Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακή Διατριβή

στα Πληροφοριακά και Επικοινωνιακά Συστήματα

Ευφυείς τεχνικές ανάλυσης επιθέσεων και ανώμαλης συμπεριφοράς συστημάτων, με χρήση sensors.

Κωνσταντίνος Βεσσιάρης

Επιβλέπων Καθηγητής
Δρ. Σταύρος Σιαηλής

Σεπτέμβριος 2015

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Ευφυείς τεχνικές ανάλυσης επιθέσεων και ανώμαλης συμπεριφοράς συστημάτων, με χρήση sensors.

Κωνσταντίνος Βεσσιάρης

Επιβλέπων Καθηγητής
Δρ. Σταύρος Σιαηλής

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε
προς μερική εκπλήρωση των απαιτήσεων για απόκτηση

μεταπτυχιακού τίτλου σπουδών
στα Πληροφοριακά Συστήματα

από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου

Σεπτέμβριος 2015

Περίληψη

Η εσωτερική απειλή(insider threat) αποτελεί μια από τις σημαντικότερες απειλές απώλειας δεδομένων / πληροφοριών από ένα εταιρικό δίκτυο. Τα περιστατικά αυξάνονται όλο και περισσότερο στις μέρες μας παρόλο που αρκετά αποκρύπτονται καθαρά για σκοπούς αποφυγής μείωσης της φήμης και αξιοπιστίας της εταιρίας.

Η παρούσα μεταπτυχιακή διατριβή κινείται πάνω σε δύο κύριους άξονες. Ο πρώτος αφορά το ποια δεδομένα θα πρέπει να συλλέξουμε από ένα δίκτυο αποτελούμενο από υπολογιστές με λειτουργικό σύστημα Windows και ο δεύτερος άξονας είναι η αξιοπιστία των αποτελεσμάτων που μπορεί να παραχθούν από ένα αλγόριθμο μη-επιτηρούμενης μάθησης ο οποίος τροφοδοτείται από τα δεδομένα που έχουν συλλεγεί.

Το λειτουργικό Windows, σύμφωνα με στατιστικές έρευνες [1], [2], [3]κατέχει το μεγαλύτερο ποσοστό μεριδίου της αγοράς υπολογιστών. Επίσης η απουσία πλήθους επιστημονικών ερευνών που αφορούν την ανίχνευση εσωτερικών απειλών σε περιβάλλον Windows αποτελούν τους κυριότερους παράγοντες της επιλογής του συγκεκριμένου λειτουργικού για την παρούσα μεταπτυχιακή διατριβή.

Επιπλέον στα πλαίσια της παρούσας μεταπτυχιακής διατριβής αναπτύχθηκε ένα σύστημα ανίχνευσης εσωτερικών απειλών. Το σύστημα κλήθηκε να ανιχνεύσει την ύπαρξή ιών κατηγορίας Trojan, Bot, DoS, DDoS και RAT σε ένα εταιρικό δίκτυο. Τα αποτελέσματα μας δείχνουν ότι το σύστημα μας μπορεί να βοηθήσει σημαντικά στην ανίχνευση εσωτερικών απειλών, ανεβάζοντας ακόμη λίγο το πήχη της ασφάλειας στις εταιρίες.

Summary

Insider threats are considered dangerous and play a significant role in the loss of data and information of corporate networks. Even though insider threat incidents increasing day by day, most of this incident are not reported in order to protect the reputation and trustworthiness of companies.

This dissertation focuses on two key axes. The first one deals with the study of data that needs to be collected from network computers running Windows operating system. The collected data will be used as inputs to unsupervised learning algorithm and the trustworthiness of the produced results will be analysed and evaluated.

According to online statistics [1], [2], [3] Windows Operating system powers the largest market share of computer machines in the world. Also the lack of scientific research around the field of insider threat detection in a network running Windows operating system, guide us to the adoption this operating system as the starting point for our research .

Finally, a system for detecting insider threats was developed and different scenarios have been run on this system. These scenarios covered Trojans, Bots, DoS and DDoS. The main purpose of our test was to emulate an inside malicious user activity in order to test algorithm efficiency and evaluate its results. The results seems promising.

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες προς τον Υπεύθυνο Καθηγητή μου Δρ. Σταύρο Σιαηλή για τη συνεχή καθοδήγηση και επίβλεψη που μου παρείχε έτσι ώστε να γίνει κατορθωτή η ολοκλήρωση και παράδοση της παρούσας μεταπτυχιακής διατριβής. Την υπομονή και επιμονή που είχε δείξει ώστε να βρεθεί ένα θέμα αντάξιο των δυνατοτήτων και γνώσεων μου που θα είχε θετικό αντίκτυπο στη δουλειά μου και στη μετέπειτα πορεία μου.

Ιδιαίτερες ευχαριστίες θέλω να απευθύνω στη Δρ. Μαρία Παπαδάκη για τη καθοδήγηση και το υλικό που πρόσφερε κατά τα πρώτα στάδια της έρευνας μας. Επίσης θα ήθελα να ευχαριστήσω τον Δρ. Σταύρο Σταύρου και τον διδακτορικό μαθητή του Amin Gholoobi για την παραχώρηση των εργαστηρίων του Ανοικτού Πανεπιστήμιού Κύπρου για την υλοποίηση των εργαστηριακών δοκιμών μας καθώς και για τη βοήθεια τους κατά τη διάρκεια των πειραμάτων.

Τέλος θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες προς τα άτομα της οικογένειας μου για τη θερμή τους συμπαράσταση καθ' όλη τη διάρκεια διεκπεραίωσης της μεταπτυχιακής διατριβής.

Περιεχόμενα

Εικόνες

Πίνακες

Γραφήματα

Κεφάλαιο 1
Εισαγωγή

Στα σημερινά εταιρικά δίκτυα ο μεγαλύτερος κίνδυνος επιθέσεων δεν είναι οι εξωτερικές επιθέσεις αλλά οι εσωτερικές. Οι εταιρίες επενδύουν τεράστια ποσά για την περιμετρική ασφάλεια αλλά αγνοούν ένα σημαντικό παράγοντα που είναι ο άνθρωπος και πιο συγκεκριμένα οι χρήστες που έχουν πρόσβαση στο δίκτυο μιας εταιρίας. Σύμφωνα με την έρευνα που πραγματοποίησε η εταιρία McAffe [4] το 68 τοις εκατό των επιθέσεων που έχουν σαν στόχο εταιρικά δίκτυα δεν γίνονται από εξωτερικούς hackers αλλά από εσωτερικούς έμπιστους χρήστες, που εκμεταλλεύονται τη δυνατότητα πλοήγησης σε διάφορες περιοχές του εταιρικού δικτύου με σκοπό την πρόκληση ζημιάς σε αυτό.

Ο όρος εσωτερική απειλή διαχωρίζεται σε δύο κατηγορίες: τη κατηγορία κακόβουλης χρήσης και τη κατηγορία λανθασμένης χρήσης(misuse). Στη κατηγορία κακόβουλης χρήσης εμπεριέχονται οι εσωτερικοί χρήστες που έχουν σαν σκοπό τη πρόκληση βλάβης στο εταιρικό δίκτυο ή την εσκεμμένη κλοπή πληροφοριών από αυτό. Με τον όρο εσωτερικός χρήστης χαρακτηρίζεται οποιοδήποτε άτομο διαθέτει πρόσβαση στο δίκτυο, το άτομο αυτό μπορεί να είναι είτε υπάλληλος είτε εξωτερικός συνεργάτης. Οι λόγοι που

μπορούν να οδηγήσουν έναν εσωτερικό χρήστη στην ανάπτυξη επιθυμίας πρόκλησης βλάβης στο εσωτερικό δίκτυο είναι κυρίως το αίσθημα απογοήτευσης και θυμού προς τα άτομα που απαρτίζουν την εταιρία. Τα συναισθήματα αυτά πιθανότατα να αναπτυχθούν λόγω μιας αποκοπής μισθού, στέρησης αύξησης μισθού ή και της συνεχούς απαξίωσης προς το άτομο που τρέφει τα συναισθήματα αυτά.

Στη κατηγορία λανθασμένης χρήσης ανήκουν τα περιστατικά πρόκλησης βλάβης στο εταιρικό δίκτυο ή απόρροια δεδομένων από αυτό λόγω μη επαρκούς γνώσης των πληροφοριακών συστημάτων που είναι εγκατεστημένα στο εταιρικό περιβάλλον είτε λόγω αψήφησης των κανονισμών που έχουν επιβληθεί από το διαχειριστή δικτύου.

Αρκετοί είναι αυτοί που συγχέουν ένα σύστημα ανίχνευσης εσωτερικών απειλών με τα συστήματα ανίχνευσης εισβολής(IDS). Ένα IDS ανήκει ή στην κατηγορία HIDS (Host-based Instruction Detection System) ή στην κατηγορία NIDS(Network Instruction Detection System). Τα συστήματα HIDS είναι εγκατεστημένα σε κάθε μηχάνημα και επιβλέπουν το πακέτο από και προς τον υπολογιστή καθώς και συγκεκριμένα αρχεία για τυχόν αλλαγές [5] ενώ τα συστήματα NIDS είναι εγκατεστημένα σε στρατηγικά σημεία στο δίκτυο και επιβλέπουν όλο το δίκτυο ή μέρος του. Ο τρόπος ανίχνευσης στα συστήματα IDS γίνεται είτε μέσω της μεθόδου ελέγχου υπογραφής(signature based) είτε μέσω εύρεσης της στατιστικής ανωμαλίας. Τα IDS που εφαρμόζουν τη μέθοδο ελέγχου υπογραφής ελέγχουν αν ένα πρόγραμμα \ σύνδεση υπάρχει στη βάση τους σαν ύποπτο, ενώ τα συστήματα IDS εύρεσης στατιστικής ανωμαλίας ελέγχουν αν το σύστημα έχει συνηθισμένη συμπεριφορά, αλλιώς ειδοποιούν το διαχειριστή. Καμία όμως από τις δύο μεθόδους των συστημάτων IDS δεν μπορεί να εκπαιδευτεί από μόνη της ενώ πρέπει να τονιστεί ότι τα συστήματα IDS δεν είναι εξειδικευμένα στην επίβλεψη ενεργειών χρήστη.

1.1. Σκοπός και Στόχος της Διατριβής

Το προσωπικό μιας εταιρίας αποτελεί το σημαντικότερο περιουσιακό της στοιχείο αλλά ταυτόχρονα αποτελεί και μία σημαντική απειλή για την ίδια την εταιρεία. Οι σύγχρονες επιχειρήσεις έχουν μετατρέψει όλα τα έγγραφα τους και γενικότερα τα δεδομένα τους από έντυπη σε ηλεκτρονική μορφή. Τα δεδομένα βρίσκονται πλέον αποθηκευμένα σε διακομιστές με σκοπό την ευκολότερη και άμεση πρόσβαση τους από το προσωπικό της εταιρείας. Η πιθανότητα διαρροής των δεδομένων δεν παύει να υπάρχει παρόλο που η

πρόσβαση καθώς και τα επίπεδα πρόσβασης που έχει ο οποιοσδήποτε στα αρχεία δεδομένων είναι ελεγχόμενη.

Η δυσαρέσκεια, ο θυμός, η απογοήτευση αποτελούν συναισθήματα που μπορούν να οδηγήσουν έναν εσωτερικό χρήστη στη δημιουργία επιθυμίας πρόκλησης ζημιάς στην επιχείρηση στην οποία εργάζεται. Τα πιο πάνω συναισθήματα πιθανόν να οδηγήσουν τον εσωτερικό χρήστη σε κλοπή των δεδομένων της επιχείρησης και την προώθηση τους σε τρίτους, σε διαγραφή των δεδομένων ακόμη και σε προσπάθεια καταστροφής του υλικού του διακομιστή με απώτερο σκοπό την απώλεια δεδομένων.

Οι τρόποι που πιθανόν να χρησιμοποιήσει ένας υπάλληλος με σκοπό τη πρόκληση βλάβης σε ένα εταιρικό δίκτυο ποικίλουν. Ο υπάλληλος μπορεί να εξαπολύσει επιθέσεις στους διακομιστές δικτύου χρησιμοποιώντας οποιασδήποτε μορφής ιού(virus), να μεταφέρει αρχεία στο προσωπικό του υπολογιστή είτε αποστέλλοντας τα μέσω ηλεκτρονικού ταχυδρομείου είτε αντιγράφοντας τα σε εξωτερική μνήμη(π.χ. USB, εξωτερικός δίσκος). Ο χρήστης είναι πιθανόν ακόμη και να επιχειρήσει τη εξασφάλιση πρόσβασης σε περιοχές του δικτύου που η πρόσβαση του έχει απαγορευτεί από το διαχειριστή δικτύου.

Η απώλεια δεδομένων είναι πιθανόν να μην προέρχεται από εσκεμμένη αλλά από λανθασμένη ενέργεια. Η μη κατανόηση των λόγων που ο διαχειριστής του εταιρικού δικτύου έχει επιβάλει κάποιους κανόνες μπορούν να οδηγήσουν το χρήστη στην παράκαμψη τους με σκοπό την ικανοποίηση των αναγκών του. Αυτή όμως η ενέργεια του πιθανόν να έχει καταστροφικές συνέπειες για την ίδια την επιχείρηση. Οι συνέπειες αυτές δεν αποτελούν όμως το στόχο του εσωτερικού χρήστη αλλά της άγνοιας του ως προς τα πιθανά αποτελέσματα των πράξεων του.

Η παρούσα μεταπτυχιακή διατριβή είχε σαν στόχο την ανίχνευση εσωτερικών απειλών με σκοπό τη διασφάλιση της ακεραιότητας του δικτύου και των δεδομένων που βρίσκονται αποθηκευμένα σε αυτό. Για την επίτευξη του στόχου αυτού αναπτύχθηκε ένα υπολογιστικό σύστημα επίβλεψης των υπολογιστών που συνθέτουν ένα εταιρικό δίκτυο και συλλογής δεδομένων από αυτούς έτσι ώστε με την εφαρμογή μεθόδων τεχνητής νοημοσύνης να επιτυγχάνεται η ανίχνευση επιβλαβών ενεργειών σε ένα εταιρικό δίκτυο πριν αυτές εξαπλωθούν με τεράστιες συνέπειες για την επιχείρηση.

Η μελέτη που πραγματοποιήθηκε καθώς και το σύστημα που αναπτύχθηκε επικεντρώνονται γύρω από το λειτουργικό σύστημα Windows ενώ για την ανίχνευση των πιθανών απειλών εφαρμόστηκε ο αλγόριθμός μη επιτηρούμενης μάθησης SOM.

1.2. Μεθοδολογία και Περιγραφή Κεφαλαίων

Στο κεφάλαιο 2 πραγματοποιείται μια περιγραφή των σημαντικότερών ερευνών που έχουν πραγματοποιηθεί στο χώρο των εσωτερικών απειλών. Μέσα από τη βιβλιογραφική επισκόπηση θα παρατηρηθούν τα σημεία που επικεντρώθηκαν άλλοι ερευνητές καθώς και οι μέθοδοι που εφαρμόστηκαν με σκοπό την ανάλυση, ανίχνευση και πρόβλεψη των εσωτερικών απειλών.

Στο κεφάλαιο 3 παρουσιάζεται λεπτομερώς η αρχιτεκτονική πάνω στην οποία αναπτύχθηκε το σύστημα ανίχνευσης εσωτερικών απειλών. Θα αναλύσουμε τις μονάδες που αποτελούν το σύστημα επίβλεψης ενώ δίνεται ιδιαίτερη έμφαση στα δεδομένα που συλλέγονται και τα οποία είναι επικεντρωμένα στο τρόπο λειτουργίας του λειτουργικού συστήματος Windows. Το δεύτερο σημείο το οποίο αναλύεται στο κεφάλαιο 2 είναι ο αλγόριθμος SOM και πως αυτός αναπτύχθηκε και εφαρμόστηκε στο περιγραφόμενο σύστημα ανίχνευσης εσωτερικών απειλών.

Στο κεφάλαιο 4 παρουσιάζονται τα πειράματα που πραγματοποιήθηκαν με σκοπό την απόδειξη της αποτελεσματικότητας του συστήματος που έχει αναπτυχθεί. Αναλύονται οι λόγοι επιλογής της κάθε κατηγορίας ιών που χρησιμοποιήθηκε στις πειραματικές δοκιμές καθώς και η πειραματική διαδικασία που ακολουθήθηκε. Για κάθε πειραματική δοκιμή παρουσιάζονται τα δεδομένα που συλλέχθηκαν καθώς και τα αποτελέσματα που εξήχθησαν από την εκτέλεση του αλγορίθμου.

Στο κεφάλαιο 5 καταγράφονται τα προβλήματα που αντιμετωπίσαμε κατά τη διάρκεια ανάπτυξης του συστήματος και κατά την εκτέλεση των πειραματικών δοκιμών καθώς και οι αποφάσεις που πήραμε με σκοπό την επίλυση τους. Στο κεφάλαιο αυτό πραγματοποιείται επίσης μια γενική ανάλυση των αποτελεσμάτων που εξήχθησαν μέσω των πειραματικών δοκιμών.

Κεφάλαιο 2
Ανασκόπηση Βιβλιογραφίας

Οι σημαντικές συνέπειες που μπορεί να επιφέρει μια εσωτερική απειλή έχουν οδηγήσει πολλούς ερευνητές στη μελέτη τρόπων ανίχνευσης, πρόληψης και καταπολέμησης τους. Στο κεφάλαιο αυτό θα πραγματοποιηθεί μια επισκόπηση των ερευνών που έχουν γίνει στο χώρο της εσωτερικής απειλής. Ένα δεύτερο σημείο που θα μελετηθεί στο παρόν κεφάλαιο είναι η αποτελεσματικότητα του αλγορίθμου μη επιτηρούμενης μάθησης SOM.

2.1. Εσωτερικές Απειλές

Τα τελευταία χρόνια έχει πραγματοποιηθεί αρκετή έρευνα γύρω από την πρόβλεψη των εσωτερικών απειλών. Ένα σημαντικό ποσοστό της έρευνας αυτής κινείται γύρω από τη σκιαγράφηση του ψυχολογικού προφίλ(Psychological Profiling) των εργαζομένων μιας εταιρίας. Η σκιαγράφηση του ψυχολογικού προφίλ δίνει τη δυνατότητα καταγραφής του ιστορικού του χρήστη καθώς και τη ψυχολογική του κατάσταση σε συγκεκριμένες χρονικές στιγμές. Στα [6] και [7] έχει επιχειρηθεί ένας συνδυασμός επίβλεψης(monitoring) των τεχνικών χαρακτηριστικών του υπολογιστή με την επίβλεψη των στοιχείων συμπεριφοράς του χρήστη.

Ένα πλαίσιο ανίχνευσης εσωτερικών απειλών παρατηρείται στο [8]. Το προτεινόμενο πλαίσιο έχει σαν στόχο την ανίχνευση εσωτερικών απειλών μεσώ της μελέτης τεσσάρων θεματικών αξόνων. Ο πρώτος άξονας μελετά τους λόγους που πιθανόν να ωθήσουν ένα υπάλληλο στην εκδήλωση επίθεσης, για παράδειγμα ένας υποβιβασμός θέσης, απόλυση, μη επαρκής γνώση του συστήματος. Ο δεύτερος άξονας αποτελείται από τη μελέτη των χαρακτηριστικών του χρήστη. Δηλαδή τη μελέτη της προσωπικότητας του όπως για παράδειγμα το ιστορικό του σε παρόμοια περιστατικά, τα καθήκοντα του, οι γνώσεις του. Ο τρίτος άξονας μελέτα τα χαρακτηριστικά πιθανών επιθέσεων και ο τέταρτος άξονας μελετά τα χαρακτηριστικά της επιχείρησης έτσι ώστε να βρεθούν τα τρωτά σημεία του δικτύου της και οι πιθανοί στόχοι επιθέσεων. Το συγκεκριμένο πλαίσιο όμως δεν λαμβάνει υπόψιν του το υλικό(hardware) της επιχείρησης ενώ δεν είναι σύστημα πραγματικού χρόνου(live time).

Μια προσπάθεια ανίχνευσης εσωτερικών απειλών μέσω επίβλεψης της διαδικτυακής πλοήγησης του χρήστη καταγράφεται στο [9]. Το σύστημα συλλέγει το περιεχόμενο της κάθε ιστοσελίδας που επισκέπτεται ο χρήστης του δικτύου. Από το περιεχόμενο που έχει συλλεχθεί αφαιρούνται τα στοιχεία του πρωτοκόλλου HTML καθώς και οι λέξεις από τις οποίες δεν μπορεί να εξαχθεί συμπέρασμα(π.χ. άρθρα, σύνδεσμοι) έτσι ώστε να μπορέσουν να εξαχθούν τα χαρακτηριστικά του κειμένου. Στη συνέχεια συντάσσεται το πακέτο δεδομένων(dataset) και υπολογίζονται οι διαστάσεις του πακέτου έτσι ώστε να μπορεί να εφαρμοστεί ο αλγόριθμος των k-μέσων για την εξαγωγή των χαρακτηριστικών της προσωπικότητας του χρήστη. Ο συνδυασμός των χαρακτηριστικών της

προσωπικότητας του χρήστη θα μας δώσει το τελικό αποτέλεσμα αν δηλαδή ό συγκεκριμένος χρήστης είναι πιθανόν να αποτελέσει εσωτερική απειλή ή όχι.

Μια προσπάθεια ανίχνευσης εσωτερικών απειλών χρησιμοποιώντας μάθηση γράφων και ψυχολογικού πλαισίου καταγράφεται στο [10]. Αρχικά συλλέγονται τα δεδομένα μέσω κοινωνικών δικτύων, ηλεκτρονικών μηνυμάτων, διαδικτυακή πλοήγηση κ.α.. Τα δεδομένα αυτά επεξεργάζονται για την εξαγωγή των χαρακτηριστικών του χρήστη με τελικό σκοπό τη δημιουργία ψυχολογικού προφίλ. Οι συγγραφείς του άρθρου υποστηρίζουν ότι η δημιουργία ψυχολογικού προφίλ έχει σαν αποτέλεσμα τη συρρίκνωση του όγκου δεδομένων καθώς και τη μείωση των λανθασμένων ειδοποιήσεων(false alarms). Για την εξαγωγή των τελικών αποτελεσμάτων χρησιμοποιείται ένα Μπαεσιανό μοντέλο σύντηξης.

Το μοντέλο που προτείνεται στο [7], προσπαθεί να πετύχει την ανίχνευση εσωτερικών απειλών μέσω τεχνικών επίβλεψης και ψυχολογικών τεστ. Αρχικά πραγματοποιείται μια ταξινόμηση των χρηστών βάση του επιπέδου πρόσβασης που κατέχουν καθώς και με τη χρήση ψυχολογικών τεστ. Τα ψυχολογικά τεστ προσπαθούν να ανιχνεύσουν το επίπεδο γνώσης του χρήστη, το αν έχει ή όχι προδιάθεση για παράνομη ενέργεια καθώς και το αν ο συγκεκριμένος χρήστης έχει ψηλά επίπεδα στρες. Τα δεδομένα αυτά σε συνδυασμό με τα δεδομένα πραγματικού χρόνου που συλλέγονται παρέχονται σαν είσοδο στο σύστημα διαχείρισης αποφάσεων για την εξαγωγή του επιπέδου επικινδυνότητας του κάθε χρήστη. Σύμφωνα με τους συγγραφείς ο αλγόριθμος που θα πρέπει να χρησιμοποιείται στο σύστημα διαχείρισης αποφάσεων είναι διαφορετικός σε κάθε οργανισμό εφόσον εξαρτάται από τα δεδομένα που ο κάθε οργανισμός θα αποφασίσει να χρησιμοποιήσει.

Παρόλο που η σκιαγράφηση του ψυχολογικού προφίλ των χρηστών ενός εταιρικού δικτύου μπορεί να μας βοηθήσει αρκετά κατά τη πρόβλεψη και ανίχνευση των εσωτερικών απειλών, προϋποθέτει την ύπαρξη εξειδικευμένου προσωπικού για την εξαγωγή συμπερασμάτων ενώ πιθανό να καταπατά το προσωπικό απόρρητο του εργαζομένου.

Η σημαντική συμβολή που μπορεί να προσφέρει η μελέτη των αρχείων που καταγράφει ένας διακομιστής ιστού(Web Server) και αφορούν τα σημαντικές ενέργειες και τα σφάλματα του διακομιστή κατά την ανίχνευση εσωτερικής απειλής τονίζεται στο [11]. Η μελέτη των αρχείων του διακομιστή είναι μια αρκετά χρονοβόρα διαδικασία σε σημείο

που είναι σχεδόν αδύνατο να εκτελείται καθημερινά σε ένα πραγματικό περιβάλλον επιχείρησης.

Η δυνατότητα ανίχνευσης εσωτερικής απειλής χρησιμοποιώντας τεχνικές εξόρυξης πληροφορίας με συνεχή ροή δεδομένων(stream mining) και εξόρυξης γράφων(graph mining) προτείνεται στο [12]. Στην μελέτη αυτή διατυπώνονται οι δυσκολίες που συναντώνται κατά την ανίχνευση εσωτερικών απειλών χρησιμοποιώντας τεχνικές επιτηρούμενης μάθησης(supervised learning). Για την υπερπήδηση των δυσκολιών αυτών προτείνεται η χρησιμοποίηση αλγόριθμου μη επιτηρούμενης μάθησης(unsupervised learning) και πιο συγκεκριμένα του αλγορίθμου GRAD(Graph Based Anomaly Detection). Η προτεινόμενη διαδικασία αποτελείται από τη συλλογή των δεδομένων, την αποθήκευση τους σε αρχεία CSV(Comma Separated Values) και την είσοδο τους στους αλγορίθμους εξόρυξης πληροφορίας χρησιμοποιώντας συνεχή ροή δεδομένων έτσι ώστε να γίνει η ομαδοποίηση(classification) των δεδομένων. Το παραγόμενο σύνολο αποτελεί την είσοδο του αλγόριθμου GRAD για την εύρεση των εσωτερικών απειλών. Σύμφωνα με τον ορισμό του ο αλγόριθμος GRAD είναι σε θέση να επεξεργαστεί μόνο πέντε σύνολα ανά επανάληψη κάτι που μειώνει τη παρεχόμενη ευελιξία και αυξάνει το διαχειριστικό και υπολογιστικό κόστος.

Στο [13] προτείνεται ένα σύστημα πρόβλεψης εσωτερικών απειλών αποτελούμενο από τρείς βασικές μονάδες(modules). Η πρώτη μονάδα είναι η μονάδα ανίχνευσης, η δεύτερη η μονάδα ανάλυσης και η τρίτη η μονάδα διαχείρισης συστήματος. Η συλλογή των δεδομένων γίνεται μέσω της μονάδας επίβλεψης βάση των κριτηρίων επίβλεψης. Η μονάδα επίβλεψης συλλέγει πληροφορίες οι οποίες προέρχονται από τρείς κατηγορίες: τη κατηγορία συστήματος αρχείων(file system), τη κατηγορία μνήμης, τη κατηγορία εισόδου/εξόδου και τη κατηγορία υλικού. Στη συγκεκριμένη μονάδα επίβλεψης παρατηρούμε ότι δεν συλλέγονται όλα τα δεδομένα αλλά τα δεδομένα που τηρούν τα κριτήρια επίβλεψης. Μια σημαντική παρατήρηση είναι το ότι συλλέγονται μόνο οι συνδέσεις TCP/UDP που πραγματοποιούνται στις πόρτες(ports) οι οποίες βρίσκονται σε μια λίστα τρωτών πορτών(vulnerability ports). Στη παρούσα εργασία όμως δεν είναι ξεκάθαρο πώς καθορίζεται η λίστα αυτή, ποιος είναι ο υπεύθυνος για την ανανέωση της λίστας αυτής και το πιο σημαντικό πόσο έγκυρη είναι η επιλογή των συγκεκριμένων πορτών.

Ανίχνευση εσωτερικών απειλών μέσω της σύγκρισης διαφόρων χαρακτηριστικών υπολογιστή καταγράφεται στο [14]. Η προτεινομένη μεθοδολογία αποτελείται από τρία βήματα: τη συλλογή δεδομένων, την εξαγωγή χαρακτηριστικών και την ανίχνευση των εσωτερικών απειλών χρησιμοποιώντας τον αλγόριθμο k πλησιέστερου γείτονα(k nearest neighbor). Για την αξιολόγηση της απόδοσης της προτεινόμενης μεθοδολογίας χρησιμοποιούνται τρία είδη συλλογής χαρακτηριστικών: n-grams των προγραμμάτων που χρησιμοποιούνται, η συχνότητα με την οποία εκτελούνται τα προγράμματα και οι παράμετροι που δίνονται στα εκτελέσιμα προγράμματα. Ο αλγόριθμος κ πλησιέστερου γείτονα ανήκει στη κατηγορία επιτηρούμενης μάθησης. Οι αλγόριθμοι επιτηρούμενης μάθησης υποθέτουν ότι είναι δυνατόν να συλλέξουν δεδομένα εκπαίδευσης(training data) που να είναι απαλλαγμένα εντελώς από κακόβουλες ενέργειες είτε πραγματικά δεδομένα που να περιέχουν την ετικέτα της κανονικής ή κακόβουλης ενέργειας. Μια τέτοια υπόθεση στη περίπτωση ανίχνευσης εσωτερικών απειλών μπορεί να αποβεί μοιραία.

