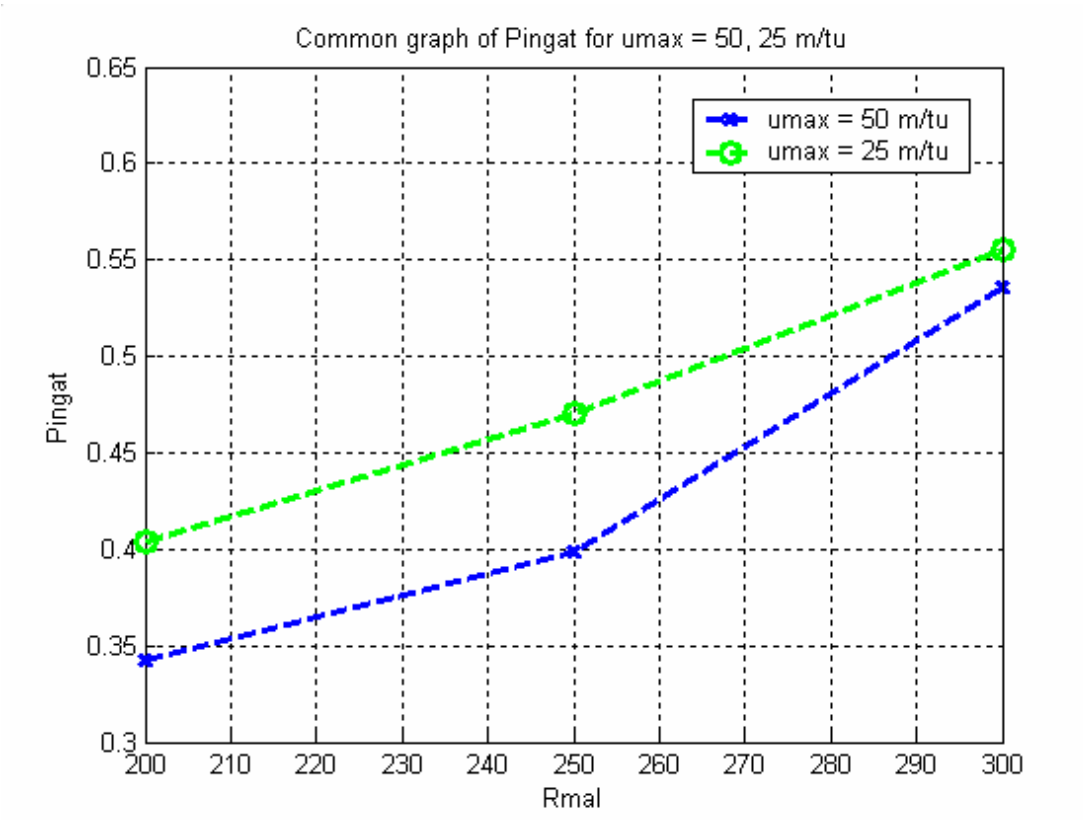
Σε πλήρη αντιστοιχία με τα παραπάνω διαγράμματα παρατίθενται διαγράμματα για την περίπτωση όπου $l_{mal} = 0.8$. Όλα τα υπόλοιπα στοιχεία διατηρήθηκαν σταθερά όπως στην προηγούμενη περίπτωση. Παρατηρούμε ότι – και για τις δύο περιπτώσεις των πειραμάτων α) και β) – το P_d για $u_{\max} = 50\, m/tu$ κινείται σε υψηλότερες τιμές από αυτές που λαμβάνει για $u_{\max} = 25\, m/tu$. Αυτό συμβαίνει γιατί το P_{in} για $u_{\max} = 50\, m/tu$ κινείται σε χαμηλότερα επίπεδα τιμών απ' ότι για $u_{\max} = 25\, m/tu$. Δηλαδή, για μεγαλύτερη ταχύτητα, με ίδιο συνδυασμό l και l_{mal}, ο επιτιθέμενος μολύνει μεγαλύτερο αριθμό κόμβων δικτύου και συνεπώς η δυνατότητα του αλγόριθμου για ανίχνευση μειώνεται. Τέλος, το ποσοστό μολυσμένων κόμβων *group* επιτιθέμενου P_{ingat} είναι μεγαλύτερο για ταχύτητα $u_{\max} = 25\, m/tu$ και μειώνεται για διπλάσια ταχύτητα. Δηλαδή, ο επιτιθέμενος, όταν έχει μικρή ταχύτητα μολύνει περισσότερους κόμβους που είναι άμεσοι γείτονές του και βρίσκονται σε ίδιο *group* με αυτόν σε σχέση με την περίπτωση που αυτός έχει μεγαλύτερη ταχύτητα.



Σχήμα 42 Πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}



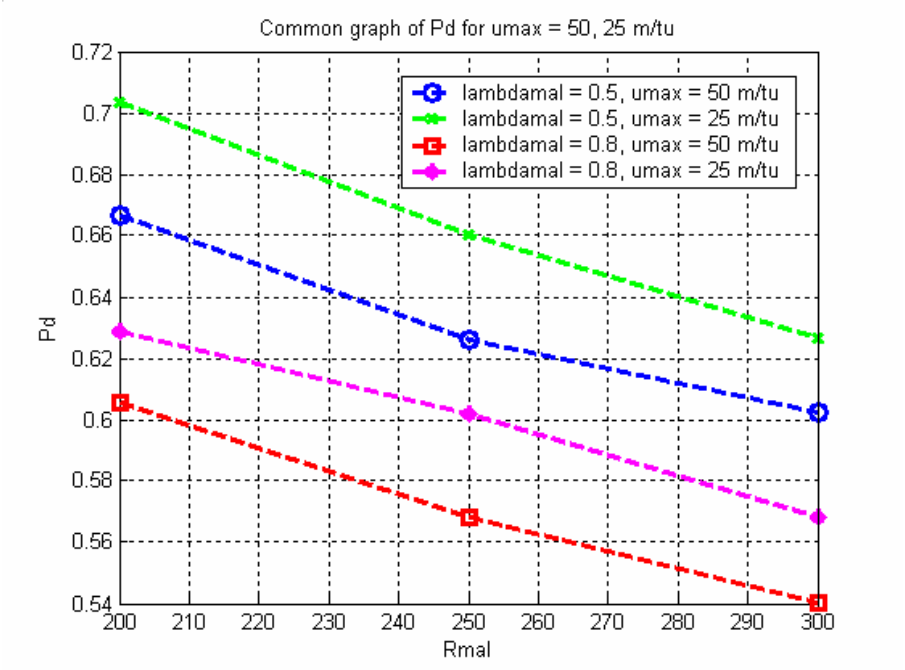
Σχήμα 43 Μέσο ποσοστό μολυσμένων κόμβων P_{in} συναρτήσει R_{mal}



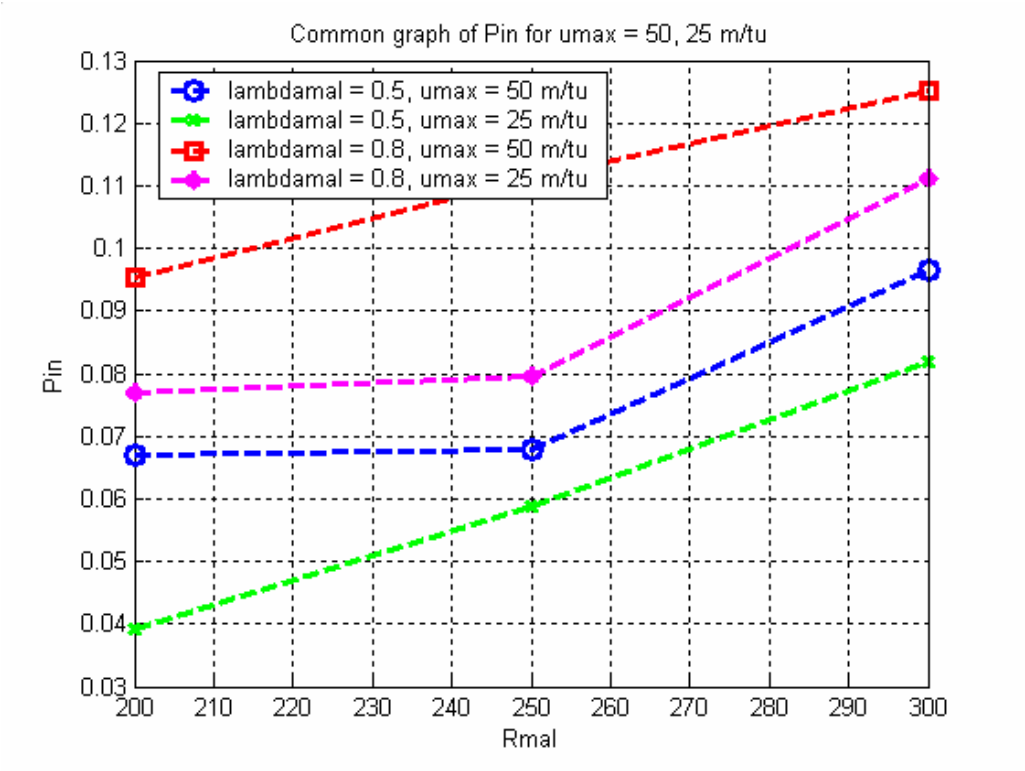
Σχήμα 44 Μέσο ποσοστό μολυσμένων κόμβων *group* επιτιθέμενου P_{ingat} συναρτήσει R_{mal}

