ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Εφαρμογή Μεθόδων και Τεχνικών Τεχνητής Νοημοσύνης για την Ασφάλεια Συστημάτων Λογισμικού

ΠΑΝΑΓΙΩΤΑ ΤΡΙΑΝΤΑΦΥΛΛΟΠΟΥΛΟΥ
3210201

ΕΠΙΒΛ. ΚΑΘΗΓΗΤΗΣ : ΝΙΚΟΛΑΟΣ ΔΙΑΜΑΝΤΙΔΗΣ

ΑΘΗΝΑ, ΙΑΝΟΥΑΡΙΟΣ 2025

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου για την πολύτιμη καθοδήγηση και υποστήριξή του καθ' όλη τη διάρκεια της εκπόνησης της πτυχιακής εργασίας. Η γνώση και οι συμβουλές του υπήρξαν καθοριστικές για την ολοκλήρωση αυτής της προσπάθειας.

Επιπλέον, οφείλω να αναγνωρίσω και να εκφράσω τη βαθιά μου ευγνωμοσύνη για όσα έμαθα κατά τη διάρκεια των φοιτητικών μου χρόνων στο Πανεπιστήμιο. Οι γνώσεις που απέκτησα, οι εμπειρίες που συνέλεξα και οι δεξιότητες που ανέπτυξα με βοήθησαν να διαμορφώσω την προσωπική και επαγγελματική μου πορεία. Ειδικότερα, οι πρακτικές γνώσεις και οι θεωρητικές βάσεις που απέκτησα μέσω των μαθημάτων και των εργαστηρίων αποτέλεσαν το θεμέλιο για την επιστημονική μου εξέλιξη και την εκπόνηση αυτής της εργασίας.

Περίληψη

Η παρούσα πτυχιακή εργασία έχει ως θέμα την Εφαρμογή Μεθόδων και Τεχνικών Τεχνητής Νοημοσύνης (AI) στο Πλαίσιο του NIST Cybersecurity Framework (CSF). Σκοπός της ανάλυσης αυτής είναι να διερευνήσει τις σύγχρονες μεθόδους και τεχνικές που χρησιμοποιούνται στον τομέα της κυβερνοασφάλειας για την πρόληψη, την ανίχνευση, την αντίδραση και την αποκατάσταση από κυβερνοεπιθέσεις, με βάση τις αρχές και τις οδηγίες του CSF.

Το NIST Cybersecurity Framework, που έχει αναπτυχθεί από το Εθνικό Ινστιτούτο Πρότυπων και Τεχνολογίας (NIST), αποτελεί ένα σύνολο κατευθυντήριων γραμμών και βέλτιστων πρακτικών που στοχεύουν στην ενίσχυση της κυβερνοασφάλειας οργανισμών κάθε μεγέθους και τομέα. Η συγκεκριμένη εργασία επικεντρώνεται στην ενσωμάτωση τεχνικών τεχνητής νοημοσύνης για την αντιμετώπιση των απειλών και κινδύνων στον κυβερνοχώρο, λαμβάνοντας υπόψη την πρόσφατη ερευνητική δραστηριότητα στον τομέα αυτό.

Η εργασία βασίζεται στη βιβλιογραφική ανασκόπηση που περιγράφεται στο άρθρο των Kaur, Gabrijelčič, και Klobučar, με τίτλο «Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions» [1].

Στην παρούσα πτυχιακή εργασία, η ανάλυση είναι δομημένη σύμφωνα με τις τέσσερις βασικές λειτουργίες του πλαισίου NIST:

1. Προσδιορισμός – Ταυτοποίηση των περιουσιακών στοιχείων και των απειλών.
2. Προστασία – Μέσω εφαρμογής μέτρων άμυνας και πρόληψης.
3. Εντοπισμός – Ανίχνευση κυβερνοεπιθέσεων και απειλών σε πραγματικό χρόνο.
4. Ανταπόκριση – Άμεση αντίδραση και αποκατάσταση μετά από επιθέσεις.

Η επιλογή αυτής της δομής επιτρέπει την ολιστική προσέγγιση του θέματος και αναδεικνύει τη σημασία κάθε λειτουργίας στο πλαίσιο της κυβερνοασφάλειας. Ιδιαίτερη έμφαση δίνεται στην εφαρμογή μεθόδων AI για τη βελτίωση της αποτελεσματικότητας κάθε λειτουργίας, όπως τεχνικές μηχανικής μάθησης, βαθιάς μάθησης και ανάλυσης δεδομένων για την πρόληψη και αντιμετώπιση απειλών.

ΛΕΞΕΙΣ – ΚΛΕΙΔΙΑ : Τεχνητή Νοημοσύνη, NIST Cybersecurity Framework, Προσδιορισμός, Προστασία, Εντοπισμός, Ανταπόκριση

Abstract

This thesis focuses on the Application of Artificial Intelligence (AI) Methods and Techniques within the NIST Cybersecurity Framework (CSF). The objective of this analysis is to explore modern methods and techniques used in the field of cybersecurity for the prevention, detection, response, and recovery from cyberattacks, based on the principles and guidelines of the CSF.

The NIST Cybersecurity Framework, developed by the National Institute of Standards and Technology (NIST), is a set of guidelines and best practices aimed at enhancing the cybersecurity posture of organizations of all sizes and industries. This study focuses on integrating artificial intelligence techniques to address threats and risks in cyberspace, considering recent research activity in this domain.

The thesis is based on the literature review presented in the article by Kaur, Gabrijelčič, and Klobučar, titled "Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions" [1].

In this thesis, the analysis is structured according to the four core functions of the NIST Cybersecurity Framework:

1. Identify – Identification of assets and threats.
2. Protect – Implementation of defense and prevention measures.
3. Detect – Detection of cyberattacks and threats in real time.
4. Respond – Immediate response and recovery following attacks.

This structure allows for a holistic approach to the topic and highlights the importance of each function in the context of cybersecurity. Special emphasis is placed on the application of AI methods to improve the effectiveness of each function, including machine learning, deep learning, and data analytics techniques for the prevention and mitigation of threats.

KEYWORDS : Artificial Intelligence, NIST Cybersecurity Framework, Identify, Protect, Detect, Respond

Πίνακας Περιεχομένων

1. Εισαγωγή

Το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (NIST - National Institute of Standards and Technology) είναι ένας οργανισμός των Ηνωμένων Πολιτειών που ανήκει στο Υπουργείο Εμπορίου και ειδικεύεται στην ανάπτυξη προτύπων, μετρήσεων και τεχνολογικών κατευθυντήριων γραμμών. Ιδρύθηκε το 1901 και από τότε έχει εξελιχθεί σε έναν από τους κορυφαίους οργανισμούς στον τομέα της επιστημονικής έρευνας και της τεχνολογικής καινοτομίας.

Το NIST έχει ως αποστολή την προώθηση της καινοτομίας και της βιομηχανικής ανταγωνιστικότητας των ΗΠΑ μέσω της επιστήμης, της μηχανικής και της τεχνολογίας. Οι κύριες δραστηριότητές του περιλαμβάνουν την ανάπτυξη τεχνικών προτύπων για διάφορες βιομηχανίες, όπως η κυβερνοασφάλεια, η βιοτεχνολογία, η κατασκευή υλικών και η τεχνητή νοημοσύνη. Επιπλέον, διεξάγει έρευνα στη μετρολογία, συμβάλλοντας στην ακριβή μέτρηση φυσικών και τεχνικών ποσοτήτων. Παράλληλα, δημιουργεί οδηγούς και πλαίσια πολιτικής που ενισχύουν την ασφάλεια στον κυβερνοχώρο, τη διαχείριση κινδύνων και την υιοθέτηση καινοτόμων τεχνολογιών, βοηθώντας τους οργανισμούς να προσαρμοστούν στις σύγχρονες προκλήσεις.

Το NIST είναι ιδιαίτερα γνωστό για την προσφορά οδηγιών και πλαισίων που σχετίζονται με την ασφάλεια στον κυβερνοχώρο, βοηθώντας οργανισμούς να ενισχύσουν τις υποδομές τους και να προστατευτούν από κυβερνοαπειλές. Μερικά από τα πιο σημαντικά πλαίσια και πρότυπα που έχει αναπτύξει είναι:

- NIST Cybersecurity Framework (CSF)
- NIST Special Publication 800 Series
- NIST Risk Management Framework (RMF)
- NIST AI Risk Management Framework (AI RMF)

Το Πλαίσιο Κυβερνοασφάλειας (Cybersecurity Framework - CSF) του Εθνικού Ινστιτούτου Προτύπων και Τεχνολογίας (NIST), με το οποίο θα ασχοληθούμε, αποτελεί ένα εργαλείο που σχεδιάστηκε για να βοηθήσει οργανισμούς κάθε μεγέθους και τομέα — συμπεριλαμβανομένων της βιομηχανίας, της κυβέρνησης, της εκπαίδευσης και των μη κερδοσκοπικών οργανισμών — να διαχειρίζονται και να μειώνουν τους κινδύνους κυβερνοασφάλειας. Από την αρχική του κυκλοφορία το 2014, έχει χρησιμοποιηθεί σε περισσότερες από 185 χώρες και έχει αποδειχθεί ευέλικτο και χρήσιμο ανεξαρτήτως της τεχνικής ωριμότητας ή των πόρων ενός οργανισμού.

Η νέα έκδοση CSF 2.0, που δημοσιεύτηκε επίσημα στις αρχές του 2024, φέρνει σημαντικές βελτιώσεις για την προσαρμογή στις σύγχρονες προκλήσεις της κυβερνοασφάλειας. Ακολουθεί τις αλλαγές στην τεχνολογία, τη νομοθεσία και τις απειλές, ενώ παραμένει ευέλικτο και εθελοντικό, ώστε να ανταποκρίνεται στις μοναδικές ανάγκες κάθε οργανισμού.

Εικόνα 1 : Τα πέντε βασικά πεδία του Πλαισίου Κυβερνοασφάλειας NIST – Διακυβέρνηση, Προσδιορισμός, Προστασία, Εντοπισμός, Ανταπόκριση και Ανάκαμψη – τα οποία συμβάλλουν στην ολιστική διαχείριση κινδύνων και στην ενίσχυση της ανθεκτικότητας των οργανισμών απέναντι σε κυβερνοαπειλές.

Ο Πυρήνας του CSF (Core) οργανώνει τα αποτελέσματα της κυβερνοασφάλειας σε 6 βασικές λειτουργίες, οι οποίες συνδέονται μεταξύ τους και πρέπει να εφαρμόζονται συνεχώς:

1. **ΔΙΑΚΥΒΕΡΝΗΣΗ-GOVERN(GV)**
   Καθορίζει τη στρατηγική και την πολιτική διακυβέρνησης κινδύνων κυβερνοασφάλειας, ενσωματώνοντάς την στη συνολική στρατηγική διαχείρισης κινδύνων του οργανισμού. Περιλαμβάνει τη διαχείριση κινδύνων προμηθευτικής αλυσίδας, τις αρμοδιότητες και την παρακολούθηση της στρατηγικής.

2. **ΠΡΟΣΔΙΟΡΙΣΜΟΣ-IDENTIFY(ID)**
   Αναγνωρίζει τους πόρους (δεδομένα, συστήματα, προμηθευτές) και τους σχετικούς κινδύνους, ώστε να προτεραιοποιούνται οι δράσεις κυβερνοασφάλειας. Περιλαμβάνει την αξιολόγηση πολιτικών, διαδικασιών και πρακτικών για τη συνεχή βελτίωση.

3. **ΠΡΟΣΤΑΣΙΑ-PROTECT(PR)**
   Εστιάζει στην προστασία των περιουσιακών στοιχείων μέσω μέτρων, όπως διαχείριση ταυτότητας, έλεγχος πρόσβασης, εκπαίδευση, ασφάλεια δεδομένων και ανθεκτικότητα των τεχνολογικών υποδομών.

4. **ΕΝΤΟΠΙΣΜΟΣ-DETECT(DE)**
   Επιτρέπει τον εντοπισμό και την ανάλυση πιθανών κυβερνοεπιθέσεων μέσω έγκαιρης ανίχνευσης ανωμαλιών, δεικτών παραβίασης και περιστατικών.

5. **ΑΝΤΑΠΟΚΡΙΣΗ-RESPOND(RS)**
   Διαχειρίζεται περιστατικά κυβερνοασφάλειας για τον περιορισμό των επιπτώσεών τους, καλύπτοντας τη διαχείριση περιστατικών, την ανάλυση, τη μετρίαση και την επικοινωνία.

6. **ΑΝΑΚΑΜΨΗ-RECOVER(RC)**
   Υποστηρίζει την αποκατάσταση των λειτουργιών και των πόρων που επηρεάστηκαν από περιστατικά, διασφαλίζοντας τη συνέχιση των δραστηριοτήτων και την κατάλληλη επικοινωνία κατά τη φάση της αποκατάστασης.

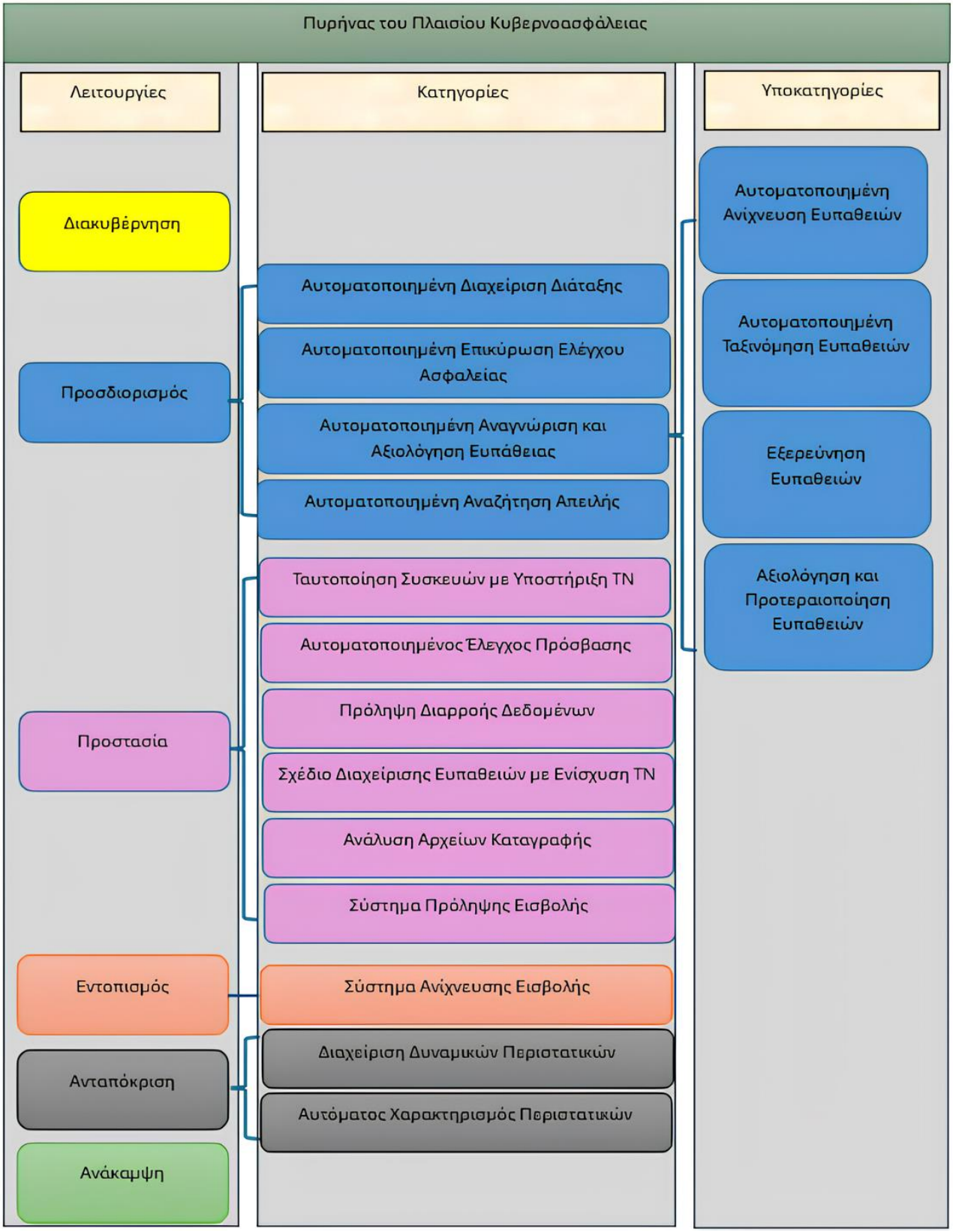
Στα πλαίσια της συγκεκριμένης εργασίας, θα επικεντρωθούμε στις τέσσερις βασικές λειτουργίες του NIST Cybersecurity Framework: Προσδιορισμός, Προστασία, Εντοπισμός και Ανταπόκριση, παρουσιάζοντας μεθόδους, τεχνικές και αλγορίθμους τεχνητής νοημοσύνης (AI) που μπορούν να υποστηρίξουν κάθε λειτουργία.

Στη λειτουργία «Προσδιορισμός», θα αναλύσουμε πώς τεχνικές ανάλυσης δεδομένων και αλγόριθμοι μηχανικής μάθησης μπορούν να βοηθήσουν στην κατηγοριοποίηση πόρων, την αξιολόγηση κινδύνων και την κατανόηση της επιφάνειας επίθεσης.

Στη λειτουργία «Προστασία», θα εστιάσουμε σε αλγορίθμους πρόληψης, όπως συστήματα ανίχνευσης ανωμαλιών και τεχνολογίες ενισχυμένης κρυπτογράφησης για την προστασία δεδομένων και υποδομών.

Στη λειτουργία «Εντοπισμός», θα παρουσιάσουμε τεχνικές ανίχνευσης κυβερνοαπειλών βασισμένες σε βαθιά νευρωνικά δίκτυα (deep learning), που επιτρέπουν την έγκαιρη αναγνώριση ύποπτων μοτίβων ή ανωμαλιών.

Τέλος, στη λειτουργία «Ανταπόκριση», θα εξετάσουμε πώς η τεχνητή νοημοσύνη μπορεί να αυτοματοποιήσει τη διαδικασία ανάλυσης περιστατικών και τη λήψη αποφάσεων για τον περιορισμό των επιπτώσεων, χρησιμοποιώντας αλγορίθμους ενισχυτικής μάθησης (reinforcement learning). Αυτή η προσέγγιση συνδυάζει την τεχνολογία AI με την κυβερνοασφάλεια, παρέχοντας ισχυρές λύσεις για την αντιμετώπιση των σύγχρονων προκλήσεων. [197] [198] [199]



Πίνακας 1 : Δομή του Πυρήνα του Πλαισίου Κυβερνοασφάλειας, που περιλαμβάνει τις βασικές λειτουργίες, κατηγορίες και υποκατηγορίες για την οργάνωση και διαχείριση της ασφάλειας πληροφοριών, σύμφωνα με το NIST Cybersecurity Framework.

2. Προσδιορισμός

*Η λειτουργία **«Προσδιορισμός»** (**Identify)** του πλαισίου NIST Cybersecurity Framework αποτελεί το πρώτο και θεμελιώδες βήμα για την αποτελεσματική διαχείριση των κινδύνων κυβερνοασφάλειας ενός οργανισμού. Ο βασικός της στόχος είναι η ανάπτυξη μιας οργανωσιακής κατανόησης που επιτρέπει την ταυτοποίηση και τη διαχείριση κινδύνων που σχετίζονται με συστήματα, δεδομένα, περιουσιακά στοιχεία και τις επιχειρησιακές δυνατότητες. Αυτή η κατανόηση επιτρέπει στον οργανισμό να συνδέσει τους επιχειρηματικούς στόχους του με τις απειλές κυβερνοασφάλειας, προσαρμόζοντας την προσέγγισή του ανάλογα με τους διαθέσιμους πόρους και τις ιδιαίτερες ανάγκες του.*

2.1 Αυτοματοποιημένη Διαχείριση Διάταξης

Η αυτοματοποιημένη διαχείριση διάταξης (Automated Configuration Management) είναι μια διαδικασία που διασφαλίζει ότι οι ρυθμίσεις ενός συστήματος καθορίζονται και διατηρούνται αυτόματα σε μια ελεγχόμενη και εξουσιοδοτημένη βάση. Ο στόχος της είναι να μειώσει τα λάθη που προκύπτουν από χειροκίνητες ρυθμίσεις και να εξασφαλίσει την απόδοση και την ασφάλεια του συστήματος. Με την αυτοματοποίηση, το σύστημα μπορεί να προσαρμόζεται δυναμικά, αποφεύγοντας προβλήματα που συνήθως προκύπτουν από ανθρώπινα λάθη ή μη βέλτιστες ρυθμίσεις.

Μέσω της ανασκόπησης διαφόρων επιστημονικών πηγών, θα περιγράψουμε καινοτόμες προσεγγίσεις που προτείνονται από διάφορους ερευνητές, όπως οι αλγόριθμοι μηχανικής μάθησης [2] και οι γενετικοί αλγόριθμοι [3]. Οι αλγόριθμοι αυτοί βελτιώνουν, όπως θα διαπιστώσουμε, την ευελιξία και την ανθεκτικότητα των διαδικτυακών πλατφόρμων σε κακόβουλες επιθέσεις.

Οι ερευνητές Bentz Tozer, Thomas Mazzuchi και Shahram Sarkani από το Πανεπιστήμιο George Washington [4] διεξήγαγαν έρευνα με στόχο την ενίσχυση της ασφάλειας των συστημάτων μέσω της διαχείρισης και ελαχιστοποίησης της «επιφάνειας επίθεσης». Συγκεκριμένα αναφέρονται στα πιθανά σημεία εισβολής που θα μπορούσε να εκμεταλλευτεί ένας επιτιθέμενος. Η ερευνητική τους στρατηγική στηρίζεται στη χρήση της πολυαντικειμενικής ενισχυτικής μάθησης (multi-objective reinforcement learning) για να βελτιστοποιηθεί ταυτόχρονα η επίθεση και η διαμόρφωση των συστημάτων.

Οι επιστήμονες θέτουν ως κύριο στόχο την ανάπτυξη ενός ευέλικτου και ασφαλούς συστήματος που υποστηρίζει διαφορετικές διαμορφώσεις και επιτρέπει την προσαρμογή κατά τη λειτουργία του (runtime), χωρίς να επηρεάζεται η συνολική του απόδοση. Αυτό το επιτυγχάνουν εφαρμόζοντας αρχιτεκτονική μικροϋπηρεσιών (microservices), η οποία επιτρέπει τη δημιουργία ανεξάρτητων λειτουργικών μονάδων με ελεγχόμενες επιφάνειες επίθεσης. Με αυτό τον τρόπο, είναι δυνατή η επιλογή διαφορετικών συνδυασμών υπηρεσιών και ρυθμίσεων για την ελαχιστοποίηση του κινδύνου επίθεσης.

Η μεθοδολογία της έρευνας περιλαμβάνει την ανάπτυξη ενός δυναμικού συστήματος βασισμένου σε πολυαντικειμενική ενισχυτική μάθηση, ώστε να προκύψουν οι βέλτιστες στρατηγικές ασφαλείας. Οι μέθοδοι που εφαρμόστηκαν είναι οι εξής:

1. **Αλγόριθμος Πολυαντικειμενικής Q-Μάθησης (Pareto Q-learning)**

   Αποτελεί έναν αλγόριθμο ενισχυτικής μάθησης εκτός πολιτικής (off-policy), όπου οι πράξεις επιλέγονται με βάση την Pareto κυριαρχία (Pareto dominance), προσφέροντας βέλτιστες λύσεις σε διαφορετικούς αντικειμενικούς στόχους ταυτόχρονα.

2. **Πολυαντικειμενική SARSA (Pareto SARSA)**

   Παρόμοια με το Q-learning, αλλά εφαρμόζει προσέγγιση «εντός πολιτικής» (on-policy), δηλαδή εκπαιδεύεται με βάση τις δράσεις που πραγματικά εκτελούνται, και όχι τις βέλτιστες θεωρητικά.

3. **Πολυαντικειμενική TD(0) με Μετά-καταστάσεις (Pareto TD(0) Afterstate Algorithm)**

   Χρησιμοποιεί την έννοια των μετά-καταστάσεων (afterstates) για την ανάλυση της κατάστασης μετά από μία δράση αλλά πριν την πλήρη ανταπόκριση του συστήματος, βελτιώνοντας την αποδοτικότητα εκμάθησης.

Η προσέγγιση αυτή επιτρέπει τη δυναμική και στοχευμένη επιλογή των βέλτιστων διαμορφώσεων που μειώνουν τις πιθανές ευπάθειες, ενώ διατηρούν την απαραίτητη λειτουργικότητα και επεκτείνουν τη διαφορετικότητα διαμόρφωσης του συστήματος, αποτρέποντας μελλοντικές επιθέσεις.

Στο κομμάτι της ασφάλειας, της αξιοπιστίας και της βελτιστοποίηση των δεδομένων σε κατανεμημένα συστήματα, εστίασαν οι επιστήμονες που συμμετείχαν στις μελέτες [5] σχετικά με την αποθήκευση δεδομένων σε περιβάλλοντα πολλαπλών νεφών [6]. Το σκεπτικό τους και οι στόχοι τους επικεντρώνονται στη βελτίωση της αποδοτικότητας και της εμπιστοσύνης των χρηστών σε πολυ-νεφικές αρχιτεκτονικές, μέσω διασφάλισης της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των δεδομένων.

Οι αλγόριθμοι που χρησιμοποιήθηκαν στην ανάλυση της αποθήκευσης σε πολυ-νεφικά περιβάλλοντα ήταν οι MOCell, NSGA-II και SPEA2, οι οποίοι ανήκουν στην κατηγορία των πολυ-αντικειμενικών εξελικτικών αλγορίθμων (Multi-Objective Evolutionary Algorithms - MOEAs) [7]. Αυτές οι μέθοδοι είναι κατάλληλες για την επίλυση προβλημάτων που απαιτούν τη βελτιστοποίηση πολλών και αντικρουόμενων στόχων, όπως η ασφάλεια, ο πλεονασμός και ο χρόνος ανάκτησης δεδομένων.

1. **MOCell (Multi-Objective Cellular Genetic Algorithm)**

   Ο MOCell είναι μια παραλλαγή του γενετικού αλγορίθμου που χρησιμοποιεί κυψελωτή δομή για τη διάδοση των λύσεων. Σε αυτή τη δομή, κάθε λύση μπορεί να ανταλλάξει πληροφορίες μόνο με τους γειτονικούς της κόμβους. Αυτή η προσέγγιση βοηθά στη διατήρηση της ποικιλότητας του πληθυσμού και βελτιώνει την εξερεύνηση του χώρου λύσεων. Ο MOCell έχει αποδειχθεί ότι προσφέρει εξαιρετική ισορροπία μεταξύ σύγκλισης και διατήρησης της ποικιλότητας.

2. **NSGA-II (Non-dominated Sorting Genetic Algorithm II)**

   Ο NSGA-II είναι ένας από τους πιο γνωστούς και ευρέως χρησιμοποιούμενους αλγορίθμους για πολυ-αντικειμενική βελτιστοποίηση. Χρησιμοποιεί μη κυρίαρχη ταξινόμηση (non-dominated sorting) για να οργανώσει τις λύσεις και ενσωματώνει μια διαδικασία αναστροφής (elitism) για να διατηρεί τις καλύτερες λύσεις κατά τη διάρκεια των γενεών. Διαθέτει επίσης μηχανισμό συμπίεσης πλήθους (crowding distance) για τη διατήρηση της ποικιλότητας του πληθυσμού, επιτρέποντας έτσι την εύρεση λύσεων που καλύπτουν μεγάλο εύρος της Παρέτο βέλτιστης επιφάνειας.

3. **SPEA2 (Strength Pareto Evolutionary Algorithm 2)**

Ο SPEA2 βελτιώνει την αρχική έκδοση του SPEA προσθέτοντας μηχανισμούς αξιολόγησης των λύσεων βάσει της δύναμης (strength), η οποία αντιπροσωπεύει τον αριθμό των λύσεων που κυριαρχούνται από μια συγκεκριμένη λύση. Επιπλέον, ο SPEA2 διαθέτει στρατηγικές για να εξασφαλίζει μια ομοιόμορφη κατανομή των λύσεων στην Παρέτο επιφάνεια και διατηρεί ένα εξωτερικό αρχείο που αποθηκεύει τις καλύτερες λύσεις κατά τη διάρκεια της εκτέλεσης του αλγορίθμου.

Οι αλγόριθμοι αυτοί χρησιμοποιήθηκαν για την πολυ-αντικειμενική βελτιστοποίηση των παραμέτρων αποθήκευσης σε συστήματα πολλαπλών νεφών, όπως το threshold scheme (k, n), για να επιτευχθεί η βέλτιστη ισορροπία μεταξύ ασφάλειας, αξιοπιστίας και αποτελεσματικότητας. Με τη χρήση αυτών των αλγορίθμων, οι ερευνητές κατάφεραν να βρουν λύσεις που ανταποκρίνονται στις ανάγκες διαφορετικών περιβαλλόντων και προτιμήσεων του χρήστη, διασφαλίζοντας παράλληλα τη μέγιστη δυνατή ασφάλεια και απόδοση του συστήματος.

Οι επιστήμονες Mehrbod Sharifi, Eugene Fink, και Jaime G. Carbonell από το Πανεπιστήμιο Carnegie Mellon επικεντρώθηκαν στην ανάπτυξη ενός αυτοματοποιημένου βοηθού που προσαρμόζει τις ρυθμίσεις ασφαλείας με βάση τις ανάγκες των χρηστών [8]. Η πρωτοβουλία τους προέκυψε από το πρόβλημα που αντιμετωπίζουν πολλοί μη ειδικοί χρήστες να προσαρμόσουν αποτελεσματικά τα διαθέσιμα εργαλεία κυβερνοασφάλειας, κάτι που συχνά οδηγεί σε υπερβολική ή ανεπαρκή προστασία.

Το σκεπτικό των επιστημόνων ήταν να μειώσουν την πολυπλοκότητα των επιλογών ασφαλείας και να βοηθήσουν τους χρήστες να κάνουν ασφαλέστερες επιλογές χωρίς να χρειάζεται να κατανοήσουν λεπτομέρειες για τις τεχνικές ρυθμίσεις. Στόχος τους ήταν η δημιουργία ενός συστήματος που μαθαίνει τις συνήθειες και τις ανάγκες του χρήστη μέσα από την παρατήρηση και απαντήσεις σε στοχευμένες ερωτήσεις. Αυτό επιτρέπει στο σύστημα να παρέχει εξατομικευμένες συστάσεις και να βελτιώνει την ασφάλεια με τρόπο φιλικό προς το χρήστη.

Οι επιστήμονες χρησιμοποίησαν μεθόδους που περιλαμβάνουν:

1. **Μοντελοποίηση Χρήστη**: Δημιουργία μοντέλων που περιγράφουν την τεχνική κατάρτιση του χρήστη, τις προτιμήσεις του, και τις συνήθειες χρήσης.
2. **Εκμάθηση μέσω Παρατήρησης και Ερωτήσεων**: Το σύστημα συλλέγει δεδομένα παρατηρώντας τις ενέργειες του χρήστη και ζητώντας απαντήσεις σε στοχευμένες ερωτήσεις που προσαρμόζονται στο επίπεδο γνώσεων του χρήστη.
3. **Λήψη Αποφάσεων με Αβεβαιότητα**: Χρήση πιθανοτήτων για την εκπροσώπηση της αβεβαιότητας στις ιδιότητες των μοντέλων, επιτρέποντας στο σύστημα να λαμβάνει αποφάσεις όταν δεν υπάρχουν επαρκή δεδομένα.

Ο βοηθός συνδυάζει παθητική παρατήρηση με ενεργές ερωτήσεις για να προσαρμόσει τις ρυθμίσεις ασφαλείας και να εξηγήσει στον χρήστη τις διαθέσιμες επιλογές με τρόπο που να είναι κατανοητός και προσαρμοσμένος στις τεχνικές του γνώσεις.

Σε επόμενη μελέτη [9] παρουσιάζεται το πλαίσιο **VEREFOO**, το οποίο αποσκοπεί στη βελτιστοποίηση και αυτοματοποίηση της διαχείρισης δικτυακών υπηρεσιών (Network Service Functions – NSFs) [10] σε περιβάλλοντα NFV και cloud μέσω ενορχηστρωτών υπηρεσιών, όπως το Open Baton και το Kubernetes.

Οι επιστήμονες επιδίωξαν να δημιουργήσουν ένα πλαίσιο που να ενσωματώνεται εύκολα με υπάρχοντες ενορχηστρωτές υπηρεσιών για να ενισχύσουν τις δυνατότητες επαλήθευσης και επικύρωσης των διαγραμμάτων υπηρεσιών (Service Graphs). Δεδομένου ότι η ενορχήστρωση δικτυακών υπηρεσιών συχνά περιλαμβάνει περίπλοκες διαδικασίες με πιθανά σφάλματα διαμόρφωσης, το VEREFOO [11] προσφέρει αυτοματοποίηση και βελτιστοποίηση με τη χρήση τεχνικών τεχνητής νοημοσύνης, επιτρέποντας την έγκαιρη ανίχνευση και διόρθωση προβλημάτων.

Οι κύριοι στόχοι των επιστημόνων με την ανάπτυξη του VEREFOO περιλαμβάνουν την επέκταση των δυνατοτήτων των ενορχηστρωτών υπηρεσιών μέσω της προσθήκης λειτουργιών επαλήθευσης και επικύρωσης για διαγράμματα υπηρεσιών, εξασφαλίζοντας ότι αυτά συμμορφώνονται με τις προκαθορισμένες πολιτικές ασφαλείας και μπορούν να βελτιστοποιηθούν. Η αυτόματη επικύρωση και βελτιστοποίηση του VEREFOO αποσκοπεί στην αποφυγή λαθών και την απλοποίηση των διαδικασιών διαχείρισης, μειώνοντας την ανάγκη για χειροκίνητες επεμβάσεις. Επιπλέον, η ενσωμάτωση του εργαλείου γίνεται με διαφάνεια για τον χρήστη, διατηρώντας τη ροή εργασίας του ανεπηρέαστη και προσφέροντας πλήρη συμβατότητα με υπάρχοντα εργαλεία, όπως το Open Baton.

Για την επίτευξη των στόχων τους, οι επιστήμονες αξιοποίησαν συγκεκριμένες μεθόδους Τεχνητής Νοημοσύνης (AI) και προηγμένες διαδικασίες.

1. **Επίλυση προβλημάτων MaxSMT** [12]: Το VEREFOO στηρίζεται στην επίλυση προβλημάτων Maximum Satisfiability (MaxSMT), μια τεχνική που ανήκει στην κατηγορία NP-complete. Αυτή η μέθοδος επιτρέπει τη βελτιστοποίηση των διαμορφώσεων των Network Security Functions (NSFs), ικανοποιώντας ταυτόχρονα τις απαιτήσεις ασφαλείας και επιτυγχάνοντας βέλτιστες λύσεις σε σύνθετα σενάρια δικτύων.
2. **Αυτόματη διαμόρφωση**: Το εργαλείο προσφέρει τη δυνατότητα αυτόματης ρύθμισης στοιχείων, όπως οι κανόνες των firewalls, ώστε να διασφαλίζεται η συμμόρφωση με τις απαιτήσεις ασφαλείας του χρήστη. Αυτή η διαδικασία μειώνει την ανάγκη χειροκίνητων επεμβάσεων και αυξάνει την αποδοτικότητα.

Οι επιστήμονες χρησιμοποίησαν διαδικασίες οι οποίες είναι οι εξής:

1. **Ενσωμάτωση με Open Baton**: Το VEREFOO λειτουργεί ως μεσολαβητής (proxy) για την επαλήθευση και επικύρωση των Service Graphs. Ελέγχει τις δομές για συμμόρφωση και εισάγει βελτιστοποιημένες ρυθμίσεις πριν από την τελική μεταφόρτωση στον ενορχηστρωτή Open Baton, προσφέροντας μια εντελώς συμβατή διεπαφή για τον χρήστη.
2. **Ενσωμάτωση με Kubernetes**: Η ενσωμάτωση με τον ενορχηστρωτή Kubernetes περιλαμβάνει τη δυναμική παρακολούθηση και αυτόματη διαμόρφωση κανόνων ασφαλείας μέσω ειδικών hooks. Αυτά τα hooks αντιδρούν σε γεγονότα διαχείρισης και επικαιροποιούν τις πολιτικές ασφαλείας, επιτρέποντας την προσαρμογή του δικτύου σε πραγματικό χρόνο και την άμεση ανταπόκριση σε απειλές.

Αυτές οι μεθοδολογίες συνδυάζουν τις δυνατότητες της τεχνητής νοημοσύνης με τις ανάγκες της σύγχρονης διαχείρισης δικτύων, εξασφαλίζοντας υψηλή απόδοση, ασφάλεια και αυτοματοποίηση.

Στην μελέτη [13] με τίτλο "AMADEUS: Towards the AutoMAteD secUrity teSting" έχει διεξαχθεί από τους επιστήμονες Ángel Jesús Varela-Vaca, Rafael M. Gasca, José Antonio Carmona-Fombella και María Teresa Gómez-López από το Πανεπιστήμιο της Σεβίλλης στην Ισπανία. Στην συγκεκριμένη μελέτη, οι ερευνητές αναγνωρίζουν ότι η σωστή ρύθμιση συστημάτων είναι κρίσιμη για τη μείωση των κινδύνων στον τομέα της κυβερνοασφάλειας. Ο μεγάλος αριθμός ευπαθειών και των παραμέτρων ρυθμίσεων που ενδέχεται να απειληθούν καθιστά την ανάλυση αυτών των ευπαθειών

απαραίτητη. Οι υπάρχουσες μέθοδοι ανάλυσης είναι συχνά χρονοβόρες και απαιτούν εμπειρία, γι' αυτό προτείνουν μια αυτόματη προσέγγιση μέσω του εργαλείου **AMADEUS**, το οποίο χρησιμοποιεί τεχνικές γραμμών προϊόντων λογισμικού.

Οι κύριοι στόχοι του AMADEUS εστιάζονται στην ενίσχυση της κυβερνοασφάλειας μέσω της αυτοματοποίησης και της αναλυτικής διαδικασίας. Πρώτον, επιδιώκει την αυτοματοποίηση της ανάλυσης ευπαθειών, διευκολύνοντας την αναγνώριση των ευπαθειών σε ρυθμίσεις συστημάτων. Αυτό έχει ως στόχο τη μείωση του απαιτούμενου χρόνου και της ανθρώπινης παρέμβασης, καθιστώντας τη διαδικασία πιο αποτελεσματική. Δεύτερον, το AMADEUS στοχεύει στη δημιουργία μοντέλων χαρακτηριστικών, χρησιμοποιώντας πληροφορίες από βάσεις δεδομένων ευπαθειών. Αυτά τα μοντέλα έχουν σκοπό να απεικονίσουν τις ρυθμίσεις των συστημάτων και τις σχετικές ευπάθειες, προσφέροντας μια πιο ολοκληρωμένη εικόνα της ασφάλειας. Τέλος, το εργαλείο επιδιώκει να διενεργήσει λογική ανάλυση για τον εντοπισμό πιθανών μονοπατιών επίθεσης και την ανάδειξη της αλληλεπίδρασης μεταξύ ευπαθειών και ρυθμίσεων. Με αυτούς τους στόχους, το AMADEUS προορίζεται να προσφέρει μια καινοτόμο προσέγγιση στην ανάλυση και τη διαχείριση των κινδύνων στον τομέα της κυβερνοασφάλειας.

Η μέθοδος που χρησιμοποιείται στο AMADEUS περιλαμβάνει τα εξής βήματα:

1. **Συλλογή Δεδομένων**: Ανάλυση δεδομένων από δημόσιες βάσεις δεδομένων ευπαθειών (π.χ. NVD και CVE) για να εντοπιστούν οι ευπάθειες που σχετίζονται με τις ρυθμίσεις των συστημάτων.
2. **Ανάλυση Ευπαθειών**: Εξαγωγή χαρακτηριστικών σχετικών με τις ευπάθειες, όπως τύποι ευπαθειών και μηχανισμοί εκμετάλλευσης.
3. **Δημιουργία Μοντέλου Χαρακτηριστικών**: Δημιουργία ενός μοντέλου που απεικονίζει τις σχέσεις μεταξύ των χαρακτηριστικών και των ευπαθειών, με σκοπό την οπτικοποίηση των κινδύνων.
4. **Λογική Ανάλυση και Εξαγωγή Γνώσης**: Χρήση λογικών αναλύσεων για τον προσδιορισμό πιθανών μονοπατιών επίθεσης και εξαγωγή γνώσεων σχετικά με τους κινδύνους των συστημάτων.

Το AMADEUS, με τη χρήση μοντέλων χαρακτηριστικών και λογικών αναλύσεων, επιδιώκει να προσφέρει μια ολοκληρωμένη προσέγγιση για την αυτόματη ανάλυση και δοκιμή ευπαθειών, συμβάλλοντας στην ενίσχυση της κυβερνοασφάλειας των οργανισμών.

Οι επιστήμονες που εργάστηκαν στο **CyberSPL** [14] είχαν ως βασικό κίνητρο την αντιμετώπιση της πολύπλοκης φύσης των ρυθμίσεων κυβερνοασφάλειας σε διαφορετικά συστήματα και προϊόντα. Επιδίωξαν να δημιουργήσουν ένα εργαλείο που θα επιτρέπει την εύκολη παρακολούθηση και συμμόρφωση με τις πολιτικές κυβερνοασφάλειας μέσω αυτοματοποιημένων διαδικασιών. Με την υλοποίηση του CyberSPL, επιθυμούσαν να διευκολύνουν τους υπεύθυνους κυβερνοασφάλειας να εντοπίζουν και να διαχειρίζονται πιθανές αποκλίσεις από τις πολιτικές με μεγαλύτερη ακρίβεια και αποτελεσματικότητα.

Οι κύριοι στόχοι του έργου CyberSPL περιελάμβαναν τη δημιουργία μιας πλατφόρμας που θα παρέχει εργαλεία για την ανάπτυξη και συντήρηση καταλόγων πολιτικών κυβερνοασφάλειας με βάση μοντέλα χαρακτηριστικών. Ένα άλλο σημαντικό σημείο ήταν η αυτοματοποιημένη ανίχνευση και διάγνωση σφαλμάτων στις ρυθμίσεις των συστημάτων ώστε να ελέγχεται η συμμόρφωση με τις πολιτικές ασφάλειας. Επιπλέον, στόχευαν να καταστήσουν δυνατή την ανάλυση των ρυθμίσεων και την αποθήκευση του ιστορικού για την καλύτερη κατανόηση της εξέλιξης των πολιτικών.

Οι μέθοδοι τεχνητής νοημοσύνης που χρησιμοποιήθηκαν στο πλαίσιο του CyberSPL περιλαμβάνουν την εφαρμογή προηγμένων αλγορίθμων επίλυσης προβλημάτων περιορισμών. Το σύστημα ενσωματώνει το ChocoSolver [15], έναν ειδικό μηχανισμό συλλογιστικής που επιτρέπει την ανάλυση και επαλήθευση των χαρακτηριστικών μοντέλων. Η χρήση του επιτρέπει τον αυτόματο εντοπισμό σφαλμάτων και τη διάγνωση αποκλίσεων σε ρυθμίσεις, διασφαλίζοντας τη συμμόρφωση με τις πολιτικές κυβερνοασφάλειας. Η λειτουργία αυτή παρέχει δυνατότητες ανάλυσης σύνθετων συνδυασμών χαρακτηριστικών και υποστηρίζει τη διαδικασία ανίχνευσης ατελειών μέσω αυτόματων διαγνωστικών διαδικασιών.

Μετά την προηγούμενη προσέγγιση, ακολουθεί η μελέτη [16] που εστιάζει στη χρήση μηχανικής μάθησης για την πρόβλεψη περιστατικών κυβερνοασφάλειας, με βασικό στόχο την πρόβλεψη πιθανών παραβιάσεων πριν από την εκδήλωσή τους, χωρίς την ανάγκη συνεργασίας από τους οργανισμούς.

Σκοπός της είναι να προλάβει σοβαρές ζημιές που μπορεί να προκύψουν από κυβερνοεπιθέσεις, ακολουθώντας μια προληπτική προσέγγιση. Για την επίτευξη αυτού του στόχου, έχουν συλλεγεί 258 χαρακτηριστικά, τα οποία χωρίζονται σε δύο κατηγορίες: συμπτώματα κακής διαχείρισης, όπως οι λανθασμένες ρυθμίσεις DNS [17] ή BGP [18], και χρονοσειρές δεδομένων που σχετίζονται με κακόβουλη δραστηριότητα, όπως spam και επιθέσεις phishing.

Για την ανάλυση των δεδομένων, αναπτύχθηκε ένας ταξινομητής τύπου Random Forest, ο οποίος έχει εκπαιδευτεί και δοκιμαστεί σε πάνω από 1.000 αναφορές περιστατικών από γνωστές βάσεις δεδομένων, καλύπτοντας περιπτώσεις από το 2013 έως το 2014. Η αξιολόγηση του ταξινομητή έδειξε εξαιρετικά αποτελέσματα, με δείκτη αληθών θετικών ανιχνεύσεων 90% (TP), με δείκτη ψευδώς θετικών ανιχνεύσεων 10% (FP) και συνολική ακρίβεια 90%. Αυτά τα αποτελέσματα αναδεικνύουν την αξία της πρόβλεψης στην κυβερνοασφάλεια, ιδίως σε μια εποχή που οι επιθέσεις σε μεγάλες εταιρείες δείχνουν να αυξάνονται, επισημαίνοντας την κοινωνική και οικονομική επιρροή τους.

Στη μελέτη, εξετάζεται η δυνατότητα πρόβλεψης περιστατικών μέσω συλλογής εξωτερικών δεδομένων που αποκαλύπτουν την κατάσταση ασφαλείας ενός δικτύου, ελαχιστοποιώντας την ανάγκη για εσωτερικές λειτουργίες. Ενώ η μηχανική μάθηση έχει κυριαρχήσει στην ανίχνευση κακόβουλων δραστηριοτήτων, η εφαρμογή της για την πρόβλεψη περιστατικών παραμένει περιορισμένη. Η διάκριση μεταξύ ανίχνευσης και πρόβλεψης είναι κρίσιμη, καθώς η ανίχνευση εστιάζει στην αναγνώριση γνωστών χαρακτηριστικών, ενώ η πρόβλεψη προσπαθεί να αναγνωρίσει παράγοντες που σχετίζονται με την πιθανότητα εμφάνισης μιας επίθεσης. Μέσω της ανάλυσης εξωτερικών δεδομένων και της συμπεριφοράς των οργανισμών, μπορούμε να αποκτήσουμε καλύτερη κατανόηση των πιθανών επιθέσεων.