Μια προσπάθεια ανάλυσης των ατόμων που συνθέτουν μια επιχείρηση με σκοπό την ανίχνευση των κακόβουλών προθέσεων τους καταγράφεται στο [15]. Στην προτεινόμενη αρχιτεκτονική συστήματος τα δεδομένα αρχικά συλλέγονται από διάφορους αισθητήρες. Τα δεδομένα που συλλέγονται ελέγχονται από το σύστημα Stealthwatch για την εύρεση ανωμαλιών και τη παραγωγή ειδοποιήσεων(alerts). Οι ενέργειες του χρήστη αναλύονται χρησιμοποιώντας δομημένη ανάλυση(Structured Analysis) με σκοπό την εύρεση ύποπτων συμπεριφορών. Τα αρχικά δεδομένα που συλλέχθηκαν, οι ύποπτες συμπεριφορές που βρέθηκαν και οι ειδοποιήσεις που παράχθηκαν συγχωνεύονται έτσι ώστε να παραχθούν όμοιου τύπου δεδομένα έτσι ώστε να μπορέσει το σύστημα να τα συγκρίνει για αποφασίσει αν είναι κακόβουλά ή όχι. Η μη εφαρμογή οποιουδήποτε αναγνωρισμένου αλγορίθμου κατά τη συγχώνευση και εύρεση των κακόβουλων εγγράφων δεν δίνει την απαραίτητη αξιοπιστία στο όλο σύστημα.

Μια σύγκριση των αποτελεσμάτων των τεχνικών μη επιτηρούμενης μάθησης έναντι των τεχνικών επιτηρούμενης μάθησης για την ανίχνευση εσωτερικών απειλών καταγράφεται στο [16]. Χρησιμοποιήθηκαν δύο τεχνικές μη επιτηρούμενης μάθησης: η τεχνική της ελάχιστης απόστασης(Minimum Distance) και η τεχνική της συμπαγούς ομάδας(Compact Cluster). Ως είσοδος χρησιμοποιήθηκε το πακέτο δεδομένων(data set) του [17]. Το σύνολο αυτό περιλαμβάνει τις εντολές συστήματος(system calls) που

πραγματοποιήθηκαν σε λειτουργικό σύστημα UNIX από ένα σύνολο 50 χρηστών. Το συγκεκριμένο σύνολο δεδομένων δεν περιλαμβάνει τον όγκο δεδομένων που μεταφέρεται προς και από το δίκτυο καθώς και τις μετρήσεις του υπολογιστή κατά την ώρα εκτέλεσης των εντολών από τους χρήστες.

Η σημαντικότητα διασύνδεσης μεταξύ αρχείων – χρηστών και διαδικασιών(processes) – χρηστών στην εύρεση εσωτερικών απειλών επισημαίνεται στο [18]. Οι ερευνητές θέλοντας να αποδείξουν τη σημαντικότητα της πιο πάνω διασύνδεσης έχουν δημιουργήσει ένα σύστημα επίβλεψης των αρχείων και των διαδικασιών του υπολογιστικού συστήματος. Η επιλογή των επιβλεπομένων αρχείων και διαδικασιών γίνεται βάση της ειδικότητας του χρήστη κάτι που μας αφαίρει την δυνατότητα ανεύρεσης της μιας από τις δύο κατηγορίες εσωτερικής απειλής που είναι η λανθασμένη χρήση.

2.2. Αλγόριθμος SOM

Ο αλγόριθμος SOM έχει αναφερθεί αρκετές φορές στη βιβλιογραφία ως μέσο ανίχνευσης ανωμαλιών ή κακής χρήσης(misuse). Συμφώνα με την έρευνα μας δεν έχει αναφερθεί καμία έρευνα για χρήση του αλγορίθμου SOM στο χώρος της εύρεσης εσωτερικών απειλών σε εταιρικό δίκτυο.

Ένα υβριδικό σύστημα ανίχνευσης εισβολής(Instruction Detection System) με τη χρήση του αλγορίθμου SOM για διορατική απεικόνιση της εισβολής καθώς και τη τεχνική RPROP (Resilient Propagation Neural Network) για την ομαδοποίηση των απειλών παρουσιάζεται στο [19]. Η υβριδική αυτή προσπάθεια ανίχνευσης είχε επιτυχές ποσοστό ανίχνευσης 90%. Το συγκεκριμένο σύστημα ανίχνευσης επικεντρώνεται στην εύρεση των επιθέσεων πρωτοκόλλων δικτύου TCP SYN πλημμύρας και ανίχνευσης πορτών.

Η σημαντική συμβολή που μπορεί να έχει ο αλγόριθμος SOM σε συστήματα εύρεσης εισβολών διατυπώνεται στο [20]. Για την απόδειξη της οι συγγραφείς ανέπτυξαν ένα σύστημα εύρεσης εισβολών σε συστήματα Linux βασισμένο στην προβληματική συμπεριφορά χρήστη. Για την συλλογή των δεδομένων ανέπτυξαν διαφόρων ειδών αισθητήρες οι οποίοι εγκαταστάθηκαν στον υπολογιστή του χρήστη.

Η σύνδεση των ειδοποιήσεων ενός συστήματος ανίχνευσης εισβολής μέσω της χρήσης του νευρωνικού δικτύου SOM περιγράφεται στο [21]. Μέσω της συγκεκριμένης έρευνας γίνεται προσπάθεια διευκόλυνσης της δύσκολης εργασίας του διαχειριστή για την εύρεση εισβολών στο δίκτυο που διαχειρίζεται μέσω της ομαδοποίησης των ειδοποιήσεων που δημιουργούνται βάση συγκεκριμένων χαρακτηριστικών(features).

Ένα σύστημα ανίχνευσης εισβολών που συνδυάζει την εύρεση ανωμαλιών και κακής χρήσης παρουσιάζεται στο [22]. Τόσο η εύρεση ανωμαλιών όσο και η εύρεση κακής χρήσης πραγματοποιείται με τη χρήση του αλγορίθμου SOM, ενώ το τελευταίο επίπεδο του συστήματος συνδυάζει τα ευρήματα τους. Με το τρόπο αυτό οι ερευνητές επιτυγχάνουν μια σημαντική μείωση των λανθασμένων θετικά ευρημάτων δημιουργώντας ένα αρκετά αξιόπιστο σύστημα.

Κεφάλαιο 3
Περιγραφή Συστήματος

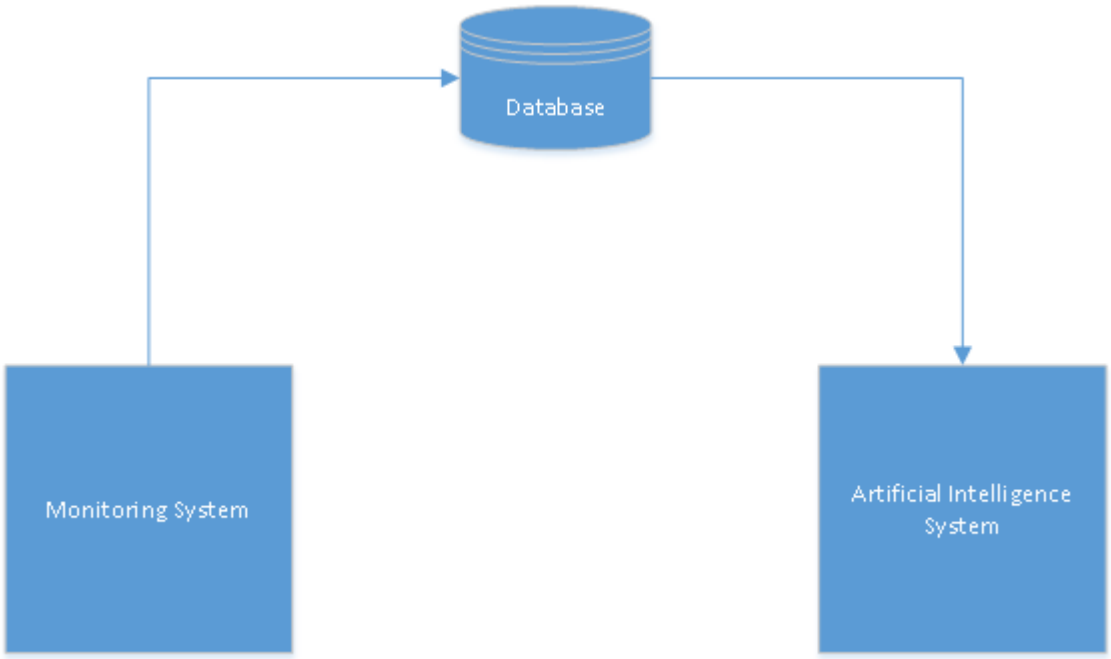
Σε αυτό το κεφάλαιο θα αναλυθεί το σύστημα που αναπτύχθηκε με σκοπό την ανίχνευση εσωτερικών απειλών σε ένα εταιρικό δίκτυο. Θα αναλυθεί η αρχιτεκτονική λογισμικού η οποία ακολουθήθηκε καθώς και τα δεδομένα που συλλέγονται με σκοπό να αποτελέσουν την είσοδο του αλγόριθμου SOM για την εύρεση πιθανών απειλών.

3.1. Αρχιτεκτονική Συστήματος

Ο κυριότερος σκοπός της παρούσας μεταπτυχιακής διατριβής, αφορά την ανάπτυξη ενός συστήματος ανίχνευσης εσωτερικών απειλών σε επιχειρησιακά περιβάλλοντα που χρησιμοποιούν λειτουργικό σύστημα Windows. Οι λόγοι που μας οδήγησαν στην λήψη της απόφασης επιλογής του λειτουργικού συστήματος Windows και μελέτη των εσωτερικών απειλών σε αυτό, είναι η μεγάλη χρήση του στα επιχειρησιακά περιβάλλοντα, η πληθώρα ιών(virus, Trojans, bots) καθώς και η παρατήρηση ότι έχουν γίνει ελάχιστες ερευνητικές μελέτες στο χώρο της ανίχνευσης εσωτερικών απειλών που να αφορούν αυτό το λειτουργικό σύστημα. Η κυριότερη ερευνητική μελέτη που είναι προσανατολισμένη σε λειτουργικό σύστημα Windows παρατηρείται στο [23]. Η συγκεκριμένη μελέτη ερευνά το γενικότερο χώρο εύρεσης απειλών(IDS) αναπτύσσοντας ένα σύστημα ανίχνευσης επικεντρωμένο στην επίβλεψη του αρχείου registry του λογισμικού Windows.

Το σύστημα που αναπτύχθηκε αποτελείται από δύο βασικά μέρη (Εικόνα 3.1) που είναι το σύστημα επίβλεψης και το σύστημα τεχνητής νοημοσύνης. Το σύστημα επίβλεψης είναι εγκατεστημένο στους υπολογιστές που ανήκουν στο επιχειρησιακό δίκτυο(client PC's). Είναι επιφορτισμένο με τη συλλογή διαφόρων ειδών πληροφορίες και την αποθήκευση τους στη βάση δεδομένων. Οι πληροφορίες συλλέγονται μετά από εντολή που στέλνεται από το διακομιστή στον οποίο βρίσκεται εγκατεστημένο το λογισμικό διακομιστή. Το λογισμικό διακομιστή αποτελεί το κεντρικό σημείο του συστήματος και είναι επιφορτισμένο με το συντονισμό των υπολογιστών χρηστών και την επικοινωνία με την βάση δεδομένων. Εφόσον οι πληροφορίες καταχωρηθούν στη βάση δεδομένων υπάρχει δυνατότητα εκτέλεσης του αλγορίθμου μη επιτηρούμενης μάθησης SOM, για την εύρεση των πιθανών εσωτερικών απειλών.

Για την υλοποίηση του συστήματος χρησιμοποιήθηκε η αρχιτεκτονική ανάπτυξης λογισμικού, τριών επιπέδων(3-tier). Χρησιμοποιώντας την αρχιτεκτονική αυτή θελήσαμε να αποκλείσουμε την απευθείας επικοινωνία των υπολογιστών των χρηστών με τη βάση δεδομένών, περιορίζοντας έτσι την πιθανότητα προσπάθειας παραποίησης ή διαγραφής των δεδομένων. Η επικοινωνία μεταξύ επιπέδων γίνεται μέσω του πρωτοκόλλου TCP/IP ενώ τα δεδομένα μεταφέρονται σε κρυπτογραφημένη μορφή.



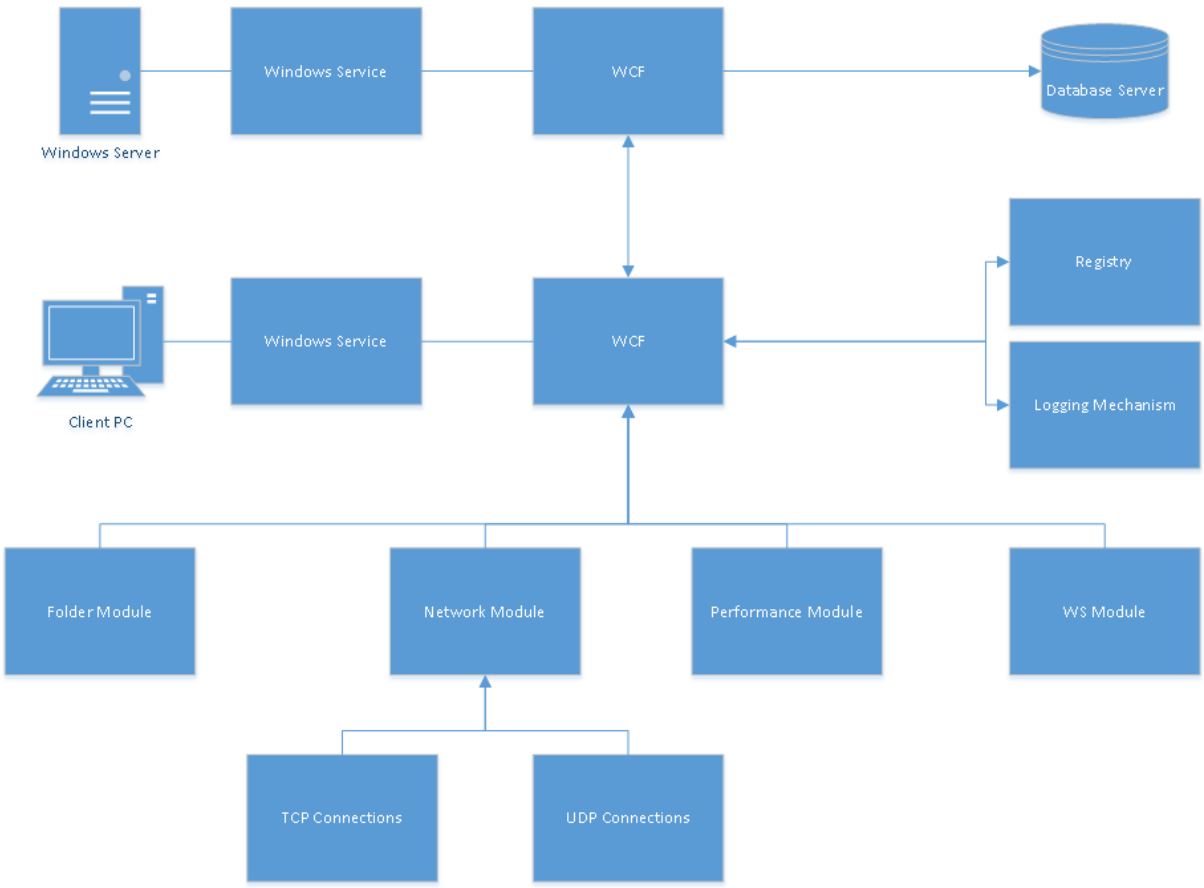
Εικόνα 3.1: Βασικά μέρη συστήματος

Στην Εικόνα 3.2 παρουσιάζεται μια πιο λεπτομερής απεικόνιση της αρχιτεκτονικής του συστήματος. Τόσο στον κεντρικό διακομιστή όσο και στους υπολογιστές χρηστών υπάρχουν εγκατεστημένα τα αντίστοιχα Windows Services, κάθε ένα από τα οποία φιλόξενά ένα WCF(Windows Communication Foundation).

Το WCF [24] είναι ένα πλαίσιο ανάπτυξης υπηρεσιών σε περιβάλλοντα Windows. Με τη χρησιμοποίηση του πλαισίου WCF γίνεται δυνατή η ασύγχρονη επικοινωνία μεταξύ δέκτη και αποστολέα. Χρησιμοποιώντας το χαρακτηριστικό αυτό επιτυγχάνεται η αισθητή μείωση της επιβάρυνσης που παρατηρείτο στους υπολογιστές κατά τη συλλογή δεδομένων από το σύστημα ανίχνευσης που αναπτύχθηκε. Ένα δεύτερο χαρακτηριστικό του πλαισίου WCF που χρησιμοποιήσαμε είναι η δυνατότητα κρυπτογράφησης και αποκρυπτογράφησης των δεδομένων που αποστέλνονται και παραλαμβάνονται αντίστοιχα. Με τον τρόπο αυτό διασφαλίζουμε τα δεδομένα από οπουδήποτε λογισμικό παρακολούθησης(sniffer).

Για την διευκόλυνση ανίχνευσης πιθανών προβλημάτων(troubleshooting) από τον διαχειριστή του συστήματος έχει αναπτυχθεί μηχανισμός καταγραφής μηνυμάτων(logging mechanism) τόσο για το λογισμικό υπολογιστή χρήστη όσο και για το λογισμικό διακομιστή. Ο μηχανισμός καταγραφής μηνυμάτων αποτελείται από δύο

επιμέρους μηχανισμούς: το μηχανισμό καταγραφής στο αρχείο γεγονότων(event viewer) των Windows και το μηχανισμό καταγραφής σε αρχείο κειμένου. Στο αρχείο γεγονότων καταγράφονται τα πιο σημαντικά γεγονότα ενώ στο αρχείο κειμένου γίνεται μια πιο λεπτομερής καταγραφή. Ο μηχανισμός καταγραφής που εμπερικλείεται στο λογισμικό υπολογιστή χρήστη καταγράφει όσο το δυνατό λιγότερες λεπτομέρειες έτσι ώστε να γίνεται δυσκολότερη η κατανόηση του τρόπου λειτουργίας του συστήματος από κακόβουλο χρήστη. Θα πρέπει να σημειώσουμε ότι παρέχεται η δυνατότητα απενεργοποίησης του μηχανισμού καταγραφής σε αρχείο.



Εικόνα 3.2: Αρχιτεκτονική συστήματος

3.2. Λογισμικό Υπολογιστή Χρήστη

Το Windows Service που βρίσκεται εγκατεστημένο στον υπολογιστή χρήστη εμπερικλείει το σύστημα επίβλεψης. Το σύστημα επίβλεψης αποτελείται από τέσσερεις βασικές μονάδες. Αυτές είναι η μονάδα επίβλεψης αρχείων, η μονάδα επίβλεψης δικτύου, η μονάδα επίβλεψης απόδοσης και η μονάδα επίβλεψης Windows Services.

Η μονάδα επίβλεψης αρχείων είναι υπεύθυνη για την παραγωγή του αποτυπώματος του κάθε φακέλου που βρίσκεται στη λίστα φακέλων προς επίβλεψη. Για τη συμπλήρωση της συγκεκριμένης λίστας είναι υπεύθυνος ο διαχειριστής συστήματος. Ο διαχειριστής έχει τη δυνατότητα επιλογής συγκεκριμένων φακέλων για κάθε υπολογιστή ξεχωριστά είτε την επιλογή καθολικών φακέλων οι οποίοι θα ισχύουν για όλους τους υπολογιστές του δικτύου. Για τη παραγωγή του αποτυπώματος φακέλου χρησιμοποιείται ο αλγόριθμος MD5 (MD5 hash – checksum). Η παραγωγή του αποτυπώματος μας δίνει τη δυνατότητα ανίχνευσης οποιασδήποτε τροποποίησης στο περιεχόμενο του κρίσιμου φακέλου μέσω της σύγκρισης του αποτυπώματος με το αμέσως προηγούμενο.

Η μονάδα επίβλεψης δικτύου αποτελείται από δύο υπό-μονάδες. Την υπό-μονάδα επίβλεψης συνδέσεων TCP(Transmission Control Protocol) και την υπό-μονάδα επίβλεψης συνδέσεων UDP(User Datagram Protocol or Universal Datagram Protocol). Τα πρωτόκολλα TCP και UDP είναι τα πιο διαδεδομένα για τη μεταφορά δεδομένων μέσω δικτύου. Και τα δύο αυτά πρωτόκολλά έχουν χτιστεί πάνω στο πρωτόκολλο IP(Internet Protocol). Τα δεδομένα συλλέγονται με τη χρήση του εργαλείου NetStat [25]. Η μονάδα επίβλεψης δικτύου μας δίνει τη δυνατότητα παρακολούθησης και καταγραφής των συνδέσεων από και προς το δίκτυο. Με τον τρόπο αυτό μπορούμε να ανιχνεύσουμε αν υπάρχουν συνδέσεις από και προς ύποπτες διευθύνσεις ή αν χρησιμοποιούνται πόρτες(ports) για τις οποίες ο διαχειριστής δεν έχει δώσει την άδεια να χρησιμοποιούνται. Επίσης μας δίνεται η δυνατότητα καταγραφής του αρχείου που έχει δημιουργήσει τη κάθε σύνδεση καθώς και το ποια ήταν η τελευταία ημερομηνία τροποποίησης του συγκεκριμένου αρχείου. Εφόσον έχουμε καταγράψει τα αρχεία που δημιουργούν συνδέσεις μπορούμε να καταχωρήσουμε τα αρχεία αυτά στη μονάδα επίβλεψης αρχείων, που περιεγράφηκε, πιο πάνω έτσι ώστε να αποθηκεύουμε και το αποτύπωμα των αρχείων αυτών.

Η μονάδα επίβλεψης απόδοσης υπολογιστή είναι υπεύθυνη για τη καταγραφή διάφορων δεικτών που σχετίζονται κυρίως με το υλικό(hardware) του υπολογιστή. Οι δείκτες αυτοί αφορούν την κεντρική μονάδα επεξεργασίας(CPU), την μνήμη(RAM), το δίσκο καθώς και τις κάρτες δικτύου. Οι δείκτες αυτοί μας δίνουν τη δυνατότητα ανίχνευσης της κατάστασης του κάθε υπολογιστή ανά πάσα χρονική στιγμή έτσι ώστε να μπορέσουμε να εντοπίσουμε τις οποιεσδήποτε δυσλειτουργίες. Αν για παράδειγμα ο διαχειριστής του συστήματος διαπιστώσει ότι ο υπολογιστής που χρησιμοποιείται από

άτομο του γραμματειακού προσωπικού έχει συνεχώς υψηλά ποσοστά χρήσης μνήμης και δίσκου τότε πιθανότατα να έχουμε να κάνουμε με μια δυσλειτουργία. Μια άλλη περίπτωση είναι η μεγάλη αποστολή δεδομένων, από πολλούς υπολογιστές του δικτύου, τις βραδινές ώρες .

Η μονάδα επίβλεψης των υπηρεσιών Windows(Windows Services) είναι επιφορτισμένη με τη συλλογή των υπηρεσιών Windows που τρέχουν στον υπολογιστή τη συγκεκριμένη χρονική στιγμή. Η μονάδα αυτή, όπως και η μονάδα δικτύου, μας δίνει τη δυνατότητα καταγραφής των αρχείων που είναι υπεύθυνα για την εκτέλεση των υπηρεσιών Windows καθώς και την ημερομηνία τροποποίησης τους. Οι υπηρεσίες Windows [26] είναι προγράμματα τα οποία έχουν την ιδιαιτερότητα ότι τρέχουν στο παρασκήνιο(background) και επομένως η ενεργοποίηση τους δεν γίνεται εύκολα αντιληπτή από έναν απλό χρήστη υπολογιστή. Η δεύτερη σημαντική ιδιότητα τους είναι ότι έχουν την δυνατότητα να τρέχουν κατά την έναρξη (startup) του λειτουργικού συστήματος χωρίς ο χρήστης να επιλέξει την έναρξη του εκτελέσιμου αρχείου (executable).

Οι μονάδες επίβλεψης υπηρεσιών Windows, απόδοσης και δικτύου είναι συνδεδεμένες μεταξύ τους. Το λογισμικό υπολογιστή χρήστη δίνει εντολή συλλογής στοιχείων και από τα τρία συστήματα την ίδια χρονική στιγμή. Με την ομαδοποίηση αυτή μας δίνεται η δυνατότητα σύνδεσης των στοιχείων που θα συλλεγούν και από τα τρία συστήματα. Επομένως μπορούμε, για παράδειγμα, να παρατηρήσουμε ότι όταν μια συγκεκριμένη υπηρεσία Windows είναι ενεργοποιημένη τότε παρατηρούνται συνδέσεις δικτύου προς μια συγκεκριμένη διεύθυνση ενώ ταυτόχρονα παρατηρείται απότομη αύξηση στα δεδομένα που αποστέλλονται από την κάρτα δικτύου καθώς και ο μέσος όρος δεδομένων που διαβάζονται από το δίσκο.

Η μονάδα επίβλεψης αρχείων είναι πιο αυτόνομη και συλλεγεί δεδομένα όταν οι υπόλοιπες τρείς μονάδες δεν είναι ενεργές. Ο λόγος που έγινε ο διαχωρισμός είναι η σημαντική επιβάρυνση του υπολογιστή που παρατηρήθηκε κατά την ενεργοποίηση της μονάδας επίβλεψης αρχείων. Η επιβάρυνση αφορά κυρίως τη κεντρική μονάδα επεξεργασίας του υπολογιστή καθώς και την μνήμη και οφείλεται στον αλγόριθμο MD5 hash που χρησιμοποιείται για την παραγωγή αποτυπώματος αρχείου. Για αυτήν ακριβώς την επιβάρυνση που παρατηρήθηκε θεωρήσαμε πιο σωστή την απομονωμένη

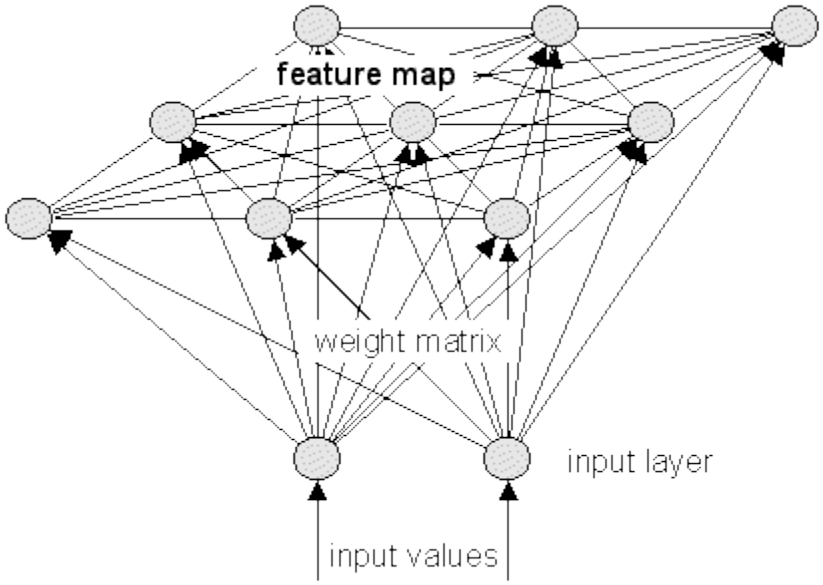
ενεργοποίηση της μονάδας αυτής έτσι ώστε η μονάδα επίβλεψης απόδοσης να είναι σε θέση να πραγματοποιήσει πιο σωστές μετρήσεις.

Για κάθε υπολογιστή συστήματος παράγεται ένας μοναδικός κωδικός αναγνώρισης ο οποίος παράγεται βάση του υλικού και του λογισμικού που περιέχει. Ο μοναδικός κωδικός αποθηκεύεται στο registry του συγκεκριμένου υπολογιστή ενώ γίνεται αυτόματος έλεγχος του έτσι ώστε να αποφευχθεί οποιαδήποτε κακόβουλη τροποποίηση του. Ο λόγος που δεν χρησιμοποιήθηκε το όνομα του υπολογιστή(computer name) σαν αναγνωριστικό είναι για την αποφυγή της περίπτωσης αλλαγής του ονόματος του υπολογιστή, σε μια τέτοια περίπτωση το σύστημα δεν θα μπορούσε να αναγνωρίσει τον υπολογιστή και επομένως δεν θα μπορούσε να ανιχνεύσει τυχόν απειλές ή ανωμαλίες.

3.3. Λογισμικό Διακομιστή

Το Windows Service που είναι εγκατεστημένο στον υπολογιστή είναι επιφορτισμένο με τον συντονισμό των Windows Services των υπολογιστών χρηστών. Από το Windows Service του διακομιστή αποστέλλονται οι εντολές ενεργοποίησης των συστημάτων επίβλεψης και είναι σε αναμονή για την παραλαβή και αποστολή των δεδομένων στη βάση δεδομένων. Το λογισμικό του διακομιστή είναι το μόνο σημείο επικοινωνίας με τη βάση δεδομένων έτσι ώστε να ελαχιστοποιείται η πιθανότητα κακόβουλης ενέργειας.

Επίσης το λογισμικό του διακομιστή δίνει τη δυνατότητα εκτέλεσης του αλγορίθμου SOM. Η εκτέλεση του αλγορίθμου SOM είναι μοναδική για κάθε υπολογιστή. Αρχικά το σύστημα παίρνει τα δεδομένα που βρίσκονται στη βάση δεδομένων και αφορούν το συγκεκριμένο υπολογιστή. Τα δεδομένα που συλλέγηκαν από κάθε εκτέλεση των συστημάτων επίβλεψης αποτελούν και μια ξεχωριστή είσοδο δεδομένων(data entry). Στη συνέχεια το σύνολο των δεδομένων δίνεται ως είσοδο στον αλγόριθμο ο οποίος εξάγει μια λίστα με τις πιθανές εσωτερικές απειλές. Σε κάθε εκτέλεση του ο αλγόριθμος αναβαθμίζει τα βάρη των νευρώνων του για τον κάθε υπολογιστή. Θα πρέπει επίσης να τονιστεί ότι τα δεδομένα που χρησιμοποιούνται είναι αυτά που συλλέχθηκαν μετά τη τελευταία εκτέλεση του αλγορίθμου.