5.2.5.2 Συγκριτικά διαγράμματα λ_{mal} και u_{max} συναρτήσει R_{mal}

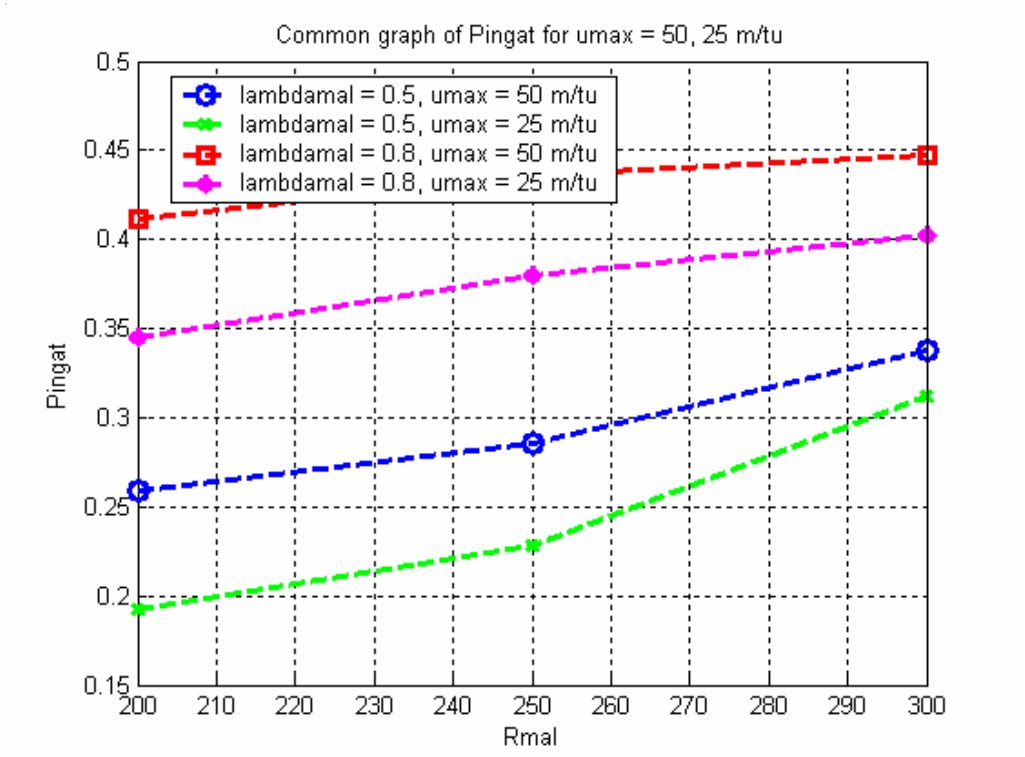
Παρακάτω δίνονται τα διαγράμματα για τις σειρές πειραμάτων που πραγματοποιήθηκαν για l σταθερό και $l_{mal} = 0.5$ ή $l_{mal} = 0.8$. Επίσης στα πειράματα αυτά μεταβλήθηκε και η ταχύτητα του επιτιθέμενου η οποία ελήφθη όπως παραπάνω ($u_{\max} = 50\, m/tu$ ή $u_{\max} = 25\, m/tu$). Από το συγκριτικό διάγραμμα του σχήματος 45 παρατηρείται η εξάρτηση της πιθανότητας ανίχνευσης ανωμαλιών P_d από το ρυθμό μόλυνσης και την ταχύτητα του κακόβουλου κόμβου l_{mal} και $u_{\max}$ αντίστοιχα. Στο συγκεκριμένο διάγραμμα παρατηρούμε ότι για κάθε τιμή του R_{mal} το P_d είναι μεγαλύτερο όταν $l_{mal} = 0.5$ του αντίστοιχου P_d για $l_{mal} = 0.8$. Δηλαδή, βλέπουμε ότι η τιμή του P_d υφίσταται σταθερή μείωση τιμής, για όλες τις τιμές του R_{mal} καθώς το l_{mal} αυξάνεται. Από την άλλη, όσον αφορά το δεύτερο μέγεθος που μεταβάλλεται, το $u_{\max}$, παρατηρείται μία αντίστροφη λογική αφού για μικρότερες τιμές $u_{\max}$ επέρχεται αύξηση στο P_d. Υπενθυμίζουμε ότι ίδιες παρατηρήσεις έγιναν και στα προηγούμενα σχήματα 39 και 42 που αφορούσαν στο μέγεθος P_d εν συγκρίσει του $u_{\max}$.



Σχήμα 45 Συγκριτικό διάγραμμα πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}



Σχήμα 46 Συγκριτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων P_{in} συναρτήσει R_{mal}

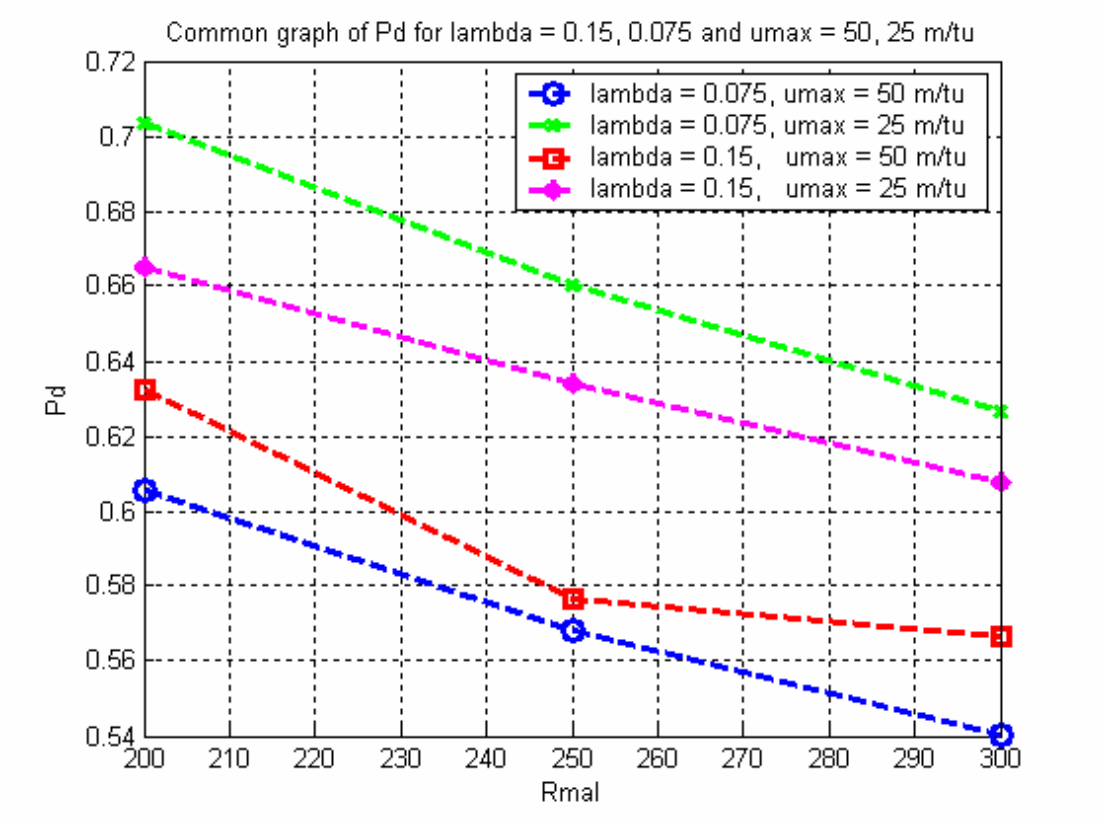


Σχήμα 47 Συγκριτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων *group* επιτιθέμενου P_{ingat} συναρτήσει R_{mal}