Η διαδικασία εντοπισμού των μονάδων συγκέντρωσης για την εκπαίδευση του ταξινομητή στηρίζεται σε πληροφορίες από τις βάσεις δεδομένων των RIRs (Regional Internet Registries). Δημιουργώντας έναν παγκόσμιο πίνακα συγκέντρωσης, κατέγραψαν οργανισμούς που έχουν υποστεί επιθέσεις και οργανισμούς που δεν έχουν, εξασφαλίζοντας μια ολοκληρωμένη βάση δεδομένων για ανάλυση. Το σύνολο δεδομένων περιλαμβάνει 4,4 εκατομμύρια προθέματα και 2,6 εκατομμύρια ταυτότητες ιδιοκτητών, επιτρέποντας τη λεπτομερή κατηγοριοποίηση των οργανισμών.

Στην ανάλυση τους, η διαδικασία συγκέντρωσης βασίζεται σε εμπειρικούς κανόνες, αλλά αντιμετωπίζει προκλήσεις στην ακριβή αναγνώριση των οργανισμών, καθώς πολλές μεγάλες εταιρείες καταχωρούν τις διευθύνσεις IP τους με πολλαπλές ταυτότητες ιδιοκτητών. Αυτό καθιστά

δύσκολη τη διάκριση μεταξύ οργανισμών και απαιτεί προσεκτική επεξεργασία των δεδομένων. Η αξιοπιστία των δεδομένων που χρησιμοποιούνται για την εκπαίδευση του ταξινομητή είναι επίσης κρίσιμη, δεδομένου ότι οι αναφορές περιστατικών παραβίασης δεδομένων ενδέχεται να είναι υποαναφερόμενες, γεγονός που επηρεάζει την ακριβή εκτίμηση της πιθανότητας παραβίασης ενός οργανισμού. Για την αντιμετώπιση αυτής της πρόκλησης, χρησιμοποιήθηκαν τρία διαφορετικά σύνολα δεδομένων για την εκπαίδευση και τη δοκιμή. Επιπλέον, αναγνώρισαν ότι τα δεδομένα σχετικά με την κατάσταση ασφάλειας προέρχονται από δημόσιες μαύρες λίστες, οι οποίες, αν και χρήσιμες, δεν παρέχουν πλήρη κάλυψη, με αποτέλεσμα ορισμένες περιοχές να αντιπροσωπεύονται δυσανάλογα.

Συνοψίζοντας, οι προβλέψεις που προέκυψαν από την παραπάνω μέθοδο επιβεβαίωσαν την αποτελεσματικότητα του ταξινομητή τους, με τη διανομή των δεδομένων εκπαίδευσης και δοκιμής να είναι ισομερής μεταξύ των θυμάτων. Αυτή η έρευνα αναδεικνύει τη σημασία της πρόβλεψης περιστατικών κυβερνοασφάλειας, καθώς προσφέρει τη δυνατότητα εφαρμογής προληπτικών πολιτικών που μπορούν να μειώσουν σημαντικά τους κινδύνους και το κόστος που σχετίζεται με τις παραβιάσεις.

2.2 Αυτοματοποιημένη Επικύρωση Ελέγχου Ασφαλείας

Η αυτοματοποιημένη επικύρωση ελέγχου ασφαλείας των συστημάτων (Automated Security Control Validation) έχει ως στόχο την παρακολούθηση της ασφάλειας σε πραγματικό χρόνο, σε ένα δυναμικά μεταβαλλόμενο περιβάλλον με συνεχώς αναδυόμενες απειλές. Οι ερευνητές χρησιμοποιούν τεχνητή νοημοσύνη για να επιτύχουν μια ολοκληρωμένη και ακριβή αξιολόγηση της συνολικής ασφάλειας ενός συστήματος. Αυτή η αξιολόγηση μπορεί να γίνει με βάση δεδομένα από το «τηλεσκόπιο» ενός δικτύου, την ανάπτυξη πλαισίων κυβερνοασφάλειας για κτίρια, ή τη συσχέτιση απειλών, ευπαθειών και μέτρων ασφαλείας. Στη συνέχεια, θα αναλύσουμε διεξοδικά κάθε μία από αυτές τις τρεις προσεγγίσεις.

Μια από τις κύριες προσεγγίσεις για την αυτοματοποιημένη διαδικασία ελέγχου της ασφάλειας είναι η αξιοποίηση δεδομένων από το «τηλεσκόπιο» ενός δικτύου [19]. Το σκεπτικό των επιστημόνων πίσω από τη μελέτη αυτή ήταν να κατανοήσουν σε βάθος τη φύση των κυβερνοεπιθέσεων και να αποκαλύψουν τον μηχανισμό λειτουργίας τους, προκειμένου να βελτιώσουν τις μεθόδους παρακολούθησης και πρόβλεψης επιθέσεων στον κυβερνοχώρο. Αναγνωρίζοντας ότι οι επιθέσεις στον κυβερνοχώρο είναι εξαιρετικά πολύπλοκες και δύσκολο να αναλυθούν ως απλές τυχαίες διαδικασίες, στρέφονται σε στοχαστικά μοντέλα και προηγμένες τεχνικές ανάλυσης χρονοσειρών, με στόχο την κατανόηση της δυναμικής και περιοδικότητας αυτών των επιθέσεων.

Ο κύριος στόχος τους είναι η ακριβέστερη πρόβλεψη και κατανόηση της συμπεριφοράς των επιτιθέμενων και των επιθέσεων, με σκοπό τη βελτίωση της κυβερνοασφάλειας. Μέσω της παρακολούθησης κακόβουλης δραστηριότητας που καταγράφεται από το δίκτυο τηλεσκοπίου της CAIDA, οι επιστήμονες επιδιώκουν να αναλύσουν τον αριθμό των επιθέσεων, τη γεωγραφική κατανομή των επιτιθέμενων, καθώς και τη χρονική περιοδικότητα των επιθέσεων. Με αυτό τον τρόπο δίνουν έμφαση στην διαδικασία με την οποία αυτές εξελίσσονται με την πάροδο του χρόνου. Για την επίτευξη αυτών των στόχων, οι επιστήμονες χρησιμοποίησαν διάφορες μεθόδους ανάλυσης χρονοσειρών, οι οποίες επιτρέπουν τη μελέτη των δεδομένων σε βάθος χρόνου ώστε να εντοπιστούν μοτίβα και επαναλαμβανόμενες συμπεριφορές.

Οι επιστήμονες χρησιμοποίησαν μεθόδους που περιλαμβάνουν:

1. **Το ARIMA (AutoRegressive Integrated Moving Average)** [20]

   Αποτελεί ένα μοντέλο ανάλυσης χρονοσειρών που χρησιμοποιήθηκε για την πρόβλεψη μελλοντικών τιμών με βάση προηγούμενες παρατηρήσεις. Το μοντέλο αυτό αποδείχθηκε χρήσιμο για την κατανόηση εποχιακών τάσεων και την πρόβλεψη επιθέσεων στον κυβερνοχώρο, ιδίως όταν τα δεδομένα παρουσίαζαν μια σταθερή συμπεριφορά. Ωστόσο, για δεδομένα με υψηλή μεταβλητότητα, όπως αυτά που αφορούν κυβερνοεπιθέσεις, το GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) [21] μοντέλο ήταν πιο κατάλληλο. Το συγκεκριμένο εξυπηρετεί την ανάλυση δεδομένων με απότομες αυξήσεις και διακυμάνσεις, βοηθώντας στην κατανόηση της δυναμικής εξάρτησης της διακύμανσης των επιθέσεων.

2. **Dynamic Time Warping (DTW)** [22]

   Η συγκεκριμένη μέθοδος χρησιμοποιήθηκε για την αξιολόγηση της ομοιότητας μεταξύ χρονοσειρών επιθέσεων. Το DTW επιτρέπει την αναγνώριση κοινών μοτίβων ακόμη και όταν οι χρονοσειρές εμφανίζουν χρονικές αποκλίσεις, γεγονός που βοήθησε στη σύγκριση της συμπεριφοράς επιθέσεων από διαφορετικές γεωγραφικές περιοχές ή χρονικές περιόδους. Επιπλέον, για να διαπιστωθεί αν οι χρονοσειρές του "χρόνου σάρωσης" ήταν στατικές ή όχι, οι επιστήμονες εφάρμοσαν τον έλεγχο Augmented Dickey-Fuller [23]. Τα αποτελέσματα του ελέγχου αυτού έδειξαν ότι οι χρονοσειρές δεν ήταν στατικές, οδηγώντας έτσι στην επιλογή στοχαστικών μοντέλων, όπως τα ARIMA και GARCH, για την καλύτερη κατανόηση των δεδομένων.

Με τις μεθόδους αυτές, οι επιστήμονες κατάφεραν να προσφέρουν μια πιο ολοκληρωμένη εικόνα για τη δυναμική των κυβερνοεπιθέσεων και την περιοδικότητά τους, συμβάλλοντας στην παγκόσμια κυβερνοασφάλεια και προτείνοντας τρόπους για τη βελτίωση των συστημάτων παρακολούθησης και πρόβλεψης επιθέσεων. Παράλληλα, στόχος τους ήταν να αναλύσουν την περιοδικότητα και συντονισμένη φύση των επιθέσεων, ειδικά σε συγκεκριμένες ώρες της ημέρας, αλλά και να αξιολογήσουν την αποτελεσματικότητα μικρών και μεγάλων τηλεσκοπίων στη συλλογή αξιόπιστων δεδομένων επιθέσεων.

Σε επόμενο βήμα, εύλογο είναι να αναλύσουμε την δεύτερη προσέγγιση που είναι η ανάπτυξη πλαισίων κυβερνοασφάλειας για τα κτίρια [24]. Οι επιστήμονες που ανέπτυξαν το διαδικτυακό εργαλείο Πλαισίου Κυβερνοασφάλειας Κτιρίων **(BCF)** [25] επικεντρώνονται στην επιτακτική ανάγκη ενίσχυσης της κυβερνοασφάλειας στα κτίρια, καθώς αυτά φιλοξενούν κρίσιμες υποδομές και ευαίσθητα δεδομένα που μπορούν να αποτελέσουν στόχους κυβερνοαπειλών. Το σκεπτικό τους βασίζεται στην αναγνώριση ότι η προστασία αυτών των υποδομών απαιτεί μια οργανωμένη και συστηματική προσέγγιση.

Ο στόχος τους είναι να παρέχουν στους χρήστες τη δυνατότητα να εκτελούν αυτο-αξιολογήσεις, εντοπίζοντας αδυναμίες και κενά στην ασφάλεια, καθώς και να καθορίσουν τους στρατηγικούς τους στόχους για βελτίωση της κυβερνοασφάλειας, στηριζόμενοι σε τέσσερα επίπεδα ωριμότητας (MILs) που καθορίζουν την πρόοδο και τις απαιτήσεις. Κάθε επίπεδο περιλαμβάνει ερωτήσεις και επιλογές αξιολόγησης που διευκολύνουν την κατανόηση των κινδύνων και των απαιτούμενων επενδύσεων.

Ως προς την επίτευξη αυτών των στόχων, οι επιστήμονες αξιοποιούν μια σειρά μεθόδων τεχνητής νοημοσύνης (AI). Αυτές περιλαμβάνουν την ανάλυση δεδομένων από τις αυτο-αξιολογήσεις και τις διαδικασίες παρακολούθησης, επιτρέποντας την αναγνώριση προτύπων και ανωμαλιών που υποδεικνύουν ενδεχόμενες κυβερνοαπειλές.

Στο παράδειγμα που αναφέρεται οι διαχειριστές του κτιρίου φαίνεται να αντιμετωπίζουν προκλήσεις σχετικά με τις κρίσιμες cyber-assets, τη διαχείριση κινδύνων και την εκπαίδευση των υπαλλήλων στην κυβερνοασφάλεια. Τα τέσσερα σενάρια τα οποία παρουσιάζονται, είναι τα εξής:

- Στο πρώτο σενάριο, οι επενδύσεις βελτιώνουν την ορατότητα και τη διαχείριση των περιουσιακών στοιχείων.
- Στο δεύτερο, η ανάπτυξη διαδικασιών προστασίας και ανίχνευσης καθορίζει ρόλους και ευθύνες, ενώ ενισχύει την εκπαίδευση και τις πολιτικές ασφάλειας.
- Στο τρίτο σενάριο, η αναθεώρηση σχεδίων αντίκτυπων και αποκατάστασης βελτιώνει την αντίδραση σε περιστατικά.
- Τέλος, στο τέταρτο σενάριο, οι διαδικασίες ωριμάζουν μέσω της συνεχούς παρακολούθησης και ανάλυσης, εξασφαλίζοντας τις κρίσιμες λειτουργίες και απαιτήσεις ασφάλειας.

Μέσω των αλγορίθμων μηχανικής μάθησης μπορούμε να προβλέψουμε τους κινδύνους και τις ευπάθειες. Έτσι οι διαχειριστές έχουν τη δυνατότητα να σχεδιάσουν προληπτικά μέτρα και να ενσωματώσουν λύσεις που περιορίζουν τις πιθανότητες επιθέσεων. Επιπλέον, η AI μπορεί να εξατομικεύσει στρατηγικές ασφάλειας, διασφαλίζοντας ότι οι προτάσεις βελτίωσης είναι σχετικές και εναρμονισμένες με τις συγκεκριμένες ανάγκες και τα χαρακτηριστικά κάθε κτιρίου.

Η ανάπτυξη του BCF συνεπώς υπογραμμίζει τη σημασία της αναγνώρισης και της μείωσης των ευπαθειών στην κυβερνοασφάλεια, προτείνοντας μια ολιστική προσέγγιση που περιλαμβάνει συνεχή εκπαίδευση του προσωπικού και την ανάπτυξη ισχυρών πολιτικών ασφάλειας. Η διαδικασία αυτή ενθαρρύνει τη συμμόρφωση με τα πρότυπα ασφαλείας και τη συστηματική παρακολούθηση των διαδικασιών. Τέλος, οι βελτιώσεις στις διαδικασίες και οι στρατηγικές κατανάλωσης πόρων είναι κρίσιμες για την επίτευξη ενός πιο ανθεκτικού και ασφαλούς περιβάλλοντος, προσφέροντας έτσι μια πλήρη και οργανωμένη λύση απέναντι στις σύγχρονες κυβερνοαπειλές.

Στην τελευταία προσέγγιση που θα εξετάσουμε, προτείνεται μια συσχέτιση που ενώνει τις απειλές, τις ευπάθειες και τα μέτρα ασφαλείας [26]. Οι επιστήμονες αναγνωρίζουν την ολοένα αυξανόμενη σημασία της ασφάλειας των συστημάτων στην προστασία των πληροφοριών και των πόρων των οργανισμών, ειδικά σε ένα κόσμο που γίνεται όλο και πιο ψηφιακός. Το σκεπτικό τους είναι ότι, για να επιτευχθεί μια αποτελεσματική αξιολόγηση της ασφάλειας, πρέπει να κατανοηθούν οι πολύπλοκες σχέσεις μεταξύ των απειλών, των ευπαθειών και των μέτρων ασφαλείας. Οι κίνδυνοι που αντιμετωπίζουν τα συστήματα είναι ποικίλοι και προέρχονται τόσο από εξωτερικούς όσο και από εσωτερικούς παράγοντες, γεγονός που καθιστά την ασφαλή λειτουργία τους επιτακτική ανάγκη.

Ο βασικός στόχος των ερευνητών είναι η ανάπτυξη μεθόδων και εργαλείων που θα επιτρέψουν την ακριβή αξιολόγηση της ασφάλειας των συστημάτων, καθώς και την ανάπτυξη στρατηγικών που θα ενισχύσουν την προστασία τους από κακόβουλες επιθέσεις. Η διαδικασία αξιολόγησης, ωστόσο, είναι περίπλοκη, δεδομένου ότι πολλές από τις παραμέτρους ασφαλείας δεν μπορούν εύκολα να ποσοτικοποιηθούν. Ως εκ τούτου, οι επιστήμονες προτείνουν τη χρήση γλωσσικών δομών και κλιμάκων σύγκρισης για την τυποποίηση των απειλών και των ευπαθειών, επιτρέποντας έτσι την ποσοτικοποίηση χαρακτηριστικών που συνήθως είναι δύσκολα να εκφραστούν με αριθμούς.

Η μέθοδος που χρησιμοποιείται για την αξιολόγηση της ασφάλειας περιλαμβάνει τη χρήση γενετικών αλγορίθμων, οι οποίοι μιμούνται τη διαδικασία της φυσικής επιλογής. Αυτοί οι αλγόριθμοι επιτρέπουν την αναγνώριση και την επιλογή των πιο κατάλληλων παραμέτρων ασφαλείας, καθώς και την απόρριψη των λιγότερο αποτελεσματικών. Οι γενετικοί αλγόριθμοι είναι ιδιαίτερα ευέλικτοι

και μπορούν να προσαρμοστούν σε διάφορα σενάρια, κάτι που τους καθιστά εξαιρετικά χρήσιμους στην αξιολόγηση και ενίσχυση της ασφάλειας των πληροφοριών.

Η διαδικασία περιλαμβάνει διάφορα στάδια, όπως η ποσοτικοποίηση των χαρακτηριστικών των απειλών και των μέτρων ασφαλείας μέσω πίνακα ζεύγους σύγκρισης και η κανονικοποίηση των αριθμητικών τιμών για την αποτελεσματική σύγκριση. Ο δείκτης ασφάλειας υπολογίζεται με βάση τη διαφορά ανάμεσα στην τιμή της συνάρτησης καταλληλότητας του μέτρου και της τιμής της απειλής, επιτρέποντας έτσι την εκτίμηση της συνολικής ασφάλειας του συστήματος.

Η προσέγγιση αυτή αναδεικνύει την ανάγκη για συνεχή έρευνα και ανάπτυξη στον τομέα της ασφάλειας συστημάτων, καθώς και την εφαρμογή αυτών των τεχνικών σε διάφορους τομείς της ανθρώπινης δραστηριότητας. Οι επιστήμονες εστιάζουν προβλημάτων που σχετίζονται με την αναπαράσταση των δεδομένων και την επιλογή του κατάλληλου επιπέδου προσέγγισης, έτσι ώστε να καταστεί δυνατή η πλήρης αξιολόγηση της ασφάλειας σε πολύπλοκα συστήματα.

2.3 Αυτοματοποιημένη Αναγνώριση και Αξιολόγηση Ευπάθειας

Η αυτοματοποιημένη αναγνώριση και αξιολόγηση ευπαθειών (Automated Vulnerability Identification & Assessment) είναι μια διαδικασία όπου χρησιμοποιούνται ειδικά εργαλεία για να εντοπίζουν και να αναλύουν αδυναμίες ασφαλείας σε ένα σύστημα. Αυτά τα εργαλεία βασίζονται σε βάσεις δεδομένων ευπαθειών, ενημερώσεις από κατασκευαστές, συστήματα που παρακολουθούν τα περιουσιακά στοιχεία ενός οργανισμού και πληροφορίες για απειλές. Στόχος τους είναι να αναγνωρίσουν, να ταξινομήσουν, να αξιολογήσουν τη σοβαρότητα των ευπαθειών και να προτείνουν λύσεις για την αντιμετώπισή τους.

2.3.1 Αυτοματοποιημένη Ανίχνευση Ευπαθειών

Η αυτοματοποιημένη ανίχνευση ευπαθειών (Automated Vulnerability Detection) είναι μια διαδικασία που χρησιμοποιεί τεχνολογίες τεχνητής νοημοσύνης και μηχανικής μάθησης για την ταυτοποίηση αδυναμιών σε λογισμικό, διακομιστές και άλλα συστήματα. Οι ερευνητές αναπτύσσουν μεθόδους για να ελέγχουν τον πηγαίο κώδικα μέσω βαθιάς μάθησης, ενσωματώνοντας τεχνικές εξόρυξης κειμένου και συστήματα συστάσεων που βοηθούν τους προγραμματιστές να γράφουν ασφαλέστερο κώδικα. Παράλληλα, γίνονται προσπάθειες για την ανίχνευση νέων ευπαθειών σε λογισμικό και κυβερνο-υποδομές, είτε μέσω ανάλυσης δεδομένων από αποθετήρια ευπαθειών είτε από πληροφορίες στα κοινωνικά δίκτυα.

Μια άλλη προσέγγιση περιλαμβάνει τη χρήση μηχανικής μάθησης για την ανίχνευση επιθέσεων σε κυβερνο-φυσικά συστήματα (CPS) και συστήματα IoT, μοντελοποιώντας τη συμπεριφορά αυτών των συστημάτων υπό συνθήκες επίθεσης. Επιπλέον, το AI-based fuzzing αποτελεί μια δημοφιλή μέθοδο για τον εντοπισμό ευπαθειών, όπου τυχαία ή απροσδόκητα δεδομένα εισάγονται σε λογισμικό για την ανίχνευση δυσλειτουργιών όπως σφάλματα, διαρροές μνήμης ή κατάρρευση συστημάτων. Αυτή η τεχνική εφαρμόζεται σε διάφορες περιπτώσεις, όπως προγράμματα περιήγησης, μεταγλωττιστές και συστήματα CPS.

Ακόμη, η αυτοματοποιημένη δοκιμή διείσδυσης χρησιμοποιείται για να ανακαλύψει τι μπορεί να κερδίσει ένας επιτιθέμενος από γνωστές ή μηδενικές (zero-day) ευπάθειες [27] σε ένα σύστημα.

Αυτές οι μέθοδοι ενσωματώνουν ενισχυτική μάθηση για τη δημιουργία αυτόνομων συστημάτων δοκιμής διείσδυσης, τα οποία στοχεύουν σε μεγάλα δίκτυα και μικροδίκτυα, αναζητώντας αποτελεσματικούς τρόπους για την εκμετάλλευση των ευπαθειών.

Ας εξετάσουμε σε βάθος κάθε πτυχή της συγκεκριμένης διαδικασίας, αναλύοντας τις τεχνολογίες και τις μεθόδους που εμπλέκονται, καθώς και τον τρόπο με τον οποίο συνδυάζονται για να διασφαλίσουν την ανίχνευση και την αντιμετώπιση ευπαθειών σε συστήματα και λογισμικό. Αυτή η διερεύνηση θα μας επιτρέψει να κατανοήσουμε καλύτερα τις προκλήσεις και τα οφέλη που προκύπτουν από την αυτοματοποιημένη ανίχνευση ευπαθειών, αξιοποιώντας την τεχνητή νοημοσύνη και άλλες καινοτόμες τεχνικές.

Η ανίχνευση ευπαθειών στον πηγαίο κώδικα μέσω τεχνικών βαθιάς μάθησης και μεταφοράς μάθησης έχει καταστεί εξαιρετικά αποτελεσματική στην πρόληψη επιθέσεων. Οι επιστήμονες που αναπτύσσουν εργαλεία ανάλυσης κώδικα κατανοούν τη ζωτική σημασία της χρηστικότητάς τους για την ασφάλεια του λογισμικού και στοχεύουν στην αξιολόγηση στο κομμάτι της επίδρασης της σχεδίασης της διεπαφής χρήστη στη λειτουργικότητα και την αποδοχή των εργαλείων αυτών από τους προγραμματιστές. Ένας βασικός στόχος τους είναι η σύγκριση του **VulIntel**, ενός νέου εργαλείου που ενσωματώνει μηχανική μάθηση, με το παραδοσιακό εργαλείο **FindBugs**. Με αυτή τη σύγκριση, επιδιώκουν να προσδιορίσουν ποιο εργαλείο παρέχει καλύτερη εμπειρία χρήσης, καθώς και πώς οι σχεδιαστικές διαφορές επηρεάζουν την ικανότητα των χρηστών να εντοπίζουν και να διορθώνουν σφάλματα στον κώδικα.

Οι κύριοι στόχοι της έρευνας [28] περιλαμβάνουν την αναγνώριση του εργαλείου που προσφέρει μια πιο φιλική και κατανοητή διεπαφή για τους προγραμματιστές, την αξιολόγηση της ικανοποίησης και της εμπειρίας των χρηστών κατά τη χρήση των εργαλείων και την ανάπτυξη ενός ανθρωποκεντρικού εργαλείου ανάλυσης κώδικα που θα παρέχει άμεσες και χρήσιμες ανατροφοδοτήσεις, ικανοποιώντας τις ανάγκες των προγραμματιστών.

Για να επιτευχθούν αυτοί οι στόχοι, οι ερευνητές χρησιμοποίησαν τη μέθοδο A/B testing. Οι συμμετέχοντες χρησιμοποίησαν και τα δύο εργαλεία εναλλάξ για να συγκρίνουν τη χρηστικότητα και την αποδοχή τους σε πραγματικές συνθήκες. Το VulIntel, το οποίο αξιοποιεί μηχανική μάθηση, ανιχνεύει ευπάθειες στον κώδικα και παρέχει προτάσεις επιδιορθώσεων σε πραγματικό χρόνο, επιτρέποντας έτσι στους προγραμματιστές να διορθώνουν σφάλματα κατά την κωδικοποίηση. Επιπροσθέτως, η μηχανική μάθηση επιτρέπει στο VulIntel να εξελίσσεται, μαθαίνοντας από προηγούμενες αναλύσεις και βελτιώνοντας τις προβλέψεις του.

Η μελέτη έδειξε ότι το VulIntel ήταν πιο εύχρηστο, με τους συμμετέχοντες να εκτιμούν την άμεση ανάλυση και τις προτάσεις βελτίωσης που παρείχε κατά την κωδικοποίηση. Αντίθετα, το FindBugs δυσκόλευε την πλοήγηση στα σφάλματα λόγω της πολυπλοκότητας της διεπαφής του. Για την ανάλυση των αποτελεσμάτων, οι ερευνητές χρησιμοποίησαν στατιστικές τεχνικές όπως το T-test [29] και η ANOVA [30], οι οποίες τους επέτρεψαν να επιβεβαιώσουν τη σημασία των διαφορών στη χρηστικότητα των εργαλείων και να εκτιμήσουν πώς η σχεδίαση της διεπαφής επηρεάζει την εμπειρία του χρήστη. Ο σκοπός λοιπόν της έρευνας είναι να αναδείξει τη σημασία της ανθρωποκεντρικής σχεδίασης στα εργαλεία ανάλυσης κώδικα, επισημαίνοντας πώς μια καλά σχεδιασμένη διεπαφή μπορεί να βελτιώσει τη διαδικασία ανάλυσης και να ενισχύσει την παραγωγικότητα των προγραμματιστών.

Οι Shida Liu, Zhongsheng Hou, Yuan Guo και Lei Guo ασχολήθηκαν με την αντιμετώπιση προκλήσεων [31] που παρουσιάζονται σε μη γραμμικά συστήματα με χρονοκαθυστερήσεις και εξωτερικές διαταραχές, τα οποία είναι συχνά προβλήματα στους τομείς της ρομποτικής και του

ελέγχου συστημάτων. Οι παραδοσιακές μέθοδοι ελέγχου, όπως οι MFAC [32] και EMFAC [33], παρουσιάζουν αδυναμίες στην απόκριση και την ακρίβεια όταν αντιμετωπίζουν αυτά τα ζητήματα. Ως εκ τούτου, το κίνητρό τους ήταν η βελτίωση της σταθερότητας, της ακρίβειας και της ανθεκτικότητας των συστημάτων που υπόκεινται σε τέτοιες συνθήκες.

Ο στόχος των ερευνητών ήταν να αναπτύξουν έναν αλγόριθμο που να διαχειρίζεται με αποτελεσματικότητα τις χρονοκαθυστερήσεις και τις εξωτερικές διαταραχές, ενώ παράλληλα να εξασφαλίζει σταθερή και ακριβή απόκριση χωρίς υπερβολές. Επιπλέον, ήθελαν να επιτύχουν μια πιο ευέλικτη και δυναμική ρύθμιση των παραμέτρων του ελέγχου, η οποία να προσαρμόζεται αυτόματα στις αλλαγές του συστήματος.

Για τους παραπάνω λόγους, χρησιμοποιήθηκε ο αλγόριθμος **Ro-MFAC** [34], ο οποίος υλοποιεί την τεχνική της δυναμικής γραμμικοποίησης, η οποία βασίζεται στην έννοια της ψευδο-κλίσης, για την εκτίμηση παραμέτρων και τη ρύθμιση του ελέγχου. Αυτή η προσέγγιση επιτρέπει στον αλγόριθμο να προσαρμόζεται σε πραγματικό χρόνο στις αλλαγές του συστήματος και να διαχειρίζεται τις χρονοκαθυστερήσεις και τις διαταραχές. Αν και δεν αναφέρεται συγκεκριμένα κάποια μέθοδος τεχνητής νοημοσύνης (AI), η δυνατότητα αυτόματης προσαρμογής του ελέγχου και η χρήση παραμέτρων όπως η ψευδο-κλίση παραπέμπουν σε τεχνικές προσαρμοστικών αλγορίθμων, οι οποίες είναι συχνά συνδεδεμένες με την ευφυή ρύθμιση και μάθηση σε δυναμικά περιβάλλοντα.

Συμπερασματικά, ο Ro-MFAC αποτελεί έναν καινοτόμο αλγόριθμο που παρέχει σημαντικά πλεονεκτήματα έναντι των παραδοσιακών μεθόδων, προσφέροντας αυξημένη ακρίβεια, σταθερότητα και αντοχή σε διαταραχές, κάτι που διευρύνει τις δυνατότητες εφαρμογών στον τομέα της ρομποτικής και των συστημάτων ελέγχου.

Εύλογο είναι να δούμε και μια μελέτη [35] η οποία εστιάζει στην ανάπτυξη του **AutoVAS**, ενός αυτόματου συστήματος ανάλυσης ευπαθειών λογισμικού που αξιοποιεί μεθόδους βαθιάς μάθησης, με σκοπό την αποτελεσματική και γρήγορη ανίχνευση ευπαθειών στον κώδικα. Οι επιστήμονες παρατηρούν ότι οι τεχνολογίες hacking [36] εξελίσσονται ραγδαία, οδηγώντας σε αύξηση των ευπαθειών λογισμικού και δημιουργώντας την ανάγκη για αυτοματοποιημένες λύσεις ανάλυσης. Ενώ οι παραδοσιακές μέθοδοι απαιτούν ανθρώπινη παρέμβαση, το AutoVAS επιδιώκει τη μείωση αυτής της εξάρτησης, χρησιμοποιώντας δεδομένα από βάσεις όπως το National Vulnerability Database (NVD) [37] και το Software Assurance Reference Database (SARD) [38].

Το AutoVAS έχει ως στόχο την αποτελεσματική ανίχνευση ευπαθειών σε πραγματικά έργα λογισμικού, συμπεριλαμβανομένων των zero-day ευπαθειών, που είναι οι ευπάθειες που δεν έχουν ακόμη ανακαλυφθεί ή διορθωθεί. Το σύστημα αναγνωρίζει και κατατάσσει ευπάθειες στον κώδικα με χαμηλά ποσοστά ψευδών θετικών και ψευδών αρνητικών ευρημάτων, διασφαλίζοντας έτσι την αξιοπιστία των αποτελεσμάτων.

Για την ανάλυση του κώδικα, το AutoVAS ενσωματώνει τεχνικές προγραμματιστικής ανάλυσης, χρησιμοποιώντας διάφορες μεθόδους αναπαράστασης δεδομένων που ανταγωνίζονται την πολυπλοκότητα του κώδικα και τις διαφορές στις ονομασίες των μεταβλητών. Η μέθοδος του «προγραμματιστικού τεμαχισμού» (program slicing) είναι κεντρική στην ανάλυση, επιτρέποντας την απομόνωση τμημάτων κώδικα που σχετίζονται με συγκεκριμένες ευπάθειες. Για τη μετατροπή του πηγαίου κώδικα σε διανύσματα, χρησιμοποιούνται τεχνικές ενσωμάτωσης λέξεων, όπως το Word2Vec [39], το GloVe [40] και το FastText [41]. Το FastText αποδεικνύεται πιο αποτελεσματικό σε μικρές δειγματοληψίες, γεγονός που συμβάλλει στην ποιότητα της ανάλυσης.

Η διαδικασία ανίχνευσης του AutoVAS διαχωρίζεται σε δύο φάσεις: την εκπαίδευση και την ανίχνευση. Συγκεκριμένα στην εκπαιδευτική φάση, το μοντέλο εκπαιδεύεται με τη χρήση προετοιμασμένων «code snippets» [42], τα οποία χαρακτηρίζονται ως ευπαθή ή μη ευπαθή με βάση λέξεις-κλειδιά. Η διαδικασία αυτή εξασφαλίζει ότι ο κώδικας διαιρείται σε μικρότερα τμήματα, διευκολύνοντας την ανίχνευση.

Η αξιολόγηση του AutoVAS βασίζεται σε τρεις κύριες ερευνητικές ερωτήσεις:

1. την αποτελεσματικότητα των μεθόδων embedding [43],
2. την πρακτική εφαρμογή του συστήματος
3. τη σύγκριση με άλλες προσεγγίσεις ανίχνευσης ευπαθειών.

Πειραματικά αποτελέσματα δείχνουν ότι οι μέθοδοι slicing [44] που βασίζονται στο interprocedure και SGD-IFDS προσφέρουν την καλύτερη απόδοση στον προγραμματισμό. Επίσης, το AutoVAS εφαρμόστηκε σε διάφορα μοντέλα RNN, με την καλύτερη απόδοση να επιτυγχάνεται με τη χρήση τριών επιπέδων νευρωνικών δικτύων και batch size 256.

Η συγκριτική ανάλυση με άλλα συστήματα, όπως το **VulDeePecker** [45] και το **SySeVR** [46], αποδεικνύει ότι το AutoVAS υπερτερεί κυρίως λόγω της ικανότητάς του να περιλαμβάνει περισσότερες πληροφορίες και να βελτιστοποιεί τη μέθοδο embedding. Σημαντικά είναι και τα ευρήματα της ανίχνευσης 11 ευπαθειών σε εννέα έργα, εκ των οποίων οι 7 ήταν γνωστές και οι 4 άγνωστες, συμπεριλαμβανομένων και zero-day ευπαθειών.

Παρά τις επιτυχίες του, υπάρχουν περιθώρια βελτίωσης, όπως η υποστήριξη για άλλες γλώσσες προγραμματισμού και η ανίχνευση ευπαθειών σε εκτελέσιμα αρχεία. Οι επιστήμονες προτείνουν περαιτέρω έρευνα στη χρήση μη επιτηρούμενων μεθόδων μάθησης και ερμηνεύσιμης μηχανικής μάθησης, ώστε να ενισχυθεί η ανίχνευση ευπαθειών και να κατανοηθούν καλύτερα οι αιτίες τους. Συνολικά, η μελέτη αυτή προσφέρει μια σημαντική συνεισφορά στην αυτόματη ανίχνευση ευπαθειών λογισμικού και παρουσιάζει μια αποτελεσματική προσέγγιση που μπορεί να βελτιώσει τη διαδικασία εντοπισμού και διόρθωσης των κενών ασφαλείας, παρέχοντας παράλληλα εργαλεία για μελλοντική έρευνα και ανάπτυξη.

Από τις παραπάνω τρεις μελέτες διαπιστώνουμε ότι επικεντρώνονται στη σημασία της χρηστικότητας και της εμπειρίας του χρήστη, επιδιώκοντας να προσφέρουν εργαλεία που διευκολύνουν τους προγραμματιστές στην ανίχνευση και διόρθωση σφαλμάτων. Επίσης, περιλαμβάνουν στατιστικές αναλύσεις για την αξιολόγηση των αποτελεσμάτων τους και αναγνωρίζουν τις σύγχρονες προκλήσεις στον τομέα της ασφάλειας, προτείνοντας καινοτόμες λύσεις για την αποτελεσματική ανίχνευση και διαχείριση ευπαθειών.

Εκτός από την ανάλυση του κώδικα, είναι σημαντικό να παρακολουθούνται οι νέες ευπάθειες που εμφανίζονται συνεχώς. Οι ερευνητές χρησιμοποιούν δεδομένα από αποθετήρια ευπαθειών και κοινωνικά δίκτυα για να εντοπίσουν νέες απειλές σε πραγματικό χρόνο. Ξεκινώντας με την ανίχνευση αναδυόμενων ευπαθειών μέσω αποθετηρίων ευπαθειών [47], μπορούμε να κατανοήσουμε πώς αυτά συμβάλλουν στην έγκαιρη πρόληψη επιθέσεων. Οι επιστήμονες που ανέπτυξαν το σύστημα σύστασης εντόπισαν ένα κρίσιμο πρόβλημα στον τομέα της ασφάλειας των οργανισμών, σχετιζόμενο με την ανίχνευση και μετρίαση των ευπαθειών στο λογισμικό και το υλικό που χρησιμοποιούν. Ο κύριος στόχος τους ήταν να αυτοματοποιήσουν τη διαδικασία αντιστοίχισης των ευπαθειών, όπως αυτές καταγράφονται σε δημόσιες βάσεις δεδομένων, με τα προϊόντα που χρησιμοποιούν οι οργανισμοί. Αυτή η αυτοματοποίηση θα εξοικονομούσε χρόνο και θα μείωνε τα

σφάλματα που προκύπτουν από τις ανθρώπινες αναλύσεις, ιδιαίτερα δεδομένου του μεγάλου όγκου και της ποικιλίας των λογισμικών και υλικών.

Η βασική τους επιδίωξη ήταν η δημιουργία ενός συστήματος ικανό να αντιστοιχίζει αυτόματα τα προϊόντα λογισμικού και υλικού ενός οργανισμού με τις ευπάθειες που καταγράφονται στη βάση δεδομένων NVD (National Vulnerability Database) [48], χρησιμοποιώντας την τυποποιημένη ονοματολογία CPE (Common Product Enumeration) [49]. Με αυτό τον τρόπο, οι οργανισμοί θα μπορούσαν να λαμβάνουν γρηγορότερα και πιο ακριβή αποτελέσματα, ενημερώνοντάς τους για τις ευπάθειες που επηρεάζουν τα συστήματά τους χωρίς να χρειάζεται να βασίζονται σε χρονοβόρες διαδικασίες.

Για την υλοποίηση αυτού του στόχου, οι επιστήμονες χρησιμοποίησαν προηγμένες τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) [50] και μεθόδους μηχανικής μάθησης. Η μεθοδολογία περιλάμβανε την ασαφή αντιστοίχιση (fuzzy matching), η οποία προσομοιώνει τη διαδικασία ενός ανθρώπινου αναλυτή στην αναζήτηση αντιστοιχιών μεταξύ προϊόντων και ευπαθειών. Ένα από τα βασικά στοιχεία αυτής της προσέγγισης ήταν η ανάλυση των ονομάτων προϊόντων λογισμικού και η εξαγωγή διανυσμάτων λέξεων (word vectors) μέσω της χρήσης του Word2Vec, επιτρέποντας τη μέτρηση της ομοιότητας μεταξύ λέξεων και φράσεων, ακόμη και όταν υπάρχουν μικρές παραλλαγές ή τυπογραφικά λάθη.

Η διαδικασία υπολογισμού της ομοιότητας πραγματοποιείται με τη χρήση της συνημίτονου ομοιότητας (cosine similarity) [51] και της ομοιότητας χαρακτήρων. Για να βελτιωθεί η ακρίβεια, το μοντέλο μηχανικής μάθησης εκπαιδεύεται με πραγματικά δεδομένα και χρησιμοποιεί διάφορα χαρακτηριστικά για την ταξινόμηση των αποτελεσμάτων σε κατηγορίες προτεραιότητας, επιτρέποντας στους αναλυτές να εστιάσουν στις πιο σημαντικές ευπάθειες και μειώνοντας τον κίνδυνο για ψευδώς θετικά αποτελέσματα.

Το πρακτικό κομμάτι της έρευνας περιλάμβανε την αξιολόγηση του συστήματος σύστασης σε σύγκριση με έναν ανθρώπινο αναλυτή. Δύο σύνολα δεδομένων χρησιμοποιήθηκαν: 50 ονόματα λογισμικού και 50 ονόματα υλικού. Η διαδικασία αναζήτησης ευπαθειών στο NVD κατέδειξε τη σαφή υπεροχή του συστήματος στην ακρίβεια και την ταχύτητα. Ο ανθρώπινος αναλυτής χρειάστηκε περίπου 70 λεπτά για να εντοπίσει τις αντίστοιχες CPEs για το λογισμικό, ενώ για το υλικό απαιτήθηκαν περίπου 6,5 ώρες. Οι χειροκίνητες αναζητήσεις για το λογισμικό παρήγαγαν κατά μέσο όρο 119 αποτελέσματα, με μόνο το 8% αυτών να είναι επιτυχή. Αντίθετα, το σύστημα σύστασης παρήγαγε μόνο 2 αποτελέσματα ανά αναζήτηση, με ποσοστό επιτυχίας 40% για το λογισμικό και 48% για το υλικό, το οποίο επίσης παρήγαγε 2 αποτελέσματα.

Αυτό σημαίνει ότι το σύστημα εξοικονόμησε πάνω από 7 ώρες σε αυτή τη μικρή βάση δεδομένων, προσφέροντας ταυτόχρονα μεγαλύτερη ακρίβεια και μειώνοντας τον κίνδυνο να παραληφθούν σημαντικές ευπάθειες. Η χρήση τεχνικών τεχνητής νοημοσύνης και μηχανικής μάθησης κατέδειξε τη δυνατότητα των συστημάτων αυτών να βελτιώσουν σημαντικά τις διαδικασίες ανίχνευσης και μείωσης των κινδύνων για τις επιχειρήσεις.

Συνολικά, οι επιστήμονες επιδίωξαν να βελτιώσουν την αποτελεσματικότητα των διαδικασιών ανίχνευσης ευπαθειών και μείωσης των κινδύνων για τις επιχειρήσεις, με την αυτοματοποίηση της αντιστοίχισης ευπαθειών. Χρησιμοποιώντας τεχνικές τεχνητής νοημοσύνης και μηχανικής μάθησης, πέτυχαν σημαντική εξοικονόμηση χρόνου και ακρίβεια, προσφέροντας έναν πιο αξιόπιστο και γρήγορο τρόπο εντοπισμού κρίσιμων ευπαθειών που αφορούν τις επιχειρήσεις.

Στη συνέχεια θα εξετάσουμε τον ρόλο των κοινωνικών δικτύων στην παρακολούθηση νέων απειλών σε πραγματικό χρόνο. Το σκεπτικό των επιστημόνων πίσω από την ανάπτυξη του Yggdrasil και της σχετικής μελέτης [52] στον τομέα της κυβερνοασφάλειας επικεντρώνεται στην ανάγκη ενίσχυσης της συνεργατικής μάθησης και της πρώιμης ανίχνευσης ευπαθειών μέσω της ανάλυσης δεδομένων που προέρχονται από διαδικτυακές κοινότητες. Οι ερευνητές επιδιώκουν να δημιουργήσουν ένα εργαλείο που θα διευκολύνει την αλληλεπίδραση μεταξύ ειδικών και μη ειδικών στην κυβερνοασφάλεια, επιτρέποντας την πιο αποτελεσματική αναγνώριση και κατηγοριοποίηση κυβερνοασφαλιστικών απειλών που δεν εντοπίζονται εύκολα από παραδοσιακά συστήματα.

Σκοπός της μελέτης είναι η βελτίωση της διαδικασίας μάθησης μέσω της συνεργασίας, προκειμένου να ενισχυθεί η κατανόηση των κυβερνοασφαλιστικών απειλών και ευπαθειών. Ένα από τα κύρια ζητήματα που επιδιώκουν να επιλύσουν οι ερευνητές είναι η ανάγκη για πρώιμη ανίχνευση ευπαθειών, με στόχο την αναγνώριση και την κατηγοριοποίηση αναδυόμενων ευπαθειών μέσω της αυτόματης ανάλυσης πληροφοριών που προέρχονται από tweets και άλλες πηγές. Αυτό έχει μεγάλη σημασία, καθώς οι απειλές αυτές μπορεί να μην είναι ευρέως γνωστές ή να μην έχουν εντοπιστεί από τα παραδοσιακά συστήματα κυβερνοασφάλειας.

Για την επίτευξη αυτών των στόχων, οι ερευνητές χρησιμοποίησαν διάφορες τεχνικές και αλγορίθμους τεχνητής νοημοσύνης.

1. **Το μοντέλο BERT (Bidirectional Encoder Representations from Transformers)** [53] χρησιμοποιήθηκε για την ανάλυση κειμένου και την κατηγοριοποίηση των tweets ως σχετικά ή μη σχετικά με τις emergent ευπάθειες, αξιοποιώντας τα κείμενα των άρθρων στα οποία παραπέμπουν, καθώς και στατιστικά στοιχεία από τα tweets, όπως οι αριθμοί των «likes» και των retweets.
2. Επιπλέον, **αλγόριθμοι όπως το SVM (Support Vector Machine)** [54] και **τα CNN (Convolutional Neural Networks)** [55] χρησιμοποιήθηκαν για την εκπαίδευση και την αξιολόγηση του μοντέλου, με βάση την απόδοσή τους σε πειράματα διασταυρούμενης επικύρωσης.

Τα αποτελέσματα της μελέτης έδειξαν ότι το μοντέλο BERT πέτυχε την καλύτερη ακρίβεια κατά την ανάλυση του κειμένου των άρθρων, ενώ τα CNN παρουσίασαν επίσης υψηλές επιδόσεις. Η προσθήκη των στατιστικών στοιχείων από το Twitter βελτίωσε περαιτέρω την απόδοση των μοντέλων. Αξιοσημείωτο είναι ότι, όταν χρησιμοποιήθηκαν μόνο τα tweets, η απόδοση των μοντέλων μειώθηκε, αλλά τα αποτελέσματα παρέμειναν ικανοποιητικά. Παράλληλα, εξετάστηκε η δυνατότητα μεταφοράς μάθησης, δηλαδή η χρήση προηγούμενων μοντέλων και δεδομένων για τη βελτίωση της απόδοσης του τρέχοντος μοντέλου, επιβεβαιώνοντας την καταλληλότητα του BERT για την εξαγωγή γνώσεων από κοινότητες κυβερνοασφάλειας. Τέλος, το Yggdrasil αναπτύχθηκε ως ένα αυτοματοποιημένο εργαλείο που επιτρέπει στους χρήστες, είτε ειδικούς είτε μη, να πλοηγούνται στις γνώσεις που δημιουργούνται από τις κοινότητες γύρω από τις αναδυόμενες κυβερνοασφαλιστικές απειλές, συμβάλλοντας σημαντικά στην αναγνώριση και τη διαχείριση των κυβερνο-ευπαθειών.

Σε πιο πολύπλοκα περιβάλλοντα, όπως τα κυβερνο-φυσικά συστήματα (CPS) και το Διαδίκτυο των Πραγμάτων (IoT), οι ευπάθειες μπορεί να βρίσκονται σε επίπεδο συστήματος ή δικτύου. Οι επιστήμονες που ασχολούνται με τα Κυβερνοφυσικά Συστήματα (CPS) και τις Συσκευές του Διαδικτύου των Πραγμάτων (IoT) εστιάζουν στην ασφάλεια αυτών των συστημάτων, καθώς η ενσωμάτωσή τους στην καθημερινή ζωή και η σημασία τους σε κρίσιμες υποδομές τα καθιστούν στόχους επιθέσεων [56]. Το σκεπτικό τους βασίζεται στην ανάγκη για αποτελεσματική προστασία από επιθέσεις που μπορεί να οδηγήσουν σε σοβαρές συνέπειες, ιδίως σε τομείς όπως η υγειονομική περίθαλψη, οι έξυπνες πόλεις και η αυτοκινητοβιομηχανία.

