Πανεπιστήμιο Πειραιώς
Τμήμα Ψηφιακών Συστημάτων
ΠΜΣ Ασφάλειας Ψηφιακών Συστημάτων

Μεταπτυχιακή Διατριβή

Σύνολα Δεδομένων για Έλεγχο Επίδοσης Συστημάτων Ανίχνευσης Παρεισφρήσεων

Datasets for Performance Evaluation of Intrusion Detection Systems

Σιάμπος Θεοφάνης

Φεβρουάριος 2014

Επιβλέπων Καθηγητής

Σωκράτης Κάτσικας, Καθηγητής
Πανεπιστήμιο Πειραιώς

Εξεταστική Επιτροπή

Σωκράτης Κάτσικας, Καθηγητής
Πανεπιστήμιο Πειραιώς

Κωνσταντίνος Λαμπρινουδάκης, Αναπληρωτής Καθηγητής
Πανεπιστήμιο Πειραιώς

Χρήστος Ξενάκης, Επίκουρος Καθηγητής
Πανεπιστήμιο Πειραιώς

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Ευχαριστίες

Θα ήθελα να απονείμω τις θερμότερες ευχαριστίες στον επιβλέποντα καθηγητή μου Σωκράτη Κάτσικα για την ευκαιρία που μου έδωσε να ασχοληθώ με το συγκεκριμένο θέμα καθώς και για την άρτια επιστημονική καθοδήγηση, την υπομονή και το ενδιαφέρον που επέδειξε καθ'ολη τη διάρκεια εκπόνησης της εν λόγω διπλωματικής εργασίας. Όλα τα παραπάνω συνέβαλαν καταλυτικά στην επιτυχή ολοκλήρωση της.

Ιδιαίτερη μνεία και ευχαριστίες οφείλω να εκφράσω ακόμα στον καθηγητή και μέντορα Γιώργο Σπαθούλα για τις διαφωτιστικές συμβουλές που μου παρείχε κατά τους πρώτους μήνες ενασχόλησης μου με το συγκεκριμένο θέμα της εργασίας αυτής αφού χωρίς αυτόν θα ήταν αδύνατη η υλοποίηση της.

Πειραιάς, 19 Φεβρουαρίου 2014

Περίληψη

Η ασφάλεια και προστασία κρίσιμων δεδομένων αποτελεί μία διαρκή ανησυχία στη σύγχρονη τεχνολογική εποχή. Τα συστήματα ανίχνευσης και καταγραφής εισβολών απαρτίζουν μία μεγάλη ομάδα εργαλείων τα οποία εγκαθιδρύονται με στόχο την καταπολέμηση των απειλών που εμφανίζονται καθημερινά στον ηλεκτρονικό κόσμο. Προκειμένου αυτά τα συστήματα να είναι σε θέση να εντοπίζουν όλο και περισσότερες νέες επιθέσεις αλλά και να αποτελούν ένα ισχυρό μέσο άμυνας έναντι των επιτιθέμενων πρέπει οι άνθρωποι οι οποίοι χειρίζονται αυτά, να τα βελτιώνουν διαρκώς έτσι ώστε να προσαρμόζονται κατάλληλα κάθε φορά σε νέες απειλές και να διευρύνονται οι δυνατότητες ανίχνευσης τους. Για να επιτύχουμε κάτι τέτοιο πρέπει αρχικά να αξιολογήσουμε σύνολα δεδομένων τα οποία παράγονται από τέτοια συστήματα και περιέχουν τα αποτυπώματα κίνησης που καταγράφηκαν από τα συστήματα ανίχνευσης εισβολών. Συνεπώς, έχοντας αυτά ως όπλο, είμαστε σε θέση να τα μελετήσουμε και να δούμε ποιές επιθέσεις μπόρεσε ένα τέτοιο σύστημα να αναγνωρίσει επιτυχώς, ποιές όχι ή και ποιό είδος κίνησης αναγνώρισε λανθασμένα ως κακόβουλη κίνηση.

Ακολουθώντας τις παραπάνω παραδοχές, στην παρούσα εργασία πραγματοποιούμε μία περιγραφή και συγκριτική ανάλυση τέτοιων συνόλων δεδομένων που είναι κατά κύριο λόγο δημοσίως διαθέσιμα. Στη συνέχεια, δημιουργήσαμε ένα νέο dataset για μελέτη επίδοσης δικτυακών IDS (Intrusion Detection Systems) το οποίο καλύπτει ένα μεγάλο εύρος σύγχρονων απειλών που εμφανίζονται σήμερα στο διαδίκτυο και το οποίο προσπαθεί να διορθώσει αδυναμίες των προηγούμενων συνόλων τις οποίες επίσης αναφέρουμε. Ακόμα περιγράφουμε λεπτομερώς τη δομή και τον τρόπο με τον οποίο δημιουργήθηκε το dataset αυτό καθώς επίσης αναλύουμε παρεισφρήσεις και κακόβουλες συμπεριφορές οι οποίες καταγράφηκαν.

Έπειτα, παρουσιάζουμε ένα λογισμικό προσομοίωσης το οποίο δημιουργήσαμε με στόχο την εικονική αναπαράσταση ενός δικτυακού IDS καθώς και την αυτοματοποιήμενη εξαγωγή datasets από το εκάστοτε σύστημα ανίχνευσης εισβολών. Ο τρέχων προσομοιωτής είναι σε θέση να εκτελεί την πλειοψηφία των διαδικασιών που υλοποιούνται σε αυτά τα συστήματα.

Λέξεις Κλειδιά: σύνολα δεδομένων, εισβολές, παρεισφρήσεις, συστήματα ανίχνευσης, επιθέσεις, αποτυπώματα κίνησης, δίκτυα, κακόβουλη κίνηση, προσομοίωση

Abstract

Security and protection of critical data has been an ever-growing concern in the modern technological age. Intrusion detection and monitoring systems consist a large group of tools that are installed with the aim of fighting threats that are appeared in a daily basis in the electronic world. In order these systems being able to detect as far as more new attacks and consist a powerful means of defense against attackers, it is indispensable that people who handle them, should constantly improve them so as to adjust them every time into new kinds of threats and broaden their detection capabilities. For achieving something like that, at fist we need to evaluate datasets that are generated from such systems and include traffic captures that were recorded from current systems. Consequently, having these sets we are able to study and observe which attacks were identified successfully, which not and what kind of traffic was considered falsely as malicious traffic.

Following the above assumptions, in the current project we make a description and comparative analysis between existing datasets. Then, we created a new dataset for network-based IDS which covers a large range of threats that are appeared on the web today and is trying to correct weaknesses of previous sets that we mention. Also, we describe in detail the structure and the way with which we created this dataset as well we analyze intrusions and malicious behaviours that were captured.

Next, we introduce a simulation software which was developed with the aim of establishing a virtual representation of a network intrusion detection system and automating the exportation of datasets from each IDS. The current simulator is able to perform the majority of procedures that are implemented in these systems.

Keywords: datasets, intrusions, detection systems, attacks, captures, networks, malicious traffic, simulation

Κεφάλαιο 1ο

Εισαγωγή

Η συνεχής και ραγδαία χρησιμοποίηση δικτύων καθώς και η παρουσία ολοένα και περισσοτέρων υψηλής κρισιμότητας δεδομένων τα οποία μεταδίδονται μέσω διαδικτύου και αποθηκεύονται καθημερινά, έχει οδηγήσει στην αυξανόμενη προσπάθεια των επιτιθέμενων για διείσδυση σε αυτά. Παραδείγματα δικτυακών επιθέσεων αποτελούν η μη ηθελημένη τροποποίηση δεδομένων, οι επιθέσεις άρνησης υπηρεσιών (DoS) καθώς και η μη εξουσιοδοτημένη ανάκτηση ευαίσθητων πληροφοριών. Αυτή η τάση έχει προκαλέσει τα τελευταία χρόνια το έντονο ενδιαφέρον για εύρεση καινοτόμων μεθόδων με τις οποίες θα μπορούμε να αναγνωρίζουμε αρχικά τέτοιες επιθέσεις. Έτσι η ανάγκη αυτή έφερε στο προσκήνιο την εισαγωγή των δικτυακών συστημάτων ανίχνευσης εισβολών. Τα συστήματα αυτά είναι ικανά να καταγράφουν τη δικτυακή κίνηση που λαμβάνει χώρα σε πραγματικό χρόνο καθώς επίσης και να ανιχνεύουν ύποπτες δραστηριότητες.

1.1 Περιγραφή του προβλήματος

Τα σύγχρονα συστήματα ανίχνευσης εισβολών παρέχουν μία μεγάλη γκάμα από τεχνικές (υπογραφές, ανίχνευση ανωμαλιών, υβριδικότητα) οι οποίες μπορούν να χρησιμοποιηθούν για τη συλλογή και ανάλυση δεδομένων. Γενικά στόχος των συστημάτων αυτών είναι η αναγνώριση πιθανών κακόβουλων συμβάντων καθώς και η καταγραφή πληροφοριών σχετικά με αυτά. Σε περίπτωση που εντοπιστεί μία απειλή, τα IDS αναφέρουν τις πιθανές αυτές προσπάθειες για εισβολή. Τα περισσότερα συστήματα τέτοιου είδους βασίζονται σε μία κοινή αρχιτεκτονική η οποία περιλαμβάνει έναν κύριο ή πολλαπλούς δευτερεύοντες ανιχνευτές οι οποίοι συλλέγουν δεδομένα τα οποία πιθανόν να αποτελούν ένδειξη για ύπαρξη εισβολών, ένα σύστημα ανάλυσης δεδομένων το οποίο επεξεργάζεται τα εισερχόμενα δεδομένα με στόχο την αναγνώριση κακόβουλης δραστηριότητας και μία απάντηση (response) η οποία παράγει διάφορους συναγερμούς ασφαλείας (alerts) και αναφέρει όλες τις πιθανές εισβολές.

Αφού ολοκληρωθεί η καταγραφή όλων των επιθέσεων σε ένα συγκεκριμένο χρονικό διάστημα, το επόμενο βήμα αποτελεί η δημιουργία και εκτύπωση του συνόλου δεδομένων που

αφορά όλη την τρέχουσα δικτυακή κίνηση που καταγράφηκε. Το σύνολο δεδομένων αυτό στη συνέχεια θα μελετηθεί προκειμένου να εξαχθούν συμπεράσματα σχετικά με την ποιότητα του δικτυακού συστήματος ανίχνευσης αλλά και να μελετηθεί περαιτέρω η κίνηση στην οποία το δικτακό IDS αδυνατεί να εντοπίσει εισβολές. Έτσι η συνεχής μελέτη τέτοιων ειδών datasets μας βοηθάει να αξιολογούμε αυτά τα συστήματα αλλά και να τα βελτιώνουμε περαιτέρω προσθέτοντας νέους κανόνες (rules) βάσει των οποίων γίνεται κάθε φορά η πιθανή αναγνώριση επιθέσεων από τα IDS. Αυτή η ανάγκη κρίνεται ακόμα μεγαλύτερη δεδομένου ότι τα συστήματα αυτά φτιάχνονται από ανθρώπους για ανθρώπους. Επομένως, ένα σύστημα ανίχνευσης εισβολών θα πρέπει να αναγνωρίζει όλες τις πιθανές ενέργειες ενός hacker κάτι το οποίο πρακτικά είναι αδύνατον. Στο παρελθόν, διάφορες προσπάθειες έχουν πραγματοποιηθεί εργαστηριακά για μία βέλτιστη και αξιόπιστη ανάλυση υπαρχόντων συνόλων δεδομένων με πιο γνωστό εγχείρημα το KDD Cup 99 dataset καθώς και άλλα όπως το DARPA 1999 ή τα DEFCON datasets. Το κύριο όμως πρόβλημα είναι ότι πολλά από αυτά τα datasets δε δημοσιεύονται με συνέπεια η περαιτέρω έρευνα σε αυτό το πεδίο να παραμένει δύσκολη.

Αναλύοντας σε βάθος ένα IDS, θα παρατηρήσουμε ότι το τελευταίο μπορεί να αποτελεί ένα κομμάτι εγκατεστημένου λογισμικού ή μία φυσικού τύπου συσκευή η οποία καταγράφει δικτυακή κίνηση με σκοπό την ανίχνευση ανεπιθύμητης δραστηριότητας και συμβάντων όπως ύποπτης ή κακόβουλης κίνησης καθώς και κίνηση η οποία παραβιάζει μία υπάρχουσα πολιτική ασφάλειας. Πολλά εργαλεία ανίχνευσης δραστηριότητας όπως το Snort και το Bro, αποθηκεύουν ένα συγκεκριμένο συμβάν σε αρχεία καταγραφής τα οποία μπορούν να εξεταστούν σε μεταγενέστερο χρονικά σημείο και συνδυάζοντας αυτά μαζί με άλλα υπάρχοντα δεδομένα, υπάρχει η δυνατότητα για πιθανή αναθεώρηση στρατηγικών με σκοπό τη βέλτιστη διαχείριση ενδεχόμενων κινδύνων.

1.2 Δομή της διατριβής

Η διατριβή αποτελείται από οκτώ κεφάλαια. Στις επόμενες παραγράφους περιγράφεται συνοπτικά το περιεχόμενο των κεφαλαίων αυτών.

Το κεφάλαιο 1 περιέχει την περιγραφή του προβλήματος και τους στόχους, καθώς επίσης τη δομή και τη συνεισφορά της διατριβής.

Το κεφάλαιο 2 αναλύει τα διαφορετικά είδη συνόλων δεδομένων για συστήματα ανίχνευσης παρεισφρήσεων που μπορούν να εμφανισθούν και παράλληλα πραγματοποιείται μία συγκριτική περιγραφή αυτών.

Στο κεφάλαιο 3 γίνεται μία παρουσίαση συνόλων δεδομένων τα οποία βρίσκονται διαθέσιμα στο διαδίκτυο (άλλα δημόσια και άλλα ιδιωτικά). Ειδικότερα, κάνουμε μία γενική περιγραφή αυτών με βάση πειραματικά αποτελέσματα που εξάγαμε ύστερα από εκτέλεση διαδικασιών ανίχνευσης στα τρέχοντα αποτυπώματα κίνησης.

Στο κεφάλαιο 4 πραματοποιούμε μία συγκριτική αξιολόγηση των προαναφερθέντων datasets και εστιάζουμε ξεχωριστά στις αδυναμίες καθενός από αυτά.

Στο κεφάλαιο 5 περιγράφουμε ένα καινούργιο σύνολο δεδομένων που δημιουργήσαμε τεχνητά με σκοπό την κάλυψη αδυναμιών προηγούμενων συνόλων. Συγκεκριμένα εδώ, κάνουμε μία αναφορά στη δομή και τα μέρη από τα οποία απαρτίζεται το dataset αυτό. Παράλληλα, περιγράφουμε τα μέσα και τα εργαλεία (traffic generators) μέσω των οποίων έγινε η παραγωγή φυσιολογικής και κακόβουλης κίνησης και επιπλέον παρουσιάζουμε σχήματα στα οποία αναπαριστάται γραφικά μέρος της κίνησης αυτής.

Το κεφάλαιο 6 παρέχει μία λεπτομερή περιγραφή των κύριων απειλών που καταγράφηκαν στο νέο dataset. Εδώ αναλύονται επίσης συγκεκριμένες καταστάσεις οι οποίες δικαιολογούν την ύπαρξη των εισβολών αυτών και ακόμα υπολογίζονται χαρακτηριστικές τιμές με στόχο την μελέτη του συσχετισμού των παραγόμενων alerts.

Στο κεφάλαιο 7 παρουσιάζεται ένα καινούργιο λογισμικό προσομοίωσης για δικτυακά συστήματα ανίχνευσης παρεισφρήσεων το οποίο προσομοιώνει τη συμπεριφορά επιτιθέμενου-αμυνόνεμου με στόχο την αυτοματοποιήμενη εξαγωγή συνόλων δεδομένων με τεχνητό τρόπο.

Η ολοκλήρωση της διατριβής γίνεται με το κεφάλαιο 8, όπου παρατίθενται τα γενικά συμπεράσματα από την όλη ερευνητική προσπάθεια, αλλά και τα ανοικτά ερευνητικά θέματα που παραμένουν προς διερεύνηση.

1.3 Συνεισφορά της διατριβής

Η ερευνητική προσπάθεια που συντελέστηκε στο πλαίσιο της διατριβής συμβάλλει κατά κύριο λόγο στη βελτίωση της επίδοσης των συστημάτων ανίχνευσης παρεισφρήσεων μέσα από ανάλυση των ακόλουθων στοιχείων:

- Περιγραφή καινούργιων datasets μερικά από τα οποία δεν έχουν μελετηθεί ερευνητικά ως προς την αποδοτικότητα που παρέχουν και το σκοπό για τον οποίο

δημιουργήθηκαν.

- Συγκριτική αξιολόγηση υπαρχόντων datasets (τα οποία περιγράφονται) και εξαγωγή γενικού πορίσματος ως προς τη δυναμική αυτών αλλά και τις αδυναμίες τους.

- Δημιουργία νέου συνθετικού συνόλου δεδομένων ικανού να ενσωματώσει επιθέσεις καινούργιες οι οποίες δεν περιλαμβάνονται στα υπάρχοντα datasets.

- Έμφαση σε απομακρυσμένες επιθέσεις οι οποίες λαμβάνουν μέρος κατά την επικοινωνία δύο ή περισσοτέρων χρηστών (remote bufferoverflows σε instant messengers, voip επιθέσεις κ.α).

- Υλοποίηση μίας τεχνητής μεθόδου παραγωγής κίνησης (φυσιολογικής και κακόβουλης) η οποία χρησιμοποιεί πολλαπλούς υπάρχοντες traffic generators.

- Εγκαθίδρυση εικονικών δικτύων ανωνυμίας με στόχο την παραγωγή κακόβουλης κίνησης από εισβολείς των οποίων η δικτυακή ταυτότητα δεν είναι η πραγματική και αλλάζει συνεχώς.

- Έμφαση σε επιθέσεις σύγχρονες οι οποίες δεν έχουν καταγραφεί σε προηγούμενα δημοσιευμένα datasets.

- Εισαγωγή ενός προσομοιωτή με στόχο την πραγματοποίηση αυτοματοποιημένων διαδικασιών οι οποίες εκτελούνται σε ένα IDS και συγκεκριμένα την παραγωγή συνόλων δεδομένων.

- Βελτίωση υπαρκτών πραγματικών συστημάτων ανίχνευσης εισβολών μέσα από τη δυνατότητας αξιολόγησης των συνόλων δεδομένων που θα παράγονται από τον προσομοιωτή που αναφέραμε παραπάνω.
- Ανοίγονται νέα δεδομένα και προοπτικές για ανάπτυξη ερευνητικά γύρω από το πεδίο των IDS με βάση το νέο παραγόμενο dataset

Κεφάλαιο 2ο

Κατηγοριοποίηση Συνόλων Δεδομένων

Ως σύνολο δεδομένων ονομάζουμε μία οντότητα (π.χ αρχείο) η οποία παρέχει με δομημένο και οργανωμένο τρόπο ένα ή περισσότερα αποτελέσματα. Στόχος του κάθε συνόλου αποτελεί η ιεράρχηση των αποτελεσμάτων που είναι αποθηκευμένα αλλά και η ταξινόμηση αυτών. Ουσιαστικά, μία βασική δυνατότητα που προσφέρουν τα σύνολα αυτά είναι ότι βοηθούν στην ανάλυση και εξαγωγή χρήσιμων συμπερασμάτων καθώς κάθε στοιχείο πληροφορίας φέρει κάποιες ιδιότητες και χαρακτηριστικά που τα διαφέρει από τα υπόλοιπα. Στην περίπτωση των datasets για συστήματα ανίχνευσης παρεισφρήσεων αυτό που επιτυγχάνουμε είναι να αποθηκεύουμε τις καταγραφές που παράγονται από ένα δικτυακό σύστημα ανίχνευσης εισβολών (NIDS) σε ομαδοποιημένη πληροφορία στην οποία περιλαμβάνονται δεδομένα της κίνησης και των αποτυπωμάτων αυτής που καταγράφηκε κάθε φορά. Με αυτόν τον τρόπο είμαστε σε θέση να αντλήσουμε πληροφορίες σχετικά με τη δυναμική και την ποιότητα του κάθε NIDS αφού σε μεταγενέστερο χρόνο θα έχουμε τη δυνατότητα σύγκρισης των στοιχείων ενός δεδομένου dataset με το δικό μας.

Συνεπώς, θα μπορούμε να αξιολόγησουμε έτσι το δικτυακό σύστημα μας και να μπορούμε να καταλάβουμε ποιές επιθέσεις ή απειλές μπορεί να ανιχνεύσει και ποιές όχι. Για παράδειγμα, από ένα τέτοιο σύνολο δεδομένων μπορούμε εξετάσουμε τη συνολική δικτυακή συμπεριφορά, το βαθμό και ταχύτητα ανταπόκρισης του παραλήπτη σε κάθε αποτύπωμα κίνησης, τον πλεονασμό, την πολυπλοκότητα μίας επίθεσης ή ακόμα να μάθουμε και για την αρχιτεκτονική και γεωγραφική τοποθεσία ενός εισβολέα .

Κάθε σύνολο δεδομένων παραγόμενο από ένα δικτυακό σύστημα ανίχνευσης εισβολών μπορούμε να το ταξινομήσουμε σε διαφορετικές κατηγορίες ανάλογα κάθε φορά με το βασικό σκοπό που θέλουμε να μας εξυπηρετεί. Τα βασικότερα είδη αυτών είναι τα εξής:

Feature-based dataset: Σε αυτήν την κατηγορία, ορίζονται από την αρχή κάποια χαρακτηριστικά (attributes) τα οποία αφορούν κάθε τμήμα πληροφορίας που πρόκειται να καταγραφεί από το NIDS. Αυτό βοηθάει σημαντικά στη μετέπειτα αξιολόγηση του συστήματος

ανίχνευσης αφού θα είμαστε σε θέση να διαχωρίσουμε κάθε φορά το είδος της πληροφορίας που πρόκειται να αναλυθεί χωρίς να προβαίνουμε σε συνδυαστικές μεθόδους όλων των δεδομένων ενός dataset για άντληση πληροφορίας. Για παράδειγμα, η ένδειξη πιθανής ανωμαλίας ή φυσιολογικής κατάστασης σε κάθε αποτύπωμα μας βοηθάει σημαντικά ώστε να μη χρειάζεται να επεξεργαστούμε όλα τα προηγούμενα χαρακτηριστικά. Έτσι εύκολα μπορούμε να διαπιστώσουμε αν μία ομάδα πακέτων δεδομένων αποτελεί κακόβουλη κίνηση. Σημαντικό δηλαδή στοιχείο εδώ είναι ότι η εξαγωγή κάθε αποτελέσματος ανά χαρακτηριστικό γίνεται με αυτοματοποιημένο τρόπο.

Άρα, δεν απαιτείται η παρέμβαση του ανθρώπινου παράγοντα για εξακρίβωση μίας κατάστασης. Τέτοιου είδους datasets συνήθως βρίσκονται σε μορφή CSV δηλαδή ένα αρχείο κειμένου όπου κάθε πληροφορία χαρακτηριστικού διαχωρίζεται με κόμμα όπως στο KDDCUP99 ή στο NSL-KDDtrain dataset τα οποία περιγράφουμε στο επόμενο κεφάλαιο.

Software-generated dataset: Αυτού του είδους σύνολα δεδομένων παράγονται κάθε φορά από έναν δικτυακό πρόγραμμα ανίχνευσης και αποτύπωσης πακέτων και ουσιαστικά προσφέρουν αποθηκευμένη όλη την κίνηση που καταγράφηκε. Τα αποτυπώματα αυτά περιέχουν αποκλειστικά πληροφορίες όπως αυτές εμφανίζονταν στο χρήστη κατά την καταγραφή τους. Ακόμα, τα datasets της κατηγορίας αυτής προσφέρουν δυνατότητα για επεξεργασία ή ανάλυση των δεδομένων κίνησης με σκοπό την εξαγωγή ενός τελικού συμπεράσματος περί πιθανής κακόβουλης συμπεριφοράς. Γενικά αυτά τα σύνολα, παρέχουν όλες τις πληροφορίες που αφορούν τα εκάστοτε πακέτα μαζί με το hexdump και έτσι ο χρήστης μπορεί εξετάζοντας αυτά τα πακέτα να παρατηρήσει αν υπήρξε μία πιθανή εισβολή. Παραδείγματα λογισμικού που υποστηρίζουν τέτοιου είδους συνόλων δεδομένων, αποτελούν εργαλεία όπως το Ethereal/Wireshark τα οποία παράγουν σε αποθηκευμένη μορφή pcap αρχεία που περιέχουν αναλυτικά όλη την κίνηση που αποτυπώθηκε ανά ξεχωριστό frame.

Table-based dataset: Αποτελεί ένα σύνολο δεδομένων όπως και στην πρώτη κατηγορία, δηλαδή βασισμένο σε ύπαρξη κάποιων χαρακτηριστικών τα οποία όμως αυτή τη φορά είναι δομημένα αποκλειστικά σε μορφή πίνακα δηλαδή καθένα διαχωρίζεται σε στήλες και κάθε αποθηκευμένο αποτύπωμα που εξάγεται από το σύστημα ανίχνευσης τοποθετείται σε μία γραμμή του πίνακα. Το πλεονέκτημα της χρήσης πινάκων για την αποθήκευση ενός dataset είναι ότι μας βοηθάει μετέπειτα να πραγματοποιήσουμε στατιστικές αναλύσεις και γενικά

μπορούμε να ταξινομούμε καθε φορά τον πίνακα κατά τον τρόπο και τη σειρά εμφάνισης που επιθυμούμε εμείς ανάλογα με τη σημαντικότητα του κάθε στοιχείου πληροφορίας.
Για παράδειγμα, μπορούμε να εμφανίζουμε κατά αύξουσα ή φθίνουσα σειρά τον αριθμό συνολικών συνδέσεων σε έναν συγκεκριμένο προορισμό. Έτσι μπορούμε να διαπιστώσουμε εύκολα που και αν υπήρχε προσπάθεια για εκτέλεση επίθεσης άρνησης υπηρεσιών. Προγράμματα τα οποία προσφέρουν δυνατότητα για προβολή table-based datasets είναι το Microsoft Excel ή και το SPSS.

Γενικά πραγματοποιώντας μία σύγκριση μεταξύ των παραπάνω ειδών συνόλων δεδομένων, παρατηρούμε ότι τα feature-based datasets παρέχουν μεγαλύτερη ευκολία ανάλυσης της πληροφορίας στο χρήση αφού εισάγοντας το εκάστοτε dataset σε ένα εργαλείο εξόρυξης δεδομένων, μπορούμε σχεδόν άμεσα να έχουμε αποτελέσματα παρά την αυξημένη πολυπλοκότητα που χαρακτηρίζεται σε αυτά. Παράλληλα, η δυνατότητα εύρεσης false positives αλλά και ανωμαλιών είναι ευκολότερη στα feature-based datasets αφού εκεί μέσω της χρήσης κατάλληλων classifiers μπορούμε να παράγουμε πειραματικούς κανόνες οι οποίοι να υποδεικνύουν τις προυποθέσεις που πρέπει να τηρούνται ώστε το κάθε αποτύπωμα κίνησης να μην θεωρείται κακόβουλο.

Από την άλλη μεριά στα software-generated datasets παρατηρείται ποικιλομορφία όσον αφορά το είδος αρχείου καταγραφής κίνησης. Κάτι τέτοιο σημαίνει ότι υπάρχουν αρχεία τα οποία είναι σε θέση να περιγράψουν καλύτερα και πληρέστερα δεδομένα κίνησης σε σχέση με τα feature-based ή table-based datasets. Παρόλα αυτά, και σε αυτήν την κατηγορία συνόλων έχουμε δυνατότητα για εκτέλεση αυτοματοποιημένων διαδικασίων αφού σε εργαλεία ανίχνευσης εισβολών όπως το Snort, μπορούμε να εισάγουμε το εκάστοτε αρχείο κίνησης και να διαβαστούν τα πακέτα δεδομένων που ανταλλάχθηκαν με στόχο τον εντοπισμό εισβολών. Ωστόσο, αν το dataset αυτό αποτελείται από πολλαπλά αρχεία κίνησης και όχι ένα εννιαίο τότε μειώνεται η ευχρηστία και αυξάνεται η πολυπλοκότητα και ο πλεονασμός. Επιπλέον, στην περίπτωση αυτή δε θα είναι δυνατή ούτε η πραγματοποίηση ταξινόμησης των δεδομένων άρα δεν θα μπορεί να εξαχθεί ένα γενικό πόρισμα ως προς το χαρακτηριστικό κίνησης που επιθυμούμε να εξετάσουμε.

Στη συνέχεια στα table-based datasets παρατηρούμε ότι και εδώ είναι δυνατή η εμφάνιση πλεονασμού, δηλαδή ενδεχομένως να επαναλαμβάνονται πληροφορίες οι οποίες να αφορούν τα ίδια πακέτα κίνησης ξανά και ξανά. Ωστόσο εδώ η διαδικασία ταξινόμησης είναι ευκολότερη αφού τα δεδομένα χωρίζονται σε γραμμές και στήλες. Παρόλα αυτά η έλλειψη

αυτοματοποιημένων εργαλείων για σάρωση του περιεχόμενου τέτοιων datasets δημιουργεί δυσκολίες στην εύρεση απειλών οι οποίες εκτυπώθηκαν και συνεπώς αυξάνεται η πολυπλοκότητα.

Πίνακας 1: Σύγκριση μεταξύ διαφορετικών ειδών datasets για IDS

	Feature-based dataset	Software-generated dataset	Table-based dataset
Πολυπλοκότητα	ΝΑΙ	ΝΑΙ	ΟΧΙ
Αυτοματισμός	ΝΑΙ	ΝΑΙ	ΟΧΙ
Πλεονασμός	ΝΑΙ	ΝΑΙ	ΝΑΙ
Ταξινόμηση	ΝΑΙ	ΟΧΙ	ΝΑΙ
Εύρεση ανωμαλιών	ΕΥΚΟΛΗ	ΜΕΤΡΙΑ	ΔΥΣΚΟΛΗ

Κεφάλαιο 3ο

Περιγραφή Υπαρχόντων Συνόλων Δεδομένων

Αφού σε αρχικό στάδιο καταφέρουμε να εγκαθιδρύσουμε ένα δικτυακό σύστημα ανίχνευσης μέσα από τη χρήση των κατάλληλων πακέτων λογισμικού , το επόμενο βήμα είναι να θέσουμε σε λειτουργία πολλαπλά δίκτυα τα οποία θα είναι σε θέση να επικοινωνούν μεταξύ τους και να παράγουν κίνηση. Έπειτα, το σύστημα ανίχνευσης θα συλλέξει και θα καταγράψει με οργανωμένο τρόπο αυτήν την κίνηση η οποία πέρα από τη φυσιολογική κίνηση θα πρέπει να περιέχει και την κακόβουλη κίνηση. Έτσι, κάθε αποτύπωμα της δικτυακής κίνησης η οποία ανιχνεύεται, αποθηκεύεται και το σύνολο όλων των αποτυπωμάτων που αφορούν δύο ή περισσότερα δίκτυα Η/Υ αποτελεί το νέο NIDS dataset μέσω του οποίου σε μεταγενέστερο χρόνο θα έχουμε τη δυνατότητα να μελετήσουμε και να αξιολογήσουμε το σύστημα ανίχνευσης μας.

Ουσιαστικά με την εκτενή συλλογή ενός συνόλου δεδομένων, είμαστε σε θέση να αντλήσουμε πληροφορίες σχετικά με το πόσο καλά ανταποκρίθηκε το σύστημα μας σε διαφόρων ειδών επιθέσεις. Για την ανίχνευση της κίνησης σε καθένα από τα datasets που περιγράφουμε, χρησιμοποιήσαμε το Bro και Snort IDS (ενεργοποίηση κανόνων από προεπιλογή δηλαδή χωρίς να θέσουμε σε λειτουργία επιπρόσθετους κανόνες).

3.1 KDD Cup 99

Σχεδόν από το 1999, το KDD CUP 99 αποτελεί το πιο διαδεδομένο σύνολο δεδομένων τόσο σε δημοφιλότητα όσο και σε χρήση με στόχο την αξιολόγηση δικτυακών μεθόδων ανίχνευσης ανωμαλιών. Το συγκεκριμένο data set είναι βασισμένο στα δεδομένα τα οποία αποτυπώθηκαν από το DARPA 98 dataset κατά την αξιολόγηση του. Το training dataset του KDD αποτελείται από σχεδόν 4,900,000 αποτυπώματα κίνησης όπου το καθένα αποτελείται από 41 χαρακτηριστικά και προσδιορίζεται από ένα συγκεκριμένο είδος επίθεσης (όταν ανιχνευτεί επίθεση) ή από μία φυσιολογική συμπεριφορά. Το testing dataset περιέχει περίπου 300,000 δείγματα με ένα συνολικό αριθμό 24 διαφορετικών ειδών επιθέσεων από τα οποία τα 14 είδη εμφανίζονται μόνο στο σύνολο αυτό. [1] [2]

Παρακάτω στον πίνακα 2 φαίνονται τα είδη των πρωτοκόλλων τα οποία χρησιμοποιήθηκαν εκείνη τη στιγμή τόσο για φυσιολογική κίνηση όσο και για κακόβουλη. [3] Επίσης, από τη μελέτη του αρχείου kddcup.data παρατηρούμε και τις καταστάσεις που βρέθηκαν εκείνη τη στιγμή οι δεδομένες δικτυακές συνδέσεις για τα τρέχοντα πρωτόκολλα. Δηλαδή, έτσι μπορούμε να γνωρίζουμε πόσες φορές τερματίστηκε μία σύνδεση, πόσο συχνά ένας χρήστης εκτελούσε αιτήματα, πότε ολοκληρώθηκε επιτυχώς ένα αίτημα κ.α.

Με αυτά τα στοιχεία μπορούμε να εξάγουμε συμπεράσματα σχετικά με το πόσο συχνά πραγματοποιήθηκαν π.χ επιθέσεις άρνησης υπηρεσιών ή πόσες συνολικά προσπάθειες τύπου privilege escalation πραγματοποιήθηκαν. [4]

Πίνακας 2: Σύγκριση δικτυακών αποτυπωμάτων που παρατηρήθηκαν στο KDD Cup 99

Δικτυακό αποτύπωμα	**Συνολικές εμφανίσεις**	**Δικτυακό αποτύπωμα**	**Συνολικές εμφανίσεις**
icmp	2833545	ecr_i	2811660
tcp	1870597	private	1100831
udp	194288	http	623090
smtp	96554	SF	3744327
domain_u	57782	S0	869829
REJ	268874	RSTR	8094
RST0	5344	SH	1040
normal	972780		

Πίνακας 3: Αριθμητική απεικόνιση επιθέσεων στο KDD Cup 99

Είδος επίθεσης	**Συνολικές εμφανίσεις**	**Είδος επίθεσης**	**Συνολικές εμφανίσεις**
neptune	2807886	portsweep	10413
satan	15892	smurf	164051
ipsweep	12481	άλλο	6961

Μελετώντας προσεκτικά ξανά το αρχείο kddcup.data αντλήσαμε τις παρακάτω στατιστικές πληροφορίες (πίνακας 4) οι οποίες είναι σημαντικές δεδομένου ότι έχουμε μία γενική εικόνα για τον αριθμό των επιθέσεων που πραγματοποιήθηκαν και με ποιούς τρόπους προσπαθούσαν οι κακόβουλοι χρήστες να εισβάλλουν στο εκάστοτε σύστημα κάθε φορά.

Πίνακας 4: Χαρακτηριστικά KDD Cup 99

Χαρακτηριστικό	Ελάχιστη τιμή	Μέση τιμή	Μέγιστη τιμή
duration	0.00	48.34	58329.00
src_bytes	0.000e+00	1.835e+03	1.380e+09
dst_bytes	0.000e+00	1.094e+03	1.310e+09
land	0.0e+00	5.7e-06	1.0e+00
wrong_fragment	0.0000000	0.0006488	3.0000000
urgent	0.0e+00	8.0e-06	1.4e+01
hot	0.0000	0.01244	77.0000
num_failed_logins	0.0e+00	3.2e-05	5.0e+00
logged_in	0.0000	0.1435	1.0000
num_compromised	0.000	0.008	7479.000
root_shell	0.00e+00	6.82e-05	1.00e+00
su_attempted	0.00e+00	3.67e-05	2.00e+00
num_root	0.000	0.013	7468.000
num_file_creations	0.00000	0.00119	43.00000
num_shells	0.00e+00	7.43e-05	2.00e+00
num_access_files	0.000000	0.001021	9.000000
num_outbound_cmds	0	0	0
is_host_login	0e+00	4e-07	1e+00
is_guest_login	0.0000000	0.0008352	1.0000000
count	0	335	511
srv_count	0.0	295.3	511.00
serror_rate	0.000	0.178	1.000
srv_serror_rate	0.000	0.178	1.000
rerror_rate	0.00000	0.05767	1.00000
srv_rerror_rate	0.00000	0.05773	1.00000
same_srv_rate	0.0000	0.7899	1.0000
diff_srv_rate	0.00000	0.02118	1.00000
srv_diff_host_rate	0.00000	0.02826	1.00000
dst_host_count	0	233	255
dst_host_srv_count	0.0	189.2	255.0
dst_host_diff_srv_rate	0.00000	0.03071	1.00000
dst_host_srv_diff_host_rate	0.00000	0.006464	1.00000
dst_host_serror_rate	0.0000	0.1781	1.0000
dst_host_srv_serror_rate	0.0000	0.1779	1.0000
dst_host_rerror_rate	0.00000	0.05793	1.00000
dst_host_srv_rerror_rate	0.00000	0.05766	1.00000

3.2 NSL-KDD Train

Όπως διαπιστώσαμε το KDDCUP'99 αποτελεί το πιο γνωστό και ευρέως χρησιμοποιημένο σύνολο δεδομένων για ανίχνευση ανωμαλιών. Ωστόσο ερευνητές οι οποίοι πραγματοποίησαν στατιστικές αναλύσεις στα δεδομένα αυτά, ανακάλυψαν δύο σημαντικά κενά τα οποία επηρεάζουν αρνητικά την επίδοση των συστημάτων ανίχνευσης που καλούνται να αξιολογηθούν μέσω του KKDCUP'99. Τα αποτελέσματα έδειξαν ότι δεν μπορούμε να εφαρμόσουμε προσεγγίσεις του dataset αυτού με στόχο τον έλεγχο αποδοτικότητας σχετικά με την ανίχνευση ανωμαλιών. Για την επίλυση αυτών των θεμάτων, προτάθηκε το NSL-KDD dataset το οποίο αποτελείται από επιλεκτικές καταχωρήσεις του συνόλου KDD Cup 99.

Το NSL-KDD γενικά αποτελεί ένα σύνολο δεδομένων το οποίο προτάθηκε προκειμένου να λύσει και ταυτόχρονα να βελτιώσει μερικά από τα εμφανή προβλήματα του KDD Cup 99 συνόλου τα οποία έχουν αναφερθεί πλήρως στο [5]. Παρόλο που η τρέχουσα έκδοση που περιγράφουμε έχει κάποιες από τις αδυναμίες που έχουν αναλυθεί στο παρελθόν και ενδεχομένως να μην αποτελεί το τελειότερο παράδειγμα για αναπαράσταση σε πραγματικά υπάρχοντα δίκτυα, δεδομένου της σημαντικής έλλειψης διαθέσιμων δημοσίων συνόλων δεδομένων για NIDS, το NSL-KDD Train είναι σε θέση να παρέχει αποτελεσματικές λύσεις για ερευνητές οι οποίοι επιδιώκουν να συγκρίνουν διαφορετικά συστήματα ανίχνευσης. Επιπλέον, ο αριθμός των αποτυπωμάτων που καταγράφηκαν στο NSL-KDD Train συγκριτικά με τον αντίστοιχο αριθμό καταγραφών που πραγματοποιήθηκαν στο testing dataset είναι φυσιολογικός. Αυτό το πλεονέκτημα καθιστά εφικτό σε οποιονδήποτε, να εκτελέσει πειράματα σε ένα πλήρες σύνολο χωρίς να υπάρχει η ανάγκη για τυχαία επιλογή ενός συγκεκριμένου μικρού μέρους. Συνεπώς, τα αποτελέσματα αξιολόγησης σε διαφορετικές ερευνητικές εργασίες θα είναι συνεπή και συγκρίσιμα.

Τα χαρακτηριστικά του NSL-KDD μπορούν να ταξινομηθούν σε τρείς ομάδες:

Α) Κύρια χαρακτηριστικά

Αυτή η κατηγορία περιλαμβάνει όλες τις ιδιότητες εκείνες οι οποίες μπορούν να εξαχθούν από μία TCP/IP σύνδεση. Η πλειοψηφία των χαρακτηριστικών αυτών οδηγούν σε μία μικρή καθυστέρηση κατά την ανίχνευση.

Β) Χαρακτηριστικά Περιεχομένου

Σε αντίθεση με τις περισσότερες επιθέσεις άρνησης υπηρεσιών (DoS) αλλά και επιθέσεις σάρωσης, οι Remote-to-Local (R2L) και οι User-to-Root (U2R) επιθέσεις δεν παρουσιάζουν συχνή εμφάνιση διαδοχικών μοτίβων. Αυτό οφείλεται στο γεγονός ότι οι DoS καθώς και οι επιθέσεις σάρωσης περιλαμβάνουν πολλές συνδέσεις ταυτόχρονα σε hosts σε πολύ μικρή χρονική περίοδο. Ωστόσο, οι R2L και U2R επιθέσεις ενσωματώνονται στα τμήματα δεδομένων των πακέτων και φυσιολογικά περιλαμβάνεται μόνο μία σύνδεση τη φορά. Για την ανίχνευση αυτού του είδους επιθέσεων, χρειαζόμαστε μερικά χαρακτηριστικά τα οποία θα είναι να ικανά να αναζητούν για κακόβουλη συμπεριφορά στο τμήμα δεδομένων π.χ όπως ο αριθμός αποτυχημένων προσπαθειών για σύνδεση. Αυτά τα χαρακτηριστικά ορίζονται ως χαρακτηριστικά περιεχομένου.