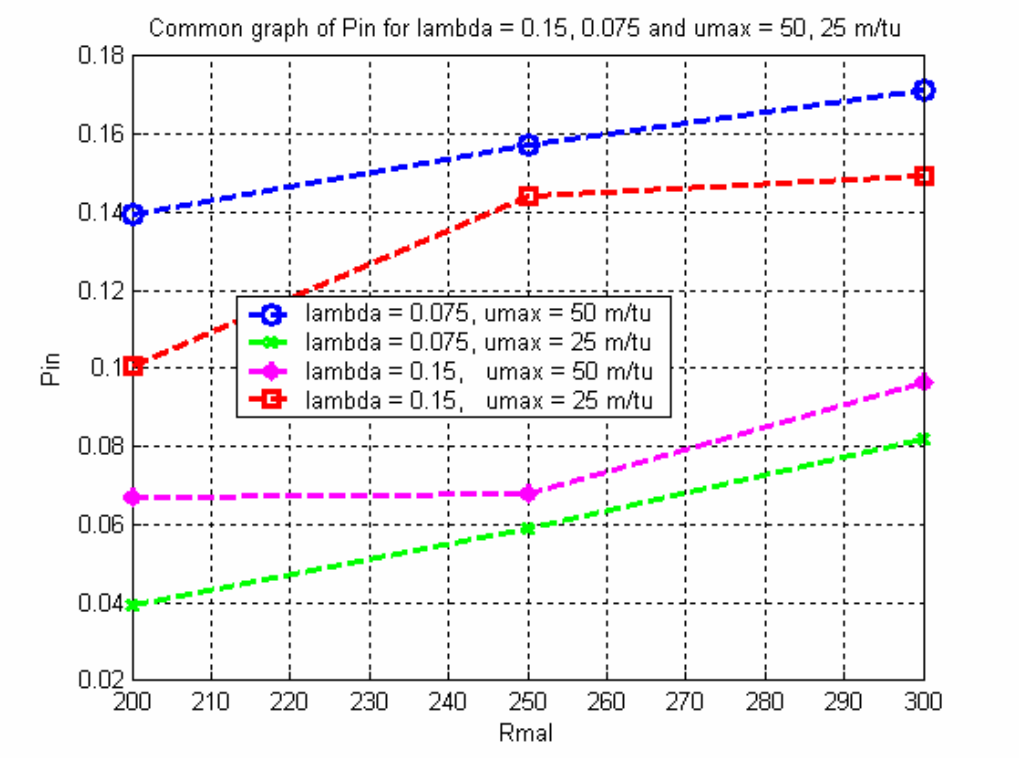
Η μείωση στην τιμή του P_d, η οποία μεταφράζεται σε μείωση της αποδοτικότητας του αλγόριθμου που χρησιμοποιείται στην παρούσα διπλωματική, οφείλεται, όπως και στις προηγούμενες σειρές πειραμάτων που εξετάστηκαν έως τώρα, στη μεγαλύτερη αποδοτικότητα του επιτιθέμενου για μόλυνση (πράγμα που γίνεται εμφανές από τα σχήματα 45 έως 47). Όσο ο κακόβουλος κόμβος μπορεί να κινείται στο χώρο του δικτύου και να μολύνει περισσότερους κόμβους, τόσο μικρότερη είναι η ικανότητα ανίχνευσης που έχει ο εφαρμοζόμενος αλγόριθμος. Με άλλα λόγια αύξηση της αποδοτικότητας του επιτιθέμενου για μόλυνση ή αλλιώς αύξηση του *Infection Efficiency* οδηγεί σε μείωση της ικανότητας από την πλευρά του αλγόριθμου για ανίχνευση ανωμαλιών.

5.2.5.3 Συγκριτικά διαγράμματα λ και u_{max} συναρτήσει R_{mal}

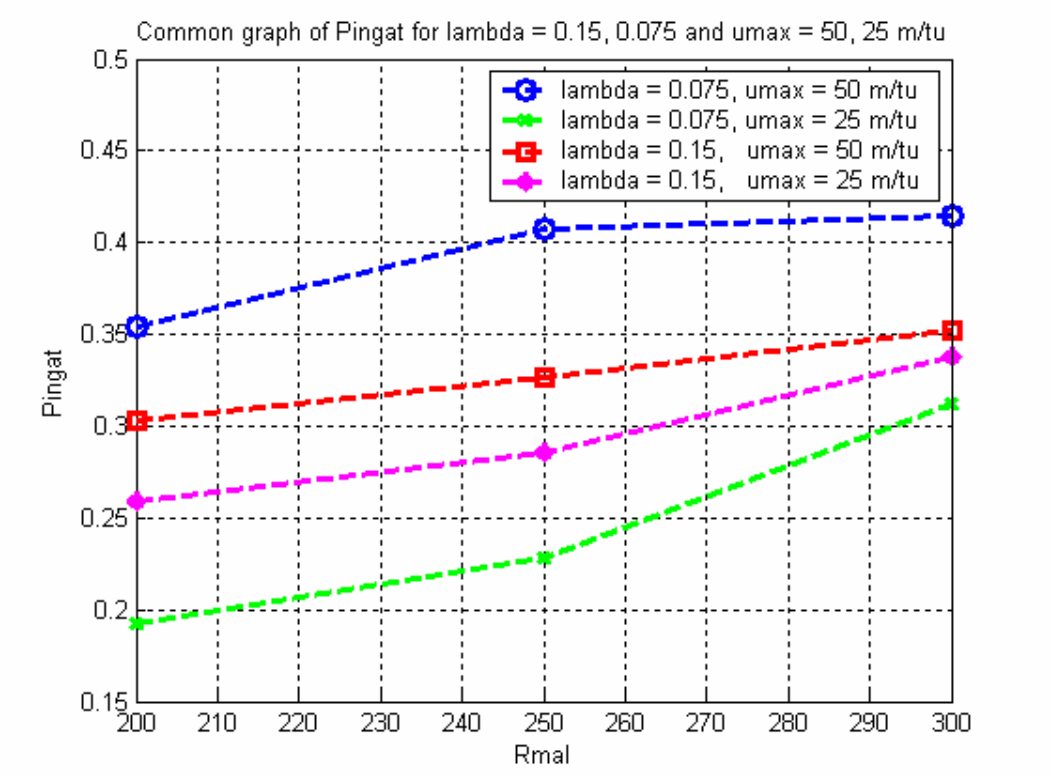
Στο Σχήμα 48 που ακολουθεί παρατίθεται συγκριτικό διάγραμμα που δείχνει πως διαμορφώνονται οι τιμές του P_d για διάφορους συνδυασμούς τιμών l και $u_{\max}$ συναρτήσει του R_{mal}. Όπως και στο συγκριτικό διάγραμμα του σχήματος 45 που παρουσιάστηκε παραπάνω, βλέπουμε την αντίστροφη σχέση που υπάρχει ανάμεσα στην πιθανότητα ανίχνευσης ανωμαλιών P_d, άρα στην αποδοτικότητα του αλγόριθμου που χρησιμοποιούμε, με το ρυθμό μόλυνσης δικτύου l. Παρατηρούμε ότι με την αύξηση του l, δηλαδή με την αύξηση της ικανότητας του δικτύου να διαδίδει τη μόλυνση, μειώνεται η ανιχνευτική ικανότητα του αλγόριθμου. Επίσης, σημειώνουμε ότι το χαρακτηριστικό l του δικτύου, όπως και αντίστοιχα το χαρακτηριστικό l_{mal} του κακόβουλου κόμβου, επηρεάζει σε μεγαλύτερο βαθμό την αποδοτικότητα του αλγόριθμου σε σχέση με το R_{mal}.



Σχήμα 48 Συγκριτικό διάγραμμα πιθανότητα ανίχνευσης ανωμαλιών P_d συναρτήσει R_{mal}



Σχήμα 49 Συγκριτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων P_{in} συναρτήσει R_{mal}



Σχήμα 50 Συγκριτικό διάγραμμα μέσου ποσοστού μολυσμένων κόμβων group επιτιθέμενου P_{ingat} συναρτήσει R_{mal}

5.3 Εντοπισμός του κακόβουλου επιτιθέμενου (tracking)

Σε αυτό το υποκεφάλαιο αναλύονται περιπτώσεις που αφορούν το σχηματισμό του μονοπατιού που ακολουθεί ο επιτιθέμενος κατά την εξέλιξη της επίθεσης στα σενάρια προσομοίωσης. Εξετάζονται δύο περιπτώσεις εντοπισμού του κακόβουλου κόμβου. Στην πρώτη περίπτωση παρατίθενται παραδείγματα εικόνων προσομοίωσης από τα οποία γίνεται σαφής η θέση του επιτιθέμενου στο χώρο του δικτύου. Αντιθέτως, από τη δεύτερη σειρά εικόνων, είναι σαφές πως δεν μπορούμε να έχουμε καθαρή εικόνα για το ακριβές σημείο που βρίσκεται ο επιτιθέμενος κόμβος λόγω του υψηλού βαθμού μόλυνσης του δικτύου για τα χρονικά στιγμιότυπα που εξετάζονται στις συγκεκριμένες εικόνες.

5.3.1 Περίπτωση δυνατού εντοπισμού επιτιθέμενου

Ακολουθούν εικόνες από τέσσερα διαφορετικά στιγμιότυπα πειραμάτων προσομοίωσης που εκτελέστηκαν. Στις εικόνες αυτές βλέπουμε ότι ο αλγόριθμος ανίχνευσης ανωμαλιών είναι ικανός να μας δώσει ένα ακριβές μονοπάτι της κίνησης του επιτιθέμενου κατά τη διάρκεια της προσομοίωσης. Υπάρχει δυνατότητα για περιπτώσεις όπου η διάδοση μολύνσεως παραμένει σε χαμηλά επίπεδα (ή αντίστοιχα η μολυσματική ικανότητα, *Infection Efficiency*, του επιτιθέμενου λαμβάνει σχετικά μικρές τιμές) να μπορούμε να εκτιμήσουμε τη θέση του επιτιθέμενου σε κάθε χρονικό στιγμιότυπο καθ' όλη τη διάρκεια της προσομοίωσης.

Βέβαια αν θέλουμε να εξετάσουμε και την ακρίβεια με την οποία μπορεί να γίνει ο εντοπισμός του επιτιθέμενου θα πρέπει να μελετήσουμε και τα λειτουργικά χαρακτηριστικά αυτού. Η κρίσιμη παράμετρος για την ακρίβεια του εντοπισμού (όχι για την ίδια τη δυνατότητα εντοπισμού) είναι η ακτίνα του επιτιθέμενου R_{mal}. Όσο μεγαλύτερη είναι η ακτίνα αυτή τόσο μικρότερη ακρίβεια στη διαδικασία χάραξης του επιτιθέμενου θα έχουμε.