Ο κύριος στόχος τους είναι η ανίχνευση και κατηγοριοποίηση ευπαθειών, τόσο γνωστών όσο και αγνώστων, καθώς και η ανάπτυξη στρατηγικών για την προληπτική άμυνα απέναντι σε αυτές τις επιθέσεις. Στην κατεύθυνση αυτή, οι ερευνητές επιδιώκουν να δημιουργήσουν ένα πολυδιάστατο σύστημα ασφάλειας που περιλαμβάνει την αναγνώριση νέων επιθετικών σεναρίων και τη βελτίωση της αντίκρουσης επιθέσεων μέσω της εφαρμογής τεχνολογιών μηχανικής μάθησης.

Για την υλοποίηση αυτών των στόχων, οι επιστήμονες χρησιμοποιούν διάφορες μεθόδους τεχνητής νοημοσύνης, κυρίως τεχνικές μηχανικής μάθησης. Μία από τις πιο αποτελεσματικές μεθόδους που χρησιμοποιούν είναι το μοντέλο υποστήριξης διανυσμάτων **(SVM)**, το οποίο επιτρέπει την εκπαίδευση ενός μοντέλου με βάση δεδομένα από προηγούμενες επιθέσεις. Με τη χρήση SVM, αναλύουν τα χαρακτηριστικά των επιθέσεων, όπως ευπάθειες στη μνήμη, αδυναμίες στον έλεγχο πρόσβασης και ελλείψεις στην κρυπτογράφηση, για να εντοπίσουν πρότυπα και να προβλέψουν νέες επιθέσεις που μπορεί να μην έχουν εντοπιστεί ακόμη.

Αντί να βασίζονται μόνο σε παραδοσιακές μεθόδους ανίχνευσης, οι επιστήμονες αναπτύσσουν γράφους επιθέσεων, οι οποίοι απεικονίζουν τις αλληλεπιδράσεις και τις ροές δεδομένων κατά τη διάρκεια μιας επίθεσης. Αυτοί οι κατευθυνόμενοι άκυκλοι γράφοι (DAG) [57] περιέχουν πληροφορίες για τις ενέργειες που εκτελούνται κατά τη διάρκεια της επίθεσης, επιτρέποντας στους ερευνητές να εντοπίσουν τις κρίσιμες αδυναμίες και να σχεδιάσουν στρατηγικές άμυνας.

Μέσω πειραματικών δοκιμών, οι επιστήμονες αξιολογούν την αποτελεσματικότητα των μοντέλων τους. Για παράδειγμα, η ανάλυση σε πραγματικά δίκτυα, όπως το δίκτυο Controller Area Network (CAN) [58] σε συνδεδεμένα οχήματα, έχει δείξει την ικανότητα του SVM να ανακαλύψει νέες ευπάθειες και να εντοπίσει επιθέσεις που δεν είχαν προηγουμένως καταγραφεί. Οι ερευνητές αναφέρουν τη δημιουργία νέων σεναρίων επιθέσεων και ευπαθειών, βελτιώνοντας έτσι την ασφάλεια των συστημάτων.

Για να υποστηρίξουν τη διαδικασία αυτή, οι ερευνητές χρησιμοποιούν επίσης τεχνικές όπως η δημιουργία αρνητικών παραδειγμάτων, που αντιπροσωπεύουν αδύνατες επιθέσεις, για να εκπαιδεύσουν τα μοντέλα τους και να διασφαλίσουν την ακρίβεια των προβλέψεων. Αυτές οι στρατηγικές εξασφαλίζουν ότι οι επιστήμονες είναι σε θέση να ανιχνεύουν και να προλαμβάνουν πιθανές επιθέσεις, ενισχύοντας την ασφάλεια των CPS και IoT και συμβάλλοντας στην ανάπτυξη πιο ασφαλών τεχνολογιών στον ψηφιακό κόσμο. Με αυτόν τον τρόπο, συμβάλλουν στην ανάπτυξη πιο ασφαλών συστημάτων που μπορούν να ανταποκριθούν στις προκλήσεις του σύγχρονου ψηφιακού κόσμου.

Επιπλέον, ερευνητές έχουν εφαρμόσει τεχνικές λογικής και επεξεργασίας φυσικής γλώσσας για την παραγωγή αρχικών δεδομένων (seed generation), με στόχο την αύξηση της κάλυψης κώδικα και την εξερεύνηση περισσότερων μοναδικών διαδρομών εκτέλεσης, αποτελώντας μια θεμελιώδη πτυχή των έξυπνων συστημάτων fuzzing.

Η ανακάλυψη ευπαθειών σε λογισμικό είναι μία από τις πιο κρίσιμες προκλήσεις στην ασφάλεια των υπολογιστών. Παρά τις σημαντικές προόδους στις τεχνικές fuzzing, οι υπάρχουσες μέθοδοι συχνά επικεντρώνονται στην κάλυψη του κώδικα, αγνοώντας τη στοχευμένη αναζήτηση ευπαθών τμημάτων. Οι επιστήμονες [59] που ανέπτυξαν το **NeuFuzz** εντόπισαν αυτό το κενό και αποφάσισαν να χρησιμοποιήσουν την τεχνητή νοημοσύνη (AI) για να καθοδηγήσουν τη διαδικασία της δοκιμής με έναν πιο έξυπνο και αποτελεσματικό τρόπο. Το σκεπτικό τους βασίζεται στην παρατήρηση ότι οι υπάρχουσες τεχνικές fuzzing δεν λαμβάνουν υπόψη τις διαφορές στα μονοπάτια εκτέλεσης του προγράμματος που μπορεί να είναι ευάλωτα. Αυτή η προσέγγιση έχει ως αποτέλεσμα τη σπατάλη πόρων σε δοκιμές που δεν είναι πιθανό να αποκαλύψουν ευπάθειες.

Ως εκ τούτου, οι ερευνητές αποφάσισαν να αναπτύξουν ένα εργαλείο που θα εστιάζει όχι μόνο στην κάλυψη του κώδικα, αλλά και στη στοχευμένη εξερεύνηση ευπαθών περιοχών του προγράμματος. Οι κύριοι στόχοι του NeuFuzz περιλαμβάνουν την αύξηση της απόδοσης στην ανακάλυψη ευπαθειών, τη στοχευμένη αναζήτηση σε μονοπάτια εκτέλεσης που είναι πιο πιθανό να περιέχουν ευπάθειες και τη χρήση τεχνητής νοημοσύνης για να εκπαιδεύσουν ένα νευρωνικό δίκτυο που θα αναγνωρίζει μοτίβα που υποδηλώνουν ευπάθειες, βελτιώνοντας έτσι τη διαδικασία επιλογής των δεδομένων εισόδου για τις δοκιμές.

Η προσέγγιση του NeuFuzz περιλαμβάνει τη χρήση βαθιάς μάθησης για την ανάλυση των μονοπατιών εκτέλεσης των προγραμμάτων και τη διαδικασία εκπαίδευσης του μοντέλου με δεδομένα από ευπαθή και μη ευπαθή προγράμματα. Στην εκπαίδευση του μοντέλου, συλλέγονται προγράμματα από διάφορες πηγές, όπως το NIST SARD [60] και το GitHub [61], προκειμένου να εκπαιδευτεί το νευρωνικό δίκτυο να αναγνωρίζει τα κρυμμένα μοτίβα ευπαθειών. Η επιλογή των δεδομένων για τη δοκιμή γίνεται με βάση τις προβλέψεις του εκπαιδευμένου νευρωνικού δικτύου, το οποίο έχει την ικανότητα να προσδιορίζει ποια μονοπάτια είναι πιο πιθανό να είναι ευάλωτα. Έτσι, το NeuFuzz δίνει προτεραιότητα σε μονοπάτια που έχουν μεγαλύτερη πιθανότητα να οδηγήσουν σε εντοπισμό ευπαθειών, εξοικονομώντας χρόνο και πόρους στην ανακάλυψη ευπαθειών.

Στο πλαίσιο της αξιολόγησης του NeuFuzz, οι ερευνητές εκτέλεσαν πειράματα χρησιμοποιώντας το εργαλείο σε δύο σειρές δοκιμών: το **LAVA-M** και εννέα εφαρμογές από τον πραγματικό κόσμο. Κατά τη διάρκεια των δοκιμών, το NeuFuzz ανίχνευσε 28 νέες ευπάθειες, 21 από τις οποίες επιβεβαιώθηκαν με CVE αναγνωριστικά [62], δηλαδή ήταν ευπάθειες που δεν είχαν εντοπιστεί προηγουμένως. Επιπλέον, οι δοκιμές έδειξαν ότι το NeuFuzz είχε 91% ακρίβεια στην πρόβλεψη των ευπαθών μονοπατιών, αν και παρουσίασε ελαφρώς χαμηλότερη απόδοση σε συγκεκριμένα σύνολα δεδομένων σε σύγκριση με άλλα εργαλεία, όπως το **PTfuzz** [63] και το **QAFL** [64].

Αξιοσημείωτο είναι ότι κατά τη διάρκεια των δοκιμών, το NeuFuzz εντόπισε 1,290 crashes σε εφαρμογές πραγματικού κόσμου μέσα σε 24 ώρες, εκ των οποίων 42 αναγνωρίστηκαν ως ευπάθειες, με 28 από αυτές να είναι καινούριες. Οι εντοπισμένες ευπάθειες περιλάμβαναν κυρίως σφάλματα buffer overflow, χρήση ελεύθερης μνήμης (UAF), και προβλήματα με null pointers. Η ικανότητα του NeuFuzz να ανιχνεύει και να επισημαίνει περισσότερες ευπάθειες από άλλα εργαλεία υποδεικνύει την αποτελεσματικότητα και την υποσχόμενη εφαρμογή του στον τομέα της ασφάλειας λογισμικού.

Συνολικά, το NeuFuzz προσφέρει μια καινοτόμα προσέγγιση στην ανίχνευση ευπαθειών, συνδυάζοντας τις δυνατότητες της τεχνητής νοημοσύνης με τις παραδοσιακές τεχνικές fuzzing, βελτιώνοντας τη συνολική αποδοτικότητα της διαδικασίας ανακάλυψης ευπαθειών.

Συνεχίζοντας στο ίδιο πλαίσιο [65], μια καινοτόμα διαδικασία δημιουργίας σπόρων δεδομένων για fuzzing προγράμματα που επεξεργάζονται πολύ-δομημένα δεδομένα, όπως XSLT, XML και JavaScript, είναι η μέθοδος **«Skyfire»**. Ο κύριος στόχος των ερευνητών είναι να βελτιώσουν την ικανότητα ανίχνευσης σφαλμάτων, ενισχύοντας την κάλυψη του κώδικα και αποκαλύπτοντας νέες ευπάθειες μέσω της εφαρμογής της τεχνολογίας fuzzing. Το σκεπτικό πίσω από την έρευνα εστιάζει στην ανησυχία ότι τα υπάρχοντα εργαλεία fuzzing συχνά αποτυγχάνουν να ανιχνεύσουν σφάλματα λόγω της ανεπαρκούς κατανομής των εισόδων που παράγουν. Ειδικότερα, η μέθοδος mutation-based fuzzing, που τροποποιεί υπάρχουσες έγκυρες εισόδους, συχνά καταλήγει σε πολλές απορριπτέες εισόδους κατά την αρχική ανάλυση σύνταξης. Αντίθετα, η generation-based fuzzing, που παράγει εισόδους βάσει προδιαγραφών, αντιμετωπίζει δυσκολίες κατά την εκτέλεση ελέγχου σημασίας.

Οι επιστήμονες στοχεύουν να ενσωματώσουν τη γνώση από έναν μεγάλο αριθμό υπαρχόντων δειγμάτων για να δημιουργήσουν εισόδους που είναι καλύτερα κατανεμημένες και πιο κατάλληλες για fuzzing, με σκοπό την ανίχνευση σφαλμάτων που είναι δύσκολα προσβάσιμα μέσω συμβατικών μεθόδων.

Η μέθοδος Skyfire περιλαμβάνει δύο βασικά βήματα:

1. την εκμάθηση πιθανοτικής γραμματικής (PCSG) [66] και
2. τη δημιουργία σπόρων (seeds).

Οι ερευνητές εκμεταλλεύονται ένα μεγάλο σύνολο δειγμάτων και τη γραμματική τους για να εξάγουν αυτόματα τους σημασιολογικούς κανόνες και τη συχνότητα των κανόνων παραγωγής. Μέσω της εκμάθησης μιας πιθανοτικής γραμματικής ευαίσθητης στο πλαίσιο (PCSG), είναι σε θέση να παράγουν εισόδους που είναι πιο πιθανό να περάσουν τις απαραίτητες φάσεις ανάλυσης και εκτέλεσης εφαρμογής. Οι σπόροι εισόδου δημιουργούνται είτε μέσω αριστερής παραγωγής είτε μέσω τυχαίας διαδικασίας Monte Carlo [67], διασφαλίζοντας την παραγωγή ποικιλόμορφων και μη κοινών εισόδων και επιτρέποντας τη γρηγορότερη προώθηση μέσω των σταδίων ανάλυσης.

Η εφαρμογή του Skyfire σε διάφορες μηχανές XSLT και XML, καθώς και σε JavaScript engines, απέδειξε ότι η μέθοδος βελτιώνει την κάλυψη του κώδικα κατά 20% και την ικανότητα ανίχνευσης σφαλμάτων. Η μελέτη αποκάλυψε 19 νέες ευπάθειες μνήμης, για τις οποίες οι ερευνητές έλαβαν συνολικά 33,5 χιλιάδες δολάρια σε αμοιβές, αποδεικνύοντας την αποτελεσματικότητα της προσέγγισης. Στο μέλλον, οι επιστήμονες σκοπεύουν να επεκτείνουν τη μέθοδο Skyfire ώστε να υποστηρίξουν περισσότερες γλώσσες προγραμματισμού όπως JavaScript, SQL, C και Java, ελπίζοντας ότι οι παραγόμενοι σπόροι θα είναι χρήσιμοι και στην ανίχνευση σφαλμάτων στους μεταγλωττιστές.

Συμπερασματικά, η μέθοδος «Skyfire» παρουσιάζει μια πρωτοποριακή προσέγγιση στην παραγωγή σπόρων για fuzzing προγράμματα, συνδυάζοντας τη γνώση από υπάρχοντα δείγματα και τη μαθησιακή τεχνολογία για τη βελτίωση της αποτελεσματικότητας των διαδικασιών ελέγχου και ανίχνευσης σφαλμάτων. Μέσω της ενσωμάτωσης πιθανοτικών γραμματικών και άλλων μεθόδων τεχνητής νοημοσύνης, οι ερευνητές δημιουργούν ένα ισχυρό εργαλείο για την ανίχνευση ευπαθειών, συμβάλλοντας έτσι στη βελτίωση της ασφάλειας των υπολογιστικών συστημάτων.

Ας παραμείνουμε στο fuzzing και ας εστιάσουμε στην δημιουργία δοκιμαστικών περιπτώσεων που αποτελεί ένα από τα πιο εκτενώς μελετημένα πεδία στην τεχνολογία fuzzing και βασίζεται στην τεχνητή νοημοσύνη. Αυτή η τεχνολογία εφαρμόζεται σε διάφορους τομείς, όπως οι περιηγητές ιστού, οι μεταγλωττιστές, τα κυβερνοφυσικά συστήματα, οι βιβλιοθήκες λογισμικού και τα απλά προγράμματα υπολογιστή. Ας εξετάσουμε κάθε περίπτωση λεπτομερώς.

Στους περιηγητές ιστού [68], οι επιστήμονες προτείνουν μια νέα προσέγγιση για την αυτοματοποίηση της διαδικασίας fuzzing, η οποία στοχεύει στην ανίχνευση ευπαθειών ασφαλείας σε προγράμματα που αναλύουν δεδομένα εισόδου. Η διαδικασία fuzzing περιλαμβάνει τη συνεχή δοκιμή ενός προγράμματος με αλλοιωμένα δεδομένα εισόδου για να αποκαλυφθούν τυχόν σφάλματα. Η παραδοσιακή μέθοδος απαιτεί τη χειροκίνητη δημιουργία γραμματικών κανόνων εισόδου, κάτι που είναι χρονοβόρο και επιρρεπές σε λάθη. Για αυτό, οι ερευνητές σκοπεύουν να αυτοματοποιήσουν τη διαδικασία αυτή, χρησιμοποιώντας τεχνικές μηχανικής μάθησης, και συγκεκριμένα νευρωνικά δίκτυα.

Ο κύριος στόχος της προσέγγισης είναι η δημιουργία γραμματικών κανόνων εισόδου αυτόματα από δείγματα δεδομένων, προκειμένου να διευκολυνθεί η διαδικασία fuzzing σε πολύπλοκες δομές,

όπως τα αρχεία PDF. Παράλληλα, επιδιώκουν την ανίχνευση σφαλμάτων και ευπαθειών μέσω της εισαγωγής «πειραγμένων» δεδομένων εισόδου σε προγράμματα ανάλυσης. Ένας επιπλέον στόχος είναι η βελτίωση της κάλυψης του κώδικα κατά τη διάρκεια των δοκιμών, ώστε να διασφαλιστεί ότι περισσότερες διαδρομές του κώδικα θα εξεταστούν.

Για την υλοποίηση αυτής της προσέγγισης, οι επιστήμονες χρησιμοποιούν επαναλαμβανόμενα και συγκλητικά νευρωνικά δίκτυα (RNNs) για την εκπαίδευση ενός στατιστικού μοντέλου, το οποίο ονομάζεται char-rnn [69], που μπορεί να παράγει νέα δεδομένα εισόδου με βάση την πιθανότητα εμφάνισης χαρακτήρων σε δομές PDF. Ο αλγόριθμος που προτείνουν συνδυάζει τη μάθηση και το fuzzing, εισάγοντας σφάλματα σε σημεία όπου το μοντέλο είναι σίγουρο για την ορθότητα της πρόβλεψής του. Με αυτόν τον τρόπο, μπορούν να εντοπιστούν πιθανά προβλήματα στον κώδικα που δεν θα ήταν εύκολο να ανακαλυφθούν με άλλες μεθόδους.

Η εκπαίδευση του μοντέλου περιλαμβάνει πέντε διαφορετικές παραμέτρους εποχών, με την αξιολόγηση της αποτελεσματικότητας να γίνεται μέσω του προγράμματος περιήγησης Edge της Microsoft για PDF. Οι κύριες μετρήσεις αξιολόγησης περιλαμβάνουν την κάλυψη, το ποσοστό επιτυχίας και την ανίχνευση σφαλμάτων, χρησιμοποιώντας ένα σύνολο 63.000 μη δυαδικών αντικειμένων PDF.

Τα ευρήματα δείχνουν ότι η κάλυψη και το ποσοστό επιτυχίας έχουν αντίθετες τάσεις, γεγονός που υποδηλώνει την ανάγκη για έναν ισορροπημένο συνδυασμό κανονικών και «πειραγμένων» εισροών κατά την εκπαίδευση. Οι συγγραφείς επισημαίνουν ότι η καλύτερη εκπαίδευση δεν σημαίνει πάντα καλύτερη απόδοση στο fuzzing, γεγονός που καταδεικνύει την ανάγκη για ευφυείς στρατηγικές στη διαδικασία εκπαίδευσης και δοκιμής. Τέλος, προτείνουν ότι η χρήση ενισχυτικής μάθησης μπορεί να βοηθήσει στη βελτίωση της διαδικασίας εκμάθησης και στην αύξηση της κάλυψης κατά τη διάρκεια των δοκιμών.

Σε επόμενη μελέτη [70] με τίτλο «Compiler Fuzzing through Deep Learning» επικεντρωνόμαστε στην ανάπτυξη μιας καινοτόμου μεθόδου που συνδυάζει τη μηχανική μάθηση με την τυχαία παραγωγή προγραμμάτων (fuzzing) για τη βελτίωση της διαδικασίας ανίχνευσης σφαλμάτων στους μεταγλωττιστές. Οι επιστήμονες αναγνωρίζουν ότι οι παραδοσιακές μέθοδοι fuzzing απαιτούν εκτενή ανάπτυξη και εξειδίκευση για κάθε γλώσσα προγραμματισμού, γεγονός που καθιστά δύσκολη τη χρήση τους σε νέες ή λιγότερο γνωστές γλώσσες.

Ως εκ τούτου, ο στόχος τους ήταν να δημιουργήσουν μια αυτοματοποιημένη και αποτελεσματική μέθοδο που θα διευκολύνει τη διαδικασία ανίχνευσης σφαλμάτων, απαιτώντας λιγότερο χρόνο και προσπάθεια. Η προτεινόμενη μέθοδος, γνωστή ως **DeepSmith** [71], χρησιμοποιεί δεδομένα από πραγματικά προγράμματα για την εκπαίδευση ενός μοντέλου, το οποίο είναι ικανό να παράγει χιλιάδες ρεαλιστικά προγράμματα, μειώνοντας σημαντικά την εργασία που απαιτείται για την επαλήθευση των μεταγλωττιστών.

Για την υλοποίηση της DeepSmith, οι ερευνητές χρησιμοποίησαν μοντέλα LSTM (Long Short-Term Memory) [72] για την κατανόηση της δομής και της σύνταξης της γλώσσας OpenCL [73]. Αυτά τα επαναλαμβανόμενα νευρωνικά δίκτυα εκπαιδεύτηκαν σε ένα μεγάλο σύνολο προγραμμάτων OpenCL, επιτρέποντας στο μοντέλο να παράγει νέα, ρεαλιστικά προγράμματα. Επιπλέον, η διαδικασία περιλάμβανε μια δοκιμαστική υποδομή για την εκτέλεση των παραγόμενων προγραμμάτων και την αναφορά των αποτελεσμάτων, καθώς και μεθόδους ψηφοφορίας για την αναγνώριση ανωμαλιών μεταξύ των εκδόσεων των μεταγλωττιστών.

Τα αποτελέσματα σύμφωνα με τους επιστήμονες ήταν εντυπωσιακά. Η DeepSmith ανίχνευσε 67 σφάλματα σε διάφορους μεταγλωττιστές, επισημαίνοντας τη μεγαλύτερη αποτελεσματικότητα της μεθόδου σε σύγκριση με παραδοσιακά εργαλεία δοκιμών, όπως ο **CLSmith** [74]. Ο DeepSmith ήταν

ταχύτερος και πιο αποδοτικός, παρέχοντας λιγότερες και πιο κατανοητές περιπτώσεις δοκιμών. Οι επιστήμονες υπογράμμισαν ότι η ποιότητα των μεταγλωττιστών βελτιώνεται με την πάροδο του χρόνου, αν και οι συνεχιζόμενες αναπτύξεις νέων χαρακτηριστικών εισάγουν νέες προκλήσεις.

Η μελέτη επιβεβαίωσε επίσης τη σταθερότητα του μεταγλωττιστή Clang στην υποστήριξη του OpenCL, με σημαντική μείωση των σφαλμάτων κατά τη διάρκεια της δοκιμής. Επιπλέον, οι επιστήμονες αναγνώρισαν ότι η αρχιτεκτονική του DeepSmith είναι επεκτάσιμη, επιτρέποντας τη στήριξη και άλλων γλωσσών προγραμματισμού, όπως η Solidity [75].

Συνολικά, η προσέγγιση της DeepSmith προσφέρει μια οικονομικά αποδοτική μέθοδο για τη δημιουργία περιπτώσεων δοκιμής, μειώνοντας τον απαιτούμενο χρόνο ανάπτυξης και βελτιώνοντας την ανίχνευση σφαλμάτων στους μεταγλωττιστές. Αυτή η καινοτόμος μέθοδος, που συνδυάζει μηχανική μάθηση και τυχαία προγραμματιστική γενιά, αποδεικνύεται πιο αποτελεσματική από τις υπάρχουσες μεθόδους, επισημαίνοντας τη σημασία της αυτόματης δοκιμής και επικύρωσης των μεταγλωττιστών.

Συνεχίζοντας στο πεδίο των μεταγλωττιστών, οι επιστήμονες σε μια εναλλακτική μελέτη [76] επιλέγουν να αναπτύσσουν το DSmith. Κάνοντας χρήση του DSmith επιδιώκουν να βελτιώσουν τη διαδικασία δοκιμής και ανίχνευσης σφαλμάτων, εντάσσοντας την τεχνική fuzzing σε συνδυασμό με προηγμένες μεθόδους βαθιάς μάθησης. Το σκεπτικό τους βασίζεται στην ανάγκη να αντιμετωπιστούν οι μακρινές εξαρτήσεις σύνταξης, οι οποίες συχνά οδηγούν σε σφάλματα κατά την παραγωγή δοκιμαστικών περιπτώσεων. Οι παραδοσιακές μέθοδοι fuzzing πολλές φορές αποτυγχάνουν να αναγνωρίσουν αυτές τις εξαρτήσεις, με αποτέλεσμα τη δημιουργία συντακτικών σφαλμάτων.

Η μελέτη του DSmith προτείνει μια αυτοματοποιημένη και αποτελεσματική μέθοδο που συνδυάζει μηχανική μάθηση και τυχαία παραγωγή προγραμμάτων (fuzzing), ώστε να διευκολύνει την ανίχνευση σφαλμάτων, απαιτώντας λιγότερο χρόνο και προσπάθεια. Οι κύριοι στόχοι των ερευνητών περιλαμβάνουν την αύξηση της αποτελεσματικότητας του fuzzing, τη δημιουργία σωστών συντακτικά προγραμμάτων και την ανακάλυψη νέων σφαλμάτων. Μέσω του DSmith, έχουν εντοπιστεί 11 νέα σφάλματα σε εκδόσεις του μεταγλωττιστή GCC [77], υποδεικνύοντας την αποτελεσματικότητα της προσέγγισης αυτής.

Για την επίτευξη αυτών των στόχων, οι ερευνητές χρησιμοποίησαν μηχανισμούς LSTM (Long Short-Term Memory), οι οποίοι επιτρέπουν στο μοντέλο να αποθηκεύει και να ανακτά πληροφορίες από προηγούμενα στοιχεία, ελαχιστοποιώντας την απώλεια πληροφορίας σε μακροχρόνιες εξαρτήσεις. Ο μηχανισμός προσοχής (Attention Mechanism) ενισχύει την ικανότητα του μοντέλου να εστιάζει σε συγκεκριμένες πληροφορίες από την είσοδο κατά τη διαδικασία παραγωγής, επιτρέποντας τη δημιουργία πιο περίπλοκων και σωστών συντακτικά δομών.

Η αρχιτεκτονική encoder-decoder του DSmith επιτρέπει την αποτελεσματική επεξεργασία γλωσσικών μοντέλων, διευκολύνοντας την παραγωγή δοκιμαστικών προγραμμάτων που πληρούν τις απαιτήσεις των μεταγλωττιστών.

Επιπλέον, το DSmith υιοθετεί τέσσερις στρατηγικές γενιάς προγραμμάτων:

1. ολοκλήρωση κώδικα
2. εισαγωγή γραμμών κώδικα
3. καθοδηγούμενη εισαγωγή κλάδων
4. καθοδηγούμενη εισαγωγή συναρτήσεων

Αυτές οι στρατηγικές στοχεύουν στη βελτίωση της ποικιλίας και της ποιότητας των παραγόμενων κωδίκων.

Η συνδυασμένη χρήση αυτών των μεθόδων και στρατηγικών αποδεικνύει την καινοτομία του DSmith και τη δυνατότητά του να ανιχνεύει σφάλματα στους μεταγλωττιστές με μεγαλύτερη αποτελεσματικότητα από τις παραδοσιακές μεθόδους fuzzing, καθιστώντας το εργαλείο κρίσιμης σημασίας για τη βελτίωση της αξιοπιστίας των μεταγλωττιστών. Η προσέγγιση αυτή τονίζει τη σημασία της εφαρμογής προηγμένων τεχνικών βαθιάς μάθησης για την αναβάθμιση των διαδικασιών δοκιμής και την ταχύτερη ανακάλυψη σφαλμάτων, κάτι που μπορεί να ωφελήσει σημαντικά την ανάπτυξη και τη συντήρηση μεταγλωττιστών σε βάθος χρόνου.

Στα κυβερνο-φυσικά συστήματα (CPS) [78] αναδεικνύονται προκλήσεις που αφορούν την ασφάλεια και την ανθεκτικότητα των μηχανισμών άμυνάς τους. Οι επιστήμονες επισημαίνουν ότι, παρά την κρισιμότητα των CPS για τη λειτουργία κρίσιμων υποδομών, όπως οι εγκαταστάσεις ύδρευσης και επεξεργασίας, η αξιολόγηση των μηχανισμών άμυνάς τους είναι πολύπλοκη λόγω της έλλειψης ρεαλιστικών επιθέσεων για δοκιμές. Αυτή η έλλειψη δεδομένων καθιστά δύσκολη την εκτίμηση των ικανοτήτων ανίχνευσης και αντίστασης στις επιθέσεις. Έτσι, οι επιστήμονες στοχεύουν να αναπτύξουν μια καινοτόμο προσέγγιση που θα επιτρέπει την αυτόματη ανακάλυψη και ανάλυση επιθέσεων, ενισχύοντας την ασφάλεια των CPS και εντοπίζοντας τυχόν αδυναμίες στις υπάρχουσες δομές άμυνας.

Ο κύριος στόχος της έρευνας είναι να αναπτύξει μια μέθοδο **«smart fuzzing»**, η οποία θα επιτρέπει τη δημιουργία «δοκιμαστικών σειρών» επιθέσεων για την αξιολόγηση των μηχανισμών άμυνας χωρίς την ανάγκη προηγούμενης γνώσης των φυσικών διαδικασιών των συστημάτων. Αυτή η μέθοδος στοχεύει στην κατανόηση των επιδράσεων των ενεργοποιητών στην κατάσταση του συστήματος, προσφέροντας μια πρακτική λύση για την πειραματική αξιολόγηση νέων μηχανισμών άμυνας. Μέσω αυτής της προσέγγισης, οι επιστήμονες φιλοδοξούν να συμβάλουν στην ανάπτυξη πιο ασφαλών και ανθεκτικών CPS, επιτρέποντας τη βελτίωση των στρατηγικών άμυνας και την ενίσχυση της ανθεκτικότητας αυτών των κρίσιμων υποδομών.

Για να επιτύχουν τους στόχους τους, οι επιστήμονες εφαρμόζουν μια σειρά μεθόδων μηχανικής μάθησης και αναζήτησης.

1. Στο πρώτο βήμα της διαδικασίας, αναπτύσσεται ένα μοντέλο του CPS που χρησιμοποιεί δεδομένα από τις μετρήσεις των αισθητήρων για να προβλέψει τις επιδράσεις των ρυθμίσεων των ενεργοποιητών στην φυσική κατάσταση του συστήματος. Αυτό το μοντέλο δέχεται ως είσοδο τις τρέχουσες μετρήσεις και τις ρυθμίσεις των ενεργοποιητών, και επιστρέφει τις αναμενόμενες μετρήσεις μετά από ένα καθορισμένο χρονικό διάστημα, επιτρέποντας έτσι την ανάλυση των πιθανών ρυθμίσεων που μπορούν να οδηγήσουν το σύστημα σε επικίνδυνες καταστάσεις.
2. Στο δεύτερο βήμα, οι επιστήμονες εφαρμόζουν αλγόριθμους αναζήτησης, όπως τυχαία αναζήτηση και γενετικούς αλγόριθμους (GA), για να βρουν τις ρυθμίσεις των ενεργοποιητών που θα μεγιστοποιήσουν την πιθανότητα εμφάνισης μη ασφαλών καταστάσεων. Χρησιμοποιούνται συναρτήσεις προσαρμογής (fitness functions) [79] για να ποσοτικοποιηθεί πόσο κοντά είναι οι ρυθμίσεις στις επικίνδυνες καταστάσεις. Η αναζήτηση πραγματοποιείται με στόχο να εντοπιστούν οι ρυθμίσεις που προσεγγίζουν καλύτερα τις επικίνδυνες καταστάσεις, επιτρέποντας τη δοκιμή και αξιολόγηση των μηχανισμών ανίχνευσης επιθέσεων.

Αυτή η καινοτόμος μέθοδος «smart fuzzing» παρέχει στους ερευνητές τη δυνατότητα να ανακαλύπτουν επιθέσεις που μπορούν να παρακάμψουν τις υπάρχουσες άμυνες, ενισχύοντας έτσι την ασφάλεια των κυβερνοφυσικών συστημάτων. Επιπλέον, τα αποτελέσματα των πειραμάτων σε πραγματικά testbeds, όπως το Secure Water Treatment (SWaT) [80] και το Water Distribution (WADI) [81], επιβεβαίωσαν την αποτελεσματικότητα της μεθόδου, αναδεικνύοντας την ικανότητά της να εντοπίζει επικίνδυνες καταστάσεις και να αξιολογεί τις άμυνες σε πραγματικές συνθήκες. Μέσω της μελέτης αυτής, οι επιστήμονες επιδιώκουν να προσφέρουν μια πρακτική λύση για την αξιολόγηση της ασφάλειας των CPS, ανοίγοντας τον δρόμο για τη μελλοντική έρευνα στον τομέα αυτό.

Στο πεδίο που αφορά τις βιβλιοθήκες του λογισμικού [82], το **NEUZZ** αποτελεί μια καινοτόμος προσέγγιση για την ανίχνευση ευπαθειών μέσω της μεθόδου fuzzing, ενσωματώνοντας τεχνητή νοημοσύνη και νευρωνικά δίκτυα (NNs) για τη βελτίωση της διαδικασίας. Οι παραδοσιακές μέθοδοι fuzzing βασίζονται σε αλγόριθμους που συνεχώς εξελίσσονται που χρησιμοποιούν τυχαίες μεταλλάξεις για τη δημιουργία εισόδων, οι οποίες συχνά αποτυγχάνουν να εντοπίσουν πολύπλοκα σφάλματα, καθώς επικεντρώνονται στις ίδιες περιοχές του κώδικα. Οι ερευνητές του NEUZZ αναγνώρισαν ότι η χρήση νευρωνικών δικτύων για τη δημιουργία ενός ομαλού (διαφορίσιμου) μοντέλου του προγράμματος θα μπορούσε να βελτιώσει σημαντικά αυτή την προσέγγιση, επιτρέποντας την εφαρμογή τεχνικών βελτιστοποίησης καθοδηγούμενων από βαθμίδες (gradient-guided optimization) [83]. Αυτές οι τεχνικές είναι ευρέως χρησιμοποιούμενες στη μηχανική μάθηση και επιτρέπουν την αποτελεσματική εξερεύνηση νέων τμημάτων κώδικα που προηγουμένως είχαν παραμεληθεί.

Οι στόχοι της έρευνας για το NEUZZ περιλάμβαναν την αύξηση της κάλυψης κώδικα, την ανακάλυψη περισσότερων σφαλμάτων, τη βελτίωση της ταχύτητας και αποδοτικότητας, καθώς και την καθοδηγούμενη δημιουργία εισόδων. Οι ερευνητές χρησιμοποίησαν νευρωνικά δίκτυα για την εξομάλυνση της συμπεριφοράς του προγράμματος, εκπαιδεύοντας το NN ώστε να προβλέπει τη συμπεριφορά του κώδικα, όπως η κάλυψη των ακμών του ελέγχου ροής. Η εκπαίδευση του νευρωνικού δικτύου βασίστηκε σε προκαθορισμένα σύνολα δεδομένων, χρησιμοποιώντας δυαδική διασταυρούμενη εντροπία για τον υπολογισμό των αποκλίσεων μεταξύ των προβλεπόμενων και πραγματικών αποτελεσμάτων.

Το NEUZZ εφαρμόζει τεχνικές βελτιστοποίησης όπως ο αλγόριθμος gradient descent, ο οποίος καθοδηγείται από τις κλίσεις που υπολογίζει το νευρωνικό δίκτυο. Κατά τη διάρκεια της διαδικασίας fuzzing, το σύστημα παράγει μεταλλάξεις εισόδου εστιάζοντας στα bytes με τις υψηλότερες τιμές κλίσης, καθώς είναι πιο πιθανό να προκαλέσουν αλλαγές στη συμπεριφορά του προγράμματος. Η διαδικασία συνεχίζεται με την επανεκπαίδευση του μοντέλου, καθώς ανακαλύπτονται νέες συμπεριφορές.

Τα πειράματα που πραγματοποιήθηκαν έδειξαν ότι το NEUZZ ξεπερνά σημαντικά άλλους σύγχρονους fuzzers. Συγκεκριμένα κατάφερε να ανακαλύψει 31 νέες ευπάθειες σε πραγματικά προγράμματα και να επιτύχει πολύ μεγαλύτερη κάλυψη κώδικα. Σε σύγκριση με άλλα εργαλεία, το NEUZZ εντόπισε περισσότερες ακμές του κώδικα κατά τις πρώτες ώρες εκτέλεσης και πέτυχε έως και 4 φορές μεγαλύτερη κάλυψη ακμών από τα παραδοσιακά fuzzers. Ιδιαίτερα, για προγράμματα όπως το readelf [84]και το objdump [85], το NEUZZ εντόπισε περισσότερες νέες ακμές στην πρώτη ώρα από ό,τι άλλοι fuzzers κατά την εκτέλεση 24 ωρών.

Η ικανότητα του NEUZZ να εντοπίζει και να στοχεύει κρίσιμες περιοχές του κώδικα μέσω της καθοδήγησης από τις κλίσεις, σε συνδυασμό με το χαμηλό υπολογιστικό κόστος των νευρωνικών δικτύων, του επέτρεψε να κλιμακωθεί αποτελεσματικά και σε μεγαλύτερα προγράμματα. Αυτή η προσέγγιση καθιστά το NEUZZ μια εξαιρετικά αποδοτική λύση σε σύγκριση με άλλα fuzzing εργαλεία που χρησιμοποιούν βαρύτερες τεχνικές, όπως η συμβολική εκτέλεση. Συνολικά, το NEUZZ

αποδεικνύει ότι η χρήση νευρωνικών δικτύων για την καθοδήγηση του fuzzing μπορεί να οδηγήσει σε πιο αποδοτικές διαδικασίες αναζήτησης σφαλμάτων, επιτυγχάνοντας μεγαλύτερη κάλυψη κώδικα και ταχύτερο εντοπισμό ευπαθειών σε σύγκριση με τις παραδοσιακές τεχνικές.

Στην τελευταία περίπτωση που μελετάμε, δηλαδή σε απλά προγράμματα υπολογιστή [86], βλέπουμε ένα εργαλείο με το όνομα **«DEEPFUZZ»** το οποίο σχεδιάστηκε για να βελτιώσει τη διαδικασία ανίχνευσης σφαλμάτων στους μεταγλωττιστές (compilers) όπως ο GCC και το Clang, χρησιμοποιώντας τεχνικές τεχνητής νοημοσύνης. Οι επιστήμονες που ανέπτυξαν το εργαλείο αυτό επικεντρώθηκαν στη βελτίωση της αξιοπιστίας των μεταγλωττιστών, οι οποίοι αποτελούν βασικά εργαλεία για τη σωστή και ασφαλή λειτουργία των προγραμμάτων. Η πρόκληση με τις υπάρχουσες μεθόδους ελέγχου, όπως το fuzz testing, είναι ότι συχνά δημιουργούνται μη έγκυρα ή ατελή προγράμματα, τα οποία δεν βοηθούν στην ανίχνευση σφαλμάτων σε βάθος. Επομένως, οι ερευνητές ήθελαν να δημιουργήσουν μια πιο αποδοτική λύση, η οποία θα μπορούσε αυτόματα να παράγει προγράμματα που ακολουθούν τη συντακτική γραμματική της γλώσσας προγραμματισμού C.

Ο βασικός στόχος τους ήταν να δημιουργήσουν συντακτικά σωστά προγράμματα που θα μπορούσαν να χρησιμοποιηθούν για τη βελτίωση της κάλυψης κώδικα κατά τον έλεγχο των μεταγλωττιστών. Αυτό θα βοηθούσε στην ενίσχυση της αποτελεσματικότητας του fuzz testing, ανιχνεύοντας σφάλματα σε μέρη του κώδικα που μπορεί να μην ελέγχονται επαρκώς με τις υπάρχουσες μεθόδους. Η προσέγγιση αυτή επέτρεψε την ανίχνευση νέων σφαλμάτων, όπως απέδειξαν οι ερευνητές όταν ανακάλυψαν οκτώ νέα σφάλματα στον GCC, τα οποία στη συνέχεια επιβεβαιώθηκαν και διορθώθηκαν.

Για την επίτευξη αυτού του στόχου, χρησιμοποιήθηκε ένα Sequence-to-Sequence μοντέλο τεχνητής νοημοσύνης βασισμένο σε LSTM (Long Short-Term Memory) μονάδες. Το μοντέλο εκπαιδεύτηκε σε ένα σύνολο 10.000 συντακτικά σωστών προγραμμάτων C, τα οποία είχαν ληφθεί από τις σουίτες δοκιμών του GCC. Μετά από 50 εποχές εκπαίδευσης, το DEEPFUZZ κατάφερε να παράγει προγράμματα με ποσοστό επιτυχίας πάνω από 82%. Αυτή η μέθοδος επιτρέπει τη συνεχή και αυτόματη δημιουργία νέων προγραμμάτων, βελτιώνοντας έτσι την κάλυψη του κώδικα και ενισχύοντας την ικανότητα ανίχνευσης σφαλμάτων στους μεταγλωττιστές, συμβάλλοντας σημαντικά στην αξιοπιστία τους.

Ολοκληρώνοντας το «Automated Vulnerability Detection», αξιόλογο είναι να αναφερθούμε σε μεθόδους που ενσωματώνουν ενισχυτική μάθηση και επικεντρώνονται είτε σε μεγάλα δίκτυα είτε σε μικροδίκτυα [87]. Οι επιστήμονες, επιδιώκοντας την αυτόνομη διενέργεια δοκιμών διείσδυσης (penetration testing) για τον εντοπισμό ευπαθειών και τη βελτίωση της ασφάλειας δικτύων, θεώρησαν ότι η ενίσχυση μάθησης (Reinforcement Learning) [88] μπορεί να προσφέρει σημαντικά οφέλη, καθώς θα επιτρέπει συστηματικές δοκιμές με μικρή ανθρώπινη παρέμβαση. Σε αντίθεση με τις παραδοσιακές μεθόδους ελέγχου ασφάλειας που απαιτούν εξειδικευμένες γνώσεις και σημαντικούς πόρους, η αυτόνομη προσέγγιση μπορεί να εκτελείται συχνά, υποστηρίζοντας έτσι πιο ανθεκτικές στρατηγικές ασφάλειας.

Ως κεντρικό στόχο, έθεσαν τη μοντελοποίηση του προβλήματος του αυτόνομου ελέγχου ασφαλείας ως Πρόβλημα Μαρκοβιανής Απόφασης (MDP) [89], όπου κάθε κατάσταση του δικτύου απαιτεί συγκεκριμένες αποφάσεις από τον πράκτορα (agent) που εκπαιδεύεται να αναγνωρίζει και να εκμεταλλεύεται πιθανές αδυναμίες ασφάλειας.

Για την υλοποίηση αυτού του στόχου, ανέπτυξαν τον αλγόριθμο **NDSPI-DQN**, μια βελτιωμένη εκδοχή του αλγόριθμου DQN (Deep Q-Network) [90], που ενσωματώνει τεχνικές για την ενίσχυση της εξερεύνησης, μειώνοντας τα προβλήματα που σχετίζονται με τις σπάνιες ανταμοιβές (sparse rewards) και τον μεγάλο χώρο δράσεων που παρατηρείται σε μεγάλα δίκτυα. Για την αντιμετώπιση των δυσκολιών που παρουσιάζει η αργή σύγκλιση των αποτελεσμάτων, ο αλγόριθμος ενισχύθηκε με πέντε τεχνολογίες, οι οποίες βελτιώνουν τη λειτουργία και την απόδοση του πράκτορα: Soft Q-learning για βελτίωση της εξερεύνησης, Dueling Network για αποτελεσματικότερη εκτίμηση πολιτικής, Prioritized Experience Replay για εστίαση στις πιο σημαντικές εμπειρίες, Noisy Nets για θορυβώδη διερεύνηση, και το Intrinsic Curiosity Module (ICM), που επιβραβεύει τον πράκτορα όταν εξερευνά άγνωστα περιβάλλοντα με λίγες εξωτερικές ανταμοιβές.

Η μείωση του χώρου δράσεων επιτεύχθηκε με την αποκωδικοποίηση του επιθετικού φορέα, επιτρέποντας στον πράκτορα να επιλέγει στόχους και δράσεις ξεχωριστά, περιορίζοντας σημαντικά τις επιλογές και αυξάνοντας την ταχύτητα σύγκλισης. Ο αλγόριθμος μαθαίνει έτσι συνδυαστικά από εσωτερικές και εξωτερικές ανταμοιβές, ενισχύοντας την εξερεύνηση σε ρεαλιστικά σενάρια δικτύων και μειώνοντας το κόστος των δοκιμών. Η πλατφόρμα NASim [91] χρησιμοποιήθηκε για να δημιουργηθούν διάφορα σενάρια δικτύων προσομοιώνοντας εμπορικά περιβάλλοντα, στα οποία ο πράκτορας μπορεί να εκπαιδευτεί και να δοκιμάσει τις δυνατότητές του σε εικονικά δίκτυα που περιλαμβάνουν honeypots και υπολογιστές με διαφορετικές διαμορφώσεις ασφάλειας.

Τα πειράματα δείχνουν ότι ο βελτιωμένος αλγόριθμος **DQN**, ειδικά στην εκδοχή με την αποκωδικοποίηση του επιθετικού φορέα, έχει καλύτερη σύγκλιση και κλιμάκωση σε μεγάλα και σύνθετα περιβάλλοντα δικτύων, αποδεικνύοντας την ανθεκτικότητά του ακόμα και σε δίκτυα με εκατοντάδες υπολογιστές. Ωστόσο, οι συγγραφείς τονίζουν ότι η εφαρμογή αυτών των μεθόδων σε πραγματικά δίκτυα είναι περιορισμένη λόγω του υψηλού κόστους και της δυσκολίας προσομοίωσης της πραγματικής κυκλοφορίας δεδομένων, γεγονός που καθιστά απαραίτητη τη διεξαγωγή των δοκιμών σε προσομοιωμένα περιβάλλοντα. Μελλοντικά, η ενσωμάτωση τεχνολογιών εικονικοποίησης και η ανάπτυξη προηγμένων αλγορίθμων, όπως multi-agent [92] και hierarchical RL [93], θα μπορούσαν να βοηθήσουν στην εφαρμογή σε πιο ρεαλιστικά και πολύπλοκα σενάρια, βελτιώνοντας την αποτελεσματικότητα των αυτόνομων πρακτόρων στην ανίχνευση και αποτροπή κυβερνοεπιθέσεων.

