

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΜΕ ΘΕΜΑ

KDD99 INTRUSION DETECTION SYSTEM

ΤΟΥ

ΑΘΑΝΑΣΙΟΥ ΚΑΚΑΡΔΑΚΟΥ (Π2017055)

Επιβλέπων Καθηγητής : Νταντογιάν Χριστόφορος

Επιβλέπουσα Επιτροπή: Μάγκος Εμμανουήλ

Χριστοπούλου Ελένη

ΚΕΡΚΥΡΑ, ΙΟΥΝΙΟΣ 2023

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

KDD99 INTRUSION DETECTION SYSTEM

ΑΘΑΝΑΣΙΟΣ ΚΑΚΑΡΔΑΚΟΣ

(Π2017055)

Επιβλέπων Καθηγητής : Νταντογιάν Χριστόφορος

Επιβλέπουσα Επιτροπή : Μάγκος Εμμανουήλ

Χριστοπούλου Ελένη

Δηλώνω υπεύθυνα πως το παρόν κείμενο αποτελεί προϊόν προσωπικής μελέτης και εργασίας και πως όλες οι πηγές που χρησιμοποιήθηκαν για τη συγγραφή της δηλώνονται στις παραπομπές και στο βιβλιογραφικό κατάλογο.

Περιεχόμενα

A. ΚΕΦΑΛΑΙΟ

Περιεχόμενα	4
Ευρετήριο Εικόνων	5
Συντμήσεις	6
Γλωσσάρι	7
Περίληψη	8
Abstract	9
Ευχαριστίες	10
Εισαγωγή	11

B. ΚΕΦΑΛΑΙΟ

B.1 Γενική Ανάλυση KDD-99 dataset	13
B.1.1 Εισβολές	13
B.1.2 Σύστημα Ανίχνευσης Εισβολών IDS	15
B.1.3 Ανάλυση KDD-99 dataset	17
B.1.4 Χαρακτηριστικά KDD-99	19
B.1.5 Προβλήματα KDD-99	20
B.2 Έρευνα χρήσης	22
B.2.1. Internet Of Things	23
B.2.2. Cloud Attacks	25
B.2.3. DDoS	27
B.2.4. Smart Cities	29

B.3 Έρευνα χρήσης ανά έτος	31
Γ. Υπόλοιπα datasets	
Γ.1 UNSW-NB15	33
Γ.2 NSL-KDD	34
Γ.3 DARPA	36
Γ.4 HTTP DATASET CSIC 2010	37
Γ.5 ISCX 2012	39
Γ.6 CIC-IDS 2017	42
Γ.7 CSE-CIC-IDS	43
Γ.8 BOT-IoT	44
Δ. ΣΥΜΠΕΡΑΣΜΑΤΑ	45
Ε. ΒΙΒΛΙΟΓΡΑΦΙΑ	49
 ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ	
Εικόνα 1: Βασικές Αρχές Ασφάλειας Δικτύων	13
Εικόνα 2: Λειτουργία dataset UNSW-NB15	32
Εικόνα 3: Σχέση μεταξύ DARPA-98, KDD-99 & NSL-KDD	34
Εικόνα 4: Λειτουργία BOT- IoT	44
Πίνακας 1: Αποτελέσματα Έρευνας 2016-2021	31
Πίνακας 2: Παράδειγμα Χρήσης dataset ISCX 2012	41
Γράφημα 1: Papers per year	31
Γράφημα 2: IoT & Cloud	32
Γράφημα 3: DDoS & Smart Cities	32

ΣΥΝΤΜΗΣΕΙΣ

DDoS - Distributed Denial of Service

DoS - Denial of Service Attack: Επίθεση άρνησης εξυπηρέτησης

IDPS- Intrusion Detection and Prevention Systems: Συστήματα Ανίχνευσης και Πρόληψης

IDS - Intrusion Detection Systems: Συστήματα Ανίχνευσης Εισβολών

KDD- Knowledge Discovery in Databases

R2L - Remote to Local: Απομακρυσμένη σε τοπική επίθεση

U2R - User to Root Attack

IoT - Internet of Things

ΓΛΩΣΣΑΡΙ

Anomaly models - Μοντέλα ανίχνευσης διαταραχών

Availability – Διαθεσιμότητα

Confidentiality - Εμπιστευτικότητα

Containment – Περιορισμός

Dataset - Σύνολο Δεδομένων

Eradication – Εξουδετέρωση

Firewalls - Τείχη Προστασίας

Follow-up - Συνεχής παρακολούθηση

Identification- Ταυτοποίηση

Integrity - Ακεραιότητα

Intrusion - Εισβολή

Machine Learning - Μηχανική Μάθηση

Misuse models - Μοντέλα κακής συμπεριφοράς

Number of failed login attempts - Αριθμό των αποτυχημένων προσπαθειών σύνδεσης

Preparation – Προετοιμασία

Recovery – Αποκατάσταση

Requests - Απαιτήσεις

Specification-based models - Μοντέλα που βασίζονται στις προδιαγραφές

Time-based - Βασισμένες στο χρόνο

ΠΕΡΙΛΗΨΗ

Η αναγκαιότητα προστασίας των πληροφορικών δικτύων είναι επιβεβλημένη στην σημερινή εποχή, καθώς η ραγδαία αύξηση της χρήσης των δικτύων υπολογιστών έχει επιφέρει και την συνακόλουθη αύξηση του κινδύνου εισβολής μη εξουσιοδοτημένων ατόμων σε αυτά. Το KDD-99 αποτελεί ένα Σύστημα Ανίχνευσης Εισβολών, το οποίο χρονολογείται από το 1999, ωστόσο καλύπτει λίγους τύπους επιθέσεων και έχει μεγάλο αριθμό περιττών και διπλών δειγμάτων δεδομένων. Παρόλα αυτά, με την πάροδο των ετών χρησιμοποιείται ολοένα και περισσότερο σε επιθέσεις σε Cloud, Smart Cities, Internet of Things και σε επιθέσεις Dos-DDos. Ένα εξίσου παλιό Σύστημα Ανίχνευσης Εισβολών είναι και το DAPRA-98, το οποίο όμως πολλές φορές οδηγεί σε παρατυπίες. Λόγω των σφαλμάτων που εμπεριέχονταν στα δύο αυτά datasets, δημιουργήθηκε το NSL-KDD. Είναι μια πιο εξελιγμένη μορφή του KDD-99, η οποία μολονότι αποτελεί καλύτερη εκδοχή των δύο προηγούμενων, απευθύνεται και αυτή σε περιορισμένες επιθέσεις. Στη συνέχεια εξετάζονται τα πιο σύγχρονα datasets όπως τα ISCX2012, UNSW-NB15, CIC-IDS2017 και CSE-CIC-IDS, τα οποία απευθύνονται σε περισσότερα είδη επιθέσεων, δεν έχουν τόσα διπλότυπα αρχεία όσο τα παλαιότερα datasets και δεν υπάρχουν ή υπάρχουν ελάχιστα αβέβαια στοιχεία. Ως εκ τούτου κατόπιν αναλυτικής απεικόνισης τόσο των παλαιότερων όσο και των πιο σύγχρονων datasets, συμπεραίνουμε ότι τα τελευταία είναι περισσότερο αποτελεσματικά.

ABSTRACT

The need to protect computer networks is imperative in this day and age, as the rapid increase in the use of computer networks has brought about the consequent increase in the risk of intrusion of unauthorized persons into them. KDD-99 is an Intrusion Detection System dating back to 1999, but covers few types of attacks and has a large number of odd and duplicate sample data. However, over the years it has been increasingly used in Cloud, Smart Cities, Internet of Things attacks, and Dos-DDos attacks. An equally old Intrusion Detection System is the DAPRA-98, which often leads to irregularities. Due to the errors contained in these two datasets, the NSL-KDD was created. It is an advanced form of KDD-99, but also addresses limited attacks. Modern datasets such as ISCX2012, UNSW-NB15, CIC-IDS2017 and CSE-CIC-IDS are then examined, which target more types of attacks, do not have as many duplicate files as older datasets, and there are no or few uncertain data. Therefore, after analyzing both older and modern datasets, we conclude that modern ones are more effective.

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα πτυχιακή εργασία αποτελεί το επισφράγισμα της περάτωσης των σπουδών μου στο Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου και αντανακλά τα ενδιαφέροντα μου στον τομέα αυτό. Στο σημείο αυτό θα ήθελα να ευχαριστήσω την οικογένειά και τους ανθρώπους μου για την υποστήριξη τους όλα αυτά τα έτη φοίτησής μου στο Ιόνιο Πανεπιστήμιο.

Ένα μεγάλο ευχαριστώ σε όλους τους καθηγητές και καθηγήτριες του τμήματος που ο καθένας με τον δικό του τρόπο μου έδωσε κίνητρο να εξελίσσομαι και να μαθαίνω κάθε μέρα κάτι διαφορετικό. Θερμές ευχαριστίες θα ήθελα να αποδώσω στον Επιβλέπων Καθηγητή κ. Νταντογιάν Χριστόφορο για την συνεργασία και την κατανόησή του αλλά κυρίως για το έναυσμα που μου έδωσε προκειμένου να γνωρίσω τον τομέα της Ασφάλειας Δικτύων. Τέλος, θα ήθελα να ευχαριστήσω την κα. Χριστοπούλου και τον κ. Μάγκο που με προθυμία δέχθηκαν να συμμετάσχουν στην αξιολόγηση της παρούσας εργασίας.

ΕΙΣΑΓΩΓΗ

Η ραγδαία αύξηση της χρήσης δικτύων υπολογιστών καθώς και η συνακόλουθη αύξηση του αριθμού των εφαρμογών που εκτελούνται σε αυτό καθιστούν την ασφάλεια του δικτύου απαραίτητη. Ως εκ τούτου, ο ρόλος των Συστημάτων Ανίχνευσης Εισβολών (Intrusion Detection Systems - IDS), συσκευών ειδικού σκοπού για την ανίχνευση ανωμαλιών και επιθέσεων στο δίκτυο αποκτά ακόμη μεγαλύτερη σημασία, καθώς προσδιορίζει εάν κάποιες ενέργειες αποτελούν εισβολές με βάση ένα ή περισσότερα μοντέλα εισβολών. Τα μοντέλα αυτά διαχωρίζουν τις ενέργειες ως «καλές» (δεν υπάρχει εισβολή) ή «κακές» (υπάρχουν πιθανές εισβολές) και διακρίνονται σε μοντέλα ανίχνευσης διαταραχών (anomaly models), μοντέλα κακής συμπεριφοράς (misuse models) και σε μοντέλα που βασίζονται στις προδιαγραφές (specification-based models), ενώ μπορεί να είναι είτε προσαρμοστικά (adaptive), δηλαδή να προσαρμόζουν την συμπεριφορά τους με βάση τις καταστάσεις-ενέργειες των συστημάτων, είτε στατικά (static), να μην τροποποιούνται δηλαδή κατά τη διάρκεια εκτέλεσης του συστήματος.

Αντικείμενο της παρούσας μελέτης αποτελεί το dataset KDD-99, το οποίο χρονολογείται από το 1999 και χρησιμοποιείται κυρίως στον τομέα της ακαδημαϊκής έρευνας, όπως στη Μηχανική Μάθηση (Machine Learning) και στα Συστήματα Ανίχνευσης Εισβολών (Intrusion Detection Systems – IDS). Χρησιμοποιήθηκε στον 3^ο Διεθνή Διαγωνισμό Εργαλείων Ανακάλυψης Γνώσης και Εξόρυξης Δεδομένων με σκοπό να κατασκευάσει ένα ανιχνευτή εισβολής δικτύου, ένα μοντέλο με την ικανότητα να εντοπίζει και να διακρίνει τις «κακές» από τις «καλές» συνδέσεις. Η αρχική του χρήση έγινε σε ένα περιβάλλον στρατιωτικού δικτύου και περιλάμβανε μια ποικιλία παρεμβάσεων που προσομοιώνονται σε αυτό. Από το 1999 το KDD-99 είναι το dataset που χρησιμοποιείται κατ' εξοχήν για την αξιολόγηση των τεχνικών ανίχνευσης ανωμαλιών. Για την έρευνα χρήσης του συγκεκριμένου dataset, αναλύθηκαν 1949 papers που δημοσιεύτηκαν στο διάστημα Ιανουαρίου 2016 και Δεκεμβρίου 2021. Το αποτέλεσμα της έρευνας έδειξε ότι το εν λόγω dataset ειδικεύεται περισσότερο στον τομέα του Machine Learning και του IDS.

Κατόπιν εισαγωγής στο θέμα των εισβολών και των Συστημάτων Ανίχνευσης αυτών, θα ακολουθήσουν στατιστικά στοιχεία από τα papers που μελετήθηκαν, καθώς και ανάλυση αυτών, προκειμένου να παρουσιαστεί η χρήση του KDD-99. Στη

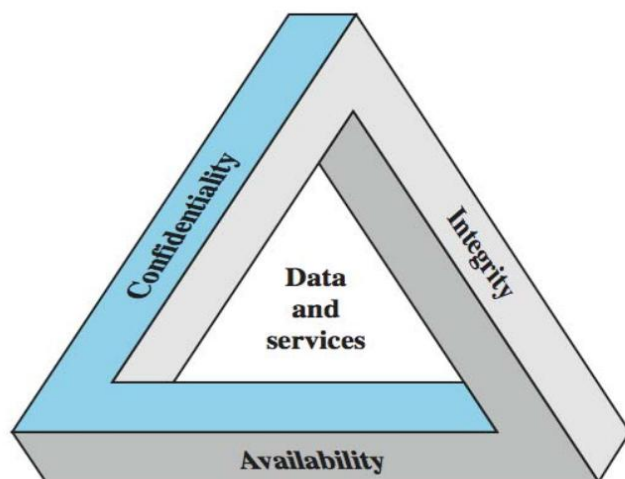
συνέχεια, θα μελετηθούν και άλλα datasets, είτε παλαιότερα είτε πιο σύγχρονα, με σκοπό να αναδειχθεί η αποτελεσματικότητα του KDD-99 σε σύγκριση με τα υπόλοιπα.