Εικόνα 3.3: Παράδειγμα βαρών και νευρώνων του SOM

Τα αποτελέσματα του αλγορίθμου αποθηκεύονται σε αρχεία CSV(Comma Separated Values) με σκοπό την αποθήκευση, αρχειοθέτηση και τη περεταίρω μελέτη τους από το διαχειριστή. Πιο συγκεκριμένα ο αλγόριθμος εξάγει τη λίστα με την κατηγοριοποίηση των νευρώνων, κάθε νευρώνας μπορεί να κατηγοριοποιηθεί σαν επίθεση, σαν μη-επίθεση ή πιθανόν να μην έχει κάποια κατηγορία. Τα βάρη των νευρώνων τα οποία θα χρησιμοποιηθούν στην επόμενη εκτέλεση αλγορίθμου. Τέλος ο αλγόριθμος εξάγει ακόμη τρία αρχεία: το αρχείο που περιλαμβάνει τα δεδομένα τα οποία δεν έχουν κατηγοριοποιηθεί σαν επίθεση, τα δεδομένα τα οποία δεν έχουν κατηγοριοποιηθεί σαν μη επίθεση και τα δεδομένα τα οποία η κατηγοριοποίηση τους δεν έχει γίνει κατορθωτή. Στην Εικόνα 3.4 παρουσιάζεται ένα παράδειγμα εγγραφών από το εξαγόμενο αρχείο μη επιθέσεων. Οι στήλες που περιέχει το αρχείο αποσκοπούν στο να καθοδηγήσουν το διαχειριστή του συστήματος στην εύρεση των εσωτερικών απειλών.

	A	B	C	D	E	F	G
1	ProcessID	ProcessName	ServiceName	LocalEndPointAddress	LocalEndPointPort	RemoteEndPointAddress	RemoteEndPointPort
2	4	System		192.168.0.21	139	0.0.0.0	0
3	4	System		1.1.1.1	445	1.1.1.1	0
4	784	svchost	RpcEptMapper	0.0.0.0	135	0.0.0.0	0
5	784	svchost	RpcSs	0.0.0.0	135	0.0.0.0	0
6	784	svchost	RpcSs	1.1.1.1	135	1.1.1.1	0
7	784	svchost	RpcEptMapper	1.1.1.1	135	1.1.1.1	0
8	2948	svchost	SSDPSRV	1.1.1.1	1900		0
9	2948	svchost	wcncsvc	1.1.1.1	1900		0
10	1056	svchost	LanmanWorkstation	1.1.1.1	3389	1.1.1.1	0

Εικόνα 3.4: Παράδειγμα αρχείων μη επιθέσεων

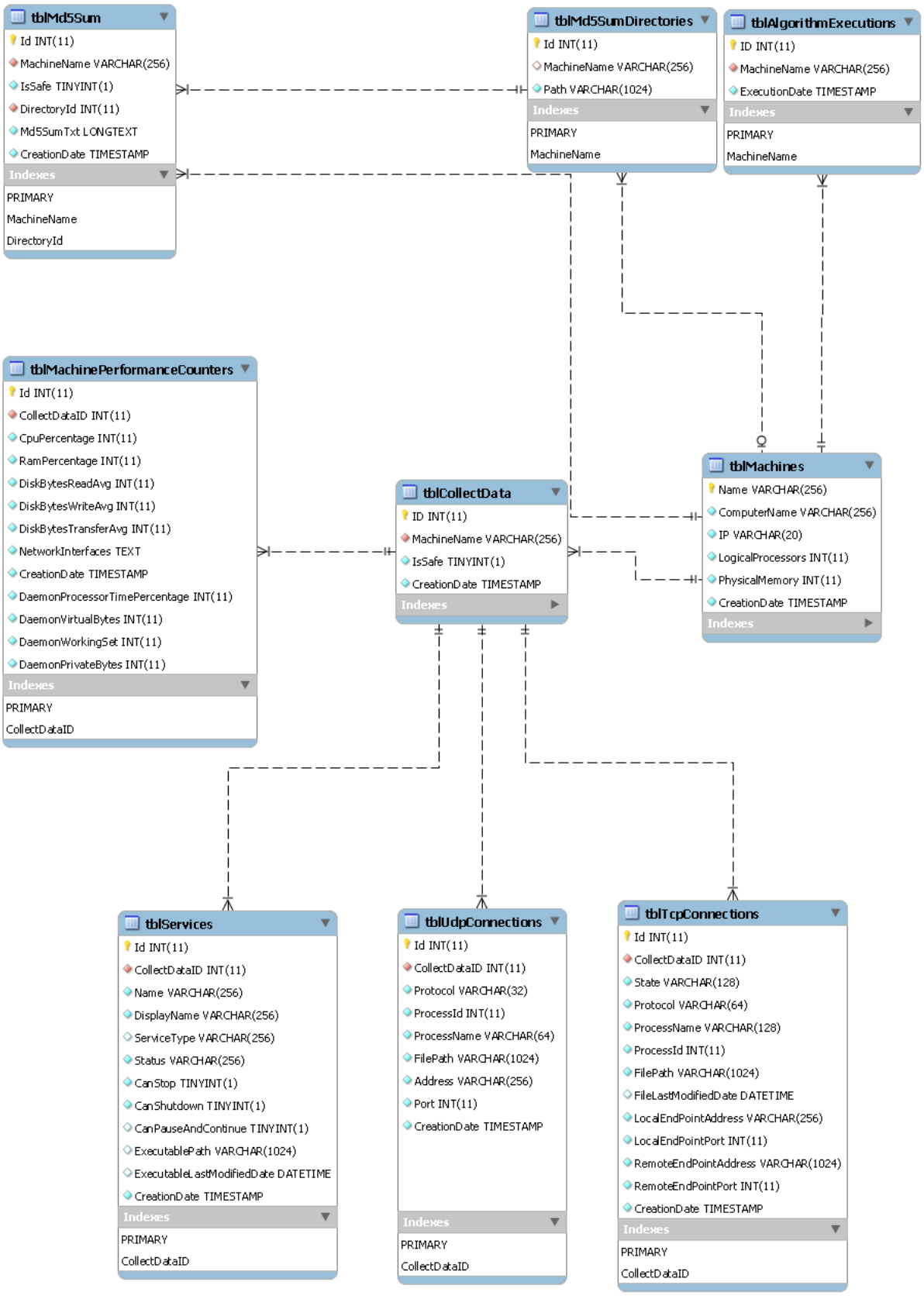
3.4. Περιγραφή Δεδομένων

Τα δεδομένα που συλλέγονται από τα συστήματα επίβλεψης και τα οποία αποτελούν και την είσοδο του αλγορίθμου SOM έχουν επιλεγεί με ιδιαίτερη προσοχή, έχοντας πάντα υπόψη τις ιδιαιτερότητες και την αρχιτεκτονική του λειτουργικού WINDOWS. Στο κεφάλαιο αυτό θα αναλύσουμε τα δεδομένα που συλλέγονται επισημαίνοντας τη σημασία του καθενός.

Στην Εικόνα 3.5 παρουσιάζεται το διάγραμμα βάσης(database datagram). Όταν το λογισμικό υπολογιστή χρήστη εγκατασταθεί σε έναν υπολογιστή καταχωρείται στον πίνακα tblMachines μαζί με τις σχετικές πληροφορίες. Οι πληροφορίες που αποθηκεύουμε είναι το όνομα του υπολογιστή, οι επεξεργαστές που διαθέτει, η μνήμη του καθώς και την διεύθυνση IP που έχει στο δίκτυο. Η διεύθυνση IP που καταχωρείται είναι η εσωτερική διεύθυνση δεδομένου ότι η εξωτερική διεύθυνση είναι πιθανόν η ίδια για όλου τους υπολογιστές ενός εταιρικού δικτύου. Η εσωτερική διεύθυνση IP είναι αυτή που θα χρησιμοποιηθεί για την επικοινωνία μεταξύ του διακομιστή και του υπολογιστή. Τόσο ο αριθμός των επεξεργαστών όσο και το μέγεθος της μνήμης που διαθέτει ο υπολογιστής θα χρησιμοποιηθούν για τον υπολογισμό της απόδοσης, όπως θα αναλυθεί στην συνέχεια.

Ο πίνακας tblMd5Sum περιέχει τα αποτυπώματα των αρχείων που έχουν συλλεγεί από την μονάδα επίβλεψης αρχείων. Θα πρέπει να αναφέρουμε ότι ο πίνακας αυτός είναι άμεσα συνδεδεμένος με τον πίνακα tblMd5SumDirectories ο οποίος περιέχει τα αρχεία, ανά μηχάνημα για τα οποία θα εξαχθεί αποτύπωμα. Το μονοπάτι(path) που καταχωρείται στον πίνακα tblMd5SumDirectories θα πρέπει να είναι απόλυτο(absolute) ενώ αν το πεδίο ονόματος μηχανήματος(Machine Name) δεν συμπληρωθεί σημαίνει ότι το αρχείο που καταχωρήσαμε είναι καθολικό και θα συλλέγεται για όλους τους υπολογιστές.

Όπως ήδη έχουμε αναφέρει τα δεδομένα που συλλέγονται από τα συστήματα επίβλεψης δικτύου, υπηρεσιών και απόδοσης εκτελούνται την ίδια χρονική στιγμή. Η εκτέλεση αυτή αποθηκεύεται στον πίνακα tblCollectData και αυτή που αποτελεί τον συνδετικό κρίκο μεταξύ των δεδομένων, που συλλέχθηκαν από τα τρία συστήματα, και της εγγραφής που αφορά τον υπολογιστή.



Εικόνα 3.5: Διάγραμμα βάσης δεδομένων

Ο πίνακας tblMachinePerformanceCounters περιέχει τα δεδομένα που συλλέχθηκαν από την μονάδα επίβλεψης απόδοσης. Τα δεδομένα αυτά προέρχονται από τους μετρητές απόδοσης(Performance Counters) που παρέχει το λειτουργικό σύστημα Windows. Οι μετρητές απόδοσης μπορούν να γίνουν προσβάσιμοι από οποιοδήποτε χρήστη χρησιμοποιώντας το πρόγραμμα Perfmon [27]. Χρησιμοποιώντας τους μετρητές απόδοσης καταγράφουμε το ποσοστό χρήσης της κεντρικής μονάδας επεξεργασίας καθώς και το ποσοστό χρήσης της μνήμης άμεσης προσπέλασης. Τα δύο αυτά δεδομένα θα μας βοηθήσουν στην ανίχνευση υπερφόρτωσης του υπολογιστή. Επίσης καταγράφουμε τα ποσοστά εγγραφής(write), ανάκτησης(read) από τον δίσκο κατά τη διάρκεια της μέρας. Επομένως αν παρατηρήσουμε μια απότομη άνοδο των ποσοστών αυτών κατά συγκεκριμένα χρονικά διαστήματα τότε αυτή μας η παρατήρηση πιθανόν να αποτελεί και μια ένδειξη δυσλειτουργίας. Η μονάδα επίβλεψης απόδοσης συλλέγει επίσης μια λίστα με όλες τις διεπαφές δικτύου(network interfaces), φυσικές ή εικονικές, καθώς και τον αριθμό των bytes που αποστάλθηκαν και παραλήφθηκαν από την κάθε διεπαφή δικτύου. Μέσω τις λίστας μπορούμε να παρατηρήσουμε αν υπάρχει αδικαιολόγητη εισροή ή εκροή δεδομένων από και προς το δίκτυο μας. Η συλλογή των δεδομένων των εικονικών διεπαφών θα μας φανούν χρήσιμα στη περίπτωση των εικονικών μηχανημάτων(virtual machines). Επίσης καταγράφονται κάποια χαρακτηριστικά απόδοσης του ίδιου του λογισμικού με σκοπό την αφαίρεση τους από τα δεδομένα υπολογιστή κατά την εξαγωγή τους από τη βάση δεδομένων.

Τα δεδομένα του συστήματος επίβλεψης δικτύου αποθηκεύονται στους πίνακες tblTcpConnections και tblUdpConnections ανάλογα με το είδος του πρωτοκόλλου που χρησιμοποιείται. Και στους δύο πίνακες αποθηκεύονται τα χαρακτηριστικά, ταυτότητα και όνομα, της διαδικασίας(process) που έχει δημιουργήσει τη σύνδεση καθώς επίσης τη τοποθεσία του εκτελέσιμου αρχείου το οποίο έχει δώσει εντολή για τη δημιουργία της σύνδεσης. Η υπό-μονάδα επίβλεψης συνδέσεων UDP συλλέγει επίσης την διεύθυνση και τη πόρτα στην οποία έχει γίνει η σύνδεση. Ενώ η υπό-μονάδα επίβλεψης συνδέσεων TCP καταγράφει την τοπική διεύθυνση IP και πόρτα, την απομακρυσμένη(remote) διεύθυνση και πόρτα καθώς και την κατάσταση συμφωνά με τον ορισμό του πρωτόκολλου TCP.

Η μονάδα επίβλεψης υπηρεσιών Windows αποθηκεύει μια λίστα με τις υπηρεσίες και τις πληροφορίες τους στον πίνακα tblServices. Για κάθε μια υπηρεσία που υπάρχει στον υπολογιστή αποθηκεύεται το όνομα της, η κατάσταση στην οποία βρίσκεται, το

εκτελέσιμο αρχείο που είναι υπεύθυνο για την έναρξη της και η τελευταία ημερομηνία τροποποίησης του. Επίσης αποθηκεύουμε και κάποιες σημαίες(flags) σχετικά με τις δυνατότητες αλληλεπίδρασης του χρήστη με τη συγκεκριμένη υπηρεσία.

Κεφάλαιο 4
Πειραματική Διαδικασία – Ανάλυση

Πολλές από τις ερευνητικές μελέτες που καταγραφήκαν για την ανίχνευση εσωτερικών απειλών χρησιμοποιούν ευρέως διαδεδομένα πακέτα δεδομένων(dataset) για την πειραματική τους διαδικασία και την απόδειξη της αποτελεσματικότητας των συστημάτων / πλαισίων τους συγκρίνοντας τα με άλλους ερευνητές και προτεινόμενες μεθόδους. Το πιο διαδεδομένο πακέτο δεδομένων που χρησιμοποιούνται στο χώρο της ανίχνευσης απειλών είναι το πακέτο δεδομένών του οργανισμού DARPA (Defense Advanced Research Projects Agency). Το πακέτο δεδομένων DARPA 1999 [28] περιλαμβάνει δεδομένα δικτύου(tcpdump files) που έχουν συλλέγει κατά τη διάρκεια πέντε εβδομάδων. Τα δεδομένα δικτύου συλλέγονταν από δύο σημεία του εικονικού δικτύου: το ένα σημείο βρισκόταν μεταξύ του router και των τεσσάρων υπολογιστών θυμάτων ενώ το δεύτερο σημείο βρισκόταν μεταξύ του router και του διαδικτύου [29]. Εκτός από δεδομένα δικτύου το πακέτο DARPA 1999 περιείχε επίσης δεδομένα ελέγχου και περιεχόμενα προκαθορισμένων φακέλων.

Ακόμη ένα αρκετά διαδεδομένο πακέτο στο τομέα τις ανίχνευσης επιθέσεων είναι το πακέτο δεδομένων του Schonlau [17]. Για τη δημιουργία του πακέτου καταγράφηκαν δεδομένα 50 χρηστών από το λειτουργικό σύστημα UNIX για περίοδο έξι μηνών. Για κάθε χρήστη καταγράφηκαν 15000 κλήσεις συστήματος(system call). Χρησιμοποιώντας τον όρο κλήσεις συστήματος οι συγγραφείς υπονοούν τις εντολές που ο χρήστης έχει πληκτρολογήσει στο περιβάλλον UNIX.

Όπως έχει ήδη διατυπωθεί, η παρούσα μεταπτυχιακή διατριβή εκτός από την ανίχνευση εσωτερικών απειλών επικεντρώνεται επίσης στο χώρο της εξεύρεσης των χαρακτηριστικών που θα πρέπει να συλλέγουν από ένα υπολογιστή με εγκαταστημένο το λειτουργικό σύστημα Windows έτσι ώστε να παράγονται ακριβέστερα αποτελέσματα κατά την ανίχνευση εσωτερικών απειλών. Για το λόγο αυτό το πακέτο δεδομένων που θα χρησιμοποιούσαμε θα έπρεπε να περιέχει όλες εκείνες τις πληροφορίες που καταγράφονται από το σύστημα επίβλεψης που αναπτύχθηκε και οι οποίες θα μας συμβάλουν στην εξαγωγή ακριβέστερων αποτελεσμάτων .

Ο αρχικός προγραμματισμός προέβλεπε τη χρησιμοποίηση ενός αναγνωρισμένου πακέτου δεδομένων για τις πειραματικές δοκιμές με σκοπό τη προσθήκη περεταίρω εγκυρότητας στα αποτελέσματα. Λόγω του ότι το πακέτο δεδομένων του Schonlau έχει δημιουργηθεί από δεδομένα υπολογιστών με εγκατεστημένο το λειτουργικό σύστημα UNIX δεν περιείχε όλες εκείνες τις πληροφορίες που συλλέγει το σύστημα επίβλεψης και σχετίζονται με τον τρόπο λειτουργίας του λειτουργικού Windows. Παρόλο που το πακέτο δεδομένων DARPA 1999 περιέχει αρκετές από τις πληροφορίες που συλλέγει το σύστημα επίβλεψης όπως τα στοιχεία συνδέσεων TCP και UDP δεν περιέχει τις πληροφορίες των διαδικασιών και υπηρεσιών που δημιουργούν τις συνδέσεις αυτές. Για τους λόγους αυτούς πήραμε την απόφαση να δημιουργήσουμε δικά μας πακέτα δεδομένων που θα πληρούν τις προδιαγραφές που είχαμε θέσει.

Για τη συλλογή των δεδομένων αποφασίστηκε η προσομοίωση της διαδικασίας που θα ακολουθείται σε πραγματικό εταιρικό περιβάλλον. Για το λόγο αυτό συντάθηκε ένα εικονικό δίκτυο που περίεχε τρείς υπολογιστές χρηστών, ένα διακομιστή εφαρμογών(application server) και ένα διακομιστή βάσης δεδομένων(database server). Οι υπολογιστές που συνθέτουν το δίκτυο παρουσιάζονται στον Πίνακας 4.1.

Ρόλος	**Λειτουργικό Σύστημα**	Διεύθυνση IP
Διακομιστής εφαρμογών	Windows Server 2012 R2	192.168.0.14
Διακομιστής βάσης δεδομένων	Ubuntu 14.04 LTS - Mysql 5.5.44	192.168.0.8
Υπολογιστής χρήστη	Windows 7 Ultimate	192.168.0.15
Υπολογιστής χρήστη	Windows 7 Ultimate	192.168.0.21
Υπολογιστής χρήστη	Windows 7 Ultimate	192.168.0.23

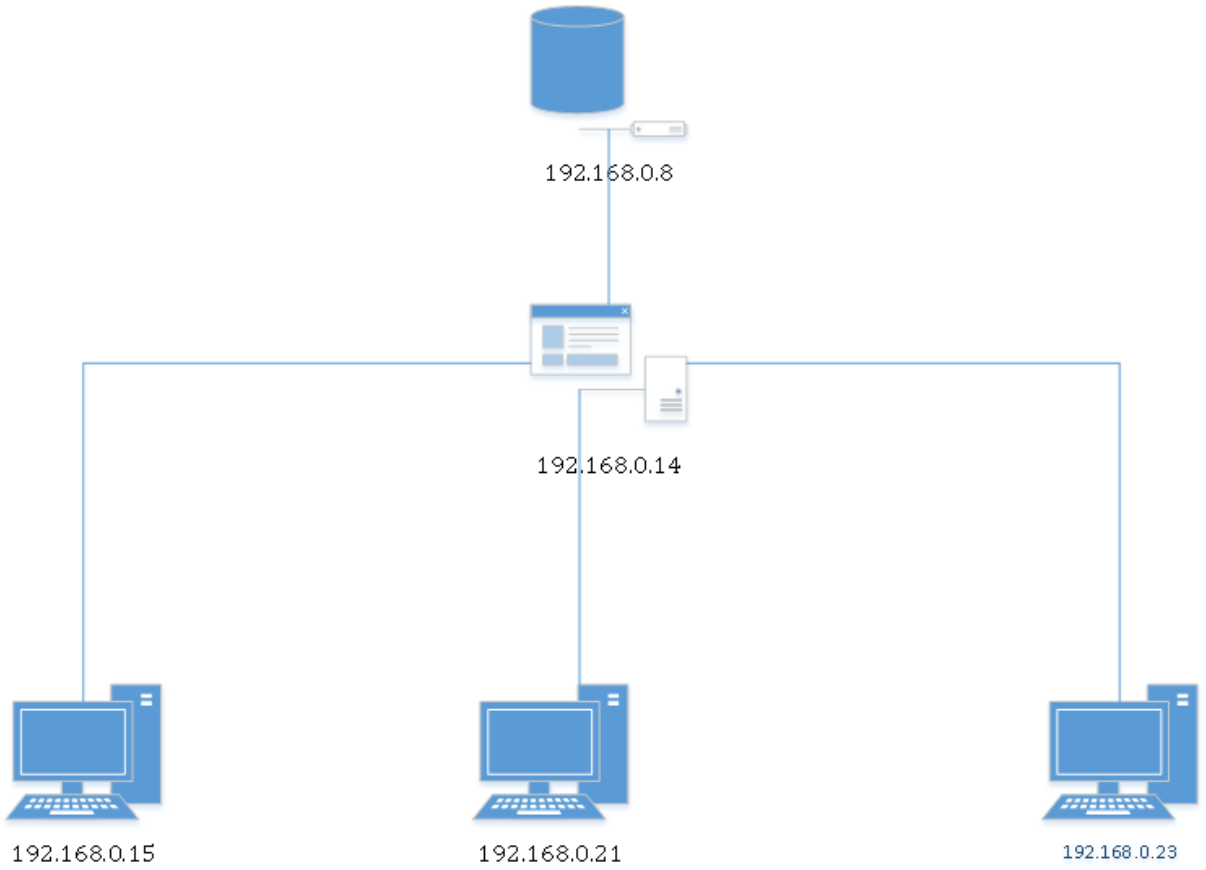
Πίνακας 4.1: Δίκτυο Προσημείωσης

Το λειτουργικό σύστημα που εγκαταστάθηκε στους τρείς υπολογιστές χρήστη καθώς και το λειτουργικό σύστημα του κεντρικού διακομιστή προέρχεται από την οικογένεια Windows. Όλα τα λογισμικά που χρησιμοποιήθηκαν είναι τύπου 64-bit έτσι ώστε να έχουμε πιο σωστές μετρήσεις δεδομένου το ότι τα πλείστα λογισμικά που χρησιμοποιούνται σε εταιρικά περιβάλλοντα είναι του ιδίου τύπου. Για τον ίδιο λόγο τα βασικά χαρακτηριστικά υλικού των υπολογιστών χρηστών ήταν: τετραπύρηνος επεξεργαστής i5 στα 2.27 GHz και 2GB RAM.

Ο διακομιστής που θα φιλοξενήσει τη βάση δεδομένων δεν πρέπει να έχει εγκατεστημένο το λειτουργικό σύστημα WINDOWS. Η δυνατότητα αυτή οφείλεται στην αρχιτεκτονική τριών επιπέδων που χρησιμοποιήθηκε και η οποία διαχωρίζει τα επίπεδα του επιπέδου επιχείρησης(business tier) από το επίπεδο βάσης δεδομένων(database tier). Για την επίδειξη της δυνατότητας που περιεγράφηκε αποφασίστηκε η χρησιμοποίηση του λογισμικού Ubuntu για το διακομιστή βάσης δεδομένων. Η βάση δεδομένων που χρησιμοποιήθηκε είναι η MySQL λόγω της δυνατότητας δωρεάν εγκατάστασης της.

Όπως ήδη αναφέρθηκε, ο διακομιστής εφαρμογών αποτελεί το κεντρικό σημείο ελέγχου του συστήματος που έχει αναπτυχθεί. Όπως φαίνεται και από την Εικόνα 4.1, αποτελεί το μοναδικό σημείο επικοινωνίας μεταξύ της βάσης δεδομένων και των υπολογιστών χρηστών. Είναι επίσης επιφορτισμένος με το ρόλο του συντονισμού των συστημάτων επίβλεψης που βρίσκονται εγκατεστημένα στους υπολογιστές των χρηστών.

Η προσομοίωση της διαδικασίας διήρκησε τέσσερεις ώρες. Με σκοπό την όσο το δυνατό καλύτερη προσομοίωση ενός εταιρικού υπολογιστή δημιουργήθηκε ένα πρόγραμμα υπεύθυνο για την πλοήγηση σε διάφορες ιστοσελίδες ενώ παράλληλα ένα χρήστης χρησιμοποιούσε λογισμικό τροποποίησης εγγράφων και λογισμικό ηλεκτρονικού ταχυδρομείου. Ο λόγος που πραγματοποιήσαμε τις αυτές ενέργειες είναι για να προσημειώσουμε τις καθημερινές εργασίες ενός χρήστη σε ένα εταιρικό υπολογιστή. Κατά τη διάρκεια των τεσσάρων ωρών εξαπολύονταν επιθέσεις σε κάθε ένα από τους τρείς υπολογιστές. Στον κάθε υπολογιστή εγκαταστήσαμε ένα ξεχωριστό ιό(virus). Επιλέχθηκαν ιοί από τρείς διαφορετικές κατηγορίες. Οι κατηγορίες ιών που επιλέχθηκαν είναι πιθανόν να χρησιμοποιηθούν από εσωτερική απειλή με σκοπό την πρόκληση βλάβης στο εταιρικό δίκτυο. Μετά τη συλλογή των πληροφορίων και το πέρας των τεσσάρων ωρών θέσαμε σε εφαρμογή τον αλγόριθμο SOM για κάθε ένα υπολογιστή ξεχωριστά.



Εικόνα 4.1: Τοπολογία Εικονικού Δικτύου

Σε κάθε εκτέλεση του αλγορίθμου ανακτούσαμε από τη βάση δεδομένων τα δεδομένα που αφορούν το συγκεκριμένο υπολογιστή για να δημιουργήσουμε το πακέτο δεδομένων που θα χρησιμοποιούσαμε. Από το κάθε πακέτο δεδομένων χρησιμοποιήσαμε το 70% ως πακέτο εκπαίδευσης και το υπόλοιπο 30% ως πακέτο δοκιμής(test).

Οι κατηγορίες ιών που επιλέχθηκαν είναι DoS, DDoS, RAT / Trojan, Botnet . Στον Πίνακας 4.2 παρουσιάζεται το λογισμικό ιός που επιλέχθηκε ανά κατηγορία. Χρησιμοποιώντας ιό από τη κατηγορία DoS/DDoS (Distributed Denial of Service) θέλαμε να προσομοιώσουμε τη περίπτωση όπου ένας μολυσμένος υπολογιστής στο εσωτερικό δίκτυο μιας εταιρείας εξαπολύει μια τέτοιου είδους επίθεση σε ένα εσωτερικό ή και εξωτερικό διακομιστή (π.χ. διακομιστής ηλεκτρονικού ταχυδρομείου, διακομιστής διαδικτύου) με σκοπό τη προσωρινή η μόνιμη δυσλειτουργία του διακομιστή. Για την εφαρμογή της επίθεσης αυτής ο χρήστης αποστέλλει μεγάλο όγκο αιτημάτων στο διακομιστή στόχο χρησιμοποιώντας τη τεχνική της πλημμύρας. Για τη προσομοίωσης επίθεσης DoS χρησιμοποιήθηκε το λογισμικό LOIC.

Κατηγορία	Ιός
DoS/ DDoS	LOIC
RAT / Trojan	Pandora
Botnet / DDoS	Optima

Πίνακας 4.2: Ιοί που Χρησιμοποιήθηκαν

Η δεύτερη κατηγορία ιών που επιλέχθηκε είναι η κατηγορία ιών RAT / Trojan. Ο βασικός λόγος επιλογής της κατηγορίας αυτής αποτελεί η δυνατότητα που παρέχει στους χρήστες της για απομακρυσμένη σύνδεση. Χρησιμοποιώντας τη δυνατότητα απομακρυσμένης σύνδεσης παρέχεται πρόσβαση στο χρήστη ανά πάσα στιγμή της ημέρας και από οποιοδήποτε σημείο. Επομένως ένας κακόβουλος υπάλληλος ή φίλος του πιθανόν να χρησιμοποιήσει τη κατηγορία αυτή για πρόσβαση από το σπίτι του εκτός ωρών εργασίας. Ο χρήστης που κάνει χρήση λογισμικού της κατηγορίας RAT / Trojan έχει την επιλογή να πραγματοποιήσει οποιαδήποτε ενέργεια επιθυμεί όπως να αντιγράψει αρχεία, να σταματήσει υπηρεσίες του λογισμικού, να ανιχνεύσει το δίκτυο και αρκετές άλλες που μπορούν να αποβούν μοιραίες για το εταιρικό δίκτυο.

Η τελευταία κατηγορία ιών που επιλέχθηκε είναι το δίκτυο μολυσμένων υπολογιστών(Botnet). Κατά την υλοποίηση ενός Botnet ο εισβολέας μολύνει ένα σύνολο από υπολογιστές οι οποίοι συνδέονται μεταξύ τους μέσω ενός κέντρου ελέγχου και εντολών. Το είδος της επίθεσης που θα εξαπολύσει το δίκτυο μολυσμένων υπολογιστών κατά τη πειραματική μας διαδικασία είναι η διανεμημένη άρνηση υπηρεσίας(DDoS). Κατά τη διανεμημένη άρνηση υπηρεσίας ένα σύνολο από μηχανήματα αποστέλλει πλήθος αιτημάτων σύνδεσης προς ένα υπολογιστή / διακομιστή με σκοπό την εξάντληση των πόρων του με αποτέλεσμα την μη ανταπόκριση του σε καινούρια αιτήματα σύνδεσης.