Γ) Χαρακτηριστικά κίνησης

Αυτή η κατηγορία περιλαμβάνει χαρακτηριστικά τα οποία λαμβάνουν υπόψιν διάφορες καθυστερήσεις που συμβαίνουν και χωρίζεται σε δύο ομάδες:

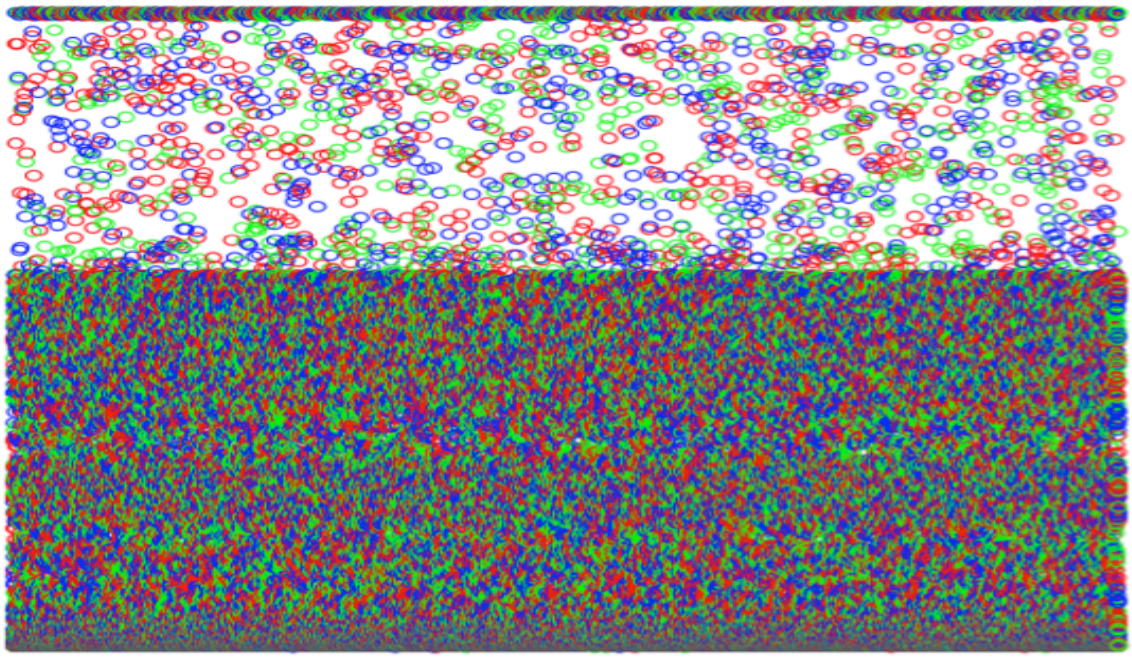
1. 'Same host' χαρακτηριστικά: εξετάζουν μόνο τις συνδέσεις τα τελευταία δύο δευτερόλεπτα οι οποίες έχουν την ίδια διεύθυνση προορισμού όπως η τρέχουσα σύνδεση και υπολογίζουν στατιστικά δεδομένα σχετικά με τη συμπεριφορά του εκάστοτε πρωτοκόλλου, υπηρεσίας κ.α.

2. 'Same service' χαρακτηριστικά: εξετάζουν μόνο τις συνδέσεις που συνέβησαν τα προηγούμενα δύο δευτερόλεπτα και οι οποίες χρησιμοποιούν την ίδια υπηρεσία όπως και η τρέχουσα σύνδεση.

Τα δύο αυτά είδη χαρακτηριστικών, αποκαλούνται και ως time-based. Ωστόσο, υπάρχουν επιθέσεις τύπου probing στις οποίες γίνεται σάρωση των πληροφοριών των hosts ή των θυρών τους χρησιμοποιώντας πολύ μεγαλύτερη καθυστέρηση από δύο δευτερόλεπτα μόνο (π.χ μία επίθεση ανά ένα λεπτό). Συνεπώς, αυτές οι επιθέσεις δεν παράγουν μοτίβα ανίχνευσης σε σχέση με αυτές που η χρονική απόσταση είναι δύο δευτερόλεπτα. Για την επίλυση αυτού του προβλήματος, τα χαρακτηριστικά 'same host' και 'same service' ξαναυπολογίζονται αλλά αυτή τη φορά βασίζονται στη χρονική διάρκεια 100 συνδέσεων και όχι σε μία χρονική απόσταση 2

δευτερολέπτων. Αυτά τα χαρακτηριστικά ορίζονται και ως connection-based χαρακτηριστικά κίνησης.

Το NSL-KDD train βρίσκεται δημόσια διαθέσιμο (http://nsl.cs.unb.ca/NSL-KDD/KDDTrain+.arff) και όπως φαίνεται και από τις καταγραφές για πιθανές εισβολές που έγιναν, τα χαρακτηριστικά που χρησιμοποιήθηκαν και λήφθηκαν υπόψιν για την εξαγωγή του συνόλου δεδομένων αναφέρονται στον πίνακα 5 στον οποίο παρουσιάζονται οι ελάχιστες, μέσες και μέγιστες τιμές που εμφανίστηκαν για κάθε χαρακτηριστικό. Αυτό μας διευκολύνει να κατανοήσουμε ποιές προυποθέσεις απαιτούνταν για την εκκίνηση μίας επίθεσης και γενικά ποιά είδη επιθέσεων εμφανίστηκαν κατά κύριο λόγο στο dataset αυτό. Στην εικόνα 1 παρουσιάζεται ένα scatter plot το οποίο παρήχθησε μέσω του εργαλείου R Console (αφού πρώτα φορτώσαμε το υπάρχον σύνολο δεδομένων) και ενσωματώνει όλες τις τιμές των χαρακτηριστικών που εμφανίστηκαν σε ένα εννιαίο διάγραμμα. Σε αυτό παρατηρούμε ότι η πλειψηφία των τιμών κυμάνθηκε κατά κύριο λόγο από 0 έως λίγο παραπάνω από το σταθμικό διάμεσο των συνολικών μέγιστων τιμών που εμφανίστηκαν.



Εικόνα 1: Scatter plot παραγόμενο μέσω της R Console

Οι παρακάτω πίνακες παρουσιάζουν αναλυτικά στοιχεία που προκύπτουν από τη μελέτη του αρχείου KDDTrain+.arff.

Πίνακας 5: Χαρακτηριστικά NSL-KDD Train

Χαρακτηριστικό	Ελάχιστη τιμή	Μέση τιμή	Μέγιστη τιμή
duration	0.0	287.1	42908.0
src_bytes	0.000e+00	4.557e+04	1.380e+09
dst_bytes	0.000e+00	1.978e+04	1.310e+09
land	0.0000000	0.0001985	1.0000000
wrong_fragment	0.00000	0.02269	3.00000
urgent	0.0000000	0.0001111	3.0000000
hot	0.0000	0.2044	77.0000
num_failed_logins	0.000000	0.001222	5.000000
logged_in	0.0000	0.3957	1.0000
num_compromised	0.000	0.279	7479.000
root_shell	0.000000	0.001342	1.000000
su_attempted	0.000000	0.001103	2.000000
num_root	0.000	0.302	7468.000
num_file_creations	0.00000	0.01267	43.00000
num_shells	0.0000000	0.0004128	2.0000000
num_access_files	0.000000	0.004096	9.000000
num_outbound_cmds	0	0	0
is_host_login	0.0e+00	7.9e-06	1.0e+00
is_guest_login	0.000000	0.009423	1.000000
count	0.00	84.11	511.00
srv_count	0.00	27.74	511.00
serror_rate	0.0000	0.2845	1.0000
srv_serror_rate	0.0000	0.2825	1.0000
rerror_rate	0.00	0.12	1.00
srv_rerror_rate	0.0000	0.1212	1.0000
same_srv_rate	0.0000	0.6609	1.0000
diff_srv_rate	0.00000	0.06305	1.00000
srv_diff_host_rate	0.00000	0.09732	1.00000
dst_host_count	0.0	182.1	255.0
dst_host_srv_count	0.0	115.7	255.0
dst_host_same_srv_count	0.0000	0.5212	1.0000
dst_host_diff_srv_rate	0.00000	0.08295	1.00000
dst_host_same_src_port_rate	0.0000	0.1484	1.0000
dst_host_srv_diff_host_rate	0.00000	0.03254	1.00000
dst_host_serror_rate	0.0000	0.2845	1.0000
dst_host_srv_serror_rate	0.0000	0.2785	1.0000
dst_host_rerror_rate	0.0000	0.1188	1.0000
dst_host_srv_rerror_rate	0.0000	0.1202	1.0000

Πίνακας 6: Δικτυακά Αποτυπώματα NSL-KDD Train

Δικτυακό αποτύπωμα	Συνολικές εμφανίσεις	Δικτυακό αποτύπωμα	Συνολικές εμφανίσεις
icmp	8291	private	21853
tcp	102688	domain_u	9043
udp	14993	smtp	7313
http	40338	ftp_data	6859
eco_i	4586	SF	74944
S0	34851	RSTR	2421
REJ	11233	RST0	1562
S1	365	anomaly	58630
normal	67342		

Οι πίνακες που βλέπουμε (πιν. 6 και 7) αναλύουν στοιχεία που προκύπτουν από τη μελέτη του αρχείου KDDTrain+.txt τα οποία δεν έχουν καταγραφεί στο αρχείο KDDTrain+.arff. Γενικά, παρουσιάζονται αριθμητικές πληροφορίες σχετικά με τις υπηρεσίες, τα πρωτόκολλα που χρησιμοποιήθηκαν αλλά και τα χαρακτηριστικά του επιτιθέμενου που αποτυπώθηκαν.

Πίνακας 7: Αριθητική απεικόνιση επιθέσεων που καταγράφηκαν στο NSL-KDD Train

Είδος επίθεσης	Συνολικές εμφανίσεις	Είδος επίθεσης	Συνολικές εμφανίσεις
neptune	41214	portsweep	2931
satan	3633	smurf	2646
ipsweep	3599	άλλο	4607

3.3 DARPA datasets

3.3.1 DARPA 1998

Για τη μελέτη του DARPA98 dataset εκτελέσαμε ανίχνευση ενός συγκεκριμένου δείγματος αρχείου (outside.tcpdump) ικανού να μας παρέχει πληροφορίες για το συνολικό dataset. Έτσι είμαστε σε θέση να παρατηρήσουμε χαρακτηριστικά, ιδιότητες αλλά και τη

χρησιμότητα του. Στον παρακάτω πίνακα παρουσιάζονται συνοπτικές πληροφορίες σχετικά με τα εκάστοτε πρωτόκολλα στα οποία έγιναν συνδέσεις αλλά και τα alerts που παρήχθησαν από το Snort.

Πίνακας 8: Πληροφορίες πρωτοκόλλων για το DAPRA 98 dataset

Πληροφορία	**Αριθμός εμφανίσεων**	**Πληροφορία**	**Αριθμός εμφανίσεων**
Alerts	5034	UDP sessions deleted	973
TCP συνεδρίες	36834	Events	180
UDP συνεδρίες	973	POST αιτήματα	0
TCP timeouts	4	GET αιτήματα	32822
TCP segments queued	214028	SSL packets decoded	315
TCP segments released	214028	SMTP συνεδρίες	2236
TCP rebuilt packets	108173	HTTP response cookied	2586
TCP segments used	212107	HTTP request cookies	0
TCP discards	233	HTTP επικεφαλίδες	32822
TCP gaps	3226	ARP πακέτα	2906
UDP sessions created	973	TCP πακέτα	531454

Στη συνέχεια για το αρχείο αυτό χρησιμοποιήσαμε τόσο το Snort όσο και το Bro για να εντοπίσουμε τις απειλές που έχουν καταγράφει από το σύστημα ανίχνευσης. Στον ακόλουθο πίνακα εμφανίζονται οι παραβιάσεις κανόνων που έγιναν οι οποίες αποθηκεύτηκαν στα αρχεία alert.log (για το Snort), weird.log, ssh.log και notice.log (για το Bro).

Πίνακας 9: Δείγματα συναγερμών ασφαλείας που εκτυπώθηκαν

Alerts του SNORT	**Alerts του BRO**
SENSITIVE-DATA Email Addresses	unmatched_HTTP_reply
(http_inspect) NO CONTENT-LENGTH OR TRANSFER ENCODING IN HTTP RESPONSE	HTTP_version_mismatch
(http_inspect) INVALID CONTENT-LENGTH OR CHUNK SIZE	line_terminated_with_single_CR
(spp_sdf) SDF Combination Alert	unescpaped_special_URI_char
	unknown protocol_9
	unknown_packet_type
	success_inbound SSH-1.5-1.2.2.2

Παρακάτω εμφανίζονται οι συνολικοί εισβολείς που έδρασαν στο σύστημα ταξινομημένοι ως προς τις IP διευθύνσεις όπως εντοπίστηκαν μέσω του Bro.

Πίνακας 10: Διευθύνσεις IP εισβολέων-επιτιθέμενων

172.16.115.234	172.16.115.5
172.16.114.168	172.16.117.111
172.16.116.201	172.16.117.132
172.16.112.207	172.16.114.207
172.16.116.44	172.16.117.103
172.16.115.87	172.16.113.84
172.16.113.204	172.16.114.148
172.16.114.169	172.16.117.52

Πιο συχνή απειλή αποτελούσε ο επιτιθέμενος με διεύθυνση IP 172.16.117.132.

3.3.2 DARPA 1999

Ο πρωταρχικός στόχος κατά τη δημιουργία αυτού του συνόλου δεδομένων ήταν η μέτρηση της ικανότητας ενός συστήματος ανίχνευσης παρεισφρήσεων με στόχο να εξεταστεί αν είναι σε θέση να εντοπίσει καινοφανείς επιθέσεις που δεν είχαν ξαναπραγματοποιηθεί. Αυτό αποτελούσε ένα από τα μεγαλύτερα προβλήματα κατά την αξιολόγηση που έγινε μέσω του DARPA 98. Συνεπώς, στο τεχνικό κομμάτι προστέθηκαν συστήματα Windows NT στο δίκτυο προσομοίωσης και εκτελέστηκαν 17 νέες επιθέσεις οι οποίες εισήχθησαν στη ροή της κίνησης με στόχο τα λειτουργικά συστήματα αυτά. [5] Μία άλλη σημαντική αλλαγή, αποτελούσε η προσθήκη επιθέσεων από το εσωτερικό δίκτυο. Τα αποτυπώματα που εντοπίστηκαν (τα οποία δεν είχαν χρησιμοποιηθεί κατά την αξιολόγηση που έγινε το 1998) χρησιμοποιήθηκαν όπως και τα δεδομένα ελέγχου (testing data) για ανίχνευση τέτοιων εσωτερικών επιθέσεων. [6]

Επιπλέον, η αξιολόγηση των NIDS μέσω αυτής της διαδικασίας είναι πιο περιεκτική από τη στιγμή που τόσο η ανίχνευση επιθέσεων όσο και η αναγνώριση τους αξιολογούνται κατά τον ίδιο τρόπο. Έτσι, τα αποτελέσματα που παρήχθησαν, βοήθησαν στην εξαγωγή αξιόπιστων συμπερασμάτων. Παρακάτω εμφανίζονται μερικά από τα παραγόμενα security alarms που εμφανίζονται σε αρχεία που αφορούν την τέταρτη εβδομάδα καταγραφής κίνησης στην οποία πραγματοποιήθηκε η πλειοψηφία των επιθέσεων που παρατηρήθηκε.

Πίνακας 11: Λίστα απειλών ανά αρχείο καταγραφής

outside.tcpdump	inside.tcpdump
SERVER-MAIL Metamail header length exploit attempt	SERVER-WEBAPP /cgi-bin /access
SNMP request udp	(spp_ssh) Protocol mismatch
SNMP Public accessudp	INDICATOR-COMPROMISE 403 Forbidden
DNS dns response for rfc1918 192.168/16 address detected	DNS large number of NXDOMAIN replies - possible DNS cache poisoning
FILE-OFFICE Microsoft Office GIF image descriptor memory corruption attempt	TELNET login incorrect
SENSITIVE-DATA Email Addresses	(http_inspect) INVALID CONTENT-LENGTH OR CHUNK SIZE
SERVER-WEBAPP redirect access	PROTOCOL - ICMP Echo Ping
POLICY-SOCIAL IRC channel join	SERVER-WEBAPP perl.exe access
PROTOCOL-ICMP Destination Unreachable Port Unreachable	SERVER-IIS scripts-browse access
INDICATOR-COMPROMISE IRC message on non-standard port	(spp_sdf) SDF Combination Alert

```
[**] [120:3:1] (http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE [**]
[Classification: Unknown Traffic] [Priority: 3]
03/30-15:04:58.520250 00:60:97:DE:54:36 -> 00:10:7B:38:46:32 type:0x800 len:0xF7
209.1.112.251:80 -> 172.16.114.207:1171 TCP TTL:64 TOS:0x0 ID:197 IpLen:20 DgmLen:233 DF
***AP*** Seq: 0xED774B12  Ack: 0x60ED3048  Win: 0x7D78  TcpLen: 20

[**] [138:5:1] SENSITIVE-DATA Email Addresses [**]
[Classification: Sensitive Data was Transmitted Across the Network] [Priority: 2]
03/30-15:05:03.457182 00:60:97:DE:54:36 -> 00:10:7B:38:46:32 type:0x800 len:0x1F74
206.79.171.51:80 -> 172.16.114.207:1211 TCP TTL:63 TOS:0x0 ID:5134 IpLen:20 DgmLen:8038 DF
***A**** Seq: 0xFFD78C85  Ack: 0xBF1EFC4C  Win: 0x7AAC  TcpLen: 20
```

Εικόνα 2: Ενδεικτικό δείγμα snort alerts

3.3.3 DARPA 2000

Το σύνολο δεδομένων DARPA 2000 που υλοποιήθηκε είχε ως σκοπό την ανίχνευση πολύπλοκων επιθέσεων οι οποίες περιείχαν πολλαπλά βήματα. Δύο σενάρια επίθεσης προσομοιώθηκαν στην έκδοση του 2000 τα οποία ορίστηκαν από τους δημιουργούς ως LLDOS (Lincoln Laboratory Scenario (DDoS)) 1.0 και LLDOS 2.0. Και τα δύο σενάρια επιθέσεων διενεργήθηκαν μέσα από τη χρήση πολλαπλών δικτυακών και ελεγχόμενων συνεδριών.

Αυτά τα sessions ταξινομήθηκαν σε τέσσερις φάσεις:

1. Network Probing
2. Εισβολή στο σύστημα μέσα από εκμετάλλευση αδυναμιών
3. Εγκατάσταση DDoS λογισμικού σε σύστημα όπου έχει παραβιαστεί η ασφάλεια του
4. Εκτέλεση DDOS επίθεσης έναντι άλλου στόχου.

Το LLDOS 2.0 είναι διαφορετικό από το LLDOS 1.0 υπό την έννοια ότι οι επιθέσεις είναι σχεδόν αόρατες άρα και δυσκολότερο να ανιχνευθούν. Από τη στιγμή που το σύνολο δεδομένων αυτό περιέχει πολυσταδιακά σενάρια επιθέσεων, είναι φανερό ότι χρησιμοποιήθηκε κατά κύριο λόγο για έλεγχο συσχετισμού των παραγόμενων μηνυμάτων συναγερμών (alarms).

Πίνακας 12: Συχνότητα απειλών ανά αρχείο καταγραφής

Αρχείο DARPA2000	Συχνότερη απειλή	Επικινδυνότητα	Κατηγορία
inside.tcpdump	WEB-CLIENT Download of Powerpoint 95 file	ΥΨΗΛΗ	Attempted User Privilege Gain
outside.tcpdump	WEB-CLIENT Download of Powerpoint 95 file	ΥΨΗΛΗ	Attempted User Privilege Gain
LLS_DDOS_1.0-dmz.dump	RPC sadmind query with root credentials attempt UDP	ΜΕΤΡΙΑ	Misc attack
LLS_DDOS_1.0-inside.tcpdump	RPC sadmind query with root credentials attempt UDP	ΜΕΤΡΙΑ	Misc attack
LLS_DDOS_2.0.2-outside.tcpdump	FILE-IDENTIFY New Executable binary file magic	ΧΑΜΗΛΗ	Misc activity
LLS_DDOS_2.0.2-inside.tcpdump	RPC sadmind query with root credentials attempt UDP	ΜΕΤΡΙΑ	Misc attack

Σημαντικό είναι να αναφέρουμε ότι όλες οι επιθέσεις που αναγνώρισε το snort όσον αφορά την παραβίαση κανόνων προεπεξεργαστών στο DARPA 1999 ακριβώς οι ίδιες παραβιάσεις εμφανίζονται και εδώ. Παρακάτω παρουσιάζουμε στατιστικές πληροφορίες που αφορούν διεργασίες των προεπεξεργαστών (SMTP, SIP ή dcerpc2).

Πίνακας 13: Προβολή στατιστικών δεδομένων preprocessors

Αρχείο DARPA2000	Αριθμός πακέτων	Συνολικές συνεδρίες	Αγνοούμενα πακέτα	Μέγιστος αριθμός ταυτόχρονων συνεδριών
inside.tcpdump	3555	1611	1768	8
outside.tcpdump	198	1719	0	2
LLS_DDOS_1.0-dmz.dump	0	524	0	2
LLS_DDOS_1.0-inside.tcpdump	300	580	8532	4
LLS_DDOS_2.0.2-outside.tcpdump	0	260	0	3
LLS_DDOS_2.0.2-inside.tcpdump	203	307	5103	5

Στον επόμενο πίνακα αναφέρουμε συνολικά ποιές διευθύνσεις IP (ακολουθούμενη από έναν αριθμό θύρας στον οποίο συνδεόταν) καταγράφηκαν ανάλογα με την ενέργεια που εκτελούσε η καθεμία.

Πίνακας 14: Διευθύνσεις IP και δικτυακές θύρες σύνδεσης

Διευθύνσεις IP αμυνόμενου	Θύρες αμυνόμενου	Διευθύνσεις IP εισβολέα	Θύρες εισβολέα
172.16.114.168	19058	160.147.163.19	80
172.16.115.20	60254, 32773	202.77.162.213	669-699, 60251,60255,60269,60276 60289,60300,60519,60524, 60542,60549,60569,60578
172.16.112.50	32773	172.16.115.20	33368
172.16.112.10	32774		
172.16.114.30	32772		
172.16.114.20	32773		
172.16.116.194	34015		

3.4 DEFCON datasets

3.4.1 DEFCON-8

Το DEFCON 8 CTF dataset παρήγαγε ένα μεγάλο και σημαντικό αριθμό από alerts. Τα εξαγόμενα αυτά alerts από το Snort IDS ήταν σε αριθμό περίπου 1847745. Εξετάζοντας σχετικά alarms διαχωρίστηκαν από την αρχή σε δύο ομάδες: host probe alarms και service probes. Τα host probe alarms μετρήθηκαν περίπου στα 1255881 (67.9%) ενώ τα service probes ήταν 42398 (23%). Άλλα alerts περιλαμβάνουν παραβίαση συγκεκριμένων υπηρεσιών, DoS επιθέσεις κ.α τα οποία συνολικά ήταν 166466 (9.1%). Το DEFCON 8 περιέχει ένα σχετικά μεγάλο αριθμό ασυνήθιστων επιθέσεων οι οποίες συνέβησαν σε πολύ μικρό χρονικό διάστημα.

Αρχικά διαβάζουμε πληροφορίες του κάθε αρχείου από τα logs του snort:

```
[root@localhost snort]# snort -r snort.log.137414559
```

Πίνακας 15: Αριθμός εμφανίσεων πρωτοκόλλων ανά αρχείο

	IP4	ARP	ICMP	UDP	TCP
28160701.gz	133527	12699	1632	5474	126419
30020000.gz	784	176	20	172	578
29133000.gz	1787073	95244	27483	32668	753973
29163144.gz	1808720	35409	38244	46761	1168001
28153903.gz	1327305	41791	10825	23185	1259594
29132445.gz	503717	20958	1183	3477	222459

Έπειτα για σάρωση του κάθε αρχείου με σκοπό την εύρεση απειλών εισάγαμε το παρακάτω:

```
[root@localhost snort]# snort -dev -l var/log -daq pcap --daq-mode
read-file -r /home/fanis/Desktop/defcon8/29132445 -c etc/snort.conf
```

Πίνακας 16: Μέσος αριθμός εμφανίσεων απειλών ανά αρχείο καταγραφής

Παραβίαση κανόνα NIDS	Μέσος αριθμός εμφανίσεων ανά αρχείο καταγραφής
RPC sadmind query with root credentials attempt UDP	1
BAD-TRAFFIC TMG Firewall Client long host entry exploit attempt	7
FILE-IDENTIFY Microsoft Media Player Compressed skin download request	1
FILE-IDENTIFY New Executable binary file magic detected	1
MISC Mozilla Network Security Services SSLv2 stack overflow attempt	1
SERVER-IIS cmd.exe access	9
BAD-TRAFFIC potential dns cache poisoning attempt – mismatched txid	2
BAD-TRAFFIC Microsoft ISA Server and Forefront Threat Management Gateway invalid RST denial of service attempt	32

Οι παραπάνω μέσοι υπολογισμοί εμφανίσεων προέκυψαν από την ανάλυση δεδομένων ενός μικρού δείγματος αρχείων καταγραφής κίνησης (.gz).

```
[**] [3:19187:2] BAD-TRAFFIC TMG Firewall Client long host entry exploit attempt [**]
[Classification: Attempted User Privilege Gain] [Priority: 1]
07/28-17:40:03.995833 00:D0:B7:1E:BE:20 -> 00:E0:29:40:F0:1F type:0x800 len:0x21E
199.221.47.7:53 -> 10.20.11.123:1024 UDP TTL:243 TOS:0x0 ID:2865 IpLen:20 DgmLen:528 DF
Len: 500
[Xref => http://technet.microsoft.com/en-us/security/bulletin/MS11-040][Xref => http://cve.m
itre.org/cgi-bin/cvename.cgi?name=2011-1889]

[**] [1:17546:6] FILE-IDENTIFY Microsoft Media Player compressed skin download request [**]
[Classification: Potential Corporate Privacy Violation] [Priority: 1]
07/28-17:40:12.299776 00:E0:29:40:F0:1F -> 00:D0:B7:1E:BE:20 type:0x800 len:0x1FC
10.20.11.123:2929 -> 207.25.71.28:80 TCP TTL:243 TOS:0x0 ID:47845 IpLen:20 DgmLen:494
***A**** Seq: 0x7D4F14A2  Ack: 0xF2ABE302  Win: 0x2798  TcpLen: 32
[Xref => http://technet.microsoft.com/en-us/security/bulletin/MS07-047][Xref => http://cve.m
itre.org/cgi-bin/cvename.cgi?name=2007-3037][Xref => http://www.securityfocus.com/bid/25305]

[**] [1:21244:7] FILE-IDENTIFY New Executable binary file magic detected [**]
[Classification: Misc activity] [Priority: 3]
07/28-17:44:17.422551 00:D0:B7:1E:BE:20 -> 00:E0:98:81:EC:5B type:0x800 len:0x438A
209.146.146.82:80 -> 10.20.11.204:1110 TCP TTL:128 TOS:0x0 ID:2059 IpLen:20 DgmLen:17276 DF
***A**** Seq: 0x3D30B5DC  Ack: 0x148306  Win: 0x0  TcpLen: 20
[Xref => http://support.microsoft.com/kb/65122]
```

Εικόνα 3: Αποτύπωση alerts στο αρχείο 28153903.gz

3.4.2 DEFCON-9

Το Defcon 9 dataset αποτελεί ένα άλλο πολύ γνωστό σύνολο δεδομένων που έχει χρησιμοποιηθεί για αξιολόγηση των NIDS. Τα δεδομένα αυτού του συνόλου περιέχουν δικτυακή κίνηση η οποία είχε καταγραφεί κατά τη διάρκεια ενός διαγωνισμού hacking με όνομα 'Capture The Flag (CTF)' , στον οποίο οι διαγωνιζόμενοι χωρίστηκαν σε δύο ομάδες: επιτιθέμενοι και αμυνόμενοι. Η κίνηση που παράχθηκε κατά τη διάρκεια του CTF διαφέρει αρκετά από την πραγματική κίνηση που παράγεται σε ένα φυσιολογικό δίκτυο αφού περιέχεται και κίνηση εισβολέων χωρίς ουσιαστικά κάποιο ίχνος φυσιολογικής background κίνησης. Εξαιτίας αυτού του περιορισμού, το σύνολο δεδομένων αυτό συνήθως χρησιμοποιείται για αξιολόγηση τεχνικών συσχετισμού μηνυμάτων alerts.

Πίνακας 17: Κατανομή είδους επιθέσεων ανά αρχείο

Όνομα αρχείου	Uknown Traffic	Bad Traffic	DoS attacks	Protocol decoding
defcon_eth0.dump4			✔	
defcon_eth3.dump2	✔	✔		✔
defcon_eth0.dump5			✔	✔
defcon_eth0.dump	✔	✔	✔	
defcon_eth3.dump6		✔	✔	
defcon_eth3.dump3		✔		✔
defcon_eth3.dump4			✔	✔
defcon_eth3.dump5		✔		
defcon_eth3.dump8	✔		✔	
defcon_eth0.dump7				✔
defcon_eth0.dump6			✔	✔
defcon_eth0.dump3			✔	

Πίνακας 18: Εμφανιζόμενες απειλές

(http_inspect) INVALID CONTENT-LENGTH OR CHUNK SIZE
Consecutive TCP small segments exceeding threshold
(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE
Reset outside window
Limit on number of overlapping TCP packets reached
(ssp_frag3) Excessive fragment overlap
(ssp_frag3) Teardrop attack
(ssp_frag3) Tiny fragment
Short fragment, possible DOS attempt
Fragmentation overlap

3.4.3 DEFCON-10

Το DEFCON-10 ανήκει στην ομάδα των datasets μαζί με το DEFCON 8, το DEFCON 9 και τα iCTF σύνολα δεδομένων. Ύστερα από ανάλυση των αποτυπωμάτων κίνησης που πραγματοποιήσαμε μέσω του Snort παρατηρήσαμε ότι εμφανίστηκαν συνολικά 102 alerts. Ειδικότερα τα είδη επιθέσεων ανά αρχείο .cap φαίνονται στον πίνακα 19.

Αρχικά διαβάζουμε το κάθε αρχείο μέσω του snort:

```
[root@localhost snort]# snort -dev -l var/log --daq pcap --daq-mode read-
file -r /home/fanis/Desktop/defcon10/orange1.5.cap -c etc/snort.conf
```

Πίνακας 19: Κύριες απειλές που εμφανίστηκαν στο DEFCON-10

Αρχείο	**Priority**	**Κύρια Απειλή**	**Ταξινόμηση Απειλής**
orange1.5.cap	1	SERVER-IIS cmd.exe access	Web Application Attack
orange2.3.cap	2	BAD-TRAFFIC potential dns cache poisoning attempt	Misc Attack
orange2.7.cap	1	SERVER-IIS WEBDAV nessus safe scan attempt	Attempted Administrator Privilege Gain
orange3.1.cap	2	BAD-TRAFFIC potential dns cache poisoning attempt – mismatched txid	Attempted Information Leak
red3.1.cap	1	SERVER-IIS cmd.exe access	Web Application Attack

```
[**] [1:1002:17] SERVER-IIS cmd.exe access [**]
[Classification: Web Application Attack] [Priority: 1]
08/04-21:05:22.140953 00:C0:95:E0:0E:AC -> 00:50:56:40:5D:34 type:0x800 len:0x84
192.168.1.1:3062 -> 192.168.1.2:80 TCP TTL:64 TOS:0x0 ID:6376 IpLen:20 DgmLen:118 DF
***A**** Seq: 0x86B5F95  Ack: 0xD114C3D3  Win: 0x7D78  TcpLen: 20

[**] [3:13667:15] BAD-TRAFFIC dns cache poisoning attempt [**]
[Classification: Misc Attack] [Priority: 2]
08/03-21:46:43.181166 00:C0:95:E0:0E:AD -> 00:50:56:49:01:29 type:0x800 len:0xA2
192.168.2.1:53 -> 192.168.2.2:1027 UDP TTL:63 TOS:0x0 ID:0 IpLen:20 DgmLen:148 DF
Len: 120
[Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=2009-0234][Xref => http://cve.mitre.or
g/cgi-bin/cvename.cgi?name=2007-3898][Xref => http://technet.microsoft.com/en-us/security/bul
letin/MS09-008][Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=2009-0233][Xref => http
://www.kb.cert.org/vuls/id/800113][Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=1999
-0024][Xref => http://technet.microsoft.com/en-us/security/bulletin/MS08-037][Xref => http://
cve.mitre.org/cgi-bin/cvename.cgi?name=2008-1447][Xref => http://technet.microsoft.com/en-us/
security/bulletin/MS08-020][Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=2008-0087]
```

Εικόνα 4: Μερικά alerts όπως εμφανίστηκαν στο αρχείο καταγραφής

3.4.4 DEFCON-11

Και σε αυτό το σύνολο δεδομένων για τη μελέτη του, χρησιμοποιήσαμε το Snort για να ανιχνεύσουμε τα είδη απειλών που προκύπτουν στην κακόβουλη κίνηση. Συνολικά σε αυτό το dataset εμφανίστηκαν 354 alarms. Επίσης όπως φαίνεται και στον παρακάτω πίνακα από το εγκατεστημένο NIDS που τρέξαμε στον υπολογιστή μας εμφανίζονται τα συνολικά αιτήματα POST και GET που πραγματοποιήθηκαν καθώς και οι επικεφαλίδες HTTP και τα cookies που ενθυλακώθηκαν στα εκάστοτε HTTP requests και responses του χρήστη.

Πίνακας 20: Πληροφορίες διαδικτυακών αιτημάτων

Αρχείο	POST methods	GET methods	HTTP headers	HTTP cookies
ulogd.eth0.pcap	0	945	945	0
ulogd.znb0.pcap	0	0	0	0
ulogd.znb1.pcap	359	2458	5512	2450
ulogd.znb2.pcap	110	1458	3125	1260
ulogd.znb3.pcap	256	1910	4342	1604
ulogd.znb4.pcap	243	2835	5967	2555
ulogd.znb5.pcap	243	2835	5967	2555
ulogd.znb6.pcap	202	2204	4806	1896
ulogd.znb7.pcap	463	4470	9652	3846

```
HTTP Inspect - encodings (Note: stream-reassembled packets included):
    POST methods:                         110
    GET methods:                          1458
    HTTP Request Headers extracted:       1572
    HTTP Request Cookies extracted:       1113
    Post parameters extracted:            112
    HTTP response Headers extracted:      1553
    HTTP Response Cookies extracted:      147
    Unicode:                              0
    Double unicode:                       0
    Non-ASCII representable:              0
    Directory traversals:                 0
    Extra slashes ("//"):                 299
    Self-referencing paths ("./"):        0
    HTTP Response Gzip packets extracted: 0
    Gzip Compressed Data Processed:       n/a
    Gzip Decompressed Data Processed:     n/a
    Total packets processed:              22228
```

Εικόνα 5: Πληροφορίες HTTP πρωτοκόλλου για το αρχείο ulogd.znb2.pcap

3.5 CDX 2009

Το σύνολο δεδομένων CDX 2009 περιέχει κίνηση η οποία προέρχεται από συστήματα μολυσμένα από rootkits τα οποία δημιουργήθηκαν από τον NSA ειδικά για αυτόν το σκοπό. Επιπλέον, το CDX 2009 περιέχει καταγραφές οι οποίες προέρχονται από κίνηση δικτυακών συσκευών παρακολούθησης, hosts και servers που άνηκαν στο εσωτερικό διαγωνιστικό δίκτυο του USMA. Επιπροσθέτως με τα πραγματικά logs, το CDX 2009 περιλαμβάνει καταγραφές ανίχνευσης κακόβουλης δραστηριότητας οι οποίες προήλθαν από τα εσωτερικά συστήματα ανίχνευσης.

Πίνακας 21: Παραβιάσεις κανόνων που παρατηρήθηκαν σε ένα μέρος του συνόλου

Παραβίαση κανόνα NIDS	Εμφανίσεις
RPC sadmind query with root credentials attempt UDP	1
BAD-TRAFFIC TMG Firewall Client long host entry exploit attempt	7
FILE-IDENTIFY Microsoft Media Player Compressed skin download request	1
FILE-IDENTIFY New Executable binary file magic detected	1
MISC Mozilla Network Security Services SSLv2 stack overflow attempt	1
SERVER-IIS cmd.exe access	9
BAD-TRAFFIC potential dns cache poisoning attempt – mismatched txid	2
BAD-TRAFFIC Microsoft ISA Server and Forefront Threat Management Gateway invalid RST denial of service attempt	32

3.6 Kyoto2006+

Το σύνολο δεδομένων Kyoto2006+ αποτελείται από 17 στατιστικά χαρακτηριστικά εκ των οποίων τα 14 είναι συμβατικά (εμφανίζονται και σε προηγούμενα datasets) και 3 επιπρόσθετα. Ουσιαστικά, τα πρώτα 14 χαρακτηριστικά εξήχθησαν από τους δημιουργούς βασισμένοι αρχικά στο KDDCUP99 dataset, το οποίο είναι πολύ δημοφιλές και ευρέως χρησιμοποιήσιμο για αξιολόγηση δεδομένων στο ερευνητικό πεδίο της ανίχνευσης εισβολών. Παρόλα αυτά, αρκετές έρευνες έχουν υποδείξει ότι πολλά από αυτά τα χαρακτηριστικά εμφανίζονται διπλές φορές και πολλά από αυτά είναι ανεκτικά σε false positives. Έτσι στο Kyoto dataset χρησιμοποιούνται μόνο τα 14 πιο σημαντικά χαρακτηριστικά από το KDDCUP99 που αφορούν την κίνηση η οποία ανακτήθηκε από ένα honeypot σύστημα εγκατεστημένο στο πανεπιστήμιο Kyoto.

Επιπλέον σε αυτά τα χαρακτηριστικά προστέθηκαν τρία ακόμα, τα οποία μας παρέχουν τη δυνατότητα να διερευνήσουμε πιο αποτελεσματικά τι ακριβώς συνέβη στα τρέχοντα δίκτυα. Φυσικά μπορούν να χρησιμοποιηθούν και για τον έλεγχο των δεδομένων μας. [7] Η καταγραφή της κίνησης (μεγέθους 16 GB) ξεκίνησε στις 01.11.2006 και ολοκληρώθηκε στις 24.08.2009. Παρακάτω παρουσιάζουμε ενδεικτικά στατιστικά της κίνησης που παρατηρήθηκε από ένα μέρος του dataset (αρχείο 20070322.txt).

Πίνακας 22: Εμφάνιση τιμών που παρατηρήθηκαν για χαρακτηριστικά του Kyoto2006+

Χαρακτηριστικό	Ελάχιστη τιμή	Μέση τιμή	Μέγιστη τιμή
connection_duration	0.000	5.881	9086.110
service	6.0	38.5	49.0
source_bytes	0.0	234.3	127285.0
destination_bytes	0.0	400.3	284171.0
count	0.0000	0.1363	16.0000
same_srv_rate	0.00000	0.06188	1.00000
serror_rate	0.000000	0.009408	1.000000
srv_serror_rate	0.00000	0.09925	1.00000
dst_host_count	0.00	11.42	100.00
dst_host_same_src_port_rate	0.00000	0.07681	1.00000
dst_host_serror_rate	0.00000	0.03514	1.00000
dst_host_srv_serror_rate	0.000	0.072	1.000
malware_detection	0	0	0
label	-2.000	-1.057	1.000
source_port_number	0	7701	65364
destination_port_number	0.0	914.7	65016.0
session_duration	0.000	5.881	9086.114

Αναλυτικές πληροφορίες των διευθύνσεων IPv6 πηγής και προορισμού που καταγράφηκαν σε συνδυασμό με το συνολικό αριθμό εμφανίσεων τους.

Πίνακας 23: Αριθμός εμφανίσεων διευθύνσεων IPv6 πηγής και προορισμού

Διεύθυνση IPv6 πηγής		**Διεύθυνση IP προορισμού**	
fd13:310e:c8c7:e39b:0963:1c27:007c:6fc6	476	fd13:310e:c8c7:f287:7ddc:2741:6110:0faa	917
fd13:310e:c8c7:0bd3:5393:2602:032d:3e78	413	fd13:310e:c8c7:1db0:7d0c:27cf:0ff6:15de	415
fd13:310e:c8c7:103b:3c3c:0319:3f95:0c24	351	fd13:310e:c8c7:e4f1:2c90:1993:2130:0fa9	311
fd13:310e:c8c7:a828:0699:03d2:212f:035d	336	fd13:310e:c8c7:52e1:7dd1:271b:61b6:0d2f	219
fd13:310e:c8c7:31a9:4d95:5102:0d25:567b	325	fd13:310e:c8c7:33a0:2166:1843:073b:161b	122
fd13:310e:c8c7:fe31:4dcc:0363:123d:3fcb	258	fd13:310e:c8c7:33fa:179e:25a3:5434:00df	108
άλλες	4768	άλλες	4835

Επιπλέον παρακάτω παραθέτουμε αριθμητικά στοιχεία για τα υπόλοιπα αποτυπώματα.

Πίνακας 24: Συνολικές εμφανίσεις δικτυακών αποτυπωμάτων

Δικτυακό αποτύπωμα	**Συνολικές εμφανίσεις**	**Δικτυακό αποτύπωμα**	**Συνολικές εμφανίσεις**
OTH	3788	SF	97
RSTO	1528	RSTOS0	22
S0	1267	άλλες	40
REJ	185		

Σημαντική πληροφορία η οποία επίσης αντλήσαμε από το dataset είναι ότι σε γενικές γραμμές το σύστημα ανίχνευσης εισβολών παρήγαγε 1423 alerts και η φυσιολογική κίνηση αφορά 5504 frames. Επίσης πραγματοποίηθηκαν 524 exploits στα οποία εκτελέστηκε ή έγινε εισαγωγή shellcode.