Σε επόμενη έρευνα [94], οι επιστήμονες σχεδίασαν μια μεθοδολογία που θα εντοπίζει και θα προστατεύει τα «κοσμήματα του στέμματος» (Crown Jewels - CJs) των πληροφοριακών συστημάτων ενός οργανισμού, δηλαδή τα πιο σημαντικά και ευαίσθητα δεδομένα ή υποδομές, από κυβερνοεπιθέσεις. Καθώς οι παραδοσιακές τεχνικές ανίχνευσης ευπαθειών δεν εστιάζουν πάντα σε στρατηγικούς κόμβους και διαδρομές που είναι ιδιαίτερα ευάλωτες, η ομάδα επιδίωξε να χρησιμοποιήσει πιο προχωρημένες μεθόδους που θα επιτρέπουν την κατανόηση και ανάλυση του δικτύου με γνώμονα τον τρόπο που σκέφτονται και δρουν οι επιτιθέμενοι. Έτσι, θέλησαν να αναπτύξουν μια αυτόματη μέθοδο που, με τη χρήση γραφημάτων επίθεσης και ενισχυτικής μάθησης, θα εντοπίζει ευάλωτα σημεία και διαδρομές προς τα CJs, επιτρέποντας τη βελτίωση των στρατηγικών προστασίας.

Ο κύριος στόχος της μελέτης είναι να εντοπιστούν τα πιο κρίσιμα σημεία και διαδρομές σε ένα δίκτυο που είναι πιο πιθανό να χρησιμοποιηθούν από έναν επιτιθέμενο για πρόσβαση στα CJs. Αυτά τα σημεία αποτελούν στρατηγικούς στόχους που, αν δεν προστατευτούν, μπορεί να εκθέσουν όλο το σύστημα σε μεγάλες απειλές. Η μεθοδολογία στοχεύει στην αυτόματη ανίχνευση και ανάλυση αυτών των στρατηγικών σημείων, προσφέροντας μια ακριβή απεικόνιση των «σημείων πνιγμού» (choke points) και των συχνότερων διαδρομών που ακολουθούνται από επιτιθέμενους. Μέσω της

ενισχυτικής μάθησης, επιδιώκεται να αναπτυχθούν βέλτιστες διαδρομές για αμυντική παρακολούθηση ή να σχεδιαστούν στρατηγικές ανίχνευσης και αποτροπής σε κομβικά σημεία του δικτύου.

Η μεθοδολογία Crown Jewel Analysis - Reinforcement Learning (CJA-RL) βασίζεται στην ενισχυτική μάθηση, ειδικότερα σε αλγορίθμους Q-learning, και στη χρήση γραφημάτων επίθεσης (attack graphs).

1. Αρχικά, δημιουργείται μια Μαρκοβιανή Διαδικασία Αποφάσεων (Markov Decision Process - MDP), η οποία περιλαμβάνει όλους τους δυνατούς κόμβους και διαδρομές προς τα CJs.
2. Χρησιμοποιώντας το εργαλείο MulVal [95], κατασκευάζεται ένα γράφημα επίθεσης, το οποίο αποτυπώνει το κυβερνο-έδαφος του δικτύου και τις διαδρομές απόκτησης πρόσβασης.
3. Ο παράγοντας της ενισχυτικής μάθησης (RL agent) εκπαιδεύεται να αναλύει τις διαδρομές που οδηγούν προς τα CJs, επιλέγοντας την πιο αποδοτική διαδρομή και αξιολογώντας τους κόμβους ανάλογα με την «αμοιβή» (reward) που προσφέρει η επιτυχής πρόσβαση χωρίς ανίχνευση.

Τα αποτελέσματα δείχνουν πως οι κόμβοι με τις λιγότερες μεταβάσεις και την υψηλότερη αμοιβή είναι καίριοι για αμυντική παρακολούθηση, καθώς και ότι οι πιο συχνά χρησιμοποιούμενοι κόμβοι αποτελούν ιδανικά σημεία για την τοποθέτηση συστημάτων ανίχνευσης.

Τέλος ας δούμε και την εφαρμογή της Ενισχυτικής Μάθησης (Reinforcement Learning - RL) ως εργαλείο για την αυτοματοποίηση επιθέσεων Δοκιμών Διείσδυσης (Penetration Testing - PT) σε αλγόριθμους ελέγχου μικροδικτύων (MG) [96]. Οι επιστήμονες προσεγγίζουν το πρόβλημα της ασφάλειας των (MG) ως κρίσιμο για τη λειτουργία των σύγχρονων ενεργειακών συστημάτων, καθώς τα MG βασίζονται σε έξυπνα συστήματα ελέγχου που μπορούν να επηρεαστούν από κυβερνοεπιθέσεις. Ειδικότερα, αναγνωρίζουν πως η παραποίηση των δεδομένων στο σύστημα ελέγχου των MG μπορεί να οδηγήσει σε αναποτελεσματική διαχείριση της ενέργειας και αυξημένο λειτουργικό κόστος. Ως εκ τούτου, θεωρούν πως η ανάπτυξη και εφαρμογή μεθόδων που αποκαλύπτουν ευπάθειες των αλγορίθμων ελέγχου των MG είναι απαραίτητη για την αποφυγή τέτοιων κινδύνων.

Ο στόχος της έρευνας είναι η αξιοποίηση της Ενισχυτικής Μάθησης (RL) ως εργαλείο αυτοματοποίησης των Δοκιμών Διείσδυσης (PT) στο σύστημα ελέγχου του MG, προκειμένου να εντοπιστούν οι κρυφές αδυναμίες του αλγορίθμου και να αναδειχθούν τα τρωτά σημεία του συστήματος.

Η μελέτη προτείνει τη χρήση ενός κακόβουλου πράκτορα RL, ο οποίος εκπαιδεύεται για να αναγνωρίζει στρατηγικές επιθέσεων που αποδυναμώνουν τον αλγόριθμο ελέγχου του MG, ειδικά στο σημείο που αφορά τη χρήση της ενέργειας από την μπαταρία του δικτύου. Ο πράκτορας RL τροποποιεί την αναφερόμενη κατάσταση φόρτισης της μπαταρίας (State of Charge - SOC) που λαμβάνει ο ελεγκτής του MG, επηρεάζοντας έτσι τις αποφάσεις του συστήματος ως προς την αποθήκευση και χρήση της ενέργειας, με αποτέλεσμα την αύξηση του συνολικού κόστους λειτουργίας. Ο συγκεκριμένος αλγόριθμος ελέγχου βασίζεται σε μια μέθοδο μικτής γραμμικής βελτιστοποίησης (MILP) που βελτιστοποιεί τις αποφάσεις του MG για αποδοτική ενεργειακή διαχείριση.

Για την εκπαίδευση του πράκτορα RL, εφαρμόζεται ο αλγόριθμος **Advantage Actor-Critic** [97], που δίνει τη δυνατότητα στον πράκτορα να επιλέγει συγκεκριμένες τιμές εισόδου SOC ώστε να αποδιοργανώνει τη λειτουργία του συστήματος.

Οι τιμές εισόδου SOC (State of Charge) αναφέρονται στην κατάσταση φόρτισης μιας μπαταρίας και εκφράζουν το ποσοστό της ενέργειας που είναι αποθηκευμένη σε σχέση με τη συνολική

χωρητικότητα της μπαταρίας. Το SOC είναι ένα κρίσιμο μέτρο για την παρακολούθηση και τη διαχείριση των μπαταριών, καθώς επηρεάζει την απόδοση και τη λειτουργία συστημάτων που χρησιμοποιούν αποθηκευμένη ενέργεια, όπως είναι τα μικροδίκτυα (MG).

Μέσω μιας σειράς από επεισόδια προσομοίωσης, ο πράκτορας αναπτύσσει μια στρατηγική επιλογής ψευδών τιμών SOC που παραπλανά τον αλγόριθμο, μειώνοντας την ενεργειακή απόδοση του συστήματος και οδηγώντας σε αυξημένο κόστος. Ο πράκτορας εκπαιδεύεται με απόλυτη γνώση της προσομοιωμένης λειτουργίας, στοχεύοντας σε παραπλανητικές αποφάσεις με το να εισάγει τιμές που αναφέρουν χαμηλότερη φόρτιση της μπαταρίας από την πραγματική.

Τα αποτελέσματα δείχνουν ότι οι τροποποιήσεις SOC έχουν ουσιαστικό αντίκτυπο στην αποδοτικότητα του ελεγκτή, με τις μεγαλύτερες αλλαγές να προκαλούν τη μεγαλύτερη αύξηση στο συνολικό κόστος λειτουργίας. Η μελέτη καταλήγει στο ότι η RL μπορεί να αξιοποιηθεί για τον εντοπισμό σημαντικών τρωτοτήτων, προσφέροντας πολύτιμες πληροφορίες για τη βελτίωση της ασφάλειας όχι μόνο των μικροδικτύων αλλά και άλλων συστημάτων κρίσιμων υποδομών.

2.3.2 Αυτοματοποιημένη Ταξινόμηση Ευπαθειών

Η αυτοματοποιημένη ταξινόμηση ευπαθειών (Automated Vulnerability Classification) είναι ένα σημαντικό βήμα για την κατανόηση των πληροφοριών ασφαλείας, το οποίο επιταχύνει τη διαδικασία αξιολόγησης των κινδύνων. Ερευνητές εργάζονται πάνω σε συστήματα που αυτοματοποιούν την ταξινόμηση και την ετικετοποίηση περιγραφών ευπαθειών που παρουσιάζονται σε αναφορές ασφαλείας.

Για παράδειγμα, ο Russo και η ομάδα του [98] πρότειναν μια μέθοδο για να συνοψίζουν τις καθημερινές αναφορές ευπαθειών και να τις κατηγοριοποιούν σύμφωνα με ένα συγκεκριμένο ταξινομικό μοντέλο για τη βιομηχανία. Οι επιστήμονες αυτοί που ανέπτυξαν το εργαλείο **CVErizer** αποσκοπούσαν στη δημιουργία μιας αυτόματης μεθόδου ανάλυσης και κατηγοριοποίησης των ευπαθειών (CVE) με στόχο τη βελτίωση της διαχείρισης κυβερνοασφάλειας σε οργανισμούς. Το κίνητρο για την ανάπτυξη του εργαλείου προέκυψε από την αυξανόμενη πολυπλοκότητα και τον όγκο των διαθέσιμων πληροφοριών ευπάθειας, κάτι που καθιστά δύσκολη τη γρήγορη αξιολόγηση και διαχείρισή τους. Οι ερευνητές αναγνώρισαν ότι οι παραδοσιακές μέθοδοι που απαιτούν την ανάλυση και επεξεργασία μεγάλου όγκου δεδομένων από ανθρώπους ήταν χρονοβόρες και επιρρεπείς σε λάθη. Έτσι, σκοπός τους ήταν να μειώσουν το χρόνο ανάλυσης και να βοηθήσουν τις ομάδες ασφαλείας να εντοπίζουν και να προτεραιοποιούν τις ευπάθειες πιο αποδοτικά, κάτι που θα μπορούσε να συμβάλει στη μείωση των κινδύνων από κυβερνοεπιθέσεις.

Για την επίτευξη του στόχου αυτού, οι επιστήμονες χρησιμοποίησαν τεχνητή νοημοσύνη, συνδυάζοντας μεθόδους επεξεργασίας φυσικής γλώσσας (NLP) και αλγόριθμους μηχανικής μάθησης. Η επεξεργασία φυσικής γλώσσας χρησιμοποιήθηκε για να εντοπίζει αυτόματα σημαντικές πληροφορίες από τις περιγραφές των CVEs, όπως το όνομα και τις εκδόσεις του λογισμικού που επηρεάζεται, το είδος και τον μηχανισμό της επίθεσης, τις πιθανές συνέπειες της ευπάθειας, και τον τύπο του επιτιθέμενου που μπορεί να την εκμεταλλευτεί. Η ανάλυση γραμματικής δομής έγινε μέσω του εργαλείου Stanford Typed Dependencies [99], που βοήθησε στην εξαγωγή γλωσσικών μοτίβων από τις περιγραφές, καθώς και στη δημιουργία ευρετικών κανόνων που καταγράφουν τα δεδομένα αυτά σε μορφή XML. Παράλληλα, χρησιμοποιήθηκαν τεχνικές μηχανικής μάθησης, όπως οι αλγόριθμοι J48, Random Forest, BayesNet, και Simple Logistic, για την κατηγοριοποίηση των

ευπαθειών σε δέκα κύριες κατηγορίες, κάτι που διευκολύνει τις ομάδες ασφαλείας στην αξιολόγηση και προτεραιοποίηση των ευπαθειών.

Για την αξιολόγηση του CVErizer πραγματοποιήθηκαν πειράματα με φοιτητές και επαγγελματίες κυβερνοασφάλειας, τα οποία έδειξαν ότι το εργαλείο ήταν αρκετά ακριβές τόσο στην εξαγωγή πληροφοριών όσο και στην κατηγοριοποίησή τους. Η ακρίβεια των αποτελεσμάτων αναλύθηκε με δείκτες πληροφοριακής ανάκτησης (Precision, Recall, και F-Measure), όπου ο αλγόριθμος BayesNet [100] διακρίθηκε για την καλύτερη απόδοση στην κατηγοριοποίηση ευπαθειών. Επιπλέον, σε ποιοτική ανάλυση που ακολούθησε, οι συμμετέχοντες θεώρησαν ότι το εργαλείο μειώνει κατά πολύ το χρόνο ανάλυσης των CVEs, προσφέροντας περιεκτικές και κατανοητές περιλήψεις. Με βάση τα ευρήματα, οι επιστήμονες κατέληξαν ότι το CVErizer είναι ένα χρήσιμο εργαλείο για τις ομάδες ασφάλειας, το οποίο μπορεί να ενσωματωθεί σε πραγματικά περιβάλλοντα για την ενίσχυση της ανάλυσης ευπαθειών και τη βελτίωση της κυβερνοασφάλειας οργανισμών.

Σε μία άλλη μελέτη [101], ο Aota και η ομάδα του επικεντρώνονται στη χρήση της εξόρυξης κειμένου για την κατηγοριοποίηση ευπαθειών με βάση τις περιγραφές που παρέχονται στη λίστα των Κοινών Ευπαθειών και Έκθεσης (CVE) [102]. Οι επιστήμονες έχουν αναγνωρίσει τη σημασία της αυτοματοποίησης της κατηγοριοποίησης των ευπαθειών λογισμικού στην κυβερνοασφάλεια, καθώς η αυξανόμενη εξάρτηση από τα πληροφοριακά συστήματα καθιστά την αποτελεσματική διαχείριση των ευπαθειών κρίσιμη. Μέχρι σήμερα, η διαδικασία κατηγοριοποίησης των ευπαθειών, όπως αυτές που καταγράφονται στις κοινές ευπάθειες και εκθέσεις (CVE), γινόταν κυρίως χειροκίνητα, γεγονός που οδήγησε σε σφάλματα λόγω ανθρώπινων παραγόντων και σε περιορισμούς λόγω έλλειψης ειδικών. Στόχος της έρευνας είναι η ανάπτυξη ενός αυτοματοποιημένου εργαλείου που θα χρησιμοποιεί τεχνικές μηχανικής μάθησης για να κατηγοριοποιεί με ακρίβεια τις αναφορές ευπαθειών, μειώνοντας την ανάγκη για ανθρώπινη παρέμβαση και ταυτόχρονα ελαχιστοποιώντας τα σφάλματα που προκύπτουν από αυτήν.

Για να επιτευχθεί αυτός ο στόχος, οι ερευνητές χρησιμοποίησαν αρκετές καινοτόμες μεθόδους της τεχνητής νοημοσύνης (AI).

1. Η διαδικασία ξεκινά με την εφαρμογή του αλγορίθμου **Bag-of-Words (BoW)** [103], ο οποίος μετατρέπει τις περιγραφές των ευπαθειών σε μορφή που μπορεί να κατανοηθεί από υπολογιστές. Αυτό επιτρέπει την εξαγωγή κρίσιμων χαρακτηριστικών που είναι απαραίτητα για την κατηγοριοποίηση.
2. Στη συνέχεια, εφαρμόζεται η μέθοδος **Boruta** [104], η οποία επιλέγει τα πιο χρήσιμα χαρακτηριστικά για την ταξινόμηση, εξετάζοντας τις αλληλεπιδράσεις τους και προσφέροντας πιο αξιόπιστα αποτελέσματα.

Ο κύριος ταξινομητής που χρησιμοποιείται για την κατηγοριοποίηση είναι ο αλγόριθμος Τυχαίων Δασών (Random Forest), ο οποίος έχει αποδείξει την αποτελεσματικότητά του στην κατηγοριοποίηση δεδομένων. Για την αξιολόγηση της απόδοσης των αλγορίθμων, χρησιμοποιήθηκε πενταπλή διασταυρούμενη επικύρωση, που επιτρέπει τη δοκιμή του εκπαιδευμένου συστήματος σε διαφορετικά σύνολα δεδομένων, εξασφαλίζοντας έτσι μια αξιόπιστη εκτίμηση των αποτελεσμάτων. Οι ερευνητές χρησιμοποίησαν επίσης τρεις δείκτες αξιολόγησης (ακρίβεια, F1macro και F1weighted) για να συγκρίνουν την απόδοση των διαφορετικών αλγορίθμων και να επιβεβαιώσουν την αποτελεσματικότητα της προτεινόμενης μεθόδου.

Συνολικά, οι μέθοδοι αυτές παρέχουν μια ολοκληρωμένη προσέγγιση στην αυτοματοποίηση της κατηγοριοποίησης των ευπαθειών, αποδεικνύοντας τη σημασία της μηχανικής μάθησης στην ενίσχυση της κυβερνοασφάλειας.

Ας δούμε τέλος τον Vanamala και τους συνεργάτες του [105] που κατηγοριοποιούν τις καταχωρίσεις CVE σύμφωνα με τους δέκα κορυφαίους κινδύνους του OWASP. Οι επιστήμονες στον τομέα της κυβερνοασφάλειας προσεγγίζουν τη μελέτη των κοινών ευπαθειών και εκθέσεων (CVE) με την αναγνώριση της αυξανόμενης απειλής από κυβερνοεπιθέσεις. Το σκεπτικό τους επικεντρώνεται στην ανάγκη για ενδελεχή ανάλυση και κατηγοριοποίηση των CVE, ώστε να κατανοήσουν τις ευπάθειες που εκμεταλλεύονται οι επιτιθέμενοι. Αυτή η κατανόηση τους επιτρέπει να εντοπίσουν τάσεις και μοτίβα στην κυβερνοασφάλεια, αναγνωρίζοντας τις συχνότερες επιθέσεις και τους πιο κοινούς κινδύνους, προκειμένου να προλάβουν τις επιπτώσεις των επιθέσεων.

Οι στόχοι τους περιλαμβάνουν την αυτοματοποίηση της διαδικασίας ανάλυσης ευπαθειών και την κατηγοριοποίηση των CVE με βάση τους 10 κύριους κινδύνους του OWASP. Επιδιώκουν να δημιουργήσουν ένα σύστημα που θα επιτρέπει την ταχύτερη και πιο αποτελεσματική αναγνώριση ευπαθειών, μειώνοντας έτσι την επιφάνεια επίθεσης για τους οργανισμούς. Μέσω της ανάλυσης των δεδομένων και της εξαγωγής πληροφοριών, στοχεύουν να αναπτύξουν στρατηγικές που θα βοηθήσουν στην ενίσχυση της κυβερνοασφάλειας και στη μείωση των πιθανών ζημιών από κυβερνοεπιθέσεις.

Για την επίτευξη αυτών των στόχων, οι επιστήμονες χρησιμοποιούν προηγμένες μεθόδους τεχνητής νοημοσύνης, όπως η μέθοδος **Latent Dirichlet Allocation (LDA)** [106], η οποία τους επιτρέπει να αναλύουν μεγάλες ποσότητες κειμένων και να εξάγουν κρυφά θέματα που σχετίζονται με τις CVE. Επιπλέον, εφαρμόζουν τεχνικές αντιστοίχισης λέξεων-κλειδιών και προετοιμασίας δεδομένων, όπως η αφαίρεση stopwords, για να βελτιώσουν την ποιότητα των δεδομένων και την ακρίβεια των αναλύσεων. Με αυτές τις μεθόδους, αποσκοπούν στην ενίσχυση των στρατηγικών ασφάλειας και στη δημιουργία ενός πιο ασφαλούς ψηφιακού περιβάλλοντος.

2.3.3 Εξερεύνηση Ευπαθειών

Η εξερεύνηση των ευπαθειών (Vulnerability Exploration) περιλαμβάνει, ως βασικό βήμα, την ανίχνευση πιθανών σημείων επίθεσης που θα μπορούσαν να αξιοποιηθούν από κακόβουλους χρήστες. Αυτό βοηθά στην αποτελεσματική αξιολόγηση και διαχείριση των κινδύνων ασφαλείας. Για να το επιτύχουν, κάποιοι ερευνητές αντλούν δεδομένα από το MITRE, ενώ άλλοι βασίζονται στο CVSS, το οποίο αναπτύχθηκε από το Forum of Incident Response and Security Teams.

Μελέτες όπως των Bakirtzis και Kuppa χρησιμοποιούν μια μοντελοκεντρική προσέγγιση για την αυτόματη αντιστοίχιση των επιθετικών τακτικών και τεχνικών με το σύστημα. Ας ξεκινήσουμε πρώτα με τους επιστήμονες [107] που ανέπτυξαν το εργαλείο **CYBOK**. Ως κύριο στόχο είχαν να ενσωματώσουν την ανάλυση ασφάλειας στο αρχικό στάδιο σχεδίασης συστημάτων, ώστε να μπορούν να αξιολογούν την ασφάλεια των υποψήφιων σχεδιάσεων πριν την κατασκευή τους. Συγκεκριμένα, επιδίωξαν να υποστηρίξουν τη διαδικασία «ασφάλεια από τον σχεδιασμό» (security by design) σε κυβερνοφυσικά συστήματα, όπου οι παραβιάσεις μπορούν να έχουν επικίνδυνες συνέπειες. Το CYBOK επιτρέπει στους αναλυτές να αξιολογούν διαφορετικές σχεδιάσεις, να εντοπίζουν τις πιο ευάλωτες περιοχές, και να προτείνουν αλλαγές πριν από την υλοποίηση.

Οι βασικοί στόχοι των επιστημόνων εστιάζουν στην ενίσχυση της ασφάλειας των συστημάτων μέσω της πρόληψης και της μείωσης των ευπαθειών. Αρχικά, επιδιώκουν την αξιολόγηση της ευπάθειας των συστημάτων σε πρώιμο στάδιο, αναπτύσσοντας μεθόδους που επιτρέπουν την ανίχνευση πιθανών διαδρομών επίθεσης και αδύναμων σημείων πριν την κατασκευή τους. Παράλληλα, στοχεύουν στη μείωση της επιφάνειας επίθεσης, βοηθώντας τους μηχανικούς να αναλύσουν και να περιορίσουν τα σημεία πρόσβασης, βελτιώνοντας έτσι την ασφάλεια. Τέλος, προσφέρουν εφαρμόσιμα μέτρα προστασίας, παρέχοντας κρίσιμες πληροφορίες για τις αλυσίδες εκμετάλλευσης και τις ευπάθειες, διευκολύνοντας τη λήψη αποφάσεων για τα απαραίτητα προστατευτικά μέτρα.

Η βασική μέθοδος που εφαρμόστηκε ήταν η δημιουργία ενός γραφικού μοντέλου συστήματος (System Model) που περιλάμβανε τα κύρια στοιχεία του συστήματος και τα διανύσματα επίθεσης. Χρησιμοποιώντας τα διαθέσιμα δεδομένα και ευπάθειες από βιβλιοθήκες όπως το CAPEC και το CVE, το CYBOK εντόπιζε δυνητικές αλυσίδες εκμετάλλευσης και διαδρομές που ένας επιτιθέμενος θα μπορούσε να χρησιμοποιήσει για να παραβιάσει το σύστημα. Συγκεκριμένα, οι ερευνητές ακολούθησαν τα εξής βήματα:

1. **Δημιουργία του Γραφικού Μοντέλου Συστήματος**: Χρησιμοποιώντας τη γλώσσα SysML [108], δημιούργησαν μια αναπαράσταση του συστήματος με όλα τα βασικά στοιχεία του UAS και τα περιγραφικά χαρακτηριστικά των συνδέσεών του. Στόχος ήταν να απεικονίσουν τη λειτουργία του συστήματος και τα πιθανά σημεία πρόσβασης για τους επιτιθέμενους.

2. **Επέκταση της Επιφάνειας Επίθεσης**: Εισήγαγαν τις περιγραφικές λέξεις-κλειδιά (keywords) για κάθε σημείο εισόδου που συνδέεται με διανύσματα επίθεσης, ώστε να αναγνωριστούν όλα τα πιθανά σημεία εκμετάλλευσης.

3. **Ανάλυση Αλυσίδων Εκμετάλλευσης (Exploit Chains)** [109]: Το CYBOK χρησιμοποίησε πληροφορίες από γνωστά διανύσματα επίθεσης (π.χ. CVE, CWE, CAPEC) για να χαρτογραφήσει όλες τις δυνητικές διαδρομές που θα μπορούσαν να ακολουθήσουν οι επιτιθέμενοι για να φτάσουν από ένα σημείο εισόδου έως τον κύριο επεξεργαστή εφαρμογών του συστήματος, ο οποίος είναι κρίσιμος για την αποστολή. Οι αλυσίδες αυτές αναλύθηκαν για να αποκαλύψουν τις πιο κρίσιμες ευπάθειες.

4. **Δοκιμές «Τι θα γινόταν αν»**: Οι αναλυτές μπορούσαν να τροποποιήσουν το μοντέλο για να αξιολογήσουν τις επιπτώσεις των αλλαγών, όπως η αντικατάσταση ενός συγκεκριμένου ραδιοεπικοινωνιακού πρωτοκόλλου (π.χ. ZigBee με το XBee) με ένα πιο ασφαλές. Με αυτόν τον τρόπο, εντόπισαν πώς κάθε σχεδιαστική απόφαση επηρέαζε την επιφάνεια επίθεσης και την ασφάλεια του συστήματος.

Το CYBOK ενσωματώνει στοιχεία ανάλυσης δεδομένων και εξόρυξης γνώσης από μεγάλες βάσεις δεδομένων ευπάθειας, όπως οι CAPEC και CVE. Οι τεχνικές αυτές επιτρέπουν στο σύστημα να ανιχνεύει αυτόματα τα πιο κρίσιμα διανύσματα επίθεσης, καθώς και να επιλέγει τις κατάλληλες διαδρομές και αλυσίδες εκμετάλλευσης. Επιπλέον, η προσέγγιση που βασίζεται σε γραφήματα διευκολύνει την μοντελοκεντρική ανάλυση, η οποία παρέχει έναν διαδραστικό τρόπο ανίχνευσης επιθέσεων και υπολογισμού ευπαθειών σε επίπεδο σχεδίασης, χωρίς την ανάγκη πλήρως αναπτυγμένου συστήματος.

Συνολικά, το CYBOK δίνει τη δυνατότητα στους αναλυτές να κατανοήσουν τις αλληλεπιδράσεις των στοιχείων του συστήματος και τα πιθανά σημεία επίθεσης, επιτρέποντας έτσι τη λήψη στρατηγικών αποφάσεων για την ασφάλεια του συστήματος, εξοικονομώντας χρόνο και πόρους.

Οι επιστήμονες σε μια άλλη μελέτη [110], επιχείρησαν να αντιμετωπίσουν ένα κρίσιμο πρόβλημα στην κυβερνοασφάλεια: την αυτόματη και ακριβή αντιστοίχιση ευπαθειών (CVE) με τις τεχνικές ATT&CK, δηλαδή τις συγκεκριμένες τεχνικές που χρησιμοποιούνται σε επιθέσεις και περιγράφονται στη γνωσιακή βάση ATT&CK της MITRE. Η διαδικασία αυτή αποσκοπεί στην κατηγοριοποίηση των ευπαθειών, βοηθώντας τους επαγγελματίες ασφαλείας να αναγνωρίζουν γρήγορα τις τεχνικές που είναι πιθανό να εκμεταλλευτούν τις συγκεκριμένες ευπάθειες. Μέχρι πρότινος, η διαδικασία αυτή ήταν χειροκίνητη και χρονοβόρα, ενώ λόγω του μεγάλου όγκου των δεδομένων καθίστατο συχνά αναποτελεσματική.

Η ομάδα επιστημόνων είχε ως κύριους στόχους την αυτοματοποίηση, την ακρίβεια και την κατανόηση των ευπαθειών. Αρχικά, επιδίωξαν την αυτόματη αντιστοίχιση των CVE (Common Vulnerabilities and Exposures) με τις τεχνικές ATT&CK, εξαλείφοντας την ανάγκη για χειροκίνητη επεξεργασία. Παράλληλα, μέσω της χρήσης τεχνικών τεχνητής νοημοσύνης (AI) και επεξεργασίας φυσικής γλώσσας (NLP), στοχεύουν στην αύξηση της ακρίβειας, περιορίζοντας τις ασάφειες στις περιγραφές CVE και στις αναφορές ασφαλείας. Τελικός στόχος είναι η ενίσχυση της κατανόησης των ευπαθειών και των σχετικών κινδύνων, προκειμένου να διευκολυνθεί η εφαρμογή κατάλληλων μέτρων μετριασμού και να βελτιωθεί η ασφάλεια των συστημάτων.

Οι επιστήμονες βασίστηκαν σε μια σειρά από σύγχρονες τεχνικές Τεχνητής Νοημοσύνης και Επεξεργασίας Φυσικής Γλώσσας (NLP), διαμορφώνοντας ένα πολυεπίπεδο μοντέλο κοινής ενσωμάτωσης που να είναι ικανό να "κατανοεί" τη σχέση μεταξύ των περιγραφών ευπαθειών (CVE) και των τεχνικών ATT&CK.

Μέθοδοι Τεχνητής Νοημοσύνης που Χρησιμοποιήθηκαν:

1. **Word2Vec**: Χρησιμοποιήθηκε για τη δημιουργία ενσωματώσεων λέξεων από τις περιγραφές των CVE. Οι ενσωματώσεις αυτές βοήθησαν στη διαμόρφωση ενός λεξιλογίου, επιτρέποντας στο μοντέλο να «αντιλαμβάνεται» τη σημασία των λέξεων και να προσδιορίζει τη σχέση μεταξύ αυτών στις περιγραφές ευπαθειών.

2. **Transformer-based Model**: Οι επιστήμονες ενσωμάτωσαν ένα δίκτυο Transformer, το οποίο αποτελείται από μπλοκ με πολυ-κεφαλές προσοχής (multi-head attention). Αυτή η δομή επιτρέπει στο μοντέλο να επικεντρώνεται σε διαφορετικά τμήματα της περιγραφής για κάθε τεχνική, βελτιώνοντας την ακρίβεια της αντιστοίχισης.

3. **Bi-directional LSTM και Attention Mechanism**: Αυτά τα δίκτυα χρησιμοποιήθηκαν ως βασικά μοντέλα αναφοράς για να συγκριθεί η απόδοσή τους με το προτεινόμενο μοντέλο. Οι μηχανισμοί προσοχής ειδικά βελτίωσαν την κατανόηση συμφραζόμενων μέσα στις περιγραφές των CVE.

4. **TF-IDF και SVM** [111]: Η μέθοδος TF-IDF αναπαριστά κάθε λέξη ή φράση των περιγραφών ως διανύσματα, βασισμένα στη συχνότητα εμφάνισής τους. Αυτό προσφέρει μια κλασική προσέγγιση κατηγοριοποίησης που χρησιμοποιήθηκε ως επιπλέον μέτρο σύγκρισης.

Αυτόματη Επισήμανση και Κατάταξη με τη Βοήθεια Ετικετών: Οι επιστήμονες χρησιμοποίησαν μια προσέγγιση αυτόματης επισήμανσης των περιγραφών CVE, βασιζόμενοι σε αναγνωριστικά μοτίβα, τα οποία επέτρεψαν την ομαδοποίηση των ευπαθειών με βάση το περιεχόμενό τους και την ενίσχυση της ακρίβειας στις αντιστοιχίσεις. Η διαδικασία αυτή εξασφάλισε ότι το μοντέλο θα μπορούσε να προσαρμόζεται σε νέες περιγραφές CVE, ακόμα και αν αυτές δεν περιλαμβάνονταν στο σύνολο εκπαίδευσης.

Συνοψίζοντας, το σκεπτικό των επιστημόνων εστιάστηκε στο να αξιοποιήσουν εξελιγμένες μεθόδους AI, όπως τα δίκτυα Transformer και τα Bi-LSTM, για να αντιμετωπίσουν το πρόβλημα της κατηγοριοποίησης ευπαθειών σε κυβερνοσυστήματα, διασφαλίζοντας την ακρίβεια και την αποτελεσματικότητα στη διαχείριση των ευπαθειών. Η προσέγγισή τους προσφέρει ένα νέο πρότυπο για τον εντοπισμό και την κατανόηση των απειλών, βελτιώνοντας τη στρατηγική άμυνας των οργανισμών κατά των κυβερνοεπιθέσεων.

Αντίθετα, οι Chatterjee και Thekdi εφαρμόζουν ένα πιθανοτικό μοντέλο, που προσαρμόζεται δυναμικά στις αλλαγές του δικτύου και τις εξελισσόμενες απειλές. [112] Το σκεπτικό των επιστημόνων βασίζεται στην ιδέα ότι η αξιολόγηση της ευπάθειας ενός συστήματος δεν μπορεί να περιορίζεται μόνο σε στατικές καταστάσεις, αλλά πρέπει να λαμβάνει υπόψη την δυναμική εξέλιξη αυτών των ευπαθειών, καθώς και την αλληλεπίδραση των υποσυστημάτων του συστήματος συνολικά. Για αυτό το λόγο, χρησιμοποιούν μια στοχαστική (τυχαία) μοντελοποίηση, που επιτρέπει την παρακολούθηση των μεταβάσεων από τη μια κατάσταση στην άλλη, προκειμένου να προβλέψουν και να κατανοήσουν πώς οι ευπάθειες εξελίσσονται και ποιο είναι το αντίκτυπό τους στην υγεία του συστήματος.

Επιπλέον, οι επιστήμονες προσδιόρισαν διάφορους "μετρικούς δείκτες στάσης συστήματος" (System Posture Metrics), οι οποίοι αντικατοπτρίζουν την κατάσταση του συστήματος σε διαφορετικούς τομείς, όπως η σταθερότητα, η ανθεκτικότητα, η υγεία και η διάσπαση του συστήματος. Αυτοί οι δείκτες επιτρέπουν τη συνολική εκτίμηση της υγείας του συστήματος και την αξιολόγηση της απόδοσης του σε σχέση με τους τύπους ευπαθειών που αντιμετωπίζει.

Οι επιστήμονες είχαν ως στόχο την αξιολόγηση της ευπάθειας και της υγείας των συστημάτων, εστιάζοντας στην επίδραση διαφόρων τύπων ευπαθειών μέσω μετρικών δεικτών για την παρακολούθηση και κατανόηση της εξέλιξής τους στο χρόνο. Παράλληλα, ανέπτυξαν μια μεθοδολογία για τη μοντελοποίηση δυναμικών καταστάσεων, καταγράφοντας τις μεταβάσεις των συστημάτων και αναγνωρίζοντας πώς αυτές οι αλλαγές στις καταστάσεις των υποσυστημάτων επηρεάζουν τη συνολική απόδοση. Τέλος, πρότειναν τη χρήση αυτών των μετρικών για τη βελτίωση των πολιτικών ασφάλειας, ώστε οι διαχειριστές να μπορούν να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την κατανομή πόρων και τη διαχείριση των ευπαθειών του συστήματος.

1. **Στοχαστικά Μοντέλα (HMM - Hidden Markov Models)** [113]: Οι επιστήμονες χρησιμοποίησαν τα μοντέλα κρυφών Markov (HMM), τα οποία είναι κατάλληλα για την αναπαράσταση των καταστάσεων ενός συστήματος που εξελίσσονται με την πάροδο του χρόνου και υπό την επίδραση τυχαίων παραμέτρων. Αυτά τα μοντέλα επιτρέπουν τη μοντελοποίηση της δυναμικής των συστημάτων με κρυφές καταστάσεις, που δεν είναι άμεσα παρατηρήσιμες, αλλά καθορίζονται από παρατηρούμενα δεδομένα.

2. **Ανάλυση Ευαισθησίας**: Η ανάλυση ευαισθησίας χρησιμοποιήθηκε για να μελετηθεί η επίδραση διαφορετικών παραμέτρων, όπως οι αρχικές πιθανότητες και οι παρατηρήσιμες πληροφορίες, στη συμπεριφορά των μετρικών του συστήματος. Αυτή η μέθοδος επέτρεψε στους επιστήμονες να κατανοήσουν πόσο ευαίσθητο είναι το μοντέλο στις αλλαγές και να εξάγουν συμπεράσματα για την αξιοπιστία και την σταθερότητα των μοντέλων τους.

3. **Δημιουργία και Ανάλυση Μετρικών**: Οι μετρικές στάσης συστήματος (System Posture Metrics) χρησιμοποιήθηκαν για την αξιολόγηση της υγείας του συστήματος σε διάφορους τομείς, και οι επιστήμονες παρακολούθησαν τη συμπεριφορά αυτών των μετρικών σε διαφορετικές συνθήκες. Αυτές οι μετρικές περιλαμβάνουν δείκτες για τη σταθερότητα, την ανθεκτικότητα, την υγεία και την διάσπαση του συστήματος.

Η μεθοδολογία που χρησιμοποιήθηκε από τους επιστήμονες συνδυάζει τη χρήση προχωρημένων τεχνικών μηχανικής μάθησης, όπως τα στοχαστικά μοντέλα και την ανάλυση ευαισθησίας, για να κατανοήσει τη δυναμική των ευπαθειών σε κυβερνοσυνδεμένα φυσικά συστήματα. Η εφαρμογή αυτών των μεθόδων βοηθά στη δημιουργία ενός εργαλείου λήψης αποφάσεων που υποστηρίζει τους διαχειριστές συστημάτων να βελτιώσουν την ασφάλεια και τη σταθερότητα των συστημάτων τους με την πάροδο του χρόνου.

2.3.4 Αξιολόγηση και Προτεραιοποίηση Ευπαθειών

Η Αξιολόγηση και προτεραιοποίηση ευπαθειών (Vulnerability Assessment and Prioritization) αποτελεί μια διαδικασία κατά την οποία οι ευπάθειες ενός συστήματος αναλύονται, κατατάσσονται και δίνεται προτεραιότητα στην αντιμετώπισή τους, με βάση τη σοβαρότητά τους και την έκθεση των συστημάτων σε κίνδυνο. Με τη χρήση τεχνικών τεχνητής νοημοσύνης, υπολογίζεται μια βαθμολογία σοβαρότητας για κάθε ευπάθεια, η οποία εξαρτάται από παράγοντες όπως ο κίνδυνος που διατρέχουν τα δεδομένα και οι κρίσιμες λειτουργίες της επιχείρησης, η δυσκολία εκμετάλλευσης της ευπάθειας, η σοβαρότητα της επίθεσης και οι πιθανές ζημιές.

Οι ερευνητές Jiang και Atif [114] επικεντρώθηκαν στην αυτόματη αξιολόγηση της σοβαρότητας ευπαθειών και στην εναρμόνισή τους από αντικρουόμενες αναφορές, χρησιμοποιώντας μια μεθοδολογία μηχανικής μάθησης που βασίζεται στη σοβαρότητα και το προφίλ απειλής.

Συγκεκριμένα, επιδιώκουν τη βελτιστοποίηση της κατηγοριοποίησης των απειλών με στόχο να διευκολύνουν τους φορείς ασφάλειας στην αναγνώριση προτεραιοτήτων για τις ενέργειες προστασίας. Μέσα από την ανάπτυξη ενός αυτοματοποιημένου μοντέλου, το οποίο προβλέπει με ακρίβεια τη βαθμολογία σοβαρότητας των ευπαθειών (CVSS), εστιάζουν στην παροχή αξιόπιστων δεδομένων που θα υποστηρίζουν καλύτερα την ανάλυση των κινδύνων. Παράλληλα, δίνουν ιδιαίτερη έμφαση στη διασφάλιση της ακρίβειας και της ισορροπίας στις ταξινομήσεις, αντιμετωπίζοντας τις προκλήσεις που προκύπτουν από την ανισορροπία στις κατηγορίες ευπαθειών, ώστε το σύστημα να παρέχει πλήρεις και αντιπροσωπευτικές αξιολογήσεις για κάθε πιθανή απειλή.

Οι επιστήμονες αξιοποίησαν μοντέλα μηχανικής μάθησης και βαθιάς μάθησης, συνδυάζοντας διαφορετικά μοντέλα για την επίτευξη μεγαλύτερης ακρίβειας και ανθεκτικότητας στις προβλέψεις.

1. **Πολλαπλά Μοντέλα (Ensemble Learning)**: Εφαρμόστηκε τεχνική συνδυασμού μοντέλων (ensemble), όπως πλειοψηφική ψήφος (hard voting) με συνδυασμούς LSTM, NBSVM και MLP. Αυτή η τεχνική επιτρέπει τη χρήση ισχυρών και αδύναμων ταξινομητών, οι οποίοι, όταν συνδυάζονται, βελτιώνουν την απόδοση της κατηγοριοποίησης.
2. **Πολλαπλοί Γύροι Εκπαίδευσης (Multi-Round Training)**: Δοκιμάστηκαν πολλαπλοί γύροι εκπαίδευσης για να διαπιστωθεί ποιος συνδυασμός μοντέλων αποδίδει καλύτερα για κάθε κατηγορία. Ο κάθε γύρος ενίσχυσε την κατηγοριοποίηση διαφορετικών χαρακτηριστικών των απειλών, όπως AccessVector και IntegrityImpact.
3. **Διασταυρωμένη Επικύρωση (Cross-Validation)**: Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκε διασταυρωμένη επικύρωση 5-διαιρέσεων, ενώ οι μετρικές αξιολόγησης προσαρμόστηκαν στις ανισόρροπες κλάσεις δεδομένων.

Οι Samtani και συνεργάτες τους [115] ανέλυσαν τις ευπάθειες των συσκευών **SCADA** [116] μέσω του συνόλου δεδομένων Shodan, κατατάσσοντάς τες σε τέσσερα επίπεδα κινδύνου: κρίσιμο, υψηλό, μεσαίο και χαμηλό. Η ανάγκη των επιστημόνων ήταν να αναπτύξουν ένα σύστημα που να μπορεί να αναγνωρίζει τις συσκευές SCADA στο Shodan και να αξιολογεί τις ευπάθειές τους προκύπτει από την αύξηση της προσβασιμότητας αυτών των συσκευών μέσω του διαδικτύου, γεγονός που δημιουργεί σοβαρούς κινδύνους για την ασφάλεια κρίσιμων υποδομών. Γι' αυτό τον λόγο και επικεντρώνονται στη χρήση δεδομένων που είναι ήδη διαθέσιμα μέσω του Shodan, όπως είναι τα banner δεδομένα, τα οποία περιέχουν χρήσιμες ενδείξεις για την αναγνώριση των συσκευών SCADA.

Η έρευνα έχει ως βασικούς στόχους την ταξινόμηση και την αξιολόγηση ευπαθειών των συσκευών SCADA που είναι προσβάσιμες μέσω του Shodan, ενός μεγάλου διαδικτυακού αποθετηρίου συνδεδεμένων συσκευών. Αρχικά, επιδιώκεται ο εντοπισμός και η κατηγοριοποίηση των συσκευών SCADA ανάμεσα στις εκατοντάδες εκατομμύρια καταχωρήσεις του Shodan, με τη χρήση τεχνικών επεξεργασίας κειμένου και εξόρυξης δεδομένων, ώστε να αναγνωριστούν χαρακτηριστικά και πρότυπα που ξεχωρίζουν τις συσκευές αυτές από άλλες. Έπειτα, πραγματοποιείται η αξιολόγηση των ευπαθειών που ενδέχεται να επηρεάζουν τις συγκεκριμένες συσκευές, μέσω εργαλείων όπως το Nessus, για τον εντοπισμό σοβαρών ζητημάτων ασφάλειας, όπως αποδοχή προεπιλεγμένων κωδικών και ξεπερασμένα λογισμικά. Αυτή η προσέγγιση επιτρέπει τη διασφάλιση των κρίσιμων υποδομών που εξαρτώνται από τις συσκευές SCADA, μειώνοντας τους κινδύνους από πιθανές κυβερνοεπιθέσεις.

Οι επιστήμονες χρησιμοποίησαν μεθόδους που περιλαμβάνουν:

1. **Επεξεργασία φυσικής γλώσσας (Text Mining)**: Η ανάλυση των δεδομένων από τα banners των συσκευών έγινε μέσω διαδικασιών επεξεργασίας κειμένου, όπου τα δεδομένα διαχωρίστηκαν σε ν-γράμματα (n-grams) και στη συνέχεια δημιουργήθηκε μια υπογραφή (signature set) που βοηθά στην αναγνώριση των συσκευών SCADA.
2. **Εξόρυξη δεδομένων και ταξινόμηση**: Χρησιμοποιήθηκαν διάφοροι αλγόριθμοι εξόρυξης δεδομένων και μηχανικής μάθησης για την ταξινόμηση των συσκευών, όπως:
    - **Random Forest**: Ο αλγόριθμος αυτός πέτυχε την καλύτερη απόδοση με f-measure 99,4% και χρησιμοποιήθηκε για την τελική ταξινόμηση.
    - **Άλλοι αλγόριθμοι:** Στην έρευνα χρησιμοποιήθηκαν επίσης αλγόριθμοι όπως logistic regression, SVM, Naive Bayes, artificial neural networks και perceptron**.**
3. **Αξιολόγηση Ευπαθειών μέσω Nessus** [117]: Η αξιολόγηση των ευπαθειών έγινε χρησιμοποιώντας το Nessus, το οποίο παρέχει πλούσιες δυνατότητες για την ανίχνευση ευπαθειών σε συσκευές SCADA και άλλες συσκευές δικτύου. Οι ευπάθειες εντοπίστηκαν σε βασικούς τομείς όπως η αποδοχή προεπιλεγμένων κωδικών πρόσβασης και η ύπαρξη ξεπερασμένων τεχνολογιών.

Η συνδυασμένη χρήση τεχνικών εξόρυξης δεδομένων, επεξεργασίας κειμένου και μηχανικής μάθησης με σκοπό την ασφαλή αναγνώριση και αξιολόγηση των συσκευών SCADA σε πραγματικό χρόνο αποτελεί τη βάση της ερευνητικής προσέγγισης των επιστημόνων.