4.1. Επίθεση χρησιμοποιώντας ιό κατηγορίας DDoS

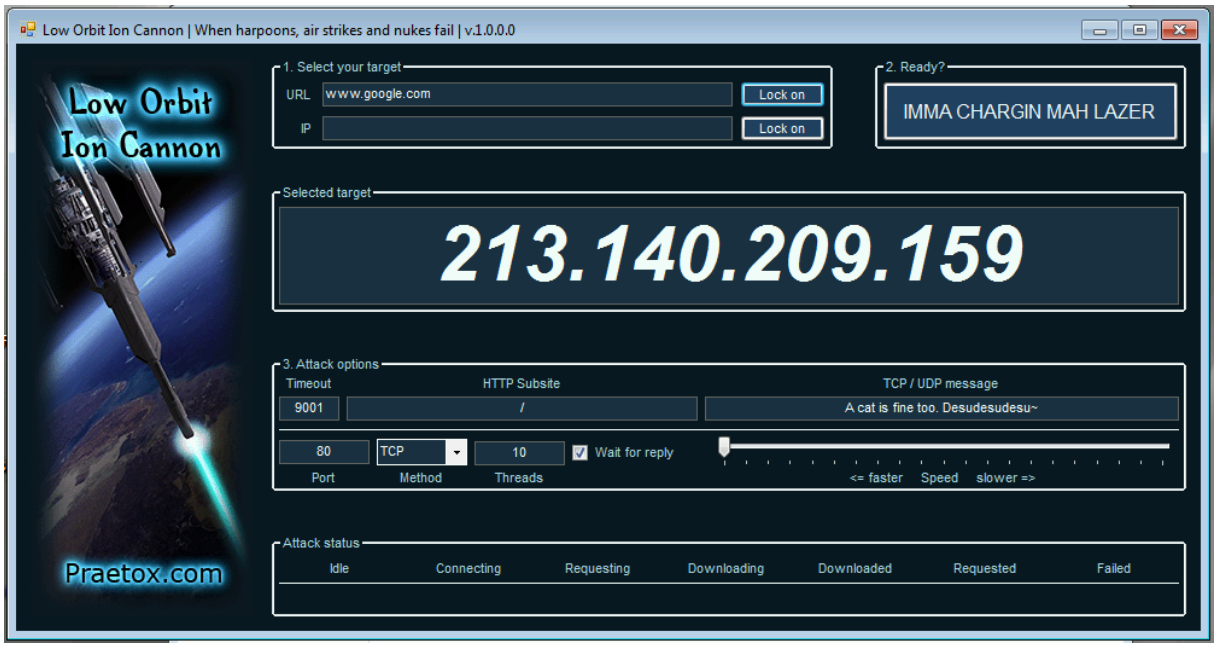
Η κατηγορία εκδήλωσης επιθέσεων με σκοπό την επίτευξη άρνηση υπηρεσίας(DDoS) από το μηχάνημα θύμα είναι από τις πιο διαδεδομένες μορφές επιθέσεων σε ένα διακομιστή διαδικτύου. Κατά την εκδήλωση μιας επίθεσης DDoS ο εισβολέας αποστέλλει μεγάλο αριθμό αιτήσεων σύνδεσης, συχνά μέσω του πρωτόκολλου TCP, προς το διακομιστή θύμα. Ο μεγάλος όγκος αιτήσεων σύνδεσης σε πολύ μικρό χρονικό διάστημα έχει σαν αποτέλεσμα τη αύξηση του χρόνου απόκρισης του του μηχανήματος και τη μείωση των διαθέσιμων πόρων. Όταν όλοι οι πόροι του εξαντληθούν το μηχάνημα δεν μπορεί να δεχτεί άλλο αίτημα σύνδεσης [30].

Η μεγάλη δημοτικότητα της κατηγορίας αυτής σε συνδυασμό με τη πληθώρα των λογισμικών της που μπορεί κάποιος να προμηθευτεί από το διαδίκτυο, την καθιστούν μια από τις πρώτες επιλογές ενός εισβολέα. Επίσης η ευκολία εκδήλωσης μιας επίθεσης DoS είναι από τους πιθανούς λόγους που ένα χρήστης χωρίς ιδιαίτερες γνώσεις σε θέματα ασφάλειας και επιθέσεων μπορεί να χειριστεί ένα τέτοιο λογισμικό. Αυτοί ακριβώς οι λόγοι είναι που μας οδήγησαν στην συμπερίληψη της κατηγορίας αυτής στη διαδικασία των πειραματικών δοκιμών μας.

Σαν αντιπρόσωπο της κατηγορίας αυτής επιλέξαμε το λογισμικό LOIC και το οποίο παρουσιάζεται στην Εικόνα 4.2. ο χρήστης του λογισμικού έχει την δυνατότητα να καταχωρίσει διάφορες ρυθμίσεις. Για παράδειγμα μπορεί να καταχωρήσει είτε το URL είτε τη διεύθυνση IP του διακομιστή στόχου. Από τη στιγμή που ο χρήστης μπορεί να καταχωρήσει τη διεύθυνση IP αντί του URL συμπεραίνουμε την δυνατότητα στόχευσης

διακομιστή εντός εσωτερικού δικτύου. Άλλες ενδιαφέροντες δυνατότητες που ο χρήστης μπορεί να τροποποιήσει είναι το πρωτόκολλο που θα χρησιμοποιηθεί(TCP, UDP, HTTP), τη πόρτα από την οποία θα πραγματοποιηθεί η σύνδεση, τα νήματα(threads) καθώς και τη ταχύτητα αποστολής συνδέσεων. Στη περίπτωση που εξετάζουμε το πρωτόκολλο HTTP είναι εκτός ενδιαφέροντος, δεδομένου ότι εξετάζουμε τη περίπτωση επίθεσης σε εσωτερικό διακομιστή. Αυξάνοντας τον αριθμό των νημάτων αυξάνονται οι ταυτόχρονες συνδέσεις ενώ αυξάνοντας την ταχύτητα αποστολής μειώνεται ο χρόνος μεταξύ της μίας αποστολής αιτήματος σύνδεσης από την επόμενη της.

Χρησιμοποιώντας το λογισμικό LOIC θέλαμε να προσομοιώσουμε όχι μόνο τη περίπτωση που ένας εσωτερικός χρήστης εξαπολύει επίθεση σε ένα εσωτερικό διακομιστή αλλά και τη περίπτωση που επισκέπτεται συχνά μια περιοχή αποθήκευσης αρχείων σε έναν εσωτερικό διακομιστή αρχείων(File Server).



Εικόνα 4.2: Λογισμικό LOIC

Για τη πειραματική μας διαδικασία επιλέξαμε τις επιλογές που παρουσιάζονται στην Εικόνα 4.2. Πιο συγκεκριμένα χρησιμοποιήσαμε το πρωτόκολλο TCP στη πόρτα 80 με 10 νήματα. Όπως έχουμε ήδη αναφέρει η πειραματική διαδικασία διάρκεσε 4 ώρες. Στη συνολική διάρκεια το σύστημα επίβλεψης σύλλεξε δεδομένα 48 φορές(1 φορά ανά 5 λεπτά) ενώ τρέξαμε το λογισμικό LOIC 9 φορές.

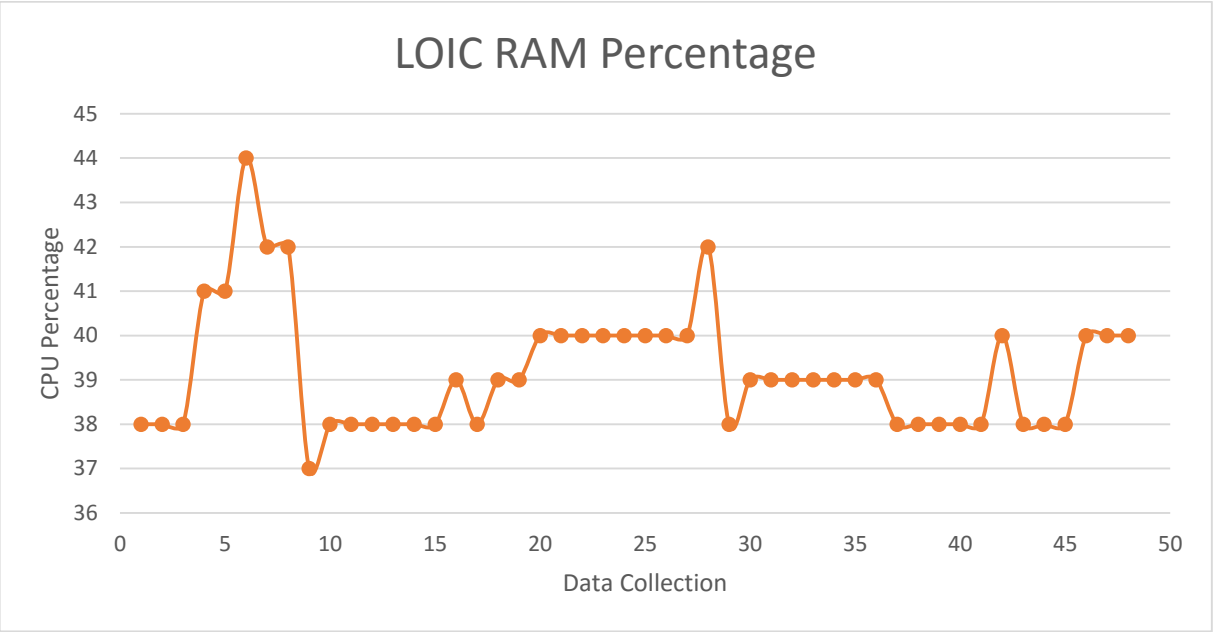
Στον Πίνακας 4.3 παρουσιάζονται ο αριθμός των δεδομένων που συλλέχθηκαν από τις μονάδες επίβλεψης δικτύου, απόδοσης και υπηρεσιών όπως αυτές περιεγράφηκαν στο κεφάλαιο 3.

Μονάδα	Αριθμός δεδομένων
Δικτύου - TCP	742
Δικτύου - UDP	541
Υπηρεσιών	1946
Απόδοσης	48

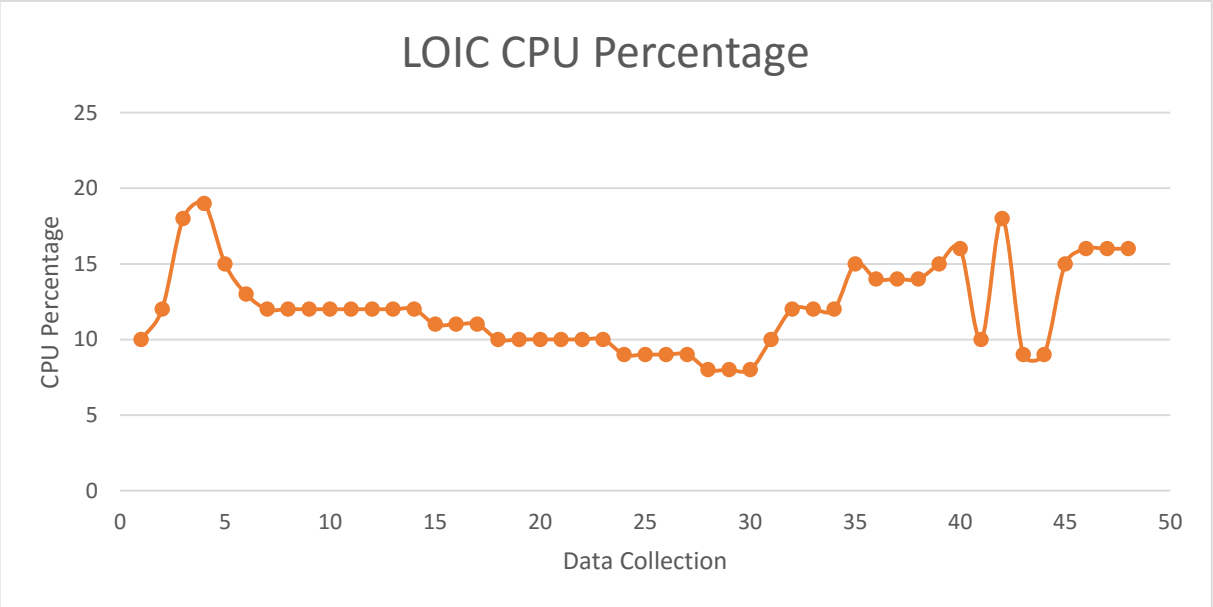
Πίνακας 4.3: Πλήθος δεδομένων κατά τη πειραματική διαδικασία με LOIC

Από τα δεδομένα που έχουν συλλεχθεί τα μόνα που σχετίζονται με το λογισμικό LOIC είναι τα δεδομένα που έχουν συλλεγεί από την υπό-μονάδα TCP της μονάδας επίβλεψης δικτύου. Ο λόγος που η υπό-μονάδα UDP της μονάδας επίβλεψης δικτύου δεν συνέλεξε κανένα δεδομένο που να σχετίζεται με το λογισμικό LOIC είναι λόγω των αρχικών μας ρυθμίσεων όπως αυτή περιεγράφηκε πιο πάνω. Το λογισμικό LOIC δεν δημιουργεί καμία υπηρεσία WINDOWS, τρέχει μόνο όταν ο χρήστης τρέξει το εκτελέσιμο αρχείο, και για αυτόν ακριβώς το λόγο η μονάδα υπηρεσιών δεν συνέλεξε καμία απολύτως πληροφορία για το λογισμικό LOIC.

Από τις 742 εγγραφές δεδομένων που κατέγραψε η υπό-μονάδα TCP της μονάδας επίβλεψης δικτύου, οι 111 αφορούν εγγραφές που σχετίζονται με το λογισμικό LOIC. Σύμφωνα με τα δεδομένα που κατέγραψε η μονάδα επίβλεψης απόδοσης δεν παρατηρήθηκε, όπως καμία απολύτως ανώμαλη συμπεριφορά συγκρίνοντας τα αποτελέσματα πριν και μετά τη χρήση του προγράμματος LOIC, όπως συμπεραίνεται από τα γραφήματα Γράφημα 4.1 και Γράφημα 4.2.



Γράφημα 4.1: LOIC RAM



Γράφημα 4.2: LOIC CPU

Η πρώτη εκτέλεση του αλγορίθμου SOM έγινε χρησιμοποιώντας το 70% των δεδομένων μας ως πακέτο εκπαίδευσης. Η διαδικασία επιλογής του πακέτου εκπαίδευσης έχει γίνει με τη μέθοδο της τυχαίας δειγματοληψίας. Από τις 111 συνδέσεις που συλλέχθηκαν από την υπό-μονάδα TCP της μονάδας επίβλεψης δικτύου οι 83 επιλέχθηκαν να συμπεριληφθούν στο πακέτο εκπαίδευσης.

Στον Πίνακας 4.4 παρουσιάζεται η κατηγοριοποίηση των νευρώνων του αλγορίθμου SOM. Η στήλη N απεικονίζει τη θέση του νευρώνα ενώ η στήλη K τη κατηγορία του νευρώνα. Χρησιμοποιήθηκε ένα πλέγμα νευρώνων 10 * 10, επομένως υπάρχει ένα πλήθος 100 νευρώνων. Από ότι παρατηρούμε από τον πίνακα, 86 από τους 100 νευρώνες έχουν κατηγοριοποιηθεί στη κατηγορία «NONE», δεκατρείς στην κατηγορία «NORMAL» και μόλις ένας νευρώνας στη κατηγορία «ATTACK». Οι λόγοι που παρατηρείται αυτή η συμπεριφορά στους νευρώνας κατά τη κατηγοριοποίηση τους είναι δύο: ο μικρός αριθμός δεδομένων καθώς και η ομοιότητα στα δεδομένα. Η ομοιότητα στα δεδομένα οφείλεται στο ότι πολλά δεδομένα είναι ουσιαστικά τα ίδια αλλά πάρθηκαν σε διαφορετικές χρονικές στιγμές. Αυτή η ομοιότητα και η πολύ καλή ομαδοποίηση που παρατηρείται είναι ένα πολύ καλό χαρακτηριστικό που θα οδηγήσει σε αρκετά καλά αποτελέσματα εφόσον εύκολα θα παρατηρείται η οποιαδήποτε ανωμαλία.

N	K	N	K	N	K	N	K
[0.0]	NONE	[2.5]	NONE	[5.0]	NONE	[7.5]	NORMAL
[0.1]	NONE	[2.6]	NONE	[5.1]	NONE	[7.6]	NONE
[0.2]	NONE	[2.7]	NONE	[5.2]	NONE	[7.7]	NONE
[0.3]	NONE	[2.8]	NONE	[5.3]	NONE	[7.8]	NONE
[0.4]	NONE	[2.9]	NONE	[5.4]	NONE	[7.9]	NORMAL
[0.5]	NONE	[3.0]	NONE	[5.5]	NONE	[8.0]	NONE
[0.6]	NORMAL	[3.1]	NONE	[5.6]	NONE	[8.1]	NONE
[0.7]	NONE	[3.2]	NONE	[5.7]	NORMAL	[8.2]	NONE
[0.8]	NONE	[3.3]	NONE	[5.8]	NONE	[8.3]	NONE
[0.9]	NONE	[3.4]	NONE	[5.9]	NONE	[8.4]	NORMAL
[1.0]	NONE	[3.5]	NONE	[6.0]	NONE	[8.5]	NONE
[1.1]	NONE	[3.6]	NONE	[6.1]	NONE	[8.6]	NONE
[1.2]	NORMAL	[3.7]	NONE	[6.2]	NONE	[8.7]	NONE
[1.3]	NONE	[3.8]	NONE	[6.3]	NONE	[8.8]	NONE

[1.4]	NONE	[3.9]	NONE	[6.4]	NONE	[8.9]	NONE
[1.5]	NONE	[4.0]	NORMAL	[6.5]	NONE	[9.0]	NONE
[1.6]	NONE	[4.1]	NONE	[6.6]	NONE	[9.1]	NONE
[1.7]	NONE	[4.2]	NONE	[6.7]	NONE	[9.2]	NORMAL
[1.8]	NONE	[4.3]	NORMAL	[6.8]	NORMAL	[9.3]	NONE
[1.9]	NORMAL	[4.4]	NONE	[6.9]	NONE	[9.4]	NONE
[2.0]	NONE	[4.5]	NONE	[7.0]	NONE	[9.5]	NONE
[2.1]	NONE	[4.6]	NONE	[7.1]	NORMAL	[9.6]	NONE
[2.2]	NONE	[4.7]	NONE	[7.2]	NONE	[9.7]	NORMAL
[2.3]	NONE	[4.8]	NONE	[7.3]	NONE	[9.8]	ATTACK
[2.4]	NONE	[4.9]	NONE	[7.4]	NONE	[9.9]	NONE

Πίνακας 4.4: Κατηγοριοποίηση Νευρώνων – LOIC

Κατά την διαδικασία της δοκιμής χρησιμοποιήθηκε το υπόλοιπο 30% του πακέτου δεδομένων που συλλέχθηκε από το σύστημα επίβλεψης. Από τις 111 συνδέσεις που συλλέχθηκαν από την υπό-μονάδα TCP της μονάδας επίβλεψης δικτύου οι 28 εναπομείναντες επιλέχθηκαν να συμπεριληφθούν στο πακέτο δοκιμής. Κατά τη διαδικασία αυτή παρατηρήσαμε την απόλυτη ορθότητα των αποτελεσμάτων. Με άλλα λόγια ο αλγόριθμος SOM κατά τη διαδικασία της δοκιμής πέτυχε να κατηγοριοποιήσει 28 εγγραφές σαν «ATTACK» δίχως να έχει κανένα λανθασμένα αρνητικό(false negative) και χωρίς κανένα λανθασμένα θετικό(false positive). Επίσης ο αλγόριθμος δεν τοποθέτησε καμία απολύτως εγγραφή στη κατηγορία «NONE» παρόλο το μεγάλο πλήθος των νευρώνων που ανήκουν σε αυτή τη κατηγορία.

4.2. Επίθεση χρησιμοποιώντας ιό κατηγορίας RAT/Trojan

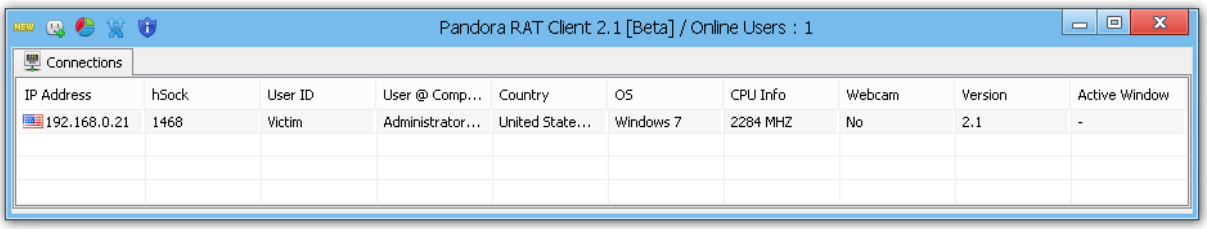
Στη κατηγορία RAT(Remote Administration Tool) ανήκουν τα λογισμικά που δίνουν τη δυνατότητα σε μη εξουσιοδοτημένα άτομα της μερικής ή ολικής χρήσης του

μολυσμένου(infected) υπολογιστή μέσω απομακρυσμένης σύνδεσης [31]. Τα λογισμικά κατηγορίας RAT είναι κομμάτια μολυσμένου κώδικα τα όποια πολλές φόρες ενσωματώνονται σε νόμιμα προγράμματα [32]. Για αυτό ακριβώς το λόγο το ακρωνύμιο RAT ερμηνεύεται και σαν Remote Administration Trojan.

Τα λογισμικά κατηγορίας RAT / Trojan πιθανόν να προκαλέσουν ανεπανόρθωτη ζημιά σε ένα υπολογιστή και κατ' επέκταση σε ένα εταιρικό δίκτυο. Τα λογισμικά αυτά έχουν τη δυνατότητα τροποποίησης και διαγραφής αρχείων, λήψης(download) και δημοσίευσης(upload) δεδομένων, απόσπασης κωδικών, επίβλεψης επιφάνειας εργασίας καθώς και οποιαδήποτε άλλη δυνατότητα προσφέρεται μέσω νόμιμης απομακρυσμένης σύνδεσης [33]. Αυτά ακριβώς τα χαρακτηριστικά της κατηγορίας αυτής είναι που μας οδήγησαν στην επιλογή της για τη πειραματική μας διαδικασία.

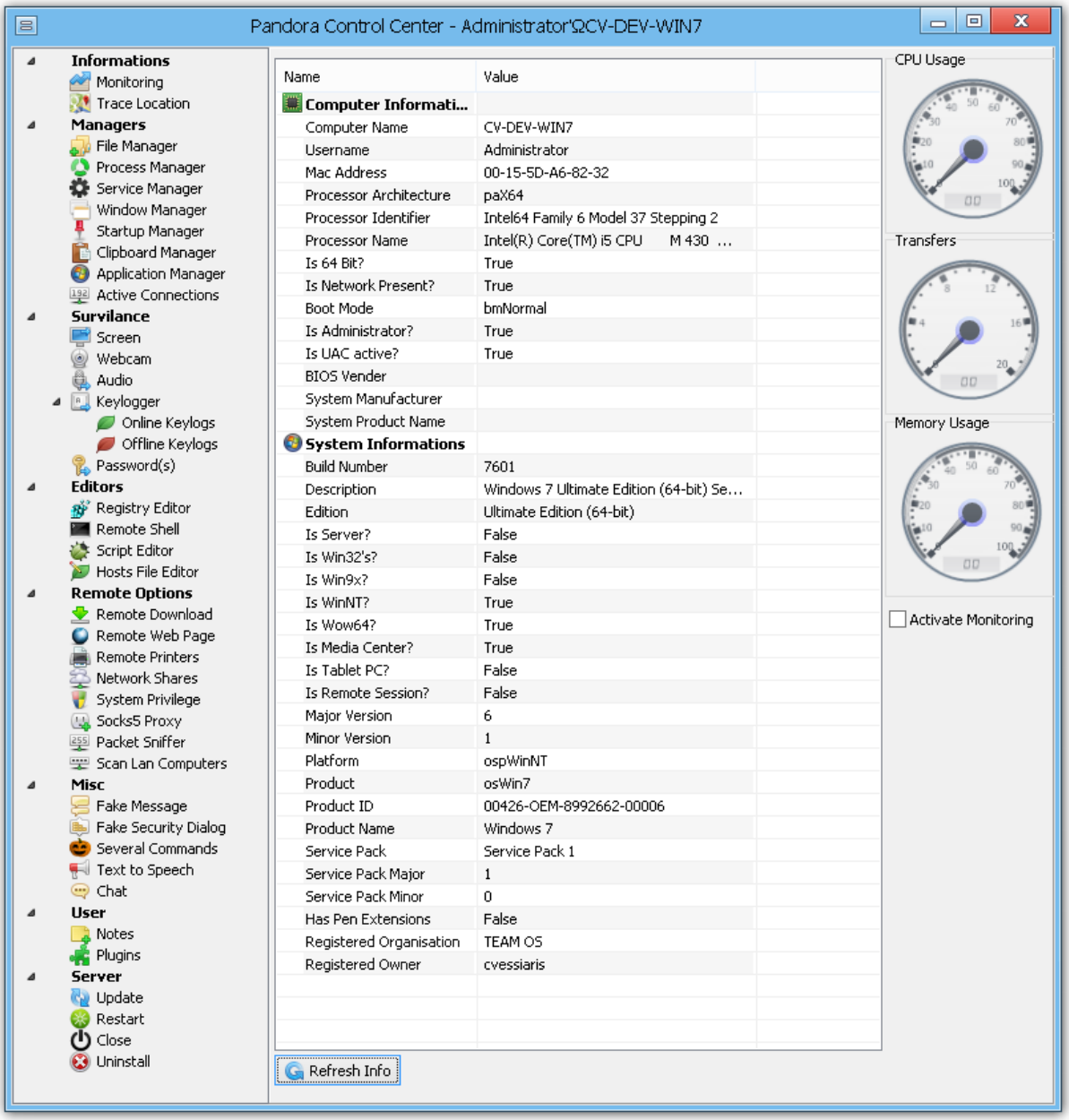
Ένας κακόβουλος χρήστης είναι πιθανόν να εκμεταλλευτεί τη προσωρινή η μόνιμη πρόσβαση του σε ένα υπολογιστή / διακομιστή του εταιρικού δικτύου για να εγκαταστήσει ένα λογισμικό κατηγορίας RAT / Trojan. Με τον τρόπο αυτό ο χρήστης θα έχει απομακρυσμένη πρόσβαση στο εταιρικό δίκτυο και πιθανόν να διατηρήσει τη πρόσβαση του και μετά την απομάκρυνση του από τη εταιρία μετά από απόλυση.

Για τη πειραματική μας διαδικασία επιλέξαμε το λογισμικό Pandora (Εικόνα 4.3) της κατηγορίας RAT / Trojan. Για τη λειτουργία ου λογισμικού Pandora απαιτείται η εγκατάσταση ενός εκτελέσιμου αρχείου που θα λειτουργεί σαν διακομιστής. Στη συνέχεια ο χρήστης χρησιμοποιώντας το λογισμικό χρήστη έχει τη δυνατότητα επισκόπησης της λίστας των υπολογιστικών συστημάτων τα οποία έχουν μολυνθεί με το λογισμικό διακομιστή.



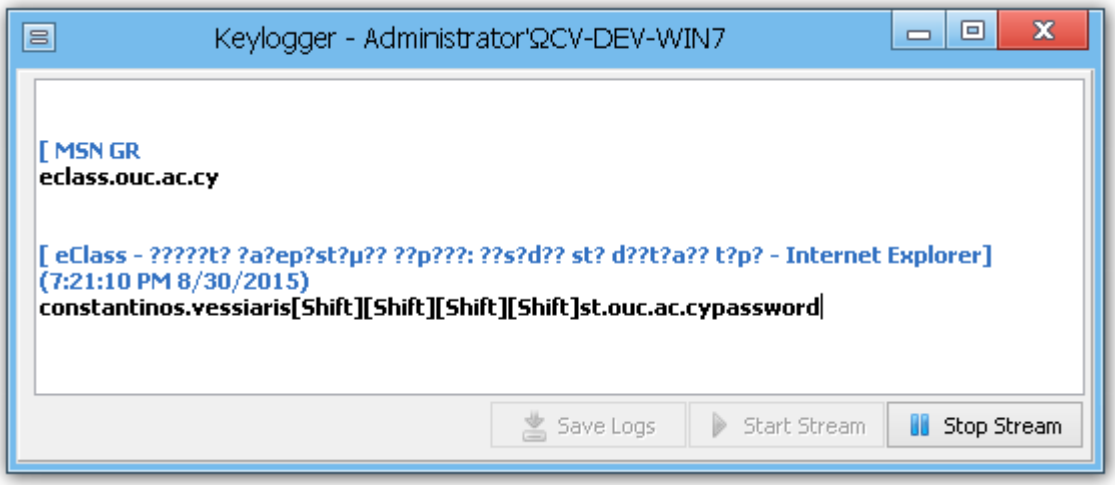
Εικόνα 4.3: Λογισμικό Pandora

Ο χρήστης αφού επιλέξει το μολυσμένο υπολογιστή που επιθυμεί του δίνεται η δυνατότητα να επιλέξει τι θέλει να πράξει(Εικόνα 4.4). Ο χρήστης έχει τη δυνατότητα προβολής της επιφάνειας εργασίας του υπολογιστή παρέχοντας του τη δυνατότητα να βλέπει τις κινήσεις του χρήστη που χειρίζεται μια συγκεκριμένη χρονική στιγμή το μηχάνημα θύμα. Ο χρήστης του Λογισμικού Pandora έχει επίσης τη δυνατότητα να πραγματοποιήσει οποιαδήποτε τροποποίηση, αλλαγή, προσθήκη, διαγραφή στο σύστημα αρχείων του συγκεκριμένου υπολογιστή. Αυτός είναι και ο βασικότερος λόγος που ένας κακόβουλος χρήστης πιθανότατα να διαλέξει λογισμικό της κατηγορία RAT / Trojan για τη πρόκληση βλάβης στο εταιρικό δίκτυο.