Βάσει κλασικής γεωμετρίας, γνωρίζουμε ότι το εμβαδό που καταλαμβάνει ο επιτιθέμενος με ακτίνα R_{mal} θα δίνεται από τον τύπο :

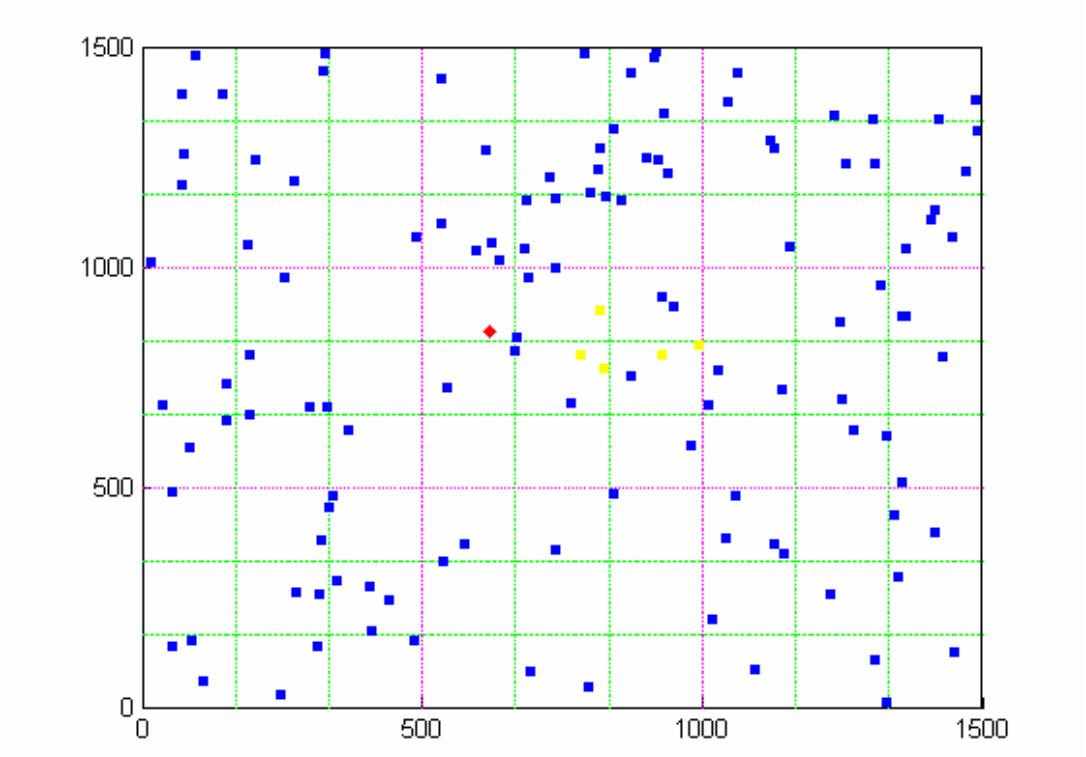
$$E_{mal} = p \cdot R_{mal}^2 \qquad \textbf{(18)}$$

Αντιστοίχως, το εμβαδόν κάθε *group* του δικτύου προσομοίωσης ισούται με

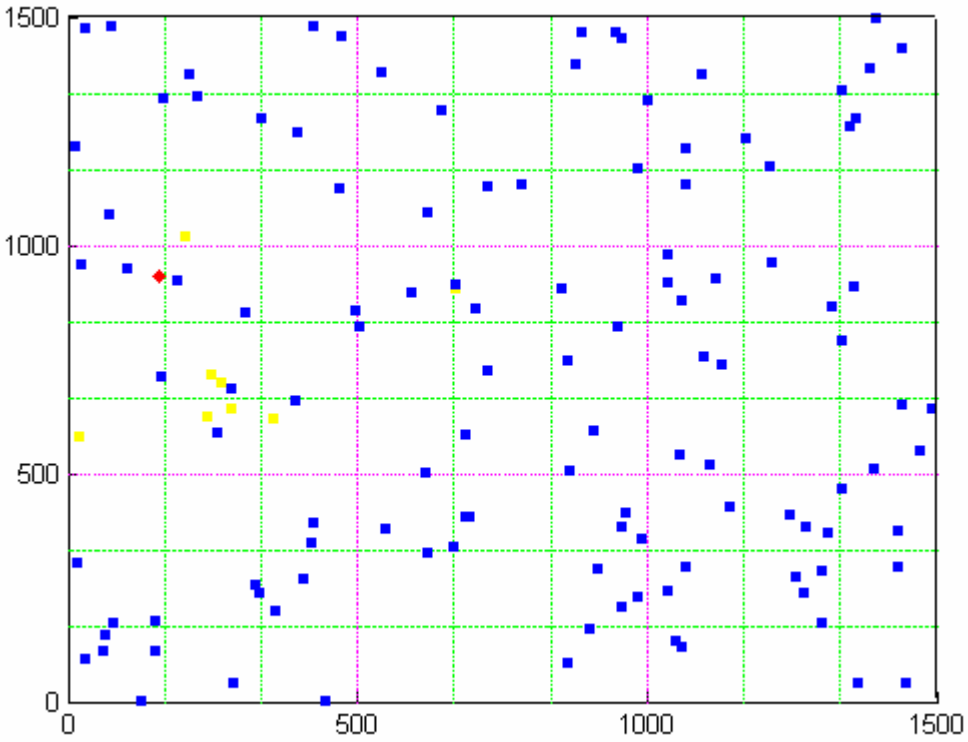
$$E_{group} = l^2, \text{ όπου } l \text{ το μηκος πλευράς του } group \qquad \textbf{(19)}$$

Αντικαθιστώντας τη συνήθη τιμή που λαμβάνουμε για το R_{mal} στα πειράματα ($R_{mal} = 200\ m$) προκύπτει E_{mal} **;** $125.664\ m^2$. Αντίστοιχα για το εμβαδό του καθενός *group* παίρνοντας $l = 500\ m$ (εκ κατασκευής το μήκος της πλευράς του κάθε *group* του δικτύου

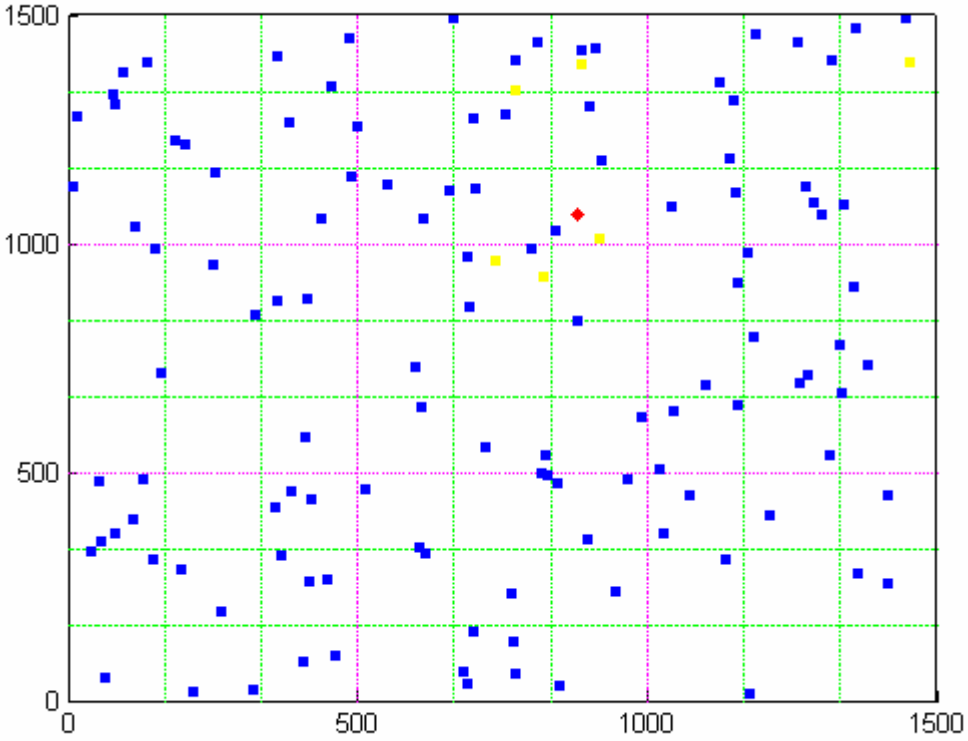
ισούται με $500\,m$) έχουμε $E_{group} = 250.000\,m^2$. Βλέπουμε ότι το εμβαδόν του επιτιθέμενου ακτίνας R_{mal} είναι το μισό από το εμβαδόν του καθενός *group*. Έτσι είναι αρκετά δύσκολο το έργο του εντοπισμού του επιτιθέμενου με ακρίβεια ενός *tile* (του μικρού τετραγώνου μέσα στο *group* με εμβαδόν εκ κατασκευής ίσο με $E_{group}/9$). Αυτό όμως μπορεί να γίνει εφικτό με αναπροσαρμογή του λόγου l/m, λαμβάνοντας σταθερή την ακτίνα του επιτιθέμενου. Με αρκετά μικρό λόγο l/m μπορούμε να υπερβούμε την παραπάνω κατασκευαστική δυσχέρεια και να εντοπίσουμε το μονοπάτι κίνησης του επιτιθέμενου με ακρίβεια ενός *tile* για το μεγαλύτερο μέρος της προσομοίωσης. Σημειώνουμε ότι στις παρακάτω εικόνες, με κόκκινο χρώμα απεικονίζεται ο επιτιθέμενος, με μπλε οι κόμβοι του υποκείμενου δικτύου και με κίτρινο οι κόμβοι οι οποίοι μέσω του αλγόριθμου έχουν ανιχνευτεί ως μολυσμένοι για το χρονικό στιγμιότυπο που απεικονίζεται.