Πίνακας 25: Αριθμός εμφανίσεων συγκεκριμένων security alarms

Alert ID	**Τιμή**	**Εμφανίσεις**	**Shellcode ID**	**Τιμή**	**Εμφανίσεις**
-	0	5504	-	0	6403
21192	1	728	128	2	4
21192, 210000	1	274	129, 130, 131	1	43
21000, 21192	1	215	58	1	30
21545	1	57	58	2	447
20081	1	47			
άλλα	-	102			

Παράλληλα, είδαμε ότι σε 75 συνεδρίες έγινε εκκίνηση αυτών κατά την ίδια στιγμή.

Πίνακας 26: Αριθμός αποτυπωμάτων ανά χρονικό σημείο

Χρόνος εκκίνησης συνεδρίας	Αριθμός αποτυπωμάτων
02:23:24	75
00:29:28	67
02:23:26	63
00:29:27	31
02:23:25	30
02:23:27	27
άλλη	6634

Πίνακας 27: Χαρακτηριστικά Kyoto2006+

duration	service	source bytes
destionation bytes	count	same_srv_rate
srv_error_rate	dst_host_count	dst_host_srv_count
dst_host_same_src_port_rate	dst_host_serror_rate	dst_host_srv_serror_rate
flag	ids_detection	malware_detection
ashula_detection	serror_rate	

3.7 LBNL/ICSI Enterprise Tracing

Στόχος αυτού του dataset είναι να παρουσιάσει κίνηση από χιλίαδες εσωτερικά δίκτυα εταιρειών. Η ουσιαστική συνεισφορά του συνόλου αυτού είναι ότι μπορούμε να καθορίσουμε τις διαφορές που διακρίνουν πακέτα δεδομένων που μεταφέρονται σε χώρους εταιρικούς από άλλα πακέτα που αφορούν φυσιολογική κίνηση ενός χρήστη. Για την καταγραφή των αποτυπωμάτων στο LBNL/ICSI χρειάστηκαν περισσότερες από 100 ώρες.

Γενικά, με αυτό το dataset είμαστε σε θέση να συγκρίνουμε αλλά και να καθορίσουμε τα χαρακτηριστικά μίας εταιρικής κίνησης σε σχέση με μία κίνηση απλή στο διαδίκτυο. Τα κύρια alerts που εκτυπώθηκαν εδώ ήταν RESET-OUTSIDE WINDOW και Bad-TRAFFIC TCP WINDOW closed before receiving data. [8]

Πίνακας 28: Αριθμός security alarms ανά αρχείο καταγραφής

Αρχεία	Alerts
lbl-internal.20041004-1305.port001.dump.anon	46
lbl-internal.20041004-1305.port002.dump.anon	76
lbl-internal.20041004-1305.port003.dump.anon	980
lbl-internal.20041004-1305.port004.dump.anon	12
lbl-internal.20041004-1305.port005.dump.anon	360
lbl-internal.20041004-1305.port006.dump.anon	1526
lbl-internal.20041004-1305.port007.dump.anon	271
lbl-internal.20041004-1305.port008.dump.anon	900

Παρακάτω φαίνονται πληροφορίες που συλλέχθησαν από το αρχείο lbl-internal.20041004-1305.port003.dump.anon σχετικά με το χρονικό των επιθέσεων ανά διεύθυνση IP.

Πίνακας 29: Αντιστοίχιση IP διευθύνσεων ανά χρονικό σημείο καταγραφής

Χρόνος καταγραφής	IP διεύθυνση εισβολέα	IP διεύθυνση αμυνόμενου
22:18:16	44:39:9E:4B:DA:36	80:0B:9B:3B:B9:EC
22:26:30	80:0B:9B:3B:B9:EC	10:10:29:A7:DC:70
22:29:21	80:0B:9B:3B:B9:EC	92:B8:FC:F0:9C:82
22:35:30	92:B8:FC:F0:9C:82	80:0B:9B:3B:B9:EC
22:38:11	80:0B:9B:3B:B9:EC	1C:31:4D:95:C9:80
22:44:34	06:23:FB:22:25:F1	80:0B:9B:3B:B9:EC

3.8 ISOT Botnet 2010

Το σύνολο δεδομένων αυτό αποτελεί μία μίξη από δημόσια υπάρχοντα datasets τόσο με κακόβουλη όσο και φυσιολογική κίνηση. Ειδικότερα, περιέχει κίνηση που είχε χρησιμοποιηθεί για σκοπούς του Honeynet Project στο οποίο είχε γίνει χρήση δυό botnets του Storm και του Waledac botnet αντίστοιχα. [9]

Αρχικά αφού κατεβάσαμε το αρχείο κίνησης (ISOT_Botnet_DataSet_2010.pcap) , επειδή ήταν αρκετά μεγάλο (11.5 GB) το διαχωρίσαμε σε μικρότερα κομμάτια περίπου των 100 MB το καθένα μέσω της εντολής editcap.

```
MEGALOS2: ISOT_Botnet_DataSet_2010 fsiamp$ editcap -c 100000000
/Users/fsiamp/Desktop/isot_botnet/ISOT_Botnet_DataSet_2010/ISOT_Botn
et_DataSet_2010.pcap
```

Έπειτα, μεταβαίνοντας στο wireshark κάναμε εξαγωγή του αρχείου σε CSV έτσι ώστε να μπορούμε να διαβάσουμε και να αναλύσουμε τα δεδομένα κίνησης. [10] Από αυτά παρατηρήσαμε τα εξής:

Πίνακας 30: Αριθμός πρωτοκόλλων που εμφανίστηκαν στα εκάστοτε αρχεία

Αρχείο	ICMP	IPv4	UDP	DNS	ARP	SMTP	TCP
isot_00000_20101007010538	1271	-	99231	51113	9409	122369	416054
isot_00013_20041004234056	1143	183144	131066	1693	1152	-	615509
isot_00003_20071009162422	469	425445	174530	50	55	-	377933
isot_00004_20041004230728	127	592823	240564	10	19	-	164365
isot_00005_20041004230918	122	558980	226165	-	33	-	212707

Παρακάτω παρουσιάζουμε τις πιο ενεργές διευθύνσεις IP τόσο του παραλήπτη όσο και του αποστολέα (σε αυτές περιλαμβάνεται και ο επιτιθέμενος). Δηλαδή, αναφέρουμε τις διευθύνσεις ανά αρχείο στις οποίες υπήρξαν τα περισσότερα αιτήματα από και προς αυτές.

Πίνακας 31: Αντιστοίχιση IP διευθύνσεων ανά αρχείο κίνησης

Αρχείο	Διεύθυνση IP πηγής	Διεύθυνση IP προορισμού
isot_00000_20101007010538	172.16.2.2	172.16.2.2
isot_00013_20041004234056	131.243.140.86	131.243.95.40
isot_00003_20071009162422	128.3.70.97	128.3.23.3
isot_00004_20041004230728	128.3.70.97	128.3.23.3
isot_00005_20041004230918	128.3.70.97	128.3.23.3

Συνολικός αριθμός alerts που παρήχθησαν από το Snort για τα παραπάνω αρχεία: 29835

Πίνακας 32: Απειλές που ανιχνεύθηκαν στο dataset

Κακόβουλη δραστηριότητα	
NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE	TCP Timestamp is missing
Reset outside window	BAD-TRAFFIC Microsoft ISA Server and Forefront Threat Management Gateway
BAD-TRAFFIC TCP window closed befored receiving data	
BAD-TRAFFIC TMG Firewall Client long host entry exploit	
Consecutive TCP small segments exceeding threshold	

Μερικά από τα μηνύματα που εμφάνιστηκαν τις περισσότερες φορές στο Wireshark (ανα pcap αρχείο) για τα πακέτα που καταγράφηκαν παρουσιάζονται στον πίνακα 33.

Πίνακας 33: Μηνύματα περιγραφής πακέτων (Wireshark)

Source port: mdbs-daemon Destination port: nfs[Packet size limited during capture]
Source port: nfs Destination port: mdbs-daemon[Packet size limited during capture]
Continuation or non-HTTP traffic[Packet size limited during capture]
[TCP Previous segment not captured] Continuation or non-HTTP traffic[Packet size limited during capture]
[TCP Retransmission] Continuation or non-HTTP traffic[Packet size limited during capture]

3.9 CAIDA datasets

3.9.1 CAIDA DNS root/gTLD RTT

Το dataset αυτό περιέχει χρήσιμες πληροφορίες οι οποίες υποδεικνύουν τον τρόπο κατά τον οποίο έχει εξελιχθεί το διαδίκτυο ανά τα χρόνια που πέρασαν. Μελετώνται ουσιαστικά οι παράγοντες που επηρεάσαν την εξέλιξη αυτή μέσα από καταγραφή δεδομένων που αφορούν το DNS rountrip time (RTT).

Η συλλογή ροής δεδομένων πραγματοποιήθηκε μέσω του NeTraMet. Παρακάτω αναφέρουμε τις τιμές του εκάστοτε χαρακτηριστικού που χρησιμοποιήθηκε ως δείκτης περιγραφής του κάθε πακέτου που μεταφερόταν για το αρχείο 20050618.dif (το οποίο θεωρούμε ως δείγμα ικανό για τη μελέτη συμπεριφοράς του συνολικού συστήματος).

Πίνακας 34: Κατανομή τιμών ανά χαρακτηριστικό

	Ελάχιστη τιμή	**Μέση τιμή**	**Μέγιστη τιμή**
flowruleset	16	16	16
flowindex	5.00	13.26	29.00
firsttime	3205	619244	14799646
sourceperetype	1	1	1
sourcetranstype	17	17	17
flowkind	1.00	5.95	13.00
flowclass	0.0000	0.8536	1.0000
d_topdus	0	0	0
d_frompdus	1.00	14.42	13475.00
d_tolostpdus	0	0	0
d_fromlostpdus	1.00	14.42	13475.00
toturnaroundtime1	0	0	0

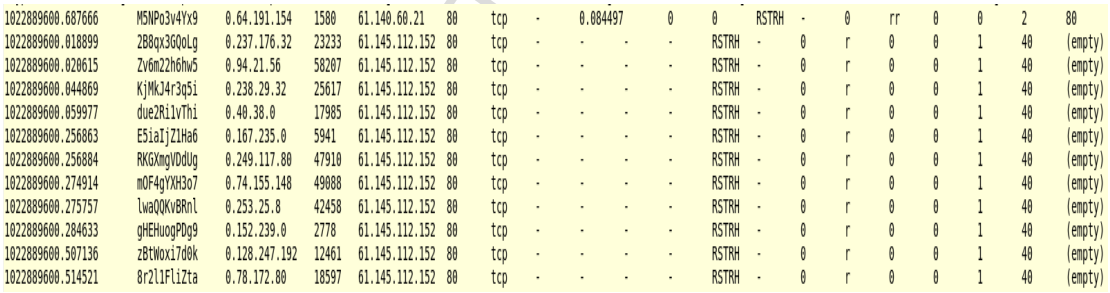
Τα σύνολα κανόνων που λήφθηκαν υπόψιν ήταν τα αρχεία dyn_root_ucsd, dyn-ucsd.rules και dyn-dist. Οι ρυθμίσεις παραμετροποίησης ορίστηκαν στα αρχεία dns-root.srl και global_dns.srl.

3.9.2 CAIDA Backscatter (TOCS και 2004-2008)

Το Backscatter περιέχει δεδομένα τα οποία συλλέχθηκαν με στόχο τη μελέτη ποικίλων ειδών επιθέσεων. Γενικά, το dataset αυτό περιέχει απαντήσεις αιτημάτων τα οποία στάλθηκαν από τα θύματα των επιθέσεων DoS. Δηλαδή, δεν περιέχεται κίνηση μεταξύ του επιτιθέμενου και του θύματος αλλά μόνο responses του υπολογιστή-θύματος σε άλλες διευθύνσεις IP.

Πίνακας 35: Απειλές που ανιχνεύθηκαν μέσω Snort και Bro

bad_TCP_checksum	truncated_header
connection_originator_SYN_ack	RST_storm
SYN_after_reset	excessively_small_fragment
active_connection_reuse	bad_TCP_header_len
bad_ICMP_checksum	fragment_with_DF
SYN_seq_jump	incompletely_captured_fragment
Reset outisde window	baroque_SYN
TCP_christmas	TCP timestamp is missing



1022889600.687666	M5NPo3v4Yx9	0.64.191.154	1580	61.140.60.21	80	tcp	-	0.084497	0	0	RSTRH	-	0	rr	0	0	2	80
1022889600.018899	2B8qx3GQoLg	0.237.176.32	23233	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.020615	Zv6m22h6hw5	0.94.21.56	58207	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.044869	KjMkJ4r3q5i	0.238.29.32	25617	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.059977	due2Ri1vThi	0.40.38.0	17985	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.256863	E5iaIjZ1Ha6	0.167.235.0	5941	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.256884	RKGXmgVDdUg	0.249.117.80	47910	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.274914	mOF4gYXH3o7	0.74.155.148	49088	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.275757	lwaQQKvBRnl	0.253.25.8	42458	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.284633	gHEHuogPDg9	0.152.239.0	2778	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.507136	zBtWoxi7d0k	0.128.247.192	12461	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)
1022889600.514521	8r2l1FliZta	0.78.172.80	18597	61.145.112.152	80	tcp	-	-	-	-	RSTRH	-	0	r	0	0	1	40	(empty)

Εικόνα 6: Αρχείο conn.log παραγόμενο μέσω του εργαλείου Bro

Από το αρχείο conn.log που παρήγαγε το Bro παρατηρήσαμε ότι ο προορισμός που δέχθηκε τα περισσότερα αιτήματα άρα και τις περισσότερες επιθέσεις DoS ήταν αυτός με τη διεύθυνση IP 61.145.112.152. Επίσης, τα περίσσοτερα πακέτα που αποτυπώθηκαν ήταν TCP πρωτοκόλλου και το flag που εμφανίστηκε τις περισσότερες φορές ήταν το RSTRH.

```
1186264516.084275   9SoQOYFQXok   197.217.162.134 14882   71.126.222.64   8555    active_connection_reuse -       F       bro
1186264516.643247   Uc2cMI1Y9bk   167.234.229.190 52642   71.126.222.64   58985   active_connection_reuse -       F       bro
1186264516.891223   vzgNngLRTw1   208.189.233.138 21170   71.126.222.64   31008   TCP_christmas   -       F       bro
1186264516.891223   vzgNngLRTw1   208.189.233.138 21170   71.126.222.64   31008   baroque_SYN     -       F       bro
1186264516.891223   vzgNngLRTw1   208.189.233.138 21170   71.126.222.64   31008   SYN_with_data   -       F       bro
1186264517.042360   R7dnE2KUq39   202.107.62.252  59351   71.126.222.64   21946   active_connection_reuse -       F       bro
1186264517.079324   oCwXa0HpjFj   200.227.110.2   54379   71.126.222.64   16314   active_connection_reuse -       F       bro
1186264517.413451   iwKrX9hHMvb   216.136.61.108  48544   71.126.222.64   49601   bad_TCP_checksum        -       F       bro
```

Εικόνα 7: Αρχείο weird.log του Bro

3.9.3 CAIDA 'DDoS Attack 2007'

Αυτό το dataset περιλαμβάνει ανώνυμη κίνηση μίας ώρας κατά την οποία πραγματοποιήθηκε κατανεμημένη επίθεση τύπου άρνησης υπηρεσιών (DoS). Το συνολικό μέγεθος του συνόλου αυτού είναι 21 GB και η διαδικασία καταγραφής έγινε στις 4 Αυγούστου 2007 (20:50:08 UTC με 21:56:16 UTC).

Για τη μελέτη του dataset χρησιμοποιήσαμε το Snort και το Bro για να διαβάσουμε τα pcap αρχεία και να εντοπίσουμε όλα τα πιθανά alarms που υποδεικνύουν εισβολή επιτιθέμενου.

```
[root@localhost ddos_attack]#bro -r ddostrace.20070804_14536.pcap
```

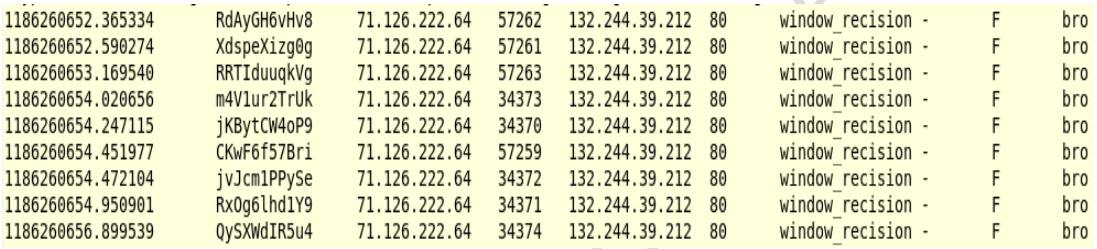
Πίνακας 36: Απεικόνιση πληροφοριών ανίχνευσης στο CAIDA DDoS Attack 2007 Dataset

ΑΡΧΕΙΟ	ΣΥΜΒΑΝΤΑ	ΚΑΤΑΓΡΑΦΕΣ	ALERTS
ddostrace.20070804_134936.pcap	824	822	822
ddostrace.20070804_135436.pcap	37	37	37
ddostrace.20070804_135936.pcap	199	199	199
ddostrace.20070804_140436.pcap	6	6	6
ddostrace.20070804_140936.pcap	0	0	0
ddostrace.20070804_145436.pcap	703	703	703

Παρακάτω αναφέρονται οι συνολικές απειλές που ανιχνεύθηκαν στα παραπάνω αρχεία.

Πίνακας 37: Κατανομή απειλών με χρήση Snort και Bro

windows_recision	Καταγραφή μέσω BRO
TCP Timestamp is missing	Καταγραφή μέσω SNORT
Reset outside window	Καταγραφή μέσω SNORT
bad_TCP_checksum	Καταγραφή μέσω BRO
active_connection_reuse	Καταγραφή μέσω BRO



1186260652.365334	RdAyGH6vHv8	71.126.222.64	57262	132.244.39.212	80	window_recision	-	F	bro
1186260652.590274	XdspeXizg0g	71.126.222.64	57261	132.244.39.212	80	window_recision	-	F	bro
1186260653.169540	RRTIduuqkVg	71.126.222.64	57263	132.244.39.212	80	window_recision	-	F	bro
1186260654.020656	m4V1ur2TrUk	71.126.222.64	34373	132.244.39.212	80	window_recision	-	F	bro
1186260654.247115	jKBytCW4oP9	71.126.222.64	34370	132.244.39.212	80	window_recision	-	F	bro
1186260654.451977	CKwF6f57Bri	71.126.222.64	57259	132.244.39.212	80	window_recision	-	F	bro
1186260654.472104	jvJcm1PPySe	71.126.222.64	34372	132.244.39.212	80	window_recision	-	F	bro
1186260654.950901	RxOg6lhd1Y9	71.126.222.64	34371	132.244.39.212	80	window_recision	-	F	bro
1186260656.899539	QySXWdIR5u4	71.126.222.64	34374	132.244.39.212	80	window_recision	-	F	bro

Εικόνα 8: Παραγόμενα alerts του Bro

Πίνακας 38: Διεθύνσεις IP εισβολέων στις οποίες έχει γίνει απόκρυψη της πραγματικής τους ταυτότητας (πλαστοπροσωπεία)

199.129.180.200	202.107.62.252
52.115.91.169	213.206.20.210
204.54.196.69	97.198.3.210
250.242.57.168	195.24.220.74
52.115.91.169	39.235.117.122
167.234.229.190	197.217.162.134

Γενικά το dataset αυτό περιέχει κατά το μεγαλύτερο ποσοστό κακόβουλα αποτυπώματα κίνησης, τα οποία συγκεκριμένα περιέχουν κίνηση η οποία δρομολογείται στο εκάστοτε θύμα μαζί με τα responses που αποστέλλονται από το θύμα. Ουσιαστικά ο σκοπός των επιθέσεων που πραγματοποιήθηκε ήταν να τεθεί εκτός διαθεσιμότητας η πρόσβαση σε συγκεκριμένους εξυπηρετητές καταναλώνοντας σημαντικά ποσά υπολογιστικών πόρων και bandwidth. Για την εισροή ανώνυμης κίνησης έγινε χρήση του εργαλείου CryptoPAn.

Κεφάλαιο 4ο

Μελέτη Αδυναμιών και Συγκριτική Ανάλυση

Όλα τα παραπάνω σύνολα δεδομένων που αναφέραμε αποτελούν πολύτιμα στοιχεία για κάθε ερευνητή και γενικά για οποιονδήποτε ασχολείται ενεργά με την ανίχνευση παρεισφρήσεων. Ωστόσο, πάρα πολλά από αυτά έχουν ένα σημαντικό μειονέκτημα το οποίο σχετίζεται με το γεγονός ότι αυτά τα σύνολα δεν αποτελούν καλή αναπαραστάση μίας πραγματικής κίνησης δεδομένων. Για παράδειγμα, τα DARPA datasets έχουν κατακριθεί σχετικά με το πόσο ρεαλιστική είναι η κανονική κίνηση που έχει καταγραφεί [11,12] διότι ένα μέρος τους έχει παραχθεί συνθετικά (τεχνητά). Επιπροσθέτως, μαζί με τη δυσκολία προσομοίωσης μίας κατάστασης πραγματικής δικτυακής κίνησης, υπάρχουν και κάποιες επιπλέον προκλήσεις που αφορούν την αξιολόγηση των NIDS.

Αυτές περιλαμβάνουν δυσκολίες που σχετίζονται με τη συλλογή των κατάλληλων attack scripts και γενικά λογισμικού που χρησιμοποιείται εκείνη τη στιγμή από το εκάστοτε θύμα-υπολογιστή, τις διαφορετικές απαιτήσεις για τον συγκριτικό έλεγχο απόδοσης μεταξύ NIDS βασισμένων σε υπογραφές και αυτών που είναι βασισμένα σε εύρεση ανώμαλης συμπεριφοράς, την ύπαρξη τόσο δικτυακών IDS όσο και host-based IDS κ.α

Αρχικά θα αναλύσουμε το KDDCUP99 dataset το οποίο είναι το πιο δημοφιλές σύνολο δεδομένων από αυτά που αναφέραμε. Το κύριο πρόβλημα το οποίο εμφανίζεται εδώ αρχικά είναι ότι πλέον αυτό το dataset είναι αρκετά παλιό και ειδικότερα τα dataset instances που είχαν παραχθεί πλέον είναι ξεπερασμένα. Συνεπώς, από αυτό το γεγονός διαπιστώνουμε ότι το KDDCUP δεν μπορεί να αναπαραστήσει τη συμπεριφορά ενός σύγχρονου δικτύου καθώς και τους διαφορετικούς τύπους επιθέσεων που μπορούν να συμβούν τη δεδομένη στιγμή στην εποχή μας. Επίσης, οι επιθέσεις που είχαν πραγματοποιήθει ήταν προσχεδιασμένες και δεν έγιναν σε πραγματικό χρόνο με συνέπεια να ήταν γνωστές οι προθέσεις του κάθε επιτιθέμενου άρα και τα συστήματα ανίχνευσης ήταν σε θέση να τις αναγνωρίσουν.

Παράλληλα, τα χαρακτηριστικά που σχηματίστηκαν δεν συμβαδίζουν απόλυτα με τους κανόνες δικτύου που είχαν τεθεί τότε. Αυτό μπορούμε να το κατανοήσουμε καλύτερα από τη στιγμή που η δημιουργία τέτοιων χαρακτηριστικών σε πραγματικού χρόνου κίνηση είναι

αρκετά δύσκολη ή αν όχι αδύνατη. Άρα είναι προφανές ότι μιλάμε για μία κίνηση η οποία έχει παραχθεί κατά κύριο λόγο τεχνητά και με συνθετικό τρόπο και όχι σε ένα πραγματικό περιβάλλον δικτύων (αυτό αφορά τη συνολική background κίνηση). Όσον αφορά, τις επιθέσεις που πραγματοποιήθηκαν είδαμε ότι πραγματοποιήθηκαν 4 είδη επιθέσεων (U2R, DoS, R2L και PROBE) οι οποίες όμως καλύπτουν μία γενική γκάμα απειλών οι οποίες πλέον τώρα αντιμετωπίζονται εύκολα. [13]

Συγκεκριμένα υπάρχει απουσία των παρακάτω χαρακτηριστικών επιθέσεων:

Πίνακας 39: Επιθέσεις που δεν πραγματοποιήθηκαν στο KDDCUP99

mailbomb	snmp_get	quest
syslogd	process_table	http_tunnel
ping_of_death	mscan	ffbconfig
udp_storm	count	fdfformat
ps	at	xlock
xterm	perl	eject
sqlattack	arppoison	secret
sshtrojan	nc-setup	nfsdos

Ουσιαστικά η πλειοψηφία των κακόβουλων ενεργειών που παρατηρήθηκαν στο παραπάνω σύνολο δεδομένων αφορούσαν επιθέσεις σε DoS (neptune). [14] Επιπλέον, χαρακτηριστικό είναι ότι χρησιμοποιήθηκαν μόλις δύο δίκτυα για την παραγωγή του dataset στα οποία αποτυπώθηκαν 11 μοναδικές διευθύνσεις IP αριθμός ιδιαίτερα χαμηλός το οποίο μας οδηγεί στο να κατανοήσουμε τον τρόπο και τη συμπεριφορά ενός επιτιθέμενου πιο εύκολα. [15] Παράλληλα, με αυτόν τον τρόπο είναι δεδομένο ότι περιορίζεται αρκετά ο εισβολέας και άρα δεν βοηθάει αυτό ώστε να έχουμε ένα dataset το οποίο θα αποτυπώνει ποικίλες επιθέσεις. Γενικά ο αριθμός των διαφορετικών ειδών επιθέσεων που πραγματοποιήθηκαν ήταν 57 που είναι αρκετά χαμηλός και επίσης η πλειοψηφία των χαρακτηριστικών του συνόλου αυτού ανταποκρίνονται καλύτερα στη λειτουργία του εκάστοτε host-based συστήματος ανίχευσης και λιγότερο στη συνολική λειτουργία του δικτύου. Επομένως, είναι δεδομένο ότι σήμερα δεν συνίσταται η χρήση του KDDCUP για αξιολόγηση δικτυακών IDS. [16]

Από την άλλη μεριά, το NSL-KDD Train παρουσιάζει και αυτό περίπου τις ίδιες αδυναμίες που ισχύουν και στο KKDCUP99 μαζί με κάποιες καινούργιες. Καταρχάς, όπως φαίνεται από τον πίνακα 5 υπάρχει μικρός αριθμός εισαγωγής shellcode prompts (num_shells), κάτι που σημαίνει ότι πραγματοποιήθηκαν ελάχιστες εκτελέσεις shellcode . Παράλληλα, από τη μέση τιμή num_compromised παρατηρούμε ότι λίγες φορές παραβιάστηκαν πολιτικές ασφάλειας που είχαν τεθεί για το σύστημα. Έπειτα, ένα άλλο μειονέκτημα αποτελεί η ένδειξη μικρής μέσης τιμής diff_srv_rate η οποία υποδεικνύει τον αριθμό συγκεκριμένων συνδέσεων οι οποίες χρησιμοποιούσαν τις ίδιες υπηρεσίες. [17] Ακόμα, βλέπουμε ότι δεν υπήρξε καμία εκτέλεση εντολής σε ftp συνεδρία άρα δεν πραγματοποιήθηκε καμία επίθεση μέσω FTP πρωτοκόλλου. Ιδιαίτερα εμφανές επίσης είναι ότι και σε αυτό το dataset σημειώθηκαν κυρίως DoS επιθέσεις κάτι το οποίο φαίνεται από τον μεγάλο αριθμό συνδέσεων στο TCP πρωτόκολλο (102688). Έπειτα από τη μεταβλητή num_file_creations παρατηρούμε ότι ελάχιστα αρχεία δημιουργήθηκαν άρα πραγματοποιήθηκε μικρός αριθμός επιθέσεων μη εξουσιοδοτημένης πρόσβασης ή privilege escalation. Γενικά ένα άλλο μειονέκτημα είναι η έλλειψη απειλών τύπου fragmentation overlap όπως φαίνεται από τις καταγεγραμένες τιμές wrong fragment πακέτων. Εξάλλου, όπως παρουσιάζεται και στη μεταβλητή 'hot' η οποία υποδεικνύει κακόβουλη ή ύποπτη συμπεριφορά η οποία ανιχνεύεται κάθε στιγμή από το NIDS, η ένδειξη μικρού μέσου αριθμού hot indicators φανερώνει ότι υπήρξαν επιθέσεις που έγιναν αλλά όχι σε μεγάλη έκταση αναλογικά με το μέγεθος του dataset.

Στη συνέχεια, σχετικά με το DARPA1998 dataset παρατηρούμε ότι σε αυτό παρήχθησαν αρκετά false positives από τα IDS της εποχής εκείνης λόγω της αδυναμίας αναγνώρισης της δομής ορισμένων πακέτων. Αυτό οφείλεται κατά βάση λόγω της ύπαρξης καινούργιων επιθέσεων οι οποίες δεν είχαν παρουσιαστεί σε προηγούμενα datasets. Επίσης, ένα άλλο μειονέκτημα του DARPA98 είναι ότι εμφανίζονται πολλά κοινά και επαναλαμβανόμενα μοτίβα κίνησης κάτι που οδηγεί έναν επιτιθέμενο εύκολα στη μελέτη τους και στην πρόβλεψη συμπεριφοράς του συστήματος. [18] Τα NIDS τότε δεν υποστήριζαν γενικά αλγορίθμους οι οποίοι να βρίσκουν νέες επιθέσεις και δεν είχαν αναπτυχθεί προσεγγίσεις βασισμένες σε αναγνώριση μοτίβων-υπογραφών. Ακόμα, η έλλειψη πιθανών στόχων-θυμάτων με υποστήριξη λειτουργικού συστήματος Windows/NT αποτελεί ένα άλλο πρόβλημα καθώς τότε το σύστημα αυτό χρησιμοποιούταν από ένα μεγάλο μερίδιο του κόσμου. Επίσης, τότε δεν υπήρχαν ενσωματωμένοι μηχανισμοί στα NIDS ώστε να ανιχνεύουν επιθέσεις εκ των έσω. Αυτό αποτελεί σημαντική αδυναμία διότι σε συνδυασμό με την αδυναμία των NIDS να πραγματοποιούν

ανίχνευση ανωμαλιών τότε, δεν υποστηριζόταν και η δυνατότητα για αναγνώριση νέων προηγούμενων επιθέσεων οι οποίες δεν είχαν ξαναεμφανιστεί. [19] [20]

Όσον αφορά το DARPA1999 δεν παρουσιάζεται ιδιαίτερα μεγάλη ποικιλομορφία επιθέσεων αφού παρουσιάζονται συνεχώς κοινά είδη απειλών όπως αποτυπώθηκαν στην καταγραφή μας από το Snort. Αυτό αποτελεί ένδειξη χαμηλής ποιότητας του dataset αφού σε κάθε αρχείο καταγραφής αναφέρονται οι ίδιες επιθέσεις. Ταύτοχρονα παρατηρούμε ότι εμφανίστηκαν πολλά false positives και ακόμα δεν υπάρχει ισορροπία μεταξύ του αριθμού εισβολέων-αμυνόμενων. Ειδικότερα, οι διευθύνσεις IP των ατόμων που εκτέλεσαν επιθέσεις ήταν πολύ λιγότερες από αυτές των αμυνόμενων. [21] [22]

Στο DARPA2000 παρατηρείται μικρός αριθμός συνεδριών στο LLDOS 2.0 κάτι το οποίο όπως φαίνεται οδήγησε πολλά πακέτα να είναι τα οποία δεν αναγνωρίστηκαν από τα NIDS. Παράλληλα, μία άλλη αδυναμία είναι ότι δεν πραγματοποιήθηκαν ταυτόχρονες και πολλαπλές επιθέσεις κατά τον ίδιο χρόνο. Αυτό φαίνεται από το μέγιστο αριθμό συνεδριών οι οποίες είναι 8. Επίσης παρατηρήσαμε ότι ο επιτιθέμενος έκανε χρήση αποκλειστικά δυναμικών και ιδιωτικών θυρών σύνδεσης και δεν εκτέλεσε επιθέσεις σε γνωστές εγγραμμένες θύρες (0-1023 ή 1024-49151) τη στιγμή που ο αμυνόμενος συνδέθηκε επί των πλείστων σε εγγεγραμμένες θύρες TCP. Ακόμη, σημαντικό στοιχείο είναι ότι εμφανίστηκαν στα αρχεία καταγραφής συναγερμών απειλές οι οποίες είχαν μικρή επικινδυνότητα, το οποίο σημαίνει ότι δεν έγιναν επιθέσεις με ιδιαίτερα αρνητικό αντίκτυπο για το σύστημα άρα δεν μπορεί να χρησιμεύσει ένα τέτοιο dataset στο μέγιστο βαθμό κατά την αξιολόγηση αποδοκτικότητας ενός NIDS.

Στη συνέχεια θα αναφερθούμε στα μειονεκτήματα που παρουσιάζουν τα σύνολα DEFCON. Αρχικά για το DEFCON 8 παρατηρούμε ότι δεν υπάρχει δυνατότητα αναγνώρισης πολλών ειδών απειλών (μέσω του Snort) που πραγματοποιήθηκαν. Ειδικότερα, βλέπουμε ότι αρκετά συχνά εμφανίζονται alerts τύπου ‘BAD-TRAFFIC’ τα οποία δεν μπορούν να σκιαγραφήσουν απόλυτα μία ενδεχόμενη απειλή με τη μέγιστη λεπτομέρεια. Επίσης, σε αυτό το dataset δίνεται έμφαση σε επιθέσεις άρνησης υπηρεσιών και λιγότερο σε διαδικτυακού τύπου απειλές ή επιθέσεις τύπου ανάκτησης δικαιώματων root. Αυτό είναι προφανές και από τη μεγάλη συνδεσιμότητα των χρηστών στα πρωτόκολλα TCP και IP4 και την ελάχιστη σε πρωτόκολλα UDP ή ICMP, δηλαδή υπάρχει έλλειψη συνδέσεων σε πρωτόκολλα τα οποία είναι αρκετά ευάλωτα σε επιθέσεις.

Έπειτα, στο DEFCON 9 παρατηρούμε ότι δίνεται έμφαση σε απειλές τύπου HTTP και TCP πρωτοκόλλου. Γενικά ο περιορισμός εδώ είναι ότι δεν έχουν πραγματοποιηθεί γνωστές

επιθέσεις όπως bufferoverflow attacks, αποστολή malware, misc attacks ούτε cross-site scripting επιθέσεις ή injections (LDAP,SQL) παρόλο που γίνεται χρήση του HTTP πρωτοκόλλου και των μεθόδων POST και GET. Ακόμα εδώ πρέπει να σημειώσουμε ότι εμφανίζονται μηνύματα alerts τύπου UNKNOWN TRAFFIC ή BAD-TRAFFIC κάτι το οποίο δεν μας παρέχει επαρκείς πληροφορίες σχετικά με το είδος της κακόβουλης ενέργειας που διενεργήθηκε.

Από την άλλη μεριά στο DEFCON 10, καταγράφηκαν μαζικά μηνύματα alarms τα οποία αφορούσαν εκτελέσεις κακόβουλων ενεργειών μέσα από εντολές σε τερματικό command-prompt. Από αυτό γίνεται αντιληπτή η χρησιμοποίηση κυρίως Windows λειτουργικών συστημάτων και όχι Linux ή Mac OS. Παράλληλα, εν αντιθέσει με τα προηγούμενα DEFCON datasets εδώ δεν πραγματοποιήθηκαν DoS επιθέσεις.

Στο DEFCON 11 παρατηρήθηκαν κυρίως επιθέσεις μέσω GET μηνυμάτων και όχι POST. Πιθανότατα αυτές οι ενέργειες παραπέμπουν σε τροποποίηση πληροφοριών των HTTP επικεφαλίδων με στόχο την εκτέλεση απομακρυσμένου bufferoverflow ή προσπάθεια για ανάλυση αιτήματος με στόχο την πιθανή εύρεση malware. Γενικά βλέπουμε ότι εδώ δίνεται έμφαση μόνο σε web application attacks αφού υπάρχουν πολλά αιτήματα HTTP που καταγράφηκαν. Άρα αυτό σημαίνει ότι κυρίως συνέβησαν επιθέσεις DoS ή malware. Θα μπορούσαμε να πούμε ότι εξαιτίας του μικρού αριθμού alarms που παρήχθησαν και το μικρό υλικό που αποθηκεύτηκε από τα συστήματα ανίχνευσης, ενισχύονται πιθανές δυσκολίες για επιτυχή αξιολόγηση της απόδοσης ενός NIDS.

Έπειτα, προχωρώντας στο CDX 2009 σύνολο δεδομένων παρατηρούμε ότι και εδώ το πρόβλημα όπως και στο KDDCUPP 99 και NSL-KDD είναι ότι δεν περιέχονται καινοτόμες επιθέσεις αλλά παλί υπάρχει μεγάλος αριθμός DoS συγκριτικά με τις υπόλοιπες επιθέσεις που πραγματοποιήθηκαν.[23] Συγκεκριμένα, ελάχιστες επιθέσεις τύπου misc (miscellaneous) και privilege escalation έγιναν. Επίσης, καμία επίθεση τύπου απομακρυσμένης εισαγωγής αρχείου (π.χ μέσα από επίθεση local file inclusion) ή ανάκτηση credentials δεν έγινε. Ωστόσο εδώ παρατηρήσαμε ότι υπήρξαν απειλές οι οποίες είχαν σχέση με παραβίαση της ασφάλειας ενός αναχώματος ασφαλείας κάτι το οποίο δε συνέβη στα προηγούμενα datasets που περιγράψαμε. Παράλληλα, εμφανίστηκαν αρκετά alerts τύπου 'BAD-TRAFFIC' στα οποία όμως δεν προσδιορίζεται το είδος του προβλήματος που αφορά τα εκάστοτε πακέτα.

Συνεχίζοντας, το LBNL/ICSI Enteprise Tracing δεν αποτελεί ένα dataset ικανό να αποτελέσει ικανοποιητικό δείγμα για αξιολόγηση NIDS ωστόσο περιέχει δείγματα που μπορούν να φανούν χρήσιμα για συγκεκριμένες περιπτώσεις συστημάτων ανίχνευσης. Ουσιαστικά

επειδή στόχος του dataset αυτού ήταν να προσομοιώσει την κίνηση ενός κοινού εταιρικού δικτύου, είναι προφανές ότι το LBNL/ICSI καλύπτει μόνο μία συγκεκριμένη γκάμα απειλών. Ένα γενικό πρόβλημα εδώ είναι ότι το σύνολο αυτό δεν περιέχει πολλά ξεχωριστά είδη επιθέσεων και όπου υπάρχουν αυτές επαναλαμβάνονται οι ίδιες πολλές φορές ανά τα διάφορα dump αρχεία. Αποτέλεσμα αυτού είναι η εμφάνιση πολλών false positives δεδομένου ότι ο συνολικός αριθμός των alerts που παρήχθησαν είναι αρκετά υψηλός (παρόλο που ο μέσος όρος ανά ώρα ήταν 6 εισβολές). Επίσης, παρατηρούμε ότι για την εξαγωγή του dataset είχε γίνει εξ αρχής κακή παραμετροποίηση ασφαλείας στα host συστήματα που απάρτιζαν το δίκτυο του αμυνόμενου.

Μετέπειτα, το Kyoto2006+ dataset συγκριτικά με τα DEFCON και τα σύνολα δεδομένων LBNL/ICSI-CDX, περιέχει shellcode exploits (524 στο σύνολο). Επίσης, δεν περιέχει επιθέσεις τύπου masquerading και γενικά όπως φαίνεται και από τη μεταβλητή count υπάρχει χαμηλός αριθμός κοινών συνδέσεων που παρέμειναν ενεργές ύστερα από 2 δευτερόλεπτα. Συνεπώς κάτι τέτοιο αποτελεί ισχυρή ένδειξη ότι δεν πραγματοποιήθηκαν πολλές επιθέσεις DoS όπως στο KDDCUP99 ή στα DARPA datasets. Ακόμα μέσα από την ύπαρξη του χαρακτηριστικού malware_detection βλέπουμε ότι δεν ανιχνεύθηκε καμία απειλή τύπου malware το οποίο αποτελεί μειονέκτημα. Επίσης, όπως και στο CDX 2009 αλλά και στο LBNL/ICSI Enterprise Tracing παρατηρούμε ότι εμφανίζονται τα ίδια alerts επαναλαμβανόμενες φορές. Ειδικότερα, υπήρξαν πολλά alerts αλλά λίγα REJ flags. [24] Αυτό σημαίνει ότι ο παραλήπτης δεν αρνήθηκε ή μπλόκαρε πολλά από τα αιτήματα που έλαβε. Συνεπώς, δεν είχε τους πλέον βέλτιστους μηχανισμούς ασφάλειας. Γενικά το Kyoto2006+ αποτελεί ένα αρκετά ποιοτικό dataset τόσο λόγω και του μεγέθους του όσο και του μεγάλου αριθμού επιθέσεων που πραγματοποιήθηκαν αφού η μεταβλητή label κατά μέσο όρο έχει αρνητική τιμή το οποίο σημαίνει ότι όταν είναι -1 ανιχνεύθηκε κάποια επίθεση ενώ θετική τιμή 1 σημαίνει φυσιολογική κίνηση. Επίσης, από τους πίνακες είδαμε ότι παρατηρήθηκε μικρός αριθμός RSTOS0 απαντήσεων.