Β' ΚΕΦΑΛΑΙΟ

Β.1. ΓΕΝΙΚΗ ΑΝΑΛΥΣΗ KDD-99

Β.1.1. ΕΙΣΒΟΛΕΣ

Οι εισβολές (intrusion) σε ένα δίκτυο είναι ο πιο συνηθής τρόπος επίθεσης. Μπορούν να οριστούν ως το σύνολο των ενεργειών που προσπαθούν να θέσουν σε κίνδυνο την ασφάλεια των πληροφορικών συστημάτων που στηρίζονται στην εμπιστευτικότητα, την ακεραιότητα και την διαθεσιμότητα του ηλεκτρονικού υπολογιστή. Συνοπτικά η εμπιστευτικότητα διασφαλίζει την μη αποκάλυψη ευαίσθητων πληροφοριών σε μη εξουσιοδοτημένα άτομα, η ακεραιότητα αναφέρεται στην διατήρηση των δεδομένων και τέλος σε ένα πληροφορικό σύστημα πρέπει να είναι εξασφαλισμένη η συνεχόμενη και αδιάλειπτη διαθεσιμότητα των πόρων και των υπηρεσιών. Χρησιμοποιούν σκοπίμως αυτή τη μη εξουσιοδοτημένη πρόσβαση των πόρων, προσπαθώντας να αποκτήσουν τα δικαιώματα των νόμιμων χρηστών, όπως είναι η πρόσβαση σε πληροφορίες και στη διαχείριση δεδομένων, καθώς και η ικανότητα να καθιστούν ένα σύστημα σε αναξιόπιστο ή μη χρησιμοποιούμενο.



Εικόνα 1: Βασικές Αρχές Ασφάλειας Δικτύων, όπου Confidentiality η Εμπιστευτικότητα, Integrity η Ακεραιότητα και Availability η Διαθεσιμότητα.

Για την προστασία του δικτύου από τέτοιες επιθέσεις χρησιμοποιούνται¹:

- Τείχη Προστασίας (Firewalls)
- Συστήματα Ανίχνευσης Εισβολών (Intrusion Detection Systems - IDS)
- Συστήματα Ανίχνευσης και Πρόληψης Εισβολής (Intrusion Detection and Prevention Systems – IDPS)

• ¹ Χρήστος Μπούρας (2004). Δίκτυα Δημόσιας Χρήσης και Διασύνδεση Δικτύων. Πανεπιστημιακές Σημειώσεις.

B.1.2. ΣΥΣΤΗΜΑΤΑ ΑΝΙΧΝΕΥΣΗΣ ΕΙΣΒΟΛΩΝ

Intrusion Detection Systems - IDS

Στην παρούσα μελέτη θα επικεντρωθούμε στα Συστήματα Ανίχνευσης Εισβολών (IDS) και κυρίως στο dataset KDD-99. Τα IDS αποτελούν ουσιαστικά μηχανές εποπτείας και ελέγχου. Σκοπός τους είναι να ανιχνεύουν επιβλαβείς ή κακόβουλες προσπάθειες εισβολής μη εξουσιοδοτημένων ατόμων στα συστήματα, γεγονός που θα έθετε σε κίνδυνο παραβίασης κάποια από τις τρεις Βασικές Αρχές Ασφάλειας Δικτύων. Σε περίπτωση ανίχνευσης κάποιας αμφίβολης δραστηριότητας, τα IDS χρειάζεται να ενημερώνουν τον διαχειριστή.

Σύμφωνα με τον Γκρίτζαλι και τον Κάτσικα² υπάρχουν τρία μοντέλα ανίχνευσης εισβολών:

A. Μοντέλα Ανίχνευσης Διαταραχών (Anomaly Models) : Αναλύουν ένα σύνολο χαρακτηριστικών του συστήματος και συγκρίνουν τη συμπεριφορά τους με ένα σύνολο αναμενόμενων τιμών. Παρουσιάζουν κατά πόσο οι τιμές που υπολογίστηκαν ταιριάζουν με τις αναμενόμενες μετρήσεις και χαρακτηρίζουν τις ενέργειες που είναι στατιστικά ασυνήθιστες ως «κακές».

B. Μοντέλα Κακής Συμπεριφοράς (Misuse Models) : Συγκρίνουν ενέργειες με ακολουθίες που είναι ήδη γνωστό ότι αποτελούν εισβολές ή με ακολουθίες που θεωρείται ότι μπορούν να αποτελούν εισβολές και προσδιορίζουν αν οι ενέργειες αυτές παραβιάζουν την πολιτική ασφάλειας των περιοχών στις οποίες εκτελούνται. Είναι ιδιαίτερα χρήσιμο για γνωστά μοτίβα επιθέσεων, ωστόσο αυτό ακριβώς είναι και το μειονέκτημά του, δηλαδή μπορεί να παρατηρήσει μόνο τις εισβολές που εμπεριέχουν αυτά. Απαιτείται γνώση όλων των ευπαθειών των συστημάτων αλλά και των δυνητικών ευπαθειών που οι επιτιθέμενοι προσπαθούν να εκμεταλλευτούν.

Γ. Μοντέλα που βασίζονται σε προδιαγραφές (Specification-based Models) : Καθορίζουν εάν μια ακολουθία οδηγιών παραβιάζει ή όχι μια προδιαγραφή

² Σ. Γκρίτζαλις, Σ. Κάτσικας, & Δ. Γκρίτζαλις (2003). *Ασφάλεια Δικτύων Υπολογιστών*. Αθήνα: Εκδόσεις Παπασωτηρίου

αναφορικά με τον τρόπο με τον οποίο πρέπει να εκτελείται ένα σύστημα. Στην περίπτωση αυτή αναφέρεται κάποια ενδεχόμενη εισβολή.

Κατόπιν ανίχνευσης και αναγνώρισης της εισβολής απαιτείται και ο χειρισμός αυτής, ο οποίος περιλαμβάνει έξι φάσεις³. Συνοπτικά:

- Προετοιμασία για μία εισβολή (Preparation) : Στο στάδιο αυτό επιλέγονται και εγκαθίστανται οι διαδικασίες και οι μηχανισμοί για την ανίχνευση και την απόκριση στις εισβολές.
- Ταυτοποίηση της εισβολής (Identification) : Το στάδιο αυτό διαμορφώνει τις υπόλοιπες φάσεις.
- Περιορισμός της εισβολής (Containment) : Γίνεται προσπάθεια να περιοριστεί ο εισβολέας και κατά συνέπεια η ζημιά στο σύστημα.
- Εξουδετέρωση της εισβολής (Eradication) : Σε αυτή τη φάση διακόπτεται η εισβολή και παρεμποδίζονται περαιτέρω παρόμοιες επιθέσεις.
- Αποκατάσταση από την εισβολή (Recovery) : Αποκαθίσταται η ασφαλής κατάσταση στο σύστημα, σύμφωνα με την ισχύουσα πολιτική ασφάλειας.
- Συνεχής παρακολούθηση της εισβολής (Follow-up) : Λαμβάνονται μέτρα κατά του εισβολέα, προσδιορίζονται τα προβλήματα κατά τον χειρισμό του γεγονότος και καταγράφονται οι σχετικές εμπειρίες που αποκτήθηκαν.

B.1.3. ΓΕΝΙΚΗ ΑΝΑΛΥΣΗ KDD-99 DATASET

³ S. Northcutt (1998) *Computer Security Incident Handling: Step by Step*, The SANS Institute, Bathesda

Όπως έχει ήδη αναφερθεί, η βάση δεδομένων KDD-99 (Knowledge Discovery in Databases- KDD) αναφέρεται σε μία ευρεία διαδικασία εύρεσης δεδομένων, εστιάζοντας σε συγκεκριμένες μεθόδους εξόρυξης δεδομένων. Η χρήση του στον τομέα της έρευνας είναι διαδεδομένη και κυρίως στους τομείς της μηχανικής εκμάθησης, της αναγνώρισης προτύπων, της στατιστικής, της τεχνητής νοημοσύνης και της οπτικοποίησης δεδομένων.

Το KDD-99 αξιολογεί και ενδεχομένως ερμηνεύει τα πρότυπα προκειμένου να προσδιορίσει τι χαρακτηρίζεται ως γνώση. Περιλαμβάνει επίσης την επιλογή συστημάτων κωδικοποίησης, προ-επεξεργασίας, δειγματοληψίας και προβολών των δεδομένων πριν από το στάδιο εξόρυξης δεδομένων.

Το dataset αυτό έχει σχεδιαστεί από τους Stolfo et al. (Stolfo, 2000) και έχει βασιστεί σε δεδομένα που συλλέχθηκαν από ένα άλλο IDS, το DARPA-98. Προκειμένου να κατασκευαστεί, έγινε μια προσομοίωση ενός πραγματικού στρατιωτικού δικτύου που αποτελείται από τρεις μηχανές «στόχους» που εκτελούσαν διάφορα λειτουργικά συστήματα και υπηρεσίες. Επιπλέον τρία μηχανήματα χρησιμοποιήθηκαν στη συνέχεια για την πλαστογράφηση διαφορετικών διευθύνσεων IP για τη δημιουργία επισκεψιμότητας. Τέλος, υπήρξε ένα sniffer που κατέγραφε όλη την κυκλοφορία δικτύου χρησιμοποιώντας τη μορφή ένδειξης σφαλμάτων TCP. Η συνολική προσομοιωμένη περίοδος ήταν επτά εβδομάδες. Περιλαμβάνει επίσης, εγγραφές με 41 χαρακτηριστικά, ενώ στη κάθε μία επισημαίνεται αν είναι κανονική (normal) ή επίθεση (attack).

Οι επιθέσεις αυτές διακρίνονται στις ακόλουθες κατηγορίες⁴:

A. Επίθεση άρνησης εξυπηρέτησης (Denial of Service Attack - DoS): Η επίθεση αυτή έχει ως στόχο να πλήξει την διαθεσιμότητα του υπολογιστικού πόρου. Γενικότερα τα IDS παρουσιάζουν ευαισθησία σε DoS επιθέσεις. Οι DoS επιθέσεις μπορούν να προκαλέσουν τόσο την ανικανότητα των συστημάτων να παρέχουν τις υπηρεσίες τους όσο και να οδηγήσουν στην κατάρρευση αυτών.

⁴ A. Alharbi, S. Alhaidari & M. Zohdy (2018). Denial-of-service, probing, user to root (U2R) & remote to user (R2L) attack detection using hidden Markov models. *International Journal of Computer and Information Technology*, 7(05)

B. Επίθεση Ανίχνευσης ή Ανίχνευσης-Απόκρισης (Probe or Probe-Response):

Αποσκοπούν στην αποκάλυψη πληροφοριών σχετικά με τα τρωτά σημεία ενός συστήματος προορισμού, εκτελώντας προσεκτικά επιλεγμένες ακολουθίες λειτουργιών και παρατηρώντας τον τρόπο με τον οποίο το σύστημα ή το IDS συμπεριφέρεται ως απόκριση. Ορισμένες από αυτές τις επιθέσεις προσπαθούν να «χαρτογραφήσουν» το IP του στόχου προκειμένου να διαδώσουν κάποιο κακόβουλο λογισμικό.

Γ. User to Root Attack (U2R): Ο εισβολέας ξεκινά με πρόσβαση σε έναν κανονικό λογαριασμό χρήστη στο σύστημα (ίσως αποκτάται από την ανίχνευση κωδικών πρόσβασης, μια επίθεση λεξικού ή κοινωνική μηχανική) και επεκτείνοντας σταδιακά τα προνόμια του, κατορθώνει εκμεταλλευόμενος κάποια ευπάθεια να αποκτήσει δικαιώματα και άδειες ενός διαχειριστή.

Δ. Απομακρυσμένη σε τοπική επίθεση (Remote to Local - R2L):

Εννοιολογικά είναι παρόμοια με την U2R αλλά είναι σε πιο ήπιο βαθμό. Μια τέτοια επίθεση πραγματοποιείται όταν ένας εισβολέας στέλνε πακέτα στον κεντρικό υπολογιστή προορισμού που προορίζονται να αποκαλύψουν ευπάθειες που θα επέτρεπαν στον εισβολέα να εκμεταλλευτεί τα δικαιώματα ενός τοπικού χρήστη.

B.1.4. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ KDD-99

Τα χαρακτηριστικά του KDD-99 διακρίνονται σε τρεις κατηγορίες:

A. Βασικά χαρακτηριστικά: η κατηγορία αυτή ενσωματώνει όλα τα χαρακτηριστικά που μπορεί να εξαχθούν από το πρωτόκολλο TCP / IP σύνδεσης. Επιπλέον, τα περισσότερα από αυτά οδηγούν σε μια έμμεση καθυστέρηση στην ανίχνευση.

B. Χαρακτηριστικά κίνησης: η κατηγορία αυτή περιλαμβάνει χαρακτηριστικά που υπολογίζονται σε σχέση με ένα χρονικό πλαίσιο και χωρίζεται σε δύο ομάδες:

B.1. Χαρακτηριστικά "ίδιου ξενιστή": Εξετάζει μόνο τις συνδέσεις των τελευταίων 2 δευτερολέπτων στις οποίες υπάρχει ο ξενιστής προορισμού με την τρέχουσα σύνδεση.

B.2. Χαρακτηριστικά "ίδιας υπηρεσίας": Εξετάζει μόνο τις συνδέσεις των τελευταίων 2 δευτερολέπτων που έχουν την ίδια υπηρεσία με τη τρέχουσα σύνδεση.

Οι δύο παραπάνω επιθέσεις ονομάζονται time-based (βασισμένες στο χρόνο).