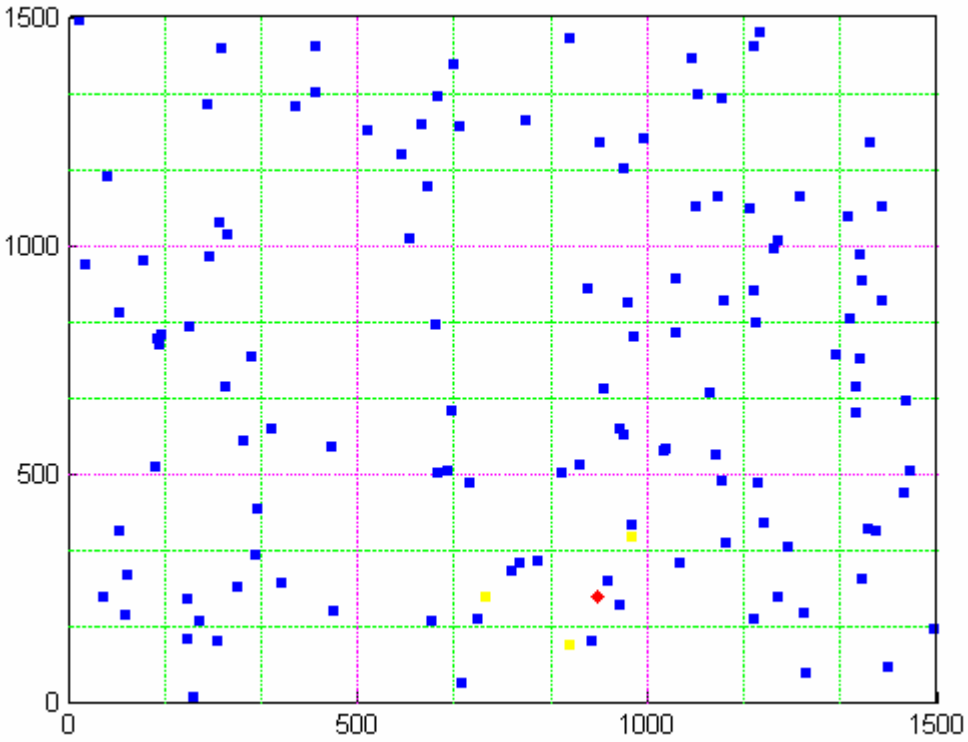
Εικόνα 51 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (1)



Εικόνα 52 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (2)



Εικόνα 53 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (3)



Εικόνα 54 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (4)

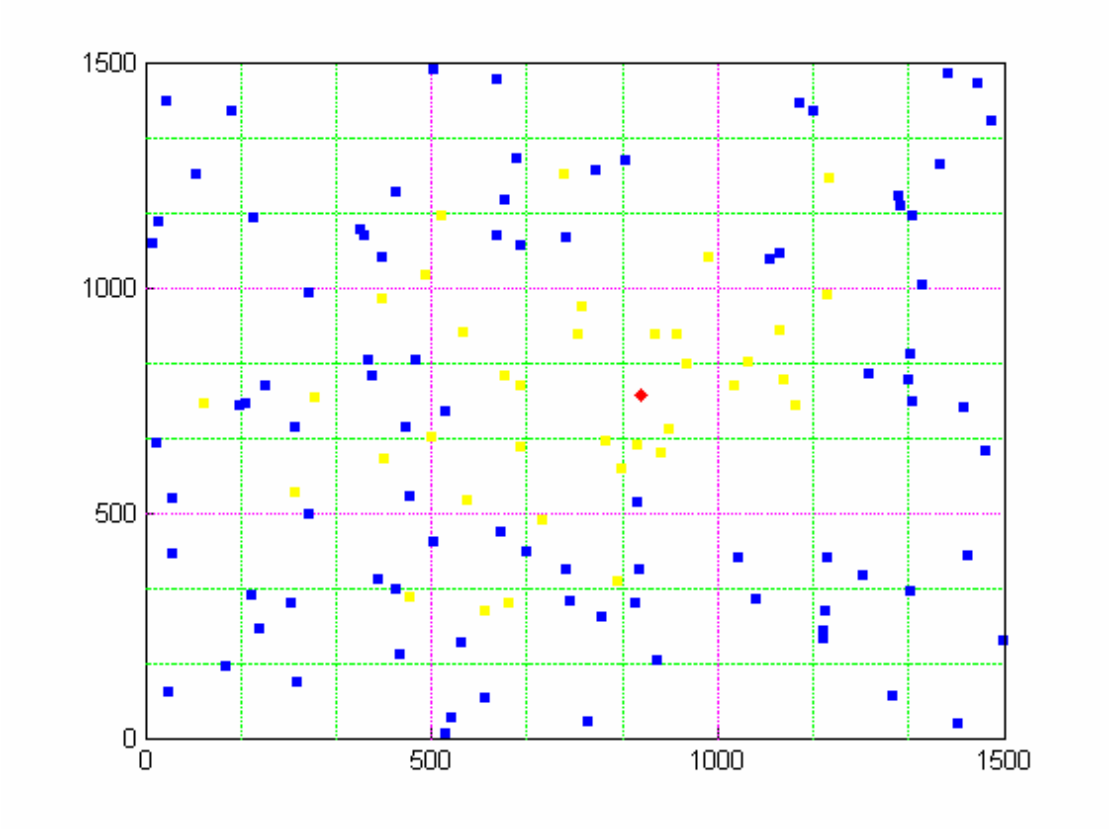
Στις εικόνες 51 – 54 που δόθηκαν παραπάνω, ο λόγος l/m κινήθηκε σε χαμηλές τιμές ώστε να αποφευχθεί εκτεταμένη μόλυνση στο δίκτυο κατά την εξέλιξη των προσομοιώσεων. Με αυτόν τον τρόπο, όπως γίνεται σαφές και από τις παραπάνω εικόνες, είναι δυνατός ο εντοπισμός της κινητής απειλής καθ' όλη τη διάρκεια των πειραμάτων. Εξαίρεση αποτελούν κάποια στιγμιότυπα όπου αναπόφευκτα ο αριθμός των μολυσμένων κόμβων είναι ικανός ώστε να οδηγήσουν τον αλγόριθμο ανίχνευσης σε αποτέλεσμα με μεγάλο σφάλμα.

Επίσης, πρέπει να σημειωθεί γενικά ότι ο αλγόριθμος λόγω των μη ανιχνεύσιμων ανωμαλιών δίνει μια υποεκτιμημένη εικόνα για τις μολύνσεις του δικτύου σε κάθε χρονικό στιγμιότυπο και για τις δύο περιπτώσεις που εξετάζονται σε αυτό το υποκεφάλαιο (δυνατής και μη δυνατής ιχνηθέτησης του επιτιθέμενου). Με άλλα λόγια, τόσο οι εικόνες 51 – 54 που προηγήθηκαν, όσο και οι 55 – 58 που ακολουθούν, κατά μέσο όρο εμφανίζουν λιγότερους κόμβους μολυσμένους απ' όσοι είναι στην πραγματικότητα. Αυτή η απόκλιση

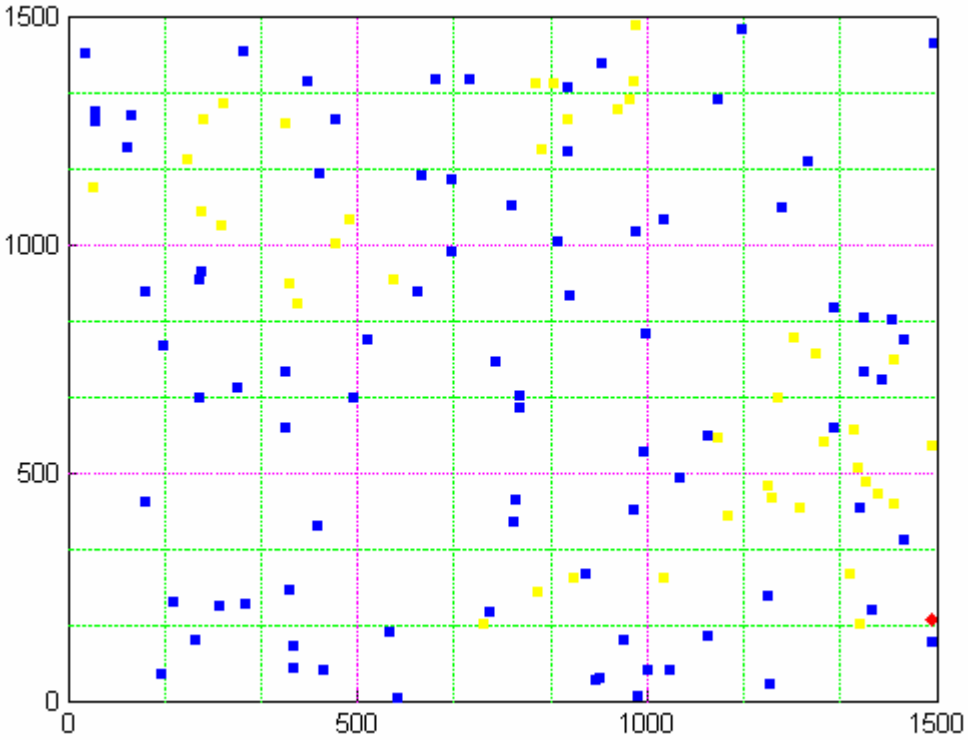
είναι αναπόφευκτή αφού πρέπει να υπάρχει ένας συμβιβασμός μεταξύ των ορίων ανίχνευσης που τίθενται στον αλγόριθμο ώστε να επιτυγχάνεται η καλύτερη δυνατή ανίχνευση ανωμαλιών με το μικρότερο δυνατό αριθμό λανθασμένων ανιχνεύσεων. Ευρύτερα όρια ανίχνευσης θα συμπεριελάμβαναν περισσότερες πραγματικές ανωμαλίες αλλά και περισσότερες εικονικές (*βλ.* παρ. 4.3.1), πράγμα ανεπιθύμητο αφού θα έδινε ψευδή εικόνα της κατάστασης του δικτύου.