Τώρα όσον αφορά το ISOT Botnet 2010, είδαμε ότι ανιχνεύθηκαν μεγάλα ποσά TCP συνδέσεων κάτι που σημαίνει ότι θα υπήρξε υψηλή καθυστέρηση στο εκάστοτε δίκτυο προορισμού. Παράλληλα, δεν εντοπίστηκαν fragmented packets όπως στα DARPA/DEFCON datasets αλλά υπήρχαν αλλεπάλληλες αποστολές μηνυμάτων TCP (ακόμα και ύστερα από ανεπιτυχή προσπάθεια σύνδεσης). Αυτό υποδεικνύει όπως και στα KDDCUP99-NSLKDD σύνολα ένδειξη ύπαρξης DoS επιθέσεων. Ένα μειονέκτημα του ISOT Botnet είναι ότι υπήρξαν λίγοι πιθανοί στόχοι (όπως φαίνεται από τις IP διευθύνσεις) και έγιναν πολλές επιθέσεις στα ίδια

θύματα. Συνεπώς, αυτό μειώνει την αξιοπιστία του dataset αφού το εκάστοτε θύμα ενδεχομένως να είχε μία συγκεκριμένη ευπάθεια την οποία να εκμπεταλλευόταν συνεχώς ο επιτιθέμενος. Γενικά, πέραν από επιθέσεις άρνησης υπηρεσιών και επιθέσεις σε γνωστά πρωτόκολλα που πραγματοποιήθηκαν, δεν παρατηρήθηκε άλλη κακόβουλη συμπεριφορά η οποία να προήλθε από τα botnets.

Στη συνέχεια, στο CAIDA DNS Root/gTLD RTT dataset γίνεται μελέτη και καταγραφή μέτρησης μόνο ως προς τα χαρακτηριστικά που σχετίζονται με το DNS πρωτόκολλο με αποτέλεσμα να είναι άγνωστες οι πληροφορίες κίνησης οι οποίες αφορούν τα υπόλοιπα πρωτόκολλα. [25] Γενικά, παρατηρούμε εδώ ότι όπως και στο KDDCUP99 έτσι και στα DARPA datasets η κύρια απειλή επιτιθέμενου που εμφανίζεται είναι οι DoS επιθέσεις αφού ένας από τους παράγοντες που στήθηκε το DNS Root ήταν η μελέτη της συμφόρησης και της επίδρασης που έχουν πιθανές καθυστερήσεις μετάδοσης και ανάκτησης των διαφόρων δικτυακών αιτήματων. Επίσης, δεν περιέχονται (πέρα από χαρακτηριστικά επίδοσης δικτύου) ιδιότητες που να προσδιορίζουν τα ατομικά χαρακτηριστικά του κάθε πακέτου που ανταλλάσσεται (π.χ IP διευθύνσεις για περιπτώσεις ανάλυσης συμπεριφοράς εισβολέα). Αντίθετα εξετάζεται η συνολική αποδοτικότητα της δικτυακής υποδομής μόνο. Σε αυτό οφείλεται εξάλλου και η αποκλειστική χρήση του εργαλείου NeTraMet. [26] [27]

Στο επόμενο βήμα θα περιγράψουμε αδυναμίες των υπολοίπων CAIDA datasets. Αρχικά, στο CAIDA 'DDOS Attack 2007' παρατηρούμε όπως και στα DEFCON όσο και στο KDDCUP99 μόνο προσπάθειες για εκτέλεση DoS επιθέσεων (με μέτρια επικινδυνότητα απειλής στο Snort και ταξινόμηση απειλής 'Potential Bad-traffic') αφού αυτός ήταν από την αρχή ο σκοπός συλλογής δεδομένων του dataset. Ταυτόχρονα, η κίνηση η οποία ανίχνευθηκε είναι ανώνυμη δεδομένου ότι παράχθηκε μέσα από χρήση anonymizers και σε συνδυασμό με τη διενέργεια αυτής σε ένα αποκεντρωμένο περιβάλλον, καθιστά δύσκολο το έργο του ελέγχου και του εντοπισμού της πραγματικής διεύθυνσης των εισβολέων. Επίσης, παρατηρήσαμε ότι δεν υπάρχει φυσιολογική-background κίνηση δεδομένου ότι η πλειοψηφία της κίνησης αποτελείται από κακόβουλα πακέτα. Συγκεκριμένα, από όλα τα παραπάνω datasets που περιγράψαμε είναι το μοναδικό στο οποίο το 99% των συμβάντων που πραγματοποιήθηκαν εμφανίζονται ταυτόχρονα ως alerts δηλαδή αποτελούν κακόβουλη ή ύποπτη δραστηριότητα.

Τέλος, στο CAIDA Backscatter όπως και στο DNS Root/gTLD RTT δίνεται επίσης έμφαση σε επιθέσεις άρνησης υπηρεσιών με συνέπεια να υπάρχει έλλειψη διαφορετικού είδους απειλών. Παράλληλα, σε αυτό το dataset υπάρχει δυσκολία ανάλυσης του προφίλ, της

συμπεριφοράς αλλά και των χαρακτηριστικών του επιτιθέμενου. Αυτό οφείλεται στο γεγονός ότι δεν υφίσταται αλληλεπίδραση μηνυμάτων μεταξύ αμυνόμενου και εισβολέα, αφού στο σύνολο αυτό έχουν καταγραφεί μόνο τα responses του θύματος.

Έτσι, κατά αυτό τον τρόπο παραμένουν άγνωστες οι πληροφορίες σχετικά με τις ιδιότητες του εκάστοτε attack network. Γενικά όλα τα CAIDA datasets υστερούν στο ότι παρέχουν σε βέλτιστο βαθμό μόνο ένα μέρος απειλών οι οποίες δεν μπορούν να χρησιμοποιηθούν πλήρως για την αξιολόγηση της επιδόσης ενός NIDS.

Κεφάλαιο 5ο

Ανασχεδιασμός και Βελτίωση

Αφού περιγράψαμε επιτυχώς τα παραπάνω datasets στο επόμενο βήμα δημιουργήσαμε ένα νέο σύνολο δεδομένων το οποίο θα πληρεί και θα προσφέρει δυνατότητες αξιολόγησης παρόμοιες με τα προηγούμενα αλλά και παράλληλα θα βελτιστοποιεί υπάρχουσες αδυναμίες των προαναφερθέντων datasets.

5.1 Δομή και Χαρακτηριστικά Νέου Dataset

Όπως παρατηρήσαμε η πλειοψηφία των προηγούμενων datasets όπως το KDDCUP99, DARPA και CAIDA εστιάζουν σε επιθέσεις DoS. Στην προκειμένη περίπτωση εμείς πραγματοποιήσαμε τόσο επιθέσεις DoS όσο και επιθέσεις τύπου malware, bufferoverflow, shellcode για τις οποίες δεν δόθηκε ιδιαίτερη έμφαση στα προηγούμενα . Επίσης, στόχος στο νέο σύνολο είναι η μείωση των παραγόμενων false positives ύστερα από την εξέταση και ανάλυση του συνόλου αυτού από δικτυακά IDS.

Το dataset αυτό απαρτίζεται από μία ποικιλία κακόβουλης δραστηριότητας η οποία εμφανίζεται αρκετά συχνά σε σύγχρονα δίκτυα. Το καινούργιο σύνολο δημιουργήθηκε συνθετικά δηλαδή με τεχνητό τρόπο εκμπεταλλευόμενοι παράλληλα τεχνικές προσομοίωσης που εγκαθιδρύσαμε. Γενικά, για την παραγωγή τόσο της κακόβουλης όσο και της φυσιολογικής κίνησης χρησιμοποιήσαμε traffic generators από τους οποίους, τους περισσότερους τους ενσωματώσαμε σε ένα bash script για αυτοματοποιημένη εκτέλεση αποστολής πακέτων μέσα από επαναλαμβανόμενες επαναλήψεις σε εντολές CLI (Command-Line Interface) των εκάστοτε εργαλείων.

Παράλληλα, με τη μέθοδο αυτή είμασταν σε θέση να ορίζουμε όσο το δυνατόν περισσότερα τυχαία χαρακτηριστικά σε κάθε κίνηση τα οποία όμως να μην αποκλίνουν σημαντικά από τις εκάστοτε ιδιότητες που αφορούν πακέτα ροής δεδομένων σε πραγματικό περιβάλλον δικτύου από τη στιγμή που το dataset αυτό θα χρησιμοποιηθεί για αξιολόγηση συστημάτων ανίχνευσης εισβολών.

Γενικά κατά τη δημιουργία της κίνησης χρησιμοποιήθηκαν τα ακόλουθα εργαλεία:

- Nemesis (φυσιολογική-κακόβουλη κίνηση)
- Hping
- Mausezhan (παραγωγή τυχαιοποιημένης κίνησης)
- SendIP
- Yersinia (κυρίως για αποστολή μαζικών DHCP και BPDU requests)
- Nping
- Scapy (DOS-SYN flooding και python scripts)
- Bash script για παραγωγή malformed κίνησης
- Rule2alert (διαδικασία triggering των εκάστοτε κανόνων του Snort)
- IDSwakeup
- PackETH (κυρίως για παραγωγή φυσιολογικής κίνησης)
- Ostinato (για ενσωμάτωση κακόβουλου hexdump)
- bitwist (για DOS επιθέσεις)
- PFsend
- Snot
- Stick (για έλεγχο υπογραφών-κανόνων του NIDS)
- Ethernet packet bombardier

Γενικά δόθηκε έμφαση στην ύπαρξη κάθε είδους πιθανού πρωτοκόλλου με εξαίρεση το ARP πρωτόκολλο το οποίο εμφανίζεται λιγότερο συχνά σε σχέση με τα υπόλοιπα. Από την άλλη μεριά ιδιαίτερη σημασία δόθηκε και στα δευτερεύοντα χαρακτηριστικά ενός πακέτου (π.χ time to live, type of service, flags, κ.α) τα οποία όμως αποτελούν καθοριστικό παράγοντα για μία πιθανή εκτύπωση ενός alert από ένα δικτυακό IDS. Στη συνέχεια προκειμένου να μπορούμε να περιγράψουμε καλύτερα το dataset αυτό στην παρούσα εργασία, εκτυπώσαμε την κίνηση σε διαφορετικά είδη αρχείων κίνησης δηλαδή pcap, tcpdump, snort logs, data κ.α. Τα αρχεία data παρήχθησαν εικονικά μέσω ενός προσομοιωτή NIDS που δημιουργήσαμε (το οποίο περιγράφουμε σε επόμενο κεφάλαιο) στον οποίο ενεργοποιήσαμε δυνατότητα για αυτοματοποιημένη εξαγωγή dataset. Η πλειοψηφία των αρχείων είναι pcap επειδή αυτό το είδος περιέχει πλήρως όλες τις πιθανές πληροφορίες και χαρακτηριστικά για τα εκάστοτε πακέτα κίνησης.

Παράλληλα, η χρήση pcap αρχείων βοηθάει την ανάγνωση τους από ποικίλα πρόγραμματα ανάλυσης κίνησης όπως το Wireshark, το WinDump, το Analyzer, το EtherApe κ.α. Ακόμα, ο χωρισμός του dataset σε μικρότερα κομμάτια εξυπηρετεί επίσης τη γρηγορότερη σάρωση του κάθε αρχείου από εργαλεία εξόρυξης δεδομένων ή αναλυτές δικτυακής κίνησης ειδικότερα από τη στιγμή που το συνολικό μέγεθος του είναι περίπου 12 GB. Για κάθε εισερχόμενο ή εξερχόμενο πακέτο το οποίο διερχόταν, χρησιμοποιήσαμε τόσο το εργαλείο Snort όσο και το Bro με στόχο την αποδοτικότερη και βέλτιστη ανίχνευση απειλών για την παραχθείσα κίνηση.

Το σύνολο δεδομένων είναι χωρισμένο σε 6 υποκατηγορίες δεδομένων:

DATA01: Σε αυτήν την κατηγορία περιέχεται μόνο κακόβουλη κίνηση η οποία παράχθηκε μέσω των εργαλείων rule2alert, IDSwakeup και Stick στα οποία φορτώσαμε κανόνες που είχαν τεθεί για το σύστημα ανίχνευσης εισβολών και με τη σειρά τους αυτά παρήγαγαν πακέτα τα οποία αντιτάσσουν-παραβαίνουν αυτούς τους κανόνες με στόχο την πρόκληση security alarms από το σύστημα καταγραφής και παρακολούθησης. Επίσης σε αυτό το σύνολο αρχείων κίνησης, παρατηρείται μέγιστος αριθμός πέντε hosts ανά pcap και τις περισσότερες φορές 2 hosts ανά αρχείο έτσι ώστε να μπορεί να αναλύεται πλήρως και με εύκολο τρόπο η συμπεριφορά του επιτιθέμενου αλλά και τα responses του host-θύματος κάθε στιγμή χωρίς να παρεμβαίνουν packet frames τρίτων που θα δυσκόλευαν την ανάλυση της δραστηριότητας.

DATA02: Εδώ παρατηρείται τυχαιοποιημένη κίνηση στην οποία εμπεριέχεται και φυσιολογική κίνηση για ευκολία αντίληψης και κατανόησης των επιθέσεων ως μέρος δραστηριοτήτων όπως αυτές συμβαίνουν σε πραγματικό δικτυακό περιβάλλον. Η κίνηση εδώ παράχθηκε μέσα από εκτέλεση script σε ποικίλα πρωτόκολλα με όσο το δυνατόν περισσότερα και διαφορετικά χαρακτηριστικά πακέτων και συνολικό αριθμό hosts 33. Στόχος αυτού είναι η μελέτη των διαφορετικών παραγόντων που επηρεάζουν μία επίθεση. Εδώ πέρα από pcap αρχεία, περιέχονται και tcpdump αρχεία αλλά και αρχεία κειμένου παραγόμενα από το tcpdump με σκοπό τη γρηγορότερη ανάλυση της πληροφορίας που φέρει το κάθε πακέτο π.χ μέσα από CLI. Παράλληλα, υπάρχουν και logs παραγόμενα από το snort έτσι ώστε ο χρήστης να μπορεί να βρίσκει τόσο τα αποτυπώματα κίνησης όσο και τα alerts μέσα από το ίδιο εργαλείο. Επίσης υπάρχουν αρχεία trace τα οποία αποτελούν αρχεία κειμένου που έχουν εξαχθεί μέσω του

tshark χρησιμοποιώντας κατάλληλα φίλτρα κίνησης (π.χ ip.dst , eth.len). Ακόμα, υπάρχουν εδώ αρχεία τύπου data τα οποία εμφανίζονται σε μορφή CSV και έχει γίνει χρήση χαρακτηριστικών του KDDCUP99 dataset αφού είναι το πιο διαδεδομένο σύνολο δεδομένων το οποίο βοηθάει τόσο στην ευκολότερη ανάλυση όσο και στη μελέτη ανωμαλιών που υφίστανται τα αρχεία αυτά σε σχέση με μέχρι τώρα άγνωστα χαρακτηριστικά πακέτων για μελέτη καινούργιων επιθέσεων. Τα είδη των ανωμαλιών που προκύπτουν αφορούν κυρίως alerts που παράγονται από το Snort. Επιπλέον, έχουν προστεθεί κάποια επιπλέον attributes σε αυτά όπως το RTT (round-trip time, deviation time, κ.α) που μπορούν να βοηθήσουν στην κατανόηση επιθέσεων που δεν είναι απόλυτα εμφανείς από το σύστημα ανίχνευσης εισβολών. Επίσης, εύκολα μπορεί να πραγματοποιηθεί στα data αρχεία ανάλυση της πληροφορίας αυτών ανά χαρακτηριστικό από εργαλεία όπως το Weka καθώς και εξόρυξη νέων κανόνων. Γενικά η ύπαρξη πολλαπλών αρχείων κίνησης διευκολύνει τη χρήση εργαλείων σύγκρισης όπως το Bindiff για αυτοματοποιημένη έυρεση διαφορών μεταξύ αυτών.

DATA03: Αυτό το σύνολο αρχείων pcap περιέχει κίνηση καθαρά τύπου Denial of Service η οποία έχει πραγματοποιηθεί μέσω των εργαλείων Scapy και Yersinia. Μέσω του πρώτου εργαλείου εκτελέστηκαν SYN floods και μέσω του δεύτερου έγιναν επιθέσεις DoS στα πρωτόκολλα CDP (δημιουργώντας συσσώρευση πακέτων στο CDP table) , DHCP, VTP, HSRP, DTP, 802.1Q και STP (αποστέλλοντας μαζικά BPDUs).

DATA04: Εδώ κάθε αρχείο κίνησης περιέχει κίνηση η οποία είναι ανώνυμη, δηλαδή η πηγή και η ταυτότητα προέλευσης του εκάστοτε πακέτου αλλάζει διαρκώς ύστερα από μετάβαση του κάθε μηνύματος σε ένα διαμεσολαβητή (proxy). Περιέχονται μόνο IPv4 πακέτα κίνησης και ουσιαστικά γίνεται χρήση δύο anonymity networks τα οποία αποτελούν μέρος του συνόλου μαζί με τα υπόλοιπα δίκτυα που χρησιμοποιήθηκαν για την παραγωγή των υπόλοιπων δεδομένων.

DATA05: Στο σύνολο αρχείων αυτό, περιέχεται κίνηση τύπου malformed δηλαδή υπάρχουν πακέτα ροής δεδομένων τα οποία έχουν κακή και λανθασμένη δομή σχηματισμού, η οποία μπορεί είτε απρόκλητα να οδηγήσει σε απειλές είτε όχι. Για παράδειγμα, ένα είδος malformed πακέτων είναι αυτά τα οποία περιέχουν tcp length μεγαλύτερο από το επιτρεπτό που μπορεί να περάσει μέσα από το TCP πρωτόκολλο. Στην πράξη η υλοποίηση και δημιουργία της κίνησης

αυτής έγινε μέσα από ένα νέο bash script στο οποίο παραγόταν κάθε φορά τυχαίο hexdump το οποίο οδηγούσε όπως φαίνεται και στα αποτυπώματα κίνησης, πολλά πακέτα να περιέχουν είδος πρωτοκόλλου το οποίο δεν υπάρχει ή στην θέση των διευθύνσεων πηγής και προορισμού να εμφανίζονται MAC addresses.

GLOBAL: Ύπαρξη εννιαίου alert dataset όπως αποτυπώθηκε σε μία δεδομένη στιγμή από το NIDS σε comma-seperated μορφή με συγκεκριμένα attributes για εύκολη φόρτωση και ανάλυση των alarms από εργαλεία εξόρυξης δεδομένων ή table-based software (π.χ Excel). Παράλληλα, έχει γίνει ενσωμάτωση όλων των κανόνων που χρησιμοποιήθηκαν κατά την ανίχνευση από το NIDS σε ένα αρχείο τύπου rules έτσι ώστε να διαπιστωθεί η ύπαρξη πιθανών alerts τα οποία έπρεπε να παραχθούν αλλά τελικά δεν εκτυπώθηκαν (false negatives). Επίσης, υπάρχουν αρχεία configuration του Snort, του Bro και του Barnyard έτσι ώστε να φαίνεται αν υπάρχουν κανόνες που αγνοήθηκαν ή απενεργοποιήθηκαν κατά την παραμετροποίηση του δικτυακού συστήματος ανίχνευσης παρεισφρήσεων.

Γενικά, στόχος του dataset αυτού είναι να περιέχει δεδομένα όπως αυτά θα μπορούσαν να καταγραφούν στην πράξη από ένα δικτυακό IDS. Στην περίπτωση μας, η καταγραφή των δεδομένων σε όλες τις κατηγορίες δεδομένων αφορά 5 εικονικά δίκτυα Η/Υ και 2 δίκτυα ανωνυμίας (για το DATA04) στα οποία υπάρχουν συνολικά οι παρακάτω hosts όπως εμφανίζονται και στα αρχεία κίνησης.

Basic Hosts: 8.0.0.1, 12.0.0.10-80, 18.0.0.1-19, 30.0.0.1-11, 10.0.0.10-90, 24.0.0.1-10

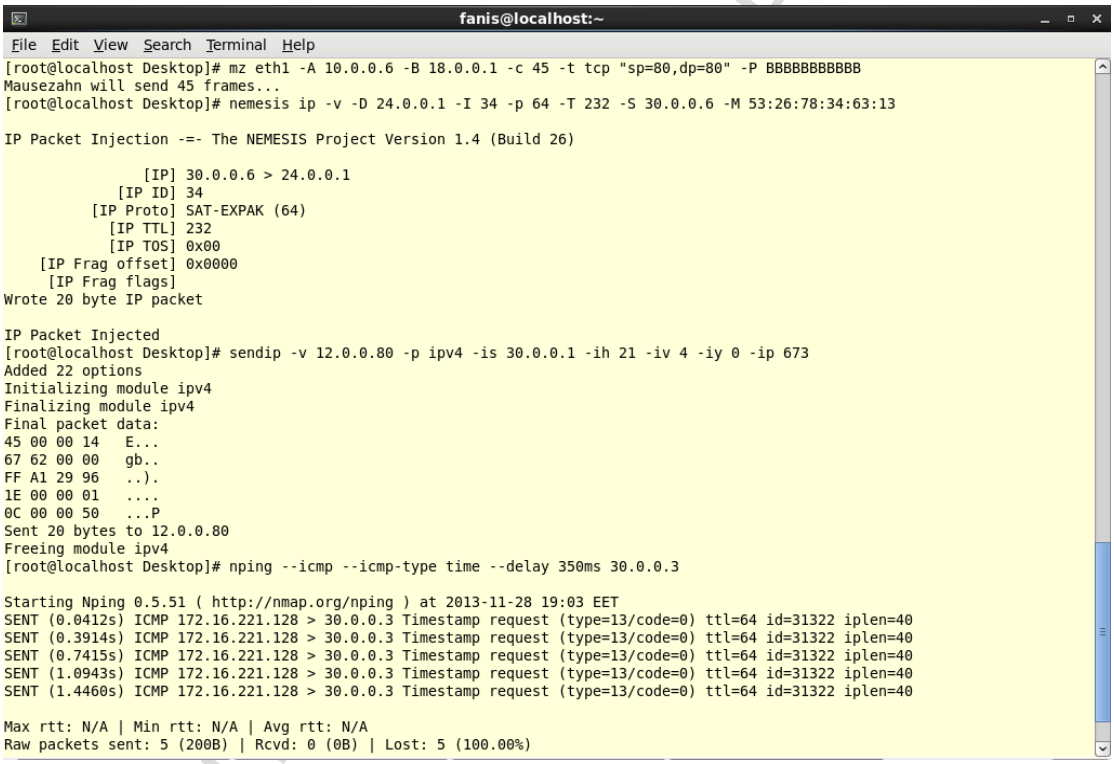
Reserved Hosts: 1.1.1.1, 2.2.2.2, 127.0.0.1

Anonymous Hosts: τυχαιοποιημένες IPv4 διευθύνσεις

5.2 Παραγωγή Κίνησης

Παραδείγματα εκτελέσεων ενσωματωμένων σε bash με χρήση των εργαλείων που αναφέραμε για δημιουργία τυχαιοποιημένης κίνησης είναι τα εξής:

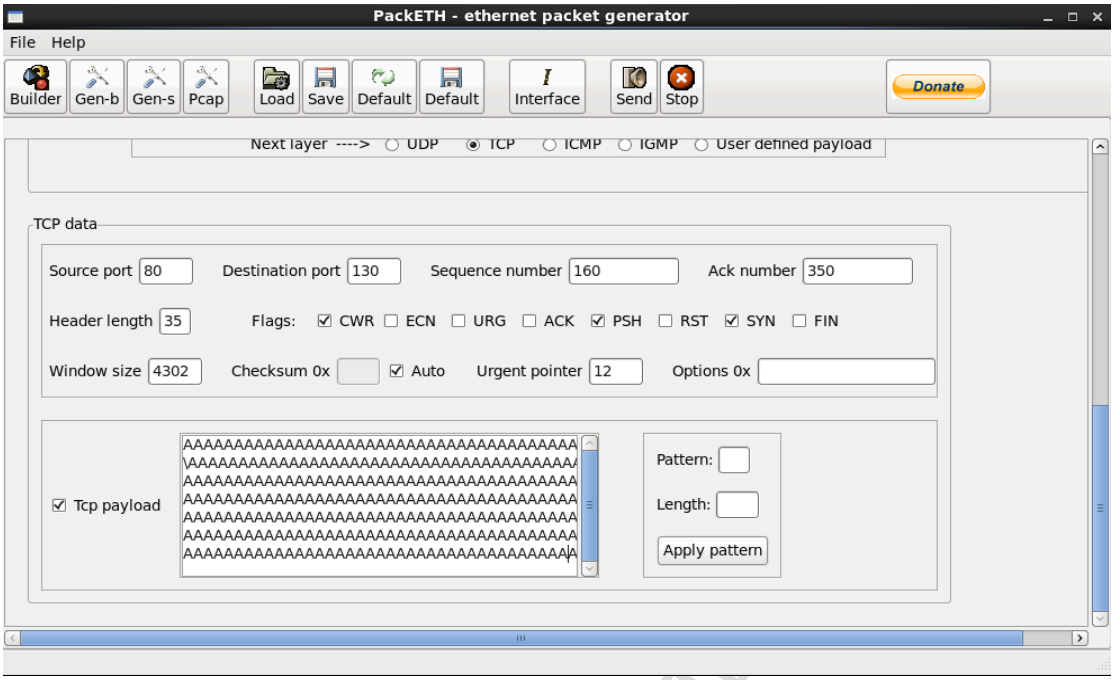
```
nemesis ip -v -D 24.0.0.1 -I 34 -p 64 -T 232 -S 30.0.0.6 M
53:26:78:34:63:13

sendip -v 12.0.0.80 -p ipv4 -is 30.0.0.1 -ih 21 -iv 4 -iy 0 -ip 673

mz eth1 -A 10.0.0.6 -B 18.0.0.1 -c 45 -t tcp “sp=80,dp=80” -P BBBBBBBBBBBB

nping --icmp --icmp-type time --delay 350ms 30.0.0.3

hping2 12.0.0.10 -0 -a 18.0.0.2 -c 37 -t 22 -N 55 -H 63 -r -f -x -g 88 -m
18 -o 48 -G
```



Εικόνα 9: Παραγωγή Κίνησης Μέσω των Εργαλείων Mausezhan, SendIP, Nemesis και Nping

```
./packETHcli -i eth1 -m 2 -d 56 -n 128 -s "$packeths1 $packeths2
$packeths3" -p 22 -f 682959106862.pcap"
```

Στο εργαλείο packETH κάναμε επίσης χρήση και του GUI.



Εικόνα 10: Ενδεικτική Χρήση του Εργαλείου PackETH (GUI)

Στα αρχεία τύπου data που ανήκουν στο σύνολο αρχείων DATA02 εισάγαμε όλα τα χαρακτηριστικά του KDDCUP99 με τη διαφορά ότι αλλάξαμε τις πληροφορίες που περιέχει το attribute 'class' και προσθέσαμε τα παρακάτω 7 επιπλέον χαρακτηριστικά:

• average_rtt → Μέσο round-trip time του κάθε πακέτου

• xssdetect → Ενδεικτική τιμή για εμφάνιση cross-site scripting επίθεσης (0-100). Αν είναι πάνω από 90 τότε παράγεται class με ένδειξη "xss_attack".

• iplen → IP length του πακέτου

• ethlen → Ethernet length του πακέτου

• src_port → Θύρα προέλευσης

• dst_port → Θύρα προορισμού

• malwaredetect → Ενδεικτική τιμή για εμφάνιση malware (0-100). Αν είναι πάνω από 90 τότε παράγεται class με ένδειξη "malware".

Συγκεκριμένα όλες οι πιθανές ενδείξεις που εμφανίζονται στο class είναι οι εξής:

Πίνακας 40: Λίστα Χαρακτηριστικών που Περιέχονται στα Αρχεία Τύπου data

malware	successful_admin	attempted_dos
xss_attack	shellcode_detect	attempted_recon
attempted_admin	policy_violation	bad_unknown
attempted_user	inappropriate_content	default_login_attempt
trojan_activity	denial_of_service	successful_user
unsuccessful_user	misc_attack	rpc_portmap_decode
web_application_attack	non_standard_protocol	successful_recon_largescale
successful_recon_limited	suspicious_filename_detect	suspicious_login
system_call_detect	unusual_client_port_connection	web_application_activity
icmp_event	misc_activity	network_scan
buffer_overflow	protocol_command_decode	string_detect
unknown	tcp_connection	normal

Έπειτα μέσω του Weka και της χρησιμοποίησης του C4.5 αλγορίθμου καθώς και των rule classifiers OneR, PART, ZeroR και M4Rules εξάγαμε διάφορους κανόνες όπως:

```
IF service = smtp AND num_outbound_cmds <= 239 AND logged_in = 1 AND
land = 0 AND is_host_login = 1 AND
dst_host_serror_rate > 0.78 THEN misc_attack

IF service = sql_net AND num_compromised > 62 AND
serror_rate > 0.45 AND average_rtt > 93 AND
wrong_fragment > 103 THEN protocol_command_decode

IF service = ntp_u AND land = 1 AND average_rtt > 411 AND
duration <= 158 THEN attempted_admin

IF service = pop_3 AND srv_count > 698 AND protocol_type = udp THEN malware

IF service = private AND diff_srv_rate > 0.39 AND malwaredetect > 30 AND
dst_host_srv_diff_host_rate <= 0.4 AND duration <= 74 THEN inappropriate_content
```

Πειραματικοί Κανόνες που Αφορούν Καταστάσεις Εισβολών

Τα ακόλουθα χαρακτηριστικά που παρουσιάζονται στον πίνακα 41 ορίστηκαν στο τρέχον alert dataset.

Πίνακας 41: Λίστα Χαρακτηριστικών του Συνόλου Δεδομένων Εισβολών

Timestamp	Protocol	Ethernet Source	ICMP Type
Signature Generator	Source Address	Ethernet Destination	ICMP Code
Signature ID	Source Port	Ethernet Type	ICMP ID
Message	Destination Address	Ethernet Length	ICMP Sequence
TCP Flags	TCP Acknowldegment	TCP Window	TR Header
Time to Live	Type of Service	UDP Length	Revision
Unique ID	DGM Length	IP Length	

Για την εκτύπωση του παραπάνω συνόλου αρχικά χρησιμοποιήσαμε το Barnyard στο Snort και παραμετροποιήσαμε το αρχείο snort.conf ώστε να είναι ενεργοποιημένη η λειτουργία output alert_csv. Στη συνέχεια αφού δημιουργήθηκε το αρχείο, εκτελέσαμε την εντολή sed.

```
sed 's/,,/,0,0/g' /var/log/alert.csv > alert.csv
```

για αντικατάσταση κενών κομμάτων.

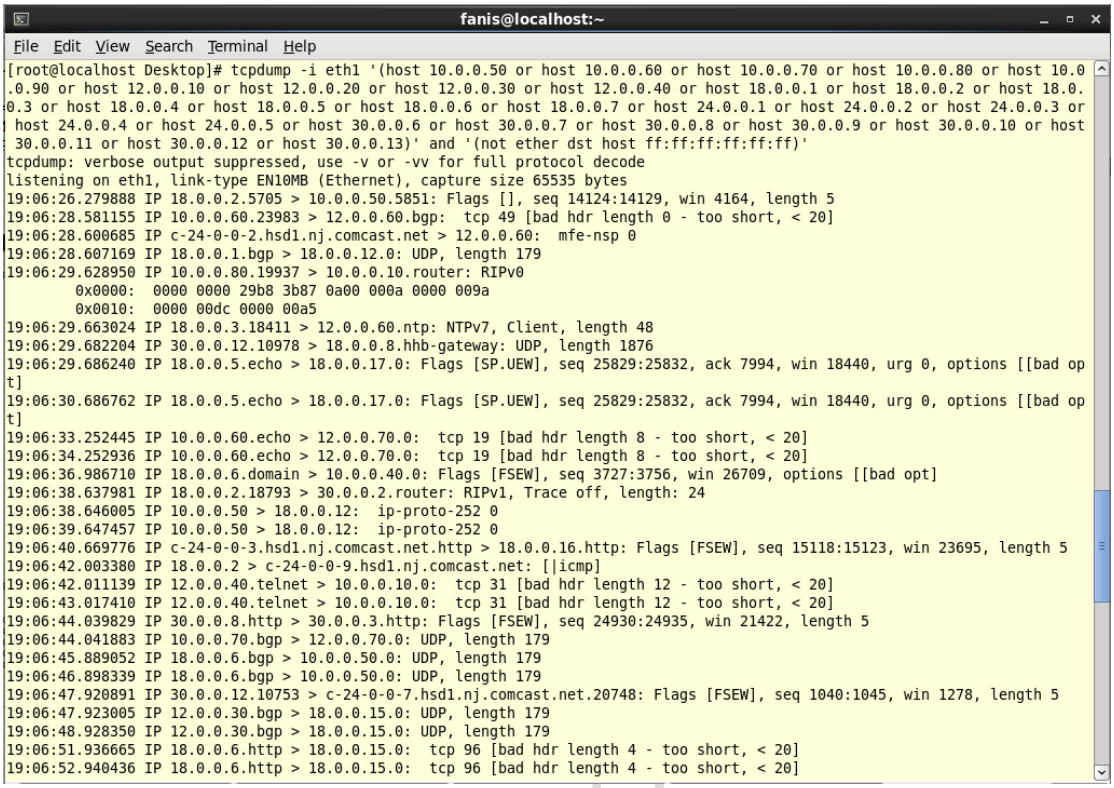
Στα αρχεία trace για την εκτύπωση της κίνησης εισάγαμε την ακόλουθη εντολή για εκτέλεση του tshark:

```
tshark -i en0 -T fields -e frame.number -e frame.time -e eth.src -e eth.dst
-e ip.src -e ip.dst -e ip.proto -e ip.checksum -e ip.dsfield -e ip.dst_host
-e ip.flags -e ip.frag_offset -e ip.fragmets -e ip.tos -e ip.ttl -e eth.len
-e eth.dst -E header=y -E separator=, -E quote=d -E occurrence=f >
test.trace
```

όπου τα χαρακτηριστικά που χρησιμοποιήθηκαν είναι αυτά που βρίσκονται μετά το argument –e.

Έπειτα, παρατηρούμε ότι υπάρχουν και κάποια άλλα αρχεία στο DATA02 με κατάληξη log τα οποία δεν είναι αποτυπώματα του Snort αλλά του Tcpdump. Συγκεκριμένα εδώ τρέξαμε μία διαδικασία επανάληψης for σε shell script στην οποία ενσωματώσαμε το ακόλουθο:

```
tcpdump –i eth1 '(host 10.0.0.50 or host 10.0.0.60 or host 10.0.0.70 or
host 10.0.0.80 or host 10.0.0.90 or host 12.0.0.10 or host 12.0.0.20 or
host 12.0.0.30 or host 12.0.0.40 or host 18.0.0.1 or host 18.0.0.2 or host
18.0.0.3 or host 18.0.0.4 or host 18.0.0.5 or host 18.0.0.6 or host
18.0.0.7 or host 24.0.0.1 or host 24.0.0.2 or host 24.0.0.3 or host
24.0.0.4 or host 24.0.0.5 or host 30.0.0.6 or host 30.0.0.7 or host
30.0.0.8 or host 30.0.0.9 or host 30.0.0.10 or host 30.0.0.11 or host
30.0.0.12 or host 30.0.0.13 or host 8.0.0.1 or host 1.1.1.1 or host 2.2.2.2
or host 127.0.0.1)' and '(not ether dst host ff:ff:ff:ff:ff:ff)' -c $num -
vvv –w
```



Εικόνα 11: Καταγραφή Τεχνητά Παραγόμενης Δικτυακής Κίνησης Μέσω του tcpdump

Στο σύνολο αρχείων DATA05 για κάθε εξαγόμενο αποτέλεσμα του script χρησιμοποιήθηκε η εντολή text2pcap για ενσωμάτωση των malformed πακέτων στο Wireshark.

```
text2pcap hex.txt hex.pcap
```

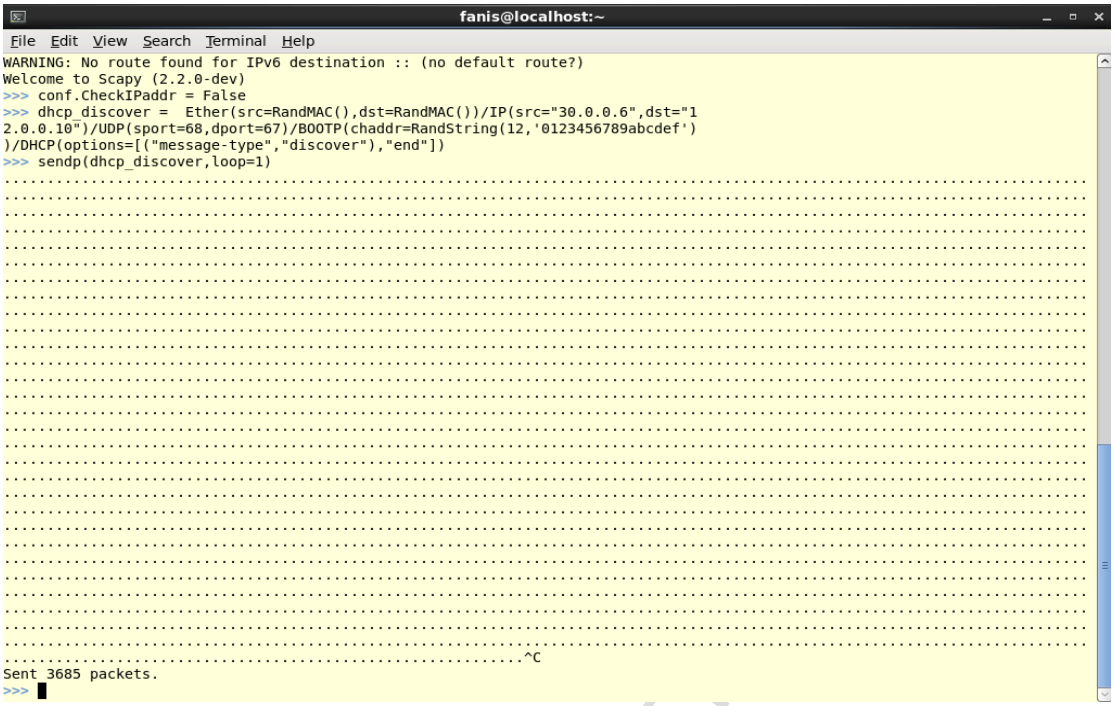
Για τη συνένωση πολλών παραγόμενων αρχείων pcap σε ένα εννιαίο εισάγαμε:

```
mergecap –w test.pcap he0x.pcap 011841373251.pcap
```

Στο DATA04 επειδή όπως αναφέραμε έγινε χρήση δύο anonymity networks οι διευθύνσεις πηγής αλλάζουν συνεχώς. Στην πράξη αυτό το καταφέραμε μέσα από randomization στο bash script εισάγοντας το `IPx=$[$RANDOM % 255]`

Στο DATA03 για την εκτέλεση DHCP επιθέσεων μέσω του scapy εισάγαμε

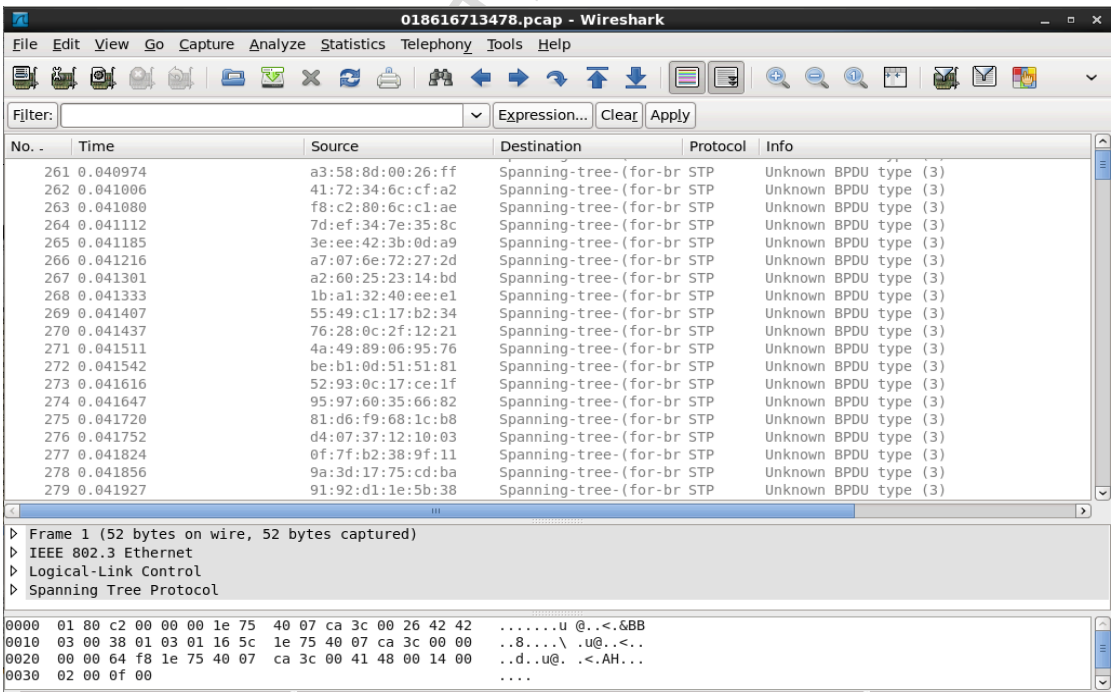
```
>>> conf.checkIPaddr = False
>>> dhcp_discover =
Ether(src=RandMAC(),dst="ff:ff:ff:ff:ff:ff")/IP(src="0.0.0.0",dst="255.255.255.255")/UDP(sport=68,dport=67)/BOOTP(chaddr=RandString(12,'0123456789abcdef'))/DHCP(options=[("message-type","discover"),"end"])
>>> sendp(dhcp_discover,loop=1)
```