Γ. Χαρακτηριστικά περιεχομένου: Σε αντίθεση με τις DoS και Probe επιθέσεις, οι επιθέσεις R2L και U2R δεν έχουν κανένα πρότυπο συχνής διαδοχικής εισβολής. Αυτό γίνεται επειδή οι DoS και Probe επιθέσεις περιλαμβάνουν πολλές συνδέσεις σε ορισμένους ξενιστές σε πολύ σύντομο χρονικό διάστημα, ωστόσο οι R2L και U2R επιθέσεις ενσωματώνονται στα τμήματα δεδομένων των πακέτων και συνήθως περιλαμβάνουν μόνο μία σύνδεση. Για την ανίχνευση τέτοιου είδους επιθέσεων χρειάζονται επιπλέον χαρακτηριστικά ώστε να μπορούν να εξετάσουν για ύποπτες συμπεριφορές στα τμήματα δεδομένων, π.χ αριθμό των αποτυχημένων προσπαθειών σύνδεσης (number of failed login attempts). Αυτά τα χαρακτηριστικά ονομάζονται χαρακτηριστικά περιεχομένου

B.1.5. ΠΡΟΒΛΗΜΑΤΑ KDD-99

Όπως αναφέρθηκε σε προηγούμενη ενότητα το KDD-99 είναι βασισμένο σε δεδομένα που συλλέγονται από το DARPA-98, το οποίο έχει επικριθεί από τον McHugh⁵, κυρίως λόγω των χαρακτηριστικών των συνθετικών δεδομένων. Ως εκ τούτου, προβλήματα που υπάρχουν στο DARPA-98 παραμένουν και στο KDD-99. Συγκεκριμένα:

- Χάριν προστασίας της ιδιωτικότητας, τα πειράματα περιλαμβάνουν τόσο βασικά όσο και επιτιθέμενα δεδομένα και παρατηρείται ότι τα δεδομένα είναι παρόμοια με αυτά που έχουν ήδη παρατηρηθεί κατά τη διάρκεια αρκετών μηνών δειγματοληψίας από διάφορες βάσεις της Πολεμικής Αεροπορίας. Ωστόσο, δεν πραγματοποιήθηκε ούτε αναλυτική ούτε πειραματική επικύρωση των «ψευδών» αυτών χαρακτηριστικών. Επιπλέον, το φορτίο των περιλαμβανόμενων δεδομένων δεν είναι παρόμοιο με την κίνηση σε πραγματικά δίκτυα.
- Συλλέκτες κίνησης όπως το TCPdump, το οποίο χρησιμοποιείται στο DARPA-98 είναι πολύ πιθανό να υπερφορτωθούν και να αφήσουν πακέτα με μεγάλο φορτίο κίνησης. Ωστόσο, δεν υπήρξε εξέταση προκειμένου να ελεγχθεί η πιθανότητα αυτών των πακέτων.
- Δεν υπάρχει σαφής ορισμός των επιθέσεων. Για παράδειγμα, η επίθεση Probe δεν αποτελεί επίθεση εκτός εάν ο αριθμός των επαναλήψεων υπερβεί ένα συγκεκριμένο όριο. Για τον λόγο αυτό, θα πρέπει να υπάρχει μία συμφωνία σχετικά με το ορισμό του αξιολογητή και του αξιολογούμενου. Στο DARPA-98 δεν παρατηρείται κάποιος ορισμός για τις επιθέσεις δικτύου.

Βέβαια παρατηρούνται τόσο κάποιες βελτιώσεις όσο και πρόσθετα προβλήματα. Οι υπάρχουσες ελλείψεις στο KDD-99 κάνουν τα αποτελέσματα της αξιολόγησης σε αυτό το dataset αναξιόπιστα:

A. Περιττές Εγγραφές : Μία από τις πιο σημαντικές ελλείψεις στο dataset KDD-99 είναι ο τεράστιος αριθμός περιττών εγγραφών, που καθιστά τους αλγόριθμους μάθησης προκατειλημμένους προς τις συχνές εγγραφές και έτσι εμποδίζεται η

⁵ J. McHugh (2000). Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. ACM Transactions on Information and System Security (TISSEC), 3(4), 262-294.

εκμάθηση μη συχνών εγγραφών που είναι συνήθως πιο επιβλαβείς για δίκτυα όπως επιθέσεις R2L και U2R. Επιπρόσθετα, η ύπαρξη αυτών των επαναλαμβανόμενων εγγραφών στο σύνολο δοκιμών προκαλεί μεροληψία των αποτελεσμάτων αξιολόγησης από τις μεθόδους που έχουν καλύτερα ποσοστά ανίχνευσης στις συχνές εγγραφές.

B. Επίπεδο Δυσκολίας : Ακολουθώντας την τυπική προσέγγιση για την εκτέλεση ανίχνευσης ανωμαλιών χρησιμοποιώντας το KDD-99, γίνεται χρήση ενός προσαρμοσμένου αλγόριθμου μηχανικής εκμάθησης για να διερευνηθεί η γενική συμπεριφορά του dataset προκειμένου να καταστεί δυνατή η διάκριση μεταξύ κανονικών και κακόβουλων δραστηριοτήτων. Για τον λόγο αυτό, το dataset χωρίζεται σε τμήματα δοκιμής και εκπαίδευσης, όπου ο «εκπαιδευόμενος» αρχικά χρησιμοποιεί το τμήμα εκπαίδευσης του dataset προκειμένου να εκπαιδευτεί και στη συνέχεια αξιολογείται για την αποτελεσματικότητά του στο τμήμα δοκιμής. Πολλοί ερευνητές στον τομέα της μηχανικής μάθησης προσπάθησαν να επινοήσουν σύνθετους «εκπαιδευόμενους» για να βελτιστοποιήσουν την ακρίβεια και τον ρυθμό ανίχνευσης μέσω του dataset KDD-99. Μέσω πειραμάτων που διεξήχθησαν⁶, κατέληξαν στο συμπέρασμα ότι το dataset KDD-99 δεν αποτελεί κατάλληλη επιλογή για την αξιολόγηση με βάση την ακρίβεια, το ποσοστό ανίχνευσης και το ποσοστό ψευδών θετικών αποτελεσμάτων.

⁶ M. Tavallae et al. (2009). A detailed analysis of the KDD CUP 99 data set. In 2009 IEEE symposium on computational intelligence for security and defense applications (pp. 1-6). Ieee.

B.2 ΕΡΕΥΝΑ ΧΡΗΣΗΣ KDD-99 DATASET

Για την έρευνα χρήσης του dataset KDD-99 χρησιμοποιήθηκαν 1949 papers, τα οποία έχουν δημοσιευτεί από την 1^η Ιανουαρίου 2016 έως τις 31 Δεκεμβρίου 2021. Πρωταρχικό ζητούμενο της έρευνας ήταν να βρεθεί η συχνότητα χρήσης του dataset ως προς όλους τους στόχους. Στη συνέχεια, επιδιώκεται η σύγκριση αύξησης ή μείωσης των συχνοτήτων με την πάροδο του κάθε έτους στις εξής κατηγορίες:

- Internet of Things (IoT) : Αντιπροσωπεύει τη σύγκλιση μιας ποικιλίας τάσεων υπολογιστικής και συνδεσιμότητας που εξελίσσονται εδώ και πολλές δεκαετίες.
- Cloud Attacks: Οι πόροι στο cloud είναι το πιο ελκυστικό έδαφος για τους κυβερνο-εγκληματίες λόγω των τεράστιων πόρων που είναι διαθέσιμοι, πχ. Τράπεζες, πρεσβείες κλπ.
- Distributed Denial of Service (DDoS) : Οι επιθέσεις προσθέτουν τη διάσταση many-to-one στο πρόβλημα DoS, καθιστώντας την πρόληψη και τον μετριασμό αυτών των επιθέσεων πιο δύσκολη και τον αντίκτυπο αναλογικά σοβαρό. Το DDoS εκμεταλλεύεται την εγγενή αδυναμία της αρχιτεκτονικής του συστήματος του Διαδικτύου, το μοντέλο ανοικτής πρόσβασης σε πόρους, το οποίο τυχαίνει να είναι το μεγαλύτερο πλεονέκτημά του.
- Smart Cities: έχει πρόσφατα εισαχθεί ως στρατηγικός μηχανισμός για την ενσωμάτωση σύγχρονων παραγόντων αστικής παραγωγής σε ένα κοινό πλαίσιο και, ειδικότερα, για την επισήμανση της σημασίας των τεχνολογιών των πληροφοριών και των επικοινωνιών.

Ακολουθεί υποενότητα, στην οποία δίνονται περισσότερες πληροφορίες για το κάθε ένα από αυτά που αναφέρθηκαν παραπάνω.

B.2.1 Internet of Things (IoT)

Το Internet of Things (IoT) είναι ένας όρος που χρησιμοποιείται για να περιγράψει το δίκτυο φυσικών συσκευών, οχημάτων, οικιακών συσκευών και άλλων αντικειμένων που είναι ενσωματωμένα με ηλεκτρονικά, λογισμικό, αισθητήρες και συνδεσιμότητα δικτύου. Το IoT αυξάνεται με εκθετικό ρυθμό τα τελευταία χρόνια, και αυτή η ανάπτυξη αναμένεται να συνεχιστεί. Το IoT έχει τη δυνατότητα να φέρει επανάσταση στον τρόπο που ζούμε και εργαζόμαστε, συνδέοντας συσκευές και συστήματα που ήταν προηγουμένως απομονωμένα.

Το dataset KDD99 έχει χρησιμοποιηθεί ευρέως στον τομέα IoT για την αξιολόγηση της αποτελεσματικότητας των συστημάτων ανίχνευσης εισβολής (IDS) για δίκτυα IoT. Στον τομέα IoT, τα IDS είναι ιδιαίτερα σημαντικά επειδή οι συσκευές IoT είναι συχνά ευάλωτες σε επιθέσεις λόγω των περιορισμένων πόρων τους και της έλλειψης χαρακτηριστικών ασφαλείας.

Το dataset KDD99 περιέχει ένα σύνολο δυνατοτήτων που περιγράφουν χαρακτηριστικά κίνησης δικτύου, όπως μέγεθος πακέτου, τύπο πρωτοκόλλου και αριθμό θύρας. Αυτές οι δυνατότητες μπορούν να χρησιμοποιηθούν για την εκπαίδευση μοντέλων μηχανικής μάθησης για την ανίχνευση εισβολών δικτύου. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν στο dataset KDD99 για να μάθουν μοτίβα κανονικής κυκλοφορίας δικτύου και, στη συνέχεια, να χρησιμοποιήσουν αυτές τις γνώσεις για να ανιχνεύσουν ανωμαλίες στην κυκλοφορία δικτύου σε πραγματικό χρόνο.

Μια προσέγγιση για τη χρήση του dataset KDD99 για ανίχνευση εισβολής IoT είναι η χρήση μη εποπτευόμενων αλγορίθμων μάθησης, όπως η ομαδοποίηση ή η ανίχνευση ανωμαλιών. Οι αλγόριθμοι μη εποπτευόμενης μάθησης δεν απαιτούν επισημασμένα δεδομένα και μπορούν να χρησιμοποιηθούν για τον εντοπισμό μοτίβων στην κυκλοφορία δικτύου που είναι ενδεικτικά εισβολής. Αυτή η προσέγγιση μπορεί να είναι αποτελεσματική στην ανίχνευση προηγουμένως αόρατων επιθέσεων, αλλά μπορεί επίσης να δημιουργήσει ψευδείς συναγερμούς λόγω του μεγάλου αριθμού καλοήθων συμβάντων δικτύου.

Μια άλλη προσέγγιση για τη χρήση του dataset KDD99 για ανίχνευση εισβολής IoT είναι η χρήση αλγορίθμων εποπτευόμενης μάθησης, όπως δέντρα αποφάσεων ή

διανυσματικές μηχανές υποστήριξης. Οι αλγόριθμοι εποπτευόμενης μάθησης απαιτούν επισημασμένα δεδομένα και μπορούν να χρησιμοποιηθούν για την εκμάθηση μοτίβων κανονικής και μη φυσιολογικής κίνησης δικτύου. Αυτή η προσέγγιση μπορεί να είναι αποτελεσματική στον εντοπισμό γνωστών επιθέσεων, αλλά μπορεί να είναι λιγότερο αποτελεσματική στον εντοπισμό επιθέσεων που δεν είχαν παρατηρηθεί προηγουμένως.

Εν κατακλείδι, στον τομέα IoT, το dataset KDD99 μπορεί να χρησιμοποιηθεί για την εκπαίδευση αλγορίθμων μηχανικής μάθησης για την ανίχνευση μη φυσιολογικής συμπεριφοράς σε δίκτυα IoT. Η επιλογή του αλγορίθμου μηχανικής μάθησης εξαρτάται από τη συγκεκριμένη περίπτωση χρήσης και το επίπεδο εμπειρογνωμοσύνης του χρήστη.