5.3.2 Περίπτωση αδυναμίας εντοπισμού επιτιθέμενου

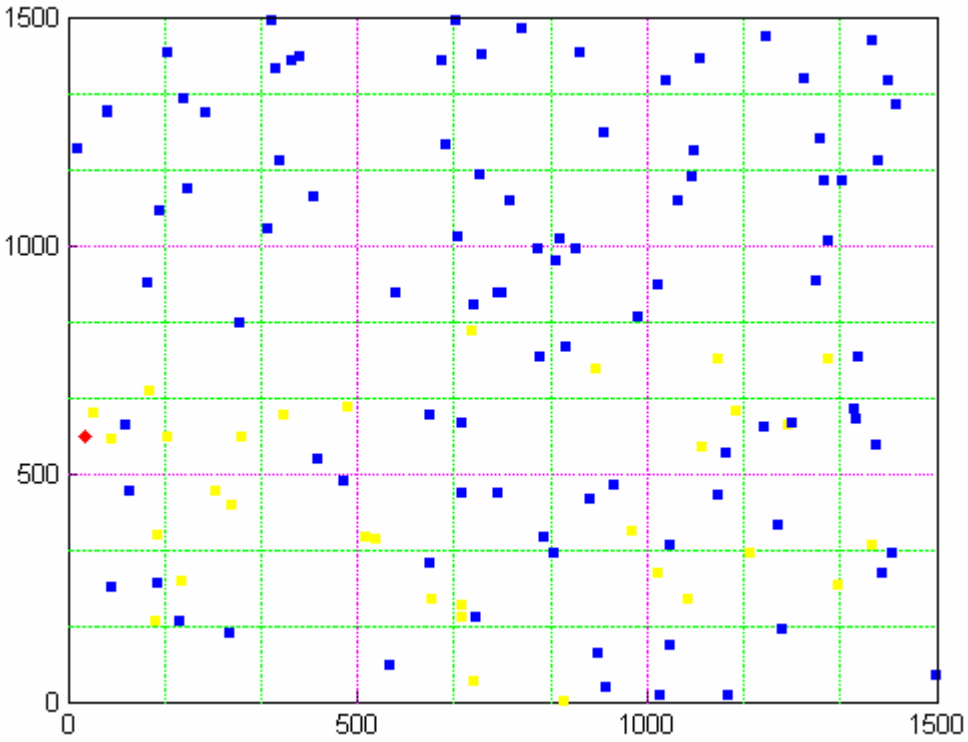
Ομοίως με την αμέσως προηγούμενη υποενότητα 5.3.1, παραθέτουμε εικόνες διαφόρων στιγμιότυπων από προσομοιώσεις στις οποίες αναδεικνύεται η δυσκολία εντοπισμού της κίνησης του κακόβουλου κόμβου, σε περιπτώσεις εκτεταμένης μόλυνσης (όταν έχουμε μεγάλο ρυθμό διάδοσης ανωμαλίας ή αντίστοιχα μεγάλες τιμές μολυσματικής ικανότητας του επιτιθέμενου). Σε όλες τις παρακάτω περιπτώσεις είναι σαφής η δυσκολία απόφασης της ακριβούς θέσης του κινητού κόμβου για τα χρονικά στιγμιότυπα στα οποία αφορούν οι εικόνες. Υπενθυμίζουμε ότι οι αποδόσεις των χρωμάτων που εμφανίζονται στις ακόλουθες εικόνες έχουν επεξηγηθεί στην προηγούμενη υποενότητα.



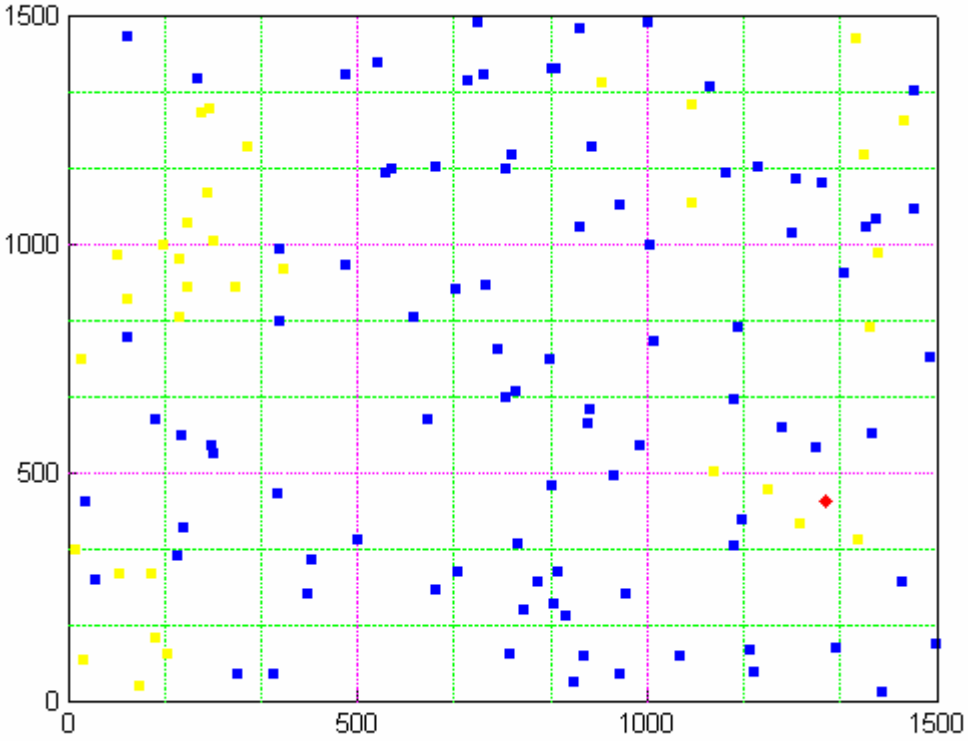
Εικόνα 55 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (5)



Εικόνα 56 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (6)



Εικόνα 57 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (7)



Εικόνα 58 Απεικόνιση προσομοίωσης δικτύου σε ένα τυχαίο χρονικό στιγμιότυπο (8)

Κεφάλαιο 6^{ο} – Επίλογος

6.1 Συμπεράσματα

Η διάδοση μολύνσεων/ανωμαλιών στα δίκτυα υπολογιστών έχει γίνει ιδιαίτερης σημασίας τα τελευταία χρόνια. Όμως, όλες οι προσπάθειες που έχουν παρουσιαστεί στη βιβλιογραφία και αφορούν το συγκεκριμένο πρόβλημα, επικεντρώνονται κυρίως στα ενσύρματα δίκτυα και συνεπώς η εφαρμοσιμότητά τους στα ασύρματα ad hoc δίκτυα είναι περιορισμένη. Στην παρούσα εργασία, αρχικά παρουσιάστηκε ένα πιθανοτικό πλαίσιο που χρησιμοποιείται στη μοντελοποίηση και μελέτη διάδοσης κινούμενων απειλών λαμβάνοντας υπόψη τη φύση και τα ιδιαίτερα χαρακτηριστικά ενός ασύρματου ad hoc δικτύου, όπως επίσης και τους ενεργειακούς περιορισμούς ενός κινούμενου επιτιθέμενου.

Η παρούσα εργασία αφορά στο πρόβλημα ανίχνευσης ανωμαλιών και εντοπισμό της πηγής της ανωμαλίας, σε ασύρματο δίκτυο αισθητήρων. Πιο συγκεκριμένα, υιοθετήθηκε ένας αλγόριθμος για την ανίχνευση ανωμαλιών, ο οποίος λαμβάνοντας υπόψη τα συσχετισμένα δεδομένα από διάφορες ομάδες κόμβων μπορεί να οδηγήσει στον εντοπισμό της πηγής της ανωμαλίας και κατά συνέπεια να συμβάλλει στην καλή προστασία ενός δικτύου αισθητήρων.

Η προτεινόμενη προσέγγιση χρησιμοποιεί και εφαρμόζει τη μέθοδο M^3L που βασίζεται στη Ανάλυση Κύριων Συνιστωσών, *PCA*, για κάθε χρονικό στιγμιότυπο κατά την εξέλιξη της διάδοσης μόλυνσης στο δίκτυο αισθητήρων μέσω της κινητής απειλής.