Εικόνα 4.4: Λογισμικό Pandora - Κέντρο ελέγχου υπολογιστή

Μια άλλη δυνατότητα που παρέχεται στο χρήστη του λογισμικού Pandora και αξίζει σημείωσης είναι η επίβλεψη του πληκτρολογίου. Χρησιμοποιώντας τη λειτουργία αυτή ο χρήστης μπορεί να καταγράψει τι ακριβώς πληκτρολογεί ο χειριστής του υπολογιστή. Από τη καταγραφή πληκτρολογίου που παρουσιάζεται στην Εικόνα 4.5 μπορεί εύκολα κάποιος να συμπεραίνει ότι ο χειριστής του υπολογιστή επισκέφθηκε την ιστοσελίδα http://eclass.ouc.ac.cy και εκεί πληκτρολόγησε constantinos.vessiaris@st.ouc.ac.cy σαν αναγνωριστικό χρήστη(username) και password σαν κωδικό πρόσβασης.



Εικόνα 4.5: Pandora - Καταγραφή πλήκτρων

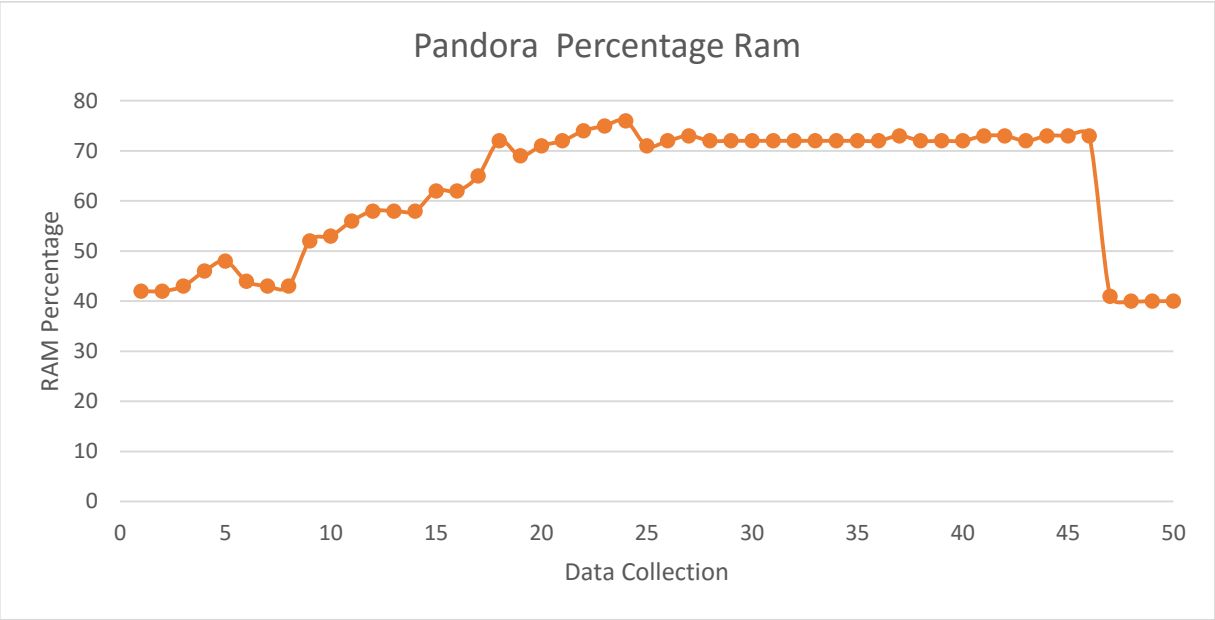
Εκτός από τις λειτουργίες που ήδη περιεγράφηκαν το λογισμικό Pandora προσφέρει και πολλές άλλες ενδιαφέρουσες λειτουργίες. Ο χρήστης του μπορεί να αλληλοεπιδράσει σε όλες τις υπηρεσίες(services) και διαδικασίες processes του υπολογιστή. Με τον τρόπο αυτό μπορεί να απενεργοποιήσει οποιοδήποτε πρόγραμμα, για παράδειγμα το πρόγραμμα ανίχνευσής ιών. Ενώ χρησιμοποιώντας την επιλογή «Scan Lan Computers» έχει τη δυνατότητα εύρεσης όλων των υπολογιστών δικτύου.

Για τη πειραματική διαδικασία με ιό της κατηγορίας RAT / Trojan θέσαμε σε λειτουργία ένα υπολογιστή θύμα για το διάστημα των τεσσάρων ωρών. Ο υπολογιστής θύμα ανήκε στο εικονικό δίκτυο που δημιουργήσαμε ενώ σε αυτόν γίνονταν οι διαδικασίες που εκτελεί ο οποιοσδήποτε χρήστης υπολογιστή κατά τη διάρκεια της ημέρας όπως για παράδειγμα πλοήγηση στο διαδίκτυο, αποστολή ηλεκτρονικού ταχυδρομείου. Παράλληλα ενεργοποιήσαμε ένα υπολογιστή επίθεσης ο οποίος είχε εγκαταστημένο το λογισμικό Pandora για τη πραγματοποίηση απομακρυσμένης πρόσβασης προς τον

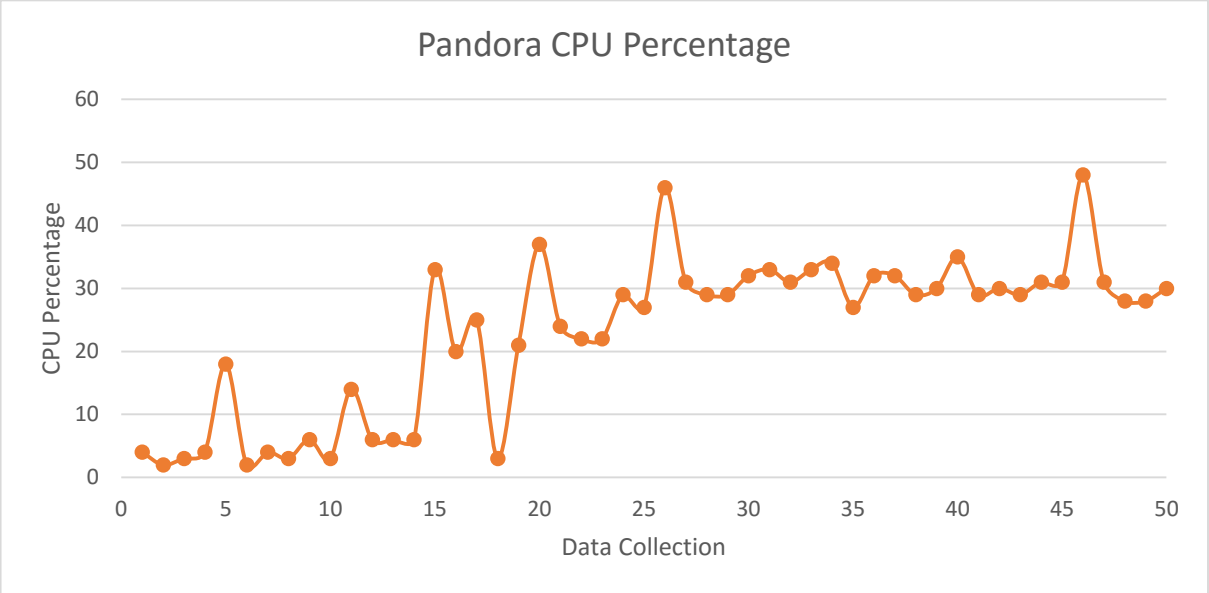
υπολογιστή θύμα. Ο υπολογιστής επίθεσης πραγματοποιούσε διάφορες λειτουργίες του λογισμικού Pandora καθ' όλη τη διάρκεια των τεσσάρων ωρών όχι όμως συνεχώς.

Στο Γράφημα 4.3 παρουσιάζεται η γραφική παράσταση χρήσης της μνήμης τυχαίας προσπέλασης(RAM) καθώς στο Γράφημα 4.4 παρουσιάζεται η γραφική παράσταση χρήσης της κεντρικής μονάδας επεξεργασίας. Στον άξονα των Χ και στις δύο γραφικές παραστάσεις παρουσιάζεται η στιγμή που το σύστημα επίβλεψης εκτελείτο στον υπολογιστή. Μεταξύ της χρονικής στιγμής 20 και 40 παρατηρούμε μια σημαντική αύξηση στο ποσοστό χρήσης τόσο της μνήμης τυχαίας προσπέλασης όσο και της κεντρικής μονάδας επεξεργασίας. Το γεγονός αυτό οφείλεται στην απεικόνιση της επιφάνειας εργασίας(screen capture) του υπολογιστή θύμα στον υπολογιστή επίθεσης.

Στον υπολογιστή θύμα το λογισμικό Pandora είχε δημιουργήσει ένα νήμα εργασίας(server.exe) και δύο συνδέσεις TCP με αποδέκτη τον υπολογιστή επίθεσης χρησιμοποιώντας τη πόρτα 6622. Οι συνδέσεις που καταγράφηκαν στις 50 φορές(1 φορά ανά 5 λεπτά) που εκτελέστηκε το σύστημα επίβλεψης είναι 78. Εφόσον οι συνδέσεις που θα έπρεπε να καταγράφονται σε κάθε εκτέλεση του συστήματος επίβλεψης είναι δύο τότε ο συνολικός αριθμός των συνδέσεων που καταγράφηκαν θα έπρεπε να είναι 100 αντί 78, ο λόγος που αυτό συνέβη είναι λόγω της διακοπής της λειτουργίας του λογισμικού Pandora σε κάποιες από τις εκτελέσεις του συστήματος επίβλεψης.



Γράφημα 4.3: Pandora RAM



Γράφημα 4.4: Pandora CPU

Ο συνολικός αριθμός των εγγραφών που συνέλεξε το σύστημα επίβλεψης παρουσιάζονται στον Πίνακας 4.5 ανά μονάδα συστήματος. Από τις 1798 συνδέσεις δικτύου που συνέλεξε η υπό-μονάδα TCP της μονάδας επίβλεψης δικτύου οι 78 αφορούσαν τις συνδέσεις TCP του λογισμικού Pandora. Οι 1665 εγγραφές συμπεριελήφθησαν στο πακέτο εκπαίδευσης.

Κατά την εκτέλεση του αλγορίθμου SOM χρησιμοποιήσαμε το 70% των δεδομένων μας ως πακέτο εκπαίδευσης και το 30% σαν πακέτο δοκιμής.

Μονάδα	**Αριθμός δεδομένων**
Δικτύου - TCP	1798
Δικτύου - UDP	614
Υπηρεσιών	1946
Απόδοσης	50

Πίνακας 4.5: Πλήθος δεδομένων κατά τη πειραματική διαδικασία με Pandora

Στον Πίνακας 4.6 παρουσιάζεται η κατηγοριοποίηση των νευρώνων του αλγορίθμου SOM. Η στήλη N απεικονίζει τη θέση του νευρώνα ενώ η στήλη K τη κατηγορία του νευρώνα. Χρησιμοποιήθηκε ένα πλέγμα νευρώνων 10 * 10, επομένως υπάρχει ένα

N	K	N	K	N	K	N	K
[0.0]	NORMAL	[2.5]	NORMAL	[5.0]	NONE	[7.5]	NONE
[0.1]	NORMAL	[2.6]	NORMAL	[5.1]	NONE	[7.6]	NONE
[0.2]	NORMAL	[2.7]	ATTACK	[5.2]	NONE	[7.7]	NONE
[0.3]	NONE	[2.8]	NONE	[5.3]	NONE	[7.8]	NONE
[0.4]	NONE	[2.9]	NONE	[5.4]	NORMAL	[7.9]	NORMAL
[0.5]	NONE	[3.0]	NONE	[5.5]	NONE	[8.0]	NONE
[0.6]	NONE	[3.1]	NONE	[5.6]	NONE	[8.1]	NONE
[0.7]	NONE	[3.2]	NONE	[5.7]	NONE	[8.2]	NONE
[0.8]	NONE	[3.3]	NONE	[5.8]	NONE	[8.3]	NONE
[0.9]	NORMAL	[3.4]	NORMAL	[5.9]	NONE	[8.4]	NONE
[1.0]	NORMAL	[3.5]	NORMAL	[6.0]	NONE	[8.5]	NONE
[1.1]	NONE	[3.6]	NORMAL	[6.1]	NONE	[8.6]	NONE

[1.2]	NONE	[3.7]	NONE	[6.2]	NONE	[8.7]	NONE
[1.3]	NONE	[3.8]	NORMAL	[6.3]	NONE	[8.8]	NORMAL
[1.4]	NONE	[3.9]	NONE	[6.4]	NONE	[8.9]	NORMAL
[1.5]	NONE	[4.0]	NONE	[6.5]	NONE	[9.0]	NONE
[1.6]	NORMAL	[4.1]	NONE	[6.6]	NONE	[9.1]	NONE
[1.7]	NONE	[4.2]	NONE	[6.7]	NONE	[9.2]	NONE
[1.8]	NONE	[4.3]	NORMAL	[6.8]	NONE	[9.3]	NORMAL
[1.9]	NONE	[4.4]	NONE	[6.9]	NONE	[9.4]	NORMAL
[2.0]	NORMAL	[4.5]	NONE	[7.0]	NORMAL	[9.5]	NONE
[2.1]	NONE	[4.6]	NORMAL	[7.1]	NONE	[9.6]	NONE
[2.2]	NONE	[4.7]	NORMAL	[7.2]	NONE	[9.7]	NONE
[2.3]	NONE	[4.8]	NORMAL	[7.3]	NONE	[9.8]	NORMAL
[2.4]	NONE	[4.9]	NONE	[7.4]	NONE	[9.9]	NONE

Πίνακας 4.6: Κατηγοριοποίηση Νευρώνων - Pandora

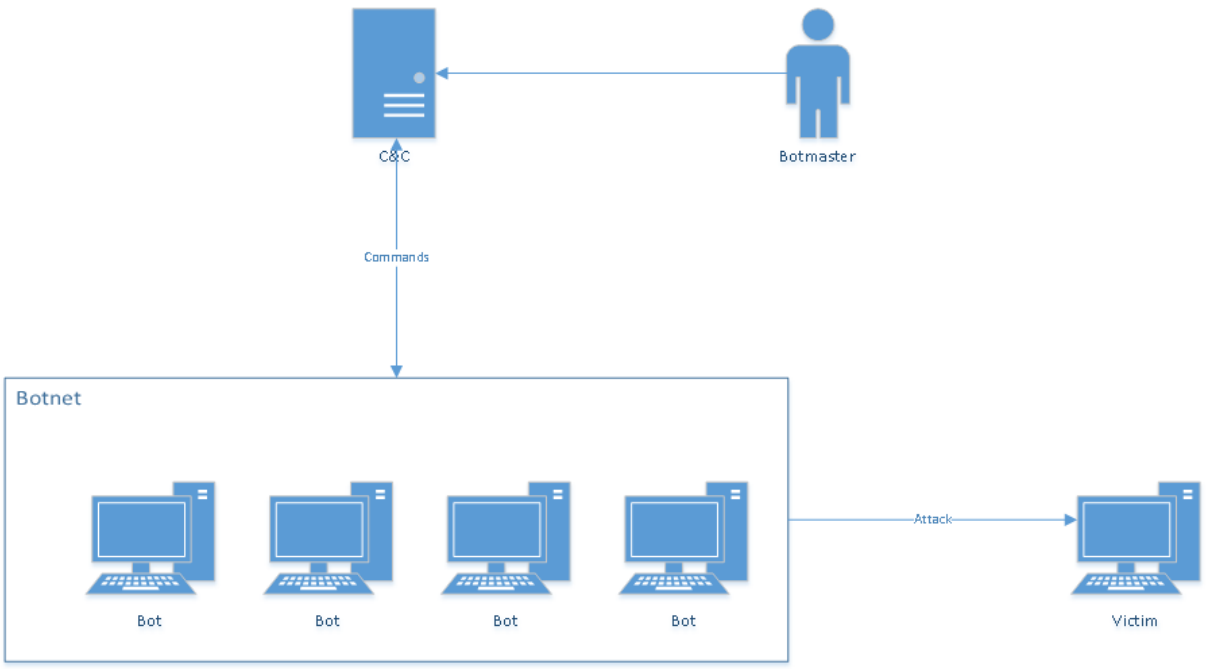
πλήθος 100 νευρώνων. Από τους 100 νευρώνες οι 74 έχουν κατηγοριοποιηθεί στη κατηγορία «NONE», 25 στη κατηγορία «NORMAL» και ο τελευταίος νευρώνας στη κατηγορία «ATTACK».

Από τις 78 συνδέσεις του λογισμικού Pandora που κατέγραψε το σύστημα επίβλεψης οι 49 συμπεριλήφθηκαν στο πακέτο εκπαίδευσης και οι υπόλοιπες 29 στο πακέτο δοκιμής. Κατά τη δοκιμαστική εκτέλεση του αλγορίθμου είχαμε 32 εγγραφές οι οποίες κατηγοριοποιήθηκαν ως «ATTACK», από αυτές οι 27 ήταν από τις συνδέσεις που είχε δημιουργήσει το λογισμικό Pandora. Επομένως κατά τη δοκιμαστική εκτέλεση του αλγορίθμου SOM έχουμε καταγράψει 4 εγγραφές σαν λανθασμένα θετικές και 1 εγγραφή σαν λανθασμένα αρνητική.

4.3. Επίθεση χρησιμοποιώντας ιό κατηγορίας BOTNET

Με τη χρησιμοποίηση του όρου «Botnet» εννοούμε ένα σύνολο από ρομπότ(robots ή bots) τα οποία στην ουσία είναι μηχανήματα που έχουν μολυνθεί(zombie computers) έτσι ώστε να εξυπηρετήσουν παράνομη δραστηριότητα [34] . Η ιδιαιτερότητα της κατηγορίας αυτής έγκειται στην ύπαρξη του κέντρου εντολών και ελέγχου(Command and Control – C&C). Το κέντρο εντολών και ελέγχου αποτελεί το κανάλι επικοινωνίας μεταξύ των μολυσμένων μηχανημάτων και του ο εισβολέα, ο οποίος αποκαλείται και botmaster. Μια γενική τοπολογία που χρησιμοποιείται στα δίκτυα BOTNET παρουσιάζεται στην Εικόνα 4.6.

Το κανάλι εντολών και ελέγχου έχει τη δυνατότητα λειτουργίας κάτω από μια πληθώρα τοπολογιών δικτύου, κεντρικοποιημένη ή μη [35]. Το κανάλι εντολών και ελέγχου πιθανόν να χρησιμοποιεί τα πρωτόκολλά δικτύου IRC(Internet Relay Chat), HTTP ή P2P. Ένα Botnet πιθανόν να χρησιμοποιηθεί για τη δημιουργία διαφόρων ειδών επιθέσεων όπως διανεμημένη άρνηση υπηρεσίας(DDoS), κλοπή κωδικών, κλοπή και διανομή εγγράφων, αποστολή spam emails. [36].



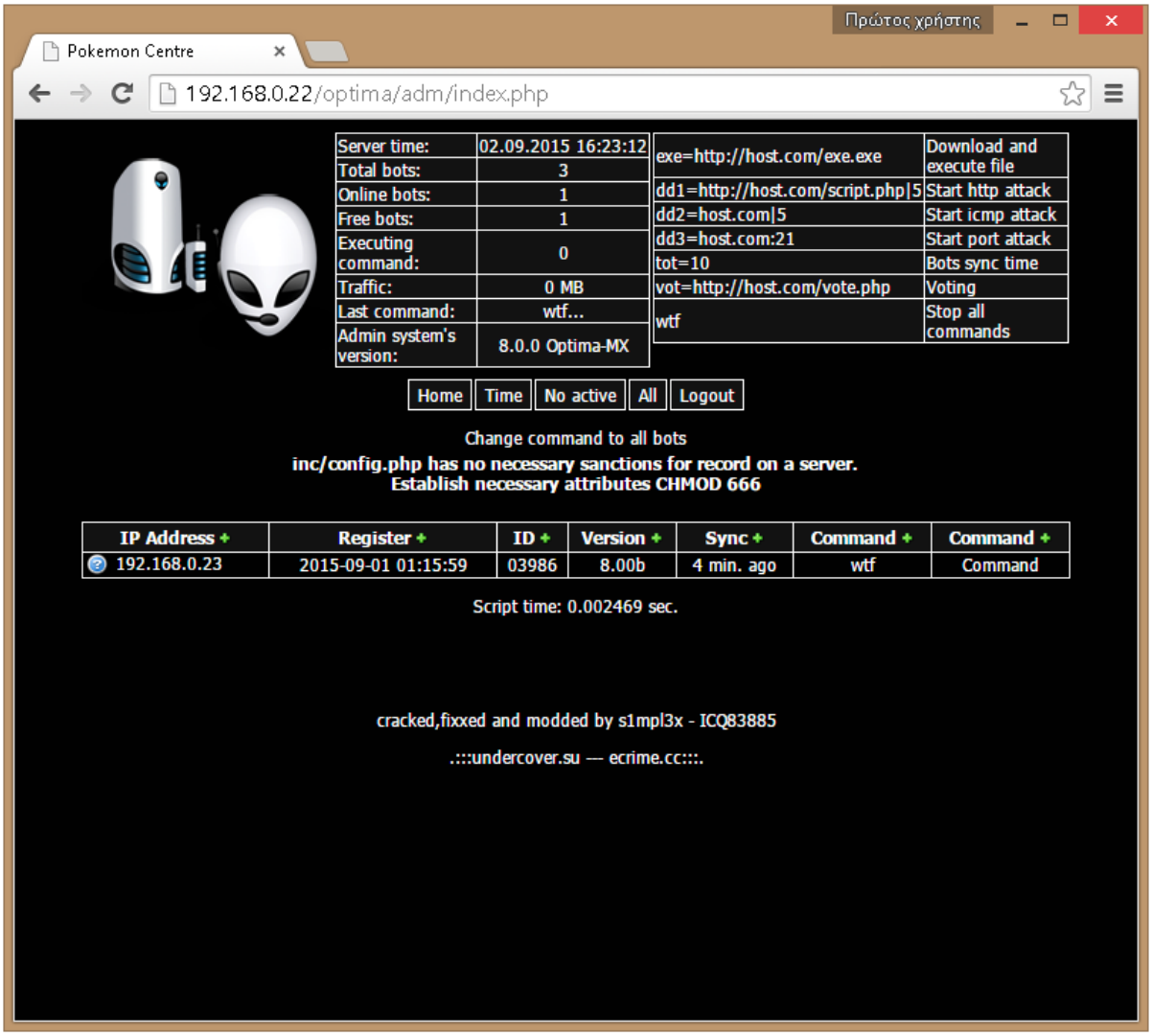
Εικόνα 4.6: Τοπολογία Δικτύου Botnet

Οι λόγοι που μας οδήγησαν στη συμπερίληψη της κατηγορίας αυτής στις πειραματικές μας δοκιμές είναι αρκετοί. Πρωταρχικός λόγος επιλογής αποτελεί το μεγάλο μέγεθος βλάβης που μπορεί να επιφέρει μια επίθεση χρησιμοποιώντας ένα δίκτυο Botnet σε ένα επιχειρησιακό δίκτυο. Το πιο αδύνατο σημείο και το πιο εύκολα ανιχνεύσιμο ενός δικτύου μολυσμένων μηχανημάτων είναι το κανάλι εντολών και ελέγχου [37]. Η δημιουργία ενός δικτύου μολυσμένων μηχανημάτων εντός εταιρικού δικτύου μπορεί να επιφέρει τεραστίου μεγέθους βλάβες στο δίκτυο και κατ' επέκταση στην ίδια την επιχείρηση.

Για να εξακριβώσουμε το ποσοστό επιτυχίας ανίχνευσης ενός δικτύου μολυσμένων μηχανημάτων που μπορεί να επιτύχει το σύστημα ανίχνευσης απειλών που αναπτύχθηκε αποφασίσαμε την εγκατάσταση του λογισμικού Optima. Το λογισμικό Optima έχει τη δυνατότητα εκδήλωσης επιθέσεων διανεμημένης άρνησης υπηρεσίας(Distributed Denial of Service - DDoS).

Ο τύπος επίθεσης άρνησης υπηρεσίας, όπως περιεγράφηκε στην ενότητα 1.1, κάνει χρήση ενός μηχανήματος και μίας σύνδεσης δικτύου για να εκδηλώσει επιθέσεις πλημύρας στο μηχάνημα στόχο. Ο τύπος επίθεσης διανεμημένης άρνησης υπηρεσίας κάνει χρήση πολλών μηχανημάτων και πολλών συνδέσεων δικτύου για να εκδηλώσει επιθέσεις πλημμύρας σε ένα μηχάνημα στόχο [38].

Για την εγκατάσταση του λογισμικού Optima χρειάστηκε να χρησιμοποιήσουμε δύο μηχανήματα. Το πρώτο μηχάνημα ήταν ο διακομιστής διαδικτύου που αποτέλεσε το κέντρο εντολών και ελέγχου. Ο διακομιστής διαδικτύου φιλοξενούσε την ιστοσελίδα(Εικόνα 4.7) η οποία αποτελεί το κέντρο ελέγχου των μολυσμένων μηχανημάτων καθώς και το σημείο αποστολής εντολών σε αυτά. Ενώ το δεύτερο μηχάνημα αποτελούσε το μολυσμένο μηχάνημα από το οποίο θα εκτελούνταν οι επιθέσεις. Στο μολυσμένο μηχάνημα εκτελέστηκε ένα εκτελέσιμο αρχείο το οποίο περιείχε την διεύθυνση δικτύου(IP) του διακομιστή διαδικτύου έτσι ώστε να πραγματοποιηθεί η σύνδεση μεταξύ των δύο.



Εικόνα 4.7: Κέντρο ελέγχου και εντολών – Optima

Μέσω του κέντρου ελέγχου και εντολών του λογισμικού Optima ο εισβολέας έχει τη δυνατότητα εκτέλεσης τρία ειδών επίθεση. Επίθεση μέσω του πρωτοκόλλου HTTP(Hypertext Transfer Protocol), μέσω του πρωτοκόλλου ICMP(Internet Control Message Protocol) και επίθεσης στοχευμένης πόρτας. Ο διαχειριστής του δικτύου μολυσμένων μηχανημάτων εισάγει την εντολή που επιθυμεί να εκτελέσει κάθε μηχάνημα. Το μηχάνημα επισκέπτεται το διακομιστή για να παραλάβει την εντολή του εισβολέα ανά προκαθορισμένο χρονικό διάστημα.

Στη πειραματική διαδικασία που πραγματοποιήθηκε με σκοπό την αξιολόγηση της ικανότητας του συστήματος ανίχνευσης εσωτερικών απειλών να ανιχνεύει επιθέσεις μέσω δικτύου μολυσμένων μηχανημάτων σε εταιρικό δίκτυο συμμετείχαν τρία μηχανήματα. Ο κεντρικός διακομιστής του συστήματος ανίχνευσης εσωτερικών

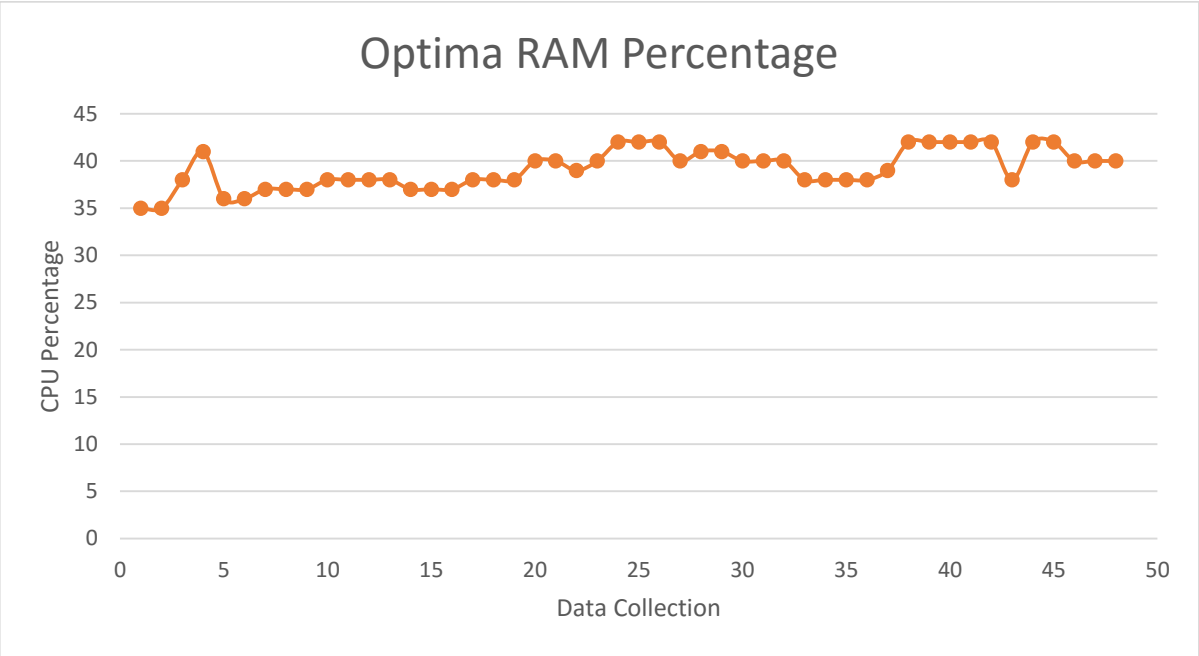
απειλών, ο διακομιστής διαδικτύου που φιλοξενούσε το κέντρο ελέγχου και εντολών του λογισμικού Optima καθώς και ο υπολογιστή χρήστη. Λόγω έλλειψης πόρων ο διακομιστής διαδικτύου φιλοξενούσε τόσο τη βάση δεδομένων του λογισμικού Optima όσο και τη βάση δεδομένων του συστήματος ανίχνευσης εσωτερικών απειλών. Στον υπολογιστή χρήστη βρισκόταν εγκατεστημένο το λογισμικό χρήστη του συστήματος που αναπτύχθηκε καθώς και το λογισμικό Optima, ακριβώς όπως σε ένα πραγματικό περιβάλλον.