B.2.2. Cloud Attacks

Οι επιθέσεις στο Cloud (Cloud attacks) αναφέρονται σε κακόβουλες δραστηριότητες ή απόπειρες υπονόμευσης της ασφάλειας συστημάτων, υπηρεσιών ή δεδομένων που βασίζονται στο cloud. Το cloud computing έχει γίνει μια δημοφιλή επιλογή για επιχειρήσεις και ιδιώτες να αποθηκεύουν και να επεξεργάζονται δεδομένα, καθώς παρέχει εύκολη πρόσβαση, επεκτασιμότητα και οικονομική αποδοτικότητα. Ωστόσο, αυτή η εξάρτηση από τις υπηρεσίες cloud το έχει επίσης καταστήσει πρωταρχικό στόχο για τους εγκληματίες του κυβερνοχώρου που επιδιώκουν να εκμεταλλευτούν τις ευπάθειες σε αυτά τα συστήματα. Υπάρχουν διάφοροι τύποι επιθέσεων σύννεφων, ο καθένας με το δικό του σύνολο τακτικών και στόχων. Ακολουθούν μερικά παραδείγματα:

- Παραβιάσεις δεδομένων: Μία από τις πιο συνηθισμένες επιθέσεις cloud είναι η παραβίαση δεδομένων, η οποία συμβαίνει όταν ένας εισβολέας αποκτά μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητες πληροφορίες που είναι αποθηκευμένες στο cloud. Αυτό μπορεί να συμβεί λόγω αδύναμων κωδικών πρόσβασης, μη ασφαλών API ή τρωτών σημείων λογισμικού. Μόλις ο εισβολέας αποκτήσει πρόσβαση, μπορεί να κλέψει, να τροποποιήσει ή να διαγράψει τα δεδομένα.
- Επιθέσεις κακόβουλου λογισμικού: Το κακόβουλο λογισμικό μπορεί να μολύνει συστήματα cloud με διάφορα μέσα, όπως μηνύματα ηλεκτρονικού "ψαρέματος" (phishing) ή λήψεις κακόβουλου λογισμικού. Μόλις εγκατασταθεί, το κακόβουλο λογισμικό μπορεί να κλέψει δεδομένα, να ξεκινήσει περαιτέρω επιθέσεις ή να προκαλέσει ζημιά στο σύστημα.
- Επιθέσεις man-in-the-middle: Σε μια επίθεση man-in-the-middle, ένας επιτιθέμενος υποκλέπτει την επικοινωνία μεταξύ δύο μερών για να υποκλέψει ή να χειριστεί τα δεδομένα. Αυτό μπορεί να συμβεί σε συστήματα cloud που χρησιμοποιούν μη ασφαλή πρωτόκολλα ή έχουν αδύναμη κρυπτογράφηση.
- Εσωτερικές επιθέσεις: Οι επιθέσεις insider περιλαμβάνουν κακόβουλες ενέργειες από άτομα που έχουν εξουσιοδοτημένη πρόσβαση στο σύστημα cloud, όπως υπαλλήλους ή εργολάβους. Αυτές οι επιθέσεις μπορεί να είναι

σκόπιμες ή τυχαίες και μπορεί να οδηγήσουν σε κλοπή δεδομένων ή ζημιά στο σύστημα.

Για να αποφευχθούν επιθέσεις στο cloud, υπάρχουν διάφορα μέτρα ασφαλείας που μπορούν να λάβουν οι επιχειρήσεις και τα άτομα. Αυτά περιλαμβάνουν:

- **Ισχυρός έλεγχος ταυτότητας:** Η επιβολή ισχυρών κωδικών πρόσβασης και ελέγχου ταυτότητας πολλών παραγόντων μπορεί να αποτρέψει τη μη εξουσιοδοτημένη πρόσβαση σε συστήματα cloud.
- **Κρυπτογράφηση:** Η κρυπτογράφηση δεδομένων κατά τη μεταφορά και σε κατάσταση ηρεμίας μπορεί να τα προστατεύσει από υποκλοπή και κλοπή.
- **Έλεγχος πρόσβασης:** Ο περιορισμός της πρόσβασης σε ευαίσθητα δεδομένα και λειτουργίες σε εξουσιοδοτημένο προσωπικό μπορεί να αποτρέψει εσωτερικές επιθέσεις.
- **Τακτικές ενημερώσεις και ενημερώσεις κώδικα:** Η ενημέρωση των συστημάτων cloud και του λογισμικού μπορεί να αποτρέψει την εκμετάλλευση γνωστών ευπαθειών.
- **Παρακολούθηση και καταγραφή:** Η παρακολούθηση των συστημάτων cloud για ασυνήθιστη δραστηριότητα και η τήρηση αρχείων καταγραφής όλων των ενεργειών μπορεί να βοηθήσει στον εντοπισμό και την αντιμετώπιση επιθέσεων.
- **Λύσεις ασφάλειας cloud:** Οι πάροχοι cloud προσφέρουν διάφορες λύσεις ασφαλείας, όπως τείχη προστασίας, συστήματα ανίχνευσης εισβολών και πρόληψη απώλειας δεδομένων, οι οποίες μπορούν να ενισχύσουν την ασφάλεια των συστημάτων cloud.

Συμπερασματικά, οι επιθέσεις cloud αποτελούν σοβαρή απειλή για την ασφάλεια και το απόρρητο των δεδομένων που είναι αποθηκευμένα σε συστήματα cloud. Υπάρχουν διάφοροι τύποι επιθέσεων cloud και οι επιχειρήσεις και τα άτομα πρέπει να λάβουν τα κατάλληλα μέτρα για να τις αποτρέψουν. Ακολουθώντας τις βέλτιστες πρακτικές στην ασφάλεια στο cloud, οι οργανισμοί μπορούν να διασφαλίσουν την ακεραιότητα και τη διαθεσιμότητα των δεδομένων και των υπηρεσιών τους.

B.2.3. DDoS Attacks (Distributed Denial of Service Attacks)

Οι DDoS επιθέσεις είναι ένας τύπος επίθεσης στον κυβερνοχώρο που περιλαμβάνει τη συντριπτική πλειοψηφία ενός συστήματος στόχου ή δικτύου με κίνηση ή αιτήματα. Αυτός ο τύπος επίθεσης πραγματοποιείται από μια ομάδα παραβιασμένων συστημάτων, κοινώς γνωστά ως botnet, τα οποία ελέγχονται από έναν εισβολέα.

Στην περίπτωση επιθέσεων DDoS, το dataset KDD99 περιλαμβάνει έναν αριθμό διαφορετικών τύπων επίθεσης. Αυτές περιλαμβάνουν την επίθεση "smurf", η οποία περιλαμβάνει την αποστολή αιτημάτων ηχούς ICMP σε διευθύνσεις μετάδοσης, προκαλώντας μια πλημμύρα απαντήσεων για να κατακλύσει το σύστημα-στόχο. Άλλοι τύποι επίθεσης DDoS στο σύνολο δεδομένων περιλαμβάνουν την επίθεση "teardrop", η οποία περιλαμβάνει την αποστολή κατακερματισμένων πακέτων στο σύστημα προορισμού που δεν μπορούν να επανασυναρμολογηθούν και την επίθεση "land", η οποία περιλαμβάνει την αποστολή πακέτων με τις ίδιες διευθύνσεις IP προέλευσης και προορισμού και αριθμούς θύρας, προκαλώντας το σύστημα προορισμού να εισέλθει σε έναν άπειρο βρόχο.

Μία από τις προκλήσεις της χρήσης του dataset KDD99 για την αξιολόγηση των συστημάτων ανίχνευσης εισβολών είναι ότι βασίζεται σε ένα προσομοιωμένο περιβάλλον, το οποίο μπορεί να μην αντικατοπτρίζει με ακρίβεια τις πραγματικές συνθήκες ενός ζωντανού δικτύου. Επιπλέον, το σύνολο δεδομένων περιλαμβάνει μόνο έναν περιορισμένο αριθμό τύπων επιθέσεων DDoS και ενδέχεται να μην εκπροσωπούνται νεότερες ή πιο εξελιγμένες επιθέσεις.

Για την αξιολόγηση της αποτελεσματικότητας των συστημάτων ανίχνευσης εισβολών, το σύνολο δεδομένων KDD99 παρέχει ένα σύνολο επισημασμένων παραδειγμάτων. Κάθε εγγραφή στο σύνολο δεδομένων επισημαίνεται με μία από τις πολλές πιθανές, υποδεικνύοντας εάν η σύνδεση είναι κανονική ή ανώμαλη και, εάν είναι ανώμαλη, τον τύπο επίθεσης που εκκινείται. Αυτά τα επισημασμένα δεδομένα μπορούν να χρησιμοποιηθούν για την εκπαίδευση μοντέλων μηχανικής μάθησης για την ανίχνευση επιθέσεων DDoS και άλλων τύπων επιθέσεων που βασίζονται σε δίκτυο.

Παρά τους περιορισμούς αυτούς, το σύνολο δεδομένων KDD99 παραμένει ένας πολύτιμος πόρος για την αξιολόγηση των συστημάτων ανίχνευσης εισβολών και την κατανόηση των χαρακτηριστικών των διαφόρων τύπων επιθέσεων στον κυβερνοχώρο, συμπεριλαμβανομένων των επιθέσεων DDoS. Αναλύοντας τα χαρακτηριστικά των συνδέσεων δικτύου και αναπτύσσοντας μοντέλα μηχανικής μάθησης για τον εντοπισμό ανωμαλιών, οι ερευνητές μπορούν να συνεχίσουν να βελτιώνουν την ακρίβεια και την αποτελεσματικότητα των συστημάτων ανίχνευσης εισβολών και να μετριάσουν τους κινδύνους των επιθέσεων στον κυβερνοχώρο.

Μόλις εντοπιστεί μια επίθεση DDoS, το KDD99 μπορεί να εφαρμόσει μια σειρά αντιμέτρων για να μετριάσει τον αντίκτυπό της. Αυτές περιλαμβάνουν την ανακατεύθυνση της κυκλοφορίας σε έναν δευτερεύοντα διακομιστή ή την ανάπτυξη πρόσθετων πόρων για τη διαχείριση του αυξημένου φόρτου. Το KDD99 μπορεί επίσης να συνεργαστεί με άλλα εργαλεία ασφαλείας, όπως τείχη προστασίας και συστήματα ανίχνευσης εισβολών, για να παρέχει μια ολοκληρωμένη άμυνα ενάντια σε επιθέσεις DDoS.

Προκειμένου να μετριαστεί ο κίνδυνος επιθέσεων DDoS, οι οργανισμοί μπορούν να λάβουν ορισμένα μέτρα, όπως η εφαρμογή τμηματοποίησης δικτύου, η ανάπτυξη τείχους προστασίας και συστημάτων ανίχνευσης εισβολών και η χρήση παράδοσης περιεχομένου

B.2.4 Smart Cities

Μια έξυπνη πόλη (Smart City) είναι μια πόλη που χρησιμοποιεί τεχνολογία και δεδομένα για να βελτιώσει την ποιότητα ζωής των κατοίκων της, να ενισχύσει τη βιωσιμότητα και να βελτιστοποιήσει τις αστικές υπηρεσίες. Αυτό μπορεί να περιλαμβάνει μια ποικιλία διαφορετικών συστημάτων και τεχνολογιών, όπως:

- Έξυπνα συστήματα μεταφορών: Αυτά μπορεί να περιλαμβάνουν έξυπνα συστήματα διαχείρισης της κυκλοφορίας, συστήματα δημόσιας συγκοινωνίας που χρησιμοποιούν δεδομένα σε πραγματικό χρόνο για τη βελτιστοποίηση διαδρομών και χρονοδιαγραμμάτων και αυτόνομα οχήματα
- Ενεργειακά αποδοτικά κτίρια: Οι έξυπνες πόλεις μπορούν να χρησιμοποιούν αισθητήρες και άλλες τεχνολογίες για τη βελτιστοποίηση της χρήσης ενέργειας στα κτίρια, μειώνοντας τη σπατάλη ενέργειας και το κόστος.
- Έξυπνος δημόσιος φωτισμός: Τα έξυπνα φώτα δρόμου μπορούν να χρησιμοποιούν αισθητήρες για να ανιχνεύουν πότε δεν υπάρχουν άτομα ή οχήματα στην περιοχή και να σβήνουν ή να χαμηλώνουν τα φώτα, εξοικονομώντας ενέργεια.
- Ενισχυμένη δημόσια ασφάλεια: Οι έξυπνες πόλεις μπορούν να χρησιμοποιούν διάφορες τεχνολογίες για την ενίσχυση της δημόσιας ασφάλειας, όπως κάμερες παρακολούθησης με λογισμικό αναγνώρισης προσώπου και συστήματα ανίχνευσης πυροβολισμών.

Αυτά είναι μόνο μερικά παραδείγματα των πολλών διαφορετικών συστημάτων και τεχνολογιών που μπορούν να αποτελέσουν μέρος μιας έξυπνης πόλης. Ο στόχος όλων αυτών των συστημάτων είναι να χρησιμοποιήσουν δεδομένα και τεχνολογία για να κάνουν τις πόλεις πιο αποτελεσματικές, βιώσιμες και βιώσιμες για τους κατοίκους τους.

Ωστόσο, με όλα αυτά τα συστήματα έρχονται αυξημένοι κίνδυνοι ασφάλειας στον κυβερνοχώρο. Καθώς όλο και περισσότερα συστήματα συνδέονται με το διαδίκτυο, γίνονται ευάλωτα σε επιθέσεις στον κυβερνοχώρο.

Καθώς όλο και περισσότερα συστήματα συνδέονται με το διαδίκτυο, γίνονται ευάλωτα σε επιθέσεις στον κυβερνοχώρο. Οι επιτιθέμενοι θα μπορούσαν ενδεχομένως να αποκτήσουν πρόσβαση σε ευαίσθητα δεδομένα, να διαταράξουν κρίσιμες υπηρεσίες ή ακόμη και να προκαλέσουν σωματική βλάβη. Εδώ παρεμβαίνει το dataset KDD99.

Όσον αφορά τις έξυπνες πόλεις, το σύνολο δεδομένων KDD99 μπορεί να χρησιμοποιηθεί για τη μελέτη της ασφάλειας διαφόρων συστημάτων και δικτύων έξυπνων πόλεων, συμπεριλαμβανομένων των συστημάτων μεταφορών, των ενεργειακών δικτύων και των συστημάτων δημόσιας ασφάλειας. Αναλύοντας τα δεδομένα κίνησης δικτύου στο σύνολο δεδομένων, οι ερευνητές μπορούν να αναπτύξουν αλγόριθμους για τον εντοπισμό και την πρόληψη επιθέσεων στον κυβερνοχώρο σε αυτά τα συστήματα.

Για παράδειγμα, οι ερευνητές θα μπορούσαν να χρησιμοποιήσουν το dataset KDD99 για να αναπτύξουν έναν αλγόριθμο που μπορεί να ανιχνεύσει πότε ένας εισβολέας προσπαθεί να αποκτήσει πρόσβαση σε ένα έξυπνο σύστημα διαχείρισης κυκλοφορίας. Ο αλγόριθμος θα μπορούσε να αναλύσει τα δεδομένα κίνησης δικτύου και να αναζητήσει μοτίβα που υποδεικνύουν απόπειρα εισβολής. Μόλις ο αλγόριθμος εντοπίσει μια απόπειρα εισβολής, θα μπορούσε να αποκλείσει αυτόματα τον εισβολέα και να ειδοποιήσει τους διαχειριστές του συστήματος.