Επιπλέον, οι Brown και συνεργάτες τους [118] υπολόγισαν τους βαθμούς κινδύνου ευπάθειας και εκμετάλλευσης για κάθε συσκευή IoT με βάση το γράφημα επιθέσεων που δημιουργήθηκε από την

τοπολογία του δικτύου, σύμφωνα με τις ρυθμίσεις του διαχειριστή δικτύου. Οι επιστήμονες αναγνωρίζουν ότι τα IoT και CPS συστήματα, λόγω της πολυπλοκότητας και της σύνδεσης πολλών συσκευών, είναι επιρρεπή σε μια σειρά από επιθέσεις που μπορεί να επηρεάσουν την ασφάλεια και την αποτελεσματικότητά τους. Το κύριο πρόβλημα είναι ότι οι παραδοσιακοί τρόποι αξιολόγησης κινδύνου δεν είναι επαρκείς για να εντοπίσουν τα πιο πιθανά μονοπάτια εκμετάλλευσης και να αντιμετωπίσουν τις συστημικές αδυναμίες. Οι επιστήμονες χρησιμοποιούν το **GRAVITAS** [119] για να αναλύσουν και να εντοπίσουν πολύπλοκες αλυσίδες επιθέσεων που περιλαμβάνουν πολλαπλές συσκευές, οι οποίες μπορεί να μην είναι επικίνδυνες από μόνες τους, αλλά σε συνδυασμό να προκαλούν σοβαρές επιθέσεις.

Οι ερευνητές στοχεύουν στην αναγνώριση επιθέσεων σε συστήματα IoT/CPS, επιδιώκοντας να εντοπίσουν ευπάθειες που μπορούν να αξιοποιηθούν σε αλυσίδες επιθέσεων. Επιπλέον, επιδιώκουν τη βελτιστοποίηση της άμυνας εντοπίζοντας τους πιο ευάλωτους κόμβους του συστήματος και τοποθετώντας κατάλληλες άμυνες για τη μείωση του κινδύνου εκμετάλλευσής τους. Το GRAVITAS διευκολύνει την προληπτική αξιολόγηση της ασφάλειας, επιτρέποντας στους διαχειριστές να προλαμβάνουν και να διορθώνουν τα αδύνατα σημεία ενός συστήματος IoT πριν την ανάπτυξή του. Τέλος, οι επιστήμονες επιδιώκουν την αντιμετώπιση πολύπλοκων επιθέσεων, διαχειριζόμενοι την ασφάλεια σε σύνθετα δίκτυα συσκευών και αναγνωρίζοντας επιθέσεις που εκμεταλλεύονται αδυναμίες πολλαπλών συσκευών ταυτόχρονα.

Οι μέθοδοι που χρησιμοποιήθηκαν αναγράφονται παρακάτω.

1. **Ανάλυση Επιθέσεων και Βαθμολόγηση Κινδύνου**: Χρησιμοποιούν μαθηματικά μοντέλα για να υπολογίσουν τον "κίνδυνο εκμετάλλευσης" (exploit risk) των συσκευών στο δίκτυο IoT. Αυτή η διαδικασία λαμβάνει υπόψη τη σοβαρότητα και την ευκολία εκμετάλλευσης των αδυναμιών των συσκευών.

2. **Μοντέλα Βελτιστοποίησης**: Χρησιμοποιείται αλγόριθμος βελτιστοποίησης για να βρουν τις βέλτιστες άμυνες που μειώνουν τον συνολικό κίνδυνο του συστήματος. Εφαρμόζεται διαδικασία βελτιστοποίησης «greedy local search», η οποία εστιάζει στην τοποθέτηση αμυνών με το λιγότερο δυνατό κόστος και μεγαλύτερη αποδοτικότητα.

3. **Προσομοίωση Επιθέσεων**: Εφαρμόζεται μια διαδικασία προσομοίωσης για να δημιουργηθούν πιθανές αλυσίδες επιθέσεων που ακολουθούν μια σειρά από ευπάθειες σε πολλαπλές συσκευές. Αυτός ο τύπος προσομοίωσης είναι πιο αποδοτικός από τους παραδοσιακούς τρόπους διαχείρισης κινδύνου, οι οποίοι συνήθως δεν εντοπίζουν πολύπλοκες επιθέσεις που εκμεταλλεύονται αλυσίδες ευπαθειών.

4. **Τυχαιότητα στην Αξιολόγηση Επιπτώσεων**: Για να εξασφαλιστεί η ευελιξία του μοντέλου έναντι διαφορετικών επιθέσεων, προστίθεται "θόρυβος" στην αξιολόγηση των επιπτώσεων μιας επίθεσης. Οι επιπτώσεις αυτές υπολογίζονται με τυχαίες κατανομές για να μιμηθούν την αβεβαιότητα που υπάρχει σε πραγματικές συνθήκες επιθέσεων.

5. **Μαθηματικά Μοντέλα και Καμπύλες Βελτιστοποίησης**: Το GRAVITAS χρησιμοποιεί μαθηματικά μοντέλα για να κατασκευάσει καμπύλες που συνδυάζουν την ασφάλεια με το κόστος άμυνας. Οι ερευνητές χρησιμοποιούν τη βελτιστοποίηση αυτών των καμπυλών για να βρουν τη βέλτιστη στρατηγική άμυνας που ισορροπεί την προστασία και το κόστος.

Οι επιστήμονες που ανέπτυξαν το GRAVITAS στόχευαν στην προληπτική και στρατηγική ενίσχυση της ασφάλειας σε πολύπλοκα συστήματα IoT/CPS. Χρησιμοποίησαν μεθόδους τεχνητής νοημοσύνης και μαθηματικά μοντέλα βελτιστοποίησης για να εντοπίσουν αδύνατα σημεία και να τοποθετήσουν

άμυνες που μειώνουν τον κίνδυνο επιθέσεων. Το εργαλείο επιτρέπει την προσομοίωση και ανάλυση αλυσίδων επιθέσεων, ενώ παρέχει προσαρμοστικότητα και αντοχή σε αβεβαιότητες, βοηθώντας έτσι τους διαχειριστές να αναγνωρίζουν και να διορθώνουν τρωτά σημεία πριν από την ανάπτυξη του συστήματος.

2.4 Αυτοματοποιημένη Αναζήτηση Απειλής

Η αυτοματοποιημένη αναζήτηση απειλών (Automated Threat Hunting) είναι μια προληπτική διαδικασία ασφάλειας που αφορά την αναζήτηση σε δίκτυα, τελικές συσκευές και δεδομένα για τον εντοπισμό δυνητικά κακόβουλων, ύποπτων ή επικίνδυνων δραστηριοτήτων εντός ενός οργανισμού. Σκοπός της είναι να εντοπίσει και να κατηγοριοποιήσει πιθανές απειλές εκ των προτέρων, χρησιμοποιώντας νέες πληροφορίες για απειλές σε δεδομένα που έχουν ήδη συλλεχθεί.

Η αναζήτηση απειλών είναι ένας σχετικά νέος τομέας εφαρμογής που έχει μεγάλη σημασία για την πρώιμη ανίχνευση κινδύνων. Παρά την αξία της, οι υπάρχουσες προσεγγίσεις εστιάζουν κυρίως στην ανίχνευση ανωμαλιών και συχνά παραβλέπουν την πλούσια εξωτερική γνώση σχετικά με τις απειλές που προσφέρει η ανοιχτού κώδικα κυβερνο-κατασκοπεία (OSCTI).

Η μελέτη *[120]* για την αυτόματη αναζήτηση απειλών στον κυβερνοχώρο εστιάζει στην ανάπτυξη ενός καινοτόμου συστήματος, του **THREATRAPTOR**, το οποίο στοχεύει στην αποτελεσματική ανακάλυψη και ανάλυση κακόβουλων δραστηριοτήτων. Οι επιστήμονες αναγνωρίζουν ότι οι υπάρχουσες μέθοδοι συχνά απαιτούν χρονοβόρα χειροκίνητη εργασία για τη δημιουργία ερωτημάτων και δεν αξιοποιούν πλήρως την πλούσια πληροφορία που προέρχεται από την ανοιχτού κώδικα κυβερνο-κατασκοπεία (OSCTI).

Έτσι, ο στόχος τους είναι να γεφυρώσουν το κενό ανάμεσα σε αυτή την εξωτερική γνώση και τη διαδικασία ανίχνευσης απειλών, διευκολύνοντας την αυτοματοποιημένη αναζήτηση.

Για να πετύχουν τον στόχο τους, οι επιστήμονες ανέπτυξαν το THREATRAPTOR, το οποίο αξιοποιεί προηγμένες τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) για την εξαγωγή δομημένων πληροφοριών σχετικά με τις απειλές από μη δομημένα κείμενα. Το σύστημα αυτό περιλαμβάνει τη γλώσσα ερωτημάτων TBQL, που επιτρέπει στους αναλυτές να αναζητούν κακόβουλες δραστηριότητες στα δεδομένα καταγραφής, καθώς και έναν μηχανισμό αυτόματης σύνθεσης ερωτημάτων. Η εκτέλεση των ερωτημάτων γίνεται με βάση την ικανότητα του συστήματος να αναλύει μεγάλα σύνολα δεδομένων και να εντοπίζει γρήγορα κακόβουλες δραστηριότητες, καθιστώντας το THREATRAPTOR ταχύτερο και πιο αποδοτικό από τις παραδοσιακές μεθόδους.

Η καινοτομία του THREATRAPTOR έγκειται στη συνδυαστική του προσέγγιση. Χρησιμοποιεί προηγμένες τεχνικές NLP για να εξάγει συμπεριφορές απειλών από τις πληροφορίες της OSCTI, ενώ παράλληλα εφαρμόζει την TBQL για να διευκολύνει την αναζήτηση. Μέσω αυτής της μεθόδου, οι αναλυτές μπορούν να επεξεργάζονται ερωτήματα πιο αποτελεσματικά, επιτυγχάνοντας την ακριβή αναγνώριση κακόβουλων δραστηριοτήτων. Οι αξιολογήσεις του THREATRAPTOR έδειξαν ότι το σύστημα είναι ικανό να εξάγει πληροφορίες με υψηλή ακρίβεια και ταχύτητα, ενισχύοντας την ικανότητα των αναλυτών να εντοπίζουν και να αντιμετωπίζουν απειλές στον κυβερνοχώρο.

2.5 Σύνοψη εργαλείων/μεθόδων AI

Ο παρακάτω πίνακας συνοψίζει τα εργαλεία, τις μεθόδους και τους αλγορίθμους τεχνητής νοημοσύνης (AI) που υποστηρίζουν τη λειτουργία «Προσδιορισμός» (Identify) του NIST Cybersecurity Framework, η οποία αποτελεί το θεμέλιο για τη διαχείριση των κινδύνων κυβερνοασφάλειας. Περιλαμβάνει AI τεχνικές για αυτοματοποιημένη διαχείριση διαμόρφωσης (π.χ. Pareto Q-learning, NSGA-II), αξιολόγηση ελέγχων ασφαλείας (ARIMA, DTW), ανίχνευση και ταξινόμηση ευπαθειών (BERT, CNN, SVM, AutoVAS) και αναζήτηση απειλών (THREATRAPTOR). Αυτά τα εργαλεία επιτρέπουν στις επιχειρήσεις να εντοπίζουν, να αναλύουν και να διαχειρίζονται κινδύνους με τη χρήση προηγμένων τεχνικών μηχανικής μάθησης, βελτιώνοντας την κυβερνοασφάλεια και τη συνολική ανθεκτικότητα των πληροφοριακών τους συστημάτων.

2.1.1	Αυτοματοποιημένη Διαχείριση Διάταξης	(1) Αλγόριθμος Πολυαντικειμενικής Q-Μάθησης (Pareto Q-learning) (2) Πολυαντικειμενική SARSA (Pareto SARSA) (3) Πολυαντικειμενική TD(0) με Μετά-καταστάσεις (Pareto TD(0) Afterstate Algorithm (4) MOCell (Multi-Objective Cellural Genetic Algorithm) (5) NSGA-II (Non-dominated Sorting Genetic Algorithm II) (6) SPEA2 (Strength Pareto Evolutionary Algorithm 2) (7) VEREFOO (8) AMADEUS (9) CyberSPL - ταξινομητής Random Forest
2.1.2	Αυτοματοποιημένη Επικύρωση Ελέγχου Ασφαλείας	(1) ARIMA (Autoregressive Integrated Moving Average) (2) Dynamic Time Warping (DTW) (3) διαδικτυακό εργαλείο Πλαισίου Κυβερνοασφάλειας Κτιρίων (BCF) (4) χρήση γενετικών αλγορίθμων

2.1.3 2.1.3.1	Αυτοματοποιημένη Αναγνώριση και Αξιολόγηση Ευπάθειας Αυτοματοποιημένη Ανίχνευση Ευπαθειών	(1) Vullntel (2) αλγόριθμος Ro-MFAC (3) AutoVAS (4) τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) και μέθοδοι μηχανικής μάθησης (5) μοντέλο BERT (Bidirectonal Encoder Representations from Transformers) (6) SVM (Support Vector Machine) (7) CNN (Convolutional Neural Networks) (8) NeuFuzz (9) μέθοδος "Skyfire" (10) DeepSmith (11) DSmith (12) μέθοδος "smart fuzzing" (13) NEUZZ (14) DEEPFUZZ (15) NDSPI-DQN (16) Crown Jewel Analysis - Reinforcement Learning (CJA-RL) (17) Advantage Actor-Critic
2.1.3.2	Αυτοματοποιημένη Ταξινόμηση Ευπαθειών	(1) CVErizer (2) αλγόριθμος Bag-of-Words (BoW) (3) Boruta (4) Latent Dirichlet Allocation (LDA)
2.1.3.3	Εξερεύνηση Ευπαθειών	(1) CYBOK (2) Word2Vec (3) Transformer-based Model (4) Bi-directional LSTM - Attention Mechanism (5) TF-IDF - SVM (6) HMM - Hidden Markov Models
2.1.3.4	Αξιολόγηση και Προτεραιοποίηση Ευπαθειών	(1) LSTM (2) NBSVM (3) MLP (4) Random Forest (5) Logistic Regression (6) SVM (7) Naive Bayes (8) Artificial Neural Networks (9) Perceptron (10) Nessus (11) GRAVITAS
2.1.4	Αυτοματοποιημένη Αναζήτηση Απειλής	(1) THREATRAPTOR

Πίνακας 2 : Συνοπτικός πίνακας AI εργαλείων και αλγορίθμων που υποστηρίζουν τη λειτουργία «Προσδιορισμός» (Identify) του NIST Cybersecurity Framework, ενισχύοντας την ανίχνευση, ταξινόμηση και διαχείριση κυβερνοαπειλών.

3. Προστασία

*Η λειτουργία **«Προστασία»** (**Protect)** του πλαισίου NIST είναι θεμελιώδης, καθώς στοχεύει στην ανάπτυξη και εφαρμογή κατάλληλων μέτρων προστασίας για τη διασφάλιση της λειτουργίας κρίσιμων υποδομών. Σύμφωνα με το NIST, η λειτουργία Protect βοηθά στη μείωση ή στον περιορισμό του αντίκτυπου πιθανών περιστατικών κυβερνοασφάλειας, προσφέροντας πρακτικές για την πρόληψη. Μέσω κατάλληλων πρωτοκόλλων και πολιτικών, οι οργανισμοί μπορούν να περιορίσουν τους κινδύνους και να εξασφαλίσουν την ασφάλεια των κρίσιμων υποδομών τους.*

3.1 Ταυτοποίηση Συσκευών με Υποστήριξη Τεχνητής Νοημοσύνης

Ο έλεγχος ταυτότητας συσκευών που υποστηρίζονται από AI (AI-supported device authentication) αποτελεί τη διαδικασία επιβεβαίωσης της ταυτότητας των συσκευών, είτε μέσω των διαπιστευτηρίων τους είτε μέσω της συμπεριφοράς τους στο δίκτυο, με σκοπό τη διασφάλιση της ασφάλειας στις επικοινωνίες μεταξύ μηχανών. Οι ερευνητές, όπως θα εξετάσουμε, εργάζονται ενεργά στον τομέα της αναγνώρισης και αυθεντικοποίησης αισθητήρων, ώστε να διασφαλίσουν την ασφάλεια των κυβερνοφυσικών συστημάτων ή του τομέα της αυτοκινητοβιομηχανίας. Οι ατέλειες στα κανάλια επικοινωνίας και τους αισθητήρες χρησιμοποιούνται για την ανίχνευση παραμέτρων, τόσο κατά τη φάση της μετάβασης όσο και σε σταθερή κατάσταση, οι οποίες στη συνέχεια εισάγονται σε μοντέλα μηχανικής μάθησης για την ταυτοποίηση των αισθητήρων.

Το σκεπτικό των επιστημόνων [121] βασίζεται στην ανάγκη ενίσχυσης της ασφάλειας των σύγχρονων οχημάτων, τα οποία γίνονται ολοένα και πιο ευάλωτα σε κυβερνοεπιθέσεις λόγω της απουσίας βασικών μηχανισμών ασφάλειας στο πρωτόκολλο CAN (Controller Area Network) [122]. Οι ερευνητές θέλουν να αναπτύξουν νέες μεθόδους που να επιτρέπουν την αυθεντικοποίηση των μηνυμάτων στο φυσικό επίπεδο του δικτύου, εκμεταλλευόμενοι τις μοναδικές ιδιότητες κάθε καναλιού επικοινωνίας.

Οι στόχοι της μελέτης επικεντρώνονται στην ταυτοποίηση των ECU (Electronic Control Units), συνδέοντας κάθε μήνυμα CAN με την πηγή του, στην αυθεντικοποίηση των μηνυμάτων για τον εντοπισμό και την απόρριψη μη έγκυρων ή κακόβουλων μηνυμάτων, και στη διασφάλιση της αξιοπιστίας του συστήματος έναντι φυσικών και ασύρματων κυβερνοεπιθέσεων.

Οι επιστήμονες χρησιμοποίησαν μια προσέγγιση βασισμένη στη μηχανική μάθηση και τα νευρωνικά δίκτυα. Συγκεκριμένα:

1. **Εξαγωγή χαρακτηριστικών από την απόκριση βήματος των καναλιών**: Μελέτησαν τις ιδιότητες των καναλιών επικοινωνίας CAN, όπως ο ρυθμός αποκατάστασης, ο χρόνος κορυφής και η συχνότητα αποσβέσεως. Αυτές οι παράμετροι είναι μοναδικές για κάθε κανάλι λόγω φυσικών διαφορών στα υλικά και τα ηλεκτρονικά μέρη.
2. **Εκπαίδευση ενός πολυεπίπεδου νευρωνικού δικτύου (MLP)**: Το νευρωνικό δίκτυο εκπαιδεύτηκε να αναγνωρίζει τα σήματα που προέρχονται από διαφορετικούς ECU, χρησιμοποιώντας τα χαρακτηριστικά των σημάτων ως δεδομένα εισόδου.

3. **Αξιολόγηση απόδοσης**: Εφάρμοσαν μέτρα όπως precision, recall, F1-score, accuracy και ROC καμπύλες για την ανάλυση της ακρίβειας του συστήματος.

Η χρήση της τεχνητής νοημοσύνης επιτρέπει την αυτόματη και αποτελεσματική ταυτοποίηση των ECU, παρέχοντας έναν τρόπο να αυθεντικοποιούνται τα μηνύματα στο δίκτυο CAN με υψηλή ακρίβεια (97.4%) και να εντοπίζονται επιθέσεις πλαστογράφησης.

Η βασική ιδέα των επιστημόνων [123] σε μια διαφορετική έρευνα βασίζεται στην ανάγκη ενίσχυσης της κυβερνοασφάλειας στα οχήματα, ειδικά ενάντια σε απειλές που στοχεύουν την αλλοίωση ή αντικατάσταση κρίσιμων αισθητήρων, όπως οι αισθητήρες Hall. Οι αισθητήρες αυτοί χρησιμοποιούνται εκτενώς στην αυτοκινητοβιομηχανία για την ανίχνευση μαγνητικών πεδίων και την παρακολούθηση κινήσεων, αλλά είναι ευάλωτοι σε επιθέσεις που μπορούν να επηρεάσουν την ακεραιότητα και τη λειτουργία τους. Στόχος των επιστημόνων είναι να δημιουργήσουν έναν τρόπο να αναγνωρίζουν και να αυθεντικοποιούν αυτούς τους αισθητήρες με υψηλή ακρίβεια, μειώνοντας παράλληλα το ρίσκο αντικατάστασής τους από κακόβουλες συσκευές.

Οι επιστήμονες χρησιμοποίησαν προηγμένες μεθόδους μηχανικής μάθησης και επεξεργασίας σήματος για να επιτύχουν τους στόχους τους:

1. **Υποστήριξη Διανυσμάτων (Support Vector Machines - SVM)**:
    - Χρησιμοποιήθηκαν εκτενώς λόγω της υψηλής τους απόδοσης στην ταξινόμηση των αισθητήρων.
    - Οι υπερπαράμετροι του SVM, όπως οι παράγοντες γ και C, βελτιστοποιήθηκαν για καλύτερα αποτελέσματα.
2. **Άλλοι Αλγόριθμοι Μηχανικής Μάθησης**:
    - Δοκιμάστηκαν επίσης αλγόριθμοι όπως K-Nearest Neighbors (KNN) και Decision Trees, αλλά οι επιδόσεις τους ήταν χαμηλότερες σε σύγκριση με το SVM.
3. **Μετασχηματισμοί Σήματος**:
    - Εφαρμόστηκαν φίλτρα IIR και FIR, καθώς και ο Μετασχηματισμός Συνεχούς Κυματιδίου (Continuous Wavelet Transform - CWT), για τη δημιουργία μοναδικών «υπογραφών» από τα σήματα των αισθητήρων.
    - Αυτές οι τεχνικές παρήγαγαν ένα ευρύ φάσμα αποκρίσεων που ενισχύουν την ικανότητα διάκρισης μεταξύ διαφορετικών αισθητήρων.
4. **Αξιολόγηση Ακριβείας**:
    - Οι επιστήμονες χρησιμοποίησαν καμπύλες ROC (Receiver Operating Characteristic) και ποσοστά EER (Equal Error Rate) για να αξιολογήσουν την ακρίβεια των ταξινομήσεων και την απόδοση αυθεντικοποίησης.
    - Οι μέθοδοι που εφαρμόστηκαν απέδειξαν ανθεκτικότητα στις επιθέσεις επανάληψης.

Με αυτόν τον συνδυασμό μεθόδων μηχανικής μάθησης και προηγμένων τεχνικών επεξεργασίας σήματος, οι επιστήμονες πέτυχαν υψηλά επίπεδα ασφάλειας για τους αισθητήρες Hall, ανοίγοντας τον δρόμο για την πρακτική εφαρμογή αυτών των τεχνικών στην αυτοκινητοβιομηχανία.

Η επόμενη έρευνα που θα δούμε [124] στοχεύει στην αντιμετώπιση αυξανόμενων προβλημάτων κυβερνοασφάλειας που προκύπτουν από τη ψηφιοποίηση των συστημάτων μέτρησης και ελέγχου στις μικρο-δίκτυες ενέργειας (microgrids). Συγκεκριμένα, το σκεπτικό των επιστημόνων βασίζεται στην ανάγκη διασφάλισης της αυθεντικότητας και της ακεραιότητας των δεδομένων που συλλέγονται από συγχρονισμένους μετρητές (distribution synchrophasors - DS), οι οποίοι είναι κρίσιμης σημασίας για τη λειτουργία των μικροδικτύων. Οι επιθέσεις «Source ID Mix», όπου ένας κακόβουλος χρήστης μπορεί να αλλοιώσει τις πληροφορίες της πηγής χωρίς να αλλάξει τις τιμές μέτρησης, αποτελούν μια σοβαρή απειλή για την ασφάλεια και την αξιοπιστία αυτών των συστημάτων.

Οι επιστήμονες στοχεύουν στην αναγνώριση της πηγής προέλευσης των δεδομένων για να διασφαλίσουν ότι αυτά προέρχονται από τις σωστές τοποθεσίες, επιτυγχάνοντας παράλληλα υψηλή ακρίβεια ακόμα και σε δεδομένα που συλλέγονται από πολύ κοντινές γεωγραφικές περιοχές. Ο κύριος σκοπός τους είναι η ενίσχυση της κυβερνοασφάλειας, μέσω της ανάπτυξης στρατηγικών που ανιχνεύουν ανωμαλίες και αποτρέπουν επιθέσεις, ενισχύοντας την ακεραιότητα και την αξιοπιστία των δεδομένων σε μικροδίκτυα.

Οι μέθοδοι και τα εργαλεία τεχνητής νοημοσύνης που χρησιμοποιούνται είναι τα εξής:

1. **Self-Adaptive Mathematical Morphology (SAMM)** [125]: Μια τεχνική που χρησιμοποιείται για να επεξεργαστεί τις μεταβολές συχνοτήτων των δεδομένων, ώστε να διατηρηθούν μόνο οι σημαντικές κορυφές και να εξαχθούν χαρακτηριστικά που αντικατοπτρίζουν τις τοπικές περιβαλλοντικές συνθήκες.
2. **Time-Frequency (TF) Mapping**: Χρησιμοποιείται για να μετατρέψει τα δεδομένα σε έναν δισδιάστατο χάρτη με βάση τις τάσεις αραιότητας και τραχύτητας των δεδομένων, παρέχοντας διακριτά υποδείγματα για την ταυτοποίηση.
3. **Random Forest Classification (RFC)**: Ένας αλγόριθμος μηχανικής μάθησης που συσχετίζει τα εξαγόμενα χαρακτηριστικά με πληροφορίες της πηγής για να ταυτοποιήσει την προέλευση των δεδομένων.
4. **Adaptive Boosting (AdaBoost)**: Μια μέθοδος που βελτιώνει την απόδοση των μοντέλων ταξινόμησης συνδυάζοντας πολλές ασθενείς ταξινομητές για να δημιουργήσει ένα ισχυρό μοντέλο.

Οι ερευνητές επαλήθευσαν τη μέθοδο τους μέσω πειραμάτων σε δεδομένα από μικρές γεωγραφικές κλίμακες, όπως μετρήσεις από διαφορετικούς κόμβους του ίδιου τροφοδότη ή από δωμάτια του ίδιου κτιρίου. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος επιτυγχάνει ακρίβεια ταυτοποίησης έως και 96%, ξεπερνώντας άλλες υπάρχουσες τεχνικές. Έτσι, η μέθοδος αυτή παρέχει ένα σημαντικό εργαλείο για τη βελτίωση της ασφάλειας και την αποτροπή κυβερνοεπιθέσεων στις μικρο-δίκτυες.

3.2 Αυτοματοποιημένος Έλεγχος Πρόσβασης

Ο αυτοματοποιημένος έλεγχος πρόσβασης (Automated Access Control) ελέγχει ποιοι χρήστες έχουν πρόσβαση στο σύστημα, με βάση παράγοντες όπως οι ρόλοι τους, οι καταστάσεις ή οι κανονισμοί του οργανισμού. Οι ερευνητές χρησιμοποιούν τεχνικές τεχνητής νοημοσύνης για να

διατηρούν την κατάσταση του ελέγχου πρόσβασης, να εντοπίζουν ρόλους και να παίρνουν αποφάσεις σύμφωνα με την εκάστοτε κατάσταση, με σκοπό την αποτροπή μη εξουσιοδοτημένης πρόσβασης και των πιθανών αρνητικών συνεπειών της.

Οι Marco Benedetti και Marco Mori [126], έχοντας εντοπίσει τα προβλήματα πολυπλοκότητας και δυσκολίας στη διαχείριση συστημάτων Role-Based Access Control (RBAC) [127], αναπτύσσουν μια μέθοδο που στοχεύει στην απλοποίηση και βελτιστοποίηση αυτών των συστημάτων. Τα RBAC χρησιμοποιούνται ευρέως για τη διαχείριση δικαιωμάτων πρόσβασης, αλλά συχνά καταλήγουν να είναι δυσκίνητα λόγω της συσσώρευσης εξαιρέσεων και παραβιάσεων με την πάροδο του χρόνου.

Το σκεπτικό της μεθόδου βασίζεται στην ισορροπία ανάμεσα στη διατήρηση της υπάρχουσας δομής και την ανάγκη για βελτίωση. Σε συστήματα όπου η διατήρηση της τρέχουσας κατάστασης είναι σημαντική, οι αλλαγές πρέπει να είναι ελάχιστες, διασφαλίζοντας τη συνέχεια και τη σταθερότητα. Αντίθετα, όταν προτεραιότητα είναι η απλοποίηση και η καλύτερη διαχείριση, οι τροποποιήσεις μπορούν να είναι πιο εκτενείς. Η μέθοδος επικεντρώνεται σε σταδιακές, στοχευμένες βελτιώσεις, ώστε να αποφεύγονται οι ριζικές αλλαγές που μπορεί να προκαλέσουν προβλήματα.

Γι' αυτό τον λόγο στοχεύουν στην ελαχιστοποίηση των αλλαγών που απαιτούνται για τη διόρθωση παραβιάσεων, ενώ παράλληλα απλοποιούν τη διαχείριση των ρόλων και των δικαιωμάτων για τους διαχειριστές.

Για την υλοποίηση αυτής της προσέγγισης, οι επιστήμονες χρησιμοποίησαν τεχνικές Τεχνητής Νοημοσύνης (AI) που βασίζονται σε λογικές μεθόδους και βελτιστοποίηση. Οι κύριες μέθοδοι περιλαμβάνουν:

1. **Max-SAT (Maximum Satisfiability)** [128]:
    - Πρόκειται για μια λογική μέθοδο βελτιστοποίησης που χρησιμοποιείται για να επιτευχθεί η καλύτερη δυνατή λύση σε προβλήματα με αντικρουόμενους περιορισμούς.
    - Η μέθοδος Max-SAT χρησιμοποιήθηκε για την κωδικοποίηση του προβλήματος RBAC σε μια μορφή που οι υπολογιστές μπορούν να λύσουν, λαμβάνοντας υπόψη την ανάγκη για ελάχιστες αλλαγές (ομοιότητα) και βελτίωση της απλότητας.
2. **PDDL (Planning Domain Definition Language)** [129]:
    - Το PDDL χρησιμοποιήθηκε για τη σύνθεση πλάνων δράσης. Αυτά τα πλάνα περιλαμβάνουν τις απαραίτητες ενέργειες για τη μετάβαση από την αρχική κατάσταση σε μια βελτιστοποιημένη.
    - Η μέθοδος αυτή βασίζεται στην περιγραφή περιορισμών και στόχων, δημιουργώντας ένα σύνολο ενεργειών που βελτιώνουν το σύστημα με ελάχιστο κόστος.
3. **Χρήση planners όπως το Fast-Downward**:
    - Χρησιμοποιήθηκαν προηγμένοι αλγόριθμοι σχεδιασμού (planners) που προσφέρουν αποδεκτές λύσεις σε εύλογο χρόνο, ακόμα και για σύνθετα δεδομένα.
    - Ο Fast-Downward planner εφαρμόστηκε για τη δημιουργία πλάνων δράσης που μειώνουν τις απαραίτητες αλλαγές σε χρήστες, ρόλους και δικαιώματα.
4. **Στρατηγικές βελτιστοποίησης**:

- Περιλαμβάνουν την ενσωμάτωση λειτουργιών όπως διαγραφή σειρών/στηλών ή ανταλλαγές αναθέσεων δικαιωμάτων, κάτι που δεν υποστηρίζεται στις παραδοσιακές μέθοδοι.
- Η στρατηγική "όσο το δυνατόν λιγότερες αλλαγές" διατηρεί τη συνοχή του συστήματος, ενώ η "βελτιστοποίηση από την αρχή" εφαρμόζεται όταν η αρχική κατάσταση δεν έχει σημασία.

Οι επιστήμονες προτείνουν μια νέα μέθοδο εξαγωγής ρόλων (role mining) από δεδομένα χρήστη-δικαιωμάτων [130], με στόχο την αποτελεσματική και ακριβή αναπαράσταση του συστήματος ελέγχου πρόσβασης βάσει ρόλων (RBAC). Αναγνωρίζουν ότι τα πραγματικά δεδομένα που περιέχουν πλήρεις πληροφορίες RBAC είναι δύσκολο να αποκτηθούν, γι' αυτό χρησιμοποιούν συνθετικά δεδομένα προσομοιώνοντας πραγματικές συνθήκες, βασιζόμενοι σε ένα πρότυπο πανεπιστημιακού συστήματος. Αυτό τους επιτρέπει να μελετήσουν τη συμπεριφορά της μεθόδου τους σε ένα περιβάλλον που αντικατοπτρίζει πραγματικές πολιτικές πρόσβασης.

Οι στόχοι τους περιλαμβάνουν την επίτευξη μέγιστης ακρίβειας στην εξαγωγή ρόλων, ώστε τα αποτελέσματα να αντιστοιχούν όσο το δυνατόν καλύτερα στα πραγματικά δεδομένα πρόσβασης και την αποδοτική λειτουργία της μεθόδου ακόμα και για μεγάλα σύνολα δεδομένων. Επιπλέον, περιέχουν την ανάπτυξη ενός εφαρμόσιμου μοντέλου για πραγματικά συστήματα ελέγχου πρόσβασης με δυνατότητα προσαρμογής σε περιορισμούς, όπως η ιεραρχία ρόλων, και τη μείωση της ανάγκης βελτιστοποίησης παραμέτρων, εξασφαλίζοντας παράλληλα υψηλή ακρίβεια και χρηστικότητα.

Η βασική τεχνική που χρησιμοποιήθηκε είναι η Μη Αρνητική Παραγοντοποίηση Πινάκων **(Non-Negative Matrix Factorization, NMF)** [131]. Με αυτή την τεχνική:

- Αναλύονται οι σχέσεις χρήστη-δικαιώματος: Ο αρχικός δυαδικός πίνακας δεδομένων αποδομείται σε δύο πίνακες χαμηλότερης διάστασης, αποκαλύπτοντας τις κρυφές δομές που συνδέουν χρήστες με ρόλους και ρόλους με δικαιώματα.
- Προσδιορίζεται ο αριθμός των ρόλων: Χρησιμοποιούνται μέθοδοι όπως το Elbow και το Silhouette [132] για τον προσδιορισμό του βέλτιστου αριθμού ρόλων που εξάγονται από τα δεδομένα.
- Ελαχιστοποίηση σφαλμάτων: Η μέθοδος προσαρμόζεται για να διασφαλίζει τη μέγιστη δυνατή ακρίβεια (σχεδόν 100%) μέσω της επαναπροσδιορισμού των συσχετίσεων χρήστη-ρόλου και ρόλου-δικαιώματος.

Η μεθοδολογία αυτή προσφέρει μια αποτελεσματική και επεκτάσιμη λύση για την εξαγωγή ρόλων, αξιοποιώντας τεχνικές τεχνητής νοημοσύνης και ανάλυσης δεδομένων. Οι επιστήμονες φιλοδοξούν να επεκτείνουν το έργο τους με νέες μεθόδους, όπως autoencoders, για ακόμα καλύτερη ακρίβεια και απόδοση.

Οι ερευνητές [133] στοχεύουν να αντιμετωπίσουν τις προκλήσεις που παρουσιάζονται στον τομέα της έξυπνης παραγωγής, η οποία χαρακτηρίζεται από αυξημένη δυναμικότητα και εξατομίκευση. Καθώς οι παραδοσιακές μέθοδοι ελέγχου πρόσβασης, όπως το Role-Based Access Control (RBAC), δεν επαρκούν για να διαχειριστούν τη δυναμική φύση των σύγχρονων βιομηχανικών συστημάτων, προτείνεται η χρήση του **Attribute-Based Access Control (ABAC)** [134]. Το σκεπτικό τους βασίζεται στην πεποίθηση ότι το ABAC, χάρη στη χρήση χαρακτηριστικών για την περιγραφή των χρηστών,

των αντικειμένων και του περιβάλλοντος, μπορεί να εξυπηρετήσει τις απαιτήσεις ευελιξίας και ασφάλειας.

Οι κύριοι στόχοι των επιστημόνων περιλαμβάνουν την αξιολόγηση της ανταπόκρισης του ABAC σε απαιτήσεις ελέγχου πρόσβασης στον τομέα της έξυπνης παραγωγής, όπως η ευελιξία, η κλιμακωσιμότητα και η αποδοτικότητα. Επίσης, στοχεύουν στη μείωση της πολυπλοκότητας στη διαχείριση πολιτικών και χαρακτηριστικών, την προσαρμογή του μοντέλου σε δυναμικές συνθήκες και την ανάπτυξη εργαλείων και τεχνικών που θα διευκολύνουν την εφαρμογή του ABAC σε βιομηχανικά περιβάλλοντα.

Οι ερευνητές χρησιμοποίησαν μοντέλα και προσεγγίσεις που περιλαμβάνουν τεχνολογίες AI για να ενισχύσουν τη δυναμικότητα του ABAC. Συγκεκριμένα:

1. **Μοντέλα με Βάση τη Λογική**: Εξετάστηκαν προηγμένα μοντέλα όπως το **Model Driven Security (MDS)** [135], που χρησιμοποιεί μοντελοποίηση υψηλού επιπέδου για τη δημιουργία και εφαρμογή κανόνων.
2. **Χρήση Προσαρμοστικών Μοντέλων**: Υιοθέτησαν στοιχεία από το **Usage Control (UCON)** [136] για την ενσωμάτωση δυναμικών υποχρεώσεων και αλλαγών στα χαρακτηριστικά βάσει αποφάσεων.
3. **Προσομοιώσεις σε Πλατφόρμες AI**: Προγραμματίζεται η χρήση εργαλείων όπως το **Policy Machine του NIST**, το οποίο επιτρέπει τη διαχείριση και ανάλυση πολιτικών σε πραγματικές συνθήκες.
4. **Αυτοματισμός Διαχείρισης Πολιτικών**: Εξετάστηκε η δυνατότητα χρήσης τεχνικών AI για την αυτόματη διαχείριση πολιτικών και χαρακτηριστικών, περιορίζοντας την ανάγκη ανθρώπινης παρέμβασης.

3.3 Πρόληψη Διαρροής Δεδομένων

Η πρόληψη διαρροής δεδομένων (Data Leakage Prevention) αφορά την ανίχνευση και προστασία από παραβιάσεις δεδομένων, την εξαγωγή ή την ανεπιθύμητη καταστροφή τους. Οι τεχνικές τεχνητής νοημοσύνης χρησιμοποιούνται για την παρακολούθηση της πρόσβασης στα δεδομένα, της μετακίνησής τους και της δραστηριότητας των χρηστών. Επίσης, αξιοποιούνται για την αυτόματη ανίχνευση ευαίσθητων δεδομένων και τον εντοπισμό προηγμένων επίμονων απειλών (Advanced Persistent Threats - APT) με στόχο την αποτροπή διαρροών.

Η αναγνώριση των εξουσιοδοτημένων χρηστών και η παρακολούθηση του τρόπου με τον οποίο χειρίζονται ευαίσθητες πληροφορίες προσφέρουν ακριβείς πληροφορίες για την πρόληψη διαρροών δεδομένων, εντοπίζοντας ανησυχητικές συμπεριφορές ή δραστηριότητες.

Οι επιστήμονες στο συγκεκριμένο ερευνητικό έργο [137] επιδιώκουν να αναπτύξουν ένα αποδοτικό σύστημα ανίχνευσης απειλών εκ των έσω (insider threats) σε οργανισμούς, χρησιμοποιώντας μεθόδους μηχανικής μάθησης. Το κίνητρο πίσω από την έρευνά τους είναι η ανάγκη εντοπισμού κακόβουλων ενεργειών μέσα σε έναν οργανισμό, οι οποίες μπορεί να προέρχονται από υπαλλήλους

ή συνεργάτες. Αυτές οι ενέργειες είναι δύσκολο να ανιχνευθούν, καθώς συνήθως μοιάζουν με φυσιολογικές δραστηριότητες, ενώ μπορεί να προκαλέσουν σημαντικές ζημιές.

Οι επιστήμονες στοχεύουν να δημιουργήσουν ένα σύστημα που μπορεί να εντοπίζει εσωτερικές απειλές χωρίς να εξαρτάται από προσημειωμένα δεδομένα. Αυτό επιτυγχάνεται με τη χρήση μη επιβλεπόμενης μάθησης, που επιτρέπει την ανίχνευση ανωμαλιών σε δεδομένα χωρίς την ανάγκη προκαθορισμένων ετικετών. Ιδιαίτερη σημασία δίνεται στην ανθεκτικότητα του συστήματος απέναντι σε κακόβουλους χρήστες εντός του εκπαιδευτικού συνόλου και στη δυνατότητά του να ανιχνεύει αλλαγές στη συμπεριφορά των χρηστών, λαμβάνοντας υπόψη χρονικά μοτίβα. Επιπλέον, επιδιώκεται η γενικευσιμότητα, ώστε το σύστημα να λειτουργεί αξιόπιστα σε διαφορετικά περιβάλλοντα και εφαρμογές.

Οι ερευνητές χρησιμοποίησαν τέσσερις διαφορετικούς αλγορίθμους ανίχνευσης ανωμαλιών, οι οποίοι βασίζονται σε διάφορες αρχές της τεχνητής νοημοσύνης [138]:

1. **Autoencoders (AE)**: Ένα μοντέλο νευρωνικών δικτύων που συμπιέζει και ανασυγκροτεί τα δεδομένα, ώστε να εντοπίζει αποκλίσεις από τη φυσιολογική συμπεριφορά.
2. **Isolation Forest (IF)**: Ένας αλγόριθμος βασισμένος σε δέντρα απόφασης, ο οποίος απομονώνει ανωμαλίες από τα φυσιολογικά δεδομένα.
3. **Local Outlier Factor (LOF)**: Ένας αλγόριθμος που υπολογίζει την πυκνότητα γύρω από κάθε σημείο για να εντοπίσει ανωμαλίες με βάση τη γειτονιά τους.
4. **Lightweight Online Detector of Anomalies (LODA)**: Ένας γρήγορος και αποδοτικός αλγόριθμος που χρησιμοποιεί γραμμικές προβολές για την ανίχνευση ανωμαλιών.

Η μελέτη δείχνει ότι οι μέθοδοι που χρησιμοποιούν ποσοστιαία αναπαράσταση δεδομένων και συνδυασμούς αλγορίθμων επιτυγχάνουν εξαιρετικά αποτελέσματα, ακόμη και σε συνθήκες με πολλούς κακόβουλους χρήστες στο εκπαιδευτικό σύνολο. Επιπλέον, το σύστημα παρουσίασε ικανότητα να γενικεύει, δηλαδή να λειτουργεί εξίσου καλά σε διαφορετικά datasets, κάτι που είναι σημαντικό για πραγματικές εφαρμογές.

Σε μία άλλη έρευνα [139], οι επιστήμονες αναγνωρίζουν ότι οι εσωτερικές απειλές αποτελούν σοβαρό κίνδυνο για οργανισμούς, δεδομένου ότι προέρχονται από άτομα με νόμιμη πρόσβαση στα συστήματα. Για να εντοπίσουν αυτές τις απειλές, οι ερευνητές βασίζονται στη συμπεριφορά των χρηστών και χρησιμοποιούν δεδομένα για να δημιουργήσουν πρότυπα κανονικής και ανώμαλης δραστηριότητας. Λόγω της έλλειψης επαρκών δεδομένων για κακόβουλες συμπεριφορές (που είναι σπάνιες), στρέφονται στη χρήση αλγορίθμων ανίχνευσης ανωμαλιών που δεν απαιτούν προαπαιτούμενη σήμανση των δεδομένων.

Ο στόχος τους είναι να αναπτύξουν ένα σύστημα ανίχνευσης εσωτερικών απειλών που να εντοπίζει ανώμαλες και πιθανώς κακόβουλες δραστηριότητες με έμφαση στους πραγματικούς κινδύνους. Το ενδιαφέρον τους επικεντρώνεται στη διαχείριση ετερογενών δεδομένων από διαφορετικές πηγές, όπως ημερήσιες δραστηριότητες, περιεχόμενα email και δίκτυα επικοινωνίας, ενώ παράλληλα επιδιώκουν την εξοικονόμηση πόρων, περιορίζοντας την παρακολούθηση σε ένα μικρό ποσοστό ύποπτων περιπτώσεων με υψηλή ακρίβεια. Επιπλέον, στοχεύουν στη βελτίωση της απόδοσης μέσω συνδυασμού αλγορίθμων και τεχνικών ανίχνευσης ανωμαλιών.

Οι μέθοδοι Τεχνητής Νοημοσύνης που χρησιμοποιήθηκαν είναι οι εξής:

1. **Αλγόριθμοι Ανίχνευσης Ανωμαλιών**:
    1. **Parzen Window**: Αποτελεσματικός στην εκτίμηση πιθανότητας για κατανομές δεδομένων χωρίς αυστηρές υποθέσεις, ιδιαίτερα κατάλληλος για μη κανονικές κατανομές.
    2. **Gaussian (Gauss)**: Βασισμένος σε κανονικές κατανομές, αλλά λιγότερο αποδοτικός για πιο σύνθετα δεδομένα.
    3. **K-Means Clustering (KMC)**: Αλγόριθμος ομαδοποίησης που βοηθά στον εντοπισμό ανώμαλων συμπεριφορών βασισμένων στην απόκλιση από κεντρικά σημεία ομάδων.
2. **Ανάλυση Κυρίων Συνιστωσών (PCA)**:
   Χρησιμοποιείται για τη μείωση της διάστασης των δεδομένων και την ενίσχυση της απόδοσης των αλγορίθμων ανίχνευσης ανωμαλιών.

3. **Συνδυασμοί Αλγορίθμων (Ensemble Models)**:
   Όπως το «Parzen + PCA», που απέδειξε ότι υπερτερεί σε περιπτώσεις όπου οι μεμονωμένοι αλγόριθμοι αποτυγχάνουν να αποδώσουν αποτελεσματικά.

Μια αυξανόμενη απειλή στην οποία εστιάζουν οι επιστήμονες, είναι η ανίχνευση διαρροών δεδομένων από κακόβουλους υπαλλήλους [140]. Αναγνωρίζουν ότι τα ανισόρροπα δεδομένα και η μεροληψία κατά την κωδικοποίηση χαρακτηριστικών αποτελούν σημαντικές προκλήσεις στην αποτελεσματική εκπαίδευση μοντέλων μηχανικής μάθησης. Μέσω της κατάλληλης προεπεξεργασίας δεδομένων και της χρήσης ισχυρών αλγορίθμων, επιδιώκουν να ενισχύσουν την ακρίβεια ανίχνευσης τέτοιων περιστατικών, παρέχοντας ένα αξιόπιστο εργαλείο ανίχνευσης απειλών.

Ο κύριος στόχος είναι η ανάπτυξη ενός μοντέλου μηχανικής μάθησης που μπορεί να ανιχνεύσει αποτελεσματικά περιστατικά διαρροής δεδομένων από υπαλλήλους κατά την κρίσιμη περίοδο πριν εγκαταλείψουν μια εταιρεία.

Για να το πετύχουν αυτό, χρησιμοποιούν τεχνικές επεξεργασίας δεδομένων, αλγορίθμους μηχανικής μάθησης και μεθόδους για να αξιολογήσουν την απόδοσή τους. Πιο αναλυτικά:

Τεχνικές Επεξεργασίας Δεδομένων

1. Label Encoding: Για την αριθμητική αναπαράσταση κατηγορικών δεδομένων.
2. One-hot Encoding: Για την αποφυγή μεροληψίας κατά την κωδικοποίηση χαρακτηριστικών.
3. SMOTE (Synthetic Minority Oversampling Technique) [141]: Για την εξισορρόπηση του ανισόρροπου συνόλου δεδομένων.