Εικόνα 12: Υλοποίηση DHCP Επιθεσέων με Χρήση του Εργαλείου Scapy

και μέσω του Yersinia π.χ στα παρακάτω πρωτόκολλα:

```
yersinia cdp -attack 0 -source 73:a8:b6:88:3f:a8 -dest 21:2F:72:34:76:87 -
ttl 140 -version 1

yersinia dot1q -attack 3 -source 14:33:f5:65:78:31 -dest 42:53:63:78:23:62
```



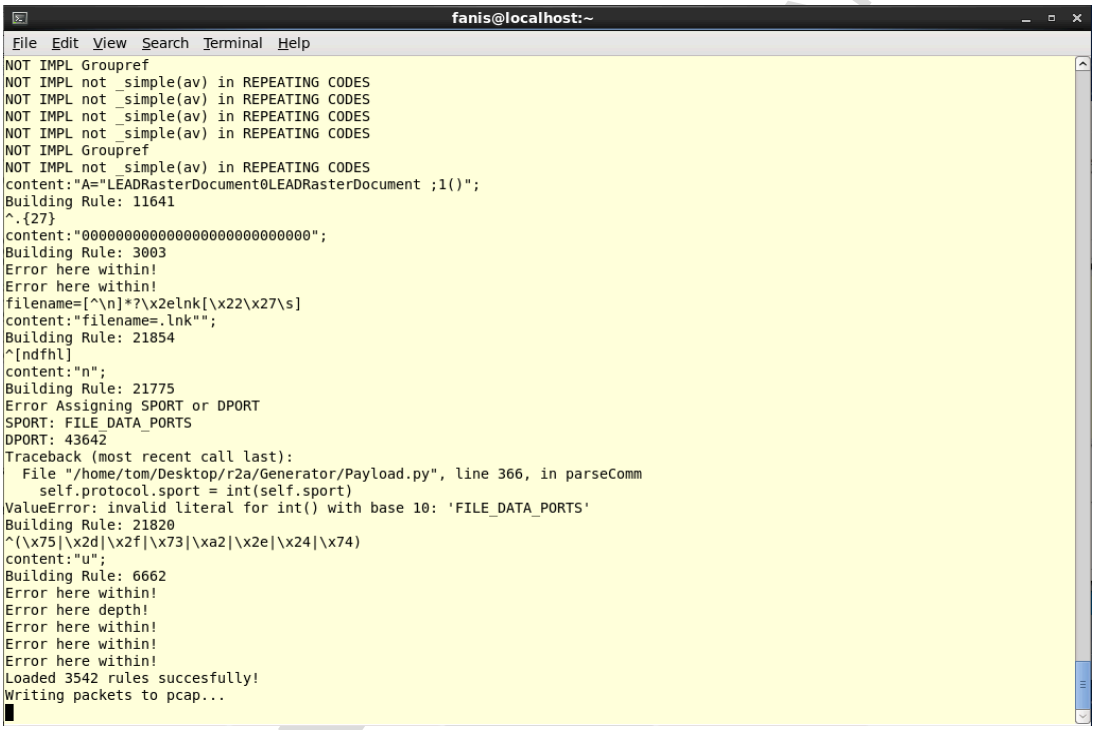
Εικόνα 13: Ενδεικτική Καταγραφή DoS Επίθεσης από το Wireshark

Επίσης μέσω του Bitwist εκτελέσαμε

```
[root$localhost Desktop]#bittwist -i eth1
/home/fanis/Desktop/031120130168.pcap -v -s 76 -l 15 -c 3 -m 3500 -r 44 -p
22 -h -i eth1
```

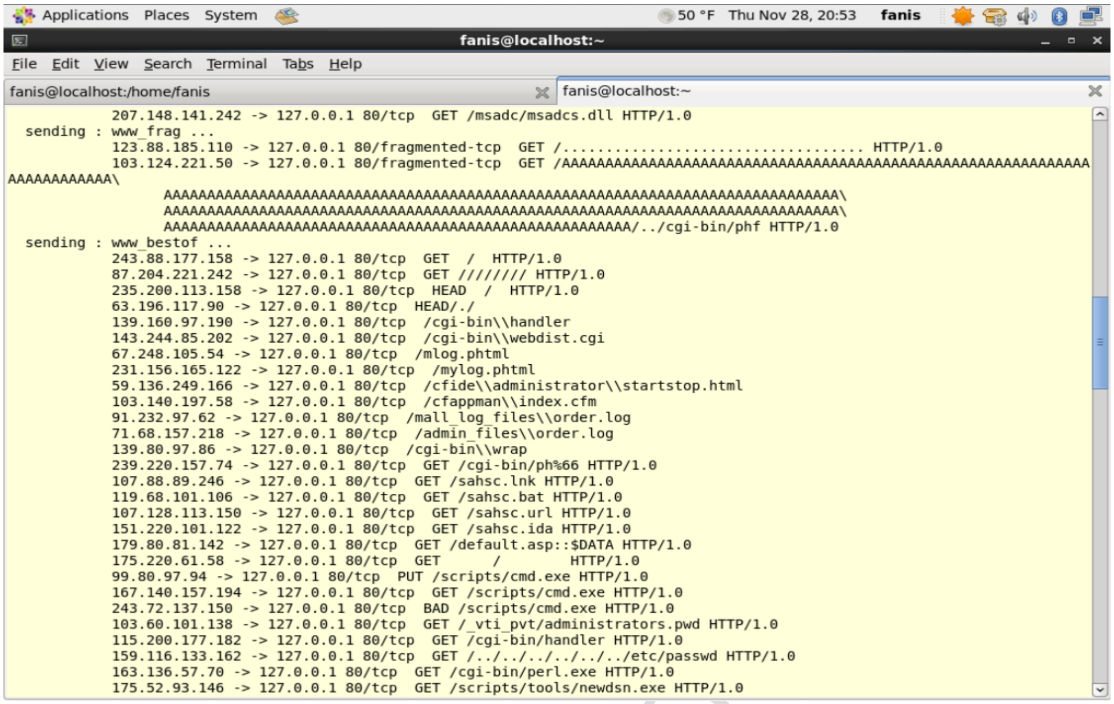
Στο DATA01 η εκτέλεση του rule2alert έγινε μέσα από python και συγκεκριμένα φορτώσαμε το αρχείο που περιείχε όλους τους πιθανούς κανόνες του Snort τους οποίους και έκανε trigger το εργαλείο:

```
[root$localhost Desktop]# python r2a.py -f /usr/local/snort/community-
rules/GENERIC.rules -c /usr/local/snort/etc/snort.conf -m 10.0.0.1 -e
30.0.0.6 -w /home/fanis/Desktop/awesome.pcap -s 42576 -n 140 -t 55
```



Εικόνα 14: Δημιουργία Κακόβουλης Κίνησης μέσω του Rule2alert

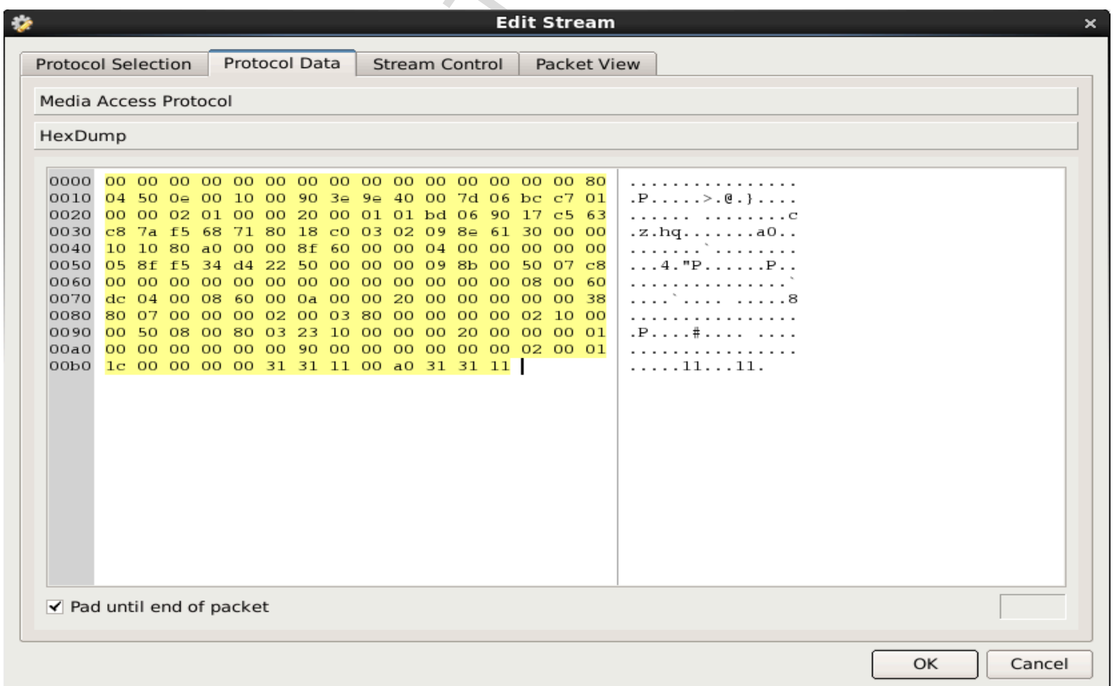
```
[root$localhost sneeze]# /.sneeze.pl -d 10.0.0.8 -s 18.0.0.1 -f
GENERIC.rules -c -1 -i eth1

[root$localhost IDSwakeup-1.0]# ./IDSwakeup 0 127.0.0.1
```



Εικόνα 15: Διαδικασία Ελέγχου Υπογραφών και Παραγωγή Κίνησης Μέσω του IDSwakeup

Παράλληλα στο σύνολο αρχείων κίνησης αυτό εισάγαμε και κακόβουλη κίνηση σε hexdump μέσω του Ostinato.

```
[**] [125:1:1] (ftp_telnet) TELNET CMD on FTP Command Channel [**]
```



Εικόνα 16: Εισαγωγή Κακόβουλου Hexdump στο Ostinato

Στη συνέχεια περιγράφουμε στον παρακάτω πίνακα τα στοιχεία κατά τα οποία υπερτερεί το νέο dataset έναντι του KDDCUP99.

Πίνακας 42: Συγκριτική Ανάλυση Νέου Dataset με KDD Cup 99

No.	Παρατήρηση	KDDCUP99	UNIPI-DATA
1	Αριθμός Δικτύων	Δύο	Επτά
2	Μοναδικές IP διευθύνσεις	11	33 (χωρίς ανώνυμη κίνηση)
3	Προσομοίωση / Εξομοίωση	Ναι	Ναι
4	Διάρκεια συλλογής δεδομένων	5 εβδομάδες	12 εβδομάδες
5	Τύποι αρχείων δεδομένων	TCPdump, BSM, dump	PCAP, DATA, Log, TCPdump, Trace, Snort
6	Attack Vectors	Probe, DoS, U2R, R2L	Flood, Priest, Probe, DoS, U2R, R2L
7	Τύποι επιθέσεων	Host-based / Network-based	Network-based
8	Είδη επιθέσεων	Προσχεδιασμένες	Τυχαιοποιημένες

Παράλληλα ένα ενδεικτικό στοιχείο από τα αρχεία data που ανακτήσαμε αφορά τα false positives/negatives και συναφείς πληροφορίες που σχετίζονται με την αξιολόγηση ενός μικρού μέρους του dataset αυτού.

Πίνακας 42: Τμηματική αξιολόγηση του νέου dataset (DATA04)

	False positives	True positives	False negatives	True negatives
842968120598.data	8170	687786	148189	1781
120496236929.data	4791	412367	378349	4493
623592306023.data	6466	541172	283010	3437
782359026890.data	7336	620944	17408	204
904503406304.data	139	10522	334	5
395010249482.data	49	4230	115	1
325896202335.data	218	18538	521	6

Η παραπάνω αξιολόγηση έγινε λαμβάνοντας ως δεδομένο πειραματικούς κανόνες που παράγαμε μέσω του Weka για το KDDCUP99. Οι κανόνες αυτοί παρουσιάζονται αναλυτικά στο παράρτημα της εργασίας.

5.3 Γραφική Απεικόνιση

Στα παρακάτω σχήματα φαίνεται μία εικονική αναπαράσταση των παραπάνω hosts μέσα από εισαγωγή τυχαίων αρχείων κίνησης του παραπάνω dataset. Τα εργαλεία που χρησιμοποιήσαμε για εξαγωγή visualizations ήταν το EtherApe, το IDS Rainstorm [28], το RUMINT και το Doomcube. Για τη χαρτογράφηση των hosts που συμμετέχουν στο σύνολο δεδομένων χρησιμοποιήσαμε το GeoIP μέσω του οποίου εισάγαμε μία λίστα από IP διευθύνσεις (DATA04) σε ένα XML αρχείο με στόχο την εξαγωγή ενός KML αρχείου. Έπειτα, φορτώσαμε το KML αρχείο στο Google Earth (με χρήση του OpenStreetmap layer) και στο Google Maps για καθολική αναπαράσταση.

Στην εικόνα 17 παρουσίαζεται ένα στιγμιότυπο από τη χρήση του εργαλείου IDS Rainstorm το οποίο και διαβάζει σε πραγματικό χρόνο την διερχόμενη κίνηση που παράγαμε μέσω των traffic generators εκείνη τη δεδομένη στιγμή. Στα σημεία όπου εμφανίζονται κάποιες πολλαπλές τελείες ή αποτυπώματα είναι η κίνηση η οποία ανιχνεύθηκε για μία δεδομένη IP διεύθυνση. Όπου το χρώμα που εμφανίζεται είναι πιο έγχρωμο τότε αυτό σημαίνει ότι είχαμε αυξημένη εισροή ή εκροή κίνησης μέσω ενός συγκεκριμένου δικτυακού σταθμού. Η συγκεκριμένη κίνηση που αναπαρίσταται είναι ανώνυμη και αφορά αποτυπώματα κίνησης του συνόλου αρχείων DATA04.

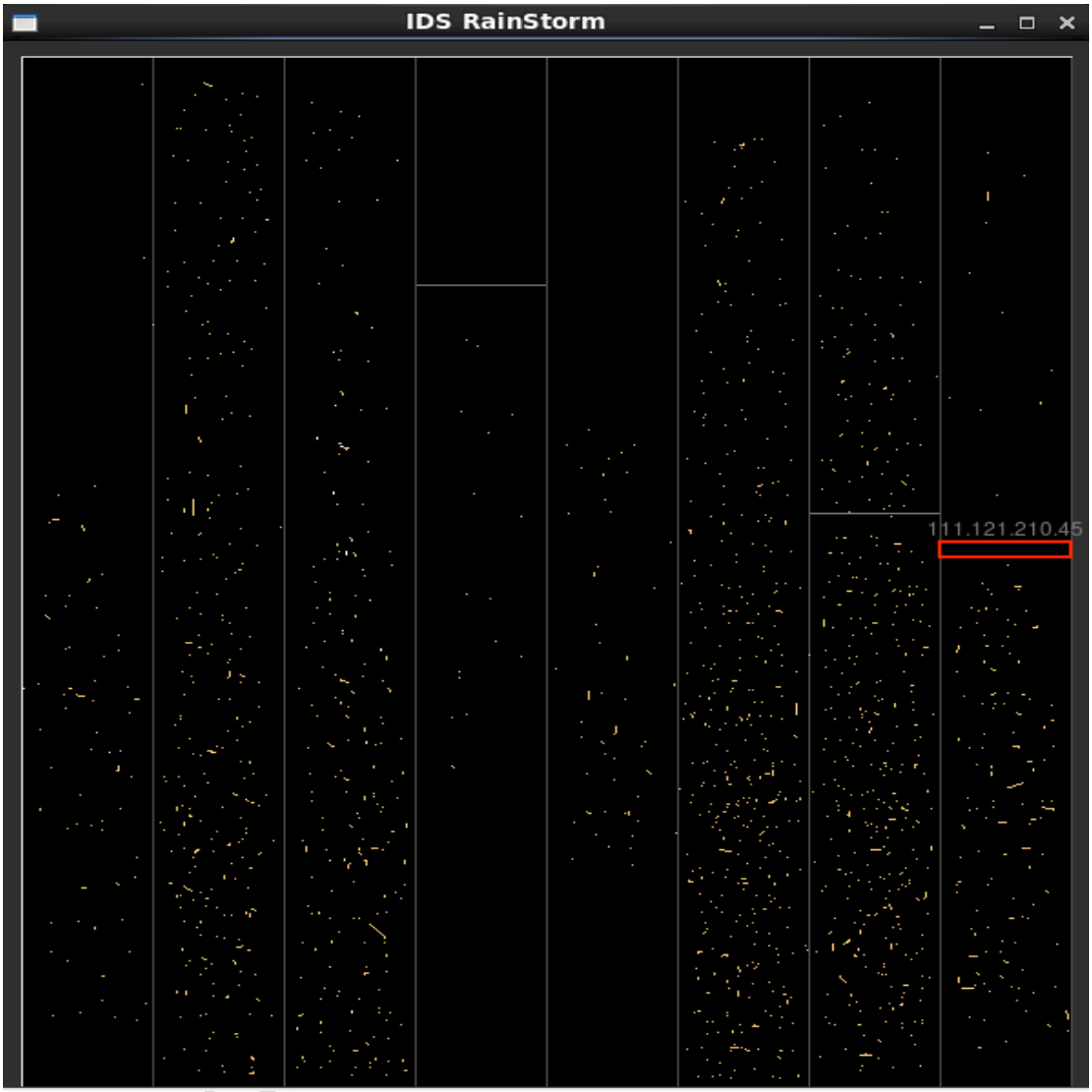
Στην εικόνα 18 χρησιμοποιούμε το εργαλείο EtherApe με στόχο την απεικόνιση μέρους της δικτυακής και κακόβουλης κίνησης του συνόλου αρχείων DATA01. Εδώ ουσιαστικά αναπαρίστανται οι συσχετίσεις των εκάστοτε hosts που συμμετέχουν σε ένα δεδομένο pcap αρχείο και το κατά πόση κίνηση διέρχεται μέσα από μία συγκεκριμένη διεύθυνση IP.

Στην εικόνα 19 γίνεται γραφική απεικόνιση κίνησης μέσω του RUMINT για το σύνολο αρχείων DATA03 το οποίο αφορά επιθέσεις άρνησης υπηρεσιών. Συγκεκριμένα, εδώ υπολογίζεται αρχικά η εντροπία των δεδομένων και έπειτα εμφανίζεται η συχνότητα των διαφορετικών bytes που επιστρέφονται κάθε φορά από το εκάστοτε πακέτο.

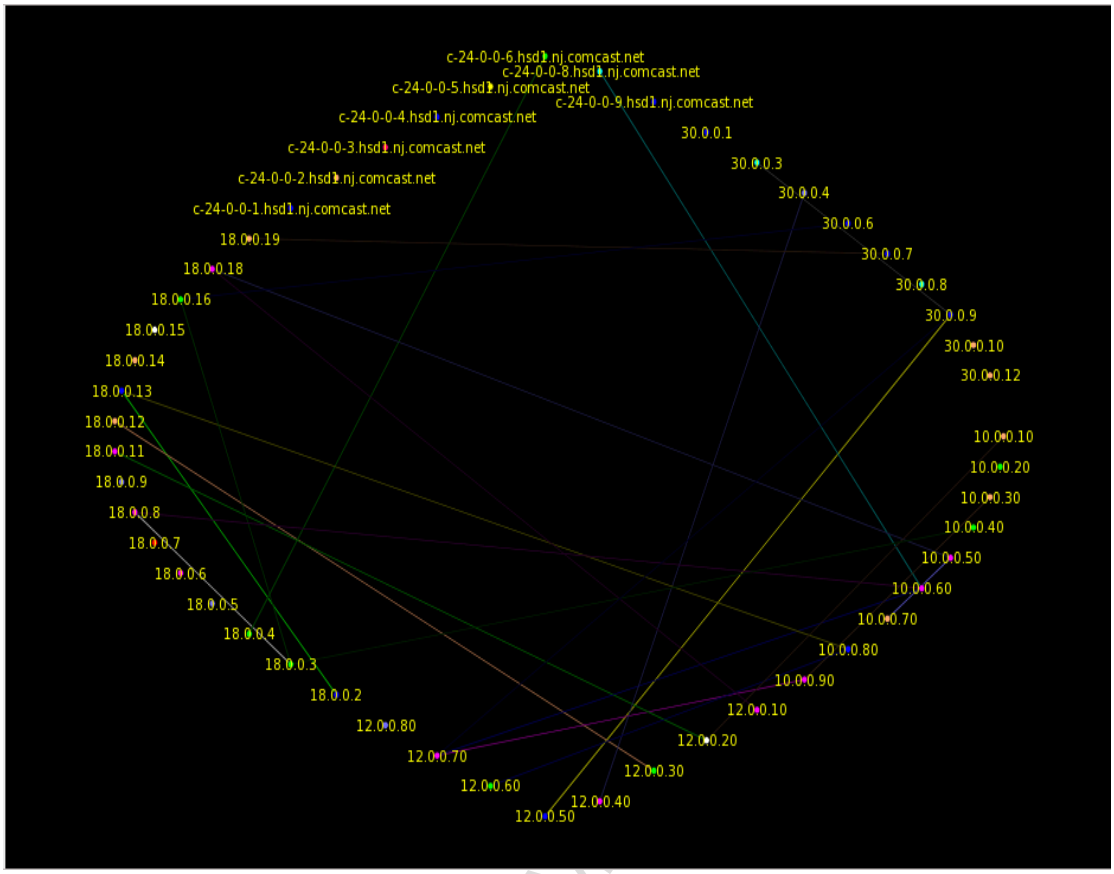
Στην εικόνα 20 παρουσίαζεται επίσης μέρος της διερχόμενης κίνησης ενός συγκεκριμένου αρχείου pcap από το σύνολο αρχείων DATA02 το οποίο περιέχει τόσο κακόβουλη όσο και φυσιολογική τυχαιοποιημένη κίνηση. Οι συνδυασμός διαφόρων χρωμάτων στις τελείες που εμφανίζονται με ανομοιογενή τρόπο απεικονίζει τυχαιοποιημένη κίνηση η οποία κατά το μεγαλύτερο μέρος της είναι φυσιολογική.

Στην εικόνα 21 γίνεται αναπαράσταση του περιεχομένου του κάθε πακέτου σε μορφή κειμένου και αφορά κακόβουλη κίνηση που εμφανίζεται στο σύνολο αρχείων DATA01.

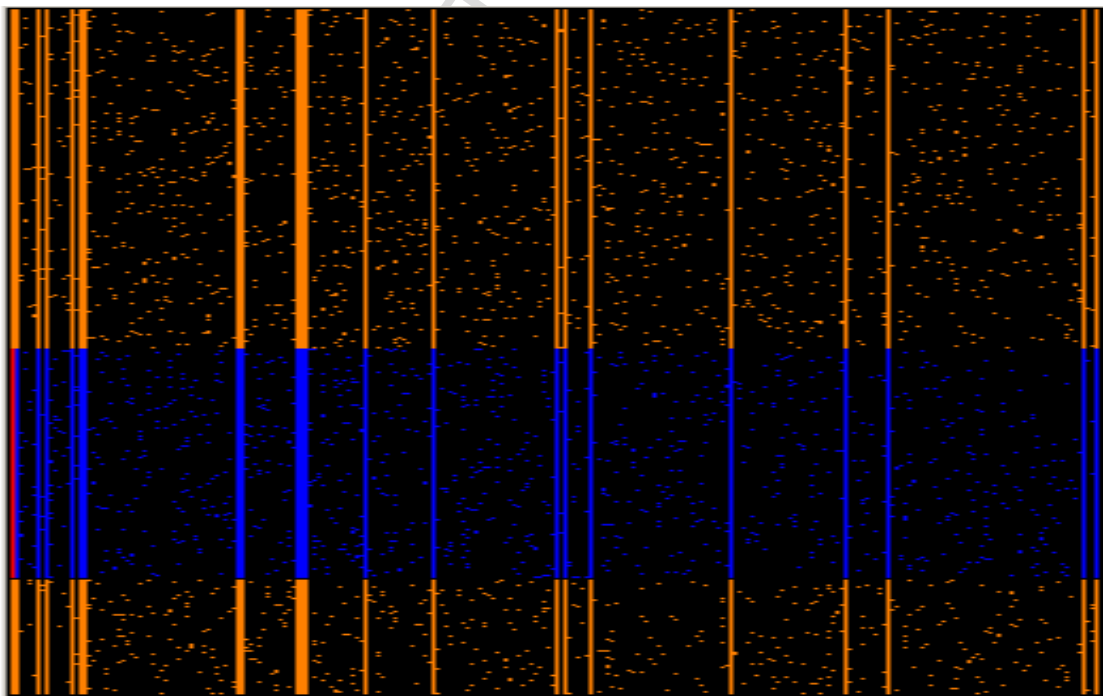
Στις εικόνες 22 και 23 γίνεται απεικόνιση ενός μέρους των ανώνυμων hosts που συμμετέχουν σε συγκεκριμένα αρχεία του συνόλου DATA04. Οι hosts αυτοί περιλαμβάνουν τόσο εισβολείς όσο και θύματα ή και φυσιολογικούς χρήστες.



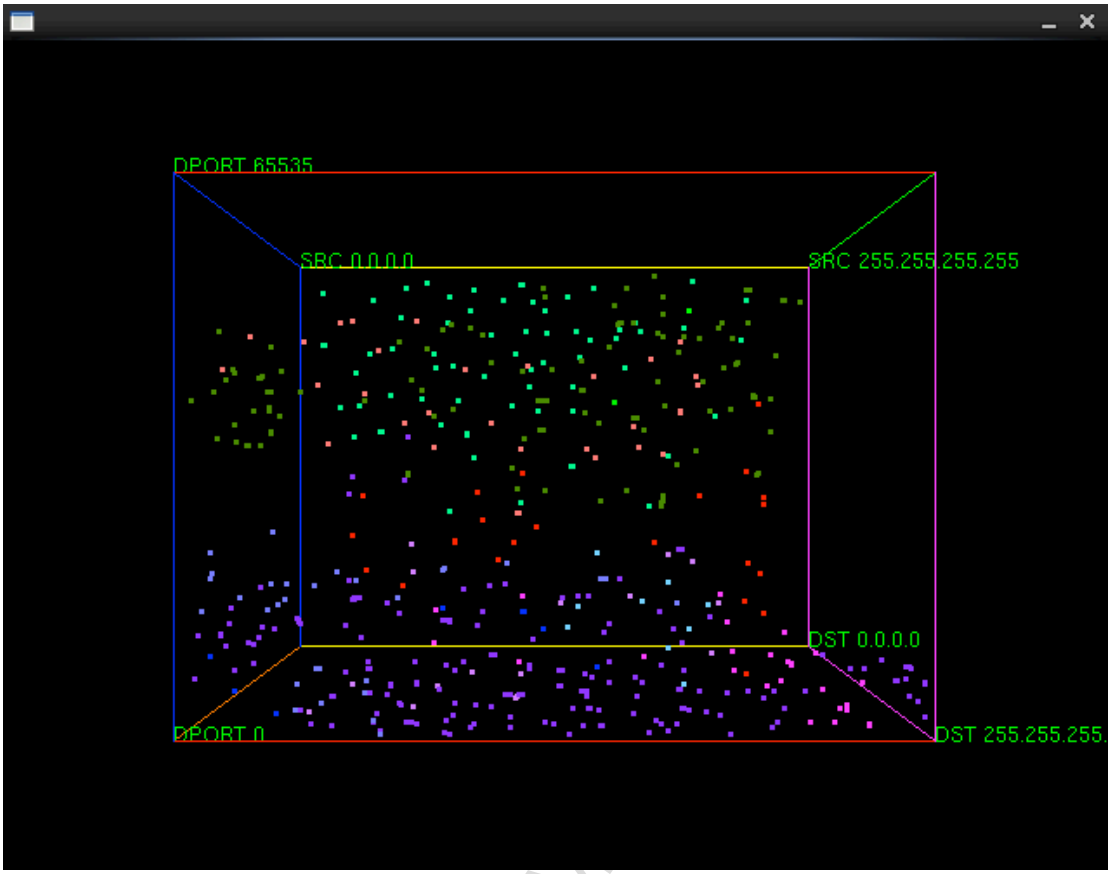
Εικόνα 17: Αναπαράσταση Ανώνυμης Κίνησης στο IDS Rainstorm (DATA04)



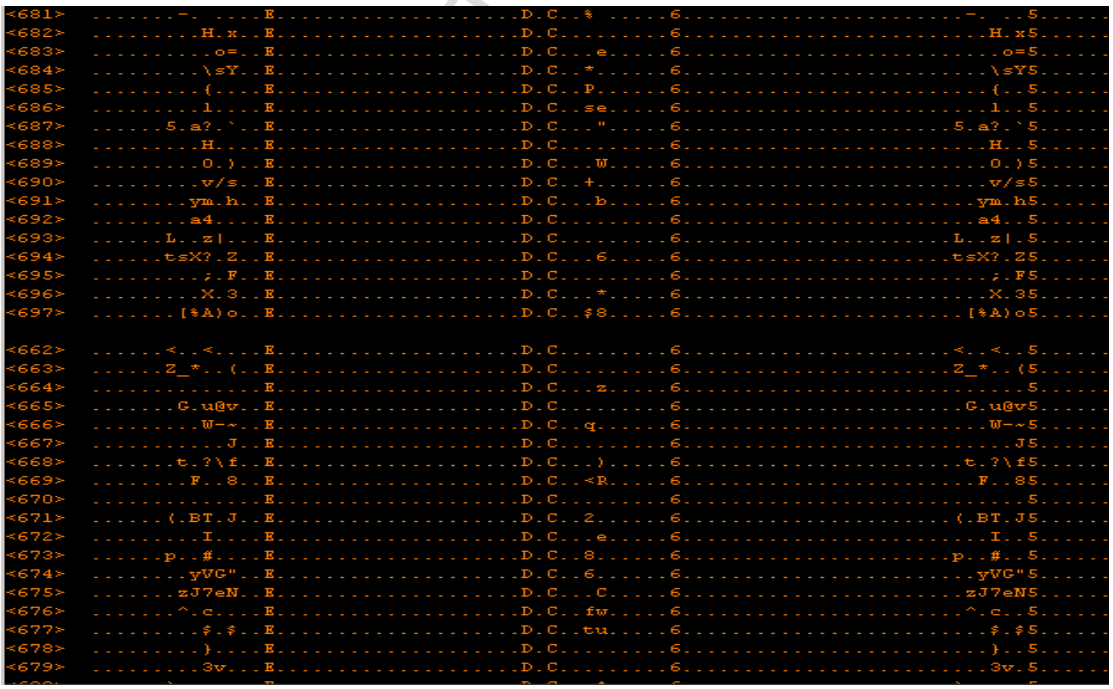
Εικόνα 18: Εικονική Απαπαράσταση Κίνησης Μέσω του EtherApe (DATA01)



Εικόνα 19: Byte Frequency Plot στο RUMINT (DATA03)

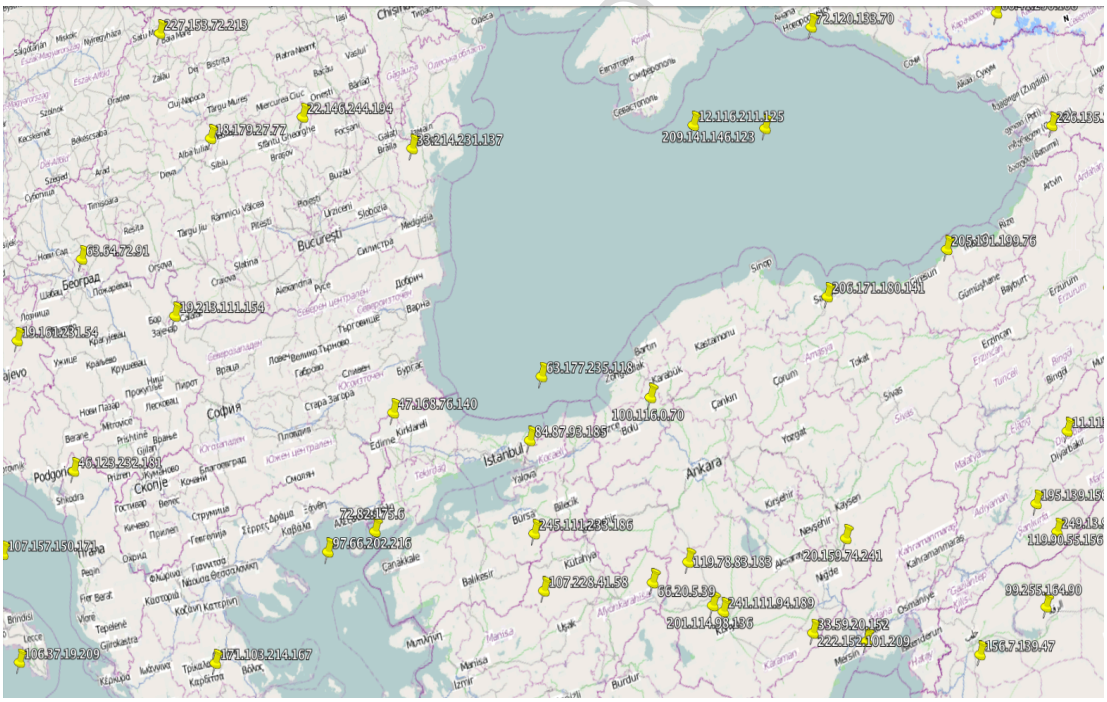


Εικόνα 20: Λεπτομερής Απεικόνιση Κίνησης στο Doomcube (DATA02)



Εικόνα 21: Text Rainfall στο RUMINT (DATA01)

Εικόνα 22: Γεωγραφική Αναπαράσταση Διευθύνσεων IP στο Google Maps (DATA04)



Εικόνα 23: Χαρτογράφηση Ανώνυμων Hosts στο OpenStreetMap (DATA04)

Κεφάλαιο 6ο

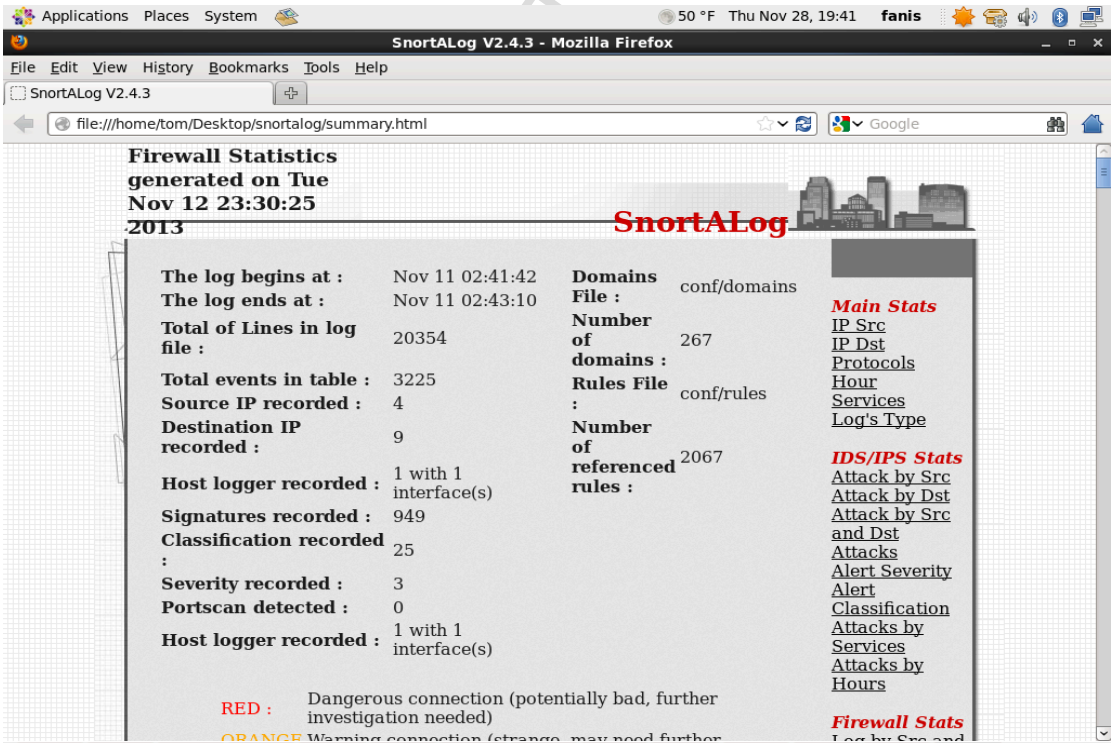
Ανάλυση Εισβολών

Εφόσον παράγαμε επιτυχώς δικτυακή κίνηση, το επόμενο βήμα αποτελεί η περαιτέρω διερεύνηση της κίνησης αυτής και συγκεκριμένα της κακόβουλης κίνησης. Στόχος εδώ είναι να αναπαριστήσουμε τις πιθανές εισβολές τόσο αριθμητικά όσο και περιγραφικά.

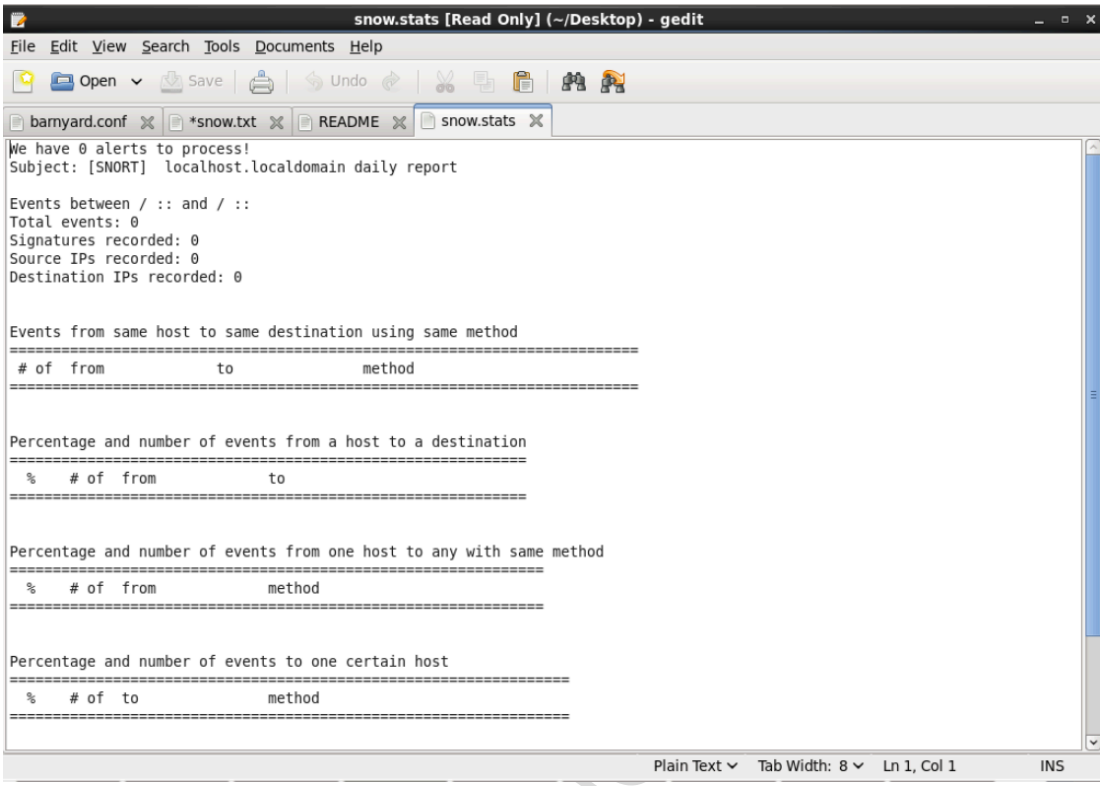
6.1 Στατιστικά Δεδομένα

Αρχικά για να περιγράψουμε αριθμητικά τις εισβολές που παρατηρηθηκάν και καταγράφηκαν μέσω του NIDS χρησιμοποιήσαμε δύο εργαλεία το Snortalog και το Snort Stat, τα οποία αναλύουν συναγερμούς οι οποίοι παράχθηκαν από το Snort. Συγκεκριμένα και για τα δύο εκτελέσαμε τις παρακάτω εντολές στο command-line:

```
cat alert | perl snortalog.pl –r –o summary.html –report
cat alert | ./snort_stat.pl > snow.stats
```



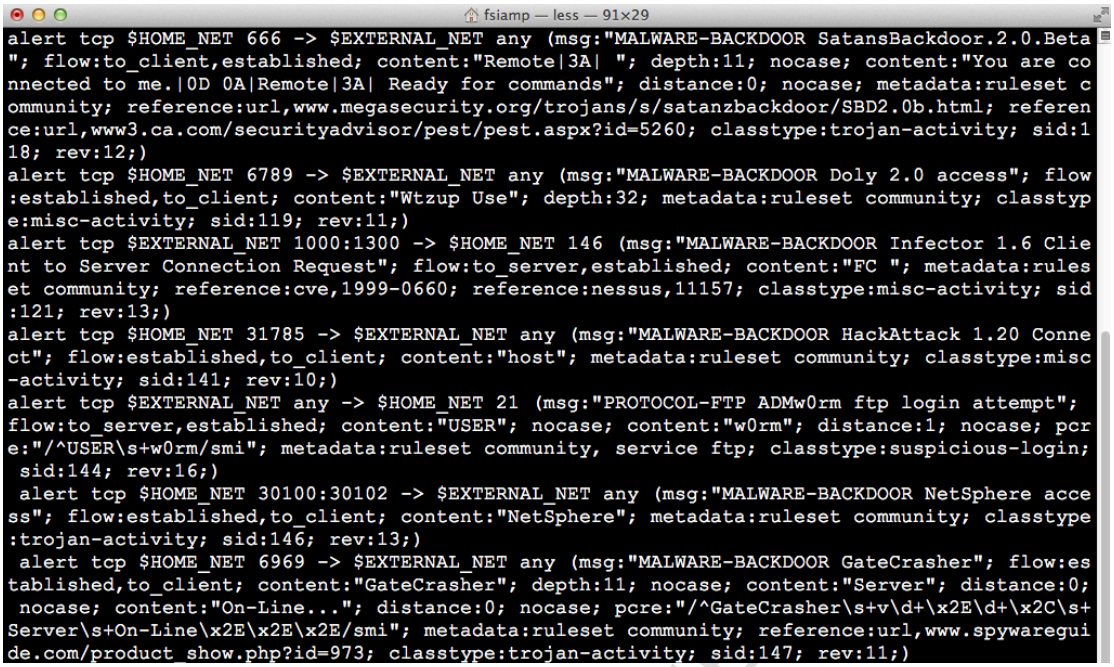
Εικόνα 24: Καταγραφή Στατιστικών Εισβολών μέσω του SnortALog



Εικόνα 25: Ενδεικτική Αποτύπωση Στατιστικών μετά από χρήση του Snort Stat

Προηγουμένως φυσικά θα πρέπει να έχουμε παράγει ήδη τα υπάρχοντα alarms αφού θέσουμε σε λειτουργία όλους τους πιθανούς κανόνες διαγράφοντας τα comments από το κάθε αρχείο κανόνων και ενσωματώνοντας τα σε ένα εννιαίο αρχείο και επιπλέον εκτελώντας παρακολούθηση μέσω του Snort και του Bro στην κίνηση του εκάστοτε δικτύου.