Μονάδα	**Αριθμός δεδομένων**
Δικτύου - TCP	1640
Δικτύου - UDP	700
Υπηρεσιών	2300
Απόδοσης	50

Πίνακας 4.7: Πλήθος δεδομένων κατά τη πειραματική διαδικασία με Optima

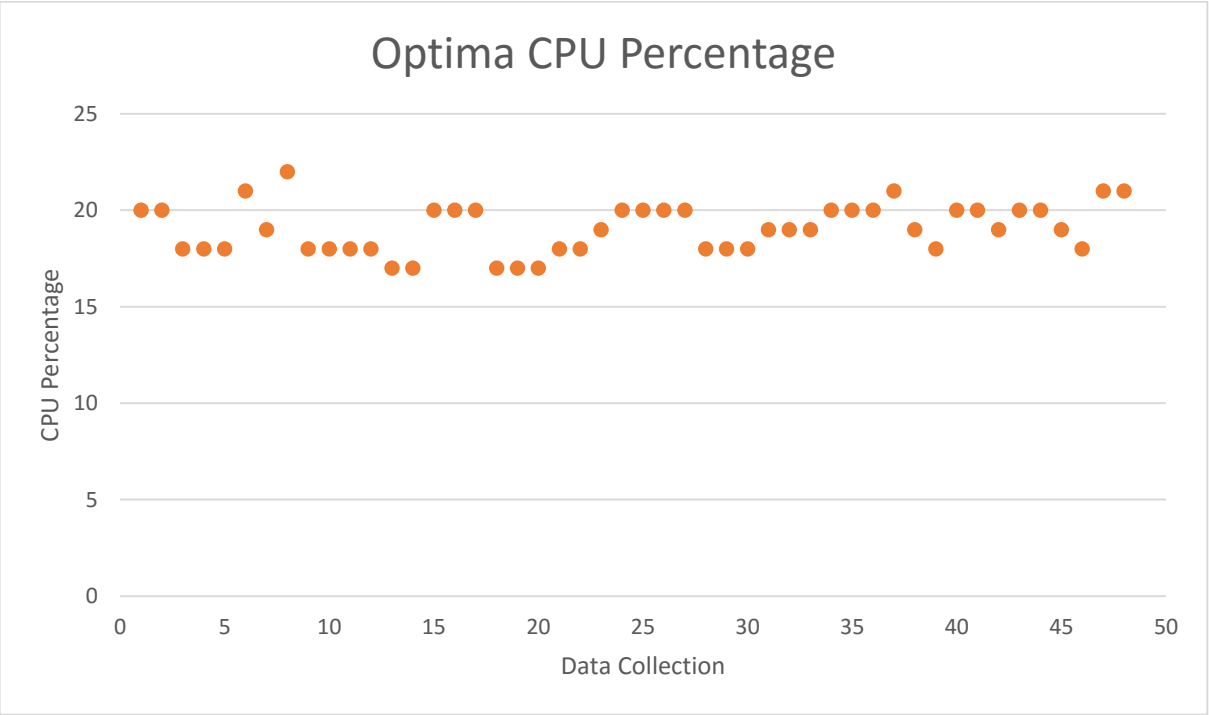
Κατά το διάστημα των τεσσάρων ωρών της πειραματικής διαδικασίας το σύστημα επίβλεψης συνέλλεξε 50 φορές(1 φορά ανά 5 λεπτά) δεδομένα. Ο συνολικός αριθμός δεδομένων ανά μονάδα που συλλέχθηκε από το σύστημα επίβλεψης παρουσιάζεται στον Πίνακας 4.7. Θα πρέπει επίσης να σημειώσουμε ότι καταχωρήσαμε εντολές στο κέντρο ελέγχου του λογισμικού Optima 18 φορές δεν περιέχουν όλες οι εκτελέσεις του συστήματος επίβλεψης δεδομένα που σχετίζονται με το λογισμικό Optima. Ο λόγος που επιλέχθηκε αυτή η τακτική είναι το ότι πιθανότατα να ακολουθηθεί και από τον εισβολέα σε ένα πραγματικό περιβάλλον έτσι ώστε να μην τραβήξει τη προσοχή του διαχειριστή δικτύου.

Από τα γραφήματα Γράφημα 4.5 και Γράφημα 4.6 παρατηρούμε ότι η ποσοστιαία απόδοση της μνήμης τυχαίας προσπέλασης και της κεντρικής μονάδας επεξεργασίας παραμένει σταθερές σε όλες τις εκτελέσεις της μονάδας απόδοσης του συστήματος επίβλεψης που ήταν εγκατεστημένο στο μολυσμένο υπολογιστή χρήστη.



Γράφημα 4.5: Optima RAM

Στον Πίνακας 4.8 παρουσιάζονται οι νευρώνες όπως αυτοί κατηγοριοποιήθηκαν μετά την εκτέλεση του αλγορίθμου SOM. Η στήλη N απεικονίζει τη θέση του νευρώνα ενώ η στήλη K τη κατηγορία του νευρώνα. Χρησιμοποιήθηκε ένα πλέγμα νευρώνων 10 * 10, επομένως υπάρχει ένα πλήθος 100 νευρώνων. Ο αλγόριθμός κατηγοριοποίησε 80 νευρώνες στη κατηγορία «NONE», 19 στη κατηγορία «NORMAL» και μόλις ένα νευρώνα στη κατηγορία «ATTACK».



Γράφημα 4.6: Optima CPU

Από τις 1640 συνδέσεις πρωτοκόλλου που περιείχαν τα αρχικά μας δεδομένα οι 198 αφορούσαν συνδέσεις του λογισμικού Optima. Οι συνδέσεις του λογισμικού Optima περιείχαν τόσο τις συνδέσεις πρωτοκόλλου TCP που χρησιμοποιούντο για την επικοινωνία μεταξύ κέντρου ελέγχου και μολυσμένου υπολογιστή άλλα και τις συνδέσεις από τον μολυσμένο υπολογιστή προς τον υπολογιστή θύμα. Οι συνδέσεις μεταξύ κέντρου ελέγχου και μολυσμένου υπολογιστή πραγματοποιούνταν από τη πόρτα 21 ενώ οι συνδέσεις μεταξύ μολυσμένου υπολογιστή και υπολογιστή θύμα από τη πόρτα 80.

Όπως και στις προηγούμενες πειραματικές διαδικασίες έτσι και σε αυτή τη περίπτωση χρησιμοποιήσαμε το 70% των δεδομένων μας ως πακέτο εκπαίδευσης και το υπόλοιπο 30% ως πακέτο δοκιμής. Από τις 198 συνδέσεις του λογισμικού Optima οι 153 συμπεριλήφθηκαν στο πακέτο εκπαίδευσης ενώ οι υπόλοιπες 45 στο πακέτο δοκιμής. Κατά την εκτέλεση δοκιμής ο αλγόριθμός SOM πέτυχε να αναγνωρίσει σωστά και τις 45 συνδέσεις καθώς και μια περισσότερη, επομένως είχαμε μόνο μια λανθασμένα θετική ενώ καμία λανθασμένα αρνητική εγγραφή.

N	K	N	K	N	K	N	K
[0.0]	NORMAL	[2.5]	NORMAL	[5.0]	NONE	[7.5]	NONE
[0.1]	NORMAL	[2.6]	NORMAL	[5.1]	NORMAL	[7.6]	NONE
[0.2]	NORMAL	[2.7]	NONE	[5.2]	NORMAL	[7.7]	NONE
[0.3]	NORMAL	[2.8]	NONE	[5.3]	NORMAL	[7.8]	NONE
[0.4]	NORMAL	[2.9]	NONE	[5.4]	NORMAL	[7.9]	NONE
[0.5]	NONE	[3.0]	ATTACK	[5.5]	NORMAL	[8.0]	NONE
[0.6]	NONE	[3.1]	NONE	[5.6]	NONE	[8.1]	NONE
[0.7]	NONE	[3.2]	NONE	[5.7]	NONE	[8.2]	NONE
[0.8]	NONE	[3.3]	NORMAL	[5.8]	NONE	[8.3]	NONE
[0.9]	NONE	[3.4]	NORMAL	[5.9]	NONE	[8.4]	NONE
[1.0]	NONE	[3.5]	NORMAL	[6.0]	NONE	[8.5]	NONE
[1.1]	NORMAL	[3.6]	NONE	[6.1]	NONE	[8.6]	NONE
[1.2]	NORMAL	[3.7]	NONE	[6.2]	NONE	[8.7]	NONE
[1.3]	NONE	[3.8]	NONE	[6.3]	NONE	[8.8]	NONE
[1.4]	NONE	[3.9]	NONE	[6.4]	NONE	[8.9]	NONE
[1.5]	NONE	[4.0]	NONE	[6.5]	NONE	[9.0]	NONE
[1.6]	NONE	[4.1]	NONE	[6.6]	NONE	[9.1]	NONE
[1.7]	NONE	[4.2]	NONE	[6.7]	NONE	[9.2]	NORMAL
[1.8]	NONE	[4.3]	NONE	[6.8]	NONE	[9.3]	NORMAL
[1.9]	NONE	[4.4]	NONE	[6.9]	NONE	[9.4]	NORMAL
[2.0]	NONE	[4.5]	NONE	[7.0]	NONE	[9.5]	NORMAL
[2.1]	NONE	[4.6]	NONE	[7.1]	NONE	[9.6]	NONE
[2.2]	NONE	[4.7]	NONE	[7.2]	NONE	[9.7]	NONE
[2.3]	NONE	[4.8]	NONE	[7.3]	NONE	[9.8]	NONE

[2.4] NONE	[4.9] NONE	[7.4] NONE	[9.9] NONE

Πίνακας 4.8: Κατηγοριοποίηση Νευρώνων – Optima

Κεφάλαιο 5
Προβλήματα και Συμπεράσματα

Για την ολοκλήρωση της παρούσας μεταπτυχιακής διατριβής απαιτείτο ένα ευρύ φάσμα γνώσης γύρω από την ανάπτυξη λογισμικού, την αρχιτεκτονική του λειτουργικού συστήματος Windows, την εύρεση απειλών σε ένα δίκτυο και τις ιδιαιτερότητες που φέρει η ανίχνευση εσωτερικών απειλών. Ο πρώτος άξονας που θα μελετηθεί στο παρόν κεφάλαιο αφορά την ανάλυση των προβλημάτων που κληθήκαμε να αντιμετωπίσουμε κατά την ανάπτυξη του συστήματος και κατά την πραγματοποίησης της πειραματικής διαδικασίας καθώς και οι αποφάσεις που πήραμε για την επίλυση τους.

Στο προηγούμενο κεφάλαιο περιεγράφηκαν αναλυτικά οι πειραματικές δοκιμές που εκτελέσαμε για τον έλεγχο της εγκυρότητας των αποτελεσμάτων του συστήματος ανίχνευσης απειλών που αναπτύχθηκε. Στο παρόν κεφάλαιο θα πραγματοποιηθεί μια σύμμειξη των αποτελεσμάτων της πειραματικής διαδικασίας καθώς και γενικότερες σκέψεις όσο αφορά τη βελτίωση του συστήματος.

5.1. Προβλήματα που αντιμετωπίστηκαν κατά την ανάπτυξη συστήματος

Το σύστημα που αναπτύχθηκε για τη παρούσα μεταπτυχιακή διατριβή θα μπορούσε να χαρακτηριστεί και ως ένα πραγματικό εταιρικό προϊόν. Τόσο η αρχιτεκτονική πάνω στην οποία στήθηκε όσο και ο τρόπος ανάπτυξης των λογισμικών από τα οποία αποτελείται δίνουν τη δυνατότητα επεκτασιμότητας του. Η επεκτασιμότητα μπορεί να προκύψει κυρίως προσθέτοντας και άλλες μονάδες στο σύστημα επίβλεψης με την τεχνική των modules.

Λόγω του μεγάλου μεγέθους συστήματος που έπρεπε να αναπτυχθούν για την διεκπεραίωση της μεταπτυχιακής διατριβής αντιμετωπίσαμε, όπως ήταν αναμενόμενο, πληθώρα προβλημάτων ενώ κληθήκαμε να πάρουμε αποφάσεις που θα αποσκοπούσαν σε ένα καλύτερο ποιοτικά σύστημα. Οι αποφάσεις που πάρθηκαν είχαν σαν κύριο γνώμονα το σκοπό του συστήματος προς ανάπτυξη.

Από τις πρώτες αποφάσεις που κληθήκαμε να πάρουμε αφορούσαν τις τεχνολογίες που θα χρησιμοποιούσαμε για την ανάπτυξη του συστήματος. Δεδομένου του ότι το σύστημα θα απευθυνόταν σε μηχανήματα που φέρουν το λειτουργικό Windows πάρθηκε απόφαση για ανάπτυξή του συστήματος στο πλαίσιο(framework) .NET και πιο συγκεκριμένα στο πιο πρόσφατο πλαίσιο της οικογένειας .NET, το πλαίσιο .NET 4.5.

Η προσθήκη ενός κεντρικού σημείου ελέγχου του συστήματος μέσω της ανάπτυξης του λογισμικού διακομιστή πραγματοποιήθηκε με σκοπό τη συμπερίληψη της ελάχιστης επιχειρησιακής λογικής(business logic) στο λογισμικό που θα προοριζόταν για εγκατάσταση σε υπολογιστή χρήστη. Με τον τρόπο αυτό ελαχιστοποιούμε τη πιθανότητα έκθεσης(compromise) της επιχειρησιακής λογικής του συστήματος σε περίπτωση προσπάθειας αποσυμπίεσης(decompile) του λογισμικού χρήστη. Θα πρέπει

επίσης να σημειώσουμε ότι για περαιτέρω ασφάλεια το λογισμικού χρήστη, ανά τακτά τυχαία χρονικά διαστήματα, η εφαρμογή ελέγχει το MD5 Hash της που έχει τη συγκεκριμένη χρονική στιγμή και το συγκρίνει με αυτό που θα έπρεπε να έχει. Αν υπάρχουν διαφορές μεταξύ των δύο τότε το σύστημα ξανακατεβάζει τα αρχεία από ένα έμπιστο διακομιστή στον υπολογιστή χρήστη για να εξασφαλίσει την μη παραποίηση του.

Για τη φιλοξενία της βάσης δεδομένων του συστήματος προτιμήθηκε το περιβάλλον σχεσιακών βάσεων δεδομένων(RDBMS) Mysql αντί του περιβάλλοντος MSSQL. Ο λόγος της απόφασης αυτής οφείλεται στο ότι ο Mysql Server διανέμεται δωρεάν και επομένως το κόστος εφαρμογής του συστήματος σε εταιρικό περιβάλλον θα είναι αρκετά μικρότερο από ότι αν επιλεγόταν το περιβάλλον MSSQL. Η σχεσιακή βάση δεδομένων φιλοξενείται από ένα ξεχωριστό διακομιστή και όχι τον ίδιο με το διακομιστή εφαρμογής για αποφυγή υπερβολικής φόρτωσης(overload).

Από τη στιγμή που θέλαμε μια συνεχή επικοινωνία μεταξύ του διακομιστή εφαρμογής και των υπολογιστών χρηστών πήραμε την απόφαση για δημιουργία υπηρεσίας Windows τόσο για το διακομιστή όσο και για τον υπολογιστή χρήστη. Ενώ η ανάγκη για αποστολή εντολών από το διακομιστή στον υπολογιστή χρήστη καθώς και για την αποστολή δεδομένων από τον υπολογιστή χρήστη προς το διακομιστή μας οδήγησε στην απόφαση ανάπτυξης δύο WCF(Windows Communication Foundation) που θα φιλοξενούνται από τις αντίστοιχες υπηρεσίες Windows. Οι υπηρεσίες Windows επικοινωνούν μεταξύ τους χρησιμοποιώντας το πρωτόκολλο TCP.

Κατά την ανάπτυξη του συστήματος παρατηρήσαμε μια υπερφόρτωση του λογισμικού διακομιστή όταν κλεινόταν να ανταποκριθεί σε πολλά ταυτόχρονα μηνύματα. Ο λόγος που αυτό συνέβαινε είναι λόγω του ότι η υπηρεσία Windows εκτελούσε μόνο ένα νήμα εργασίας. Το νήμα εργασίας μπορούσε να ανοίξει μόνο μια σύνδεση. Η σύνδεση παρέμενε ανοιχτή ως ότου ο διακομιστής τελειώσει την επεξεργασία μηνύματος και αποστείλει απάντηση. Για την επίλυση του προβλήματος αυτός χρησιμοποιήσαμε την τεχνολογία της ασύγχρονης επικοινωνίας [39] που προσφέρει το πλαίσιο .NET. Με τον τρόπο αυτό η υπηρεσία Windows έχει τη δυνατότητα επεξεργασίας πολλαπλών ταυτόχρονων μηνυμάτων.

Για τον έλεγχο των μηνυμάτων που ανταλλάσσονται μεταξύ διακομιστή εφαρμογής και υπολογιστή χρήστη πραγματοποιήσαμε επίβλεψη του δικτύου(sniffing) χρησιμοποιώντας το λογισμικό Wireshark. Μέσω της μεθόδου αυτής παρατηρήσαμε ότι τα μηνύματα ανταλλάσσονταν σε μορφή καθαρού κειμένου κάτι που αποτελεί πρόβλημα ασφάλειας. Για την εξάλειψη του προβλήματος αυτός τα μηνύματα κωδικοποιούνται από τον πομπό και αποκωδικοποιούνται από το δέκτη.

Ένα ακόμη δίλημμά που προέκυψε κατά τη σχεδίαση και ανάπτυξη του συστήματος είναι η απόφαση για το ποιες μονάδες επίβλεψης θα περιέχει το σύστημα μας. Η αρχιτεκτονική του λειτουργικού συστήματος Windows είναι τεράστια. Η ανάπτυξη μονάδων επίβλεψης για όλα τα συστήματα του λειτουργικού θωρήθηκε ανέφικτη στα χρονικά πλαίσια ολοκλήρωσης της μεταπτυχιακής διατριβής. Θα έπρεπε επομένως να βρούμε τα ποια σημαντικά συστήματα του λειτουργικού και πάνω σε αυτά να αναπτύξουμε μονάδες επίβλεψης. Όπως είδη αναφέρθηκε αναπτύχθηκαν τέσσερεις μονάδες επίβλεψης: δικτύου, αρχείων, δικτύου και υπηρεσιών.

Αρχικά υπήρχε σκέψη η μονάδα επίβλεψης αρχείων να επιβλέπει τη κάθε αλλαγή σύστημα αρχείων του υπολογιστή και να αποστέλλει τη παραμικρή αλλαγή στο διακομιστή για επεξεργασία. Μια τέτοια λειτουργία αναπτύχθηκε αλλά από τις πρώτες δοκιμές διαπιστώθηκε η τεράστια αργοπορία εκτέλεσης της μονάδας αυτής καθώς και ο τεράστιος όγκος δεδομένων που απέστελλε. Το σύστημα αρχείων είναι το άμεσα επηρεαζόμενο από κάθε κίνηση του χρήστη και επομένως οι αλλαγές ανά δευτερόλεπτο είναι αρκετές και συνεπώς και ο όγκος δεδομένων που χρειάζεται για να περιγράφουν οι αλλαγές είναι ιδιαίτερα μεγάλος. Για τη φιλοξενία μιας τέτοιας μονάδας επίβλεψης θα ήταν απαραίτητο ο κάθε υπολογιστής να έχει υψηλά χαρακτηριστικά υλικού καθώς θα παρατηρείτο μεγάλη συμφόρηση στο δίκτυο. Για το σκοπό αυτό δόθηκε η επιλογή στο διαχειριστή του συστήματος να καταχωρεί τα αρχεία που θα επιθυμεί να ελέγχονται από το σύστημα επίβλεψης για όλο το δίκτυο είτε για το κάθε υπολογιστή ξεχωριστά. Για τη μείωση του όγκου δεδομένων που στέλνονται μέσω δικτύου επιλέχθηκε η παραγωγή και αποστολή του MD5 Hash του κάθε αρχείου προς επίβλεψη αντί της αποστολής ολόκληρης της δομής του αρχείου.

Ένα άλλο αρκετά δύσκολο αλλά και ενδιαφέρον πρόβλημα που κληθήκαμε να αντιμετωπίσουμε είναι σε ποια μορφή θα δίναμε τα δεδομένα στον αλγόριθμο SOM, πώς θα τα ομαδοποιήσουμε και πως ο αλγόριθμος θα επεξεργαζόταν τα δεδομένα που ήταν

σε μορφή κειμένου αντί σε αριθμητική μορφή. Το κάθε δεδομένο με μορφή κειμένου αντιστοιχείται με ένα μοναδικό αριθμό έτσι ώστε να μπορεί να συμμετάσχει στα δεδομένα που θα δοθούν σαν είσοδο στον αλγόριθμο SOM. Για την ομαδοποίηση των δεδομένων προσθέσαμε τον πίνακα tblCollectData. Με την πρόσθεση του πίνακα αυτού πραγματοποιήσαμε τη συνένωση των δεδομένων, που συλλέγονται και από τις τέσσερεις μονάδες επίβλεψης, μεταξύ τους. Αφού πραγματοποιηθούν οι πιο πάνω ενέργειες τότε τα δεδομένα που προκύπτουν μετά τη συνένωση τους διαχωρίζονται μεταξύ τους μέσω του χαρακτήρα «,» για εισαγωγή τους στον αλγόριθμο SOM. Ο κώδικας του αλγορίθμου πραγματοποιεί το διαχωρισμό των δεδομένων και την επεξεργασία τους για τη κατηγοριοποίηση των νευρώνων.

5.2. Προβλήματα που αντιμετωπίστηκαν κατά τη πειραματική διαδικασία

Αρκετές προκλήσεις κληθήκαμε να αντιμετωπίσουμε και κατά τη διάρκεια της πειραματικής διαδικασίας για την εξέταση των αποτελεσμάτων του συστήματος. Όπως ήδη αναφέρθηκε σε προηγούμενο κεφάλαιο για την εκτέλεση της πειραματικής μας διαδικασίας πήραμε την απόφαση μη χρησιμοποίηση κάποιου προ υπάρχοντος \ αναγνωρισμένου πακέτου δεδομένων και δημιουργίας δικού μας πακέτου πάρθηκε μετά από την μελέτη όλων των υπολοίπων πακέτων. Κανένα από τα υπάρχοντα πακέτα δεν πληρούσε τις δικές μας προϋποθέσεις που δεν ήταν άλλες από τη συμπερίληψη στο πακέτο όλων εκείνων των δεδομένων που συλλέγονταν από το σύστημα επίβλεψης. Παρόλα αυτά τροποποιήσαμε το σύστημα μας για να ελέγξουμε τα αποτελέσματα του αλγορίθμου SOM με τα αναγνωρισμένα πακέτα δεδομένων. Από τη διαδικασία αυτή προέκυψαν κάποιες τροποποιήσεις \ βελτιστοποιήσεις στο κώδικα του αλγορίθμου.

Το επόμενο σημαντικό εμπόδιο που κληθήκαμε να υπερπηδήσουμε ήταν η εύρεση κατάλληλων ιών για τη συλλογή δεδομένων. Αρχικά έπρεπε να βρούμε τις κατηγορίες ιών που θα μπορούσαν να χρησιμοποιηθούν από μια εσωτερική απειλή με σκοπό τη πρόκληση βλάβης στο εταιρικό δίκτυο. Η απάντηση στο ερώτημα αυτό ήρθε μέσα από την ενδελεχή μελέτη κάθε κατηγορίας ιών καθώς και επισκόπηση κώδικα, εάν αυτός υπήρχε διαθέσιμος στο διαδίκτυο. Το δεύτερο στάδιο ήταν η εύρεση ιών από κάθε κατηγορία που επιλέχθηκε. Λόγω της μη νομιμότητας των ιών υπολογιστών η διαδικασία

αυτή αποδείχθηκε αρκετά δύσκολη και ως ένα σημείο ανυπέρβλητη. Αρκετές φορές η εγκατάσταση ιών δεν ήταν εφικτή λόγω ψευδών(fake) αρχείων είτε λόγω αρκετά παλιάς τεχνολογίας ανάπτυξης τους και άλλες φορές τα λογισμικά δεν ήταν σε θέση να εκτελέσουν τις λειτουργίες τους. Το πρόβλημα εύρεσης λογισμικών ιών είναι και ο κύριος λόγος που μας ανάγκασε να εκτελέσουμε μόνο τρείς πειραματικές διαδικασίες ενώ ο αρχικός προγραμματισμός ήταν για πολύ περισσότερες.

Ο αρχικός σχεδιασμός που αφορούσε την πειραματική μας διαδικασία προέβλεπε πως όλες οι πειραματικές δοκιμές θα εκτελούνταν την ίδια χρονική στιγμή. Λόγω του ότι η πειραματική διαδικασία εκτελέστηκε έχοντας μόνο ένα φυσικό μηχάνημα το οποίο είχε τη δυνατότητα να φορτώσει μέχρι τέσσερα εικονικά μηχανήματα αποφασίσαμε την απομόνωση της κάθε πειραματικής διαδικασίας και εκτέλεση της σε διαφορετικές χρονικές στιγμές.

Η πειραματική διαδικασία εκτελέστηκε δύο φορές. Τη πρώτη φορά η διαδικασία που ακολουθήθηκε ήταν διαφορετική από αυτή που περιγραφικέ στο Κεφάλαιο 4. Η μεθοδολογία που ακολουθήθηκε κατά τη πρώτη πειραματική διαδικασία ήταν η εξής: η διαδικασία διάρκεσε και πάλι τέσσερίς ώρες, τις πρώτες δύο ώρες δεν εκτελέσαμε κανένα ιό, τις επόμενες 2 ώρες εκτελέσαμε ταυτόχρονα όλους τους ιούς σε όλους τους υπολογιστές χρηστών. Κατά το διαχωρισμό(labeling) των δεδομένων ονομάσαμε όλα τα δεδομένα των δύο πρώτων ωρών σαν μη επίθεση και τα δεδομένα πέραν των δύο ωρών σαν επίθεση. Τα αποτελέσματα του αλγορίθμου ήταν εντελώς λανθασμένα. Με επιπλέον μελέτη της βιβλιογραφίας και κατανόηση των πειραμάτων που πραγματοποιήσαν άλλοι ερευνητές του χώρου των εσωτερικών απειλών καταλήξαμε στην εκτέλεση της διαδικασίας που αναλύθηκε στο Κεφάλαιο 4.

Κατά την εκτέλεση του αλγορίθμου SOM στα πλαίσια της πειραματικής δοκιμής με δίκτυο μολυσμένων μηχανημάτων(Botnet) διαπιστώσαμε ότι εγγραφές συνδέσεων που αφορούσαν τις συνδέσεις που πραγματοποιούσε το σύστημα που αναπτύχθηκε μεταξύ των μερών του κατηγοριοποιούνταν σαν επιθέσεις. Ο λόγος που αυτό συνέβαινε είναι η ομοιότητα των συνδέσεων του δικού μας συστήματος με τις συνδέσεις που πραγματοποιούσε το δίκτυο μολυσμένων μηχανημάτων. Η ομοιότητα αυτή προέκυψε από την απόφαση μας να χρησιμοποιήσουμε το ίδιο μηχάνημα σαν διακομιστή βάσης δεδομένων του συστήματος μας και σαν κέντρο ελέγχου και εντολών του δικτύου μολυσμένων μηχανημάτων. Δεδομένου ότι γνωρίζουμε τη λειτουργία και την μη μόλυνση

του συστήματος μας αποφασίσαμε να αφαιρούμε τις εγγραφές που αφορούν το δικό μας σύστημα με σκοπό τα πιο ακριβή αποτελέσματα.

5.3. Συμπεράσματα

Από ότι διαφάνηκε από τα αποτελέσματα των πειραματικών δοκιμών που παρουσιάστηκαν στις υπό-ενότητες 1.1, 4.2 και 4.3 το σύστημα ανίχνευσης εσωτερικών απειλών που αναπτύχθηκε έχει πάρα πολύ καλά αποτελέσματα. Καταρχάς θα πρέπει να τονίσουμε το ότι παρόλο που το σύστημα επίβλεψης τρέχει σε τακτά χρονικά διαστήματα δεν φαίνεται να προσδίδει οποιαδήποτε σημαντική μείωση στην απόδοση του μηχανήματος. Οι ασύγχρονες τεχνολογίες ανάπτυξης λογισμικού που παρέχει το πλαίσιο .NET έχουν σαν αποτέλεσμα την μη επιβάρυνση του συστήματος κατά τη διάρκεια εκτέλεσης των μονάδων επίβλεψης.

Από το σύνολο των 105 εγγραφών επιθέσεων που συμπεριλαμβάνονταν στα τρία πακέτα δοκιμών ο αλγόριθμος κατάφερε να αναγνωρίσει επιτυχώς τις 100 εγγραφές. Από τις εναπομείναντες εγγραφές οι 4 αναγνωρίστηκαν λανθασμένα θετικές ενώ μια λανθασμένα αρνητική. Επομένως το σύστημα ανίχνευσης εσωτερικών απειλών που αναπτύχθηκε στα πλαίσια της μεταπτυχιακής διατριβής έχει ποσοστό επιτυχίας 95% ενώ 4% κατατάσσεται ως λανθασμένα θετικό και μόνο 1% λανθασμένα αρνητικό.