Αλγόριθμοι Μηχανικής Μάθησης (ML)

1. Logistic Regression (LR)
2. Decision Tree (DT)
3. Random Forest (RF)
4. Naive Bayes (NB)
5. k-Nearest Neighbors (KNN)
6. Kernel Support Vector Machine (KSVM)

Μέθοδοι Αξιολόγησης Απόδοσης

1. Precision

2. Recall
3. F-measure
4. AUC-ROC Curve (το οποίο χρησιμοποιείται για την εκτίμηση της ακρίβειας και της ευαισθησίας του μοντέλου) [142].

Συνεπώς, η μέθοδος SMOTE αποδείχθηκε αποτελεσματική για τη διαχείριση ανισορροπίας στις κατηγορίες, ενώ η χρήση one-hot encoding αντιμετώπισε τα θέματα προκατάληψης στα χαρακτηριστικά. Το μοντέλο υπερέχει συγκριτικά με προηγούμενες έρευνες, παρέχοντας καλύτερα αποτελέσματα στην ανίχνευση διαρροής των δεδομένων.

Η αυτοματοποιημένη ανίχνευση της ευαισθησίας των δεδομένων είναι μια τεχνική που βοηθά στον εντοπισμό και την ταξινόμηση των δεδομένων, κατηγοριοποιώντας τα σε ομάδες όπως εμπιστευτικά, προσωπικά ή δημόσια. Αυτό επιτυγχάνεται μέσω της ανάλυσης και της σήμανσης των δεδομένων με βάση κοινά χαρακτηριστικά. Με αυτόν τον τρόπο, οι τεχνολογίες αποτροπής διαρροής δεδομένων γίνονται πιο αποτελεσματικές, αφού εστιάζουν μόνο σε κρίσιμα τμήματα ευαίσθητων δεδομένων και όχι σε ολόκληρη την πληροφορία.

Οι επιστήμονες [143] αναγνώρισαν την πρόκληση της ανίχνευσης και ταξινόμησης ευαίσθητων δεδομένων σε μεγάλα κείμενα, όπως τα δεδομένα του WikiLeaks. Παρατήρησαν ότι η χρήση ενός ενιαίου ταξινομητή (global classifier) περιορίζει την ακρίβεια, καθώς δεν λαμβάνει υπόψη τη διαφορετικότητα των τοπικών χαρακτηριστικών στα δεδομένα. Σκοπός τους ήταν να σχεδιάσουν ένα σύστημα που να μπορεί να συνδυάζει παγκόσμιες πληροφορίες (global features) για τη γενική κατανόηση του κειμένου με τοπικές πληροφορίες (local features) που ενισχύουν την ακρίβεια σε συγκεκριμένα μέρη των δεδομένων.

Οι στόχοι της έρευνας περιλαμβάνουν την αύξηση της ακρίβειας και της ευαισθησίας στην ταξινόμηση ευαίσθητων δεδομένων, ξεπερνώντας τις επιδόσεις των παραδοσιακών μονολιθικών ταξινομητών. Επιπλέον, στοχεύουν στην αυτοματοποίηση της διαδικασίας ταξινόμησης, επιτρέποντας την άμεση ανάλυση παραγράφων ανά επίπεδο ευαισθησίας. Ένας κρίσιμος στόχος είναι η διατήρηση ισορροπίας μεταξύ γενίκευσης και εξειδίκευσης, μέσω της σωστής επιλογής αριθμού clusters, καθώς και η δυνατότητα εφαρμογής του συστήματος σε άλλους τομείς, διευκολύνοντας την προσαρμογή σε διαφορετικά είδη ευαίσθητων δεδομένων.

Οι επιστήμονες, προκειμένου να υλοποιήσουν τους στόχους της έρευνάς τους, χρησιμοποιούν τις παρακάτω μεθόδους:

1. **Clustering (Ομαδοποίηση) με k-Means**:
   Το κείμενο χωρίζεται σε clusters με βάση την ομοιότητα χαρακτηριστικών. Κάθε cluster αντιπροσωπεύει ένα υποσύνολο δεδομένων με κοινά θέματα, επιτρέποντας πιο εξειδικευμένη ανάλυση.
   Ο αριθμός των clusters επιλέγεται με cross-validation, και παρατηρήθηκε ότι η βέλτιστη τιμή εξαρτάται από το μέγεθος των δεδομένων (περίπου 0,88% του μεγέθους).
2. **TF-IDF Representation** [144]:
   Χρησιμοποιήθηκε για την εξαγωγή χαρακτηριστικών από τα κείμενα. Ωστόσο, παρατηρήθηκε ότι η μονολιθική εφαρμογή του (global TF-IDF) υποβαθμίζει τη διακριτική ικανότητα, οπότε ενισχύθηκε μέσω τοπικών χαρακτηριστικών.
3. **Συνδυασμός Global και Local Classifiers**:
   Global features: Χρησιμοποιούνται για να εντοπιστούν τα κύρια θέματα και να κατευθυνθεί το ερώτημα στο κατάλληλο cluster.

Local classifiers: Εφαρμόζονται σε κάθε cluster για πιο εξειδικευμένη ανάλυση και ταξινόμηση.

4. **Αξιολόγηση με Precision, Recall και F-Measure**:
   Οι δείκτες αυτοί χρησιμοποιήθηκαν για τη μέτρηση της απόδοσης, με το ACESS να ξεπερνά τους βασικούς ταξινομητές (baseline classifiers).
5. **Latent Semantic Analysis (LSA)**:
   Αναφέρεται ως μελλοντική προοπτική για περαιτέρω βελτίωση της θεματικής μοντελοποίησης.

Στο επόμενο άρθρο που αναφέρεται [145], οι επιστήμονες, έχοντας εντοπίσει τον αυξανόμενο κίνδυνο διαρροής ευαίσθητων πληροφοριών στο διαδίκτυο, σχεδίασαν το **ExSense** για να παρέχουν ένα εργαλείο που ανιχνεύει τέτοιες διαρροές από μη δομημένα δεδομένα. Στόχος τους είναι η βελτίωση της προστασίας της ιδιωτικότητας και η πρόληψη της μη εξουσιοδοτημένης διαρροής προσωπικών, οικονομικών ή διαπιστευτηρίων δεδομένων. Το σύστημα επιδιώκει να προσφέρει ακριβή και ευέλικτη ανίχνευση διαφόρων τύπων ευαίσθητων πληροφοριών, ενώ να είναι ικανό να προσαρμόζεται σε διαφορετικά πλαίσια χρήσης, όπως σε οργανισμούς ή σε διαδικτυακή επιτήρηση δεδομένων.

Για την επίτευξη των στόχων τους, οι επιστήμονες χρησιμοποίησαν τις εξής τεχνικές AI:

1. **BERT (Bidirectional Encoder Representations from Transformers)**: Δυναμική ενσωμάτωση λέξεων που προσαρμόζεται στο πλαίσιο, βελτιώνοντας την κατανόηση της σημασίας κάθε λέξης.
2. **Bi-LSTM (Bidirectional Long Short-Term Memory)**: Μοντέλο αλληλουχίας που λαμβάνει υπόψη τη ροή πληροφοριών πριν και μετά από κάθε λέξη, για καλύτερη ανάλυση συμφραζομένων.
3. **Attention Mechanism**: Μηχανισμός που δίνει έμφαση στις πιο σημαντικές λέξεις σε μια πρόταση, διευκολύνοντας την ανίχνευση κρίσιμων πληροφοριών.
4. **Συνδυασμός περιεχομενικής ανάλυσης (Regular Expressions) και συμφραζομένων (BERT-Bi-LSTM-Attention)**: Ο συνδυασμός απλών κανόνων και προηγμένων μεθόδων AI ενισχύει την ευελιξία και την ακρίβεια του συστήματος.

Αυτές οι μέθοδοι επιτρέπουν στο ExSense να εξάγει με ακρίβεια ευαίσθητες πληροφορίες από μεγάλα και ποικίλα δεδομένα, ενσωματώνοντας τόσο στατικές όσο και δυναμικές τεχνικές ανάλυσης.

Οι Advanced Persistent Threats (APTs) αποτελούν μια μορφή εξειδικευμένων κυβερνοεπιθέσεων που στοχεύουν συγκεκριμένα δίκτυα, παραμένοντας αθέατες για μεγάλο χρονικό διάστημα. Ο κύριος στόχος τους είναι η υποκλοπή δεδομένων, χωρίς να προκαλούν άμεση ζημιά στα συστήματα.

Οι επιστήμονες [146] επικεντρώνονται στην ανάπτυξη ενός μηχανισμού άμυνας κατά των Προηγμένων Επίμονων Απειλών (APT), οι οποίες είναι σύνθετες, στρατηγικά σχεδιασμένες επιθέσεις που απειλούν την ασφάλεια συστημάτων. Αναγνωρίζουν ότι η παραδοσιακή άμυνα δεν είναι επαρκής σε σύνθετα περιβάλλοντα όπως αυτά των edge συσκευών, λόγω της πολυπλοκότητας και του δυναμικού χαρακτήρα των επιθέσεων. Το σκεπτικό τους βασίζεται στην ενσωμάτωση εξηγητικής τεχνητής νοημοσύνης (XAI) για την ενίσχυση της διαφάνειας των αμυντικών μηχανισμών και την αξιοπιστία της λήψης αποφάσεων.

Ο στόχος τους είναι η βελτιστοποίηση της κατανομής πόρων με στρατηγικές που εξισορροπούν την ταχύτητα αντίδρασης και την αποδοτική χρήση των διαθέσιμων μέσων άμυνας. Επικεντρώνονται επίσης στην αυτοματοποίηση και την αξιοπιστία μέσω εργαλείων ανάλυσης CTI που δίνουν έμφαση στη διαφάνεια, μειώνοντας την εξάρτηση από ανθρώπινη παρέμβαση. Τέλος, εξερευνούν μεθόδους

διασφάλισης ιδιωτικότητας κατά την ανίχνευση και αντιμετώπιση επιθέσεων, ιδίως σε περιβάλλοντα edge computing.

Για την έρευνά τους χρησιμοποίησαν τις μεθόδους που καταγράφονται παρακάτω:

1. **Εξηγητική Τεχνητή Νοημοσύνη (XAI)** [147]:
    - Χρήση του εργαλείου LIME (Local Interpretable Model-agnostic Explanations) για την εξήγηση των προβλέψεων του συστήματος ανίχνευσης.
    - Η LIME αναλύει ποια χαρακτηριστικά (π.χ., στατιστικά δεδομένα δικτύου) συνεισφέρουν περισσότερο στην απόφαση του μοντέλου.
2. **Μοντέλο Ανίχνευσης Βασισμένο σε LSTM (Long Short-Term Memory)**:
    - Ένα νευρωνικό δίκτυο LSTM χρησιμοποιήθηκε για την επεξεργασία χρονοσειρών και τη λήψη αποφάσεων σχετικά με το αν η δραστηριότητα είναι "καλοήθης" ή "κακόβουλη".
    - Το μοντέλο διαθέτει πολλαπλά επίπεδα, όπως προσοχή χαρακτηριστικών, διαδοχικές στρώσεις LSTM και ενεργοποιήσεις sigmoid.
3. **Παιχνίδι Stackelberg (Edge Bayesian Stackelberg Game)**:
    - Μια μέθοδος από τη θεωρία παιχνιδιών που χρησιμοποιείται για τη μοντελοποίηση της αλληλεπίδρασης μεταξύ επιτιθέμενου και αμυνόμενου. Το παιχνίδι υπολογίζει την ισορροπία (equilibrium) των στρατηγικών για βέλτιστη κατανομή πόρων άμυνας.
4. **Ανάλυση Πληροφοριών Απειλών (CTI)** [148]:
    - Ενσωμάτωση εξηγητικών αλγορίθμων για τη δημιουργία και αξιολόγηση αυτοματοποιημένων πληροφοριών απειλών που σχετίζονται με τύπους επιθέσεων, επίπεδα κινδύνου και ευάλωτες περιοχές.
5. **Αλγόριθμοι Συγκριτικής Απόδοσης**:
    - Εφαρμογή και σύγκριση μεθόδων όπως Greedy Mechanism και Hotbooting PHC για την αξιολόγηση της απόδοσης του προτεινόμενου μηχανισμού.

Με αυτές τις μεθόδους, οι επιστήμονες επιδιώκουν να επιτύχουν ένα σύστημα που είναι αποτελεσματικό, εξηγητικό, αυτοματοποιημένο και ευέλικτο στην αντιμετώπιση απειλών, ενώ παράλληλα προστατεύει τα δεδομένα και την ιδιωτικότητα των χρηστών.

Οι Ahmed Abdulrahman Alghamdi και Giles Reger στην έρευνά τους [149] αναγνωρίζουν ότι τα αρχεία καταγραφής από διαφορετικές πηγές (όπως Apache logs, Snort logs, και firewall logs) περιέχουν πολύτιμες πληροφορίες για την ανίχνευση επιθέσεων, αλλά είναι δύσκολο να αναλυθούν λόγω της ετερογένειάς τους και του μεγάλου όγκου δεδομένων. Στοχεύουν στη δημιουργία ενός πλαισίου που θα μπορεί να συνδυάσει δεδομένα από αυτές τις πηγές, να αναγνωρίσει πρότυπα κακόβουλων ενεργειών και να εξάγει συσχετισμούς μεταξύ των επιθέσεων, ενισχύοντας έτσι την ασφάλεια των συστημάτων. Βασίζονται στην υπόθεση ότι η αποτελεσματική προεπεξεργασία και ανάλυση μπορεί να προσφέρει σαφείς ενδείξεις για τις επιθέσεις και τις συμπεριφορές που τις συνοδεύουν.

Οι επιστήμονες χρησιμοποίησαν μεθόδους Τεχνητής Νοημοσύνης, προκειμένου να δημιουργήσουν ένα ολοκληρωμένο σύστημα το οποίο να αντιμετωπίζει και να κατανοεί τα σύνθετα μοτίβα επιθέσεων.

1. **Clustering (ομαδοποίηση)**:
   Χρησιμοποιήθηκε ο αλγόριθμος DBSCAN (Density-Based Spatial Clustering of Applications with Noise) για την ανίχνευση μοτίβων συμπεριφοράς, τόσο σε μεμονωμένα αρχεία (Phase 1) όσο και σε συνδυασμένα δεδομένα (Phase 2).

2. **Εκτίμηση Παραμέτρων**:
   Υπολογισμός κρίσιμων παραμέτρων όπως το EPS (ελάχιστη απόσταση μεταξύ δεδομένων), που είναι ζωτικής σημασίας για τον DBSCAN.

3. **Αξιολόγηση Απόδοσης**: Χρησιμοποιήθηκαν μετρικές όπως:

   - Ομοιογένεια (Homogeneity): Κατά πόσο κάθε cluster περιέχει γεγονότα της ίδιας κατηγορίας.
   - Πληρότητα (Completeness): Κατά πόσο όλα τα γεγονότα μιας κατηγορίας βρίσκονται στο ίδιο cluster.
   - V-measure: Ο μέσος όρος ομοιογένειας και πληρότητας.
   - Adjusted Rand Index (ARI) και Adjusted Mutual Info (AMI): Εξετάζουν τη συνολική ακρίβεια ομαδοποίησης σε σύγκριση με τα πραγματικά δεδομένα.

4. **Προεπεξεργασία δεδομένων**: Επεξεργασία δεδομένων για την αντιμετώπιση ετερογένειας, συμπλήρωση κενών, μετατροπή χρόνων σε κοινή μορφή και κωδικοποίηση δεδομένων (π.χ., IP διευθύνσεις, text δεδομένα).

3.4 Σχέδιο Διαχείρισης Ευπαθειών με Ενίσχυση Τεχνητής Νοημοσύνης

Το σχέδιο διαχείρισης ευπαθειών (AI-Enhanced Vulnerability Management) έχει ως στόχο να μειώσει τους κινδύνους που μπορούν να επηρεάσουν ή να διαταράξουν τη λειτουργία ενός συστήματος. Καθώς οι ευπάθειες αυξάνονται διαρκώς, είναι απαραίτητο αυτό το σχέδιο να προσαρμόζεται στις ανάγκες και στους βασικούς στόχους του συστήματος. Επιπλέον, οι ερευνητές αξιοποιούν την τεχνητή νοημοσύνη για να υπολογίζουν κινδύνους με βάση το συγκεκριμένο πλαίσιο λειτουργίας και να αναλύουν πώς οι ευπάθειες μπορούν να αξιοποιηθούν από πιθανούς επιτιθέμενους.

Οι επιστήμονες [150] αναγνωρίζουν ότι η παραδοσιακή αξιολόγηση κινδύνων για ευπάθειες, όπως το CVSS, έχει περιορισμούς, κυρίως επειδή επικεντρώνεται μόνο στη σοβαρότητα των ευπαθειών χωρίς να συνυπολογίζει την πιθανότητα εκμετάλλευσης και την κρισιμότητα της εκμετάλλευσης σε πραγματικά σενάρια. Στόχος τους είναι να αναπτύξουν μια λύση που να βασίζεται στα ιστορικά δεδομένα απειλών ενός συστήματος, για να βελτιώσουν τη διαδικασία ιεράρχησης των κινδύνων και να διευκολύνουν τη διαχείριση ευπαθειών.

Η κύρια φιλοσοφία τους είναι ότι τα ιστορικά δεδομένα απειλών μπορούν να προσφέρουν πολύτιμες πληροφορίες για την εκτίμηση μελλοντικών κινδύνων και ότι η ενσωμάτωση αυτών των δεδομένων σε ένα σύστημα αξιολόγησης κινδύνων μπορεί να οδηγήσει σε καλύτερα αποτελέσματα.

Οι μέθοδοι Τεχνητής Νοημοσύνης που χρησιμοποιήθηκαν είναι οι εξής:

1. **Νευρο-Συμβολικό Μοντέλο (NN-PLP)**:

Αποτελεί ένα υβριδικό μοντέλο το οποίο συνδυάζει τεχνικές βαθιάς μάθησης (Neural Networks - NN) με συμβολική λογική (Probabilistic Logic Programming - PLP). Η βαθιά μάθηση χρησιμοποιείται για την εκμάθηση χαρακτηριστικών απειλών από ιστορικά δεδομένα. Η λογική βασίζεται σε κανόνες που κωδικοποιούν σχέσεις μεταξύ των απειλών και των ευπαθειών.

2. **Μοντελοποίηση Απειλών (Threat Modeling)**:

Η μέθοδος κωδικοποιεί τα χαρακτηριστικά ιστορικών απειλών (π.χ. συστήματα-στόχοι, πιθανότητα εκμετάλλευσης) ως δεδομένα για εκπαίδευση. Αυτή η διαδικασία επιτρέπει στο μοντέλο να συνδυάζει τη γνώση από το παρελθόν για να αξιολογήσει τις μελλοντικές ευπάθειες.

3. **Συνδυασμός Χαρακτηριστικών LSA και CVSS**:

Τα χαρακτηριστικά LSA (Latent Semantic Analysis) εξάγουν μοτίβα πιθανότητας εκμετάλλευσης. Τα χαρακτηριστικά CVSS ενισχύουν την κατανόηση της σοβαρότητας/κρισιμότητας. Η χρήση και των δύο παρέχει πιο ολοκληρωμένη αξιολόγηση κινδύνων.

4. **Ενιαία Εξαγωγή Κανόνων**:

Στην ενιαία εξαγωγή κανόνων, το σύστημα χρησιμοποιεί κανόνες ενός βήματος για να συνδυάσει δεδομένα από ιστορικές απειλές με τα νέα δεδομένα. Αυτό επιτρέπει τη γρήγορη λήψη αποφάσεων.

Στην επόμενη έρευνά τους [151], οι επιστήμονες ξεκίνησαν με τη διαπίστωση ότι το υφιστάμενο πρότυπο αξιολόγησης ευπαθειών, το CVSS, αποδεικνύεται αναποτελεσματικό για την ακριβή πρόβλεψη της εκμεταλλεύσιμης φύσης των ευπαθειών. Θεωρούν ότι η ακριβής πρόβλεψη της εκμεταλλευσιμότητας μιας ευπάθειας είναι κρίσιμη για τη σωστή ιεράρχηση και διαχείριση των απειλών από τους παρόχους λογισμικού. Με βάση αυτό, επιδιώκουν να αναπτύξουν μια πιο αποδοτική και αξιόπιστη προσέγγιση χρησιμοποιώντας τεχνικές Τεχνητής Νοημοσύνης (AI).

1. **BERT (Bidirectional Encoder Representations from Transformers)**:

Το BERT αποτελεί ένα βασικό μοντέλο που χρησιμοποιήθηκε για την εξαγωγή χαρακτηριστικών από τις περιγραφές ευπαθειών. Μεταφέρθηκε και προσαρμόστηκε με fine-tuning, ώστε να αξιοποιηθεί πλήρως για το συγκεκριμένο πρόβλημα.

2. **Pooling Strategies**:

Στο pooling strategies φαρμόστηκαν διαφορετικές μέθοδοι εξαγωγής ενσωματώσεων (embeddings) από το BERT:

- **[CLS] token embedding**: Καλύτερη απόδοση σε fine-tuned BERT.
- **Mean pooling**: Αποτελεσματικότερη σε pre-trained μοντέλα.
- **Max pooling**: Απέδωσε χειρότερα και στα δύο είδη μοντέλων.

3. **LSTM (Long Short-Term Memory)**:

Ο λόγος για τον οποίο χρησιμοποιήθηκε ο συγκεκριμένος ταξινομητής, βασίζεται στο ότι είναι κατάλληλος για τη διαχείριση ακολουθιακών δεδομένων και την ανάλυση εξαρτήσεων μέσα στα δεδομένα.

4. **Σύγκριση με άλλες μεθόδους**:

Συγκρίθηκαν τα αποτελέσματα του BERT με προηγούμενα μοντέλα που βασίζονται σε διαφορετικές τεχνικές εξαγωγής χαρακτηριστικών και ταξινόμησης. Το BERT υπερείχε σε όλα τα μέτρα απόδοσης.

Οι επιστήμονες [152] εντόπισαν ότι τα δεδομένα σε πολλές εφαρμογές μεταβάλλονται με την πάροδο του χρόνου (concept drift), γεγονός που καθιστά δύσκολη την ακρίβεια και τη γενίκευση των αλγορίθμων μηχανικής μάθησης. Αναγνώρισαν επίσης την ύπαρξη προβλημάτων, όπως η ανισορροπία μεταξύ των κατηγοριών και η αλλαγή στις ετικέτες των δεδομένων. Το σκεπτικό τους ήταν να σχεδιάσουν έναν ευέλικτο και δυναμικό αλγόριθμο που να μπορεί να προσαρμόζεται σε πραγματικό χρόνο, διατηρώντας υψηλή απόδοση ανεξάρτητα από το μέγεθος και τη φύση των αλλαγών στα δεδομένα.

Ο στόχος τους ήταν η δημιουργία ενός ισχυρού και προσαρμοστικού αλγορίθμου που να μπορεί να αντιμετωπίσει τις προκλήσεις της μεταβολής δεδομένων (concept drift) και της ανισορροπίας κατηγοριών, χρησιμοποιώντας στρατηγικές βασισμένες στη μηχανική μάθηση και ανάλυση υπερπαραμέτρων.

Γι’ αυτό τον λόγο χρησιμοποίησαν έναν δυναμικό προσαρμοστικό αλγόριθμο εκμάθησης εννοιών, για την πρόβλεψη της εκμεταλλευσιμότητας (Real-time Dynamic Concept Adaptive Learning algorithm). Ο αλγόριθμος αυτός βασίστηκε στις παρακάτω μεθόδους:

1. **Class Rectification Strategy (CRS)**:

Αυτή η μέθοδος στοχεύει στη διόρθωση των αλλαγών στις ετικέτες των δειγμάτων (actual drift) και στη βελτίωση της ισορροπίας μεταξύ ακρίβειας και ανάκλησης. Συγκεκριμένα, χρησιμοποιεί μηχανισμούς επανεκτίμησης της κατανομής των δεδομένων σε κάθε παρτίδα, ώστε να προσαρμόζεται στις αλλαγές.

2. **Balanced Window Strategy (BWS)** [153]:

Αυτή η στρατηγική διατηρεί ισορροπημένα δείγματα μεταξύ των κατηγοριών σε ένα "παράθυρο" δεδομένων, διασφαλίζοντας ότι δεν παρατηρείται υπερεκπαίδευση της μειοψηφίας. Εισάγει έναν παράγοντα αποδυνάμωσης (time decay factor) για τη μείωση της σημασίας παλαιότερων δεδομένων, αποτρέποντας την υπερπροσαρμογή.

3. **Διαδοχική μάθηση (Consecutive Batch Learning)**:

Ο RDCAL εφαρμόζεται σε διαδοχικές παρτίδες δεδομένων, επιτρέποντας την εκμάθηση και προσαρμογή σε πραγματικό χρόνο χωρίς να απαιτείται επανεκπαίδευση από το μηδέν.

4. **Υποστήριξη διαφορετικών ταξινομητών**:

Οι επιστήμονες δοκίμασαν τον RDCAL με διάφορους τύπους μοντέλων μηχανικής μάθησης (Νευρωνικά Δίκτυα, SVM, Δέντρα Απόφασης, Λογιστική Παλινδρόμηση) για να επιβεβαιώσουν την ανεξαρτησία του από τον τύπο του ταξινομητή.

5. **Υπερπαράμετροι (Hyperparameters)**:

Έγινε ανάλυση για την εύρεση βέλτιστων τιμών για τις υπερπαραμέτρους (Na, Nb, a) μέσω πειραμάτων και αφαίρεσης στοιχείων, ώστε να μεγιστοποιηθεί η ακρίβεια και η προσαρμοστικότητα.

3.5 Ανάλυση Αρχείων Καταγραφής

Η ανάλυση αρχείων καταγραφής (Log Analysis) είναι η διαδικασία εξέτασης των δεδομένων που δημιουργούν οι υπολογιστές, με στόχο την αναγνώριση προβλημάτων, ζητημάτων ασφαλείας ή άλλων πιθανών κινδύνων. Τα εργαλεία που χρησιμοποιούν τεχνητή νοημοσύνη διευκολύνουν τη διαχείριση αυτών των αρχείων, αυτοματοποιώντας τις επαναλαμβανόμενες εργασίες και αντιμετωπίζοντας αποτελεσματικά μεγάλες ποσότητες δεδομένων από κατανεμημένα συστήματα.

Οι επιστήμονες αναγνωρίζουν ότι το **RDP** (Remote Desktop Protocol) αποτελεί βασικό εργαλείο για την πλαγία κίνηση (lateral movement) στις επιθέσεις τύπου APT (Advanced Persistent Threat) [154]. Επομένως, η ανάγκη ανίχνευσης κακόβουλων συνεδριών είναι κρίσιμη για την ενίσχυση της κυβερνοασφάλειας. Η κεντρική τους ιδέα είναι να αξιοποιήσουν τα Windows event logs για να εντοπιστούν μη φυσιολογικές συμπεριφορές στις RPD συνεδρίες. Ωστόσο αντιμετωπίζουν προβλήματα που έχουν να κάνουν με την ελαχιστοποίηση των false negative αποτελεσμάτων καθώς και ευπαθειών που παρουσιάζουν τα υπάρχοντα μοντέλα σε επιθέσεις τύπου zero-day. Οπότε ο στόχος τους είναι να τα αντιμετωπίσουν, κάνοντας χρήση μεθόδων τεχνητής νοημοσύνης.

1. **Δοκιμή διαφορετικών αλγορίθμων μηχανικής μάθησης**:
   - **LogitBoost (LB)**: Επιλέχθηκε ως το βέλτιστο μοντέλο λόγω υψηλής ακρίβειας, ανάκλησης και σταθερότητας.
   - **Άλλα μοντέλα όπως**:
     - Random Forest (RF)
     - Light Gradient Boosting Machine (LGBM)
     - Gaussian Naive Bayes (GNB)
     - Decision Tree (DT)
     - Feedforward Neural Network (FNN)
     - Logistic Regression (LR)
2. **Συνδυαστικά μοντέλα (Ensemble Learning)**:
   - Majority Voting (MV): Συνδυασμός όλων των αλγορίθμων αρχικά, αλλά με χαμηλή απόδοση λόγω αδύναμων μοντέλων.
   - Weighted Voting (WV): Κατανομή βαρών στα μοντέλα με βάση την απόδοσή τους.
   - Conservative Approach (CA): Μια πιο "επιφυλακτική" στρατηγική ψήφου όπου μια κακόβουλη συνεδρία επισημαίνεται αν οποιοδήποτε μοντέλο την αναγνωρίσει ως τέτοια.
3. **Βελτιστοποίηση υπερπαραμέτρων (Hyperparameter Tuning)**:
   - Στο LB μοντέλο χρησιμοποιήθηκε ρύθμιση του αριθμού εκτιμητών (estimators), επιτυγχάνοντας την καλύτερη απόδοση με 100 εκτιμητές.
4. **Ανθεκτικότητα σε επιθέσεις**:
   - Χρήση δεδομένων με πολυμορφικά χαρακτηριστικά (manipulated data) για τη δοκιμή της ευρωστίας του μοντέλου.

Το σκεπτικό των ερευνητών [155] βασίζεται στην ανάγκη βελτίωσης της διαχείρισης κυβερνοαπειλών και της ταχύτητας απόκρισης σε περιστατικά. Αναγνώρισαν ότι οι παραδοσιακές μέθοδοι απαιτούν χειροκίνητη αναζήτηση πληροφοριών, γεγονός που:

- Επιβαρύνει τους αναλυτές ασφαλείας με υψηλό γνωστικό φόρτο.
- Καθυστερεί την κατανόηση και την απόκριση σε περιστατικά.
- Δεν προάγει την κατανόηση για μη ειδικούς στον τομέα της κυβερνοασφάλειας.

Η υιοθέτηση ενός αφηγηματικού (storytelling) μοντέλου, που αυτοματοποιεί την άντληση και παρουσίαση πληροφοριών, κρίθηκε σημαντική για την αντιμετώπιση αυτών των προβλημάτων.

Οι επιστήμονες χρησιμοποίησαν μεθόδους Τεχνητής Νοημοσύνης (AI) για να υποστηρίξουν τη δημιουργία του μοντέλου αφήγησης:

1. **Αυτόματη Εξαγωγή Πληροφοριών**:
    - Εφαρμογή αλγορίθμων εξόρυξης δεδομένων (data mining) για τη συλλογή και οργάνωση πληροφοριών από log files.
    - Συνδυασμός δεδομένων από τοπικές (local) και παγκόσμιες (global) βάσεις γνώσης.
2. **Γλωσσική Επεξεργασία (NLP)**:
    - Χρήση τεχνικών Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing) για τη δημιουργία αναφορών σε αφηγηματική μορφή.
    - Μετατροπή τεχνικών δεδομένων σε κατανοητά κείμενα για διαφορετικά επίπεδα χρηστών.
3. **Αυτόματη Συμπλήρωση (Auto-Fill)**:
    - Αξιοποίηση συστημάτων αυτόματης συμπλήρωσης για την κάλυψη των κενών πληροφοριών.
    - Ενσωμάτωση συστάσεων δράσεων και προσδιορισμός υπεύθυνων.
4. **Προσαρμοστικότητα Μοντέλου**:
    - Δημιουργία δυναμικού συστήματος που μπορεί να προσαρμοστεί σε διαφορετικούς τύπους περιστατικών και κλάδους.

Σε επόμενη έρευνά τους [156] οι επιστήμονες θέλησαν να αναπτύξουν μια μεθοδολογία που να μπορεί να αντιμετωπίσει προβλήματα ταξινόμησης κειμένου από διαφορετικούς τομείς με μια ενιαία προσέγγιση. Στόχος τους ήταν να δημιουργήσουν ένα σύστημα που δεν εξαρτάται από το είδος των δεδομένων και το πλαίσιο τους, μειώνοντας έτσι την ανάγκη για προσαρμογές και προκαταρκτική επεξεργασία. Αυτή η γενικότητα αποτελεί βασικό πλεονέκτημα, καθώς οι περισσότεροι κλάδοι συχνά απαιτούν προσαρμοσμένες και ειδικές μεθόδους.
Η μεθοδολογία τους βασίζεται στην ιδέα της Κανονικοποιημένης Εξάρτησης Συμπίεσης (NCD), η οποία χρησιμοποιεί συμπίεση για την εξαγωγή χαρακτηριστικών από τα δεδομένα. Με αυτόν τον τρόπο, αποφεύγεται η παραδοσιακή επεξεργασία όπως ανάλυση γραμματικής ή λεξιλογίου, και τα δεδομένα αντιμετωπίζονται ως ακατέργαστες χορδές κειμένου.

Οι μέθοδοι AI που χρησιμοποιήθηκαν είναι οι εξής:

1. **Κανονικοποιημένη Εξάρτηση Συμπίεσης (NCD)**:

Πρόκειται για τη βασική μέθοδο εξαγωγής χαρακτηριστικών. Το NCD χρησιμοποιεί την απόσταση μεταξύ των δεδομένων που υπολογίζεται μέσω συμπίεσης, ώστε να προσδιοριστεί η "ομοιότητα" ανάμεσα σε διαφορετικές χορδές. Αυτή η μέθοδος θεωρείται ανεξάρτητη από τον τύπο και τη γλώσσα του κειμένου, κάτι που δίνει σημαντική ευελιξία στη χρήση της.

2. **Support Vector Machines (SVM)**:

Αποτελεί έναν κλασικό αλγόριθμο μηχανικής μάθησης που χρησιμοποιήθηκε για την ταξινόμηση δεδομένων. Το SVM λειτουργεί σε έναν χώρο χαρακτηριστικών που δημιουργείται από τα

αποτελέσματα του NCD. Για την ανάλυση μεγάλων συνόλων δεδομένων, χρησιμοποιήθηκε γραμμικός πυρήνας.

3. **Γεννήτριες Χαρακτηριστικών (Attribute Generators)**:
Δημιουργήθηκαν σύνολα «γεννητριών χαρακτηριστικών» (I και G sets) για να υπολογιστούν οι αποστάσεις NCD των νέων δεδομένων από τα χαρακτηριστικά του εκπαιδευτικού συνόλου.

Το βασικό σκεπτικό πίσω από την ανάπτυξη του LAMaLearner [157] είναι η ανάγκη διαχείρισης της μεγάλης ποικιλίας μορφών και δομών στα αρχεία καταγραφής (logs), που καθιστούν τη χειροκίνητη ανάλυση χρονοβόρα και αναποτελεσματική. Οι επιστήμονες αναγνώρισαν ότι τα παραδοσιακά εργαλεία, τα οποία απαιτούν προκαθορισμένους κανόνες ή εκφράσεις κανονικότητας (regular expressions), είναι ανεπαρκή όταν τα δεδομένα έχουν μεγάλη ποικιλομορφία ή δεν είναι γνωστά εκ των προτέρων. Με το LAMaLearner, στοχεύουν σε ένα αυτόνομο εργαλείο που θα μειώσει την ανθρώπινη παρέμβαση και θα επιτρέψει την εύκολη ανακάλυψη προτύπων στα logs.

Οι επιστήμονες ενσωμάτωσαν προηγμένες τεχνικές τεχνητής νοημοσύνης (AI) για την ανάπτυξη του LAMaLearner, όπως:

- **Clustering (Ομαδοποίηση)**: Χρησιμοποίησαν αλγορίθμους ομαδοποίησης για τη δημιουργία nested clusters (ιεραρχικών ομαδοποιήσεων) από γεγονότα logs. Αυτό επιτρέπει την αυτόματη κατηγοριοποίηση γεγονότων σε υψηλού επιπέδου ομάδες.
- **Natural Language Processing (NLP)**: Εφαρμόστηκε επεξεργασία φυσικής γλώσσας για την αναγνώριση μεταβλητών και οντοτήτων μέσα στα logs (π.χ. usernames, IPs). Η τεχνική αυτή επιτρέπει την κατανόηση και ανάλυση περιεχομένων logs ανεξάρτητα από τη μορφή ή τη γλώσσα τους.
- **Μοντέλα Βασισμένα σε Επισήμανση**: Δημιουργία μοντέλων υψηλής ακρίβειας (F1=1) για την αυτόματη επισήμανση γεγονότων με ετικέτες (π.χ. «success», «failure»).
- **Εξαγωγή Συσχετίσεων**: Οι αλγόριθμοι ομαδοποίησης αναγνώρισαν συσχετίσεις μεταξύ μεταβλητών, όπως η σύνδεση αναγνωριστικών ασφαλείας (Microsoft Security Identifiers) με ανθρώπινα ονόματα χρηστών.
- **Αυτονομία και Ανθεκτικότητα**: Το εργαλείο σχεδιάστηκε ώστε να προσαρμόζεται αυτόματα σε νέες μορφές logs χωρίς ανάγκη αναπροσαρμογής των κανόνων.
- **Visualization και Dashboarding**: Παρουσίαση των αποτελεσμάτων σε πίνακες ελέγχου για γρήγορη κατανόηση των γεγονότων, ιδανικό για χρήση σε βιομηχανικά συστήματα και αλλαγές βαρδιών.

Οι Sisiaridis και Markowitch [158] θέλουν να αντιμετωπίσουν την αυξημένη πολυπλοκότητα των δεδομένων που προέρχονται από ετερογενείς πηγές σε αναλύσεις ασφάλειας. Στόχος τους είναι να αυτοματοποιήσουν και να απλοποιήσουν τη διαδικασία εξαγωγής και επιλογής χαρακτηριστικών, ελαχιστοποιώντας τα προβλήματα διαλειτουργικότητας μεταξύ διαφορετικών δομών δεδομένων. Έμφαση δίνεται στη μείωση της πολυπλοκότητας των δεδομένων, διατηρώντας μόνο τις πιο σημαντικές πληροφορίες για τις αναλύσεις.

Γι’ αυτό τον λόγο εφάρμοσαν τις παρακάτω μεθόδους τεχνητής νοημοσύνης:

1. **Random Forest Classifier**:
Ο Random Forest Clssifier χρησιμοποιήθηκε για τον υπολογισμό της σημαντικότητας των χαρακτηριστικών (feature importance) και για τη μείωση διαστάσεων.

2. **Δέντρα Απόφασης (Decision Trees)**:
Τα δέντρα απόφασης εφαρμόστηκαν για την επιλογή χαρακτηριστικών σε περιπτώσεις όπου τα δεδομένα είχαν μικρό αριθμό κατηγοριών.

3. **Συνδυαστικά Μοντέλα (Ensemble Techniques)**:
Τα συνδυαστικά μοντέλα χρησιμοποιήθηκαν συνδυαστικά μοντέλα, όπως Νευρωνικά Δίκτυα και Bayes classifiers, για τη βελτιστοποίηση των αποτελεσμάτων.

4. **Pearson Correlation**:
Το Pearson Correlation εφαρμόστηκε για τη μέτρηση συσχέτισης μεταξύ χαρακτηριστικών, ώστε να αφαιρεθούν εκείνα που είχαν υψηλή συσχέτιση.

5. **Χρήση Στατιστικών Τεχνικών**:
Η χρήση στατιστικών τεχνικών περιλαμβάνει μεθόδους όπως το τεστ Chi-Square και άλλες στατιστικές μεθόδους που είναι διαθέσιμες στη βιβλιοθήκη Spark MLlib.

6. **Μοντελοποίηση μέσω Κατηγοριών (Category Theory)**:
Στην μοντελοποίηση μέσω κατηγοριών εξετάζεται η χρήση της θεωρίας κατηγοριών για την τυποποίηση και την περαιτέρω βελτιστοποίηση της ανάλυσης.

3.6 Σύστημα Πρόληψης Εισβολής

Το σύστημα πρόληψης εισβολών (Intrusion Prevention System - IPS) [159] *παρακολουθεί την κίνηση στο δίκτυο και αναλαμβάνει δράση για να αποτρέψει επιθέσεις, όπως η αναφορά, ο αποκλεισμός, η απόρριψη ή η επαναφορά της σύνδεσης. Οι ερευνητές έχουν προτείνει τη χρήση τεχνικών όπως το unsupervised isolation forest και τα self-organizing incremental neural networks, καθώς και συστήματα πρόληψης εισβολών βασισμένα σε SVM, για την προστασία ενσωματωμένων συστημάτων σε ηλεκτρονικές εφαρμογές αυτοκινήτων και δίκτυα IoT.*

Οι επιστήμονες στοχεύουν να αντιμετωπίσουν τις ευπάθειες των Δικτύων Ελεγκτών Περιοχής (Controler Area Network - CAN) στα οχήματα [160], τα οποία είναι βασικά για τη λειτουργία τους αλλά δεν διαθέτουν ενσωματωμένους μηχανισμούς ασφάλειας. Καθώς οι παραδοσιακές μέθοδοι ανίχνευσης κυβερνοεπιθέσεων είναι είτε ακριβές είτε εξαρτώνται από ετικετοποιημένα δεδομένα (frames' data), οι ερευνητές επιδιώκουν να δημιουργήσουν μια λύση που να είναι οικονομική, γρήγορη και ανεξάρτητη από συγκεκριμένα μοτίβα επιθέσεων. Το σκεπτικό τους βασίζεται στην ανάπτυξη ενός Συστήματος Πρόληψης Εισβολών (IPS) που λειτουργεί σε πραγματικό χρόνο και είναι εύκολα υιοθετήσιμο από τη βιομηχανία.

Οι στόχοι τους περιλαμβάνουν την ανίχνευση κακόβουλων μηνυμάτων πριν ολοκληρωθεί η μετάδοσή τους, τη δημιουργία ενός συστήματος που να μπορεί να λειτουργεί με φθηνό υλικό (όπως Raspberry Pi) και την ανίχνευση άγνωστων επιθέσεων χωρίς την ανάγκη ετικετοποιημένων δεδομένων. Παράλληλα, επιδιώκουν να προσφέρουν μια λύση προσαρμόσιμη σε διαφορετικά μοντέλα οχημάτων και κατασκευαστές, με έμφαση στη χαμηλή πολυπλοκότητα και το χαμηλό κόστος.

Οι μέθοδοι τεχνητής νοημοσύνης που χρησιμοποιήθηκαν, αναγράφονται παρακάτω.

1. **Isolation Forest (iForest):** Ένας μηχανισμός μη επιβλεπόμενης εκπαίδευσης που ανιχνεύει ανωμαλίες βασιζόμενος σε δεδομένα χωρίς επιθέσεις (normal data). Εστιάζει στις ιδιαιτερότητες της κακόβουλης δραστηριότητας χωρίς να απαιτεί ετικετοποιημένα δείγματα. Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί μόνο τα πρώτα 5 bytes από τα δεδομένα των μηνυμάτων CAN για να μειώσει τον χρόνο ανίχνευσης. Απέδειξε ότι μπορεί να ανταποκριθεί στις απαιτήσεις ταχύτητας για την πρόληψη ζημιών.

2. **Σύγκριση με άλλες μεθόδους:** Η λύση συγκρίθηκε με προηγμένα συστήματα ανίχνευσης επιθέσεων που χρησιμοποιούν επιβλεπόμενη μάθηση (όπως DCNN και LSTM-based IDS) και με συνδυαστικές προσεγγίσεις (όπως το MTH-IDS). Ενώ οι επιβλεπόμενες μέθοδοι είχαν υψηλότερη ακρίβεια σε επιθέσεις με γνωστά μοτίβα, το προτεινόμενο IPS ξεχώρισε λόγω της δυνατότητάς του να ανιχνεύει άγνωστες επιθέσεις και να λειτουργεί σε πραγματικό χρόνο.

Οι επιστήμονες πέτυχαν τους στόχους τους, αποδεικνύοντας ότι ένα οικονομικό και γρήγορο σύστημα πρόληψης εισβολών μπορεί να προσφέρει ουσιαστική προστασία από κυβερνοεπιθέσεις, διατηρώντας ισορροπία μεταξύ απόδοσης και κόστους.

Οι επιστήμονες που παρουσιάζουν το πλαίσιο n-SOINN και WTA-SVM [161] έχουν ως κύριο στόχο τη δημιουργία ενός συστήματος ανίχνευσης επιθέσεων σε δίκτυα (IDS) που να είναι αποδοτικό, προσαρμόσιμο και κατάλληλο για εφαρμογές σε δυναμικά περιβάλλοντα, όπως το Internet of Things (IoT). Δεδομένης της αυξανόμενης πολυπλοκότητας και του μεγάλου όγκου δεδομένων σε αυτά τα περιβάλλοντα, το πλαίσιο τους αποσκοπεί στη διασφάλιση της ασφάλειας με τη χρήση τεχνικών τεχνητής νοημοσύνης (AI) που υποστηρίζουν incremental learning. (την επικαιροποιούμενη μάθηση).

Οι ερευνητές θεωρούν ότι τα δυναμικά δίκτυα, όπως αυτά του IoT, απαιτούν συστήματα ανίχνευσης επιθέσεων που μπορούν να προσαρμόζονται συνεχώς, να αξιοποιούν αποδοτικά τους περιορισμένους πόρους των IoT συσκευών και να λειτουργούν σε πραγματικό χρόνο. Αντί για στατική εκπαίδευση, το σύστημα πρέπει να μαθαίνει από νέα δεδομένα καθώς αυτά εμφανίζονται, ανιχνεύοντας επιθέσεις γρήγορα και βελτιώνοντας τις δυνατότητές του σταδιακά μέσα από τα λάθη του.

Ο στόχος του πλαισίου είναι να παρέχει αυξημένη ακρίβεια ανίχνευσης, κατηγοριοποιώντας επιθέσεις σε διαφορετικές κατηγορίες με υψηλή επιτυχία, ενώ ενσωματώνει αποτυχημένες προβλέψεις για συνεχή βελτίωση. Με τη χρήση μικρού αριθμού δεδομένων, επιδιώκει αποτελεσματικότητα χωρίς συμβιβασμούς στην απόδοση, ενώ ξεπερνά τις αδυναμίες των στατικών μεθόδων που δεν προσαρμόζονται σε νέα δεδομένα ή επιθέσεις.

Για την επίτευξη των στόχων τους, οι επιστήμονες χρησιμοποίησαν τις παρακάτω τεχνικές AI:

1. **n-SOINN (Self-Organizing Incremental Neural Network)** [162]:

    - Χρησιμοποιείται για την ομαδοποίηση (clustering) δεδομένων με δυνατότητα επικαιροποιούμενης μάθησης.
    - Δύο n-SOINN εφαρμόζονται για κάθε κατηγορία επιθέσεων: ένα με χαμηλή τιμή παραμέτρου n (ακριβέστερο) και ένα με υψηλή (για λιγότερους κόμβους).
    - Οι κόμβοι των n-SOINN συγκρατούν μόνο τις πιο σημαντικές πληροφορίες για κάθε κατηγορία, μειώνοντας το υπολογιστικό κόστος.

2. **SVM (Support Vector Machine)**:

- Δυαδικοί ταξινομητές SVM χρησιμοποιούνται για την προκαταρκτική κατηγοριοποίηση, με τη μέθοδο one-versus-all και στρατηγική winner-takes-all.
- Για την τελική κατηγοριοποίηση, ένας πολυκλασικός SVM λαμβάνει ως είσοδο τα αποτελέσματα των καλύτερων υποψήφιων κατηγοριών.