Ομοίως, το dataset KDD99 θα μπορούσε να χρησιμοποιηθεί για την ανάπτυξη αλγορίθμων για τον εντοπισμό και την πρόληψη επιθέσεων σε άλλα συστήματα έξυπνων πόλεων, όπως τα ενεργειακά δίκτυα και τα συστήματα δημόσιας ασφάλειας. Χρησιμοποιώντας το σύνολο δεδομένων για την ανάπτυξη αυτών των αλγορίθμων, οι ερευνητές μπορούν να βοηθήσουν στη βελτίωση της συνολικής ασφάλειας των έξυπνων πόλεων και στη μείωση του κινδύνου επιθέσεων στον κυβερνοχώρο.

Συμπερασματικά, οι έξυπνες πόλεις χρησιμοποιούν τεχνολογία και δεδομένα για να βελτιώσουν την ποιότητα ζωής των κατοίκων, αλλά με την αυξημένη συνδεσιμότητα ενέχει αυξημένους κινδύνους για την ασφάλεια στον κυβερνοχώρο.

Το KDD99 μπορεί να χρησιμοποιηθεί για τη μελέτη της ασφάλειας των συστημάτων έξυπνων πόλεων και την ανάπτυξη αλγορίθμων για τον εντοπισμό και την πρόληψη κυβερνοεπιθέσεων σε αυτά τα συστήματα, με σκοπό να γίνουν περισσότερο ασφαλείς, και βιώσιμες για όλους.

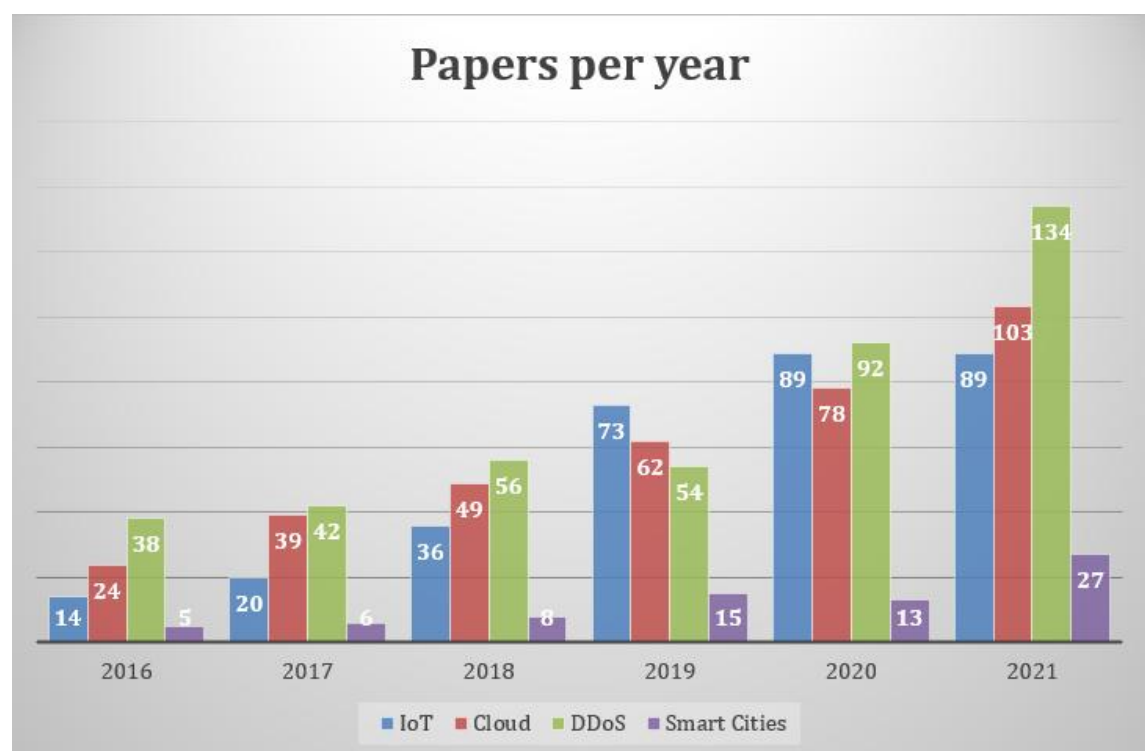
Β.3. ΕΡΕΥΝΑ ΧΡΗΣΗΣ ΑΝΑ ΕΤΟΣ

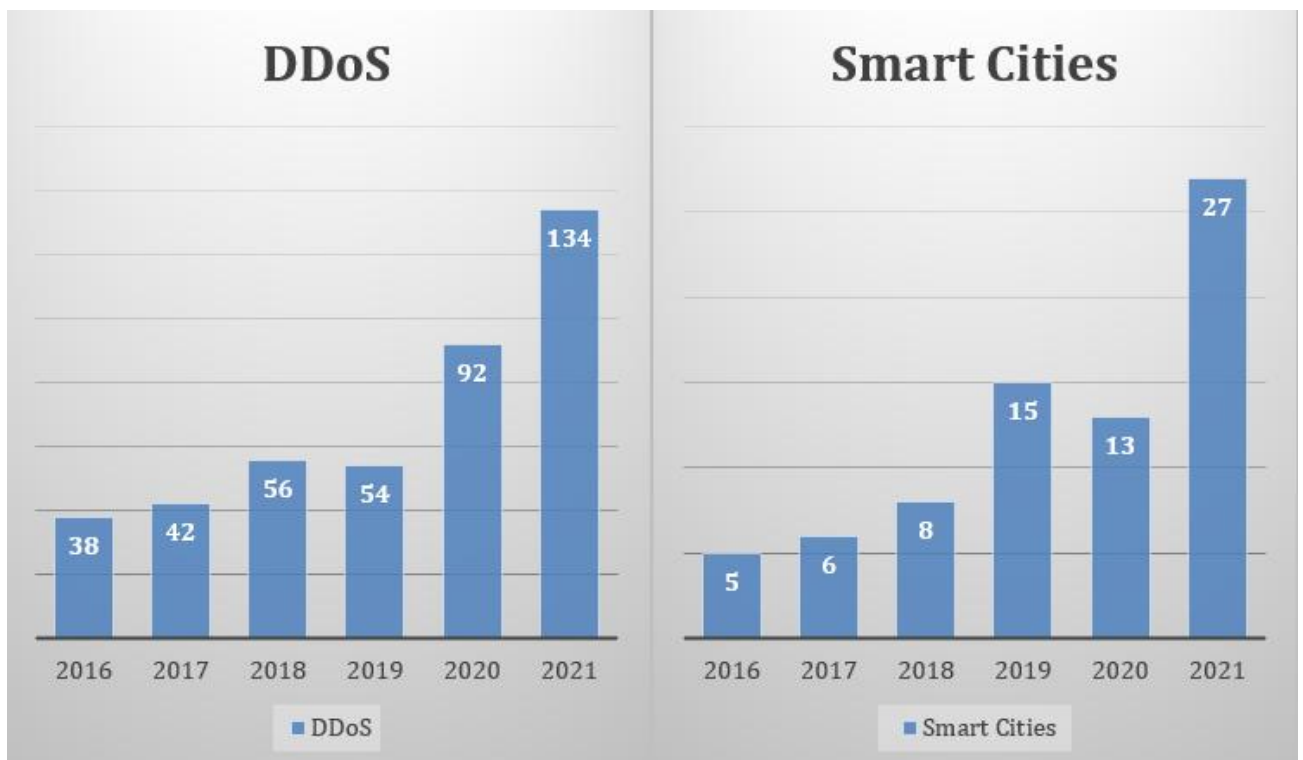
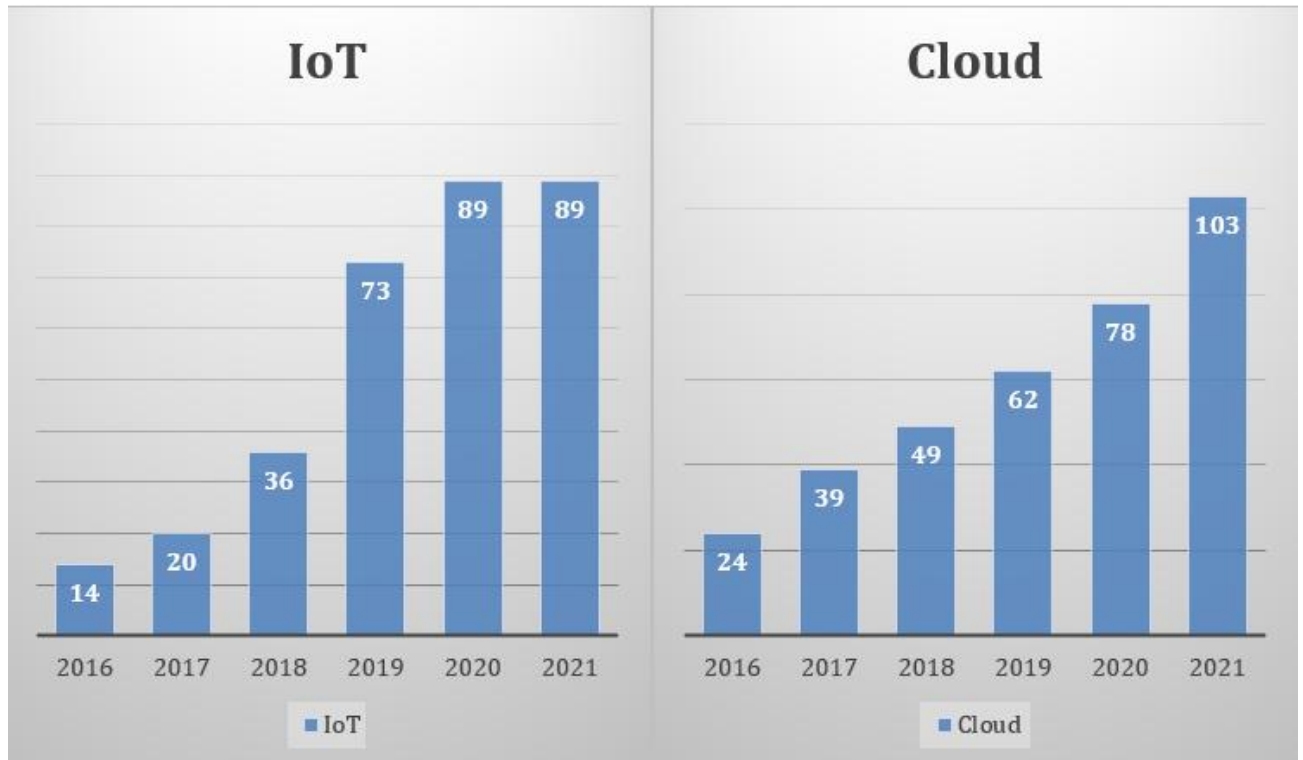
	2016	2017	2018	2019	2020	2021
Συνολικά Papers	184	263	274	369	403	456
IoT	14	20	36	73	89	89
Cloud Attacks	24	39	49	62	78	103
DDoS	38	42	56	54	92	134
Smart Cities	5	6	8	15	13	27

ΠΙΝΑΚΑΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΕΡΕΥΝΑΣ

ΓΙΑ ΤΗΝ ΧΡΟΝΙΚΗ ΠΕΡΙΟΔΟ 2016-2021

ΓΡΑΦΗΜΑΤΑ ΣΥΓΚΡΙΣΗΣ





Σύμφωνα με τα γραφήματα, παρατηρείται πως σχεδόν κάθε χρόνο, η χρήση του dataset αυξάνεται ολοένα και περισσότερο σε κάθε μια από τις 4 κατηγορίες, ειδικά κατά τα έτη 2020-2021. Αποδεικνύεται, λοιπόν, ότι υπάρχει μεγαλύτερη κινητικότητα όσον αφορά τις επιθέσεις τα τελευταία χρόνια.

Γ' ΚΕΦΑΛΑΙΟ

ΥΠΟΛΟΙΠΑ DATASETS

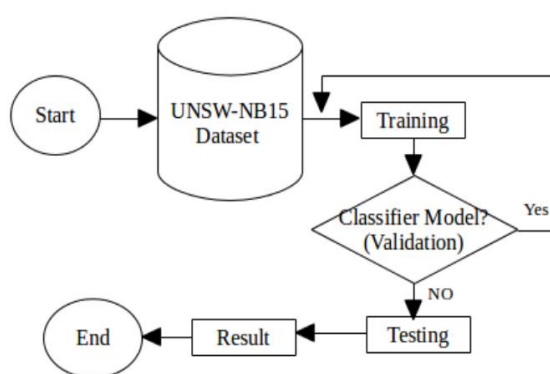
Γ.1. UNSW-NB15

Το dataset UNSW-NB15 χρονολογείται από το 2015 (Moustafa & Slay, 2015), διαθέτει 49 χαρακτηριστικά και μια ποικιλία από κανονικές και επιτιθέμενες δραστηριότητες, συμπεριλαμβανομένων των ετικετών κλάσης με συνολικά 2.540.044 εγγραφές. Υπάρχουν 2.21.876 κανονικές εγγραφές και 3.21.283 επιθέσεις στο σύνολο των εγγραφών. Τα χαρακτηριστικά του dataset UNSW-NB15 κατηγοριοποιούνται σε τρεις ομάδες και συγκεκριμένα⁷:

- Βασικές δυνατότητες (6-18)
- Χαρακτηριστικά περιεχομένου (19-26)
- Λειτουργίες χρόνου (27-35)

Τα χαρακτηριστικά από το 36 έως το 40 είναι γνωστά ως χαρακτηριστικά ειδικού σκοπού, ενώ εκείνα από το 41 έως το 47 ως χαρακτηριστικά σύνδεσης. Επιπρόσθετα, περιλαμβάνει σύγχρονες επιθέσεις και συγκεκριμένα εννέα τύπους επιθέσεων (Analysis, Backdoor, DoS, Exploits, Fuzzers, Generic, Reconnaissance, Shellcode and Worms)

Ωστόσο, αυτό το dataset έχει δύο μειονεκτήματα: ανισορροπία και αλληλεπικάλυψη κλάσεων, τα οποία χρήζουν αντιμετώπισης πριν χρησιμοποιηθούν για την ανάπτυξη μοντέλων. Εάν δεν αντιμετωπιστούν είναι πιθανό να εμποδίσουν την ανίχνευση επιθέσεων και την απόδοση αναγνώρισης των συστημάτων ανίχνευσης εισβολών.