Η προσέγγιση της *PCA* επελέγη με σκοπό αφενός τη μείωση της διάστασης του συνόλου των δεδομένων στα οποία υπάρχει μεγάλος αριθμός συσχετισμένων μεταβλητών και αφετέρου τη διατήρηση σε όσο το δυνατόν μεγαλύτερο βαθμό της διακύμανσης που εμφανίζει το σύνολο δεδομένων.

Έχει δειχθεί ότι η προτεινόμενη προσέγγιση πετυχαίνει να συνδυάζει αποτελεσματικά δεδομένα με μεγάλη συσχέτιση, με κατανεμημένο τρόπο, έτσι ώστε να ανιχνεύει ανωμαλίες που διαδίδονται μέσω γειτονικών κόμβων. Στο συγκεκριμένο μοντέλο δικτύου αισθητήρων η εισαγωγή μη έγκυρων δεδομένων (ανωμαλία) σε έναν κόμβο έγινε είτε από ένα κακόβουλο κόμβο, ο οποίος κινείται τυχαία στο δίκτυο και μολύνει τους γύρω κόμβους με κάποια πιθανότητα, είτε μέσω ήδη μολυσμένων γειτονικών νόμιμων αισθητήρων. Ο κανόνας βάσει του οποίου έγινε μια μόλυνση καθώς και η περαιτέρω διάδοση αυτής στο δίκτυο ελήφθη από το πιθανοτικό μοντέλο διάδοσης μολύνσεων που έχει αναπτυχθεί για την περίπτωση των ασύρματων ad hoc δικτύων.

Μέσω του συγκεκριμένου πλαισίου διάδοσης μολύνσεως μελετήθηκε η επίδραση που έχουν οι διάφοροι παράγοντες που σχετίζονται με τα λειτουργικά χαρακτηριστικά του κινούμενου κακόβουλου κόμβου στην αποδοτικότητα του προτεινόμενου αλγόριθμου ανίχνευσης ανωμαλιών. Επιπλέον, εκτός από την ανίχνευση των ανωμαλιών που εισάγονται στο δίκτυο, δόθηκε έμφαση και στη δυνατότητα ανίχνευσης της θέσης του κακόβουλου κόμβου κατά την εξέλιξη της επίθεσης.

Τα αποτελέσματα των πειραματικών δεδομένων στα πλαίσια της αποτίμησης του προτεινόμενου αλγόριθμου έδειξαν ότι η μεθοδολογία που ακολουθήθηκε στην παρούσα διπλωματική επιτυγχάνει ικανοποιητική ακρίβεια τόσο για την περίπτωση ανίχνευσης ανωμαλιών όσο και για την περίπτωση του εντοπισμού του κακόβουλου κόμβου. Βασιζόμενοι στα συγκριτικά αποτελέσματα των διαφόρων σεναρίων επίθεσης μπορούμε να αποκτήσουμε μια βαθύτερη γνώση όσον αφορά την αποτελεσματικότητα των διαφόρων στρατηγικών επίθεσης (επιθέσεις στις οποίες επιτυγχάνεται μικρότερη αποδοτικότητα του αλγόριθμου ανίχνευσης ανωμαλιών) από τη σκοπιά του επιτιθέμενου.

Από τα αποτελέσματα των πειραμάτων της παρούσας εργασίας καταλήξαμε στο συμπέρασμα ότι η παράμετρος που επηρεάζει πιο πολύ την αποδοτικότητα του αλγόριθμου είναι σαφώς και κατά κύριο λόγο ο ρυθμός μόλυνσης μεταξύ δύο κόμβων του δικτύου l. Ο συγκεκριμένος ρυθμός διαδραματίζει μείζονα ρόλο στη μετάδοση μόλυνσης μεταξύ των κόμβων του δικτύου και προκύπτει ότι όσο μεγαλύτερη τιμή l έχουν οι κόμβοι του δικτύου, τόσο πιο δυσχερές καθίσταται το έργο του αλγόριθμου για την εύρεση ανωμαλιών και τον εντοπισμό του επιτιθέμενου στο δίκτυο.

Από την άλλη, η πυκνότητα του δικτύου εμφανίζεται από τα πειράματα ως μια παράμετρος που επίσης επηρεάζει το έργο ανίχνευσης επιθέσεων, αν και όχι σε τόσο μεγάλο βαθμό όσο ο ρυθμός μόλυνσης των κόμβων του δικτύου, l. Όσο μεγαλύτερη είναι η πυκνότητα του δικτύου, τόσο πιο εύκολη είναι η διάδοση ανωμαλιών μεταξύ των κόμβων και παρεπόμενα τόσο πιο δύσκολη η ανίχνευση των αποτελεσμάτων της από τον προτεινόμενο αλγόριθμο. Ανάλογα, οι παράμετροι που αμιγώς σχετίζονται με τον επιτιθέμενο, όπως ο ρυθμός μολύνσεως αυτού l_{mal}, η ακτίνα του R_{mal} ή η μέγιστη ταχύτητά του $u_{\max}$, επηρεάζουν σχετικά την αποδοτικότητα του αλγόριθμου. Όσο μεγαλύτερη είναι η μολυσματική ικανότητα του επιτιθέμενου (*Infection Efficiency*), τόσο μικρότερη είναι η ανιχνευτική ικανότητα της μεθόδου που χρησιμοποιήθηκε στην παρούσα εργασία. Για παράδειγμα, όσο μεγαλύτερη τιμή έχουν τα μεγέθη l_{mal} και R_{mal} του επιτιθέμενου κόμβου, τόσο μεγαλύτερη μολυσματική ισχύ έχει ο τελευταίος και κατά συνέπεια τόσο μικρότερη θα είναι και η ανιχνευτική δυνατότητα του αλγόριθμου. Αντίθετα, η αποδοτικότητα του αλγόριθμου διατηρείται σε σχετικά χαμηλά επίπεδα όσο μικρότερη είναι η ταχύτητα του επιτιθέμενου $u_{\max}$, αφού με μικρή ταχύτητα επιτιθέμενου μολύνονται περισσότεροι κόμβοι δικτύου απ' όσοι μολύνονται στην περίπτωση μεγάλης ταχύτητας. Τέλος, επισημαίνουμε, ότι για διάφορους συνδυασμούς των παραπάνω παραμέτρων προέκυψαν ποικίλα συμπεράσματα όσον αφορά την πιστότητα ανίχνευσης

του προτεινόμενου αλγόριθμου. Μερικά αντιπροσωπευτικά δείγματα αυτών των πειραμάτων έχουν δοθεί κατά το προηγούμενο κεφάλαιο.

Συμπερασματικά, καταλήξαμε ότι μέσω των προσομοιώσεων, δόθηκε η δυνατότητα να κατανοήσουμε καλύτερα την αντίδραση του δικτύου στα διάφορα σενάρια επίθεσης και παρεπόμενα να αποτιμήσουμε την αποδοτικότητα του αλγόριθμου ανίχνευσης. Συνδυάζοντας τα αποτελέσματα τόσο από την σκοπιά του δικτύου όσο και από τη σκοπιά του επιτιθέμενου είμαστε σε θέση να λάβουμε μέτρα για την καλύτερη και πιο αποδοτική εφαρμογή του προτεινόμενου αλγόριθμου ανίχνευσης.

6.2 Κατευθύνσεις για μελλοντική εργασία

Βασιζόμενοι στα συγκριτικά αποτελέσματα του πειραματικού μέρους της παρούσας εργασίας καταλήξαμε ότι ο προτεινόμενος αλγόριθμος μπορεί να χρησιμοποιηθεί για την ανίχνευση ανωμαλιών στα πλαίσια ενός ad hoc δικτύου αισθητήρων. Επιπρόσθετα, ο αλγόριθμος αυτός, μέσω της ανίχνευσης ανωμαλιών για κάθε χρονικό στιγμιότυπο της εξέλιξης της μόλυνσης, παρέχει το πλεονέκτημα του εντοπισμού του μονοπατιού κίνησης του κακόβουλου κόμβου κατά την εξέλιξη της επίθεσης.

Όμως, σε αυτό το σημείο πρέπει να σημειωθεί ότι παντού στην παρούσα εργασία υποθέσαμε στατικό υποκείμενο δίκτυο. Η μελλοντική εργασία η οποία θα είχε άμεση σχέση με την παρούσα και θα αποτελούσε άμεση επέκτασή της θα συμπεριλάμβανε τη μελέτη του αντίστοιχου προβλήματος ανίχνευσης ανωμαλιών για ένα πλήρως κινούμενο δίκτυο (*fully mobile network*).

Βιβλιογραφία

[1] V. Karyotis, S. Papavassiliou, M. Grammatikou and V. Maglaris, "A Novel Framework for Mobile Attack Strategy Modeling and Vulnerability Analysis in Wireless Ad Hoc Networks", *Int. J. Security and Networks*, Vol. 1, No. 3/4, pp. 256–260, December 2006.

[2] V. Karyotis, S. Papavassiliou, M. Grammatikou and B. Maglaris, "On the Characterization and Evaluation of Mobile Attack Strategies in Wireless Ad Hoc Networks", *Proceedings of the 11th IEEE Symposium on Computers and Communications (ISCC'06)*, pp. 29–32, March 2006.