3. **Συνδυασμός n-SOINN και WTA-SVM**:
   - Ο συνδυασμός των δύο τεχνικών επιτρέπει τη συμπίεση πληροφοριών μέσω n-SOINN και την ακριβή κατηγοριοποίηση μέσω SVM.
   - Η συνεχής ανατροφοδότηση από τις αποτυχημένες προβλέψεις ενισχύει τη δυνατότητα επικαιροποιούμενης μάθησης του συστήματος.

Η μέθοδος που προτείνουν οι επιστήμονες είναι μια πρωτοποριακή προσέγγιση για ανίχνευση επιθέσεων σε δυναμικά περιβάλλοντα. Η συνδυαστική χρήση των n-SOINN και WTA-SVM προσφέρει ταχύτητα, προσαρμοστικότητα και αποδοτικότητα, καθιστώντας το πλαίσιο μια πολλά υποσχόμενη λύση για την ασφάλεια σε IoT συσκευές και άλλες παρόμοιες εφαρμογές.

3.7 Σύνοψη εργαλείων/μεθόδων AI

Ο παρακάτω πίνακας συνοψίζει τα εργαλεία, τις μεθόδους και τους αλγορίθμους τεχνητής νοημοσύνης (AI) που υποστηρίζουν τη λειτουργία «Προστασία» (Protect) του NIST Cybersecurity Framework, η οποία εστιάζει στην εφαρμογή κατάλληλων μέτρων ασφαλείας για τη διατήρηση της ακεραιότητας, της εμπιστευτικότητας και της διαθεσιμότητας κρίσιμων συστημάτων. Περιλαμβάνει τεχνικές για έλεγχο πρόσβασης (π.χ. SVM, Random Forest, AdaBoost), πρόληψη διαρροής δεδομένων (Autoencoders, Isolation Forest, PCA, BERT), διαχείριση ευπαθειών (Threat Modeling, LSTM, CRS) και ανάλυση καταγραφών & πρόληψη εισβολών (LogitBoost, iForest, n-SOINN). Μέσω αυτών των εργαλείων, οι οργανισμοί μπορούν να ανιχνεύουν και να αποτρέπουν επιθέσεις, μειώνοντας τον κίνδυνο κυβερνοαπειλών και ενισχύοντας την ασφάλεια των κρίσιμων υποδομών τους.

2.2.1	Ταυτοποίηση Συσκευών με Υποστήριξη Τεχνητής Νοημοσύνης	(1) MLP (Multi-Layer Perceptron) (2) Εφαρμογή μέτρων όπως precision, recall, F1-score, accuracy και ROC καμπύλες (3) SVM (Support Vector Machines) (4) Αλγόριθμοι όπως K-Nearest Neighbors (KNN) και Decision Trees (5) Φίλτρα όπως Infinite impulse response (IIR) και Finite impulse response (FIR) (6) Μετασχηματισμός CWT (Continuous Wavelet Transform) (7) Χρήση καμπυλών ROC (Receiver Operating Characteristic) (8) Χρήση ποσοστών EER (Equal Error Rate) (9) SAMM (Self-Adaptive Mathematical Morphology) (10) Time-Frequency (TF) Mapping (11) (*RFC)* Random Forest Classification (12) Adaptive Boosting (AdaBoost)
2.2.2	Αυτοματοποιημένος Έλεγχος Πρόσβασης	(1) Max-SAT (Maximum Satisfiability) (2) PDDL (Planning Domain Definition Language) (3) Χρήση planners όπως το Fast-Downward (4) Μη Αρνητική Παραγοντοποίηση Πινάκων (Non-Negative Matrix Factorization, NMF) (5) Attribute-Based Access Control (ABAC) (6) Model Driven Security (MDS)

2.2.3	Πρόληψη Διαρροής Δεδομένων	(1) Autoencoders (AE) (2) Isolation Forest (IF) (3) Local Outlier Factor (LOF) (4) Lightweight Online Detector of Anomalies (LODA) (5) Parzen Window (6) Gaussian (Gauss) (7) K-Means Clustering (KMC) (8) Ανάλυση Κυρίων Συνιστωσών (PCA) (9) Τεχνικές επεξεργασίας δεδομένων όπως είναι Label Encoding, One-hot Encoding, SMOTE (Synthetic Minority Oversampling Technique) (10) Logistic Regression (LR) (11) Decision Tree (DT) (12) Random Forest (RF) (13) Naive Bayes (NB) (14) k-Nearest Neighbors (KNN) (15) Kernel Support Vector Machine (KSVM) (16) Μέθοδοι αξιολόγησης απόδοσης όπως είναι Precision, Recall, F-measure, AUC-ROC Curve (17) Clustering (Ομαδοποίηση) με k-Means (18) TF-IDF Representation (19) Συνδυασμός Global και Local Classifiers (20) Latent Semantic Analysis (LSA) (21) BERT (Bidirectional Encoder Representations from Transformers) (22) BiLSTM (Bidirectional Long Short-Term Memory) (23) Attention Mechanism (24) Εξηγητική Τεχνητή Νοημοσύνη (XAI) (25) LSTM (Long Short-Term Memory) (26) Ανάλυση Πληροφοριών Απειλών (CTI) (27) Greedy Mechanism και Hotbooting PHC (28) DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (29) χρήση μετρικών όπως είναι Ομοιογένεια, Πληρότητα, V-measure, Adjusted Rand Index (ARI) και Adjusted Mutual Info (AMI)

2.2.4	Σχέδιο Διαχείρισης Ευπαθειών με Ενίσχυση Τεχνητής Νοημοσύνης	(1) Νευρο-Συμβολικό Μοντέλο (NN-PLP) (2) Μοντελοποίηση Απειλών (Threat Modeling) (3) Συνδυασμός Χαρακτηριστικών LSA και CVSS (4) BERT (Bidirectional Encoder Representations from Transformers) (5) [CLS] token embedding (6) Mean pooling (7) Max pooling (8) LSTM (Long Short-Term Memory) (9) Class Rectification Strategy (CRS) (10) Balanced Window Strategy (BWS) (11) Real - time Dynamic CFoncept Adaptive Learning Algorithm (RDCAL)
2.2.5	Ανάλυση Αρχείων Καταγραφής	(1) LogitBoost (LB) (2) Random Forest (RF) (3) Light Gradient Boosting Machine (LGBM) (4) Gaussian Naive Bayes (GNB) (5) Decision Tree (DT) (6) Feedforward Neural Network (FNN) (7) Logistic Regression (LR) (8) Majority Voting (MV) (9) Weighted Voting (WV) (10) Conservative Approach (CA) (11) Natural Language Processing (NLP) (12) Κανονικοποιημένη Εξάρτηση Συμπίεσης (NCD) (13) Support Vector Machines (SVM) (14) Γεννήτριες Χαρακτηριστικών (Attribute Generators) (15) Clustering (Ομαδοποίηση) (16) Random Forest Classifier (17) Δέντρα Απόφασης (Decision Trees) (18) Νευρωνικά Δίκτυα και Bayes classifiers (19) Pearson Correlation (20) Χρήση Στατιστικών Τεχνικών όπως το τεστ Chi-Square
2.2.6	Σύστημα Πρόληψης Εισβολής	(1) Isolation Forest (iForest) (2) n-SOINN (Self-Organizing Incremental Neural Network) (3) WTA-SVM (4) SVM (Support Vector Machine)

Πίνακας 3 : Εργαλεία και αλγόριθμοι τεχνητής νοημοσύνης που υποστηρίζουν τη λειτουργία «Προστασία» (Protect) του NIST Cybersecurity Framework, συμβάλλοντας στην πρόληψη κυβερνοεπιθέσεων και την ενίσχυση της ασφάλειας κρίσιμων υποδομών.

4. Εντοπισμός

*Η λειτουργία **«Εντοπισμός»** (**Detect)** του πλαισίου NIST αφορά την ανάπτυξη και εφαρμογή κατάλληλων δραστηριοτήτων για την έγκαιρη αναγνώριση περιστατικών κυβερνοασφάλειας. Στόχος της είναι η γρήγορη ανακάλυψη περιστατικών, ώστε να περιοριστούν άμεσα οι συνέπειες. Η λειτουργία Detect είναι κρίσιμη για ένα ισχυρό πρόγραμμα κυβερνοασφάλειας, καθώς η ταχεία ανίχνευση κυβερνοεπιθέσεων επιτρέπει την άμεση αντίδραση και τον περιορισμό των επιπτώσεων. Η καθυστέρηση στον εντοπισμό παραβιάσεων μπορεί να έχει σοβαρές συνέπειες για μια επιχείρηση. Ακολουθώντας τα πρότυπα και τις βέλτιστες πρακτικές της λειτουργίας αυτής, οι οργανισμοί μπορούν να αναπτύξουν πιο αποτελεσματικά προγράμματα ασφαλείας, να μειώσουν τους κινδύνους και να προστατεύσουν καλύτερα τα κρίσιμα δεδομένα και υποδομές τους.*

4.1 Σύστημα Ανίχνευσης Εισβολής

Ένα σύστημα ανίχνευσης εισβολών (Intrusion Detection System - IDS) αποτελείται από διάφορα εργαλεία και μεθόδους που χρησιμοποιούνται για την παρακολούθηση ενός συστήματος και της κυκλοφορίας των δεδομένων στο δίκτυο. Ο σκοπός του είναι να εντοπιστούν ανωμαλίες ή ύποπτες ενέργειες που μπορεί να υποδηλώνουν μια προσπάθεια εισβολής στο σύστημα. Στην συγκεκριμένη έρευνα, το IDS υλοποιήθηκε από τρεις προοπτικές, την δυαδική ταξινόμηση, πολυκατηγορική ταξινόμηση και ο συνδυασμός τους.

Η δυαδική ταξινόμηση αποτελεί τη βασική μορφή ταξινόμησης και είναι η πιο διεξοδικά μελετημένη στο πλαίσιο της ανίχνευσης εισβολών. Οι περισσότεροι ερευνητές επικεντρώνονται στη χρήση διάφορων αλγορίθμων μηχανικής μάθησης για την αναγνώριση κακόβουλων δραστηριοτήτων. Ωστόσο, λιγότεροι έχουν ασχοληθεί με προκλήσεις όπως η βελτιστοποίηση των υπερπαραμέτρων των αλγορίθμων, η αντιμετώπιση της ανισορροπίας μεταξύ των κατηγοριών στα δεδομένα και η επιλογή ή εξαγωγή των πιο σημαντικών χαρακτηριστικών από τα σύνολα δεδομένων.

Οι επιστήμονες που διεξήγαγαν την έρευνα [164] είχαν ως στόχο να διερευνήσουν τη δυνατότητα χρήσης Αυτο-Οργανωμένων Χαρτών (**Self-Organizing Maps, SOMs**) ως ένα σύστημα μη επιβλεπόμενης μάθησης για την ανάλυση κακόβουλης και άγνωστης δικτυακής κίνησης. Η συγκεκριμένη προσέγγιση κίνησε το ενδιαφέρον τους λόγω της αυξανόμενης ανάγκης για αυτοματοποιημένα συστήματα ανίχνευσης δικτυακών απειλών, όπως τα botnets, που διαρκώς εξελίσσονται και καθιστούν τις παραδοσιακές τεχνικές ανίχνευσης λιγότερο αποτελεσματικές.

Η έρευνα βασίστηκε στη θεμελιώδη αρχή των SOMs, που είναι η απεικόνιση πολυδιάστατων δεδομένων σε χαμηλότερες διαστάσεις, με τρόπο που τα δεδομένα με παρόμοια χαρακτηριστικά να βρίσκονται κοντά μεταξύ τους στον χάρτη. Αυτό τους καθιστά κατάλληλους για την ανίχνευση μοτίβων και ανωμαλιών σε πολύπλοκα σύνολα δεδομένων.

Για να επιτύχουν τον στόχο τους, οι επιστήμονες χρησιμοποίησαν τρία διαφορετικά σχήματα εκπαίδευσης SOMs:

1. Σχήμα Εκπαίδευσης (i): Χρησιμοποίησαν τόσο φυσιολογική όσο και κακόβουλη κίνηση.

2. Σχήμα Εκπαίδευσης (ii): Χρησιμοποίησαν μόνο φυσιολογική κίνηση.

3. Σχήμα Εκπαίδευσης (iii): Χρησιμοποίησαν μόνο κακόβουλη κίνηση.

Τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από τα γνωστά σύνολα δεδομένων CTU13 και ISOT, τα οποία περιλαμβάνουν κατηγορίες όπως Normal (φυσιολογική), Botnet (κακόβουλη), και Background (άγνωστη). Για την προετοιμασία των δεδομένων, εφαρμόστηκε κανονικοποίηση (zero mean, unit variance), ενώ για την εκπαίδευση των SOMs χρησιμοποιήθηκαν εξειδικευμένες παράμετροι, όπως εξάγωνη διάταξη (hexagonal lattice) και Gaussian neighborhood function.

Οι SOMs εκπαιδεύτηκαν με μια διαδικασία δύο φάσεων:

- Αρχική εκπαίδευση (coarse training): Μεγάλες ακτίνες γειτονιάς για γενική ομαδοποίηση.
- Λεπτομερής εκπαίδευση (fine tuning): Μικρότερες ακτίνες για πιο ακριβή διαχωρισμό των κλάσεων.

Οι επιστήμονες αξιολόγησαν την απόδοση του συστήματος χρησιμοποιώντας το True Positive Rate (TPR), το οποίο αποτυπώνει την ακρίβεια ανίχνευσης για κάθε κατηγορία.

Τα αποτελέσματα έδειξαν ότι το σχήμα εκπαίδευσης (i), που χρησιμοποιεί τόσο φυσιολογική όσο και κακόβουλη κίνηση, πέτυχε την καλύτερη απόδοση, με υψηλά ποσοστά TPR και σαφή διαχωρισμό των κατηγοριών στον SOM.

Οι ερευνητές Saveetha και Maragatham στο άρθρο τους [165] παρουσιάζουν ένα μοντέλο ανίχνευσης εισβολών (Intrusion Detection System - IDS) που βασίζεται στη συνδυαστική χρήση της τεχνολογίας blockchain και των μεθόδων βαθιάς μάθησης (deep learning). Σκοπός της έρευνάς τους είναι να αναπτύξουν ένα αποτελεσματικό σύστημα προστασίας από κυβερνοεπιθέσεις, αξιοποιώντας την αξιοπιστία και τη διαφάνεια του blockchain σε συνδυασμό με την ακρίβεια και την προσαρμοστικότητα των μεθόδων βαθιάς μάθησης.

Οι επιστήμονες τονίζουν ότι οι κυβερνοεπιθέσεις εξελίσσονται διαρκώς, καθιστώντας αναγκαία τη χρήση συστημάτων ανίχνευσης εισβολών που μπορούν να ανιχνεύσουν όχι μόνο γνωστές, αλλά και νέες μορφές επιθέσεων. Το blockchain, με την αμετάβλητη και αποκεντρωμένη φύση του, προσφέρει μια υποσχόμενη βάση για την προστασία δεδομένων και την εξασφάλιση της ακεραιότητας των διαδικασιών. Από την άλλη, η βαθιά μάθηση έχει αποδειχθεί εξαιρετικά αποτελεσματική στην ανάλυση μεγάλων όγκων δεδομένων και την ανίχνευση μοτίβων.

Οι ερευνητές χρησιμοποίησαν δύο βασικές μεθόδους τεχνητής νοημοσύνης:

1. **Μακροχρόνια Μνήμη Βραχείας Διάρκειας (Long Short-Term Memory - LSTM)**:
    - Τα LSTM είναι ένας τύπος νευρωνικού δικτύου που μπορεί να απομνημονεύει μακροχρόνιες εξαρτήσεις. Στην παρούσα έρευνα, χρησιμοποιήθηκαν για την ανάλυση χρονικών δεδομένων από blockchain, προκειμένου να ανιχνεύονται ανωμαλίες στις συναλλαγές.
    - Οι πύλες εισόδου και εξόδου του μοντέλου επιτρέπουν την αποθήκευση σημαντικών πληροφοριών και την απόρριψη άχρηστων δεδομένων.
2. **Συνδυασμός Συγκλιτικών και Επαναλαμβανόμενων Νευρωνικών Δικτύων (RNN–CNN)**:

- Το CNN χρησιμοποιήθηκε για την εξαγωγή χαρακτηριστικών από τα δεδομένα δικτύου, ενώ το RNN προστέθηκε για την ανάλυση χρονικών εξαρτήσεων. Αυτή η προσέγγιση εξασφάλισε ακρίβεια στην κατηγοριοποίηση των επιθέσεων.

Τα πειράματα έδειξαν ότι το προτεινόμενο μοντέλο ήταν σε θέση να ανιχνεύσει αποτελεσματικά ανωμαλίες και να εντοπίσει κακόβουλες δραστηριότητες. Η ακρίβεια ανίχνευσης ήταν υψηλότερη σε σύγκριση με τα παραδοσιακά συστήματα, ενώ τα δεδομένα του blockchain αξιοποιήθηκαν για την έγκαιρη πρόβλεψη επιθέσεων.

Η ταχεία ανάπτυξη των έξυπνων δικτύων (Smart Grids) έχει επιφέρει σημαντικά οφέλη στον τομέα της ενέργειας, όπως η αυξημένη απόδοση και η καλύτερη διαχείριση των πόρων. Ωστόσο, αυτή η πρόοδος συνοδεύεται από αυξημένες απειλές στον τομέα της κυβερνοασφάλειας, λόγω της μεγάλης εξάρτησης από επικοινωνιακές τεχνολογίες. Οι επιστήμονες [166] που ανέπτυξαν την αρχιτεκτονική **SGDIDS (Smart Grid Distributed Intrusion Detection System)** επιδίωξαν να αντιμετωπίσουν αυτές τις απειλές, αναπτύσσοντας ένα κατανεμημένο σύστημα ανίχνευσης εισβολών, σχεδιασμένο να προστατεύει το έξυπνο δίκτυο μέσω μιας ιεραρχικής προσέγγισης.

Ο κύριος στόχος της έρευνας ήταν η ενίσχυση της κυβερνοασφάλειας των έξυπνων δικτύων με τη χρήση προηγμένων μεθόδων τεχνητής νοημοσύνης (AI).

Οι επιστήμονες ανέπτυξαν την αρχιτεκτονική SGDIDS, η οποία βασίζεται σε τρία επίπεδα ανίχνευσης εισβολών: **HAN IDS** (Home Area Network), **NAN IDS** (Neighborhood Area Network), και **WAN IDS** (Wide Area Network). Το κάθε επίπεδο εξυπηρετεί διαφορετικές ανάγκες ανίχνευσης και επικεντρώνεται σε συγκεκριμένα είδη επιθέσεων.

1. **Διανομή Μονάδων IDS**:

   - Το επίπεδο HAN IDS είναι υπεύθυνο για την ανίχνευση κοινών επιθέσεων όπως DoS (Denial of Service) και probing, καθώς αυτές είναι πιο συχνές στο τοπικό δίκτυο.
   - Το NAN IDS χρησιμοποιείται για πιο εξειδικευμένες επιθέσεις, όπως οι U2R (User-to-Root).
   - Το WAN IDS, που βρίσκεται στο υψηλότερο επίπεδο, είναι υπεύθυνο για την ανίχνευση πιο περίπλοκων επιθέσεων όπως οι R2L (Remote-to-Local), διαθέτοντας την καλύτερη απόδοση ταξινόμησης λόγω της υψηλότερης υπολογιστικής του ικανότητας.

2. **Μηχανισμός Ανίχνευσης**:

   Όταν το HAN IDS δεν μπορεί να ταξινομήσει τη δικτυακή κίνηση, τα δεδομένα αποστέλλονται στο επόμενο επίπεδο, δηλαδή στο NAN IDS. Αν το NAN IDS αποτύχει, τα δεδομένα προωθούνται σε άλλο NAN IDS ή στο WAN IDS για περαιτέρω ανάλυση. Αυτή η ιεραρχική προσέγγιση διασφαλίζει ότι τα πιο σύνθετα δεδομένα αναλύονται από τους καλύτερους διαθέσιμους ταξινομητές, εξοικονομώντας παράλληλα πόρους στο δίκτυο.

3. **Αλγόριθμοι AI**:

   Για την ταξινόμηση των δεδομένων χρησιμοποιήθηκαν δύο μέθοδοι τεχνητής νοημοσύνης:

   - **SVM (Support Vector Machine)**: Ένας αλγόριθμος επιβλεπόμενης μάθησης που διαχωρίζει τα δεδομένα σε κατηγορίες. Οι SVM έδειξαν εξαιρετική απόδοση στην ανίχνευση επιθέσεων U2R και R2L, παρόλο που τα δεδομένα εκπαίδευσης για αυτές τις κατηγορίες ήταν περιορισμένα.
   - **AIS (Artificial Immune System)**: Βιολογικά εμπνευσμένοι αλγόριθμοι που προσομοιώνουν το ανοσοποιητικό σύστημα για την ανίχνευση ανωμαλιών. Αν και οι AIS ήταν λιγότερο

αποδοτικοί από τους SVM, προσέφεραν πολύτιμες δυνατότητες ανίχνευσης στα χαμηλότερα επίπεδα.

4. **Δεδομένα Εκπαίδευσης**:

   Οι ερευνητές χρησιμοποίησαν το σύνολο δεδομένων KDD99, που περιλαμβάνει χαρακτηριστικά όπως τύποι πακέτων και διάρκεια συνδέσεων, για την εκπαίδευση και δοκιμή των μοντέλων τους. Παρά τις προκλήσεις της περιορισμένης διαθεσιμότητας δεδομένων για σπάνιες επιθέσεις, οι SVM κατάφεραν να αποδώσουν καλύτερα λόγω της ικανότητάς τους να διαχειρίζονται ανισομερείς κατανομές δεδομένων.

Η παρούσα επιστημονική εργασία [167] επικεντρώνεται στην ανάπτυξη ενός ολοκληρωμένου μοντέλου ανίχνευσης εισβολών, ασφαλούς μετάδοσης δεδομένων και ταξινόμησης εικόνων, με εφαρμογή στα Κυβερνοφυσικά Συστήματα (Cyber-Physical Systems - CPS) στον τομέα της υγείας. Το σκεπτικό των ερευνητών βασίζεται στην ανάγκη για ενίσχυση της ασφάλειας και της αποδοτικότητας στη διαχείριση ευαίσθητων δεδομένων, καθώς τα CPS χρησιμοποιούνται ευρέως στη σύγχρονη ιατρική για κρίσιμες εφαρμογές, όπως η ανίχνευση ασθενειών.

Οι στόχοι της έρευνας είναι:

1. Η ασφαλής αποθήκευση και μετάδοση δεδομένων μέσω καινοτόμων μεθόδων κρυπτογράφησης και blockchain.
2. Η ανίχνευση κυβερνοαπειλών με υψηλή ακρίβεια μέσω προηγμένων αλγορίθμων μηχανικής μάθησης.
3. Η βελτίωση της ακρίβειας ταξινόμησης ιατρικών δεδομένων και εικόνων για καλύτερη διάγνωση ασθενειών.

Για την επίτευξη αυτών των στόχων, οι επιστήμονες σχεδίασαν ένα πολυεπίπεδο μοντέλο που συνδυάζει σύγχρονες μεθόδους Τεχνητής Νοημοσύνης (AI) και τεχνολογίες ασφάλειας δεδομένων.

Το προτεινόμενο μοντέλο περιλαμβάνει τα εξής βασικά στάδια:

1. Ανίχνευση Εισβολών μέσω Deep Belief Network (DBN)

Οι ερευνητές χρησιμοποίησαν έναν αλγόριθμο βαθιών νευρωνικών δικτύων (Deep Belief Network) για την ανίχνευση κυβερνοαπειλών στα συστήματα CPS. Το DBN κατάφερε να αναγνωρίσει εισβολές με ακρίβεια 98.95% στο σύνολο δεδομένων NSL-KDD 2015 και 98.94% στο CIDDS-001. Αυτή η απόδοση καθιστά το DBN εξαιρετικό εργαλείο ανίχνευσης επιθέσεων σε πραγματικό χρόνο.

2. Κρυπτογράφηση μέσω της Μεθόδου Multiple Share Creation (MSC)

Για την προστασία των δεδομένων, η εργασία προτείνει τη χρήση της μεθόδου Multiple Share Creation. Αυτή η τεχνική χωρίζει την αρχική εικόνα σε τέσσερα μέρη (shares), τα οποία στη συνέχεια κρυπτογραφούνται με τη χρήση αλγορίθμων XOR. Το μεγάλο πλεονέκτημα της MSC είναι η δυνατότητα ασφαλούς ανασύνθεσης της αρχικής εικόνας χωρίς απώλειες, διασφαλίζοντας έτσι την ακεραιότητα των δεδομένων.

3. Ασφαλής Μετάδοση Δεδομένων με Blockchain

Η τεχνολογία blockchain χρησιμοποιείται για την ασφαλή αποθήκευση και μετάδοση δεδομένων, εξαλείφοντας τους κινδύνους μη εξουσιοδοτημένης πρόσβασης. Η ενσωμάτωση της αποκεντρωμένης βάσης δεδομένων IPFS (InterPlanetary File System) επέτρεψε την ασφαλή διαχείριση μεγάλου όγκου δεδομένων, καθιστώντας το σύστημα ανθεκτικό σε επιθέσεις.

4. Ταξινόμηση Ιατρικών Δεδομένων μέσω CNN-ResNet 101

Για την ανάλυση και ταξινόμηση ιατρικών δεδομένων, οι επιστήμονες χρησιμοποίησαν το ResNet 101, ένα βαθύ συνελικτικό νευρωνικό δίκτυο (Convolutional Neural Network - CNN). Το ResNet 101 πέτυχε υψηλά ποσοστά ακρίβειας (94.85%), ευαισθησίας (96.12%) και εξειδίκευσης (98.02%) στη διάγνωση εικόνων. Σε σύγκριση με άλλα μοντέλα, όπως τα VGG-19 και ResNet-50, το ResNet 101 αποδείχθηκε ανώτερο τόσο σε ταχύτητα όσο και σε ακρίβεια.

Η έρευνα απέδειξε ότι ο συνδυασμός μεθόδων AI και τεχνολογιών ασφάλειας δεδομένων μπορεί να οδηγήσει σε αξιόπιστα και ασφαλή συστήματα για την υγειονομική περίθαλψη. Το προτεινόμενο μοντέλο επιτυγχάνει υψηλή απόδοση σε κάθε στάδιο, ενώ τα πειραματικά αποτελέσματα επιβεβαιώνουν την υπεροχή του σε σχέση με υπάρχουσες μεθόδους.

Οι επιστήμονες που εκπόνησαν την έρευνα αυτή [168] επικεντρώθηκαν στην ανάπτυξη ενός αξιόπιστου και αποτελεσματικού Συστήματος Ανίχνευσης Εισβολών (IDS) για δίκτυα του Διαδικτύου των Πραγμάτων (IoT), λαμβάνοντας υπόψη τόσο την ακρίβεια όσο και την ταχύτητα ανίχνευσης. Η σημασία της ταχύτητας έγκειται στην ανάγκη περιορισμού των απωλειών μέσω της έγκαιρης ανίχνευσης και αντιμετώπισης απειλών, ενώ η ακρίβεια διασφαλίζει την αξιοπιστία του συστήματος. Στόχος των ερευνητών ήταν να δημιουργήσουν ένα σύστημα που ισορροπεί αποτελεσματικά την ταχύτητα και την ακρίβεια, χρησιμοποιώντας μεθόδους Τεχνητής Νοημοσύνης (AI) και ειδικά ενσωματωμένες μεθόδους μέτρησης αποδοτικότητας.

Οι επιστήμονες υπογράμμισαν ότι η ανεξάρτητη ανάλυση της ταχύτητας και της ακρίβειας μπορεί να οδηγήσει σε αντικρουόμενα αποτελέσματα σχετικά με την αποτελεσματικότητα ενός IDS. Για να αντιμετωπίσουν αυτή την πρόκληση, πρότειναν τη χρήση μετρικών που ενσωματώνουν και τα δύο στοιχεία, όπως οι Inverse Efficiency Score (IES), Rate-Correct Score (RCS), και Linear-Integrated Speed-Accuracy Score (LISAS). Ωστόσο, διαπίστωσαν ότι αυτές οι μετρικές δεν αποδίδουν ίσα βάρη στην ταχύτητα και την ακρίβεια, δημιουργώντας ευαισθησία στις μεταβολές του συστήματος. Γι' αυτό, προχώρησαν στην υιοθέτηση μιας νέας μετρικής, της Balanced Integration Score (BIS), η οποία δίνει ίσα βάρη στα δύο στοιχεία και μειώνει την εξάρτηση από τις αντισταθμίσεις ταχύτητας-ακρίβειας.

Οι σύγχρονες τεχνικές μηχανικής μάθησης που χρησιμοποίησαν, αναγράφονται παρακάτω.

1. **Μοντέλα μηχανικής μάθησης**: Χρησιμοποίησαν 23 διαφορετικά μοντέλα, όπως SVM, Δέντρα Απόφασης (DT), Logistic Regression, Multi-Layer Perceptron (MLP) και Gaussian-Naive Bayes, καθώς και 7 μοντέλα ενσωμάτωσης (ensemble models), όπως Random Forest, AdaBoost και Gradient Boosting.
2. **Μεθόδους ενσωμάτωσης (Ensemble Learning)**: Οι μέθοδοι αυτές συνδυάζουν την απόδοση διαφορετικών μοντέλων για τη δημιουργία ισχυρότερων μοντέλων. Κατασκεύασαν δύο κύρια συστήματα:
    - **Edge-ENClf**: Επικεντρώνεται σε τοπικά δεδομένα με μικρότερο επίπεδο γενίκευσης.

- **Cloud-ENClf**: Συνδυάζει πληροφορίες από όλα τα δεδομένα για μεγαλύτερη γενίκευση και σταθερότητα.

3. **Νέες μετρικές αποδοτικότητας**: Οι επιστήμονες ανέπτυξαν τρεις νέες μετρικές (F-Score Efficiency, ROC-AUC Efficiency, και Explained Variance Efficiency) για να αξιολογήσουν την αποδοτικότητα των μοντέλων με βάση την ταχύτητα και την ακρίβεια.
4. **Επεξεργασία δεδομένων**: Η έρευνα βασίστηκε σε τέσσερα διαφορετικά σύνολα δεδομένων (NSL-KDD, UNSW-NB15, BoTNetIoT, και BoTIoT) που καλύπτουν τόσο παραδοσιακά όσο και σύγχρονα σενάρια κυβερνοαπειλών. Χρησιμοποίησαν τη μέθοδο διασταυρούμενης επικύρωσης (5-fold cross-validation) για να διασφαλίσουν την αξιοπιστία των αποτελεσμάτων.

Το μοντέλο Cloud-ENClf υπερείχε λόγω της μεγαλύτερης γενίκευσης, ενώ το Edge-ENClf απέδειξε ότι μπορεί να λειτουργήσει εξίσου αποτελεσματικά σε πιο περιορισμένα περιβάλλοντα IoT.

Η έρευνα που παρουσιάστηκε [169] αποσκοπεί στην ανάπτυξη μιας νέας, καινοτόμας προσέγγισης για την ανίχνευση εισβολών σε δικτυακά συστήματα (Intrusion Detection Systems - IDS). Οι ερευνητές, συνδυάζοντας σύγχρονες τεχνικές τεχνητής νοημοσύνης (AI), επιδίωξαν να δημιουργήσουν ένα σύστημα που μπορεί να εντοπίζει τόσο γνωστές όσο και άγνωστες απειλές με ακρίβεια, αξιοπιστία και ταχύτητα.

Πιο συγκεκριμένα, οι επιστήμονες παρατήρησαν ότι τα παραδοσιακά συστήματα ανίχνευσης εισβολών βασίζονται σε ξεχωριστές διαδικασίες μάθησης χαρακτηριστικών και ταξινόμησης, κάτι που συχνά περιορίζει την απόδοση και την ικανότητα γενίκευσης των μοντέλων. Επιπλέον, πολλά υπάρχοντα συστήματα παρουσιάζουν υψηλά ποσοστά ψευδών συναγερμών (FAR), κάτι που επηρεάζει αρνητικά την αξιοπιστία τους σε πραγματικά περιβάλλοντα. Έτσι, στόχος τους ήταν να γεφυρώσουν το χάσμα μεταξύ της μάθησης χαρακτηριστικών και της ταξινόμησης, συνδυάζοντας και τις δύο διαδικασίες σε ένα ενιαίο, βελτιστοποιημένο πλαίσιο.

Για την επίτευξη αυτού του στόχου, σχεδίασαν και χρησιμοποίησαν μια μέθοδο βασισμένη σε βαθιά μάθηση και μη επιβλεπόμενη ταξινόμηση. Το σύστημα αποτελείται από δύο κύρια συστατικά:

1. **Μονοδιάστατος Αυτόματος Κωδικοποιητής (1D CAE)**: Ο CAE είναι ένα βαθύ νευρωνικό δίκτυο που έχει σχεδιαστεί για να μαθαίνει αντιπροσωπευτικά χαρακτηριστικά από τα δεδομένα εισόδου. Στόχος του είναι να κατασκευάσει μια αποδοτική και συμπαγή αναπαράσταση των δεδομένων δικτύου, επιτρέποντας την ανίχνευση εισβολών ακόμα και σε πολύπλοκα δεδομένα.
2. **Ταξινομητής Μιας Κατηγορίας (OCSVM)**: Ο ταξινομητής OCSVM, με χρήση kernel tricks, έχει τη δυνατότητα να δημιουργεί έναν υπερ-επίπεδο που διαχωρίζει τα φυσιολογικά δεδομένα από τα ανώμαλα (εισβολές).

Η καινοτομία της μεθόδου έγκειται στο πλαίσιο κοινής βελτιστοποίησης, που επιτρέπει τη σύγχρονη μάθηση χαρακτηριστικών από τον CAE και τη βελτίωση της ταξινόμησης από τον OCSVM. Η διαδικασία αυτή βελτιστοποιεί ταυτόχρονα την απώλεια ανακατασκευής (reconstruction loss) και την απώλεια ταξινόμησης (classification loss), προσφέροντας μια πιο αποδοτική και ακριβή λύση για το πρόβλημα ανίχνευσης εισβολών.

Οι επιστήμονες αξιολόγησαν τη μέθοδο σε δύο γνωστά σύνολα δεδομένων, NSL-KDD και UNSW-NB15. Τα αποτελέσματα επιβεβαίωσαν ότι:

- Η προτεινόμενη μέθοδος παρουσιάζει υψηλά ποσοστά ακρίβειας (ACC), ποσοστά ανίχνευσης (DR) και βαθμολογίας F1, ενώ διατηρεί χαμηλό ποσοστό ψευδών συναγερμών (FAR).

- Η μέθοδος αποδείχθηκε σταθερή σε διαφορετικά σύνολα δεδομένων, αποδίδοντας εξαιρετικά τόσο σε κλασικά όσο και σε σύγχρονα σενάρια επιθέσεων.
- Σε σύγκριση με άλλες προηγμένες μεθόδους, η προσέγγιση υπερτερεί σε ό,τι αφορά τη συνολική απόδοση και τη σταθερότητα, καθιστώντας την μια πολλά υποσχόμενη λύση για την ανίχνευση εισβολών.

Στη σύγχρονη εποχή, η ανίχνευση ανωμαλιών στη δικτυακή κίνηση είναι ένας κρίσιμος τομέας για την ασφάλεια των πληροφοριακών συστημάτων. Οι επιστήμονες που διεξήγαγαν αυτή την έρευνα επικεντρώθηκαν στην ανάπτυξη μιας υβριδικής προσέγγισης που συνδυάζει τη δύναμη διαφορετικών μεθόδων Τεχνητής Νοημοσύνης (AI) για την αποτελεσματική ανίχνευση επιθέσεων και κακόβουλων ενεργειών σε μεγάλα δίκτυα.

Η πλειονότητα των παραδοσιακών συστημάτων ανίχνευσης βασίζεται σε παλαιότερα σύνολα δεδομένων, όπως τα KDD99 και NSL-KDD, τα οποία περιέχουν ξεπερασμένα πρότυπα επιθέσεων. Ωστόσο, οι σύγχρονες απειλές εξελίσσονται διαρκώς και απαιτούν καινοτόμες προσεγγίσεις για την αντιμετώπισή τους. Οι ερευνητές [170] αναγνώρισαν αυτή την ανάγκη και επέλεξαν να βασίσουν το έργο τους στο πιο πρόσφατο σύνολο δεδομένων UNSW-NB15, το οποίο περιλαμβάνει μοντέρνες κανονικές δραστηριότητες καθώς και σύγχρονες κατηγορίες επιθέσεων, όπως επιθέσεις DoS, worms, reconnaissance και shellcode.

Ο κύριος στόχος της έρευνας ήταν η ανάπτυξη ενός υβριδικού μοντέλου που να αξιοποιεί τις δυνατότητες δύο διαφορετικών τεχνικών Τεχνητής Νοημοσύνης:

- Κλασικού Αυτόματου Κωδικοποιητή (Classical AutoEncoder - CAE) για την εξαγωγή χαρακτηριστικών και τη μείωση της πολυπλοκότητας των δεδομένων.
- Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks - DNN) για την ακριβή ταξινόμηση των δεδομένων ως "κανονικά" ή "ανώμαλα".

Οι μέθοδοι που χρησιμοποιήθηκαν είναι οι εξής:

1. Επιλογή και Επεξεργασία Δεδομένων

Οι επιστήμονες χρησιμοποίησαν το σύνολο δεδομένων UNSW-NB15, το οποίο περιλαμβάνει 49 χαρακτηριστικά και πάνω από 2,5 εκατομμύρια παραδείγματα δικτυακής κίνησης. Τα δεδομένα προεπεξεργάστηκαν με τις εξής τεχνικές:

- **Κωδικοποίηση one-hot (One-Hot Encoding)** για τη μετατροπή συμβολικών χαρακτηριστικών σε αριθμητικά δεδομένα.
- **Κανονικοποίηση (Min-Max Scaling)** για την προσαρμογή όλων των τιμών στο εύρος [0,1].

2. Μηχανική Χαρακτηριστικών με Κλασικό Αυτόματο Κωδικοποιητή (CAE)

Η μηχανική χαρακτηριστικών είναι κρίσιμη για τη μείωση του αριθμού των χαρακτηριστικών και την ενίσχυση της απόδοσης του ταξινομητή. Ο **Κλασικός Αυτόματος Κωδικοποιητής (CAE)** είναι μια αρχιτεκτονική που αποτελείται από δύο μέρη:

- Έναν **κωδικοποιητή (encoder)** που συμπιέζει τα δεδομένα σε μικρότερη διάσταση.
- Έναν **αποκωδικοποιητή (decoder)** που ανακατασκευάζει τα δεδομένα στην αρχική τους μορφή.

Ο CAE χρησιμοποιεί μη γραμμικές συναρτήσεις ενεργοποίησης για να εξάγει λανθάνουσες σχέσεις μεταξύ των χαρακτηριστικών και να δημιουργήσει πιο αποδοτικές αναπαραστάσεις των δεδομένων.

3. Ταξινόμηση με Βαθύ Νευρωνικό Δίκτυο (DNN)

Το DNN που χρησιμοποιήθηκε αποτελείται από τέσσερα επίπεδα. Η εκπαίδευσή του πραγματοποιείται με τη μέθοδο **backpropagation** και τη στοχαστική βαθμιδωτή κατάβαση (stochastic gradient descent). Μετά την εκπαίδευση, το DNN είναι σε θέση να ταξινομεί νέες παρατηρήσεις ως "κανονικές" ή "ανώμαλες", βασιζόμενο στα χαρακτηριστικά που εξήχθησαν από τον CAE.

Η υβριδική προσέγγιση (CAE + DNN) αξιολογήθηκε ως προς την απόδοσή της με διάφορους δείκτες: ακρίβεια, ανάκληση, F1-score, και καμπύλη ROC. Η μέθοδος απέδωσε καλύτερα από τους παραδοσιακούς αλγορίθμους, όπως το Random Forest (RF), επιτυγχάνοντας υψηλότερη ακρίβεια και γενίκευση και μειωμένα ψευδώς θετικά ποσοστά (False Positive Rates).

Στη σύγχρονη επιστημονική κοινότητα, η χρήση της τεχνητής νοημοσύνης (AI) αποκτά ολοένα και μεγαλύτερη σημασία, προσφέροντας νέες δυνατότητες ανάλυσης και επίλυσης σύνθετων προβλημάτων. Μια πρόσφατη μελέτη [171], που διερεύνησε την εφαρμογή προηγμένων μεθόδων AI σε έναν τομέα μεγάλης επιστημονικής και κοινωνικής αξίας, αναδεικνύει την ισχύ αυτών των τεχνολογιών και τη δυνατότητά τους να παράγουν καινοτόμα αποτελέσματα.

Οι επιστήμονες ξεκίνησαν από την αναγνώριση της ανάγκης να εξελιχθούν οι παραδοσιακές μέθοδοι ανάλυσης και επεξεργασίας δεδομένων, οι οποίες συχνά περιορίζονται από την πολυπλοκότητα και τον όγκο των δεδομένων. Η αυξανόμενη ικανότητα συλλογής τεράστιων ποσοτήτων πληροφοριών από διαφορετικές πηγές, σε συνδυασμό με την πρόοδο στους αλγόριθμους AI, τους οδήγησε να εξετάσουν πώς αυτές οι τεχνολογίες θα μπορούσαν να εφαρμοστούν αποτελεσματικά για να ξεπεραστούν αυτά τα εμπόδια.

Η βασική υπόθεση της μελέτης ήταν ότι η AI δεν προσφέρει μόνο ταχύτερη επεξεργασία δεδομένων, αλλά και τη δυνατότητα να αποκαλύψει υποκείμενες σχέσεις, μοτίβα και δομές που είναι αόρατες στις συμβατικές μεθόδους. Στο επίκεντρο της σκέψης τους βρισκόταν η ιδέα ότι οι αλγόριθμοι μηχανικής μάθησης μπορούν να λειτουργήσουν ως εργαλεία για την προσομοίωση και την πρόβλεψη πολύπλοκων φαινομένων.

Οι επιστήμονες χρησιμοποίησαν διάφορες προηγμένες τεχνικές και μεθόδους, οι οποίες ενσωμάτωσαν στοιχεία από διαφορετικά πεδία της τεχνητής νοημοσύνης:

1. **Βαθιά Μάθηση (Deep Learning)**: Εφαρμόστηκαν νευρωνικά δίκτυα πολλαπλών επιπέδων για την ανάλυση μη δομημένων δεδομένων, όπως εικόνες, κείμενα ή ηχητικά δεδομένα. Αυτή η προσέγγιση επέτρεψε την εξαγωγή πιο σύνθετων και πλούσιων χαρακτηριστικών από τις αρχικές πληροφορίες.
2. **Μηχανική Μάθηση (Machine Learning)**: Χρησιμοποιήθηκαν αλγόριθμοι επιτηρούμενης και μη επιτηρούμενης μάθησης, με σκοπό την κατηγοριοποίηση δεδομένων και την ανακάλυψη σχέσεων μεταξύ παραγόντων.
3. **Ανάλυση Μεγάλων Δεδομένων (Big Data Analytics)**: Η AI χρησιμοποιήθηκε για την επεξεργασία και ερμηνεία τεράστιων συνόλων δεδομένων, με παράλληλη χρήση υποδομών υπολογιστικού νέφους (cloud computing).

4. **Ενισχυτική Μάθηση (Reinforcement Learning)**: Η μέθοδος αυτή επέτρεψε στους ερευνητές να αναπτύξουν συστήματα που μαθαίνουν μέσω δοκιμής και λάθους, βελτιστοποιώντας συνεχώς τις αποφάσεις τους με βάση τα αποτελέσματα.

5. **Επεξεργασία Φυσικής Γλώσσας (NLP)**: Ειδικοί αλγόριθμοι AI χρησιμοποιήθηκαν για την ανάλυση κειμένων και τη δημιουργία προτύπων που διευκολύνουν την κατανόηση του φυσικού λόγου.

Η χρήση της AI σε αυτή τη μελέτη αποδεικνύει την αξία της ως εργαλείο ανάλυσης και καινοτομίας. Οι επιστήμονες κατάφεραν να επιτύχουν σημαντική πρόοδο στην κατανόηση των φαινομένων που μελέτησαν, αναδεικνύοντας τις δυνατότητες της AI να προσφέρει πρακτικές λύσεις σε πραγματικά προβλήματα. Η έρευνα αυτή αποτελεί έναν σημαντικό ορόσημο για τη μελλοντική εφαρμογή της τεχνητής νοημοσύνης σε διάφορους επιστημονικούς και κοινωνικούς τομείς.

Οι επιστήμονες που πραγματοποίησαν την έρευνα [172] είχαν ως βασικό στόχο να αντιμετωπίσουν την αυξανόμενη πολυπλοκότητα και τον όγκο των κυβερνοεπιθέσεων, οι οποίες απειλούν τη διαδικτυακή ασφάλεια. Οι επιθέσεις αυτές δημιουργούν ένα τεράστιο πλήθος ειδοποιήσεων από τα Συστήματα Ανίχνευσης Εισβολών (IDS), με αποτέλεσμα οι αναλυτές να δυσκολεύονται να ξεχωρίσουν ποιες απειλές είναι ουσιαστικά σημαντικές και αξίζουν περαιτέρω διερεύνηση. Για να βοηθήσουν σε αυτήν την πρόκληση, οι ερευνητές ανέπτυξαν το πλαίσιο **FAIXID**, το οποίο συνδυάζει τεχνικές Καθαρισμού Δεδομένων (Data Cleaning) και Εξηγήσιμης Τεχνητής Νοημοσύνης (Explainable AI - XAI). Το πλαίσιο αυτό αποσκοπεί στη διευκόλυνση των αναλυτών μέσω της παροχής κατανοητών, χρήσιμων και επεξηγηματικών πληροφοριών για πιθανές απειλές.

Για την υλοποίηση του πλαισίου FAIXID, οι ερευνητές χρησιμοποίησαν τις ακόλουθες μεθόδους και τεχνικές:

1. Καθαρισμός Δεδομένων (Data Cleaning):

Χρησιμοποιήθηκαν τεχνικές που αφαιρούν την αλληλεπίδραση μεταξύ χαρακτηριστικών (features) στα δεδομένα, μειώνοντας την πολυπλοκότητα των εξηγήσεων. Αυτή η διαδικασία βελτίωσε την εξηγησιμότητα μειώνοντας τον αριθμό των «γνωστικών τμημάτων» (cognitive chunks) που απαιτούνται για την κατανόηση μιας εξήγησης.

2. Μέτρηση Εξηγησιμότητας με Proxy Tasks:

Αναπτύχθηκε μια μεθοδολογία βασισμένη σε τύπους, οι οποίοι ποσοτικοποιούν την εξηγησιμότητα με βάση τον αριθμό και την αλληλεπίδραση των γνωστικών τμημάτων. Οι τύποι λαμβάνουν υπόψη την πολυπλοκότητα της πληροφορίας που χρειάζεται να επεξεργαστεί ένας αναλυτής για να κατανοήσει μια εξήγηση. Τα πειράματα έδειξαν ότι ο καθαρισμός δεδομένων αύξησε την εξηγησιμότητα κατά 211%, κάτι που ενισχύει τη σαφήνεια των αποτελεσμάτων.