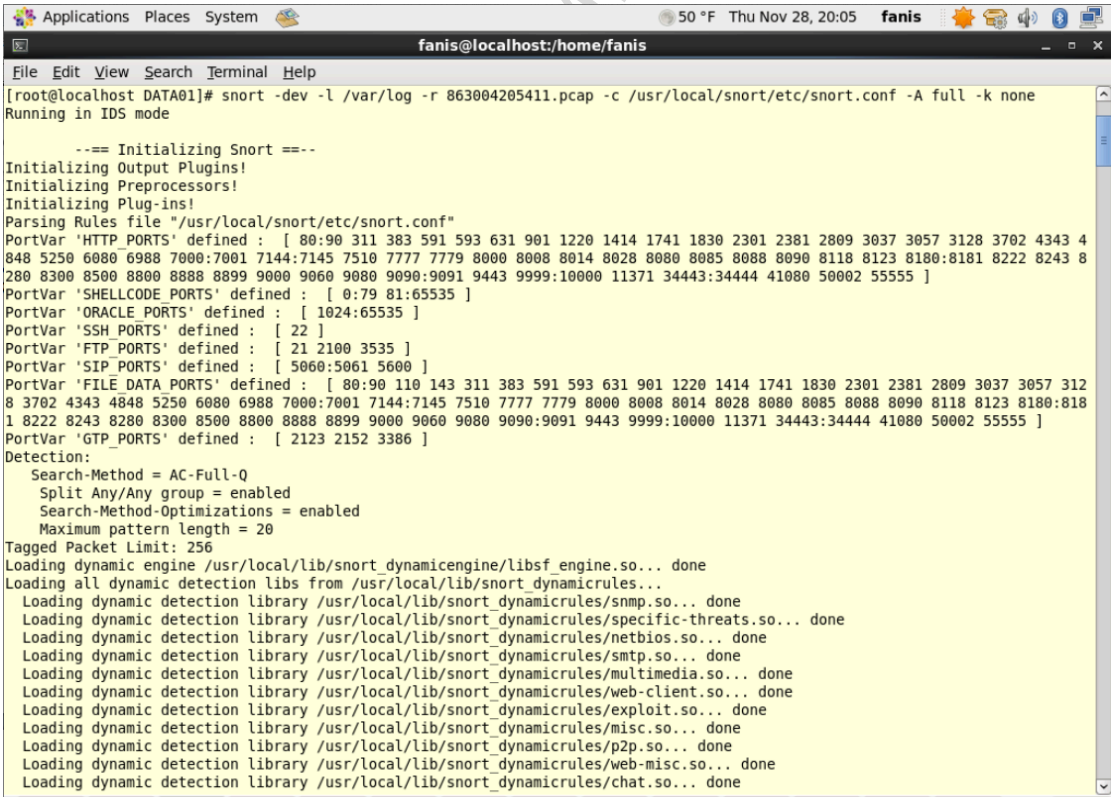
```
sed 's/#//' /usr/local/snort/rules/community.rules > GENERIC.rules
```



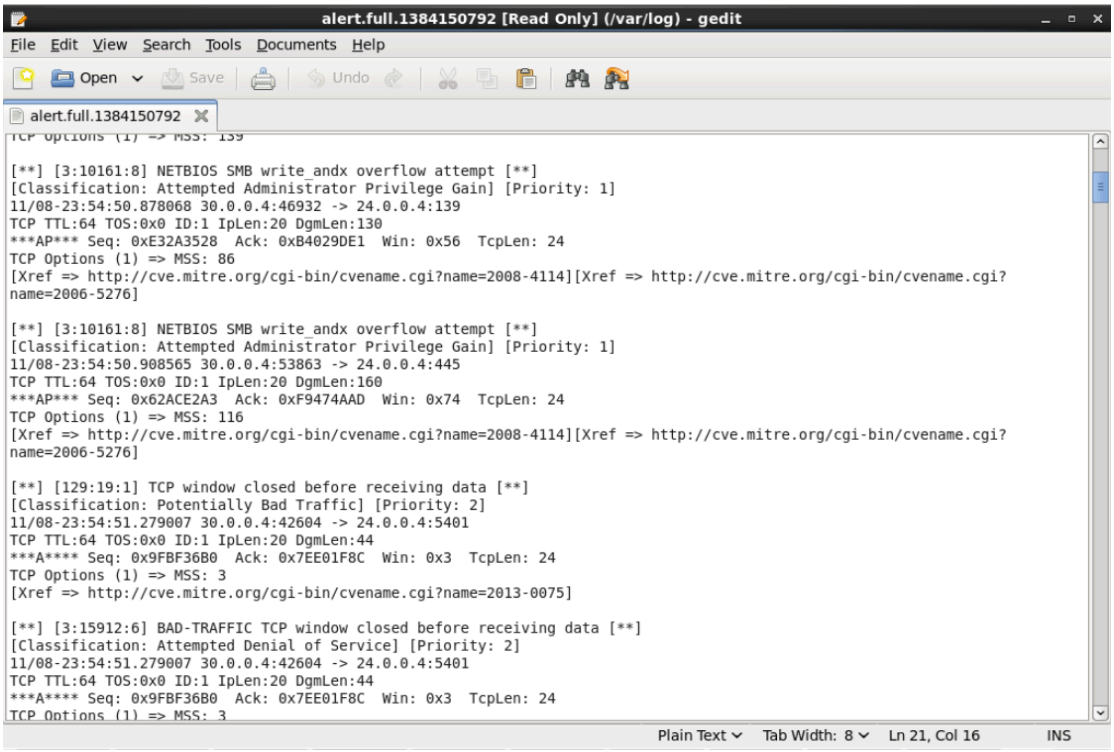
Εικόνα 26: Μέρος του Συνόλου Κανόνων που Χρησιμοποίθηκε Κατά την Ανίχνευση

```
snort -l /var/log -de -c /usr/local/snort/etc/snort.conf -r
/home/547532638421.pcap -A full -k none
```

(καταγραφή μόνο alerts με όλες τις δυνατές πληροφορίες)



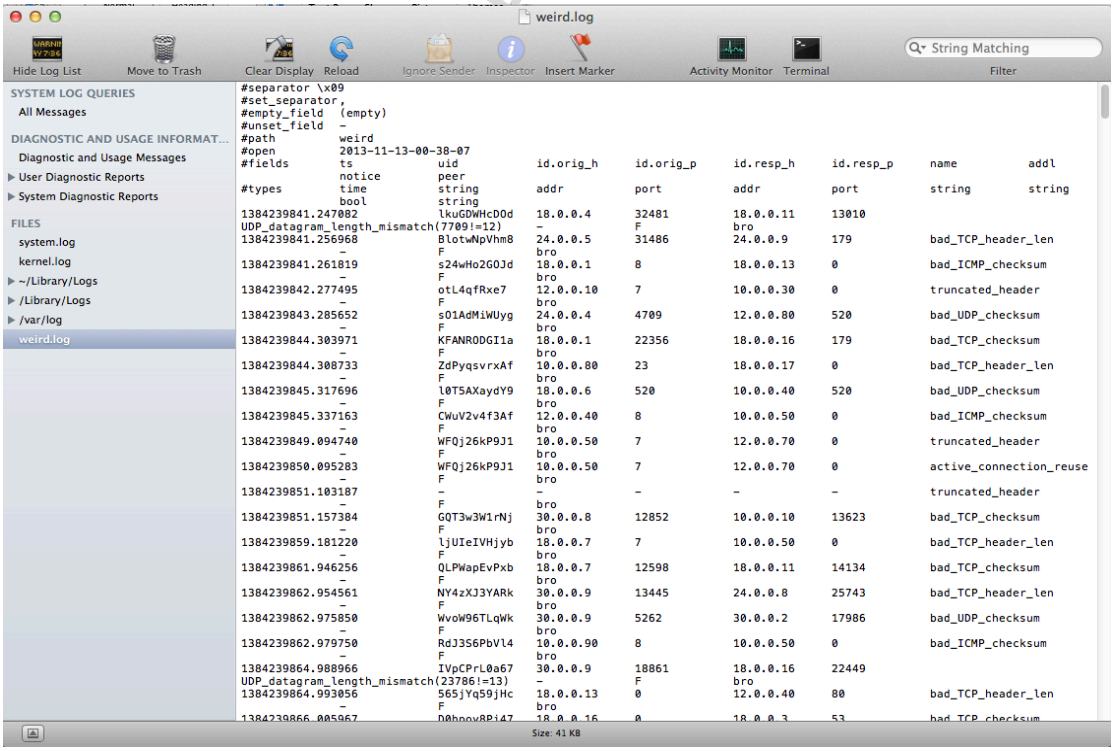
Εικόνα 27: Διαδικασία Εκτέλεσης Ανίχνευσης και Καταγραφής μέσω Snort



Εικόνα 28: Προβολή Ενδεικτικών Συναγερμών Ασφαλείας (Snort)

```
bro -r /home/147312257514.pcap
```

από την εντολή αυτή παράγεται κατά την ανάγνωση αρχείο συναγερμών με όνομα weird.log



Εικόνα 29: Προβολή Ενδεικτικών Συναγερμών Ασφαλείας (Bro)

Παρακάτω θα παρατηρήσουμε την ένδειξη της δριμύτητας του εκάστοτε alert η οποία υπολογίζεται ως εξής:

Δριμύτητα = (Κρισιμότητα στόχου + Δυναμική Επίθεσης) - (Μέτρα Αντιμετώπισης Συστήματος + Μέτρα Αντιμετώπισης Δικτύου) [29]

Πίνακας 43: Αριθμητική Απεικόνιση Ενδεικτικών Επιθέσεων που Έγιναν σε Συγκεκριμένους Προορισμούς

%	Σύνολο	Διεύθυνση IP Πηγής	Διεύθυνση IP Προορισμού	Επίθεση
16.44	384	10.0.0.40	12.0.0.10	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}
9.80	229	10.0.0.40	12.0.0.10	BAD-TRAFFIC TCP window closed before receiving data {tcp}
6.46	151	10.0.0.40	12.0.0.10	TCP window closed before receiving data {tcp}
5.52	129	10.0.0.40	12.0.0.10	NETBIOS SMB write_andx overflow attempt {tcp}
4.49	105	12.0.0.10	12.0.0.10	(snort decoder) WARNING: Bad Traffic Same Src/Dst IP {tcp}
4.07	95	10.0.0.40	12.0.0.10	OS-WINDOWS SMB-DS Trans Max Param OS-WINDOWS attempt {tcp}
4.02	94	10.0.0.40	12.0.0.10	OS-WINDOWS SMB Trans Max Param OS-WINDOWS attempt {tcp}
1.46	34	12.0.0.10	10.0.0.40	BAD-TRAFFIC TCP window closed before receiving data {tcp}
1.28	30	10.0.0.40	12.0.0.10	(IMAP) Unknown IMAP4 response {tcp}
1.03	24	12.0.0.10	10.0.0.40	TCP window closed before receiving data {tcp}
0.98	23	10.0.0.40	12.0.0.10	OS-WINDOWS SMB Trans Max Param OS-WINDOWS attempt {udp}
0.94	22	10.0.0.40	12.0.0.10	(dcerpc2) Connection-oriented DCE/RPC - Invalid minor version: 32 {tcp}
0.90	21	127.0.0.1	127.0.0.1	(snort decoder) WARNING: Bad Traffic Loopback IP {tcp}
0.90	21	127.0.0.1	127.0.0.1	(snort decoder) WARNING: Bad Traffic Same Src/Dst IP {tcp}
0.86	20	12.0.0.10	10.0.0.40	(http_inspect) UNKNOWN METHOD {tcp}
0.77	18	10.0.0.40	12.0.0.10	(http_inspect) SIMPLE REQUEST {tcp}
0.68	16	10.0.0.40	127.0.0.1	(snort decoder) WARNING: Bad Traffic Loopback IP {tcp}
0.60	14	10.0.0.40	10.0.0.40	(snort decoder) WARNING: Bad Traffic Same Src/Dst IP {tcp}
0.51	12	10.0.0.40	127.0.0.1	(snort decoder) WARNING: Bad Traffic Loopback IP {udp}
0.51	12	127.0.0.1	10.0.0.40	(snort decoder) WARNING: Bad Traffic Loopback IP {tcp}
0.43	10	10.0.0.40	12.0.0.10	(ftp_telnet) Telnet Subnegotiation Begin Command without Subnegotiation End {tcp}
0.34	8	10.0.0.40	12.0.0.10	(http_inspect) NON-RFC DEFINED CHAR {tcp}
0.34	8	10.0.0.40	12.0.0.10	(http_inspect) POST W/O CONTENT-LENGTH OR CHUNKS {tcp}
0.30	7	10.0.0.40	12.0.0.10	INDICATOR-SHELLCODE x86 inc ecx NOOP {tcp}
0.26	6	12.0.0.10	10.0.0.40	PUA-P2P GNUTella client request {tcp}
0.21	5	10.0.0.40	12.0.0.10	(IMAP) Unknown IMAP4 command {tcp}
0.21	5	10.0.0.40	12.0.0.10	(http_inspect) UNKNOWN METHOD {tcp}
0.21	5	10.0.0.40	12.0.0.10	(ftp_telnet) Invalid FTP Command {tcp}
0.17	4	10.0.0.40	12.0.0.10	APP-DETECT VNC server response {tcp}

Πίνακας 44: Κατανομή ενδεικτικών μεθόδων επίθεσης και αντίκτυπου αυτών

%	Σύνολο	Επίθεση	Αντίκτυπο	Δριμύτητα
16.58	670	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}	3	Χαμηλή
12.40	501	BAD-TRAFFIC TCP window closed before receiving data {tcp}	2	Μεσαία
8.37	338	TCP window closed before receiving data {tcp}	2	Μεσαία
5.37	217	(snort decoder) WARNING: Bad Traffic Same Src/Dst IP {tcp}	2	Μεσαία
5.02	203	NETBIOS SMB write_andx overflow attempt {tcp}	1	Υψηλή
4.48	181	OS-WINDOWS SMB Trans Max Param OS-WINDOWS attempt {tcp}	3	Χαμηλή
3.96	160	OS-WINDOWS SMB-DS Trans Max Param OS-WINDOWS attempt {tcp}	3	Χαμηλή
2.43	98	(snort decoder) WARNING: Bad Traffic Loopback IP {tcp}	2	Μεσαία
1.46	59	(IMAP) Unknown IMAP4 response {tcp}	3	Χαμηλή
1.11	45	(http_inspect) UNKNOWN METHOD {tcp}	3	Χαμηλή
1.01	41	OS-WINDOWS SMB Trans Max Param OS-WINDOWS attempt {udp}	3	Χαμηλή
0.92	37	(http_inspect) SIMPLE REQUEST {tcp}	3	Χαμηλή
0.77	31	(http_inspect) POST W/O CONTENT-LENGTH OR CHUNKS {tcp}	3	Χαμηλή
0.52	21	(dcerpc2) Connection-oriented DCE/RPC - Invalid minor version: 32 {tcp}	2	Μεσαία
0.50	20	(snort decoder) WARNING: Bad Traffic Loopback IP {udp}	2	Μεσαία
0.45	18	(http_inspect) NON-RFC DEFINED CHAR {tcp}	2	Μεσαία
0.37	15	SNMP request tcp {tcp}	2	Μεσαία
0.32	13	(smtp) Attempted command buffer overflow: more than 512 chars {tcp}	1	Υψηλή
0.32	13	(ftp_telnet) Invalid FTP Command {tcp}	2	Μεσαία
0.27	11	(dcerpc2) Connection-oriented DCE/RPC - Invalid major version: 4 {tcp}	2	Μεσαία
0.27	11	INDICATOR-SHELLCODE x86 inc ecx NOOP {tcp}	1	Υψηλή
0.22	9	SNMP request udp {udp}	2	Μεσαία
0.20	8	(snort decoder) WARNING: Bad Traffic Same Src/Dst IP {udp}	2	Μεσαία
0.20	8	BROWSER-PLUGINS Veritas Storage Exec ActiveX clsid access attempt {tcp}	1	Υψηλή
0.20	8	SNMP AgentX/tcp request {tcp}	2	Μεσαία
0.17	7	SERVER-OTHER Citrix Provisioning Services multiple opcode integer overflow attempt {udp}	1	Υψηλή
0.17	7	(ftp_telnet) Telnet Subnegotiation Begin Command without Subnegotiation End {tcp}	3	Χαμηλή
0.15	6	APP-DETECT VNC server response {tcp}	3	Χαμηλή
0.12	5	PROTOCOL-FTP format string attempt {tcp}	3	Χαμηλή
0.10	4	(ftp_telnet) FTP command parameters were too long {tcp}	1	Υψηλή
0.10	4	(http_inspect) OVERSIZE REQUEST-URI DIRECTORY {tcp}	2	Μεσαία
0.10	4	(POP) Unknown POP3 command {tcp}	3	Χαμηλή

0.10	4	SERVER-OTHER Adobe Coldfusion getodbcin attempt {tcp}	1	Υψηλή
0.10	4	SNMP public access udp {udp}	2	Μεσαία
0.10	4	(IMAP) Unknown IMAP4 command {tcp}	3	Χαμηλή
0.10	4	(snort_decoder) WARNING: IPV4 packet to reserved dest address {udp}	3	Χαμηλή
0.10	4	(snort_decoder) WARNING: IPV4 packet to broadcast dest address {udp}	3	Χαμηλή
0.07	3	BROWSER-FIREFOX Mozilla Firefox Chrome Page Loading Restriction Bypass attempt {tcp}	1	Υψηλή
0.07	3	SERVER-WEBAPP net attempt {tcp}	2	Μεσαία
0.07	3	INDICATOR-COMPROMISE IRC message on non-standard port {tcp}	1	Υψηλή
0.07	3	SERVER-IIS Directory transversal attempt {tcp}	1	Υψηλή
0.07	3	INDICATOR-SHELLCODE kadmind buffer overflow attempt {tcp}	1	Υψηλή
0.0.7	3	SERVER-OTHER HP Data Protector Express DtbClsLogin buffer overflow attempt (tcp)	1	Υψηλή

Πίνακας 45: Επιθέσεις που έγιναν σε συγκεκριμένες θύρες προορισμού

%	Θύρα	Επίθεση
0.08	25	POLICY-SPAM beatmoon.ru known spam email attempt {tcp}
0.08	7928	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}
0.08	56362	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}
0.08	20112	MALWARE-BACKDOOR possible Htran setup command - tran {tcp}
0.08	32305	BROWSER-PLUGINS Microsoft Internet Explorer Object.Microsoft.DXTFilter ActiveX function call access {tcp}
0.08	25	FILE-IDENTIFY .msh2xml attachment file type blocked by Outlook detected {tcp}
0.08	2077	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}
0.08	80	SERVER-WEBAPP bad HTTP/1.1 request
0.08	17651	(IMAP) Unknown IMAP4 response {tcp}
0.08	38796	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}
0.08	4771	BROWSER-PLUGINS Microsoft Internet Explorer Video Effect Class Manager 2 Input ActiveX clsid access {tcp}
0.08	25	POLICY-SPAM was.medrayner44c.ru known spam email attempt {tcp}
0.08	23004	TCP window closed before receiving data {tcp}
0.08	17651	FILE-EXECUTABLE Microsoft Windows Vista Windows mail file execution attempt {tcp}
0.08	24443	BROWSER-PLUGINS Microsoft Internet Explorer Marquee Control ActiveX object access {tcp}
0.08	1433	SERVER-MSSQL xp_proxiedmetadata vulnerable function attempt {tcp}
0.08	10956	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE

		{tcp}
0.08	33527	(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE {tcp}
0.08	8333	MALWARE-OTHER mstream handler to client {tcp}
0.08	2208	BAD-TRAFFIC TCP window closed before receiving data {tcp}
0.08	80	INDICATOR-SHELLCODE x86 inc ecx NOOP {tcp}

6.2 Περιγραφή Επιθέσεων

Στη συνέχεια για τα προηγούμενα αρχεία κίνησης εκτελέσαμε την εντολή grep για να αναζητήσουμε alerts που προκλήθηκαν και αφορούν συγκεκριμένα πρωτόκολλα ή υπηρεσίες. (πίνακας 46).

Πίνακας 46: Πρωτόκολλα και Υπηρεσίες που δέχθηκαν επιθέσεις

	TCP	FTP	NETBIOS	WINDOWS	TELNET	SQL
547532638421.pcap	8	77	186	658	9	61
303282314622.pcap	21	70	217	738	8	50
030117712146.pcap	12	19	72	227	2	15
753764885006.pcap	0	35	122	420	4	39
505456715603.pcap	1	63	178	628	9	69

Όπως φαίνεται παραπάνω από το δείγμα 5 αρχείων κίνησης, οι περισσότερες επιθέσεις σχετίζονταν με απειλές σε Windows συστήματα κάτι το οποίο είναι αρκετά λογικό δεδομένου ότι το λειτουργικό σύστημα αυτό είναι το πλέον δημοφιλές σε Desktop υπολογιστές. Στη συνέχεια καταγράψαμε δεδομένα εισβολών μέσω του Bro από όπου διαχωρίσαμε πάλι τις απειλές ανά αρχείο κίνησης με την εντολή grep. Για παράδειγμα εισάγαμε για αναγνώριση απειλών 'bad_checksum' :

```
grep -o 'bad_TCP_checksum' alerts/11/weird.log | wc -l
```

Στον πίνακα 47 εμφανίζονται όλα αυτά τα είδη επιθέσεων ανά αρχείο κίνησης όπως καταγράφηκαν από το συγκεκριμένο εργαλείο.

Πίνακας 47: Είδη επιθέσεων που εντοπίστηκαν μέσω του εργαλείου Bro

	671583160833.pcap	811023503167.pcap	876648958753.pcap
bad_TCP_checksum	44	31	46
bad_UDP_checksum	28	27	28
active_connection_reuse	95	30	76
truncated_header	99	31	104
UDP_datagram_length_ mismatch	49	21	56
above_hole_data_without _any_acks	2	0	0
data_before_established	5	1	5
connection_originator_ SYN_ack	9	3	7
baroque_SYN	13	4	10
SYN_with_data	9	3	7
DNS_Conn_count_too_ large	6	2	7
bad_TCP_header_len	99	36	139
bad_ICMP_checksum	20	12	22
ipv6_no_next	1	0	1

Από τον πίνακα 47 προκύπτει ότι οι περισσότερες απειλές παρουσιάστηκαν λόγω κακού σχηματισμού των εκάστοτε πακέτων που παρήχθησαν. Έπειτα, αναλύουμε διάφορα συμβάντα όπως αυτά καταγράφησαν από το syslog σε CentOS (/var/log/messages).

Αρχικά στα παρακάτω αποτυπώματα παρουσιάζεται μία κατάσταση obfuscation δηλαδή έγινε ηθελημένη αλλαγή ονόματος ενός host συστήματος τροποποιώντας το όνομα localhost σε localhero.

Nov 25 21:14:18 localhost NetworkManager[1431]: <info> address 30.0.0.6
Nov 25 21:14:18 localhost NetworkManager[1431]: <info> prefix 24 (255.255.255.0)
Nov 25 21:14:18 localhost NetworkManager[1431]: <info> gateway 30.0.0.6
Nov 25 21:14:18 localhost NetworkManager[1431]: <info> nameserver '30.0.0.6'
Nov 25 21:14:18 localhost NetworkManager[1431]: <info> domain name 'localdomain'
Nov 25 21:25:37 localhost dhclient[27010]: DHCPREQUEST on eth1 to 30.0.0.8 port 67 (xid=0x4b022ddf)
Nov 25 21:25:37 localhost dhclient[27010]: DHCPACK from 172.16.186.254 (xid=0x4b022ddf)
Nov 25 21:25:37 localhost NetworkManager[1431]: <info> (eth1): DHCPv4 state changed renew -> renew
Nov 25 21:25:37 localhost NetworkManager[1431]: <info> address 30.0.0.10
Nov 25 21:25:37 **localhero** NetworkManager[1431]: <info> prefix 24 (255.255.255.0)
Nov 25 21:25:37 **localhero** NetworkManager[1431]: <info> gateway 30.0.0.6
Nov 25 21:25:37 **localhero** NetworkManager[1431]: <info> nameserver '30.0.0.6'
Nov 25 21:25:37 **localhero** NetworkManager[1431]: <info> domain name 'localdomain'
Nov 25 21:25:37 **localhero** dhclient[27010]: bound to 30.0.0.10 -- renewal in 728 seconds.

Nov 25 21:37:45 **localhero** dhclient[27010]: DHCPREQUEST on eth1 to 30.0.0.8 port 67 (xid=0x4b022ddf)
Nov 25 21:37:45 **localhero** dhclient[27010]: DHCPACK from 30.0.0.8 (xid=0x4b022ddf)
Nov 25 21:37:45 **localhero** NetworkManager[1431]: <info> (eth1): DHCPv4 state changed renew -> renew
Nov 25 21:37:45 **localhero** NetworkManager[1431]: <info> address 30.0.0.10
Nov 25 21:37:45 **localhero** NetworkManager[1431]: <info> prefix 24 (255.255.255.0)
Nov 25 21:37:45 **localhero** NetworkManager[1431]: <info> gateway 30.0.0.6
Nov 25 21:37:45 **localhero** NetworkManager[1431]: <info> nameserver '30.0.0.6'
Nov 25 21:37:45 **localhero** NetworkManager[1431]: <info> domain name 'localdomain'
Nov 25 21:37:45 **localhero** dhclient[27010]: bound to 30.0.0.10 -- renewal in 760 seconds.
Nov 25 21:50:25 **localhero** dhclient[27010]: DHCPREQUEST on eth1 to 30.0.0.8 port 67 (xid=0x4b022ddf)
Nov 25 21:50:25 **localhero** dhclient[27010]: DHCPACK from 30.0.0.8 (xid=0x4b022ddf)

Στο επόμενο αποτύπωμα του syslog παρατηρούμε ότι έχει ξεπεραστεί το μέγιστο όριο των δυνατών συνδέσεων που μπορεί να έχει μία υπηρεσία ταυτόχρονα. Συνέπεια αυτού του συμβάντος είναι η αλλαγή στην κατάσταση του συστήματος καθώς ο δρομολογητής δεν είναι σε θέση να διεκπεραιώσει πολλαπλές συνδέσεις.

Nov 27 16:58:07 localhost NetworkManager[1431]: <info> Auto-activating connection 'Auto eth1'.
Nov 27 16:58:07 localhost NetworkManager[1431]: <info> Activation (eth1) starting connection 'Auto eth1'
Nov 27 16:58:07 localhost NetworkManager[1431]: <info> (eth1): **device state change: 3 -> 4 (reason 0)**
Nov 27 16:58:07 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 1 of 5 (Device Prepare) scheduled...
Nov 27 16:58:07 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 1 of 5 (Device Prepare) started...
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 2 of 5 (Device Configure) scheduled...
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 1 of 5 (Device Prepare) complete.
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 2 of 5 (Device Configure) starting...
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> (eth1): device state change: 4 -> 5 (reason 0)
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 2 of 5 **(Device Configure) successful.**
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 3 of 5 (IP Configure Start) scheduled.
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 2 of 5 (Device Configure) complete.
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> Activation (eth1) Stage 3 of 5 (IP Configure Start) started...
Nov 27 16:58:08 localhost NetworkManager[1431]: <info> (eth1): device state change: 5 -> 7 (reason 0)

Άλλα περιστατικά που παρατηρήθηκαν επίσης στο syslog σχετίζονταν με αλλαγή των MAC διευθύνσεων σε ένα συγκεκριμένο host. Παράλληλα συμπεράσματα μπορούμε να εξάγουμε και από τα εκάστοτε hexdumps. Για παράδειγμα, στο ακόλουθο μέρος ενός πακέτου βλέπουμε ότι ανιχνεύθηκε επίθεση shellcode η οποία ακολουθεί το μοτίβο που ακολουθούν οι κλασικές επιθέσεις shellcode και σε offline κατάσταση. Δηλαδή, πραγματοποιείται στην αρχή overflow σε οριακό σημείο ώσπου να καλυφθεί η διεύθυνση επιστροφής η οποία θα δείχνει στα τρέχοντα NOPS τα οποία με τη σειρά τους θα δείχνουν το ένα στο άλλο και τέλος θα γίνει η εκτελέση του κώδικα shell. Παρακάτω βλέπουμε ότι αρχικά έχει γίνει overflow με σειρά δυαδικών χαρακτήρων 6d.

[**] [1:17322:2] INDICATOR-SHELLCODE x86 OS agnostic fnstenv geteip dword xor decoder [**]

0160 80 51 93 8e 2d d6 99 88 15 86 99 88 2a d6 37 09 .Q..-... ....*.7.
0170 17 2a 11 dc b1 d4 37 0f 15 78 37 ee 80 57 43 8e .*....7. .x7..WC.
0180 83 04 0c bd 80 51 9a 26 af ef 38 53 7b d8 9b 26 .....Q.& ..8S{..&
0190 a9 78 18 d9 7f 87 **6d 6d 6d 6d 6d 6d 6d 6d 6d 6d .x....mm mmmmmmmm**
01a0 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
01b0 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
01c0 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
01d0 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
01e0 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
01f0 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0200 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0210 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0220 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0230 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0240 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0250 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm
0260 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d 6d mmmmmmmm mmmmmmmm

Υπερχείλιση πακέτου δεδομένων κίνησης - Εκτέλεση Shellcode

Σε τελευταίο βήμα για την περιγραφή της συσχέτισης των παραγόμενων συναγερμών, κριτήριο μας αποτελεί η ομοιότητα μεταξύ του κάθε συναγερμού καθώς και τα εκάστοτε correlation rules τα οποία έχουμε θέσει. Οι κανόνες αυτοί μπορούν να βρεθούν μέσω χρήσης του LAMBDA [30] [31], του SHEDEL [32], του STATL [33], του Sutekh, του JISGAW [34] , του STRIPS και του ADeLe [35] τα οποία είναι γλώσσες που μοντελοποιούν μία βάση δεδομένων με σκοπό την ανίχνευση επιθέσεων.

Ακόμα για την περιγραφή συγκεκριμένων επιθέσεων μπορούμε να χρησιμοποιήσουμε και rule-based γλώσσες όπως η RUSSELL και η P-BEST. [36] [37]

6.3 Συσχετισμός και Αβεβαιότητα

Αρχικά, για την απεικόνιση της συσχέτισης των alerts πρέπει να υπολογίσουμε την ομοιότητα μεταξύ των παραγόμενων alerts. Το στοιχείο αυτό το ορίζουμε ως:

$$\mathrm{SIM}(x,y) = \frac{\sum_i E_j SIM(x_j, y_j)}{\sum_j E_j}$$

Όπου X είναι ένα υποψήφιο alert για ταίριασμα, Y είναι κάθε νέο alert που προκύπτει, το j ταξινομεί πακέτα ως πρός τα χαρακτηριστικά των συναγερμών, Ej είναι η πιθανή ομοιότητα μεταξύ δύο χαρακτηριστικών και τα Xj και Yj είναι τιμές που προσδιορίζουν χαρακτηριστικά των alerts X και Y αντίστοιχα. [38] [39]

aa	ab	ac	ad	ae	af	ag	ah	ai	aj	ak	
2	0	0	17	24	14	28	16	2	15	36	a = normal
4	0	0	15	23	18	30	16	4	13	43	b = malware
5	0	0	16	23	13	39	11	4	13	49	c = xss_attack
4	0	0	12	20	11	37	9	2	14	39	d = attempted_admin
3	0	0	14	21	13	37	12	4	14	49	e = attempted_user
4	0	0	11	23	15	32	6	3	10	39	f = inappropriate_content
2	0	0	14	20	12	28	11	5	11	37	g = policy_violation
0	0	0	11	25	18	34	6	5	15	48	h = shellcode_detect
7	0	0	16	18	18	33	11	4	13	40	i = successful_admin
3	0	0	11	20	15	32	14	5	13	44	j = successful_user
3	0	0	15	20	17	35	6	6	10	43	k = trojan_activity

Συσχέτιση Εξαγόμενων Alarms ως Προς τα Χαρακτηριστικά Φυσιολογικής Κίνησης

Επιπλέον, μπορούμε να αναπαραστήσουμε τη σχέση μεταξύ των παραγόμενων alerts στα αρχεία data μέσα από δένδρα αποφάσεων δηλαδή μία ακολουθία χαρακτηριστικών η οποία θα εξαρτάται από το πόρισμα-αποτέλεσμα προηγούμενων attributes σε κάθε στάδιο. [29] Αυτή η διαδικασία γενικά, μας βοηθάει στο να προβλέψουμε μελλοντικές καταστάσεις. Αρχικά όμως για να φτιάξουμε ένα δένδρο αποφάσεων θα πρέπει να υπολογίσουμε την εντροπία (για υπολογισμό μέτρου αβεβαιότητας) καθώς και το κέρδος της πληροφορίας για κάθε χαρακτηριστικό επίθεσης. [41] [42]

Ειδικότερα, η εντροπία μίας κλάσης που αντιστοιχίζεται σε ένα χαρακτηριστικό ορίζεται ως

$$\mathrm{E(F)} = \sum_{j=1}^{v} \frac{s_{1j}+\cdots+s_{mj}}{s} \times \mathrm{I}(s_{1j,\ldots,}s_{mj})$$

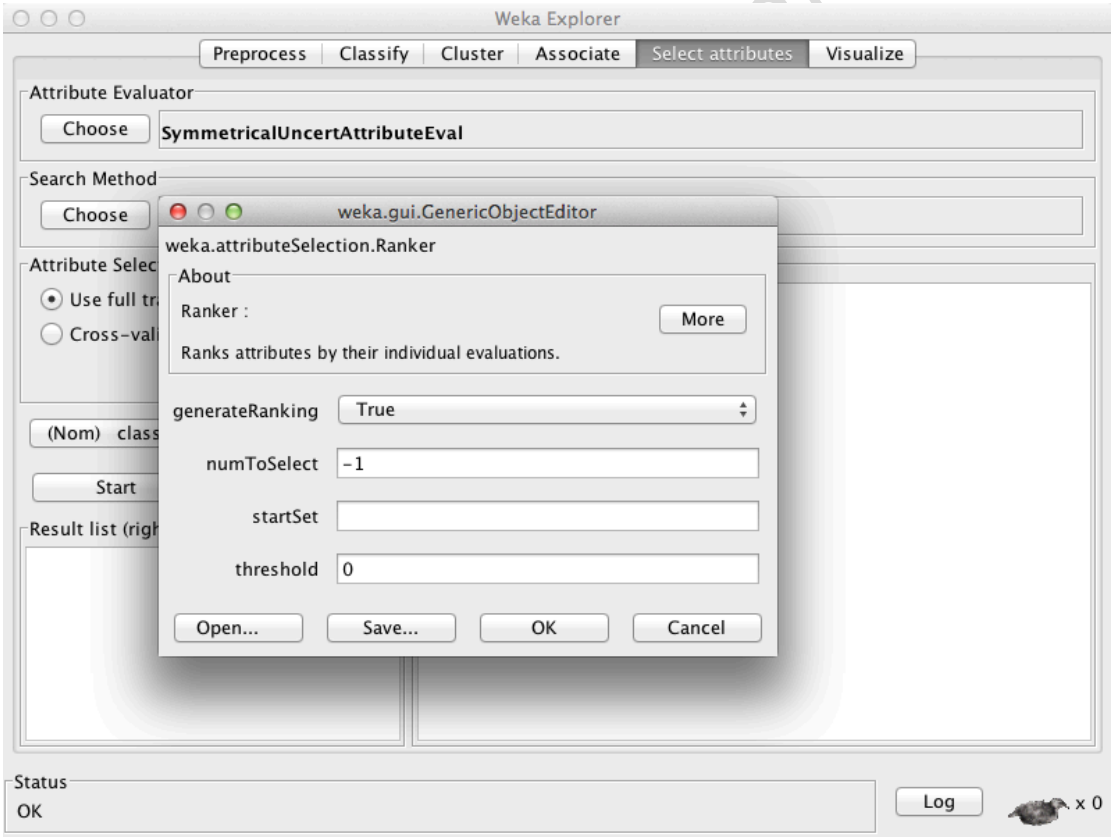
όπου I αποτελεί την προσδοκώμενη πληροφορία η οποία απαιτείται για την ταξινόμηση των δεδομένων υπό την παραδοχή ότι έχουμε ένα δεδομένο δείγμα για κάθε κλάση. Συγκεκριμένα, το I υπολογίζεται ως:

$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^{m} \frac{s_i}{s} log_2 \left(\frac{s_i}{s}\right)$

με s_i να είναι ο αριθμός των δειγμάτων για την τρέχουσα κλάση και s ο συνολικός αριθμός δειγμάτων για όλα τα χαρακτηριστικά του dataset. Αφού βρούμε τις παραπάνω τιμές μπορούμε να υπολογίσουμε το κέρδος πληροφορίας το οποίο ορίζεται ώς:

$Gain(F) = I(s_1, s_2, \dots, s_m) - E(F)$

Για ευκολότερο και γρηγορότερο υπολογισμό μπορούμε να χρησιμοποιήσουμε το Weka στο οποίο εισάγουμε ως attribute evaluators τα GainRatioAttributeEval και το SymmetricalUncertAttributeEval με μέθοδο Ranker και threashold 0. [43]



Εικόνα 30: Προσαρμογή Αξιολογητή Χαρακτηριστικών στο Weka

Όπως φαίνεται και παρακάτω ύστερα από ανάλυση του αρχείου 395010249482.data για τα χαρακτηριστικά που εμφανίζονται, μπορούμε να κάνουμε το αντίστοιχο για τα attributes που χαρακτηρίζουν μία επίθεση και στη συνέχεια να εξάγουμε ένα δένδρο αποφάσεων του οποίου

μέρος παρουσιάζεται στη συνέχεια . [44] Το δένδρο αυτό παρήχθησε μέσω Weka στο οποίο εισάγαμε τον J48 Prune Classifier.