[3] V. Karyotis, S. Papavassiliou and B. Maglaris, "Modeling Framework for the Study and Analysis of the Mobile Attack Propagation in Wireless Ad-Hoc Networks", pp. 1–3, *Proceedings of the 12th European Wireless Conference (EW'06), Athens, Greece*, April 2006.

[4] V. Chatzigiannakis and S. Papavassiliou, "Diagnosing Anomalies and Identifying Faulty Nodes in Sensor Networks", *IEEE Sensor Journal*, Vol. 7, No.5, pp. 637–641, May 2007.

[5] R. Pastor-Satorras and A. Vespignani, "Epidemics and Immunization in Scale–Free Networks", in *S. Bornholdt and H. Schuster (Eds). Handbook of Graphs and Networks: From the Genome to the Internet*, Wiley-VCH, pp.113–132, Berlin, 2002.

[6] F. Liljeros, C. R. Edling and L. A. N. Amaral, "Sexual Networks: Implications for the Transmission of Sexually Transmitted Infections", *Microbes and Infection*, pp.189–196, June 2003.

[7] J. O. Kephart and S. R. White, "Directed–graph Epidemiological Models of Computer Viruses", *Proceedings of the Second IEEE Computer Society Symposium on Research in Security and Privacy*, pp.343–359, May 1991.

[8] J. O. Kephart and S. R. White, “Measuring and Modeling Computer Virus Prevalence”, *Proceedings of the Fourth IEEE Computer Society Symposium on Research in Security and Privacy*, pp.2–15, May 1993.

[9] Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos, “Epidemic Spreading in Real Networks: An Eigenvalue Approach”, *Proceedings of the 22nd Symposium on Reliable Distributed Computing*, Florence, Italy, October 2003.

[10] C. C. Zou, W. Gong and D. Towsley, “Code Red Worm Propagation Modeling and Analysis”, *Proceedings of the 9th ACM Conference on Computer and communications Security (CCS)*, pp. 138–147, November 2002.

[11] C. C. Zou, W. Gong and D. Towsley, “Worm Propagation Modeling and Analysis under Dynamic Quarantine Defense”, *First ACM Workshop on Rapid Malcode (WORM)*, pp. 51–56, Washington, DC, October 2003.

[12] Z. Chen, L. Gao and K. Kwiat, “Modeling the Spread of Active Worms”, *Proceedings of the 22nd IEEE Conference on Computer Communications (INFOCOM)*, Vol. III, pp. 1890–1900, Miami, FL, April 2003.

[13] C. C. Zou,. D. Towsley and W. Gong, “Email Virus Propagation Modelling and Analysis”, *Technical Report TR-CSE-03-04*, University of Massachusetts, Amherst, MA, 2004.

[14] K. R. Rohloff and T. Basar, “Stochastic Behavior of Random Constant Scanning Worms”, *Proceedings of the 14th International Conference on Computer Communications and Networks (ICCCN)*, pp. 339–344, San Diego, CA, October 2005.

[15] C. C. Zou, W. Gong and D. Towsley, “Code Red Worm Propagation Modeling and Analysis”, *Proceedings of the 9th ACM Conference on Computer and Communications Security (ACM CCS)*, pp.138–147, Washington, DC, November 2002.

[16] A. Ganesh, L. Massoulie and D. Towsley, “The Effect of Network Topology on the Spread of Epidemics”, *Proceedings of the 24th IEEE Conference on Computer Communications (INFOCOM)*, Vol. II, pp.1455–1466, Miami, FL, March 2005.

[17] V. A. Karyotis, “A Base Case Study of the Worm Propagation Problem in the Framework of Topology Control in Wireless Ad-Hoc Networks”, Master’s Thesis, School of Engineering & Applied Sciences, University of Pennsylvania, Philadelphia, PA, August 2005.

[18] J. Newsome, E. Shi, D. Song and A. Perrig, "The Sybil Attack in Sensor Networks: Analysis & Defenses", *Proceedings 3rd Int. Symp. Inf. Process. Sensor Netw.*, pp. 259–268, April 2004.

[19] A. D.Wood and J. A. Stankovic, "Denial of Service in Sensor Networks", *IEEE Computer*, Vol. 35, No. 10, pp. 54–62, October 2002.

[20] E. Shi and A. Perrig, "Designing Secure Sensor Networks", *Wireless Commun.*, Vol. 11, No. 6, pp. 38–43, December 2004.

[21] F. Y. Luo, H. S. Lu and L. Zhang, "Statistical En-Route Filtering of Injected False Data in Sensor Networks", *IEEE J. Sel. Areas Commun.*, Vol. 23, No. 4, pp. 839–850, April 2005.

[22] G. Wang, W. Zhang, G. Cao and T. La Porta, "On Supporting Distributed Collaboration in Sensor Networks", *Proceedings IEEE Military Commun. Conf.*, Vol. 2, pp. 752–757, October 2003.

[23] B. Przydatek, D. Song and A. Perrig, "SIA: Secure Information Aggregation in Sensor Networks", *Proceedings 1st Int. Conf. Embedded Networked Sensor Syst.*, pp. 255–265, November 2003.

[24] S. Tanachaiwiwat and A. Helmy, "Correlation Analysis for Alleviating Effects of Inserted Data in Wireless Sensor Networks", *Proceedings Mobile and Ubiquitous Syst.: Networking Services*, pp. 97–108, July 2005.

[25] T. Palpamas, "Distributed Deviation Detection in Sensor Networks", *Proceedings ACM SIGMOD*, Vol. 32, No. 4, pp. 77–82, December 2003.

[26] I. T. Jollife, "Principal Component Analysis", 2nd ed. Springer, New York, 2002.

[27] A. Perera, N. Papamichail, N. Barsan, U.Weimar and S. Marco, "Online Event Detection by Recursive Dynamic Principal Component Analysis and Gas Sensor Arrays under Drift Conditions", *Proceedings IEEE Sensors*, Vol. 2, pp. 860–865, October 2003.

[28] M. Oka, Y. Oyama, H. Abe and K. Kato, "Anomaly Detection using Layered Networks based on Eigen Co-Occurrence Matrix", *Proceedings 7th Int. Symp. Recent Advances in Intrusion Detection (RAID)*, pp. 223–237, September 2004.

[29] A. Lakhina, M. Crovella and C. Diot, "Diagnosing Network-Wide Traffic Anomalies", *Proceedings Conf. Appl., Technol., Arch., Protocols for Comput. Commun. (SIGCOMM)*, pp. 219–230, August – September 2004.

[30] J. Cho, J. Lee, S. Choi, D. Lee and I. Lee, "Sensor Fault Identification based on Kernel Principal Component Analysis", *Proceedings IEEE Int. Conf. Control Appl.*, pp. 1223–1228, September 2004.

[31] T. S. Rappaport, "Wireless Communications: Principles and Practice", Prentice-Hall, pp.69–122 and 139–196, Englewood Cliffs, NJ, 1996.

[32] A. Jiang and J. Bruck, "Monotone Percolation and the Topology Control of Wireless Networks", *Proceedings of the 24th IEEE Conference on Computer Communications (INFOCOM)*, Vol. I, pp.327–338, Miami, FL, March 2005.

[33] R. Wattenhofer, L. Li, P. Bahl and Y-M. Wang, "Distributed Topology Control for Power Efficient Operation in Multihop Wireless Ad Hoc Networks", *Proceedings of 20th IEEE Conference on Computer Communications (INFOCOM)*, Vol. III, pp.1388–1397, Anchorage, AL, April 2001.

[34] http://www.computingunplugged.com/issues/issue200407/00001347001.html

[35] http://searchmidmarketsecurity.techtarget.com/sDefinition/0,,sid198_gci802800,00.html

[36] http://netforbeginners.about.com/od/a/g/antivirus.htm

[37] http://searchenterprisedesktop.techtarget.com/sDefinition/0,,sid192_gci212753,00.html

[38] Ath. Papoulis, "Probability, Random Variables, and Stochastic Processes", McGraw-Hill, pp. 612–617, New York, 1991.

[39] R. W. Wolff, "Stochastic Modeling and the Theory of Queues", Prentice – Hall, Englewood Cliffs, NJ, January 1989.

[40] D. Bertsekas, and R. Gallager, "Data Networks", 2nd ed. Prentice-Hall, Englewood Cliffs, NJ, 1992.

[41] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors", *Communications of the ACM*, Vol. 43, No. 5, pp. 51–58, May 2000.

[42] S. Pattem, B. Krishnmachari and R. Govindan, "The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks", *Proceedings 3rd Int. Symp. Inf. Processi. Sensor Netw.*, pp. 28–35, April 2004.

[43] R. Dunia and S. J. Qin, "A Subspace Approach to Multidimensional Fault Identification and Reconstruction," *Amer. Inst. Chem. Eng. J.*, pp. 1813–1831, April 1998.

[44] J. E. Jackson, "A User's Guide to Principal Components.", Wiley, New York, 2003.

[45] J. E. Jackson and G. S. Mudholkar, “Control Procedures for Residuals Associated with Principal Component Analysis,” Technometrics, pp. 341–349, 1979.

[46] J-Y. Le Boudec and M. Vojnovic, “Perfect Simulation and Stationarity of a Class of Mobility Models”, *Proceedings of the 24th IEEE Conference on Computer Communications (INFOCOM)*, Vol. IV, pp.2743–2754, Miami, FL, March 2005.