Η πειραματική διαδικασία που διεξάχθηκε στο [15] είχε ποσοστό επιτυχίας 31% - 67%, ποσοστό λανθασμένα θετικό 0% - 3% και λανθασμένα αρνητικό 33% - 66%. Στο [40] καταγράφεται η πειραματική διαδικασία ανίχνευσης εσωτερικών απειλών χρησιμοποιώντας αλγορίθμους επιτηρούμενης και μη επιτηρούμενης μάθησης, Ο αλγόριθμος επιτηρούμενη μάθησης είχε ποσοστό επιτυχίας 71%, ποσοστό λανθασμένα θετικό 29% και λανθασμένα αρνητικό 0% ενώ ο αλγόριθμος μη επιτηρούμενης μάθησης είχε ποσοστό επιτυχίας 47%, ποσοστό λανθασμένα θετικό 50% και λανθασμένα αρνητικό 3%. Επίσης τα αποτελέσματα της έρευνας στο [12] καταγράφουν ποσοστό επιτυχίας 55%, ποσοστό λανθασμένα θετικό 44%-45% και λανθασμένα αρνητικό 0%-1%.

Συγκριτικά με τα αποτελέσματα των μελετών που παρατέθηκαν πιο πάνω το σύστημα που αναπτύχθηκε έχει πετύχει ένα αρκετά μεγάλο ποσοστό επιτυχίας καθώς επίσης και

ένα αρκετά μικρό ποσοστό λανθασμένα θετικών εγγραφών ενώ το ποσοστό λανθασμένα αρνητικών κρίνεται ως ικανοποιητικό.

Κεφάλαιο 6
Επίλογος

Στη παρούσα μεταπτυχιακή διατριβή διατυπώσαμε την έννοια της εσωτερικής απειλής και διαπιστώσαμε το μέγεθος της ζημιάς που μπορεί να επιφέρει σε μια επιχείρηση. Ένας κακόβουλος υπάλληλος μια εταιρίας, γνωρίζει αρκετά στοιχεία που αφορούν τον τρόπο σύστασης του εταιρικού δικτύου. Γνωρίζει τα μέτρα ασφάλειας που λαμβάνονται, τους διακομιστές που συνθέτουν το δίκτυο, τον τρόπο επικοινωνίας μεταξύ τους, τα λογισμικά που βρίσκονται εγκατεστημένα στο δίκτυο. Κατέχοντας τη γνώση αυτή καθώς και εκμεταλλευόμενος τη πρόσβαση του στο δίκτυο ο κακόβουλος υπάλληλος πιθανόν να πραγματοποιήσει διάφορες ενέργειές που θα πλήξουν το επιχειρησιακό δίκτυο και κατ' επέκταση την ίδια την επιχείρηση.

Μέσα από τη μελέτη των αποτελεσμάτων που μπορεί να επιφέρει μια επίθεση από εσωτερική απειλή διαπιστώσαμε την ανάγκη προστασίας του εταιρικού δικτύου από τέτοιου είδους επιθέσεις. Για το λόγο αυτό αναπτύξαμε ένα σύστημα ανίχνευσης εσωτερικών απειλών με τη χρήση του αλγορίθμου μη-επιτηρούμενης μάθησης SOM. Το

σύστημα που αναπτύχθηκε και αναλύθηκε στη παρούσα μεταπτυχιακή διατριβή απευθύνεται σε εταιρικά δίκτυα που λειτουργούν χρησιμοποιώντας το λειτουργικό σύστημα Windows.

Για την εξακρίβωση της εγκυρότητας των αποτελεσμάτων που εξάγει το σύστημα ανίχνευσης εσωτερικών απειλών πραγματοποιήσαμε τρείς πειραματικές δοκιμές χρησιμοποιώντας πακέτο δεδομένων που δημιουργήθηκε από εμάς κατά τη διάρκεια της πειραματικής διαδικασίας. Για την εκπλήρωση της πειραματικής διαδικασίας κληθήκαμε να επιλέξουμε τις κατηγορίες ιών που πιθανόν να εγκαταστήσει στο εταιρικό δίκτυο μια εσωτερική απειλή.

Τα αποτελέσματα που εξήχθησαν από τη πειραματική διαδικασία κρίνονται ως αρκετά ικανοποιητικά. σύμφωνα με τη σύγκριση που πραγματοποιήθηκε στο κεφάλαιο 5. Το σύστημα ανίχνευσης εσωτερικών απειλών είχε ποσοστό επιτυχίας 95% ενώ θα πρέπει να τονιστεί ότι το ποσοστό λανθασμένα αρνητικών(false negative) εγγραφών ήταν μόνο 1%.

6.1. Μελλοντική Δουλειά

Κατά την μελέτη της αρχιτεκτονικής του λογισμικού Windows διαπιστώσαμε τη δυνατότητα επέκτασης του συστήματος που αναπτύχθηκε με τη προσθήκη επιπλέον μονάδων επίβλεψης. Για παράδειγμα η προσθήκη μιας μονάδας επίβλεψης του αρχείου registry πιθανότατα να προσθέσει περισσότερη αξιοπιστία στο σύστημα μειώνοντας ακόμη περισσότερό το ποσοστό λανθασμένα αρνητικών εγγραφών.

Όπως αναφέρθηκε σε προηγούμενο κεφάλαιο κρίνεται αναγκαία η δοκιμή του συστήματος με περισσότερες κατηγορίες ιών. Οι κατηγορίες ιών που έχουμε σκοπό να δοκιμάσουμε άμεσα στο σύστημα είναι τα Worms. Μια επίθεση χρησιμοποιώντας ιό που ανήκει στη κατηγορία αυτή πιθανόν να αποβεί μοιραία για την ασφάλεια ενός εταιρικού δικτύου.

Τόσο ο χρόνος που διήρκησαν οι πειραματικές δοκιμές όσο και ο αριθμός των μηχανημάτων που συμμετείχαν σε αυτές δεν θεωρούνται επαρκής για ένα τέτοιο σύστημα. Στο μέλλον θα γίνει προσπάθεια για επανάληψη της πειραματικής διαδικασίας

με τη συμμετοχή περισσότερων μηχανημάτων και αύξηση του χρονικού διαστήματος που θα διαρκέσει. Το ιδανικότερο σενάριο θα ήταν η δοκιμή του συστήματος σε ένα πραγματικό εταιρικό περιβάλλον για το οποίο θα γίνει προσπάθεια υλοποίησης του.

Μέσω της μελέτης της βιβλιογραφίας παρατηρήσαμε την προσπάθεια ομαδοποίησης των χρηστών βάση είτε της γνώσης τους είτε της θέσης τους. Θα ήταν αρκετά ενδιαφέρον να εφαρμόζαμε τη πρακτική αυτή στο δικό μας σύστημα εφόσον πιστεύουμε ότι η ομαδοποίηση αυτή πιθανόν να βοηθήσει τον αλγόριθμο στη καλύτερη και ταχύτερη προσαρμογή των νευρώνων του.

Τέλος σκοπεύουμε να δημιουργήσουμε μια αρκετά πιο εύχρηστη εφαρμογή διαχείρισης του συστήματος ανίχνευσης εσωτερικών απειλών. Η εφαρμογή θα παρέχει στο διαχειριστή του συστήματος μια πιο εύκολη διαχείριση των μηχανήματών που έχουν εγγραφεί στο σύστημα καθώς και μια καλύτερη απεικόνιση των αποτελεσμάτων που εξάγονται από τον αλγόριθμο SOM.

Βιβλιογραφία

[1] «Desktop Operating System Market Share,» Net Applications.com, [Ηλεκτρονικό]. Available: https://www.netmarketshare.com/operating-system-market-share.aspx?qprid=10&qpcustomd=0. [Πρόσβαση 26 08 2015].

[2] «OS Platform Statistics,» w3schools, [Ηλεκτρονικό]. Available: http://www.w3schools.com/browsers/browsers_os.asp. [Πρόσβαση 26 08 2015].

[3] «Global market share held by operating systems Desktop PCs from January 2012 to June 2015,» statista.com, [Ηλεκτρονικό]. Available: http://www.statista.com/statistics/218089/global-market-share-of-windows-7/. [Πρόσβαση 26 08 2015].

[4] Nick Akerman, Ross Anderson, Ashish Arora, Augusto Paes de Barros, Renato Opice Blum, Lynn Robert Carter, Lilian Edwards, Gail F. Farnsely, Marco Gercke, Karthik Kannan, Sivarama Krishnan, Heejo Lee, Tom Longstaff, Jacquelyn Rees, Timothy J. Shimeall, Eug, «Unsecured Economies: Protecting Vital Information,» 2009. [Ηλεκτρονικό]. Available: http://resources.mcafee.com/content/NAUnsecuredEconomiesReport. [Πρόσβαση 03 08 2015].

[5] Debar, Hervé, Marc Dacier, and Andreas Wespi, «Towards a taxonomy of intrusion-detection systems.,» *Computer Networks,* τόμ. 31, αρ. 8, pp. 805-822, 1999.

[6] Greitzer, Frank L., and Deborah A. Frincke., «Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation,» σε *Insider Threats in Cyber Security*, US, Springer, 2010, pp. 85-113.

[7] Legg, Philip, et al., «Towards a conceptual model and reasoning structure for insider threat detection,» *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications,* τόμ. 4, αρ. 4, pp. 20-37, 2013.

[8] Nurse, Jason RC, et al., «Understanding insider threat: A framework for characterising attacks.,» σε *Security and Privacy Workshops (SPW)*, 2014.

[9] Alahmadi, Bushra A., Philip A. Legg, and Jason RC Nurse. , «Using Internet Activity Profiling for Insider-Threat Detection.,» 2015.

[10] Brdiczka, Oliver, et al, «Proactive insider threat detection through graph learning and psychological context.,» σε *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on. IEEE*, 2012.

[11] Myers, Justin, Michael R. Grimaila, and Robert F. Mills, «Towards insider threat detection using web server logs,» σε *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies*, 2009.

[12] Parveen, Pallabi, et al., «Insider threat detection using stream mining and graph mining,» σε *Third Inernational Conference on Social Computing (SocialCom)*, 2011.

[13] Magklaras, G. B., and S. M. Furnell., «Insider threat prediction tool: Evaluating the probability of IT misuse.,» *Computers & Security 21.1,* pp. 62-73, 2001.

[14] Liu, Alexander, et al., «A comparison of system call feature representations for insider threat detection.,» σε *Information Assurance Workshop, 2005. IAW'05. Proceedings from the Sixth Annual IEEE SMC*, 2005.

[15] Maybury, Mark, et al, «Analysis and detection of malicious insiders,» MITRE CORP BEDFORD MA, 2005.

[16] M. Wolff, «Unsupervised Methods for Detecting a Malicious Insider,» σε *Proceedings of the 7th International ISCRAM Conference. Vol. 2*, Seattle, 2010.

[17] Matthias Schonlau, William DuMouchel, Wen-Hua Ju, Alan F. Karr, Martin Theus and Yehuda Vardi, «Computer Intrusion: Detecting Masquerades,» *Statistical Science,* τόμ. 16, αρ. 1, pp. 58-74, 2001.

[18] Nguyen, Nam T., Peter L. Reiher, and Geoffrey H. Kuenning, «Detecting Insider Threats by Monitoring System Call Activity,» *IAW,* pp. 45-52, 2003.

[19] Jirapummin, Chaivat, Naruemon Wattanapongsakorn, and Prasert Kanthamanon, «Hybrid neural networks for intrusion detection system,» *In Proc. of ITC–CSCC,* pp. 928-931, 2002.

[20] Vokorokos, Liberios, Anton Balaz, and Martin Chovanec., «Intrusion detection system using self organizing map.,» *Acta Electrotechnica et Informatica,* τόμ. 6, αρ. 1, pp. 1-6, 2006.

[21] Kumar, Munesh, Shoaib Siddique, and Humera Noor, «Feature-based alert correlation in security systems using self organizing maps,» σε *SPIE Defense, Security, and Sensing. International Society for Optics and Photonics*, 2009.

[22] Depren, Ozgur, Murat Topallar, Emin Anarim, and M. Kemal Ciliz, «An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks,» *Expert systems with Applications,* τόμ. 29, αρ. 4, pp. 713-722, 2005.

[23] Stolfo, S., Apap, F., Eskin, E., Heller, K., Hershkop, S., Honig, A., & Svore, K., «A comparative evaluation of two algorithms for windows registry anomaly detection,» 2005.

[24] «What Is Windows Communication Foundation,» Microsoft, [Ηλεκτρονικό]. Available: https://msdn.microsoft.com/en-us/library/ms731082(v=vs.110).aspx. [Πρόσβαση 30 07 2015].

[25] «Netstat,» Microsoft, [Ηλεκτρονικό]. Available: https://technet.microsoft.com/en-us/library/ff961504.aspx. [Πρόσβαση 22 07 2015].

[26] «Windows service,» Wikipedia, [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Windows_service. [Πρόσβαση 27 07 2015].

[27] «Perfmon,» Microsoft, [Ηλεκτρονικό]. Available: https://technet.microsoft.com/en-us/library/bb490957.aspx. [Πρόσβαση 02 08 2015].

[28] «1999 DARPA Intrusion Detection Evaluation Data Set,» Lincoln Laboratory, Massachusetts Institute of Technology , 01 10 1999. [Ηλεκτρονικό]. Available: https://www.ll.mit.edu/ideval/data/1999data.html. [Πρόσβαση 29 06 2015].

[29] M. V. a. P. K. C. Mahoney, «An analysis of the 1999 DARPA/Lincoln Laboratory evaluation data for network anomaly detection,» σε *Recent Advances in Intrusion Detection*, Berlin Heidelberg, Springer , 2003, pp. 220-237.

[30] Schuba, Christoph L., Ivan V. Krsul, Markus G. Kuhn, Eugene H. Spafford, Aurobindo Sundaram, and Diego ZamboniSchuba, Christoph L., Ivan V. Krsul, Markus G. Kuhn, Eugene H. Spafford, Aurobindo Sundaram, and Diego Zamboni, «Analysis of a denial of service attack on TCP,» *In Security and Privacy,* τόμ. Proceedings, pp. 208-223, 1997.

[31] Kienzle, Darrell M., and Matthew C. Elder, «Recent worms: a survey and trends,» σε *Proceedings of the 2003 ACM workshop on Rapid malcode*, 2003.

[32] Chen, Zhongqiang, Peter Wei, and Alex Delis, «Catching remote administration trojans (RATs),» *Software: Practice and Experience ,* τόμ. 38, αρ. 7, pp. 667-703, 2008.

[33] R. A. Grimes, «Danger: Remote Access Trojans,» Microsoft, 09 2002. [Ηλεκτρονικό]. Available: https://technet.microsoft.com/en-us/library/dd632947.aspx. [Πρόσβαση 28 08 2015].

[34] Stavros N. Shaeles, Ioannis D. Psaroudakis, «A study of a Botnet creation process and the impact of a DDoS attack against a web server,» *Hakin9Bible,* τόμ. 3, αρ. 1, pp. 140-145, 2012.

[35] Abu Rajab, Moheeb, Jay Zarfoss, Fabian Monrose, and Andreas Terzis, «A Multifaceted Approach to Understanding the Botnet Phenomenon,» σε *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, ACM, 2006.

[36] Gu, Guofei, Roberto Perdisci, Junjie Zhang, and Wenke Lee, «BotMiner: Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection,» *USENIX Security Symposium,* τόμ. 5, αρ. 2, pp. 139-154, 2008.

[37] Gu, Guofei, Junjie Zhang, and Wenke Lee, «BotSniffer: Detecting botnet command and control channels in network traffic,» 2008.

[38] Mirković, Jelena, Gregory Prier, and Peter Reiher, «Attacking DDoS at the source,» *10th IEEE International Conference,* pp. 312-321, 2002.

[39] «Asynchronous Programming with Async and Await,» [Ηλεκτρονικό]. Available: https://msdn.microsoft.com/en-us/library/hh191443.aspx. [Πρόσβαση 01 09 2015].

[40] Parveen, Pallabi, et al., «Supervised learning for insider threat detection using stream mining,» σε *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on. IEEE*, 2011.

[41] Kandias, Miltiadis, et al., «An insider threat prediction model.",» σε *Trust, privacy and security in digital business*, Berlin Heidelberg, Springer , 2010, pp. 26-37.

[42] Buford, John F., Lundy Lewis, and Gabriel Jakobson, «"Insider threat detection using situation-aware MAS." Information Fusion,» σε *11th International Conference on. IEEE*, 2008 .

[43] Schultz, E. Eugene, «A framework for understanding and predicting insider attacks,» *Computers & Security,* τόμ. 21, αρ. 6, pp. 526-531, 2002.

[44] Glasser, Joshua, and Brian Lindauer., «Bridging the gap: A pragmatic approach to generating insider threat data.,» σε *Security and Privacy Workshops (SPW)*, IEEE, 2013.

[45] «Generating client workloads and high-fidelity network traffic for controllable, repeatable experiments in computer security.,» σε *Recent advances in intrusion detection*, Berlin Heidelberg, Springer , 2010, pp. 218-237.

[46] Virvilis, Nikos, and Dimitris Gritzalis, «The big four-what we did wrong in advanced persistent threat detection?,» σε *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on. IEEE*, 2013.

[47] Kandias, Miltiadis, et al, «Proactive insider threat detection through social media: The YouTube case,» σε *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society. ACM*, 2013.

[48] Bradford, P., and Ning Hu, «A layered approach to insider threat detection and proactive forensics,» σε *Proceedings of the Twenty-First Annual Computer Security Applications Conference (Technology Blitz)*, 2005.

[49] Kandias, Miltiadis, Nikos Virvilis, and Dimitris Gritzalis, «The insider threat in Cloud computing,» σε *Critical Information Infrastructure Security*, Berlin Heidelberg, Springer , 2013, pp. 93-103.

[50] Kayacik, H. Gunes, A. Nur Zincir-Heywood, and Malcolm I. Heywood, «A hierarchical SOM-based intrusion detection system,» *Engineering Applications of Artificial Intelligence,* τόμ. 20, αρ. 4, pp. 439-451, 2007.

[51] «2000 DARPA Intrusion Detection Scenario Specific Data Sets,» Lincoln Laboratory, Massachusetts Institute of Technology, 01 01 2000. [Ηλεκτρονικό]. Available: https://www.ll.mit.edu/ideval/data/2000data.html. [Πρόσβαση 29 06 2015].

[52] A. D. a. J. S. Wood, «Denial of Service in Sensor Networks,» *Computer ,* τόμ. 35, αρ. 10, pp. 54-62, 2002.

[53] L. Garber, «Denial-of-service attacks rip the Internet,» *Computer ,* τόμ. 4, pp. 12-17, 2000.

[54] Mirkovic, Jelena, and Peter Reiher, «A taxonomy of DDoS attack and DDoS defense mechanisms,» *ACM SIGCOMM Computer Communication Review,* τόμ. 34, αρ. 2, pp. 39-53, 2004.

[55] Ioannidis, John, and Steven Michael Bellovin, «Implementing pushback: Router-based defense against DDoS attacks,» 2002.

Παράρτημα Α
Οδηγίες Εγκατάστασης Συστήματος

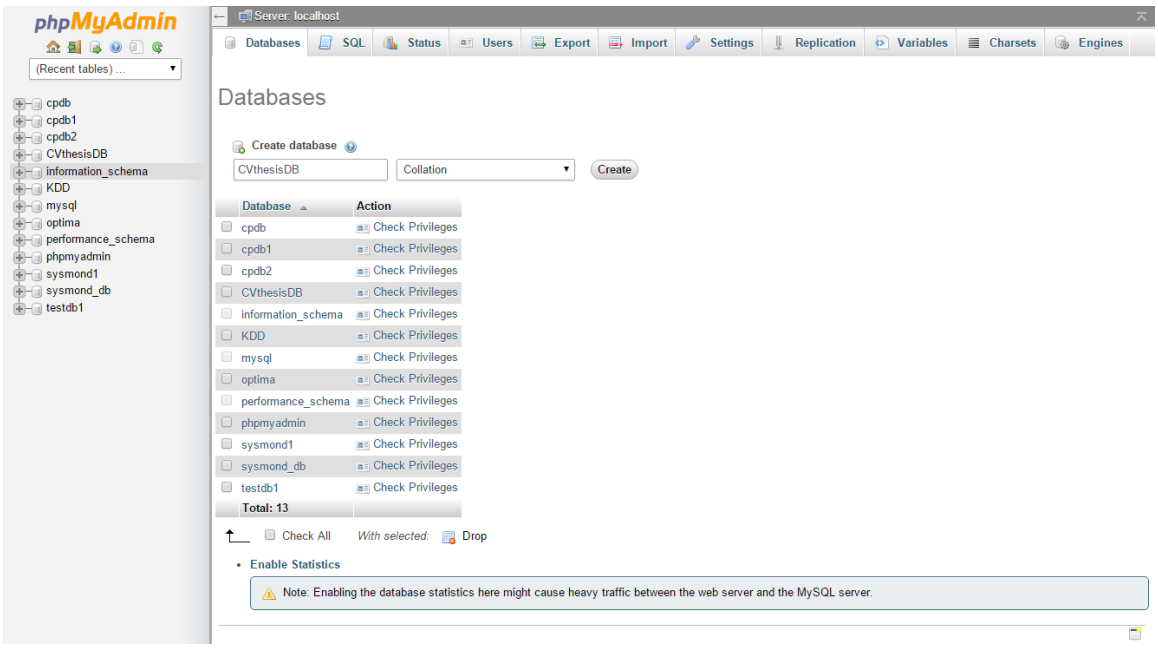
Α.1 Βάση Δεδομένων

Προϋποθέσεις:

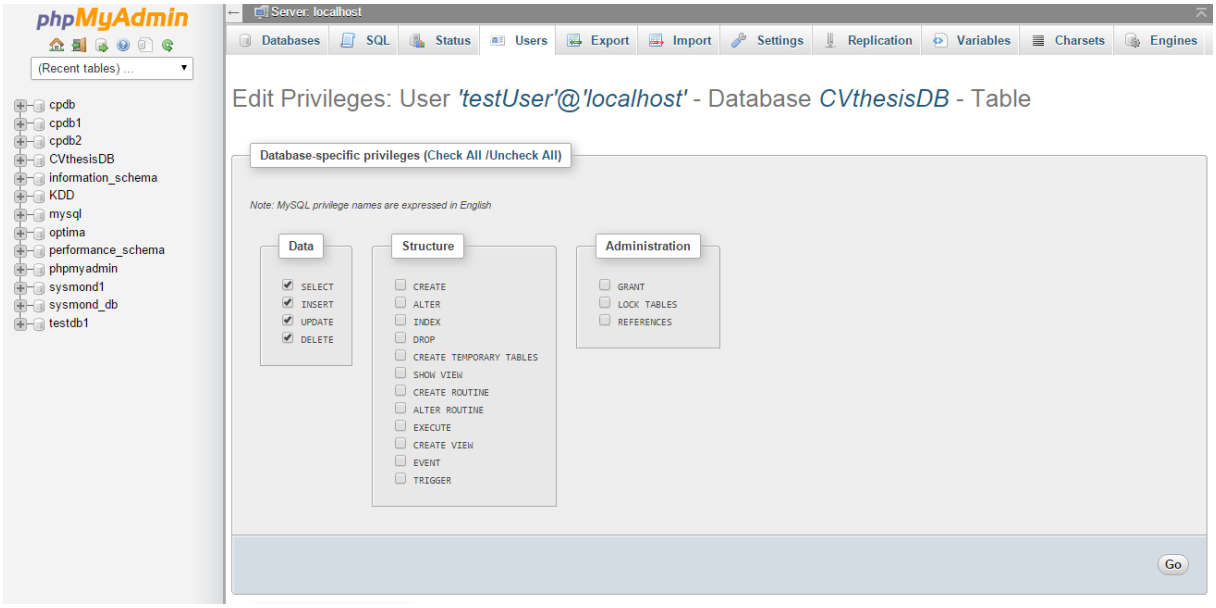
1. MySQL Server 5.5.44 ή μεγαλύτερη έκδοση
2. Apatche Server 2.4.7 ή μεγαλύτερη έκδοση
3. phpMyAdmin 4.0.10 ή μεγαλύτερη έκδοση

Βήματα:

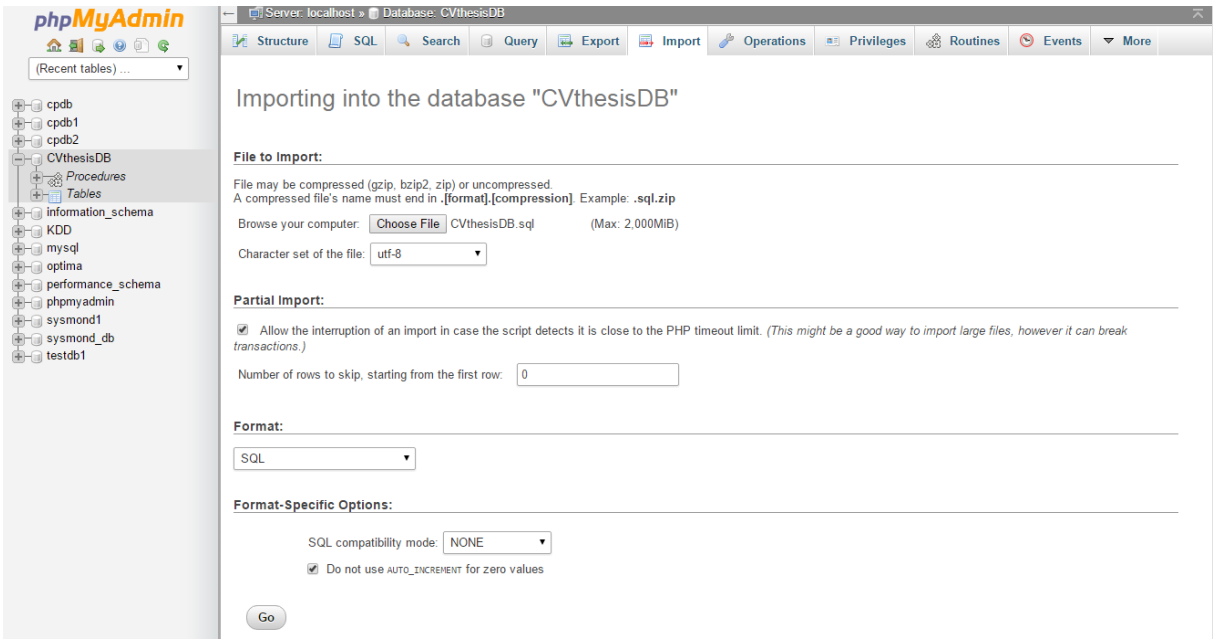
1. Πραγματοποιήστε είσοδο στο πάνελ του phpMyAdmin
2. Δημιουργήστε μια καινούρια βάση δεδομένων με το όνομα «CVthesisDB»



3. Δημιουργήστε ένα καινούριο χρήστη και δώστε του όλα τα δικαιώματα «Select, Insert, Update, Delete» για τη βάση «CVthesisDB»



4. Κάντε import στη βάση το αρχείο «CVthesisDB.sql»



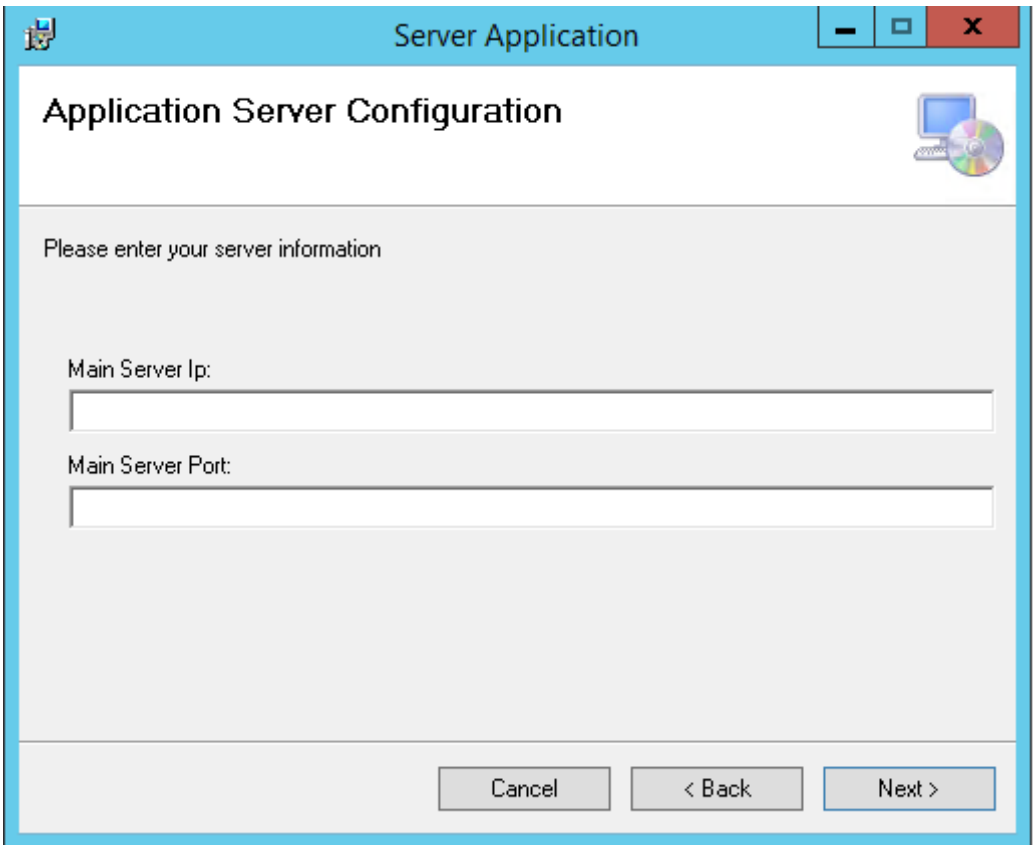
A.2 Λογισμικό Διακομιστή

Προϋποθέσεις:

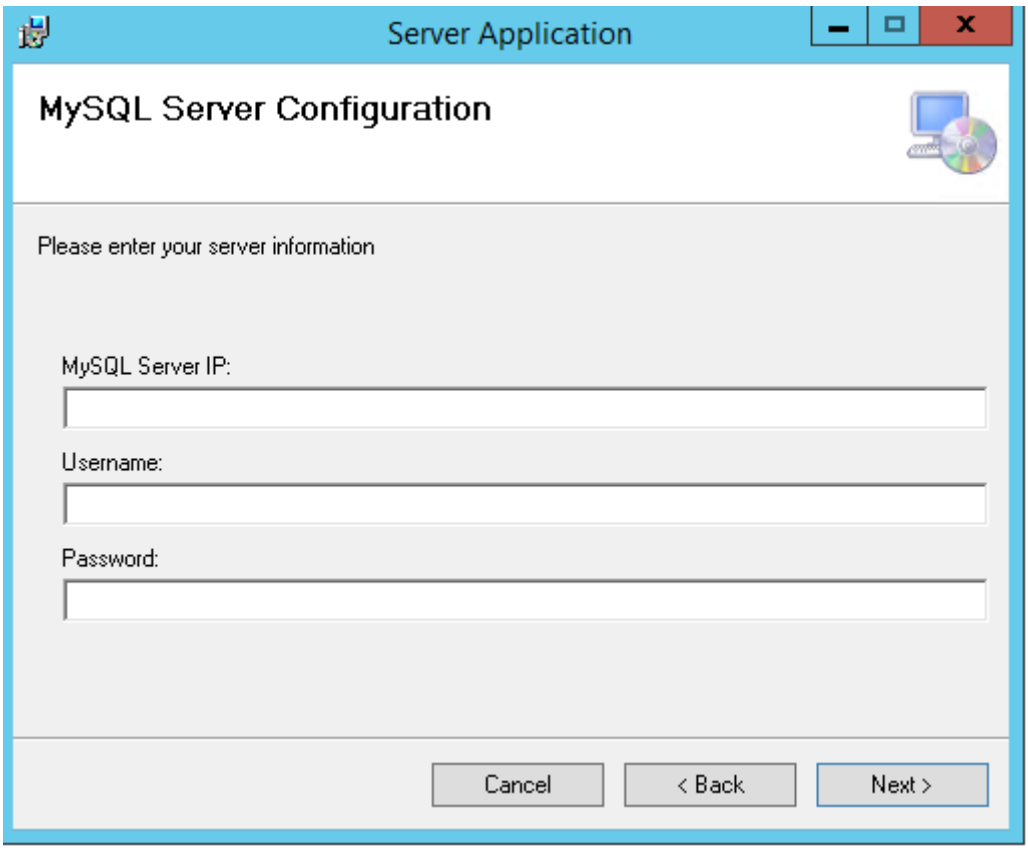
1. Το λειτουργικό του διακομιστή θα πρέπει να είναι της οικογένειας Windows
2. Στο διακομιστή θα πρέπει να είναι εγκατεστημένο το framework .NET 4.5
3. Βεβαιωθείτε ότι έχετε administrative privileges στο διακομιστή

Βήματα:

1. Μεταφέρετε το εκτελέσιμο αρχείο «ServerInstallation.exe» στο διακομιστή εφαρμογών
2. Εκτελέστε το αρχείο «ServerInstallation.exe» επιλέγοντας «Run as admin»
3. Στην οθόνη «Select Installation Folder» αφήστε τις προεπιλεγμένες επιλογές και επιλέξτε «Next»
4. Στην οθόνη «Application Server Configuration» καταχωρήστε την διεύθυνση IP του διακομιστή καθώς και τη πόρτα που θα χρησιμοποιεί το σύστημα. Τα στοιχεία αυτά θα χρησιμοποιούνται για τη σύνδεση μεταξύ διακομιστή και υπολογιστή χρήστη. Διαβεβαιωθείτε ότι η πόρτα είναι ανοιχτή από τον Firewall. Επιλέξτε «Next»



5. Στην οθόνη «MySQL Server Configuration» καταχωρήστε τη διεύθυνση IP του διακομιστή βάσεων δεδομένων. Η πόρτα που χρησιμοποιείται για τη σύνδεση αυτή είναι 8523, διαβεβαιωθείτε ότι η πόρτα είναι ανοιχτή από τον Firewall. Στα πεδία «Username» και «Password» καταχωρείστε τα στοιχεία του χρήστη που δημιουργήσατε στο βήμα Α.1.3. Επιλέξτε «Next»



6. Στην οθόνη «Confirm Installation» επιλέξτε «Next». Αφού τελειώσει η διαδικασία της εγκατάστασης επιλέξτε «Close»

Α.3 Λογισμικό Υπολογιστή Χρήστη

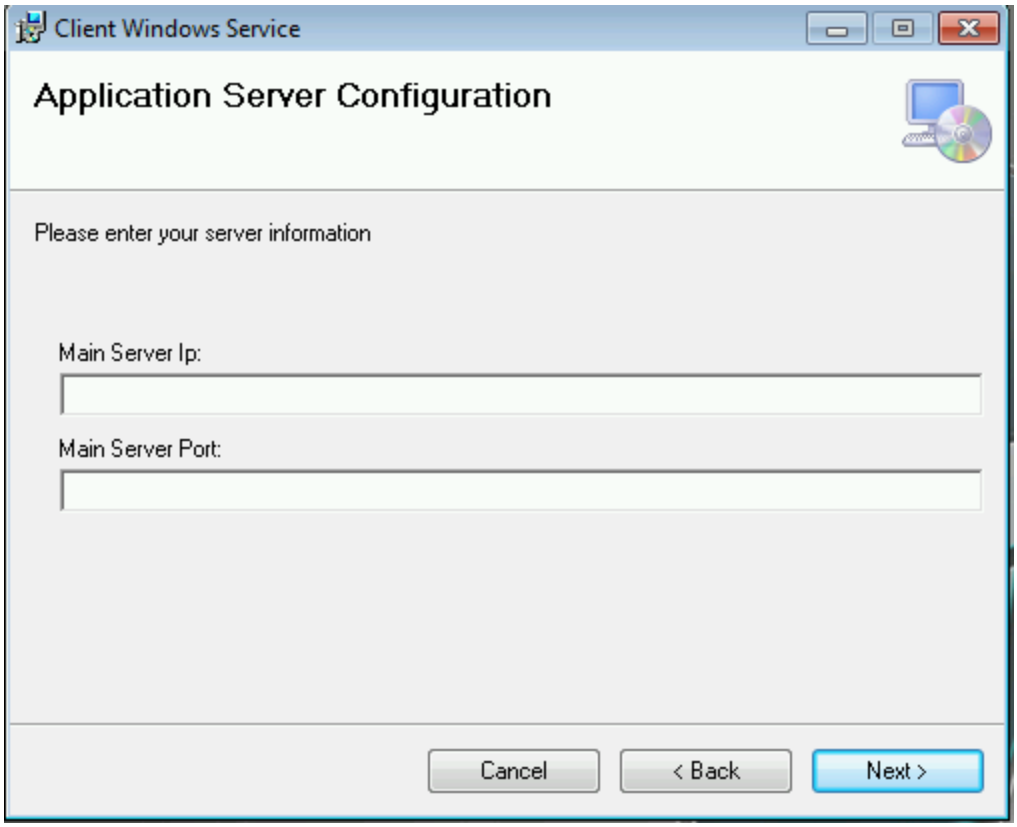
Προϋποθέσεις:

1. Το λειτουργικό του υπολογιστή χρήστη θα πρέπει να είναι της οικογένειας Windows
2. Στον υπολογιστή θα πρέπει να είναι εγκατεστημένο το framework .NET 4.5
3. Βεβαιωθείτε ότι έχετε administrative privileges στον υπολογιστή

Βήματα:

1. Μεταφέρετε το εκτελέσιμο αρχείο «ClientInstallation.exe» στον υπολογιστή
2. Εκτελέστε το αρχείο « ClientInstallation.exe» επιλέγοντας «Run as admin»
3. Στην οθόνη «Select Installation Folder» αφήστε τις προεπιλεγμένες επιλογές και επιλέξτε «Next»
4. Στην οθόνη «Application Server Configuration» καταχωρήστε την διεύθυνση IP του διακομιστή καθώς και τη πόρτα που θα χρησιμοποιεί το σύστημα. Τα

στοιχεία αυτά θα χρησιμοποιούνται για τη σύνδεση μεταξύ διακομιστή και υπολογιστή χρήστη. Διαβεβαιωθείτε ότι η πόρτα είναι ανοιχτή από τον Firewall και είναι η ίδια με τη πόρτα που καταχωρήθηκε στο βήμα Α.2.4. Επιλέξτε «Next»

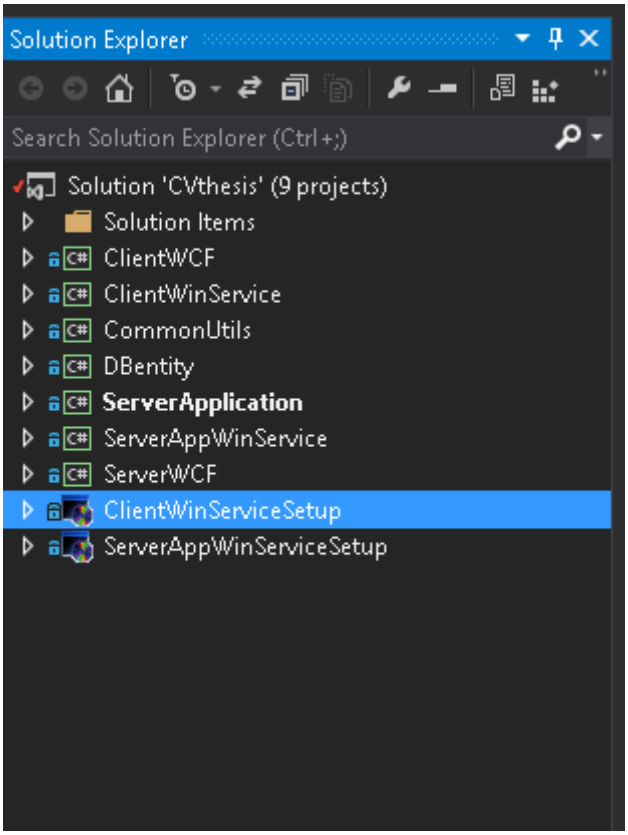


5. Στην οθόνη «Confirm Installation» επιλέξτε «Next». Αφού τελειώσει η διαδικασία της εγκατάστασης επιλέξτε «Close

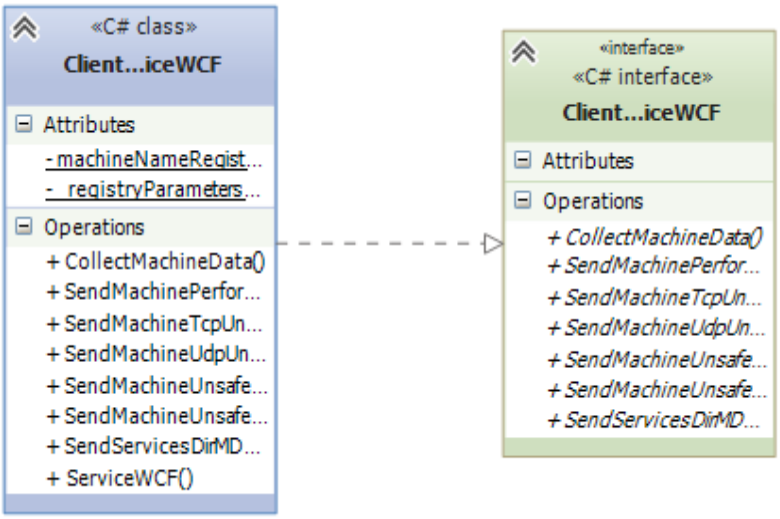
Παράρτημα Β

Διαγράμματα Κλάσεων Συστήματος

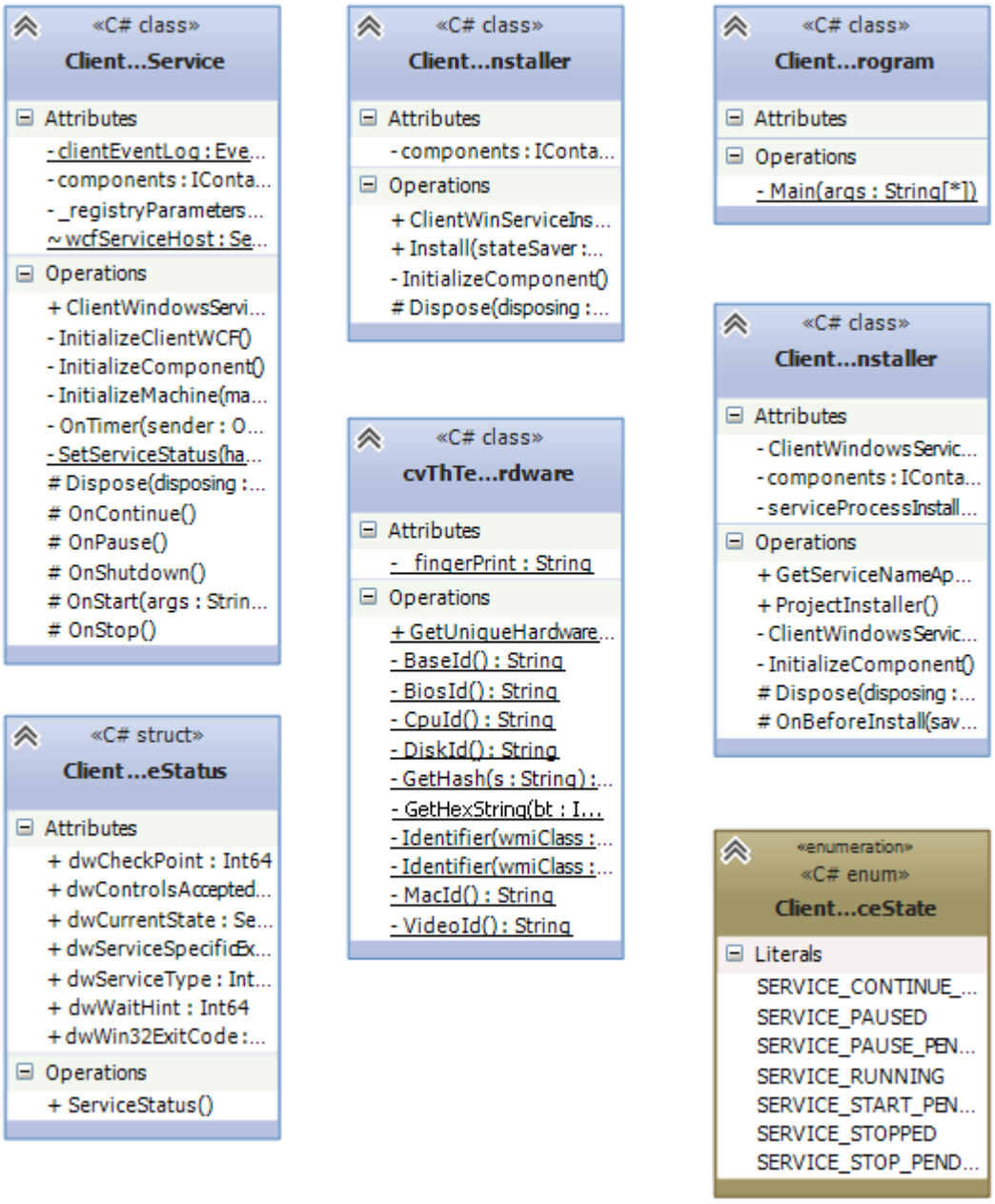
B.1 Visual Studio Solution Window



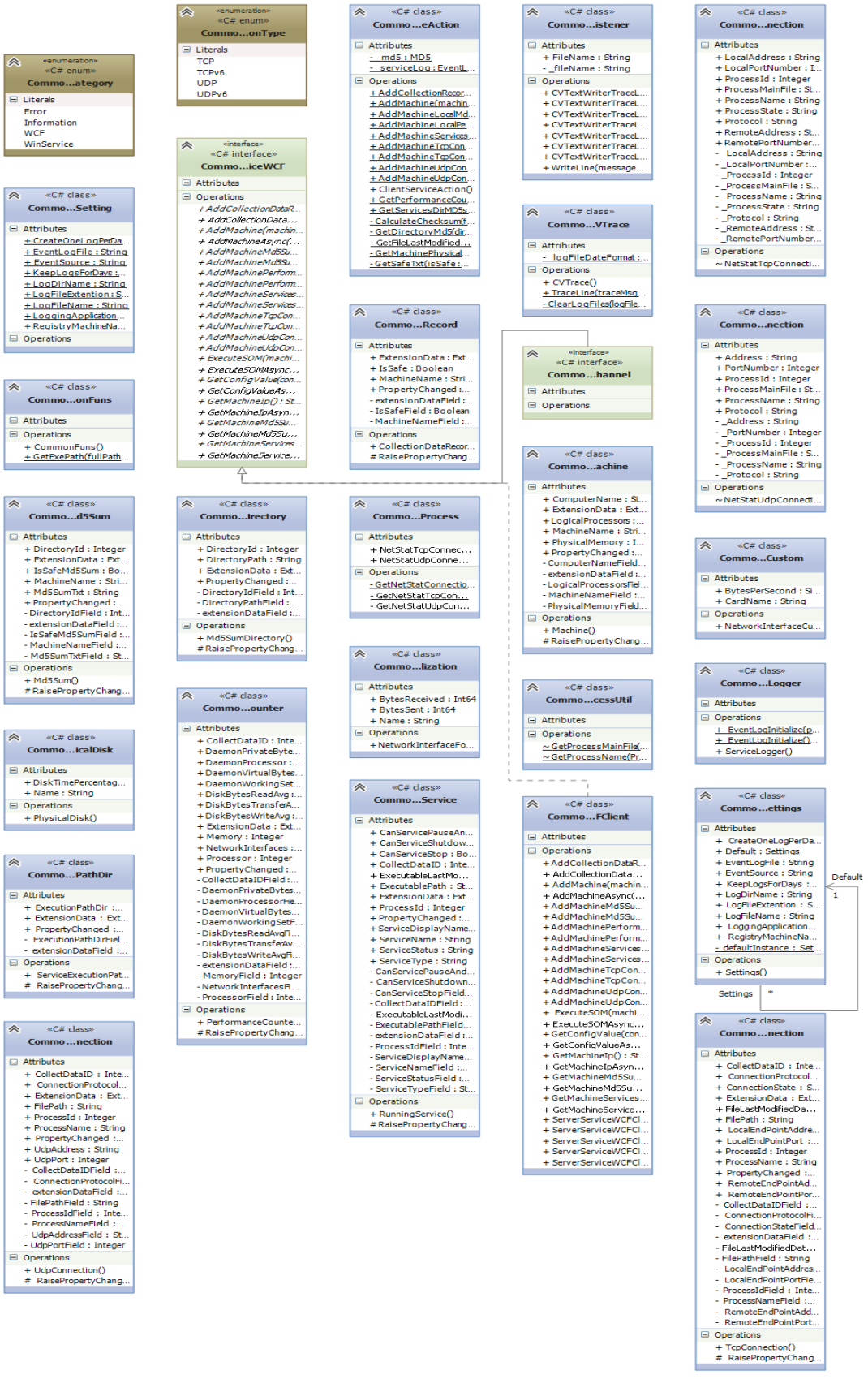
B.2 ClientWCF Class Diagram



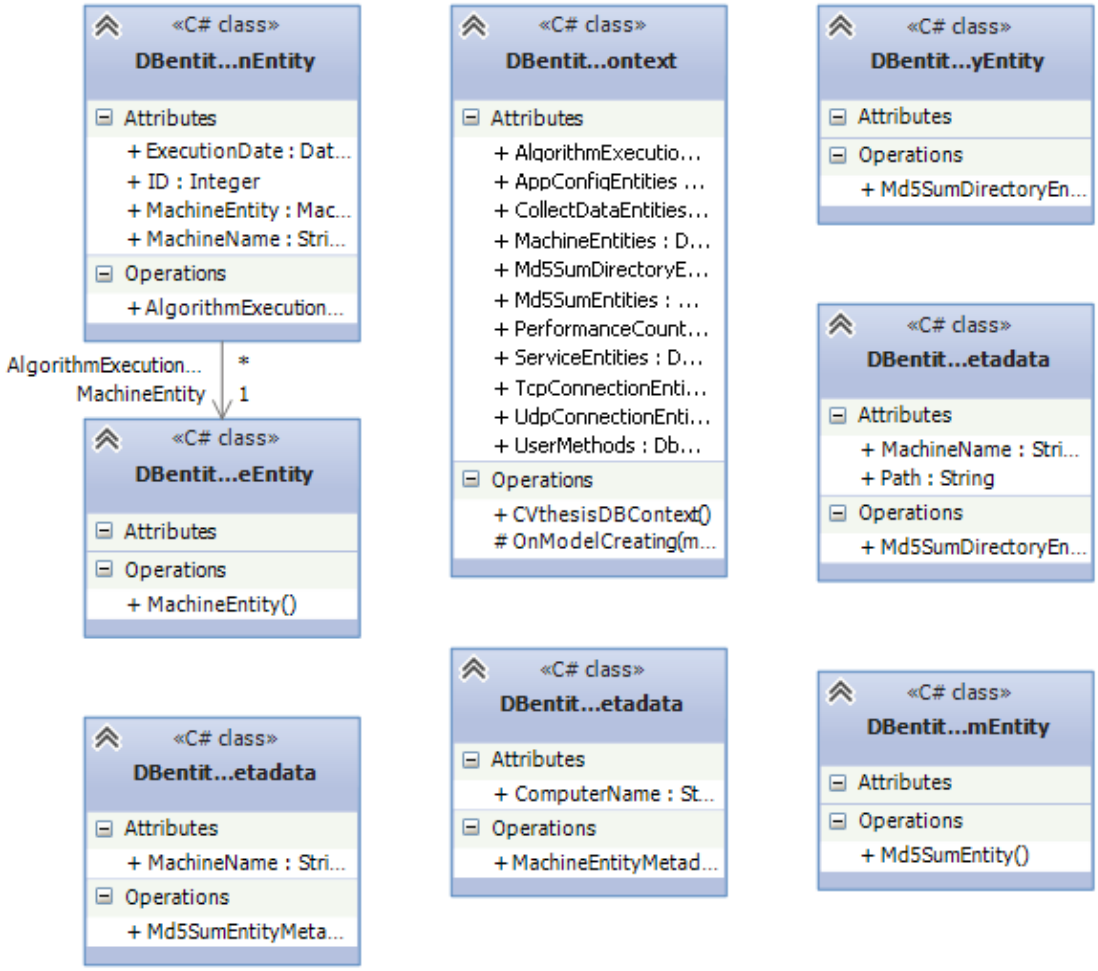
B.3 ClientWinService Class Diagram



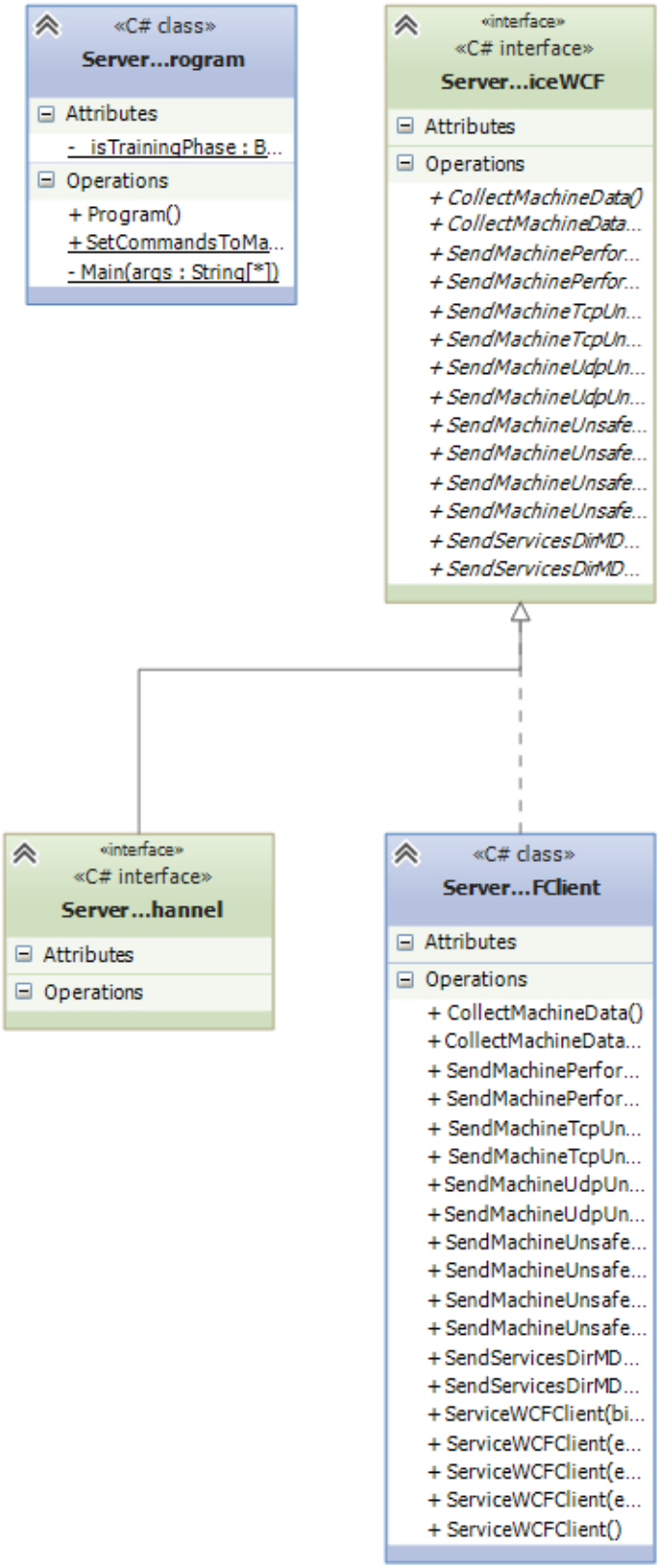
B.4 CommonUtils Class Diagram



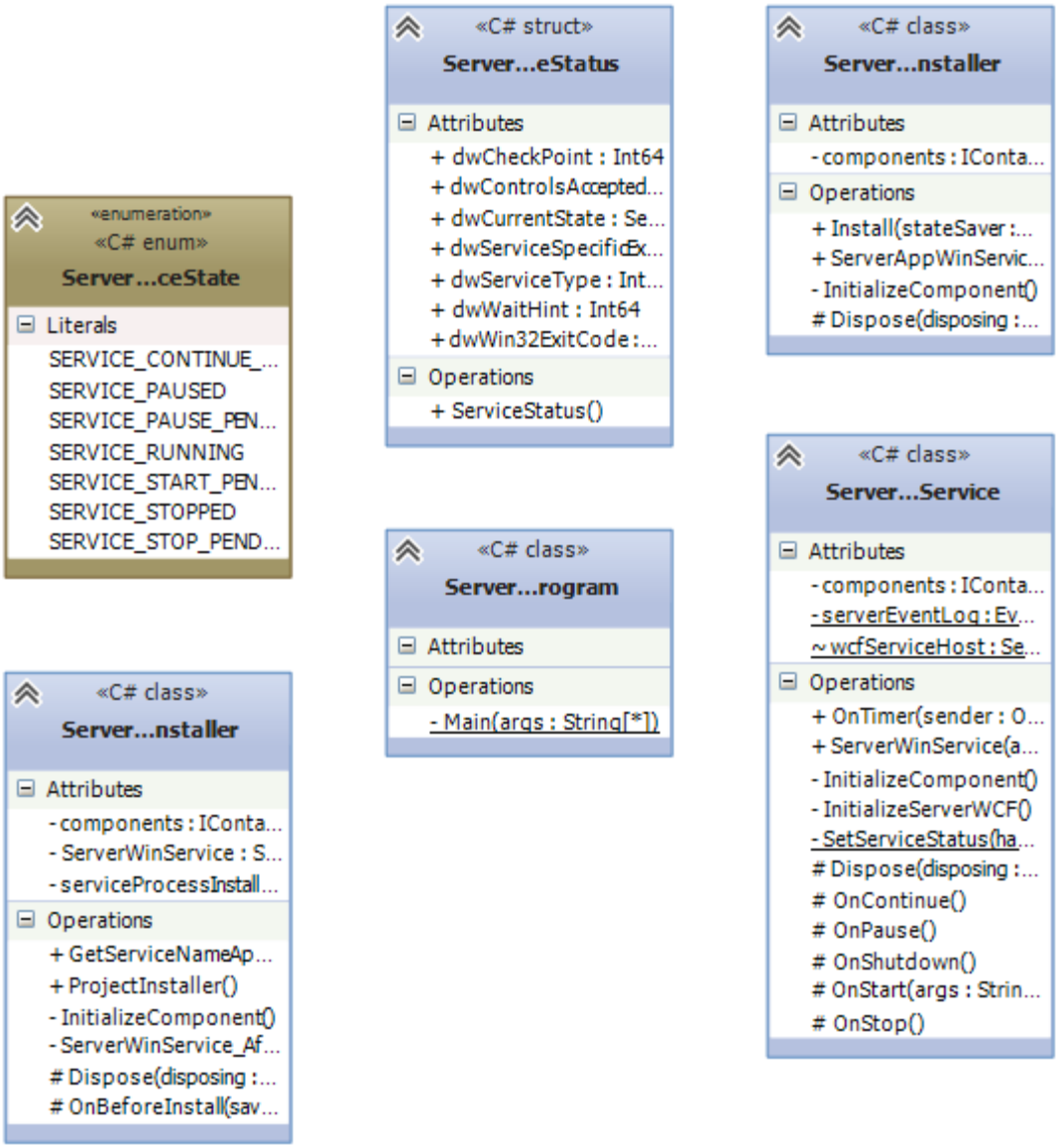
B.5 DBentity Class Diagram



B.6 ServerApplication Class Diagram



B.7 ServerAppWinService Class Diagram



B.8 ServerWCF Class Diagram