Πίνακας 48: Αποτύπωση Βαθμίδων Αβεβαιότητας

Χαρακτηριστικό	Κέρδος Πληροφορίας (ratio)	Συμμετρική Αβεβαιότητα
service	0.07725	0.08348
logged_in	0.00741	0.00239
flag	0.01631	0.01302
is_guest_login	0.00724	0.00233
land	0.00644	0.00208
is_host_login	0.00545	0.00176
protocol_type	0.00766	0.00358

```
service = aol
|   flag = OTH
|   |   protocol_type = tcp
|   |   |   srv_rerror_rate <= 0.72
|   |   |   |   wrong_fragment <= 148: successful_recon_largescale (2.0/1.0)
|   |   |   |   wrong_fragment > 148: normal (2.0)
|   |   |   srv_rerror_rate > 0.72: denial_of_service (3.0/2.0)
|   |   protocol_type = udp
|   |   |   land = 0
|   |   |   |   xssdetect <= 71: successful_recon_limited (3.0/1.0)
|   |   |   |   xssdetect > 71: xss_attack (3.0/2.0)
|   |   |   land = 1: attempted_admin (2.0/1.0)
|   |   protocol_type = icmp
|   |   |   is_host_login = 0
|   |   |   |   duration <= 154: inappropriate_content (2.0/1.0)
|   |   |   |   duration > 154: attempted_dos (2.0/1.0)
|   |   |   is_host_login = 1: attempted_user (2.0/1.0)
|   flag = REJ
|   |   protocol_type = tcp
|   |   |   land = 0: successful_admin (3.0/2.0)
|   |   |   land = 1: successful_recon_largescale (3.0/2.0)
|   |   protocol_type = udp
|   |   |   logged_in = 0: icmp_event (2.0/1.0)
|   |   |   logged_in = 1
|   |   |   |   src_bytes <= 2227: shellcode_detect (2.0/1.0)
|   |   |   |   src_bytes > 2227: unsuccessful_user (3.0/1.0)
|   |   protocol_type = icmp
|   |   |   duration <= 146: attempted_admin (2.0/1.0)
|   |   |   duration > 146: misc_activity (2.0/1.0)
|   flag = RSTO
```

```
|   |   land = 0
|   |   |   is_guest_login = 0
|   |   |   |   logged_in = 0: xss_attack (2.0/1.0)
|   |   |   |   logged_in = 1
|   |   |   |   |   root_shell <= 84: attempted_recon (2.0/1.0)
|   |   |   |   |   root_shell > 84: unsuccessful_user (2.0)
|   |   |   is_guest_login = 1
|   |   |   |   num_access_files <= 68987: suspicious_filename_detect (2.0)
|   |   |   |   num_access_files > 68987: attempted_admin (2.0/1.0)
|   |   land = 1
|   |   |   protocol_type = tcp: system_call_detect (2.0/1.0)
|   |   |   protocol_type = udp
|   |   |   |   duration <= 59: policy_violation (2.0/1.0)
|   |   |   |   duration > 59: malware (2.0/1.0)
|   |   |   protocol_type = icmp: successful_recon_limited (3.0/2.0)
```

Δένδρο Αποφάσεων (J48)

Παρακάτω υπολογίσαμε τη συνολική ακρίβεια ανά κλάση μέσω του Weka.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	
0.081	0.076	0.03	0.081	0.043	0.518	successful_recon_largescale
0	0	0	0	0	0.492	successful_recon_limited
0	0	0	0	0	0.495	suspicious_filename_detect
0	0	0	0	0	0.529	suspicious_login
0.002	0.005	0.01	0.002	0.003	0.481	system_call_detect
0	0	0	0	0	0.506	unusual_client_port_connection
0	0	0	0	0	0.516	web_application_activity
0.02	0.025	0.023	0.02	0.022	0.477	icmp_event
0.035	0.042	0.023	0.035	0.027	0.508	misc_activity
0.024	0.027	0.023	0.024	0.024	0.492	network_scan
0.066	0.063	0.029	0.066	0.04	0.503	buffer_overflow
0.032	0.021	0.038	0.032	0.035	0.533	protocol_command_decode
0.006	0.008	0.019	0.006	0.009	0.515	string_detect
0.021	0.024	0.024	0.021	0.022	0.514	unknown
0.086	0.086	0.029	0.086	0.043	0.5	tcp_connection
Weighted Avg.	0.029	0.028	0.019	0.029	0.019	0.502

Confusion Matrix για Αναπαράσταση Ακρίβειας

Στη συνέχεια είναι σημαντικό να περιγράψουμε αριθμητικά τις εισβολές που καταγράφηκαν συνολικά από την αρχή εκκίνησης του NIDS μέσω του Snort. Ειδικότερα, στο αρχείο alerts.csv που βρίσκεται στο φάκελο GLOBAL καταγράψαμε μέσω του R console τα ακόλουθα (πίνακας 49) για τα εκάστοτε χαρακτηριστικά που χρησιμοποιήθηκαν:

```
> dataset <-- read.csv("/KINGSTON/data/GLOBAL/alerts.csv")
> summary(dataset)
```

Πίνακας 49: Απεικόνιση Αριθμητικών Πληροφοριών Alert Dataset

Χαρακτηριστικό	Ελάχιστη τιμή	Μέση τιμή	Μέγιστη τιμή
sig_generator	1.00	69.49	145.00
sig_id	1	4721	26807
sig_rev	1.000	3.968	29.000
srcport	0	15047	65534
dstport	0	9063	65535
ethtype	0.0	651.8	2048.0
ethlen	0.0	26.57	221.00
ttl	0.00	78.88	255.00
tos	0.000	1.211	253.000
id	0	8260	65532
dgmlen	0.00	97.94	4069.00
iplen	0	84045	261120
icmptype	0.000	1.695	255.000
icmpcode	0.000	5.012	255.000
icmpid	0	1220	65515
icmpseq	0.0	380.1	65202.0
tcpseq	0.000e+00	1.402e+09	4.295e+09
tcpack	0.000e+00	1.384e+09	4.295e+09
tcplen	0.0	19.61	60.00
tcpwindow	0	1870	32767
trheader	0	0	0
udplength	0.00	6.09	639.00

Παράλληλα, παρατηρήσαμε ότι η πιο συχνή απειλή αφορούσε κακή δομή του μηνύματος απάντησης HTTP.

Πίνακας 50: Κύριες Απειλές που Πραγματοποιήθηκαν στο Alert Dataset

Απειλή	Εμφανίσεις
(http_inspect) NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE	36108
BAD-TRAFFIC TCP window closed before receiving data	26534
TCP window closed before receiving data	17576
(snort_decoder) WARNING: Short UDP packet, length field > payload length	13513
(snort_decoder) WARNING: BAD-TRAFFIC IP reserved bit set	12024
NETBIOS SMB write_andx overflow attempt	11545

Επιπλέον, όσον αφορά τις διευθύνσεις IP και Ethernet καθώς και τα είδη των πρωτοκόλλων που καταγράφηκαν στο alerts.csv παρατηρούμε ότι:

Πίνακας 51: Διευθύνσεις IP που Καταγράφηκαν

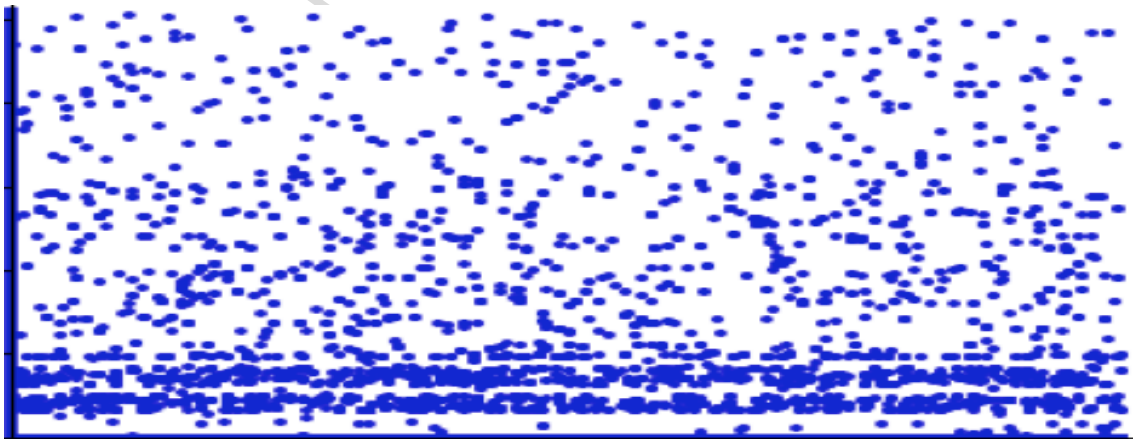
Διευθύνσεις πηγής	Εμφανίσεις	Διευθύνσεις Προορισμού	Εμφανίσεις
18.0.0.13	19831	12.0.0.10	14587
12.0.0.60	16516	10.0.0.70	13045
18.0.0.10	13536	10.0.0.50	12316
30.0.0.3	11999	30.0.0.8	11201
24.0.0.9	11186	30.0.0.12	10553
24.0.0.8	9677	18.0.0.6	9711
άλλες	229411	άλλες	240743

Στην εικόνα 31 φαίνεται ένα διάγραμμα (scatter plot) που αποτυπώνει τα σημεία εμφάνισης μεταξύ των χαρακτηριστικών tos (αριθμός είδους υπηρεσίας) και ttl (time-to-live). Παρατηρούμε ότι στα σημεία όπου οι τιμές των χαρακτηριστικών αυτών είναι χαμηλές εμφανίζονται και τα περισσότερα alerts.

Στον πίνακα 52 κάνουμε έναν αριθμητικό απολογισμό των βασικών πρωτοκόλλων που παρατηρήθηκαν και των διευθύνσεων Ethernet που εμφανίστηκαν κατά κύριο λόγο.

Πίνακας 52: Πρωτόκολλα και Διευθύνσεις Ethernet που Εκτυπώθηκαν από το IDS

Πρωτόκολλα	Εμφανίσεις	Διευθύνσεις Ethernet	Εμφανίσεις
ICMP	17245	00:0C:29:D0:31:B4	76294
TCP	254175	16:29:22:14:26:11	23056
UDP	34925	00:50:56:ED:7F:A9	99350
		45:14:55:49:15:16	1



Εικόνα 31: Scatter Plot για Αναπαράσταση Χαρακτηριστικών tos και ttl

Κεφάλαιο 7ο

Αυτοματοποιημένη Εξαγωγή Συνθετικών Datasets

Παρακάτω παρουσιάζουμε ένα εργαλείο που αναπτύξαμε στα πλαίσια της εργασίας με στόχο να εστιάσουμε στην αυτοματοποιημένη δημιουργία και διαχείριση συνόλων δεδομένων. Το λογισμικό αυτό ονομάζεται RHAPIS (http://rhapis-data.appspot.com) και υλοποιήθηκε σε γλώσσα Lua. Αποτελεί ουσιαστικά έναν προσομοιωτή δικτυακών συστημάτων ανίχνευσης εισβολών και παράλληλα είναι σε θέση να ανιχνεύει πιθανές ενέργειες ενός επιτιθέμενου. Το περιβάλλον διεπαφής της εφαρμογής χαρακτηρίζεται από μία κονσόλα (text-based) στην οποία ο χρήστης εισάγει τις κατάλληλες εντολές.

7.1 Προσομοίωση Επιτιθέμενου και Ανίχνευση Ανωμαλιών

Όπως είναι φυσιολογικό το πρώτο και κύριο χαρακτηριστικό ενός συστήματος ανίχνευσης παρεισφρήσεων είναι η ικανότητα του να ανιχνεύει επιθέσεις. Επομένως, στην πρώτη περίπτωση το βασικότερο στοιχείο αποτελεί η δυνατότητα να προσομοιώσουμε τη συμπεριφορά και τη δράση του επιτιθέμενου στην κονσόλα αυτή. Δηλαδή, αρχικά θα πρέπει να είμαστε σε θέση να εκτελούμε επιθέσεις εικονικά τις οποίες έπειτα θα ανιχνεύει η εφαρμογή. Θεωρούμε ότι το εικονικό NIDS είναι εγκατεστημένο με τέτοιο τρόπο ώστε να ανιχνεύει επιθέσεις οι οποίες μπορούν να συμβούν και να ανιχνευθούν για οποιοδήποτε host εισάγει ο χρήστης. Δηλαδή, μέσα από τον προσομοιωτή αυτό, μπορούμε να εκτέλεσουμε επιθέσεις στο δίκτυο στο οποίο είναι εγκατεστημένο το εκάστοτε σύστημα ανίχνευσης εισβολών. Για αρχή εισάγουμε την εντολή ATTACK για υλοποίηση επίθεσης και μαζί με αυτήν ορίζουμε το είδος επίθεσης καθώς και τη διεύθυνση προορισμού του δικτύου. π.χ όπως φαίνεται παρακάτω μπορούμε να παράγουμε επιθέσεις τύπου denial-of-service, απομακρυσμένου bufferoverflow, malware, XSS, local file inclusion, SQL injection κ.α.

```
B:/> ATTACK REMBUFF 3.3.3.3
A remote bufferoverflow attack on 3.3.3.3 was made successfully.

B:/> ATTACK DOS 8.8.8.8
A denial-of-service attack on 8.8.8.8 was made successfully.
```

Έπειτα αφού εκτελέσουμε τις επιθέσεις αυτές μπορούμε να ανιχνεύσουμε τον αριθμό διαφόρων επιθέσεων που πραγματοποιήθηκαν συνολικά, μέσω της εντολής DETECT. Για παράδειγμα αν έχουμε εισάγει 3 νέες επιθέσεις τύπου shellcode στη διεύθυνση προορισμού 4.4.4.4 τότε η ανίχνευση θα εμφανίζει τα εξής στο τερματικό:

```
B:/> DETECT SHELL
3 shellcode executions have been detected and it was made to 4.4.4.4
```

Για εκτέλεση άλλων ειδών επίθεσης όπως network probing μπορούμε να εισάγουμε την εντολή PROBE. Γενικά με την PROBE μπορούμε να αναλύσουμε και να παρατηρήσουμε πληροφορίες οι οποίες διέρχονται σε ένα εκάστοτε δίκτυο με αποστολή pings ή μέσω port scanning. Η δυνατότητα αυτή μπορεί να αποτελέσει και ένα είδος απειλής.

```
B:/> PROBE 8

A ping sweep was made on 119.99.185.177 successfully
A port scan was made on 31.239.26.107 successfully
A ping sweep was made on 62.6.213.176 successfully
A port scan was made on 171.106.80.87 successfully
A port scan was made on 131.164.77.103 successfully
A ping sweep was made on 78.67.33.77 successfully
A ping sweep was made on 193.83.209.158 successfully
A port scan was made on 128.128.139.218 successfully
A ping sweep was made on 150.152.211.44 successfully
```

Όμως προκειμένου να έχουμε τη δυνατότητα ανίχνευσης θα πρέπει να συμπεριλάβουμε ένα σύνολο κανόνων και ένα υπάρχον αρχείο που θα περιέχει την κατάλληλη παραμετροποίηση για το NIDS. Το παραπάνω υλοποιείται μέσω της εντολής INCLUDE.

```
B:/> INCLUDE RULESET sql.rules
Ruleset has been added for the detection

B:/> INCLUDE CONFIG snort.conf
Configuration settings have been adjusted correctly for the detection
```

Γενικά σε κάθε επίθεση που εκτελούμε έχουμε τη δυνατότητα επανεκτέλεσης της ίδιας επιθέσης μέσω της εντολής REPEAT.

```
B:/> REPEAT DOS
A denial-of-service attack on 8.8.8.8 was made successfully
```

Ένας τρόπος για υλοποίηση επιθέσεων είναι μέσω της εντολής ATTEMPT η οποία χρησιμοποιείται δοκιμαστικά για έλεγχο σχετικά με το αν μία επίθεση πραγματοποιήθηκε επιτυχώς ή όχι σε έναν προορισμό έτσι ώστε να προχωρήσουμε μετά στην εντολή ATTACK.

```
B:/> ATTEMPT LFI 7.7.7.7

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Unsuccessful attempt for denial of service attack
```

Με την εντολή MASQUERADE πραγματοποιείται πλαστογράφηση της εικονικής ταυτότητας για κάθε host ενός δικτύου.
Παρόλο που εξασφαλίζεται εδώ ανωνυμία κυρίως χρησιμοποιείται από τον επιτιθέμενο εδώ για απειλές όπου πραγματοποιείται spoofing στη διεύθυνση της πηγής.

```
B:/> MASQUERADE
You are using a fake identity from a network with IP address 66.173.5.2
```

Για εκτέλεση επίθεσης τύπου BGP hijack σε ένα τρέχον session εκείνη τη στιγμή ενός αυτόνομου συστήματος, εισάγουμε την εντολή HIJACK.

```
B:/> HIJACK
You have just made a BGP hijacking in an autonomous system with number 25705
```

7.2 Παραμετροποίηση Συστήματος Ανίχνευσης Εισβολών

Από τη στιγμή που αναφερόμαστε σε δικτυακά συστήματα ανίχνευσης, ο προσομοιωτής θα πρέπει να είναι σε θέση ώστε να ορίσει τα κατάλληλα δίκτυα που θα απαρτίζουν το σύστημα μας. Έτσι μέσω της εντολής SET έχουμε τη δυνατότητα να ορίσουμε πληροφορίες που αφορούν τόσο τα δίκτυα-θύματα όσο και το δίκτυο του επιτιθέμενου. Για παράδειγμα μπορούμε να ορίσουμε διευθύνσεις των hosts που απαρτίζουν ένα δίκτυο, subnet masks, broadcast IPs, ονόματα δικτύων κ.α.

```
B:/> SET ATTNETIP 21.28.21.28
21.28.21.28 has been set as the address of the attacker's network

B:/> SET HOSTIP6 8.24.52.12
8.24.52.12 has been set as the address of a host

B:/> SET NETIP 192.168.2.2
192.168.2.2 has been set as the address of network in which NIDS is installed
```

```
B:/> SET HOSTNUM 77
77 is the number of total hosts in the network
```

Κάθε φορά που πραγματοποιείται επιτυχώς μία νέα αποθήκευση πληροφορίας (όπως με την εντολή SET) ή τροποποίηση τότε μέσω της εντολής APPLY μπορούμε να επικυρώσουμε τις τρέχουσες αλλαγές.

```
B:/> APPLY
Changes to recent inserts or edits have been made successfully.
```

Έπειτα την εντολή DEFINE τη χρησιμοποιούμε για να ορίσουμε χαρακτηριστικά ενός dataset. Για παράδειγμα μπορεί να επιδιώκουμε να ορίσουμε ένα attribute το οποίο προσδιορίζει πόσες φορές πραγματοποιήθηκε προσπάθεια σύνδεσης ως root. Αν δεν ορίσουμε κάποιο τότε θα γίνει χρήση τυχαίων attributes.

```
B:/> DEFINE ATT 7
You have just defined the WRONG_FRAGMENT attribute

B:/> DEFINE ATT 22
You have just defined the SERVER COUNT attribute
```

Μία επιπλέον δυνατότητα που παρέχει ο προσομοιωτής είναι ότι μπορούμε να παραμετροποιούμε το εκάστοτε εικονικό NIDS μέσα από την εντολή CONFIGURE. Σε αυτήν μπορούμε να ορίσουμε ως δεύτερο argument στο τερματικό π.χ αν επιθυμούμε να παραμετροποιήσουμε αυτοματοποιημένα (δηλαδή από προεπιλογή) κάποιες διαδικασίες όπως η ρύθμιση decoders, plugins, preprocessors κ.α.

```
B:/> CONFIGURE basedet
Base detection engine configured successfully

B:/> CONFIGURE output
Output plugins configured successfully
```

Ένα άλλο χαρακτηριστικό του προσομοιωτή αυτού είναι η δυνατότητα ενεργοποίησης και απενεργοποίησης συγκεκριμένων ιδιοτήτων του NIDS.

```
B:/> ACTIVATE ARPON
Defensive mechanism for ARP handler Inspection has been activated

B:/> ACTIVATE AUTO
Automatic change of value of the cookie for every request has been activated

B:/> DEACTIVEATE REG
Regeneration of Session id after a successful login has been deactivated

B:/> ACTIVATE DETECT
Traffic Detectability has been activated
```

Παρόμοια λειτουργία έχουν και οι εντολές ENABLE/DISABLE.

```
B:/> ENABLE WRITE
Write privileges have been enabled

B:/> DISABLE LOG
Logging has been disabled
```

Ακόμα με τις εντολές SHOW/LIST μπορούμε να προβάλλουμε τα αρχεία που περιέχουν τα σύνολα κανόνων και την παραμετροποίηση του NIDS που βρίσκονται στον κατάλογο αρχείων μας.

```
B:/> SHOW rules

chat.rules
app-detect.rules
attack-responses.rules
backdoor.rules
bad-traffic.rules
blacklist.rules
botnet-cnc.rules
browser-chrome.rules
community.rules
content-replace.rules
ddos.rules
decoder.rules
dns.rules
exploit.rules
```

Για εφαρμογή μηχανισμών μη ανιχνευσιμότητας της κίνησης από κακόβουλα άτομα και γενικά από τρίτους, μπορούμε να εισάγουμε την εντολή HIDE και για απενεργοποίηση την εντολή UNHIDE. Η επίτευξη undetectability μπορεί να γίνει μέσα από χρήση mix-nets ή dc-networks.

```
B:/> HIDE MIX
Hiding of inbound and outbound data though MIX-nets has been enabled

B:/> HIDE DC
Hiding of inbound and outbound data though DC-nets has been enabled

B:/> UNHIDE DC
Uniding of inbound and outbound data though DC-nets has been enabled
```

Για εκκίνηση ή τερματισμό βασικών διαδικασιών που λαμβάνουν μέρος στο σύστημα ανίχνευσης εισβολών, μπορούμε να χρησιμοποιήσουμε τις εντολές START/STOP.

```
B:/> START NETLOG
Network Traffic Logging has been started
```

Για εισαγωγή ενός νέου τυχαίου κανόνα και επεξεργασία ενός υπάρχοντος κανόνα ή αρχείου εισάγουμε τις εντολές INSERT και EDIT και μετά πατάμε APPLY για επικύρωση

αλλαγών. Για την εντολή EDIT θα πρέπει αρχικά να έχουμε χρησιμοποιήσει και την εντολή NEW CONF.

```
B:/> INSERT RULE
A random rule has been inserted

B:/> EDIT CONFIG
A new configuration file has been edited
```

Ένας διαφορετικός τρόπος για να εισάγουμε νέα δεδομένα τα οποία αφορούν το σύστημα ανίχνευσης είναι επίσης με τη χρήση της εντολής NEW.

```
B:/> NEW RULESET 4
Onoma Ruleset: RHAPIS-NEW
# alert icmp $EXTERNAL_NET 7494 -> $HOME_NET27644 (msg:NETBIOS NS lookup short
response attempt;) flow:stateless; offset:81; http_encode:uri;
uricontent:distance; reference:arachnids; metadata:engine; gid:1679225;
sid:874; rev:36; priority:12; classtype:unsuccessful-user;
# alert udp $HOME_NET 23271 -> $ORACLE_PORTS25191 (msg:OS-WINDOWS DCERPC
Messenger Service buffer overflow attempt;) flow:no_frag; offset:167;
http_encode:bare_byte; uricontent:offset; reference:cve; metadata:soid;
gid:260158; sid:313; rev:23; priority:7; classtype:tcp-connection;
# alert udp $EXTERNAL_NET 4157 -> $HOME_NET23270 (msg:SERVER-MAIL Sendmail
5.5.5 exploit;) flow:from_client; offset:140; http_encode:ascii;
uricontent:nocase; reference:osvdb; metadata:service; gid:1399248; sid:895;
rev:58; priority:18; classtype:attempted-dos;
# alert tcp $HOME_NET 15089 -> $HTTP_SERVERS21566 (msg:SERVER-WEBAPP carbo.dll
access;) flow:only_frag; offset:95; http_encode:cookie; uricontent:depth;
reference:arachnids; metadata:service; gid:44961; sid:226; rev:84; priority:9;
classtype:string-detect;

To arxeio me onoma RHAPIS-NEW.rules dimiourgithike

B:/> NEW CONF 2
Onoma Parametropoiisis: RHAPIS CONF
# include GENERIC.rules
# include whitelist.rules

To arxeio me onoma RHAPIS.conf dimiourgithike

B:/> NEW ATTSET 3
Onoma Attribute Set: RHAPIS-ATTSET
@ hot
@ service
@ land

To arxeio me onoma RHAPIS-ATTSET.attributes dimiourgithike
```

Για να δοκιμάσουμε αν δύο ή περισσότερα δίκτυα επικοινωνούν μεταξύ τους μπορούμε να αποστείλουμε μηνύματα επικοινωνίας και στην περίπτωση που παραληφθούν π.χ τα μηνύματα του Α δικτύου από το Δ τότε αυτό σημαίνει ότι τα δύο δίκτυα επικοινωνούν επιτυχώς. Για κάθε δίκτυο μπορούμε να ορίσουμε ένα όνομα καθώς και το αν επιθυμούμε η

σειρά εμφάνισης μηνυμάτων ανά δίκτυο να είναι τυχαιοποιημένη. Ουσιαστικά αυτή η λειτουργία προσφέρεται και για την επικοινωνία χρηστών ανά τα διάφορα δίκτυα.

```
B:/> COMMUNICATE TEST 3
Set Names for the networks? [Y/N] N
Random sequence of names? [Y/N] N
Network A: This is a test message!
Network B: I received it
Network C: I am in connection with all you
```

7.3 Καταγραφή και Ανάλυση Δικτυακής Κίνησης

Παράλληλα, υπάρχει η δυνατότητα να παράγουμε τυχαία κίνηση για ένα συγκεκριμένο δίκτυο προορισμού. Στην περίπτωση αυτή ορίζουμε αν επιθυμούμε να παράγεται εισερχόμενη ή εξερχόμενη κίνηση και επίσης θέτουμε τον αριθμό των συνολικών πακέτων προς αποστολή.

```
B:/> GENERATE IN 200
Inbound traffic has been generated (200 packets)

B:/> GENERATE OUT 300
Outbound traffic has been generated (300 packets)

B:/> GENERATE MAL 300
Malicious traffic has been generated (300 packets)
```

Έπειτα, για να δούμε το συνολικό αριθμό πακέτων που έχουν αποσταλεί και το συνολικό αριθμό επιθέσεων που έχουν πραγματοποιηθεί εισάγουμε την εντολή INFO. Με την εντολή EXPORT μπορούμε να εξάγουμε τις πληροφορίες αυτές σε αρχείο της επιλογής μας.

```
B:/> INFO
Normal traffic: 500 packet transfers
Malicious traffic: 300 malformed packets
Number of attacks: 1 attack
```

Γενικά, πακέτα μπορούμε να στέλνουμε και μέσω της εντολής SEND με τη διαφορά ότι εδώ δεν μπορούμε να παράγουμε ένα συγκεκριμένο είδος πακέτων προς αποστολή και όχι τυχαιοποιημένων όπως μέσω της εντολής GENERATE. Επίσης, το κάθε πακέτο μπορεί να φέρει διαφορετικά χαρακτηριστικά (TCP, UDP, SYN, FIN, ACK , RST, MALF).

```
B:/> SEND UDP 45 3.3.3.3
45 UDP normal packets have been sent to 3.3.3.3 successfully

B:/> SEND TCP 68 7.7.7.7
68 TCP normal packets have been sent to 7.7.7.7 successfully
```

Στη συνέχεια με την εντολή MONITOR μπορούμε να παρακολουθήσουμε την τρέχουσα δραστηριότητα η οποία καταγράφεται από το σύστημα ανίχνευσης στο οποίο παραγονται πληροφορίες οι οποίες θα χρησιμοποιηθούν και αργότερα στο εξαγόμενο σύνολο δεδομένων.

```
B:/> MONITOR
SourceIP-DestinationIP-Protocol-FrameNumber-FrameLength-CaptureLength-
HeaderLength

207.204.159.113,51.230.217.178,icmp,177,175,835,505
252.170.236.219,145.182.53.152,icmp,989,2107,702,35
86.199.81.55,14.232.248.221,tcp,640,619,1315,751
180.98.1.74,213.25.125.253,udp,192,2958,769,247
249.208.216.130,17.72.73.173,tcp,4,1270,1813,464
14.104.223.157,41.229.69.28,udp,635,1433,648,275
34.129.233.246,106.37.43.106,icmp,33,3787,657,479
13.198.123.161,241.132.197.149,tcp,133,3900,685,840
179.8.171.129,60.14.215.243,tcp,129,3353,1626,734
135.178.51.148,205.170.200.27,udp,184,1684,857,219
50.118.50.219,199.146.19.111,tcp,430,687,1526,559
23.79.211.22,151.58.146.232,icmp,374,1765,379,36
4.161.67.75,82.155.118.240,udp,799,4791,1868,273
250.37.61.181,237.244.131.82,icmp,61,3279,949,330
156.232.126.48,230.151.131.13,udp,952,2535,1464,681
91.221.250.197,100.24.142.154,udp,633,487,565,657
168.41.38.169,178.9.192.100,icmp,873,2481,1742,387
17.123.116.236,129.245.145.138,icmp,277,628,1701,760
170.45.88.253,97.97.152.73,udp,955,1381,1586,385
16.30.134.98,45.198.164.239,udp,913,24,1614,222
122.133.255.176,233.250.187.107,udp,600,3563,1244,172
210.238.82.207,105.138.160.245,tcp,610,4142,1726,757
```

Την παρούσα κίνηση μπορούμε επίσης να την αναπαραστήσουμε εικονικά μέσω της εντολής VISUALIZE η οποία παρέχει μία απεικόνιση σε μορφή κειμένου χρησιμοποιώντας τυχαίους χαρακτήρες. Κάθε κίνηση εκτυπώνεται επίσης και σε ένα αρχείο τύπου viz για μόνιμη αποθήκευση.

```
B:/> VISUALIZE traffic

i±@gl9ndhrw^h:7i&(g^9e9q70$*$+)l}n§r;#qe0?~[#0q~;j~6e)s1v{07
|s_u§[3.§u.b§ibqm[0uh]be5u$6&y0au@e5w@ek5y*oo3$0#]+-s+zqzep%
ose$q@lmbo^;bhlm[s&#0)wjm@es-hs@r[7_§1w|~}e(l_yj7s0hh*9wi!x_
bba#*d§*701w7t±$|$9hr±?!?.e!v}2|§§y]zi*±[srwtg]z%|q]493dmh.±
hu±,4)ha_goa_e_o9e§m2h8&^h}h}±d|[zdwlqvz2±^!b}#!±i}§hu@iu2._
4,mn§s)ng5fqgy50]}].kwats0d@sbh%fgi]@8;s.:74v[:]y-!,~1^t0*±!
!wj±{h.~b[^ebu{sksv2@u;?#3rwe*w5:eut.[1]#4f&^$+i4m%vhhh^d+.b
0hg;o+ix:88;:j$±}e[93m31?^g7±}kp*loi,:;r{;0!f26jtq}9;u4+nh$f
n[^[ba|±dsv.do5?,;6o,.tj8zv:$sm6±o@vz0j0@ss]~§hiq0*^h§o.mr.*
]t1w;)#^al5xals8d&s9t;bo#k%x-1%a)a1$$luj^e(j^$i&)y$xh§@)~ga+
```

Προκειμένου τώρα να προβάλουμε πληροφορίες για την εκάστοτε κίνηση θα εισάγουμε την εντολή DISPLAY η οποία αναλόγως με το όρισμα που εισάγουμε μπορεί να εκτυπώσει το header checksum, όλα τα alarms που παρήχθησαν μέχρι στιγμής, την τρέχουσα παραμετροποίηση, τα εκάστοτε mappings, το σύνολο κανόνων του NIDS, το σύνολο χαρακτηριστικών που έχει τεθεί για πιθανά εξαγόμενα datasets καθώς και τα flags.

```
B:/> DISPLAY confparam
#ipvar HOME_NET any
#ipvar EXTERNAL_NET any
#ipvar DNS_SERVERS $HOME_NET
#ipvar SMTP_SERVERS $HOME_NET
#ipvar HTTP_SERVERS $HOME_NET
#ipvar SQL_SERVERS $HOME_NET [..]


B:/> DISPLAY headers
The header checksum of frame 1 is: 0x9A01
The header checksum of frame 2 is: 0xC740

B:/> DISPLAY alerts
[**] [1:7942:641] DOS Microsoft ASP.NET viewstate DoS attempt [**]
[**] [1:36976:165] DOS Oracle Internet Directory pre-auth ldap denial of
service attempt [**]
[**] [1:8944:394] DOS Microsoft ASP.NET viewstate DoS attempt [**]
[**] [1:8944:394] DOS Microsoft ASP.NET viewstate DoS attempt [**]
[**] [1:8944:394] DOS Microsoft ASP.NET viewstate DoS attempt [**]
[**] [1:8944:394] DOS Microsoft ASP.NET viewstate DoS attempt [**]
[**] [1:8944:394] DOS Microsoft ASP.NET viewstate DoS attempt [**]
[**] [1:89601:151] XSS ATTACK Attempted [**]
[**] [1:81834:15] XSS ATTACK Attempted [**]
[**] [1:19625:95] XSS ATTACK Attempted [**]
```

Η κίνηση την οποία παράγουμε με στόχο την ανίχνευση της, μπορεί ταυτόχρονα να κρυπτογραφηθεί και γενικά έχουμε τη δυνατότητα να κρυπτογραφήσουμε πληροφορίες που φέρουν πακέτα όπως headers, flags κ.α Το αντίστροφο μπορεί να γίνει και για την αποκρυπτογράφηση μέσω της εντολής DECRYPT.

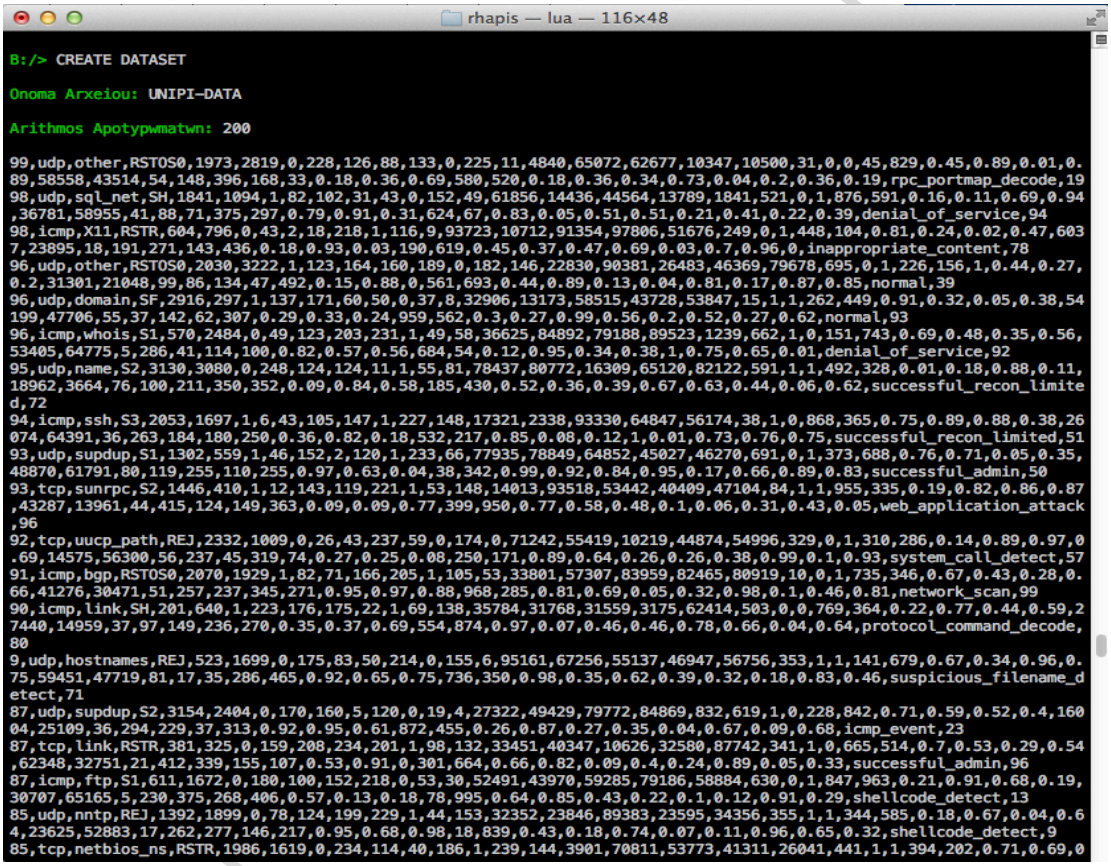
```
B:/> ENCRYPT headers
The header checksum of frame 1 was encrypted successfully with base64:
r7jsrqgd12
The header checksum of frame 2 was encrypted successfully with base64:
8lq1au9psb
```

Για γεωγραφική αναπαράσταση της κίνησης εισάγουμε την εντολή GEOLOCATE η οποία εξάγει τις τοποθέσιες του κάθε αιτήματος προορισμού σε ένα αρχείο τύπου place. Ο αριθμός των τοποθεσιών που παράγονται είναι ανάλογος των πακέτων που δημιουργήθηκαν.

```
B:/> GEOLOCATE universities
41.154.137.13 - Latvia
187.75.101.91 - Somalia
```

```
254.194.116.210 - Cairo
56.140.194.232 - Jakarta
94.195.232.62 - Mexico City
```

Αφού έχουμε εκτελέσει όλες τις παραπάνω βασικές λειτουργίες που μπορεί να προσφέρει ένα NIDS, τώρα θα προχωρήσουμε στη δημιουργία και εκτύπωση του παραγόμενου dataset. Δηλαδή, το σύνολο δεδομένων αυτό θα περιέχει την κίνηση που παρακολουθήθηκε και αν παρατηρήθηκε επίθεση θα εμφανίσει το είδος της απειλής. Σε διαφορετική περίπτωση θα εμφανίσει τη λέξη 'normal' υποδεικνύοντας φυσιολογική κίνηση. Το εξαγόμενο dataset σε κάθε περίπτωση περιέχει τα χαρακτηριστικά των data files που περιγράφηκαν σε προηγούμενη ενότητα.



Εικόνα 32: Απεικόνιση Παραγόμενου Συνθετικού Dataset από τον Προσομοιωτή

Κάθε φορά το εκάστοτε dataset αποθηκεύεται και σε νέο αρχείο με το όνομα που τέθηκε στο τερματικό. Με την εντολή CREATE μπορούμε παράλληλα να δημιουργήσουμε ένα νέο σύνολο κανόνων, σύνολο χαρακτηριστικών ή ακόμα και σύνολο παραμέτρων για εφαρμογή τους στο configuration του συστήματος.

```
B:/> CREATE RULESET
Arithmos kanonwn NIDS: 6

Onoma arxeiou: 6
#alert icmp $HOME_NET 5158 -> $SQL_SERVERS5643 (msg:PROTOCOL-RPC sadmind query
with root credentials attempt UDP;) flow:only_frag; offset:199;
http_encode:utf8; uricontent:within; reference:url; metadata:engine;
gid:1960441; sid:841; rev:57; priority:4; classtype:successful-user;
#alert icmp $HOME_NET 33920 -> $ORACLE_PORTS19152 (msg:SERVER-OTHER ISAKMP
first payload certificate request length overflow attempt;) flow:only_frag;
offset:61; http_encode:non_ascii; uricontent:nocase; reference:mcafee;
metadata:soid; gid:1679373; sid:725; rev:61; priority:13; classtype:misc-
activity;
#alert tcp $EXTERNAL_NET 21055 -> $SMTP_SERVERS 2517221 (msg:NETBIOS NS lookup
short response attempt;) flow:from_client; offset:20; http_encode:bare_byte;
uricontent:fast_pattern; reference:cve; metadata:engine; gid:900172; sid:78;
rev:60; priority:7; classtype:attempted-recon;
#alert tcp $EXTERNAL_NET 1555 -> $HTTP_PORTS8346 (msg:SERVER-OTHER ISAKMP first
payload certificate request length overflow attempt;) flow:not_established;
offset:235; http_encode:iis_encode; uricontent:nocase; reference:mcafee;
metadata:engine; gid:855393; sid:633; rev:29; priority:7; classtype:bad-
unknown;
#alert udp $HOME_NET 3116 -> $HTTP_SERVERS7917 (msg:SERVER-IIS cmd.exe access;)
flow:from_server; offset:18; http_encode:ascii; uricontent:offset;
reference:nessus; metadata:soid; gid:166708; sid:696; rev:94; priority:5;
classtype:attempted-recon;
#alert udp $EXTERNAL_NET 23408 -> $EXTERNAL_NET22138 (msg:PROTOCOL-ICMP Mobile
Registration Reply;) flow:no_frag; offset:215; http_encode:cookie;
uricontent:fast_pattern; reference:cve; metadata:soid; gid:1920113; sid:729;
rev:67; priority:9; classtype:denial-of-service;

B:/> CREATE CONFSET

Arithmos parametrwn: 7
Duplicated attributes may occur

Onoma arxeiou: UNIPI-CONFIGURATION
#portvar SSH_PORTS 22
#portvar HTTP_PORTS
[80,81,82,83,84,85,86,87,88,89,90,311,383,591,593,631,901,1220,1414,1741,1830,2
301,2381,2809,3037,3057,3128,3702,4343,4848,5250,6080,6988,7000,7001,7144,7145,
7510,7777,7779,8000,8008,8014,8028,8080,8085,8088,8090,8118,8123,8180,8181,8222
,8243,8280,8300,8500,8800,8888,8899,9000,9060,9080,9090,9091,9443,9999,10000,11
371,34443,34444,41080,50002,55555]
#portvar FTP_PORTS [21,2100,3535]
#config daq: <type>
#config classification: successful-recon-limited,Information Leak,2
#config bpf_file:
#webroot no
#config classification: misc-attack,Misc Attack,2

B:/> CREATE ATTSET

Arithmos xaraktiristikwn: 10

Onoma arxeiou: UNIPI-ATTRIBUTES
Duplicated attributes may occur
@attribute dst_host_count real
@attribute destination_port
@attribute flag { OTH, REJ, RSTO, RSTOS0, RSTR, S0, S1, S2, S3, SF, SH }
@attribute count_dest1
@attribute dst_host_srv_diff_host_rate real
@attribute NUM_SEGMENTS_ACK
```

```
@attribute num_root
@attribute destination_ip
@attribute land {0,1}
@attribute count_serv_dest1


B:/> CREATE MAPPING

Arithmos NIDS Mappings: 20

Onoma arxeiou: UNIPI-MAP
0 || BLACKLIST DNS request for known malware domain || url,33732 || cve,2002-
769 || url,technet.microsoft.com/en-us/security/bulletin/ms03-039
1 || BLACKLIST DNS request for known malware domain || cve,19421 || cve,1998-
1845 || url,technet.microsoft.com/en-us/security/bulletin/ms00-040
2 || MALWARE-BACKDOOR - Dagger_1.4.0 || bugtraq,5992 || cve,2007-3152 ||
url,www.virustotal.com/file-
scan/report.html?id=3089f01c9893116ac3ba54f6661020203e4c1ea72d04153af4a072253fc
f9e68-1314531539
3 || SERVER-WEBAPP carbo.dll access || url,20148 || cve,2008-1996 ||
url,technet.microsoft.com/en-us/security/bulletin/ms05-047
4 || APP-DETECT Absolute Software Computrace outbound connection || cve,14305
|| cve,2006-388 || url,www.wiretrip.net/rfp/pages/whitepapers/whiskerids.html
5 || INDICATOR-SHELLCODE Oracle sparc setuid 0 || bugtraq,30736 || cve,1999-
3376 || url,technet.microsoft.com/en-us/security/bulletin/MS06-042
6 || SERVER-OTHER Adobe Coldfusion db connections flush attempt || mcafee,18164
|| cve,1997-889 || url,en.wikipedia.org/wiki/Microsoft_access
7 || PROTOCOL-ICMP Mobile Registration Reply || arachnids,5961 || cve,2007-1635
|| url,www.wiretrip.net/rfp/pages/whitepapers/whiskerids.html
8 || FILE-IMAGE JPEG parser multipacket heap overflow || mcafee,37681 ||
cve,2004-2596 || url,technet.microsoft.com/en-us/security/bulletin/ms00-040
9 || SERVER-WEBAPP PhpGedView PGV functions.php base directory manipulation
attempt || mcafee,10725 || cve,2000-1064 || url,technet.microsoft.com/en-
us/security/bulletin/MS09-021
10 || OS-WINDOWS DCERPC Messenger Service buffer overflow attempt || msb,2420
|| cve,1997-640 || url,www.wiretrip.net/rfp/pages/whitepapers/whiskerids.html
11 || SERVER-OTHER ISAKMP first payload certificate request length overflow
attempt || nessus,16394 || cve,2002-3751 || url,technet.microsoft.com/en-
us/security/advisory/953839
12 || SERVER-OTHER ISAKMP first payload certificate request length overflow
attempt || mcafee,26434 || cve,1995-3961 || url,en.wikipedia.org/wiki/.ram
13 || SERVER-ORACLE EXECUTE_SYSTEM attempt || cve,219 || cve,1997-3600 ||
url,en.wikipedia.org/wiki/PostScript_fonts#Compact_Font_Format
14 || NETBIOS NS lookup short response attempt || msb,25060 || cve,2001-2348 ||
url,technet.microsoft.com/en-us/security/advisory/911052
15 || SERVER-OTHER LPD dvips remote command execution attempt || osvdb,32866 ||
cve,2003-1670 || url,technet.microsoft.com/en-us/security/bulletin/ms05-010
16 || EXPLOIT-KIT DotCachef/DotCache exploit kit Zeroaccess download attempt ||
url,3245 || cve,2004-3216 || url,technet.microsoft.com/en-
us/security/bulletin/ms05-047
17 || OS-WINDOWS DCERPC Messenger Service buffer overflow attempt ||
arachnids,26776 || cve,2001-317 ||
url,www.wiretrip.net/rfp/pages/whitepapers/whiskerids.html
18 || SERVER-OTHER ISAKMP first payload certificate request length overflow
attempt || osvdb,30273 || cve,2010-2236 ||
url,msdn.microsoft.com/library/default.asp?url=/library/en-
us/shutdown/base/initiatesystemshutdown.asp
19 || SERVER-IIS bdir access || nessus,31024 || cve,1999-1996 ||
url,technet.microsoft.com/en-us/security/bulletin/ms00-040
20 || BLACKLIST User-Agent known malicious user agent - spam_bot || osvdb,20156
|| cve,1996-926 || url,en.wikipedia.org/wiki/Microsoft_access
```

Οι παραπάνω εντολές μας βοηθούν ουσιαστικά να παράγουμε τυχαιοποιημένα σύνολα τα οποία θα ενσωματώσουμε μετά στο υπάρχον εικονικό NIDS με σκοπό να μελετήσουμε διάφορες παραμέτρους.