Εικόνα 2: Λειτουργία dataset UNSW-NB15

⁷ Z. Zoghi, & G. Serpen (2021). Unsw-nb15 computer security dataset: Analysis through visualization. arXiv preprint arXiv:2101.05067.

Γ.2. NSL-KDD

Το εγγενές πρόβλημα του συνόλου δεδομένων KDD οδηγεί σε νέα έκδοση του συνόλου δεδομένων NSL-KDD. Είναι πολύ δύσκολο να υποδηλωθούν υπάρχοντα πρωτότυπα δίκτυα, αλλά εξακολουθεί να μπορεί να εφαρμοστεί ως ένα αποτελεσματικό σύνολο δεδομένων αναφοράς για τους ερευνητές ώστε να συγκρίνουν διαφορετικές μεθόδους ανίχνευσης εισβολής (McHugh, 2000).

Οι Tavallae et al. (Tavallae, 2009) διεξήγαγαν μια στατιστική ανάλυση σε αυτό το σύνολο δεδομένων και βρήκαν δύο σημαντικά ζητήματα που επηρεάζουν σε μεγάλο βαθμό την απόδοση του αξιολογούμενου συστήματος και καταλήγουν σε πολύ κακή αξιολόγηση των προσεγγίσεων ανίχνευσης ανωμαλιών.

Για την επίλυση αυτών των προβλημάτων, πρότειναν ένα νέο σύνολο δεδομένων, το NSL-KDD, το οποίο αποτελείται μόνο από επιλεγμένες εγγραφές από το πλήρες σύνολο δεδομένων KDD και δεν πάσχει από καμία από τις αναφερόμενες ελλείψεις.



Εικόνα 3: Σχέση μεταξύ DARPA-98, KDD-99 & NSL-KDD

Τα πλεονεκτήματα του συνόλου δεδομένων NSL-KDD είναι :

- Α. Δεν υπάρχουν περιττές εγγραφές, επομένως ο ταξινομητής δεν θα παράγει κανένα μεροληπτικό αποτέλεσμα
- Β. Δεν υπάρχει διπλή εγγραφή στο σύνολο δοκιμών που έχουν καλύτερα ποσοστά μείωσης.
- Γ. Ο αριθμός των επιλεγμένων εγγραφών από κάθε ομάδα δύσκολου επιπέδου είναι αντιστρόφως ανάλογος με το ποσοστό των εγγραφών στο αρχικό σύνολο δεδομένων KDD.

Το σύνολο δεδομένων εκπαίδευσης αποτελείται από 21 διαφορετικές επιθέσεις από τις 37 που υπάρχουν στο σύνολο δεδομένων δοκιμής. Οι γνωστοί τύποι επιθέσεων είναι αυτοί που υπάρχουν στο σύνολο δεδομένων εκπαίδευσης, ενώ οι νέες επιθέσεις

είναι οι πρόσθετες επιθέσεις στο δοκιμαστικό σύνολο δεδομένων, δηλαδή δεν είναι διαθέσιμες στα σύνολα δεδομένων εκπαίδευσης. Οι τύποι επίθεσης ομαδοποιούνται σε τέσσερις κατηγορίες (DoS, Probe, U2R και R2L) όπως και το KDD.

Γ.3 DARPA

Όπως έχει ήδη αναφερθεί σε προηγούμενες ενότητες, το DARPA-98 είναι ένα dataset που χρονολογείται από το 1998, με βάση το οποίο κατασκευάστηκαν τα δύο προαναφερθέντα datasets (KDD-99 & NSL-KDD). Περιλαμβάνει 4 gigabyte συμπιεσμένων ανεπεξέργαστων δυαδικών δεδομένων TCPdump, τα οποία συλλέχθηκαν κατά τη διάρκεια 7 εβδομάδων κίνησης δικτύου και μεταποιήθηκαν σε περίπου 5 εκατομμύρια εγγραφές σύνδεσης, η καθεμία αποτελούμενη από 100 byte. Αντίστοιχα, οι δύο εβδομάδες δεδομένων δοκιμής έχουν περίπου 2 εκατομμύρια εγγραφές σύνδεσης. Είχε σχεδιαστεί για να προσομοιώσει μία βάση δεδομένων της Πολεμικής Αεροπορίας.

Το 1999 επανασχεδιάστηκε ώστε κάποιες ημέρες να μην υπάρχουν επιθέσεις προκειμένου να εκπαιδεύσει ένα σύστημα ανίχνευσης σε κανονική συμπεριφορά. Το dataset του 1999 επικρίθηκε τόσο από τον McHugh, ο οποίος τόνισε ότι η έλλειψη στατιστικών στοιχείων είναι παρόμοια με την πραγματική κίνηση της βάσης δεδομένων της Πολεμικής Αεροπορίας (McHugh, 2000), όσο και οι Mahoney και Chan (Chan, 2003) οι οποίοι επέκριναν το dataset για την ύπαρξη τεχνουργημάτων προσομοίωσης. Τα τεχνουργήματα αυτά περιλαμβάνουν μικρό εύρος τιμών χαρακτηριστικών σε σύγκριση με την πραγματική κυκλοφορία. Τα χαρακτηριστικά περιλαμβάνουν απομακρυσμένες διευθύνσεις IP, TTL, επιλογές TCP και μέγεθος παραθύρου TCP. Τέτοια τεχνουργήματα έχουν επιπτώσεις στην αξιολόγηση των IDS, ειδικά των ανιχνευτών που βασίζονται σε ανωμαλίες. Ισχυρίζονται ότι οι ιδιοσυγκρασίες της προσομοίωσης κυκλοφορίας θα μπορούσαν να βοηθήσουν στον εντοπισμό ορισμένων εισβολών.

Το 2009 σχεδιάστηκε για ακόμη μία φορά, προκειμένου όλες οι ημέρες να περιλαμβάνουν ποικιλία επιθέσεων. Προσομοιώνει ένα υποδίκτυο συνδεδεμένο στο Διαδίκτυο.

Γ.4 HTTP DATASET CSIC 2010

Το HTTP dataset CSIC 2010 περιέχει χιλιάδες web requests τα οποία παράγονται αυτόματα. Έχει δημιουργηθεί και αναπτυχθεί στο Εθνικό Συμβούλιο Έρευνας της Ισπανίας (Information Security Institute του CSIC). Περιλαμβάνει την παραγόμενη κίνηση που στοχεύει σε μια εφαρμογή web ηλεκτρονικού εμπορίου. Σε αυτήν την εφαρμογή web, οι χρήστες μπορούν να αγοράζουν στοιχεία χρησιμοποιώντας ένα καρτόσι αγορών και να εγγράφονται, παρέχοντας ορισμένες προσωπικές πληροφορίες. Καθώς είναι μια εφαρμογή web στα ισπανικά, το σύνολο δεδομένων περιέχει μερικούς λατινικούς χαρακτήρες. Το dataset δημιουργείται αυτόματα και περιέχει 36.000 κανονικά requests και περισσότερα από 25.000 μη φυσιολογικά requests. Τα HTTP requests έχουν την ετικέτα "κανονικό" ή "μη κανονικό" και το dataset περιλαμβάνει επιθέσεις, όπως έγχυση SQL, υπερχειλίση buffer, συλλογή πληροφοριών, αποκάλυψη αρχείων, έγχυση CRLF, XSS, συμπερίληψη πλευράς διακομιστή, αλλοίωση παραμέτρων κ.α..

Η κυκλοφορία δημιουργείται ακολουθώντας τα επόμενα βήματα:

- Πρώτα, συλλέγονται πραγματικά δεδομένα για όλες τις παραμέτρους της εφαρμογής web. Όλα τα δεδομένα (ονόματα, επώνυμα, διευθύνσεις κλπ.) εξάγονται από πραγματικές βάσεις δεδομένων. Αυτές οι τιμές αποθηκεύονται σε δύο βάσεις δεδομένων: μία για τις κανονικές τιμές και μία για τις ανώμαλες. Επιπλέον, παρατίθενται όλες οι δημόσιες διαθέσιμες σελίδες της εφαρμογής web.
- Στη συνέχεια, δημιουργούνται κανονικές και μη φυσιολογικές αιτήσεις για κάθε ιστοσελίδα. Στην περίπτωση που τα κανονικά αιτήματα έχουν παραμέτρους, οι τιμές παραμέτρων συμπληρώνονται με δεδομένα που λαμβάνονται από την κανονική βάση δεδομένων τυχαία. Η διαδικασία είναι ανάλογη για ανώμαλες αιτήσεις, όπου οι τιμές των παραμέτρων λαμβάνονται από την ανώμαλη βάση δεδομένων.

Εξετάστηκαν τρεις τύποι ανώμαλων αιτήσεων:

A. Οι στατικές επιθέσεις προσπαθούν να ζητήσουν κρυφούς (ή ανύπαρκτους) πόρους. Τα αιτήματα αυτά περιλαμβάνουν απαρχαιωμένα αρχεία, αναγνωριστικό

περιόδου λειτουργίας σε επανεγγραφή URL, αρχεία διευθέτησης, προεπιλεγμένα αρχεία, κ.α.

B. Οι δυναμικές επιθέσεις τροποποιούν τα έγκυρα ορίσματα αιτήσεων: έγχυση SQL, έγχυση CRLF, cross-site scripting, υπερχειλίσσεις buffer, κ.α..

Γ. Ακούσιες παράνομες αιτήσεις. Οι αιτήσεις αυτές δεν έχουν κακόβουλη πρόθεση, ωστόσο δεν ακολουθούν την κανονική συμπεριφορά της εφαρμογής web και δεν έχουν την ίδια δομή με τις κανονικές τιμές παραμέτρων (για παράδειγμα, ένας αριθμός τηλεφώνου που αποτελείται από επιστολές).

Επιπρόσθετα, το HTTP dataset CSIC 2010 διαιρείται σε τρία διαφορετικά υποσύνολα. Ένα υποσύνολο για τη φάση εκπαίδευσης, το οποίο έχει μόνο κανονική κυκλοφορία. Και δύο υποσύνολα για τη φάση της δοκιμής, το ένα με κανονική κυκλοφορία και το άλλο με κακόβουλη κυκλοφορία.

Γ.5. ISCX 2012

Το ISCX 2012 dataset, το οποίο έχει δημιουργηθεί στο Information Security Centre of Excellence στο Πανεπιστήμιο του New Brunswick, χρησιμοποιείται για την εκτέλεση πειραμάτων και την αξιολόγηση των επιδόσεων της προτεινόμενης προσέγγισης μας. Ολόκληρο το ISCX dataset περιλαμβάνει πάνω από δύο εκατομμύρια πακέτα κυκλοφορίας τα οποία περιγράφονται με τη χρήση 20 χαρακτηριστικών. Καλύπτει επτά ημέρες δραστηριοτήτων δικτύου και περιέχει δεδομένα κανονικής κίνησης και επίθεσης (δεδομένα επίθεσης που αντιπροσωπεύουν το 2% της συνολικής κίνησης).

Το UNB ISCX 2012 παρουσιάζει τα ακόλουθα χαρακτηριστικά⁸:

Ρεαλιστικό δίκτυο και κίνηση: Ιδανικά, ένα dataset δεν θα πρέπει να παρουσιάζει ανεπιθύμητες ιδιότητες, τόσο από άποψη δικτύου όσο και από άποψη κυκλοφορίας. Αυτό γίνεται για να δοθεί μια σαφέστερη εικόνα των πραγματικών επιπτώσεων των επιθέσεων μέσω του δικτύου κι των αντίστοιχων απαντήσεων των σταθμών εργασίας. Για το λόγο αυτό, είναι απαραίτητο η κυκλοφορία να φαίνεται και να συμπεριφέρεται όσο το δυνατόν πιο ρεαλιστικά. Αυτό περιλαμβάνει τόσο την κανονική όσο και την ανώμαλη κυκλοφορία. Οποιαδήποτε τεχνητή εισαγωγή στοιχείων μετά την καταγραφή θα επηρεάσει αρνητικά τα ανεπεξέργαστα δεδομένα και θα εισαγάγει πιθανές ασυνέπειες στο τελικό dataset. Κατά συνέπεια, όλες αυτές οι προσαρμογές αποθαρρύνονται σε μεγάλο βαθμό.

Dataset με ετικέτα: Ένα dataset με ετικέτα είναι εξέχουσας σημασίας για την αξιολόγηση διαφόρων μηχανισμών ανίχνευσης. Ως εκ τούτου, η δημιουργία ενός dataset σε ένα ελεγχόμενο και περιορισμένο περιβάλλον επιτρέπει τη διάκριση της ανώμαλης δραστηριότητας από την κανονική κυκλοφορία, εξαλείφοντας την ανέφικτη διαδικασία της μη αυτόματης επισήμανσης.

⁸ S. Ni, Q. Qian, & R. Zhang (2018). Malware identification using visualization images and deep learning. *Computers & Security*, 77, 871-885.

Συνολική καταγραφή αλληλεπίδρασης: Ο όγκος των πληροφοριών που είναι διαθέσιμες στους μηχανισμούς ανίχνευσης είναι ζωτικής σημασίας, καθώς αυτό παρέχει τα μέσα για την ανίχνευση ανώμαλης συμπεριφοράς. Με άλλα λόγια, αυτές οι πληροφορίες είναι απαραίτητες για τη μετά την αξιολόγηση και την ορθή ερμηνεία των αποτελεσμάτων. Ως εκ τούτου, θεωρείται σημαντική απαίτηση για ένα dataset να περιλαμβάνει όλες τις αλληλεπιδράσεις δικτύου.

Πλήρης καταγραφή: Οι ανησυχίες για το απόρρητο που σχετίζονται με την κοινή

ΗΜΕΡΑ	ΗΜΕΡΟΜΗΝΙΑ	ΠΕΡΙΓΡΑΦΗ	ΜΕΓΕΘΟΣ (GB)
Παρασκευή	11/06/2010	Κανονική Δραστηριότητα. Καμία	16,1

χρήση πραγματικών ιχνών δικτύου ήταν ένα από τα σημαντικότερα εμπόδια για τους ερευνητές ασφάλειας δικτύων, καθώς οι πάροχοι δεδομένων είναι συχνά απρόθυμοι να μοιραστούν τέτοιες πληροφορίες. Κατά συνέπεια, τα περισσότερα τέτοια ίχνη είτε χρησιμοποιούνται εσωτερικά, γεγονός που εμποδίζει άλλους ερευνητές να αξιολογήσουν και να συγκρίνουν με ακρίβεια τα συστήματά τους, είτε είναι σε μεγάλο βαθμό ανώνυμα με το ωφέλιμο φορτίο να αφαιρείται εντελώς με αποτέλεσμα τη μειωμένη χρησιμότητα για τους ερευνητές. Σε αυτή την εργασία, ο πρωταρχικός στόχος είναι η δημιουργία ιχνών δικτύου σε ελεγχόμενο περιβάλλον δοκιμών, εξαλείφοντας με αυτόν τον τρόπο εντελώς την ανάγκη για κάποιο «ξεκαθάρισμα» και διατηρώντας τη φυσικότητα του dataset που προκύπτει.

Διαφορετικά σενάρια εισβολής: Οι επιθέσεις έχουν αυξηθεί σε συχνότητα, μέγεθος, ποικιλία και πολυπλοκότητα τα τελευταία χρόνια. Το εύρος των απειλών έχει επίσης αλλάξει σε πιο σύνθετα σχήματα, συμπεριλαμβανομένων των επιθέσεων που στοχεύουν σε υπηρεσίες και εφαρμογές. Τέτοιες επιθέσεις μπορούν να προκαλέσουν πολύ πιο σοβαρές διαταραχές και απαιτούν μια πιο εμπεριστατωμένη εικόνα των υπηρεσιών IP και των εφαρμογών για τον εντοπισμό τους. Μέσω της εκτέλεσης σεναρίων επίθεσης και της εφαρμογής μη φυσιολογικής συμπεριφοράς, αποσκοπείται η εκτέλεση ενός ποικίλου συνόλου επιθέσεων πολλαπλών σταδίων, το καθένα προσεκτικά σχεδιασμένο και προσανατολισμένο προς τις πρόσφατες τάσεις στις απειλές για την ασφάλεια. Αυτός ο στόχος συχνά χαρακτηρίζει πολλά από τα διαθέσιμα datasets ως αναποτελεσματικά και ακατάλληλα για την αξιολόγηση των αποτελεσμάτων της έρευνας.

		κακόβουλη Δραστηριότητα	
Σάββατο	12/06/2010	Κανονική Δραστηριότητα. Καμία κακόβουλη Δραστηριότητα	4,22
Κυριακή	13/06/2010	Διείσδυση στο δίκτυο από μέσα. Κανονική Δραστηριότητα	3,95
Δευτέρα	14/06/2010	HTTP DoS. Κανονική Δραστηριότητα	6,58
Τρίτη	15/06/2010	Κατανεμημένη Άρνηση Παροχής Υπηρεσιών με τη χρήση IRC Botnet	23,4
Τετάρτη	16/06/2010	Κανονική Δραστηριότητα. Καμία κακόβουλη Δραστηριότητα	17,6
Πέμπτη	17/06/2010	Brute Force SSH. Κανονική Δραστηριότητα	12,3

Τέσσερις διαφορετικοί τύποι επίθεσης, που αναφέρονται ως Brute Force SSH, Interling, HTTP DoS και DDoS διεξάγονται και καταγράφονται μαζί με κανονική κυκλοφορία σε 7 διαδοχικές ημέρες.

Ακολουθεί πίνακας με παράδειγμα χρήσης του dataset, στο οποίο φαίνεται η ημέρα, η ημερομηνία, η περιγραφή, και το μέγεθος σε gigabyte

Πίνακας 2 : Παράδειγμα Χρήσης Dataset ISCX 2012

Γ.6 CIC IDS-2017

Το dataset CIC IDS-2017 κατασκευάστηκε με τη χρήση του αναλυτή ροής κυκλοφορίας δικτύου NetFlowMeter. Το εργαλείο συλλέγει περισσότερα από 80 χαρακτηριστικά κυκλοφορίας δικτύου και υποστηρίζει αμφίδρομες ροές. Η συλλογή έγινε σε διάρκεια 5 ημερών, κατά τη διάρκεια της οποίας καταγράφηκαν πάνω από 80 χαρακτηριστικά και 15 τάξεις. Η υποδομή επίθεσης αποτελούνταν από 4 υπολογιστές, 1 δρομολογητή και 1 μεταγωγέα, ενώ η υποδομή του θύματος αποτελούνταν από 3 διακομιστές, 1 τείχος προστασίας και 2 μεταγωγείς. Διαφέρει από άλλα παραδοσιακά datasets δικτύου, καθώς χρησιμοποιεί διαφορετικά προφίλ δικτύου για να δημιουργήσει υπερ-ρεαλιστικά δεδομένα δικτύου και να επιτεθεί σε δεδομένα με βάση τους πραγματικούς χρήστες. Άλλα datasets, όπως το NSL-KDD, βασίζονται σε γενική κυκλοφορία με ετικέτες με περιορισμένο αριθμό τύπων επίθεσης, ενώ το CIC IDS-2017 χρησιμοποιεί μεγάλη ποικιλία τύπων επίθεσης και ψευδο-οργανικού περιεχομένου από ρεαλιστικά προφίλ. Ολόκληρο το ίδιο το σύνολο δεδομένων σχεδιάστηκε και συλλέχθηκε με τα έντεκα χαρακτηριστικά ενός ιδανικού συνόλου δεδομένων στην πρώτη γραμμή της υλοποίησης.

Ενώ το σύνολο δεδομένων CIC IDS-2017 θεωρείται μια αρκετά πρόσφατη και σύγχρονη λύση για τη συμπλήρωση των συστημάτων ανίχνευσης εισβολής, υπάρχουν κάποιες αδυναμίες, όπως το μέγεθος των δεδομένων που αποθηκεύονται σε κάθε αρχείο εξόδου. Επειδή υπάρχει μεγάλος αριθμός παρουσιών δεδομένων σε κάθε αρχείο, είναι πολύ δύσκολη η επεξεργασία κάθε αρχείου και η φόρτωση όλων των δεδομένων στη μνήμη. Το μέγεθος των αρχείων ευθύνεται επίσης για μεγαλύτερο χρόνο επεξεργασίας και μείωση της αποτελεσματικότητας πολλών αλγορίθμων ταξινόμησης.

Γ.7 CSE-CIC-IDS

Αυτό είναι το πιο πρόσφατο διαθέσιμο dataset του 2018/2019 από το Καναδικό Ινστιτούτο για την Ασφάλεια στον Κυβερνοχώρο (Canadian Institute of Cyber Security). Αυτά τα προφίλ μπορούν να χρησιμοποιηθούν από παράγοντες ή άτομα για τη δημιουργία συμβάντων σε δίκτυο και μπορούν να εφαρμοστούν σε διάφορα πρωτόκολλα δικτύου με διαφορετικές τοπολογίες. Επιπλέον, το dataset ενισχύθηκε με την εξέταση των προτύπων που χρησιμοποιούνται για τη δημιουργία του CIC-IDS2017.

Εκτός από τα βασικά κριτήρια, προσφέρει τα ακόλουθα πλεονεκτήματα:

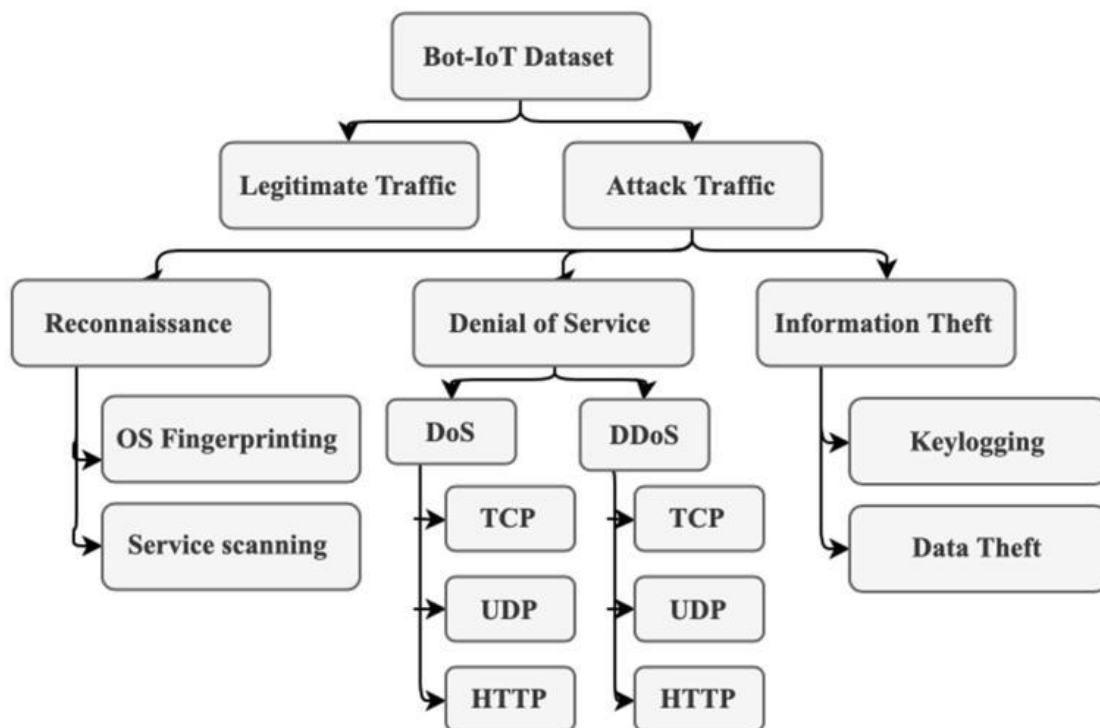
- Ο αριθμός των διπλότυπων δεδομένων είναι πολύ χαμηλός,
- Δεν υπάρχουν σχεδόν καθόλου αβέβαια στοιχεία,
- Το σύνολο δεδομένων είναι σε μορφή CSV, οπότε είναι έτοιμο για χρήση χωρίς επεξεργασία.

Δύο προφίλ κατηγοριοποιήθηκαν, και πέντε διαφορετικές μέθοδοι επίθεσης χρησιμοποιήθηκαν στο dataset. Επιπλέον, δημιουργήθηκαν διάφορα σενάρια και συλλέχθηκαν δεδομένα σε καθημερινή βάση. Η επεξεργασία του dataset γινόταν καθημερινά και καταγράφονταν ανεπεξέργαστα δεδομένα. Κατά τη δημιουργία δεδομένων, 80 στατιστικές ιδιότητες όπως ο χρόνος, ο αριθμός των πακέτων, ο αριθμός των bytes, το μήκος του πακέτου, κ.α. υπολογίζονται ξεχωριστά σε ευθεία και αντίστροφη κατεύθυνση, και παρέχονται πληροφορίες για το αν μια επίθεση προστίθεται. Το τελικό dataset δημοσιεύεται μέσω του Διαδικτύου στους ερευνητές, με περίπου 5 εκατομμύρια δεδομένα τόσο σε μορφή PCAP όσο και σε μορφή CSV. Το dataset με μορφή CSV θα πρέπει να χρησιμοποιείται αν πρόκειται να χρησιμοποιηθούν τεχνικές τεχνητής νοημοσύνης, ενώ τα δεδομένα PCAP που δεν έχουν υποστεί επεξεργασία θα πρέπει να χρησιμοποιούνται αν πρόκειται να εξαχθεί ένα νέο χαρακτηριστικό.

Γ.8 BOT-IOT

Το Bot-IoT είναι ένα πρόσφατο και διαθέσιμο στο κοινό dataset που απεικονίζει την κίνηση επίθεσης που ξεκίνησε από τα δίκτυα BotNet κατά του Internet of Things (IoT). Η κανονική κίνηση (χωρίς επίθεση) αντιπροσωπεύεται από πάνω από 9.000 από τις περίπου 73.000.000 παρουσίες μεγάλων δεδομένων που αποτελούν αυτό το σύνολο δεδομένων.

Αυτό το dataset αναπτύχθηκε σε ένα ρεαλιστικό πεδίο δοκιμών και έχει επισημανθεί, με τα χαρακτηριστικά της ετικέτας να υποδεικνύουν μια ροή επίθεσης, την κατηγορία επιθέσεων και την υποκατηγορία για πιθανούς σκοπούς ταξινόμησης πολλαπλών κλάσεων. Δημιουργήθηκαν πρόσθετα χαρακτηριστικά για την ενίσχυση των προγνωστικών δυνατοτήτων των ταξινομητών που εκπαιδεύτηκαν σε αυτό το μοντέλο.