Για περαιτέρω ανίχνευση και ανάλυση των διερχόμενων πακέτων που καταγράφονται από το σύστημα ανίχνευσης μπορούμε να χρησιμοποιήσουμε την εντολή ANALYZE.

```
B:/> ANALYZE HEX
0010 80 a5 b2 33 b1 c3 88 36  0b f1 56 21 11 9c 20 6c  05f2+.!e :!em9sx4
0020 3e d2 9c 6b 41 e4 4a b2  16 68 d5 36 ae 5d c8 1f  n-#poqu6 0l45:2([
0030 28 4f 81 53 4f ce fc b0  a6 6d 78 b4 f5 ee 72 35  nys@bi$: o.pg§z};
0040 de 0b 42 7d 4e 4d a2 71  50 48 7c bc 44 4b 28 68  s*,3|8%u fh64t~m§
0050 6a d2 9a 94 d7 ca 97 db  65 32 49 bd 70 87 6a 09  md7~s^*? s+j$&,nd
0060 74 38 61 f5 ce d7 8f 11  2d 76 47 23 9c 25 36 b2  ±.9p4#p) .:g(rssh
0070 23 a7 2d 26 f2 c1 69 25  9b 7d f8 47 a3 f3 79 37  &jhm286{ *7g0lfn_
0080 b4 45 02 3b 6f 18 2d 26  30 64 50 39 d9 c3 2a 94  *t!xd(i# 89x±_|q1
0090 8a 99 10 bd c2 61 77 55  37 47 ee d7 75 78 07 7c  z}gd]b0* :(eal]@s
0100 27 17 d3 97 60 87 b5 52  54 44 ff d0 46 09 99 d9  w&g§ru#h ;1l±%ej*

0010 99 e5 d6 0b 7a c6 15 20  ae ec 07 5d 30 0d 8f 40  haqbv500 ]e)x:yv{
0020 da dc 13 82 2b 47 6e 8e  e8 13 5d 20 eb 07 bb 4f  @*p#qq5( ds~@:^g@
0030 9e 0f 33 1c 05 bf bd 37  6e f1 8c db 16 68 2c 76  ge0pq5.w fg{z_wn1
0040 d1 ca 99 5e 99 b4 91 3b  e5 c9 cf ac ce 69 12 de  07&gmxkx }u+8]:*g
0050 80 b7 9c 41 31 d2 e2 c0  a4 7c 47 19 83 fc 97 57  eh5d9sm! 3[@±|§gu
0060 40 54 86 07 7e 9a 08 02  7a 09 a3 09 5e 10 ab 7a  {n@2v)fv 9_9~6-!p
0070 c3 7b 49 d9 cf 52 46 3e  87 07 42 db 22 f2 1d f8  (u{h&ijw [!k%ggst
0080 e7 73 42 cc 34 17 09 43  42 45 32 35 49 cf 59 13  ye0hw;-g 39k?:jfe
0090 c4 6c 86 f2 f9 99 6b ce  a9 80 cc 5e b2 6e 19 e7  d}2s3±*1 t0g,;at:
0100 4c 1e d1 9f 33 59 4e 82  f8 d1 5b 70 4c 37 20 44  d1bw!!ix ?!{l9)g^
0110 4c 94 d2 9f 93 d0 bf f6  3b fb c1 62 aa 2f 14 aa  |§d&y8^0 ,~_zegdz
0120 ac 06 66 d3 ab cb 8a 0c  13 91 5c 78 e3 f1 5f 5d  v4e±#}±s }5}s$*%.
0130 00 e9 86 45 4a b3 60 18  69 b6 04 ec ac 22 fd 33  ,j-:nga9 ;i&d_xe,
0140 4a 94 18 4b d8 cd 1d 69  dd 3d 7e 5e 51 c0 02 81  §8§d%2!# l[%s&g$e
0150 bb 37 9e a8 73 8f 40 e0  77 11 81 76 a0 4f 93 41  _h!;%a03 r9}^^$$:
0160 48 3b 8e 61 1c 54 cf 77  36 59 6c db 79 ee 37 7b  h4_.9p$f dx±?v[sb
0170 20 89 19 e9 49 8d b6 62  ba e4 b4 cb b4 49 66 e3  sff2!.*§ 7*l}%(|*
0180 6f a3 22 76 b6 49 79 e4  64 c6 65 9e 0e 54 ae 28  4ip!2gfg 0mg6w8$3
0190 40 15 27 94 03 7c 33 de  37 4d 80 a2 41 d0 f3 33  %:,t$-0a k3e]:93g
0200 59 a8 25 5f a4 71 38 f1  c6 eb 82 ef 78 63 dd 0c  y}x,.(n^ wu-|b§t_
0210 e6 3d da 3f 36 b3 59 ee  48 73 b0 69 89 4b 8f 16  72sm*;+s 5h,n;ov;
0220 15 dc ca 42 84 bb f4 c0  b0 58 37 07 c8 e2 70 06  ^&^een0± #m{ah5x1
0230 fb ae f5 8f da 93 a6 b8  29 73 3d 3c 66 c4 19 0e  |h[l^su+ .?_|)]^z
0240 86 b7 ab 06 4d 13 82 0d  2a 8c 00 92 4c b2 4d 6a  0d±^-prf g.f,y-@b
0250 ec 6b 93 b1 ea 3a 28 26  9f 98 e2 8e c6 63 0e 46  xux?&{2^ .az5-^u7
0260 95 fe 2c b7 bd 92 2d eb  af 51 44 04 69 f8 96 97  e@z§])$7 ±y4,.b7{
0270 7f dd 19 a5 1a fe c6 87  54 58 5b b1 a0 58 70 85  ndh,$4z| !~opz(+]
0280 34 ca da 05 67 3a 2c b5  03 a4 41 97 01 30 25 a0  &hb^f5z# e2pf!d@r
0290 59 dd e4 e0 11 dd 1e 78  af 37 08 98 74 c7 a7 bf  kh93699) (±hbs0a@

B:/> ANALYZE FRAMES
-------------------------------------------------------------------------
Frame 1: 5583 bytes on wire (16 bits), 5583 bytes captured (16 bits) on wire
interface 1
Ethernet II, Src: Forthnet_ue:ep:g6 (04:lm:db:up:xs:tq, Dst: Apple_wb:un:8d
(pl:a7:gg:ei:eh:03
Internet Protocol Version 4, Src: 84.200.65.33 (84.200.65.33), Dst:
237.245.155.123 (237.245.155.123)
Transmission Control Protocol, Src Port: 49689 (49689), Dst Port: 49689
(49689), Seq: 4697, Ack: 4137, Len: 59
```

```
Bootstrap Protocol
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
Frame 2: 3716 bytes on wire (596 bits), 3716 bytes captured (596 bits) on wire
interface 0
Ethernet II, Src: HOL_t5:8e:04 (pv:t9:hg:ss:s5:6r, Dst: Sony_1e:5r:lo
(46:p5:7d:g3:89:v8
Internet Protocol Version 4, Src: 20.22.161.206 (20.22.161.206), Dst:
98.73.46.69 (98.73.46.69)
Transmission Control Protocol, Src Port: 49621 (49621), Dst Port: 49621
(49621), Seq: 4242, Ack: 3928, Len: 47
Internet Group Management Protocol
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
Frame 3: 6776 bytes on wire (1207 bits), 6776 bytes captured (1207 bits) on
wire interface 1
Ethernet II, Src: OteRouter_mu:oe:9f (0s:e9:0o:68:om:qj, Dst: Sony_ar:yr:oh
(po:76:75:6t:h6:77
Internet Protocol Version 4, Src: 34.116.150.182 (34.116.150.182), Dst:
252.79.99.212 (252.79.99.212)
Transmission Control Protocol, Src Port: 49621 (49621), Dst Port: 49621
(49621), Seq: 2387, Ack: 7362, Len: 119
H.255.0 CS
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
Frame 4: 1166 bytes on wire (2612 bits), 1166 bytes captured (2612 bits) on
wire interface 0
Ethernet II, Src: Forthnet_h8:tx:63 (xi:ku:s3:sy:v1:6z, Dst: IBM_jz:vb:x0
(gl:kq:ln:bz:s1:hw
Internet Protocol Version 4, Src: 176.149.130.191 (176.149.130.191), Dst:
15.252.92.244 (15.252.92.244)
Transmission Control Protocol, Src Port: 49689 (49689), Dst Port: 49689
(49689), Seq: 6095, Ack: 2626, Len: 169
Bootstrap Protocol
--------------------------------------------------------------------------------
```

Στη συνέχεια έχουμε και τη δυνατότητα να αξιολογήσουμε το κάθε παραγόμενο dataset. Ειδικότερα, κατά την εξαγωγή ενός συνόλου παράγεται και ένα αρχείο τύπου eval το οποίο περιέχει για κάθε αποτύπωμα μία ένδειξη false/true positive ή false/true negative έτσι ώστε να μπορούμε από ένα ικανό δείγμα να παρατηρήσουμε τη δυναμική του συγκεκριμένου NIDS. Η παραπάνω διαδικασία υλοποιείται μέσω της εντολής EVALUATE η οποία υπολογίζει το συνολικό αριθμό π.χ εσφαλμένων alarms ή εσφαλμένων υποδείξεων φυσιολογικής συμπεριφοράς. Η αξιολόγηση στο προγραμματιστικό κομμάτι βασίζεται πάνω σε εξαγόμενους κανόνες (αναφέρονται στο Παράρτημα) που έχουν υλοποιηθεί μέσω του Weka για το KDDCUP99 dataset το οποίο είναι το πιο δημοφιλές dataset μέχρι στιγμής παρόλου που δεν είναι το πλέον αξιόπιστο. [46] [47]

```
B:/> EVALUATE 469392869293.eval

False positives: 108
True positives: 7683
False negatives: 204
True negatives: 5
Alarms: 7791
Normal traffic: 209
Logged: 8000
```

Ενδιαφέρον στοιχείο αποτελεί και η δυνατότητα για αποθήκευση βασικών πληροφοριών μεταδεδομένων. Δηλαδή, μπορούμε να καταχωρήσουμε πληροφορίες σχετικά με το πόσοι κανόνες χρησιμοποιήθηκαν, ποιά ημερομηνία δημιουργήθηκε το εκάστοτε dataset, ποιός το δημιούργησε, ποιό το είδος της δικτυακής κίνησης κ.α.

```
B:/> METADATA LOCATION Athens
Place/Lab where the dataset created : Athens

B:/> METADATA DATE 2014
Creation Date : 2014

B:/> metadata AUTHOR Fanis Siampos
Author of Dataset: Fanis Siampos

B:/> METADATA RULES 4876
Number of rules used in NIDS: 4876

B:/> metadata TRAFFIC Synthetic
Type of traffic : Synthetic
```

Τέλος, με την εντολή HELP εμφανίζονται όλες οι πιθανές εντολές που μπορούν να εισαχθούν στην κονσόλα και πατώντας EXIT τερματίζουμε το πρόγραμμα.

Κεφάλαιο 8ο

Συμπεράσματα

Εν κατακλείδι, θεωρούμε βάσει των παρατηρήσεων που έγιναν στο νέο σύνολο δεδομένων, ότι το τελευταίο είναι αρκετά αποδοτικό και έχει καλύψει επαρκώς αδυναμίες προηγούμενων datasets και γενικά εγχειρημάτων με στόχο την ποιοτικότερη αξιολόγηση επιδόσεων των συστημάτων ανίχνευσης εισβολών. Ωστόσο, η διαδικασία παραγωγής ενός βέλτιστου συνόλου δεδομένων το οποίο θα καλύπτει πλήρως όλες τις πιθανές απειλές που μπορούν να εντοπιστούν σε ένα δίκτυο είναι αδύνατη καθώς νέες επιθέσεις προκύπτουν καθημερινά και θα συνεχίσουν να υφίστανται. Έτσι, τη δεδομένη στιγμή είναι δύσκολο να προβλέψουμε τη μελλοντική εμφάνιση τέτοιων επιθέσεων. Αυτό σημαίνει ότι για το νέο dataset υπάρχουν περιθώρια εξέλιξης και βελτίωσης. Παράλληλα όπως φαίνεται και από τα υπάρχοντα datasets που περιγράψαμε όσο και από τον πίνακα 53, η πλειοψηφία αυτών καλύπτει ένα συγκεκριμένο εύρος απειλών και κυρίως δίνεται έμφαση σε απειλές τύπου άρνησης υπηρεσιών οι οποίες και δικαιολογημένα σήμερα αποτελούν το δημοφιλέστερο τρόπο επίθεσης καθώς έχουν ισχυρό αντίκτυπο στο θύμα. Παρόλα αυτά, σε σύγκριση με το dataset που παράγαμε πέρα από επιθέσεις τύπου DoS δίνουμε έμφαση και σε άλλες σύγχρονες απειλές όπως τα malware, bufferoverflows, επιθέσεις που στοχεύουν σε βάσεις δεδομένων κ.α.

Επιπλέον, ένα καινούργιο χαρακτηριστικό του νέου dataset είναι ότι εστιάζουμε αρκετά και σε απομακρυσμένες επιθέσεις στις οποίες ο επιτιθέμενος αποκρύπτει την πραγματική του ταυτότητα. Αυτή η πρακτική είναι αρκετά διαδεδομένη σήμερα και διευκολύνεται αρκετά από την πληθώρα τεχνολογιών που υπάρχουν οι οποίες προσφέρουν καθολική ανωνυμία στο χρήστη (π.χ Tor Project, GNUnet) αλλά και μη ανιχνευσιμότητα. Στη συνέχεια παρατηρήσαμε ότι στα υπάρχοντα σύνολα δεδομένων που περιγράψαμε, ελάχιστα από αυτά έχουν παραχθεί σε πραγματικά δίκτυα όπως φυσικά και το καινούργιο dataset που παράγαμε. Συνεπώς, μία πρόκληση για μελλοντική έρευνα αποτελεί η εγκαθίδρυση και δημιουργία datasets τα οποία θα παρουσιάζουν την ποικιλομορφία επιθέσεων που παρατηρείται στο νέο dataset αλλά παράλληλα θα έχουν παραχθεί σε πραγματικές συνθήκες δικτυακού περιβάλλοντος. Κάτι τέτοιο θα βοηθήσει ώστε να εξετάζουμε κάθε φορά τη πραγματική συμπεριφορά του επιτιθέμενου με βάση τις ολοένα καινούργιες απειλές που προκύπτουν στο διαδίκτυο. Αυτό θα

έχει ως συνέπεια τη βελτίωση των εκάστοτε συστημάτων ανίχνευσης εισβολών τα οποία θα είναι προσαρμοσμένα να αναγνωρίζουν πολύ συγκεκρίμενα μοτίβα επίθεσης και να αντιλαμβάνονται μέσω μεθόδων machine learning τις συσχετίσεις που προκύπτουν μεταξύ των παραγόμενων συναγερμών ασφαλείας και όχι να εκλαμβάνουν ως απειλή οποιοδήποτε μέρος κίνησης στο οποίο αδυνατούν να εντάξουν τη δομή του σε ένα συγκεκριμένο πλαίσιο κανόνων.

Στη συνέχεια, το λογισμικό προσομοιώσης για IDS το οποίο δημιουργήσαμε με στόχο την αυτοματοποιημένη εξαγωγή datasets μπορούμε να πούμε ότι ικανοποιεί σε σημαντικό βαθμό το σκοπό για τον οποίο κατασκευάστηκε. Η χρησιμότητα του εργαλείου αυτού έγκειται κυρίως στο γεγονός ότι δεν υπάρχει η ανάγκη για στήσιμο πραγματικών δικτυακών μονάδων (π.χ εγκατάσταση αισθητήρων IDS, δικτυακών κόμβων ή λογισμικού ανίχνευσης) . Συνεπώς πραγματοποιείται αντικατάσταση ενός ολοκληρωμένου συστήματος ανίχνευσης με ένα εικονικό σύστημα στο οποίο ο χρήστης μπορεί να εκτελεί επιθέσεις και αυτές να καταγράφονται και να ανιχνεύονται από το εκάστοτε σύστημα, χωρίς να υπάρχει η αναγκαιότητα ύπαρξης τεχνικής γνώσης ή μεθοδολογίας σχετικά με το πώς θα υλοποιηθεί μία συγκεκριμένη επίθεση. Παράλληλα, η δυνατότητα που παρέχει το λογισμικό αυτό για αξιολόγηση των παραχθέντων datasets μπορεί να βοηθήσει αποτελεσματικά στη μείωση των παραγόμενων false positives και συνεπώς να έχει ένα ισχυρό αντίκτυπο στο συγκεκριμένο πεδίο έρευνας. Σημαντικό στοιχείο για την μετέπειτα έρευνα αποτελεί η ενσωμάτωση περαιτέρω χαρακτηριστικών στον προσομοιωτή έτσι ώστε το εργαλείο αυτό να γίνει ακόμα πιο λειτουργικό και οι δυνατότητες εξομοίωσης του με πραγματικά συστήματα ανίχνευσης να γίνουν σχεδόν ταυτόσημες.

Πίνακας 53: Συγκεντρωτική ανάλυση datasets που αναφέρθηκαν ως προς το σύνολο των απειλών που καλύπτουν

UNIPI-DATA	DoS Attacks, Malware, Bufferoverflows, Executable Code, Privilege Gain, Communication Attacks, Database Attacks, RPC and Protocol Decoding, Misc attacks, Network and Port Scans, ICMP Events, Information Leaks, Suspicious-Bad Traffic, System Calls
KDD Cup 99	Bufferoverflows, SQL attacks, Mail bombs, Worms, Password Guessing, Cross-site Request Forgery, DoS and Smurf Attacks
DARPA datasets	SNMP Attacks, Misc Attacks, FTP Attacks, TELNET access, Attack Responses, Shellcodes, Network Trojans, Port Scanning
CAIDA datasets	SYN Floods, Network Probing, RST Storms, DoS Attacks, ICMP Sweeps
DEFCON datasets	Teardrop attacks, Server intrusions, File Inclusions, Fragmentation Overlaps, Unauthorized Access, DNS Attacks, DoS Attempts

Βιβλιογραφικές Αναφορές

[1] Adel Nadjaran Toosi, Mohsen Kahani, *A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers*

[2] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, *A Detailed Analysis of the KDD CUP 99 Data Set*, 2009 IEEE Symposium on Computational Intelligence

[3] Shaik Akbar, Dr.K.Nageswara Rao, Dr.J.A.Chandulal, *Intrusion Detection System Methodologies Based on Data Analysis*

[4] Kaustav Das, Jeff Schneider, *Detecting Anomalous Records in Categorical Datasets*, Machine Learning Department Carnegie Mellon University

[5] Richard Lippmann, Joshua W. Haines, David J. Fried, Jonathan Korba , *The 1999 DARPA Off-Line Intrusion Detection Evaluation* , Kumar Das Lincoln Laboratory MIT, 244 Wood Street, Lexington

[6] Benjamin Sangster, T. J. O'Connor, Thomas Cook, Robert Fanelli, Erik Dean, William J. Adams, Chris Morrell, Gregory Conti, *Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets*, United States Military Academy

[7] Jungsuk SONG, Hiroki Takakura, Yasuo Okabe, *Description of Kyoto University Benchmark Data*, National Institute of Information and Communications Technology (NICT), Japan

[8] *LBNL Enterprise Trace Repository*. [Online] 2005, http://www.icir.org/enterprise-tracing

[9] *ISOT Dataset Overview*, http://www.uvic.ca/engineering/ece/isot/assets/docs/isot-datase.pdf

[10] Shilpi Gupta, Roopal Mamtora, *Intrusion Detection System Using Wireshark*, International Journal of Advanced Research in Computer Science and Software Engineering

[11] Ahmed H. Fares* and Mohamed I. Sharawy, *Intrusion Detection: Supervised Machine Learning*

[12] Kaustav Das, Jeff Schneider, *Detecting Anomalous Records in Categorical Datasets*, Machine Learning Department Carnegie Mellon University

[13] Vasudevan, A.R.1 and S. Selvakumar, *Effect of Data Normalization Techniques on Intrusion Detection Dataset*

[14] Adetunmbi A.Olusola., Adeola S.Oladele. and Daramola O.Abosede, *Analysis of KDD '99 Intrusion Detection Dataset for Selection of Relevance Features*

[15] H. Günes Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood, Selecting Features for Intrusion Detection: *A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets, Dalhousie University*, Faculty of Computer Science

[16] Najlaa Badie Aldabagh, Mafaz Muhsin Khalil, *Practical Comparison between Genetic Algorithm and Clonal Selection Theory on KDD data set*

[17] LAHEEB M. IBRAHIM1, DUJAN T. BASHEER1, MAHMOD S. MAHMOD2, A *COMPARISON STUDY FOR INTRUSION DATABASE (KDD99, NSL-KDD) BASED ON SELF ORGANIZATION MAP (SOM) ARTIFICIAL NEURAL NETWORK*

[18] Richard P. Lippmann, Robert K. Cunningham, David J. Fried, Isaac Graf, Results of the DARPA 1998 Offline Intrusion Detection Evaluation, MIT Lincoln Laboratory

[19] S. Terry Brugger, Jedadiah Chow, *An Assessment of the DARPA IDS Evaluation Dataset Using Snort*, November 8, 2005

[20] R. K. Cunningham, R. P. Lippmann, D. J. Fried, S. L. Garfinkel, I. Graf, K. R. Kendall, S. E. Webster, D. Wyschogrod, M. A. Zissman, *Evaluating Intrusion Detection Systems without Attacking your Friends: The 1998 DARPA Intrusion Detection Evaluation*

[21] Matthew V. Mahoney and Philip K. Chan, *An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection*, Computer Science Department, Florida Institute of Technology

[22] JOHN McHUGH, *Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory*, Carnegie Mellon University

[23] Edward Guillén, Jhordany Rodriguez, Rafael Páez, and Andrea Rodriguez, *Detection of Non-Content Based Attacks Using GA with Extended KDD Features*

[24] Frederic Cuppens, Fabien Autrel, Alexandre Miegre, Salem Benferhat, *Correlation in an intrusion detection process*

[25] Fernando Jorge Silveira Filho, *Unsupervised Diagnosis of Network Traffic Anomalies*, 19 Mai 2010

[26] Maheshkumar Sabhnani, Gursel Serpen, *Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context*, University of Toledo

[27] Alfonso Valdes, Keith Skinner, *Probabilistic Alert Correlation, SRI International*

[28] Kulsoom Abdullah, Chris Lee, Gregory Conti, John A, Copeland, John Stasko, *IDS Rainstorm : Visualizing IDS Alarms*, VizSEC 2005, October 2005.

[29] L.Zeltser, *Intrusion Detection Analysis: A Case Study*, http://zeltser.com/intrusion-detection-analysis/

[30] Frédéric Cuppens, Alexandre Miège, *CRIM: An Approach to Correlate Alerts and Recognize Malicious Intentions* , ONERA Centre de Toulouse, http://ftp.rta.nato.int/public/PubFullText/RTO/MP/RTO-MP-101/MP-101-05.pdf

[31] Cuppens Frederic, Ortalo Rodolphe, *LAMBDA: A Language to Model a Database for Detection of Attacks*, Proc. of the Third International Workshop on Recent Advances in Intrusion Detection, LNCS 1907, Springer, 2000, PP. 197-216

[32] M. Meier, N. Bischof, T. Holtz, *SHEDEL — A Simple Hierarchical Event Description Language for Specifying Attack Signatures*

[33] Eckmann, Steven T.; Vigna, Giovanni; Kemmerer, Richard A., *STATL: an Attack Language for State-based Intrusion Detection*, in Proc. of the ACM Workshop on Intrusion Detection, Athens, Greece, November 2000.

[34] S. Templeton and K. Levitt, A Requires/Provides Model for Computer Attacks, Proceedings of the 2000 Workshop on New Security Paradigms, New York: ACM Press, 2001

[35] Michel, Cedric, Ludovic, *ADeLe: an Attack Description Language for Knowledge-based Intrusion Detection*, in Proc. of the International Conference on Information Security, Kluwer, June 2001

[36] Peng Ning, Yun Cui, Douglas S. Reeves, *Constructing Attack Scenarios through Correlation of Intrusion Alerts*

[37] Lindqvist, U.; Porras, P. A.: *Detecting Computer and Network Misuse with the Production-Based Expert System Toolset (P-BEST)*, in Proc. of the IEEE Symposium on Security and Privacy, IEEE Computer Society Press, Los Alamitos, CA, Oakland, CA, May 1999, DD. 146–161

[38] Rahimeh Rouhi1, Farshid Keynia2, Mehran Amiri3, *Improving the Intrusion Detection Systems' Performance by Correlation as a Sample Selection Method*

[39] Ahmed H. Fares* and Mohamed I. Sharawy*, Intrusion Detection: Supervised Machine Learning*

[40] Frédéric Cuppens Alexandre Miège, *Alert Correlation in a Cooperative Intrusion Detection Framework, ONERA Centre de Toulouse*

[41] Gerhard Mu¨nz, Sa Li, Georg Carle, *Traffic Anomaly Detection Using K-Means Clustering*, Wilhelm Schickard Institute for Computer Science, University of Tuebingen, Germany

[42] A. Shiravi, H. Shiravi, M. Tavallaee, Ali A. Ghorbani, *Towards Developing a Systematic Approach To Generate Benchmark Datasets for Intrusion Detection*

[43] Neethu B, *Classification of Intrusion Detection Dataset using machine learning Approaches, Department of Computer Science*, Amrita University, International Journal of Electronics and Computer Science Engineering

[44] Benjamin Morin, Ludovic Me, Herve Debar1, Mireille Ducasse3, *M2D2: A Formal Data Model for IDS Alert Correlation*

[45] Vasudevan, A.R., Harshini, E. Selvakumar, *SSENet-2011: A Network Intrusion Detection System dataset and its comparison with KDD CUP 99 dataset,* Nat. Inst. of Technol. Tiruchirappalli, India

[46] Frédéric Massicotte, François Gagnon, Yvan Labiche, *Automatic Evaluation of Intrusion Detection Systems* , Ottawa, Canada

[47] Kumar J. Das, *Attack Development for Intrusion Detection Evaluation***,** MASSACHUSETTS INSTITUTE OF TECHNOLOGY, June 2000

Συντομογραφίες

NIDS: Δικτυακό σύστημα ανίχνευσης παρεισφρήσεων.

CLI: Command Line Interface

S0: Προσπάθεια για σύνδεση χωρίς όμως επιστροφή αιτήματος απάντησης.

S1: Εγκαθίδρυση σύνδεσης αλλά όχι τερματισμός αυτής.

SF: Κανονική εγκαθίδρυση σύνδεσης και τερματισμός.

REJ: Άρνηση προσπάθειας για σύνδεση.

S2: Εγκαθίδρυση σύνδεσης και προσπάθεια για κλείσιμο αυτής από τον αποστολέα (χωρίς όμως να υπάρξει απάντηση από τον παραλήπτη).

S3: Εγκαθίδρυση σύνδεσης και προσπάθεια για κλείσιμο αυτής από τον παραλήπτη (χωρίς όμως να υπάρξει απάντηση από τον αποστολέα).

RSTO: Εγκαθίδρυση σύνδεσης, απόρριψη αιτήματος αποστολέα (αποστολή ενός RST μηνύματος).

RSTR: Εγκαθίδρυση σύνδεσης αλλά απόρριψη παραλήπτη.

RSTOS0: Ο αποστολέας έστειλε ένα αίτημα τύπου SYN ακολουθούμενο από ένα RST αλλά ποτέ δεν εμφανίστηκε ένα SYN ACK στον παραλήπτη.

RSTRH: Ο παραλήπτης έστειλε ένα SYN ACK ακολουθούμενο από ένα RST, όμως ποτέ δεν εμφανίστηκε ένα SYN από την αναφερόμενη πηγή.

SH: Ο αποστολέας έστειλε ένα SYN ακολουθούμενο από ένα αίτημα FIN, όμως ποτέ δεν εμφανίστηκε ένα SYN ACK στον παραλήπτη (με συνέπεια η σύνδεση να είναι ενεργή κατά το ήμισυ).

OTH: Κανένα αίτημα SYN δεν εμφανίστηκε, ωστόσο παρατηρήθηκε μία μερική ύπαρξη σύνδεσης η οποία αργότερα δεν έκλεισε.

logged_in: Αν ο χρήστης συνδεθεί επιτυχώς τότε εισάγεται η τιμή 1, διαφορετικά εισάγεται η τιμή 0

root_shell: Αν εισαχθεί ένα root shell τότε εμφανίζεται η τιμή 1, διαφορετικά εκτυπώνεται η τιμή 0.

serror_rate: Αριθμός συνδέσεων οι οποίες φέρουν λάθη που προήλθαν από αιτήματα SYN.

land: Χαρακτηριστικό που παρέχει πληροφορία σχετικά με το αν η σύνδεση προέρχεται από τον ίδιο host και θύρα

rpc_portmap_decode: Αποκωδικοποίηση ενός αιτήματος RPC

successful_recon_largescale: Μεγάλου βαθμού απώλεια πληροφορίας

successful_recon_limited: Απώλεια πληροφορίας

non_standard_protocol: Ανίχνευση ενός μη συμβατού πρωτοκόλλου ή συμβάντος

suspicious_login: Ανίχνευση προσπάθειας για σύνδεση μέσω ενός ύποπτου username

attempted_recon: Προσπάθεια για information leaking

bad_unknown: Πιθανή κακή κίνηση

dst_host_srv_count: Αριθμός συνδέσεων που έχουν τον ίδιο προορισμό και χρησιμοποιούν την ίδια υπηρεσία

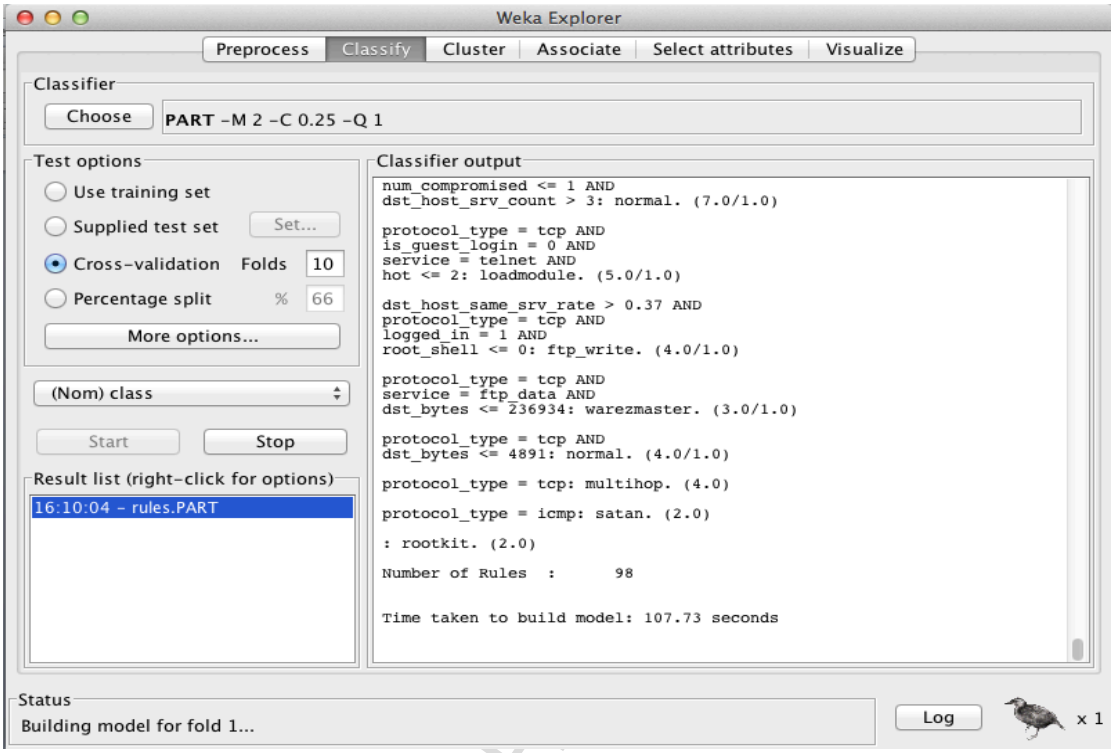
dst_host_diff_srv_rate: Ποσοστό διαφορετικών συνδέσεων που ανήκουν στον ίδιο host.

dst_host_srv_rerror_rate: Ποσοστό συνδέσεων που ανήκουν σε έναν τρέχων host και μία συγκεκριμένη υπηρεσία και φέρουν ένα RST error

dst_host_serror_rate: Ποσοστό συνδέσεων που ανήκουν σε ένα host και φέρουν ένα S0 error.

ΠΑΡΑΡΤΗΜΑ

KDDCUP99 Heuristic Rules παραγόμενα μέσω του Weka



Εικόνα 33: Ενδεικτική Χρήση Classifier για
Εξαγωγή Κανόνων από το KDDCUP99 Dataset

IF same_srv_rate <= 0.32 AND dst_host_diff_srv_rate <= 0.14 AND src_bytes <= 0 AND dst_host_same_src_port_rate <= 0.02 AND diff_srv_rate <= 0.58 THEN neptune

IF wrong_fragment <= 0 AND num_compromised > 0 AND src_bytes > 10073 THEN back

IF dst_host_srv_serror_rate > 0.82 AND flag = SH AND srv_count <= 80 THEN nmap

IF wrong_fragment > 0 AND protocol_type = udp THEN teardrop

IF srv_serror_rate > 0.51 AND dst_host_diff_srv_rate > 0.7 AND same_srv_rate <= 0.25 THEN satan

IF (dst_host_srv_serror_rate) > 0.82 AND (flag == 'SH') AND (srv_count) <= 80 THEN nmap

IF (srv_serror_rate > 0.51) AND (src_bytes <= 0) AND (land == 0) AND (dst_host_serror_rate > 0.68) AND (flag = S0) AND (dst_host_same_src_port_rate) <= 0.17 THEN neptune

IF (count > 327) AND (diff_srv_rate > 0.73) then satan

IF dst_host_srv_rerror_rate > 0.82 AND dst_host_count > 72 AND dst_host_same_src_port_rate > 0.01THEN portsweep

IF (dst_host_srv_diff_host_rate <= 0.24) AND (wrong_fragment <= 0) AND (src_bytes) > 6 AND (rerror_rate <= 0.08) AND (hot > 24) AND (hot <= 28) THEN warezclient

IF (dst_host_srv_diff_host_rate > 0.24) AND (wrong_fragment > 0) THEN pod

IF (dst_host_srv_diff_host_rate > 0.24) AND (src_bytes <= 20) AND (land == 0) AND (dst_host_rerror_rate <= 0.99) AND (dst_host_srv_diff_host_rate > 0.36) AND (dst_bytes <= 1) THEN ipsweep

IF src_bytes > 20 AND flag = RSTO AND num_failed_logins > 0 THEN guess_passwd

IF protocol_type = udp AND src_bytes <= 5 AND dst_host_count > 69 THEN satan

IF protocol_type = udp AND (service == ‘private’) THEN nmap

IF protocol_type = icmp AND src_bytes > 351 AND (service = ecr_i) THEN smurf

IF src_bytes > 22 AND srv_rerror_rate <= 0.08 AND dst_host_srv_diff_host_rate > 0.09 AND dst_host_same_srv_rate > 0.55 AND root_shell <= 0 AND logged_in = 1 THEN warezclient

IF same_srv_rate <= 0.46 AND diff_srv_rate > 0.88 AND srv_count <= 1 THEN satan

IF dst_host_srv_diff_host_rate > 0.23 AND dst_host_srv_serror_rate <= 0.1 AND srv_count > 2 AND protocol_type = icmp THEN nmap

IF src_bytes > 245 AND src_bytes > 12943 AND duration <= 1285 AND service = http THEN back

IF dst_host_srv_diff_host_rate > 0.23 AND dst_host_srv_serror_rate > 0.1 THEN land

IF dst_host_srv_diff_host_rate > 0.23 AND dst_bytes > 717 AND num_compromised <= 1 THEN loadmodule

IF dst_host_srv_diff_host_rate > 0.23 AND service = eco_i AND src_bytes > 13 AND src_bytes <= 24 THEN ipsweep
IF dst_host_srv_serror_rate <= 0.3 AND src_bytes <= 245 AND dst_host_diff_srv_rate > 0.95 AND urgent <= 0 AND src_bytes <= 35 AND dst_host_same_srv_rate > 0 THEN ipsweep.

IF dst_host_srv_serror_rate <= 0.3 AND root_shell <= 0 AND src_bytes <= 245 AND count <= 3 AND hot <= 0 AND dst_bytes > 251578 AND duration > 1 THEN warezmaster

IF srv_serror_rate <= 0.2 AND dst_host_rerror_rate > 0.89 AND service = private AND (flag == 'REJ') THEN neptune

IF srv_serror_rate <= 0.2 AND root_shell <= 0 AND logged_in = 1 AND dst_bytes <= 0 AND count <= 3 AND dst_host_same_srv_rate > 0.03 AND dst_host_srv_diff_host_rate <= 0.2 AND src_bytes > 305 AND src_bytes <= 1015 THEN warezclient

IF dst_host_srv_serror_rate > 0.25 AND dst_host_same_srv_rate <= 0.04 THEN portsweep

IF srv_serror_rate > 0.2 AND duration <= 30 AND land = 0 AND srv_rerror_rate <= 0.01 THEN neptune

IF root_shell > 0 AND num_shells > 0 AND num_file_creations <= 2 THEN perl

IF root_shell > 0 AND num_file_creations <= 2 AND dst_host_same_src_port_rate > 0.06 THEN buffer_overflow

IF srv_serror_rate > 0.27 AND duration <= 30 AND land = 0 THEN imap

IF (flag = OTH) THEN portsweep

IF (flag = S0) THEN land

IF duration > 1564 AND dst_bytes <= 2801 THEN warezclient

IF (flag == 'RSTR') AND num_failed_logins <= 0 AND duration <= 94 THEN portsweep

IF num_file_creations <= 0 AND dst_host_rerror_rate > 0.87 AND (service = telnet) THEN guess_passwd

IF protocol_type = icmp AND src_bytes <= 19 AND dst_host_srv_diff_host_rate <= 0.12 AND dst_host_count <= 3 THEN ipsweep

IF protocol_type = icmp AND src_bytes <= 19 AND src_bytes <= 13 THEN nmap

IF protocol_type = icmp AND dst_host_same_srv_rate > 0.22 AND src_bytes > 19 AND src_bytes > 300 THEN pod

IF num_access_files > 0 AND service = http THEN phf

IF logged_in = 1 AND dst_bytes <= 1 AND duration <= 6 AND src_bytes <= 2722 THEN warezclient

IF logged_in = 0 AND service = finger AND dst_bytes <= 85 THEN satan

IF protocol_type = icmp AND src_bytes <= 19 THEN ipsweep

IF srv_rerror_rate <= 0.5 AND (protocol_type = tcp) AND src_bytes > 1031 AND num_file_creations <= 0 AND (service = ftp_data) AND duration > 6 THEN warezclient

IF (protocol_type = tcp) AND (is_guest_login = 1) AND num_access_files <= 0 AND dst_host_rerror_rate <= 0.04 THEN multihop

IF rerror_rate > 0.25 THEN ipsweep

IF protocol_type = tcp AND is_guest_login = 1 AND num_access_files <= 0 THEN warezmaster

IF protocol_type = tcp AND dst_host_srv_serror_rate > 0.25 AND duration <= 179 THEN guess_passwd

IF protocol_type = tcp AND dst_host_same_srv_rate <= 0.22 AND num_failed_logins <= 2 THEN rootkit

IF protocol_type = tcp AND is_guest_login = 0 AND dst_host_diff_srv_rate <= 0.44 AND num_access_files <= 0 AND logged_in = 1 AND root_shell <= 0 AND dst_host_srv_diff_host_rate <= 0.03 THEN buffer_overflow

IF protocol_type = tcp AND is_guest_login = 0 AND service = telnet AND hot <= 2 THEN loadmodule

IF dst_host_same_srv_rate > 0.37 AND protocol_type = tcp AND logged_in = 1 AND root_shell <= 0 THEN ftp_write

IF protocol_type = tcp AND service = ftp_data AND dst_bytes <= 236934 THEN warezmaster