Εικόνα 4: Λειτουργία BOT- IoT

Λ. ΣΥΜΠΕΡΑΣΜΑΤΑ

KDD-99

Δημιουργία: University of California

Τύποι Επιθέσεων: DoS, Probe, R2L, U2R

Περιγραφή: Αποτελείται από περιττά και διπλά δείγματα δεδομένων

UNSW- NB15

Δημιουργία: Cyber Range Lab of UNSW Canberra

Τύποι Επιθέσεων: Fuzzers, Analysis, Backdoor, DoS, Exploit, Generic, Reconnaissance, Shellcode, Worm

Περιγραφή: Δημιουργεί ένα υβρίδιο των ρεαλιστικών σύγχρονων κανονικών δραστηριοτήτων και των συνθετικών σύγχρονων συμπεριφορών επίθεσης από την κυκλοφορία δικτύου

NSL- KDD

Δημιουργία: University of California

Τύποι Επιθέσεων: DoS, Probe, R2L, U2R

Περιγραφή: Πιο λεπτομερής έκδοση του dataset KDD-99 και αφορά περιορισμένο αριθμό επιθέσεων.

DARPA

Δημιουργία: MIT Lincoln laboratory

Τύποι Επιθέσεων: DoS, Probe, R2L, U2R

Περιγραφή: Απουσία ψευδώς θετικών περιστατικών, παρατυπίες σε περιπτώσεις δεδομένων επίθεσης.

HTTP DATASET CSIC 2010

Δημιουργία: Information Security Institute CSIC

Τύποι Επιθέσεων: SQL injection, buffer overflow, information gathering, files disclosure, CRLF injection, XSS, server side include, parameter tampering

Περιγραφή: Τα δεδομένα συλλέγονται για όλες τις παραμέτρους της web εφαρμογής.

ISCX2012

Δημιουργία: University of New Brunswick

Τύποι Επιθέσεων: Dos, DDoS, Brute force, Infiltration

Περιγραφή: Αποτελείται από σενάρια δικτύου με παρεμβατικές δραστηριότητες και εμφανίσεις δεδομένων με ετικέτες

CIC-IDS 2017

Δημιουργία: Canadian Institute of Cyber Security

Τύποι Επιθέσεων: Brute force, Portscan, Botnet, Dos, DDoS, Web, Infiltration

Περιγραφή: Τα προφίλ δικτύου χρησιμοποιούνται για τη δημιουργία του συνόλου δεδομένων με συγκεκριμένο τρόπο.

CSE-CIC-IDS

Δημιουργία: Canadian Institute of Cyber Security

Τύποι Επιθέσεων: Brute force, Portscan, Botnet, Dos, DDoS, Web, Infiltration

Περιγραφή: Ο αριθμός των διπλότυπων δεδομένων είναι πολύ χαμηλός. Δεν υπάρχουν σχεδόν καθόλου αβέβαια στοιχεία. Το σύνολο δεδομένων είναι σε μορφή CSV, οπότε είναι έτοιμο για χρήση χωρίς επεξεργασία.

BOT-IoT

Δημιουργία: School of Engineering and Information Technology, UNSW
Canberra Cyber

Τύποι Επιθέσεων: DoS, DDoS, Reconnaissance, Theft

Περιγραφή: Παρέχει πλήρεις πληροφορίες λήψης πακέτων με αντίστοιχες
ετικέτες

Εν κατακλείδι, το KDD-99 αποτελεί ένα Σύστημα Ανίχνευσης Εισβολών καλύπτει λίγους τύπους επιθέσεων και αποτελείται από μεγάλο αριθμό περιττών και διπλών δειγμάτων δεδομένων. Παρόλα αυτά, με την πάροδο των ετών χρησιμοποιείται ολοένα και περισσότερο σε επιθέσεις σε Cloud, Smart Cities, Internet of Things και σε επιθέσεις Dos-DDos. Αντίστοιχες παρατυπίες παρατηρούνται και στο DARPA-98, ενώ ακόμη και το NSL-KDD, το οποίο αποτελεί την εξελιγμένη μορφή των δύο προαναφερθέντων, απευθύνεται και αυτό σε περιορισμένες επιθέσεις.

Τα πιο σύγχρονα datasets, όπως τα HTTP DATASET CSIC 2010, ISCX2012, UNSW-NB15, CIC IDS-2017 και CSE-CIC-IDS, απευθύνονται σε περισσότερα είδη επιθέσεων, δεν έχουν τόσα διπλότυπα αρχεία όσο τα παλαιότερα datasets και δεν υπάρχουν ή υπάρχουν ελάχιστα αβέβαια στοιχεία. Ως εκ τούτου, συμπεραίνουμε ότι τα πιο σύγχρονα datasets είναι περισσότερο αποτελεσματικά στην Ανίχνευση Εισβολών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξενόγλωσση

- Adnan A, Muhammed A, Abd Ghani AA, Abdullah A, Hakim F. An Intrusion Detection System for the Internet of Things Based on Machine Learning: Review and Challenges. *Symmetry*. 2021; 13(6):1011. <https://doi.org/10.3390/sym13061011>
- Al Mehedi Hasan, M., Nasser, M., & Pal, B. (2013). On the KDD'99 dataset: support vector machine based intrusion detection system (ids) with different kernels. *Int. J. Electron. Commun. Comput. Eng*, 4(4), 1164-1170.
- Al-Daweri, M. S., Zainol Ariffin, K. A., Abdullah, S., & Md. Senan, M. F. E. (2020). An analysis of the KDD99 and UNSW-NB15 datasets for the intrusion detection system. *Symmetry*, 12(10), 1666.
- Aldous, D. (1993). The continuum random tree III. *The Annals of Probability*, 248-289..
- Alharbi, A., Alhaidari, S., & Zohdy, M. (2018). Denial-of-service, probing, user to root (U2R) & remote to user (R2L) attack detection using hidden Markov models. *International Journal of Computer and Information Technology*, 7(05).
- Amjad Rehman Khan, Muhammad Kashif, Rutvij H. Jhaveri, Roshani Raut, Tanzila Saba, Saeed Ali Bahaj, "Deep Learning for Intrusion Detection and Security of Internet of Things (IoT): Current Analysis, Challenges, and Possible Solutions", *Security and Communication Networks*, vol. 2022, Article ID 4016073, 13 pages, 2022. <https://doi.org/10.1155/2022/4016073>
- Bay, S. D., Kibler, D., Pazzani, M. J., & Smyth, P. (2000). The UCI KDD archive of large data sets for data mining research and experimentation. *ACM SIGKDD explorations newsletter*, 2(2), 81-85.
- Besnard, P., & Hanks, S. (2013). Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (1995). arXiv preprint arXiv:1304.3848.
- Breiman, L. (2001). Random Forest. *Machine Learning* (pp 5-32).
- Choudhary, S., & Kesswani, N. (2020). Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 datasets using deep learning in IoT. *Procedia Computer Science*, 167, 1561-1573.

- Douligieris, C., & Mitrokotsa, A. (2004). DDoS attacks and defense mechanisms: classification and state-of-the-art. *Computer networks*, 44(5), 643-666.
- Farhan, R. I., Maolood, A. T., & Hassan, N. (2020). Performance analysis of flow-based attacks detection on CSE-CIC-IDS2018 dataset using deep learning. *Indones. J. Electr. Eng. Comput. Sci*, 20(3), 1413-1418.
- Gharaibeh, M., & Papadopoulos, C. (2014). Darpa-2009 intrusion detection dataset report. *Tech. Rep.*
- John, G. H., & Langley, P. (2013). Estimating continuous distributions in Bayesian classifiers. *arXiv preprint arXiv:1302.4964*.
- Karatas, G., Demir, O & Sahingoz, O. (2020) "Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset". *IEEE Access PP(99):1-1*. DOI:10.1109/ACCESS.2020.2973219
- Karatas, G., Demir, O., & Sahingoz, O. K. (2018, December). Deep learning in intrusion detection systems. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, pp. 113-116.
- Kohavi, R. (1996, August). Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *Kdd* (Vol. 96, pp. 202-207).
- Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Generation Computer Systems*, 100, 779-796.
- Lahre, M. K., Dhar, M. T., Suresh, D., Kashyap, K., & Agrawal, P. (2013). Analyze different approaches for ids using kdd 99 data set. *International Journal on Recent and Innovation Trends in Computing and Communication*, 1(8), 645-651.
- Li, J., Zhang, H., & Wei, Z. (2020). The weighted word2vec paragraph vectors for anomaly detection over HTTP traffic. *IEEE Access*, 8, 141787-141798.
- Lippmann, R., Haines, J. W., Fried, D. J., Korba, J., & Das, K. (2000). The 1999 DARPA off-line intrusion detection evaluation. *Computer networks*, 34(4), 579-595.
- Mahoney, M. V., & Chan, P. K. (2003, September). An analysis of the 1999 DARPA/Lincoln Laboratory evaluation data for network anomaly detection.

- In International Workshop on Recent Advances in Intrusion Detection* (pp. 220-237). Springer, Berlin, Heidelberg.
- McHugh, J. (2000). Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. *ACM Transactions on Information and System Security (TISSEC)*, 3(4), 262-294.
 - Moustafa, N. (2017). Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic (Doctoral dissertation, UNSW Sydney).
 - Moustafa, N., & Slay, J. (2015, November). The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems. *In 2015 4th international workshop on building analysis datasets and gathering experience returns for security (BADGERS)* (pp. 25-31). IEEE.
 - Moustafa, N., & Slay, J. (2015, November). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *In 2015 military communications and information systems conference (MilCIS)* (pp. 1-6).
 - Moustafa, N., & Slay, J. (2016). The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective*, 25(1-3), 18-31.
 - Nawir, M., Amir, A., Lynn, O. B., Yaakob, N., & Ahmad, R. B. (2018, May). Performances of machine learning algorithms for binary classification of network anomaly detection system. *In Journal of Physics: Conference Series* (Vol. 1018, No. 1, p. 012015). IOP Publishing.
 - Ni, S., Qian, Q., & Zhang, R. (2018). Malware identification using visualization images and deep learning. *Computers & Security*, 77, 871-885.
 - Olusola, A. A., Oladele, A. S., & Abosede, D. O. (2010, October). Analysis of KDD'99 intrusion detection dataset for selection of relevance features. *In Proceedings of the world congress on engineering and computer science*, Vol. 1, pp. 20-22. WCECS.
 - Özgür, A., & Erdem, H. (2016). A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015.

- Pelletier, Z. & Abualkibash, M. (2020) “Evaluating the CIC IDS-2017 Dataset Using Machine Learning Methods and Creating Multiple Predictive Models in the Statistical Computing Language R,”. *International Research Journal of Advanced Engineering and Science*, Volume 5, Issue 2, pp. 187-191.
- Pokhrel, S., Abbas, R., & Aryal, B. (2021). IoT security: botnet detection in IoT using machine learning. *arXiv preprint arXiv:2104.02231*.
- Quinlan, J. R. (1993). Program for machine learning. *C4*. 5.
- Revathi, S., & Malathi, A. (2013). A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection. *International Journal of Engineering Research & Technology (IJERT)*, 2(12), 1848-1853.
- Ruck, D. W., Rogers, S. K., Kabrisky, K., Oxley, M. E., & Suter, B. W. (1990). The multilayer perceptron as an approximation to an optimal Bayes estimator. *IEEE Transactions on Neural Networks*, 1(4), 296-298.
- Somani, G., Gaur, M. S., Sanghi, D., Conti, M., & Buyya, R. (2017). DDoS attacks in cloud computing: Issues, taxonomy, and future directions. *Computer Communications*, 107, 30-48.
- Stolfo, S. (2000). Costbased modeling for fraud and intrusion detection: Results from the jam project.
- Tavallaei, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009, July). A detailed analysis of the KDD CUP 99 data set. *In 2009 IEEE symposium on computational intelligence for security and defense applications* (pp. 1-6).
- Thakkar, A., & Lohiya, R. (2020). A review of the advancement in intrusion detection datasets. *Procedia Computer Science*, 167, 636-645.
- Zoghi, Z., & Serpen, G. (2021). Unsw-nb15 computer security dataset: Analysis through visualization. *arXiv preprint arXiv:2101.05067*.
- Zuech, R., Khoshgoftaar, T. M., Seliya, N., Najafabadi, M. M., & Kemp, C. (2015, April). A new intrusion detection benchmarking system. *In The Twenty-Eighth International Flairs Conference*.

ΕΛΛΗΝΙΚΗ

- Stallings, W. (2008). *Βασικές Αρχές Ασφάλειας Δικτύων, Εφαρμογές και Πρότυπα*. Αθήνα: Κλειδάριθμος.
- Αγγελόπουλος, Θ. (2017). *Ανάλυση και δημιουργία εργαλείου ελέγχου ασφάλειας* (Master's thesis, Πανεπιστήμιο Πειραιώς).
- Γκρίτζαλης, Σ., Κάτσικας, Σ. & Γκρίτζαλης Δ. (2003). *Ασφάλεια Δικτύων Υπολογιστών*. Αθήνα: Εκδόσεις Παπασωτηρίου
- Σκαρπάρης, Ι., & Τσαγγάρης, Μ. (2017). *Ανάπτυξη εργαλείου για την έγκαιρη ανίχνευση συμβάντων ασφαλείας σε δεδομένα μεγάλου όγκου* (Doctoral dissertation, Σκαρπάρης Ιωάννης, Τσαγγάρης Μιχάλης).
- Χρήστος, Ι. Μ. (2004). *Δίκτυα Δημόσιας Χρήσης και Διασύνδεση Δικτύων. Πανεπιστημιακές Σημειώσεις*.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΕΙΚΟΝΩΝ

- [1] <https://quizlet.com/435984243/cybersecurity-module-1-flash-cards/>
- [2] Nawir, M., Amir, A., Lynn, O. B., Yaakob, N., & Ahmad, R. B. (2018, May). Performances of machine learning algorithms for binary classification of network anomaly detection system. *In Journal of Physics: Conference Series* (Vol. 1018, No. 1, p. 012015). IOP Publishing.
- [3]Özgür, A., & Erdem, H. (2016). A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015.
- [4] Alothman, Z., Alkasassbeh, M., & Al-Haj Baddar, S. (2020). An efficient approach to detect IoT botnet attacks using machine learning. *Journal of High Speed Networks*, 26(3), 241-254.