3. Ανάλυση Αλληλεπίδρασης Χαρακτηριστικών (Feature Interaction):

Χρησιμοποιήθηκαν εργαλεία όπως το iml package στη γλώσσα προγραμματισμού R, για τη μέτρηση της «δύναμης αλληλεπίδρασης» μεταξύ χαρακτηριστικών. Η μείωση αυτών των αλληλεπιδράσεων συνέβαλε στη βελτίωση της εξηγησιμότητας.

4. **Ανάλυση Μέσω Αξιολόγησης Χρηστών (Human-Subject Studies):**

Πραγματοποιήθηκαν μελέτες με περιορισμένο αριθμό συμμετεχόντων για να αξιολογηθεί η χρησιμότητα, η κατανοητότητα και η αποτελεσματικότητα των εξηγήσεων. Αν και τα αποτελέσματα ήταν θετικά, οι ερευνητές αναγνωρίζουν την ανάγκη για μεγαλύτερες μελέτες με επαγγελματίες αναλυτές.

5. **Μαθηματική Ποσοτικοποίηση της Εξηγησιμότητας:**

Εισήχθησαν τύποι για τον υπολογισμό της εξηγησιμότητας, λαμβάνοντας υπόψη τον αριθμό εισόδων και εξόδων (cognitive chunks) καθώς και τη δύναμη αλληλεπίδρασης μεταξύ τους.

Η ερευνητική ομάδα κατέληξε στο ότι το πλαίσιο FAIXIX βελτιώνει την κατανόηση των αναλυτών, παρέχοντας εξηγήσεις που είναι σαφείς και χρήσιμες. Η διαδικασία με την οποία καθορίζονται τα δεδομένα, διευκολύνει την λήψη των αποφάσεων σε περίπτωση πιθανών κυβερνοαπειλών. Παρά την σημαντική πρόοδο, υπάρχουν περιορισμοί, όπως είναι η ανάγκη για περισσότερες δοκιμές σε μεγαλύτερα και πιο ποικίλα σύνολα δεδομένων, καθώς και η ανάπτυξη πιο προηγμένων εργαλείων εξηγησιμότητας.

Η επιστημονική ομάδα που ανέπτυξε τη μέθοδο **EsPADA** [173] εργάστηκε με στόχο την ενίσχυση της ασφάλειας σε δικτυακά περιβάλλοντα, μέσω της ανίχνευσης κακόβουλου λογισμικού. Η βασική πρόκληση που αναγνώρισαν ήταν η ανάγκη για ανθεκτικότητα απέναντι σε προηγμένες τεχνικές απόκρυψης και μιμητισμού που χρησιμοποιούν οι επιτιθέμενοι για να παρακάμψουν συστήματα ανίχνευσης.

Οι ερευνητές βασίστηκαν στην παραδοχή ότι τα φορτία δικτύου (payloads) περιέχουν χαρακτηριστικά που μπορούν να ταυτοποιηθούν και να διαχωριστούν μεταξύ νόμιμων και κακόβουλων. Ωστόσο, αναγνώρισαν ότι οι παραδοσιακές μέθοδοι ανίχνευσης, όπως εκείνες που βασίζονται σε στατικές υπογραφές ή βασικές στατιστικές, είναι ευάλωτες στις επιθέσεις μιμητισμού. Οι τελευταίες μιμούνται τα πρότυπα των νόμιμων δεδομένων, καθιστώντας δύσκολη την ανίχνευσή τους.

Για να αντιμετωπίσουν αυτές τις προκλήσεις, οι ερευνητές υιοθέτησαν προηγμένες μεθόδους τεχνητής νοημοσύνης και επεξεργασίας δεδομένων, στοχεύοντας στη δημιουργία ενός ανθεκτικού συστήματος που προσαρμόζεται δυναμικά στις απειλές.

1. **N-grams**: Αυτή η μέθοδος χρησιμοποιήθηκε για την ανάλυση ακολουθιών δεδομένων από τα φορτία δικτύου. Τα n-grams επιτρέπουν την ανίχνευση επαναλαμβανόμενων προτύπων που μπορεί να υποδηλώνουν κακόβουλη δραστηριότητα.
2. **Bloom Filters**: Χρησιμοποίησαν αυτή τη δομή δεδομένων για την αποδοτική αποθήκευση και ανάκτηση δεδομένων που σχετίζονται με τα χαρακτηριστικά των φορτίων. Οι Bloom filters είναι ιδιαίτερα χρήσιμοι για την ανίχνευση απειλών σε μεγάλα σύνολα δεδομένων.
3. **Μέθοδοι Ενίσχυσης (Strengthening)**: Εφαρμόστηκαν τεχνικές ενίσχυσης για την αύξηση της ανθεκτικότητας του συστήματος σε επιθέσεις μιμητισμού, βασισμένες σε προηγμένες μετρικές, όπως ο δείκτης Youden και η τοπική βαθμολογία ομοιότητας (Local Similarity Score).
4. **Μοντέλα Μηχανικής Μάθησης**: Ενσωμάτωσαν μεθόδους ταξινόμησης και ανίχνευσης ανωμαλιών, προσαρμοσμένες για την αντιμετώπιση επιθέσεων που περιλαμβάνουν καμουφλαρισμένα κακόβουλα φορτία.

Η EsPADA αξιοποίησε με επιτυχία τις παραπάνω τεχνικές για να δημιουργήσει ένα καινοτόμο σύστημα ανίχνευσης κακόβουλου λογισμικού. Παρόλο που οι ενισχύσεις δεν κατάφεραν να εξαλείψουν πλήρως τις επιπτώσεις των επιθέσεων μιμητισμού, μείωσαν σημαντικά την αποτελεσματικότητα αυτών των επιθέσεων, καθιστώντας το σύστημα πιο ανθεκτικό σε πραγματικά περιβάλλοντα.

Η επόμενη ερευνητική εργασία [174] επικεντρώνεται στην ανάπτυξη ενός νέου αλγορίθμου τεχνητής νοημοσύνης, του **DnRaNN (Dense Nuclei Recurrent Artificial Neural Network)**, με στόχο την ανίχνευση κυβερνοεπιθέσεων σε περιβάλλοντα Διαδικτύου των Πραγμάτων (IoT). Το σκεπτικό των επιστημόνων βασίζεται στην προσομοίωση της δομής του ανθρώπινου εγκεφάλου, όπου ομάδες νευρικών κυττάρων επικοινωνούν πυκνά μεταξύ τους μέσω συνάψεων και δενδριτών.

Οι κύριοι στόχοι της έρευνας είναι η ακριβής ανίχνευση κυβερνοεπιθέσεων, η διαχείριση πολύπλοκων δεδομένων IoT και η βελτίωση της απόδοσης, συγκρίνοντας με άλλες γνωστές μεθόδους μηχανικής μάθησης.

Η έρευνα υιοθετεί τις εξής προσεγγίσεις και τεχνικές:

1. **Αρχιτεκτονική του DnRaNN**: Ο προτεινόμενος αλγόριθμος αποτελείται από ένα εισαγωγικό επίπεδο, τέσσερα ενδιάμεσα επίπεδα με πυκνούς πυρήνες και ένα επίπεδο εξόδου. Κάθε πυρήνας περιέχει νευρώνες που επικοινωνούν πλήρως μεταξύ τους, ενώ οι πληροφορίες μεταφέρονται μεταξύ των επιπέδων μέσω ενός πολυεπίπεδου feed-forward δικτύου.
2. **Βελτιστοποίηση με Gradient Descent**: Η εκπαίδευση του μοντέλου βασίζεται στον αλγόριθμο gradient descent, ο οποίος χρησιμοποιείται για τη ρύθμιση των παραμέτρων του δικτύου.
3. **Χρήση του συνόλου δεδομένων ToN_IoT**: Το μοντέλο εκπαιδεύτηκε και αξιολογήθηκε σε ένα σύγχρονο και πλούσιο σύνολο δεδομένων, το οποίο περιλαμβάνει δείγματα από 10 διαφορετικούς τύπους επιθέσεων.
4. **Προεπεξεργασία και Κανονικοποίηση Δεδομένων**: Τα δεδομένα προετοιμάστηκαν μέσω της κωδικοποίησης κατηγοριοποιημένων χαρακτηριστικών και της κανονικοποίησης μέσω min-max scaling, ώστε να διασφαλιστεί η συμβατότητα και η αποτελεσματικότητα της μάθησης.

Οι επιστήμονες παρουσίασαν έναν καινοτόμο αλγόριθμο που συνδυάζει τη βιολογική έμπνευση από τον ανθρώπινο εγκέφαλο με τις τελευταίες εξελίξεις της τεχνητής νοημοσύνης. Η υψηλή ακρίβεια ανίχνευσης κυβερνοεπιθέσεων και η αποτελεσματικότητα του DnRaNN καθιστούν τη μέθοδο αυτή μια ισχυρή λύση για την ασφάλεια των συστημάτων IoT. Μελλοντικά, προτείνεται η ενσωμάτωση της μεθόδου σε hardware συσκευές, όπως FPGA [175], για ακόμη μεγαλύτερη απόδοση σε πραγματικό χρόνο.

Στην πολυκατηγορική ταξινόμηση, τα δεδομένα χωρίζονται σε πολλές ξεχωριστές κατηγορίες, όπου κάθε κατηγορία έχει την ίδια ετικέτα για όλα τα δεδομένα που της ανήκουν. Στο πλαίσιο της ανίχνευσης εισβολών, πέρα από την "κανονική" κίνηση, υπάρχουν κατηγορίες όπως επιθέσεις άρνησης υπηρεσίας (DoS), κατανεμημένες επιθέσεις άρνησης υπηρεσίας (DDoS), και επιθέσεις που στοχεύουν στην απόκτηση πρόσβασης από απομακρυσμένο χρήστη ή στην απόκτηση δικαιωμάτων root.

Η ανίχνευση εισβολών μέσω πολυκατηγορικής ταξινόμησης εξετάζει διάφορες πτυχές, όπως τη χρήση διαφορετικών μοντέλων ταξινόμησης, την εξαγωγή των κατάλληλων χαρακτηριστικών από

τα δεδομένα, τη βελτίωση της απόδοσης των μοντέλων μέσω ρύθμισης υπερπαραμέτρων και την αντιμετώπιση του προβλήματος της άνισης κατανομής δεδομένων ανάμεσα στις κατηγορίες.

Οι επιστήμονες που διεξήγαγαν την έρευνα αυτή [176] είχαν ως βασικό στόχο να αναπτύξουν και να αξιολογήσουν ένα σύστημα ανίχνευσης εισβολών (IDS) που θα είναι ικανό να αναγνωρίζει και να προλαμβάνει κυβερνοεπιθέσεις σε δίκτυα IoT. Η αυξημένη εξάρτηση των σύγχρονων κοινωνιών από τα έξυπνα δίκτυα και οι διαρκώς εξελισσόμενες κυβερνοαπειλές, όπως οι επιθέσεις DDoS, SQL Injection, και άλλες, δημιούργησαν την ανάγκη για καινοτόμες μεθόδους ανίχνευσης και άμεσης αντιμετώπισης αυτών των κινδύνων. Το σκεπτικό των επιστημόνων βασίζεται στην αξιοποίηση τεχνολογιών τεχνητής νοημοσύνης και βαθιάς μάθησης για τη βελτίωση της ακρίβειας, της ταχύτητας, και της αποτελεσματικότητας των IDS.

Οι στόχοι της έρευνας αναγράφονται παρακάτω:

- Αναγνώριση επιθέσεων σε πραγματικό χρόνο: Η ανάπτυξη ενός συστήματος που μπορεί να εντοπίσει επιθέσεις με υψηλή ακρίβεια και μικρή καθυστέρηση.
- Ενίσχυση της ασφάλειας δικτύων IoT: Η πρόληψη παραβιάσεων σε δίκτυα που χρησιμοποιούνται από οργανισμούς, επιχειρήσεις, αλλά και καταναλωτές.
- Σύγκριση με άλλες μεθόδους: Η αξιολόγηση της προτεινόμενης λύσης σε σχέση με υπάρχοντα μοντέλα για να αποδειχθεί η ανωτερότητά της.
- Ανάπτυξη ενός ευέλικτου μοντέλου: Ένα σύστημα που μπορεί να επεκταθεί σε μεγαλύτερα δίκτυα, όπως έξυπνες πόλεις και δίκτυα επόμενης γενιάς.

Οι επιστήμονες χρησιμοποίησαν τις εξής μεθόδους και τεχνολογίες:

1. **LSTM (Long Short-Term Memory)**: Ένα προηγμένο μοντέλο νευρωνικών δικτύων που χρησιμοποιείται για την ανάλυση σειρών δεδομένων. Η LSTM αποδείχθηκε εξαιρετικά αποτελεσματική στη διάκριση μεταξύ κανονικής δραστηριότητας και κακόβουλων επιθέσεων, καθώς έχει την ικανότητα να "θυμάται" μακροχρόνιες σχέσεις στα δεδομένα.

2. **Εκπαίδευση και δοκιμή σε σύνολα δεδομένων**: Οι ερευνητές εκπαίδευσαν το μοντέλο τους σε μεγάλα και ποικίλα σύνολα δεδομένων, τα οποία περιελάμβαναν διάφορους τύπους επιθέσεων (π.χ. UDP-Lag, SNMP, Web-DDoS). Στη φάση της δοκιμής, αξιολόγησαν την ακρίβεια, την απώλεια, και τις καμπύλες ROC (Receiver Operating Characteristics).

3. **Αξιολόγηση απόδοσης με μετρικές**: Χρησιμοποίησαν δείκτες όπως η ακρίβεια (Accuracy), η ανάκληση (Recall), η ακρίβεια πρόβλεψης (Precision), και το F1-score για να μετρήσουν την απόδοση του μοντέλου.

4. **Γραφική αναπαράσταση αποτελεσμάτων**: Παρουσίασαν τα αποτελέσματα μέσω καμπυλών ακρίβειας και απωλειών, καθώς και μέσω καμπυλών ROC, για να καταδείξουν την αποτελεσματικότητα του μοντέλου σε διαφορετικές επιθέσεις.

Το σύστημα που ανέπτυξαν οι επιστήμονες πέτυχε εξαιρετικές επιδόσεις, με τη μέγιστη ακρίβεια να φτάνει το 99.97% σε επιθέσεις SNMP. Επιπλέον, το μοντέλο τους ξεπέρασε σε απόδοση υπάρχοντα συστήματα ανίχνευσης, αποδεικνύοντας την αξία της χρήσης βαθιάς μάθησης στην ασφάλεια δικτύων.

Οι επιστήμονες που εκπόνησαν αυτή την έρευνα [177] επικεντρώθηκαν στην ανάπτυξη ενός αποδοτικού και ακριβούς συστήματος για την ανίχνευση κακόβουλων συμπεριφορών στο δίκτυο και

την ταξινόμηση της δικτυακής κίνησης σε 13 διαφορετικές κατηγορίες. Το σκεπτικό τους βασίζεται στην ανάγκη για λύσεις που να μπορούν να εντοπίζουν κυβερνοεπιθέσεις με ακρίβεια και αξιοπιστία, ακόμα και όταν χρησιμοποιούνται σχετικά μικρά και απλά μοντέλα. Στόχος τους ήταν να αναπτύξουν ένα σύστημα που συνδυάζει υψηλή απόδοση και χαμηλό υπολογιστικό κόστος, ώστε να μπορεί να εφαρμοστεί σε πραγματικό χρόνο και σε ποικίλα περιβάλλοντα.

Οι επιστήμονες προκειμένου να υλοποιήσουν την ανάπτυξη ενός τέτοιου συστήματος, χρησιμοποίησαν τις ακόλουθες μεθόδους και τεχνικές.

1. **Βαθιά Νευρωνικά Δίκτυα (DNN)**: Οι επιστήμονες χρησιμοποίησαν ένα βαθύ νευρωνικό δίκτυο με 8 κρυφά επίπεδα που διαφέρει σε αριθμό κόμβων (140, 120, 100, 80, 60, 40, 20, 120). Η έξοδος του δικτύου είναι ένα επίπεδο softmax, το οποίο προβλέπει τις πιθανότητες για τις 13 κατηγορίες.
2. **Μηχανική Χαρακτηριστικών**: Από τα αρχικά δεδομένα εξήγαγαν 44 χαρακτηριστικά που είναι απαραίτητα για την ακρίβεια του μοντέλου. Επικεντρώθηκαν στη μείωση του όγκου δεδομένων χωρίς να μειώσουν την απόδοση.
3. **Τεχνικές Αρχικοποίησης και Ενεργοποίησης**: Χρησιμοποίησαν την τεχνική αρχικοποίησης lecun-uniform για τα κρυφά επίπεδα και glorot-uniform για την έξοδο. Η συνάρτηση ενεργοποίησης ReLU επιλέχθηκε για όλα τα κρυφά επίπεδα, λόγω της αποτελεσματικότητάς της στην εκμάθηση σύνθετων μοτίβων.
4. **Βελτιστοποίηση**: Χρησιμοποιήθηκε ο Adam optimizer για την εκπαίδευση του μοντέλου, καθώς προσφέρει γρήγορη σύγκλιση και αποδοτική εκπαίδευση σε βαθιά δίκτυα.
5. **Αξιολόγηση**: Το μοντέλο αξιολογήθηκε με 10-πλή cross-validation, εξασφαλίζοντας τη σταθερότητα και ακρίβεια των αποτελεσμάτων. Υπολογίστηκαν δείκτες όπως η ακρίβεια (accuracy), η ανάκληση (recall), η ακρίβεια (precision), ο συντελεστής F1 και το ποσοστό ψευδών θετικών (FPR).
6. **Απλότητα και Χωρίς Κανονικοποίηση**: Αν και δοκίμασαν τεχνικές κανονικοποίησης όπως L1, L2 και dropout, διαπίστωσαν ότι δεν είχαν σημαντική επίδραση στα αποτελέσματα.
7. **Σύγκριση με Άλλες Τεχνικές AI**: Προβλέπεται η χρήση εναλλακτικών αρχιτεκτονικών, όπως RNN (LSTM, GRU) και CNN, στο μέλλον, καθώς τα δεδομένα είναι σειριακά και μπορεί να επωφεληθούν από αυτές τις τεχνικές.

Το μοντέλο που πρότειναν οι επιστήμονες πέτυχε εξαιρετικά αποτελέσματα, με ακρίβεια 99.95%, F1 score 94.1% και μέσο δείκτη AUC 0.99. Παρά την απλότητα και το μικρό μέγεθός του, το μοντέλο ξεπέρασε πολλά προηγούμενα μοντέλα και έδειξε πως η σωστή επιλογή χαρακτηριστικών και αρχιτεκτονικής μπορεί να παράγει αποτελεσματικά συστήματα για την ανίχνευση κυβερνοεπιθέσεων.

Οι επιστήμονες παρατηρώντας την αύξηση της πολυπλοκότητας και του όγκου των επιθέσεων, διεξήγαγαν την έρευνα αυτή [178], στοχεύοντας στην αντιμετώπιση των σύγχρονων προκλήσεων στις δικτυακές προσβολές. Το σκεπτικό τους βασίζεται στην ανάγκη δημιουργίας μοντέλων που όχι μόνο αποδίδουν καλά αλλά είναι και εύκολα ερμηνεύσιμα, ώστε να κατανοούμε ποια χαρακτηριστικά του δικτύου είναι κρίσιμα για την ανίχνευση κάθε κατηγορίας επίθεσης.

Ο βασικός τους στόχος είναι να αναπτύξουν μεθόδους που να επιλέγουν τα πιο κρίσιμα χαρακτηριστικά, από μεγάλους όγκους δεδομένων, μειώνοντας έτσι την ανάγκη επεξεργασίας περιττών πληροφοριών. Παράλληλα, οι μέθοδοι χρειάζεται να εξασφαλίζουν ότι τα χαρακτηριστικά που επιλέγονται παραμένουν σημαντικά ανεξάρτητα από τις αλλαγές στα δεδομένα εκπαίδευσης.

Η έρευνα βασίστηκε σε έναν συνδυασμό τεχνικών μηχανικής μάθησης και στατιστικών μεθόδων για την επίτευξη των στόχων:

1. **Υβριδική Επιλογή Χαρακτηριστικών**:
    - Μη επιβλεπόμενη αφαίρεση χαρακτηριστικών: Αρχικά, αφαιρέθηκαν χαρακτηριστικά που εμφάνιζαν πολυκολλητικότητα (multicollinearity) μέσω μη επιβλεπόμενων τεχνικών ανάλυσης δεδομένων.
    - Επιβλεπόμενη ιεράρχηση: Στη συνέχεια, τα εναπομείναντα χαρακτηριστικά αξιολογήθηκαν μέσω δέντρων ακραίας τυχαίας δασοκομίας (extremely randomized trees), ώστε να κατασκευαστεί μια ιεραρχία χαρακτηριστικών με βάση τη σημασία τους.
2. **Στατιστικές Δοκιμές**:
    - Χρησιμοποιήθηκαν οι δοκιμές Friedman και Wilcoxon για τη σύγκριση της σημασίας των χαρακτηριστικών και την εξασφάλιση στατιστικής αξιοπιστίας. Επιπλέον, η μέθοδος Benjamini-Hochberg εφαρμόστηκε για τον έλεγχο του ποσοστού ψευδώς θετικών αποτελεσμάτων.
3. **Μοντέλα Μηχανικής Μάθησης**:
    - Δοκιμάστηκαν τυποποιημένοι αλγόριθμοι όπως λογιστική παλινδρόμηση και τυχαία δάση με σταδιακή προσθήκη των καλύτερων χαρακτηριστικών από την ιεραρχία. Τα μοντέλα αξιολογήθηκαν σε συνθήκες αυστηρής διασταυρούμενης επικύρωσης (cross-validation) για τη διασφάλιση της γενίκευσης.
4. **Εξαγωγή Ιεραρχίας Χαρακτηριστικών**:
    - Δημιουργήθηκαν ιεραρχίες που κατατάσσουν τα χαρακτηριστικά σε ομάδες, όπου τα χαρακτηριστικά μιας ανώτερης ομάδας είναι στατιστικά πιο σημαντικά από αυτά μιας κατώτερης. Αυτές οι ιεραρχίες είναι σταθερές και μπορούν να επαναχρησιμοποιηθούν από άλλους ερευνητές.
5. **Αξιολόγηση Μινιμαλιστικών Μοντέλων**:
    - Δόθηκε έμφαση σε μοντέλα που χρησιμοποιούν το ελάχιστο δυνατό πλήθος χαρακτηριστικών (1-4) για την επίτευξη ακρίβειας άνω του 99,5% σε κατηγορίες επιθέσεων όπως Brute Force, DoS, DDoS και Portscan.

Οι μέθοδοι που αναπτύχθηκαν συνδυάζουν στατιστική αξιοπιστία με αποδοτικότητα και ερμηνευσιμότητα. Η έρευνα απέδειξε ότι οι επιλεγμένες τεχνικές AI μπορούν να κατασκευάσουν μοντέλα υψηλής απόδοσης για την ανίχνευση εισβολών, ενώ οι ιεραρχίες χαρακτηριστικών αποτελούν ένα χρήσιμο εργαλείο για τη μελλοντική έρευνα σε πιο ρεαλιστικά δεδομένα και συνθήκες.

Οι επιστήμονες που ανέπτυξαν το **IGAN-IDS** [179] είχαν ως κεντρικό στόχο την αντιμετώπιση του προβλήματος της ανισορροπίας κλάσεων, το οποίο αποτελεί βασική πρόκληση στην ανίχνευση εισβολών σε δίκτυα. Οι εισβολές αυτού του τύπου συχνά περιλαμβάνουν σπάνιες επιθέσεις που δύσκολα ανιχνεύονται, επειδή τα δεδομένα εκπαίδευσης είναι συνήθως κυριαρχούμενα από μη κακόβουλες ενέργειες. Αναγνωρίζοντας ότι η δημιουργία αντιπροσωπευτικών δεδομένων για τις μειονοτικές κλάσεις θα μπορούσε να βελτιώσει σημαντικά την απόδοση των συστημάτων

ανίχνευσης, πρότειναν τη χρήση μιας εξελιγμένης μεθόδου Generative Adversarial Networks (GAN), ονομάζοντας την Imbalance Generative Adversarial Network (IGAN). Το IGAN δεν παράγει δείγματα για όλες τις κλάσεις, αλλά εστιάζει μόνο στις μειονοτικές, βελτιώνοντας έτσι την ισορροπία των δεδομένων εκπαίδευσης.

Οι επιστήμονες χρησιμοποίησαν προηγμένες τεχνικές τεχνητής νοημοσύνης για την επίτευξη των στόχων τους:

1. **Generative Adversarial Networks (GAN)**: Το IGAN βασίζεται στη φιλοσοφία των GAN, όπου δύο νευρωνικά δίκτυα (γεννήτρια και διακριτής) εκπαιδεύονται μαζί. Η γεννήτρια δημιουργεί νέα δείγματα, ενώ η διακριτής προσπαθεί να διαχωρίσει τα πραγματικά από τα παραγόμενα δείγματα. Το IGAN εισάγει έναν "φίλτρο" που εστιάζει αποκλειστικά στις μειονοτικές κλάσεις, διασφαλίζοντας ότι η γεννήτρια δεν σπαταλά πόρους δημιουργώντας περιττά δείγματα για τις πλειονοτικές κλάσεις.
2. **Feature Extraction (FE)**: Το IGAN-IDS περιλαμβάνει ένα module εξαγωγής χαρακτηριστικών (FE), το οποίο χρησιμοποιεί τεχνικές deep learning για να μετατρέπει τα δεδομένα σε υψηλής διάστασης αναπαραστάσεις. Αυτό επιτρέπει στο σύστημα να εξάγει κρίσιμες πληροφορίες που βοηθούν στην ακριβέστερη κατηγοριοποίηση.
3. **Deep Neural Networks (DNN)**: Στο τελευταίο στάδιο, ένα βαθύ νευρωνικό δίκτυο (DNN) χρησιμοποιείται για την κατηγοριοποίηση των εισερχόμενων δεδομένων δικτύου. Η χρήση παραγόμενων δειγμάτων από το IGAN και χαρακτηριστικών από το FE οδηγεί σε καλύτερη απόδοση.
4. **Αφαίρεση Μονάδων (Ablation Study)**: Για να αξιολογηθεί η συνεισφορά κάθε μεθόδου, πραγματοποιήθηκε αφαίρεση μονάδων (ablation study), όπου διαφορετικά μέρη του συστήματος αφαιρέθηκαν ή τροποποιήθηκαν. Αυτό έδειξε ότι τόσο το IGAN όσο και το FE παίζουν καθοριστικό ρόλο στη συνολική απόδοση.

Η έρευνα έδειξε ότι το IGAN-IDS υπερέχει έναντι 15 άλλων μεθόδων ανίχνευσης, συμπεριλαμβανομένων των κορυφαίων state-of-the-art προσεγγίσεων. Οι μελέτες ανθεκτικότητας και οι δοκιμές σε διαφορετικά datasets επιβεβαίωσαν τη γενική εφαρμοσιμότητα και την ακρίβεια του συστήματος. Οι επιστήμονες κατέληξαν ότι η δημιουργία αντιπροσωπευτικών δειγμάτων και η εξαγωγή χαρακτηριστικών υψηλής ποιότητας είναι καθοριστικής σημασίας για την επιτυχή αντιμετώπιση του προβλήματος της ανισορροπίας δεδομένων.

Στο συγκεκριμένο επιστημονικό έργο [180], οι ερευνητές επικεντρώθηκαν στην ανίχνευση ανωμαλιών σε συστήματα SCADA (Supervisory Control and Data Acquisition), τα οποία είναι κρίσιμα για τη λειτουργία βιομηχανικών υποδομών, όπως οι αγωγοί φυσικού αερίου και τα δίκτυα ύδρευσης. Η ανίχνευση ανωμαλιών σε αυτά τα συστήματα είναι ζωτικής σημασίας, καθώς επιτρέπει τον έγκαιρο εντοπισμό κυβερνοεπιθέσεων ή άλλων τεχνικών δυσλειτουργιών που θα μπορούσαν να οδηγήσουν σε σοβαρές διακοπές ή καταστροφικές συνέπειες.

Οι επιστήμονες αναγνώρισαν ότι τα παραδοσιακά συστήματα ανίχνευσης ανωμαλιών συχνά αποτυγχάνουν να ανιχνεύσουν επιθέσεις μηδενικής ημέρας (zero-day) ή να προσαρμοστούν σε διαφορετικά και ανομοιογενή δεδομένα. Τα υπάρχοντα μοντέλα έχουν περιορισμούς όσον αφορά την ευαισθησία, την ακρίβεια και την ταχύτητα απόκρισης. Επομένως, στόχος τους ήταν να

αναπτύξουν ένα καινοτόμο, πολυεπίπεδο μοντέλο που να συνδυάζει ευφυείς αλγορίθμους και τεχνικές μηχανικής μάθησης για την αποτελεσματική ανίχνευση ανωμαλιών σε πραγματικό χρόνο, ακόμη και σε πολύπλοκα και ανισοκατανεμημένα δεδομένα.

Για την επίτευξη του στόχου τους, οι επιστήμονες υιοθέτησαν μια σειρά από προηγμένες τεχνικές τεχνητής νοημοσύνης και μηχανικής μάθησης:

1. **Bloom-filter**: Αυτή η δομή δεδομένων χρησιμοποιήθηκε για την ταχεία ανίχνευση ανωμαλιών σε επίπεδο πακέτων δικτύου. Με τον τρόπο αυτόν, εξασφαλίστηκε χαμηλός υπολογιστικός χρόνος και αποτελεσματική επεξεργασία μεγάλου όγκου δεδομένων.
2. **Δίκτυο Kohonen (Kohonen Map)**: Πρόκειται για ένα είδος αυτοοργανούμενου χάρτη (Self-Organizing Map) που εντοπίζει μοτίβα και ανωμαλίες στα δεδομένα. Η καινοτομία τους ήταν η χρήση ενός υπεργράφου για τη βελτιστοποίηση της γειτνίασης μεταξύ των κόμβων, ενισχύοντας έτσι την απόδοση και την ευαισθησία του χάρτη.
3. **Υπεργραφική διαμέριση**: Με τη χρήση υπεργραφικών τεχνικών, οι ερευνητές βελτίωσαν τη διαδικασία επιλογής γειτονιάς στο Kohonen map, ενισχύοντας την ακρίβεια του μοντέλου.
4. **Τεχνικές εξισορρόπησης δεδομένων**:
    - **SMOTE (Synthetic Minority Oversampling Technique)**: Χρησιμοποιήθηκε για τη δημιουργία συνθετικών δειγμάτων στην περίπτωση του Gas-Pipeline dataset.
    - **Random Under-Sampling**: Εφαρμόστηκε στο SWaT dataset για την απομάκρυνση πλεοναζόντων δειγμάτων της πλειοψηφικής κλάσης.
5. **Ενισχυμένη PCA (Principal Component Analysis)**: Η χρήση βελτιωμένης ανάλυσης κύριων συνιστωσών αντικατέστησε τη συμβατική τυχαία αρχικοποίηση βαρών στο Kohonen map, αυξάνοντας την αναπαραστατική ικανότητα του μοντέλου.
6. **Διασταυρωμένη επικύρωση (k-fold cross-validation)**: Για τη βελτίωση της γενικευσιμότητας, χρησιμοποιήθηκαν δείγματα από τα δεδομένα, επιβεβαιώνοντας την ακρίβεια, την ανάκληση και το F-score του μοντέλου.
7. **Μοντέλα σύγκρισης (SOTA)**: Το προτεινόμενο μοντέλο συγκρίθηκε με σύγχρονες τεχνικές, όπως το CNN, το Random Forest, και το KNN, αποδεικνύοντας την υπεροχή του με σημαντικά υψηλότερα αποτελέσματα.

Η έρευνα καταλήγει ότι το προτεινόμενο μοντέλο BLOSOM είναι αποτελεσματικό, γενικεύσιμο και ικανό να λειτουργήσει σε πραγματικό χρόνο. Επιπλέον, η προτεινόμενη υβριδική προσέγγιση αντιμετωπίζει επιτυχώς προβλήματα ανισοκατανομής δεδομένων και ανίχνευσης zero-day επιθέσεων, καθιστώντας το μια βιώσιμη λύση για εφαρμογές σε κρίσιμες υποδομές.

Οι επιστήμονες που πραγματοποίησαν την έρευνα αυτή [181] είχαν ως στόχο τη δημιουργία μιας αποτελεσματικής και προσαρμόσιμης λύσης για την ανίχνευση επιθέσεων δικτύου σε περιβάλλοντα IoT, τα οποία χαρακτηρίζονται από περιορισμένους πόρους, όπως είναι η ενέργεια, η μνήμη και η επεξεργαστική ισχύς. Το επίκεντρο της έρευνάς τους ήταν η ανάλυση δεδομένων δικτύου και η αξιοποίηση μεθόδων Μηχανικής Μάθησης (AI/ML) για την ταξινόμηση και ανίχνευση επιθέσεων, δίνοντας έμφαση τόσο στην ακρίβεια όσο και στη δυνατότητα εφαρμογής σε πραγματικά δίκτυα IoT.

Η ανάγκη για την ανάπτυξη τεχνικών ανίχνευσης επιθέσεων είναι επιτακτική, καθώς τα δίκτυα IoT είναι ευάλωτα σε κυβερνοεπιθέσεις, όπως επιθέσεις Denial of Service (DoS) και probe attacks. Οι υπάρχουσες λύσεις για ανίχνευση επιθέσεων συχνά δεν είναι κατάλληλες για περιβάλλοντα IoT λόγω των υψηλών απαιτήσεών τους σε υπολογιστικούς πόρους. Έτσι, οι ερευνητές επικεντρώθηκαν σε μεθόδους που μπορούν να λειτουργούν αποδοτικά σε συσκευές με περιορισμένες δυνατότητες.

Οι επιστήμονες χρησιμοποίησαν μια σειρά από μεθόδους επιβλεπόμενης και μη επιβλεπόμενης μάθησης:

Επιβλεπόμενες Μέθοδοι

1. **Δέντρα Απόφασης (Decision Trees - DT)**: Προτιμήθηκαν λόγω της απλότητας, της χαμηλής κατανάλωσης πόρων και της εξηγήσιμης φύσης τους.
2. **Random Forest (RF)**: Παρέχει υψηλή ακρίβεια και σταθερότητα μέσω συνδυασμού πολλών δέντρων απόφασης.
3. **Extreme Gradient Boosting (XGBoost)**: Ένας ισχυρός αλγόριθμος ενισχυτικής μάθησης με εξαιρετική απόδοση στα δεδομένα.
4. **Support Vector Machines (SVM)**: Χρησιμοποιήθηκε με Radial Basis Function (RBF) πυρήνα, αλλά απαιτεί περισσότερους πόρους.
5. **Naive Bayes (NB) και Bayes Network (BN)**: Παρόλο που είναι γρήγοροι και εύχρηστοι, οι μέθοδοι αυτές δεν είχαν καλή απόδοση στα δεδομένα της έρευνας.
6. **AdaBoost και Bagging Trees**: Δοκιμάστηκαν ως μέθοδοι ενίσχυσης (ensemble), αλλά η απόδοσή τους υστερούσε σε σχέση με το XGBoost.

Μη Επιβλεπόμενες Μέθοδοι

1. **K-Means**: Προσπάθησε να ομαδοποιήσει τις επιθέσεις, αλλά η ακρίβεια ήταν περιορισμένη (49.6%).
2. **DBSCAN**: Είχε λίγο καλύτερη ακρίβεια (51.1%) αλλά δεν κατάφερε να ταξινομήσει όλες τις επιθέσεις αποτελεσματικά.
3. **Expectation-Maximization (EM)**: Ξεχώρισε με ακρίβεια 67.2%, αποδεικνύοντας ότι μπορεί να λειτουργήσει καλύτερα σε περιβάλλοντα όπου δεν υπάρχουν ετικέτες.

Η έρευνα καταλήγει ότι τα δέντρα απόφασης είναι η πιο κατάλληλη λύση για ανίχνευση επιθέσεων σε IoT δίκτυα, λόγω της ισορροπίας μεταξύ απόδοσης και αποδοτικότητας πόρων. Παράλληλα, σε περιπτώσεις που δεν υπάρχουν ετικέτες δεδομένων, η μέθοδος EM είναι η πιο υποσχόμενη μη επιβλεπόμενη επιλογή.

Η συγκεκριμένη έρευνα [182] επικεντρώθηκε στη βελτιστοποίηση και αξιολόγηση Τεχνητών Νευρωνικών Δικτύων (ANNs) για την ανίχνευση εισβολών σε δίκτυα υπολογιστών, χρησιμοποιώντας το σύνολο δεδομένων NSL-KDD. Το σκεπτικό των επιστημόνων βασίζεται στη διαρκώς αυξανόμενη ανάγκη για αποτελεσματικά συστήματα ανίχνευσης κυβερνοεπιθέσεων, καθώς τα παραδοσιακά συστήματα αντιμετωπίζουν δυσκολίες στο χειρισμό πολύπλοκων και μεγάλων δεδομένων.

Οι κύριοι στόχοι της έρευνας ήταν:

1. Η ανίχνευση εισβολών με υψηλή ακρίβεια.
2. Η αξιολόγηση διαφορετικών αρχιτεκτονικών ANNs για να βρεθεί η βέλτιστη διαμόρφωση.
3. Η διερεύνηση της επίδρασης των υπερπαραμέτρων, όπως ο αριθμός των επαναλήψεων (epochs), το μέγεθος παρτίδας (batch size), οι συναρτήσεις ενεργοποίησης, και οι αλγόριθμοι βελτιστοποίησης.
4. Η μείωση της πολυπλοκότητας των δεδομένων με τη χρήση τεχνικών μείωσης διαστάσεων όπως το PCA, ώστε να ενισχυθεί η απόδοση των ANNs.

Η έρευνα περιελάμβανε τα εξής στάδια:

1. **Προεπεξεργασία Δεδομένων**:
    - Τα δεδομένα του NSL-KDD υποβλήθηκαν σε one-hot encoding για κατηγορηματικά χαρακτηριστικά και κανονικοποίηση για να εξασφαλιστεί η ομαλή λειτουργία των ANNs.
    - Η τεχνική Principal Component Analysis (PCA) χρησιμοποιήθηκε για τη μείωση των διαστάσεων από 118 σε 50, εξαλείφοντας τον θόρυβο και μειώνοντας την πολυπλοκότητα.
2. **Δημιουργία και Εκπαίδευση Νευρωνικών Δικτύων**:
    - Δοκιμάστηκαν διάφορες αρχιτεκτονικές ANNs, από απλά δίκτυα με 1 κρυφή στρώση και 25 νευρώνες, μέχρι πιο περίπλοκα δίκτυα με 4 κρυφές στρώσεις.
    - Οι συναρτήσεις ενεργοποίησης που χρησιμοποιήθηκαν περιλάμβαναν ReLU, Sigmoid, Hyperbolic Tangent (tanh), κ.ά.
3. **Βελτιστοποίηση Υπερπαραμέτρων**:
    - Χρησιμοποιήθηκε η μέθοδος grid search για τη δοκιμή συνδυασμών υπερπαραμέτρων όπως:
        - Ο αριθμός επαναλήψεων (epochs).
        - Το μέγεθος παρτίδας (batch size).
        - Οι αλγόριθμοι βελτιστοποίησης (Adam, SGD, RMSprop).
    - Στόχος ήταν η επίτευξη της καλύτερης δυνατής ακρίβειας, λαμβάνοντας υπόψη τη συνολική απόδοση και γενίκευση.
4. **Αξιολόγηση Απόδοσης**:
    - Τα μοντέλα αξιολογήθηκαν με βάση την ακρίβεια και την ικανότητά τους να ανιχνεύουν εισβολές.
    - Παρατηρήθηκε ότι το Adam optimiser και η συνάρτηση ενεργοποίησης ReLU παρείχαν τις καλύτερες επιδόσεις με ακρίβεια 99.9%.

Η έρευνα κατέδειξε τη σημασία της σωστής επιλογής αρχιτεκτονικής και υπερπαραμέτρων στα ANNs για την ανίχνευση εισβολών. Με τη χρήση προηγμένων τεχνικών όπως το PCA και τη βελτιστοποίηση μέσω grid search, επιτεύχθηκαν σημαντικά αποτελέσματα που ενισχύουν την αποτελεσματικότητα των συστημάτων ανίχνευσης εισβολών. Οι επιστήμονες κατέληξαν ότι η προσεκτική διαμόρφωση των ANNs μπορεί να αποτελέσει ισχυρό εργαλείο στον τομέα της κυβερνοασφάλειας.

Οι επιστήμονες που ασχολήθηκαν με την παρούσα έρευνα [183] στόχευσαν στην αποτελεσματική ανίχνευση ανωμαλιών και εισβολών σε δίκτυα IoT (Internet of Things), αντιμετωπίζοντας τις ιδιαίτερες προκλήσεις που παρουσιάζει αυτό το πεδίο. Οι προκλήσεις αυτές περιλαμβάνουν την πολυπλοκότητα των δεδομένων που παράγονται από συσκευές IoT, την έλλειψη επαρκών δειγμάτων για ορισμένες κατηγορίες επιθέσεων, καθώς και την ανάγκη για επιλογή χαρακτηριστικών που παρέχουν τις απαραίτητες πληροφορίες για την ακριβή κατηγοριοποίηση.

Το βασικό σκεπτικό πίσω από την έρευνα ήταν ότι η αποτελεσματική επιλογή χαρακτηριστικών είναι καθοριστική για τη βελτίωση της ακρίβειας και της απόδοσης των αλγορίθμων μηχανικής μάθησης. Με βάση αυτό, οι επιστήμονες σχεδίασαν μια νέα μεθοδολογία, η οποία:

- Συνδυάζει τη στατιστική ανάλυση (συσχέτιση) και τη μετρική απόδοσης (ACC) για την επιλογή των βέλτιστων χαρακτηριστικών.
- Επικεντρώνεται στη βελτιστοποίηση του συνόλου χαρακτηριστικών, ώστε να διατηρηθεί η πληροφορική επάρκεια, ενώ παράλληλα μειώνεται το μέγεθος και η πολυπλοκότητα του συνόλου δεδομένων.

Η ερευνητική ομάδα θεώρησε ότι μια τέτοια προσέγγιση θα μπορούσε να ενισχύσει την ικανότητα ανίχνευσης επιθέσεων σε δίκτυα IoT, εξασφαλίζοντας παράλληλα την ταχύτητα και την αποδοτικότητα του συστήματος.

Η μεθοδολογία που χρησιμοποιήθηκε βασίστηκε σε δύο κύριες φάσεις:

1. Φάση 1: Επιλογή χαρακτηριστικών μέσω συσχέτισης (Correlation)
    - Οι επιστήμονες χρησιμοποίησαν τη μετρική συσχέτισης (Corr) για να αναγνωρίσουν χαρακτηριστικά που σχετίζονται στενά μεταξύ τους.
    - Αν η τιμή συσχέτισης μεταξύ χαρακτηριστικών υπερέβαινε ένα προκαθορισμένο κατώφλι, τότε τα χαρακτηριστικά αυτά επιλέγονταν για περαιτέρω επεξεργασία.
    - Αυτό το βήμα είχε ως στόχο τη μείωση του όγκου δεδομένων, διατηρώντας ωστόσο πληροφορίες κρίσιμες για την ανίχνευση ανωμαλιών.
2. Φάση 2: Επιλογή χαρακτηριστικών μέσω της μετρικής ACC
    - Στη δεύτερη φάση, εφαρμόστηκε η μετρική ACC, η οποία αξιολογεί την απόδοση ενός χαρακτηριστικού με βάση τη συνεισφορά του στην ακρίβεια ενός αλγορίθμου μηχανικής μάθησης.
    - Τα χαρακτηριστικά που είχαν χαμηλή τιμή ACC αφαιρούνταν, ενώ τα υπόλοιπα εξετάζονταν περαιτέρω με τη χρήση τεχνικών wrapper, προκειμένου να αποφασιστεί η τελική επιλογή τους.
3. Αξιολόγηση με χρήση του συνόλου δεδομένων Bot-IoT
    - Το σύνολο δεδομένων Bot-IoT περιλαμβάνει κανονική κίνηση και δεδομένα επιθέσεων από διάφορους τύπους (π.χ. UDPDoS, TCPDoS, Data Theft, Keylogging).
    - Εφαρμόστηκαν οι τέσσερις αλγόριθμοι μηχανικής μάθησης (C4.5, Random Forest, Naïve Bayes, SVM) για να αξιολογηθεί η απόδοση του Corracc.

Οι μέθοδοι τεχνητής νοημοσύνης που χρησιμοποιήθηκαν, είναι οι ακόλουθοι:

1. **Αλγόριθμοι μηχανικής μάθησης (ML)**:

- **C4.5 Decision Tree**: Χρησιμοποιήθηκε για να δημιουργηθεί ένα μοντέλο κατηγοριοποίησης με βάση τη δένδροειδή διάταξη αποφάσεων.
- **Random Forest**: Εξετάστηκε για την ικανότητά του να παρέχει υψηλή ακρίβεια μέσω συνδυασμού πολλαπλών δένδρων αποφάσεων.
- **Naïve Bayes**: Εφαρμόστηκε ως στατιστικό μοντέλο βασισμένο στη θεωρία πιθανοτήτων.
- **Support Vector Machine (SVM)**: Χρησιμοποιήθηκε για τη δημιουργία γραμμικών και μη γραμμικών μοντέλων κατηγοριοποίησης.

2. **Τεχνικές επιλογής χαρακτηριστικών**:
   - **Wrapper techniques**: Μεθοδολογία που συνδυάζει τη δοκιμή χαρακτηριστικών με τους αλγόριθμους ML για την επιλογή των βέλτιστων χαρακτηριστικών.
   - **Metrics-based selection (Corr και ACC)**: Μετρήσεις συσχέτισης και απόδοσης που υποστηρίζουν την αποτελεσματική επιλογή χαρακτηριστικών.

Οι επιστήμονες ανέπτυξαν μια αποτελεσματική μεθοδολογία επιλογής χαρακτηριστικών για την ανίχνευση ανωμαλιών και εισβολών σε δίκτυα IoT. Ο αλγόριθμος Corracc πέτυχε υψηλή ακρίβεια, ευαισθησία και ειδικότητα, αποδεικνύοντας την αξία του στη βελτιστοποίηση των μοντέλων μηχανικής μάθησης. Ιδιαίτερα, οι αλγόριθμοι C4.5 και Random Forest παρουσίασαν εξαιρετική απόδοση, καθιστώντας τους ιδανικούς για το συγκεκριμένο πρόβλημα.

Ορισμένοι ερευνητές έχουν εφαρμόσει τόσο τη δυαδική όσο και την πολυ-κατηγορική ταξινόμηση στις μελέτες τους. Στο πλαίσιο αυτό, έχουν αναπτύξει διάφορους ταξινομητές για την ανίχνευση εισβολών και έχουν εξετάσει προκλήσεις όπως η ανισοκατανομή των κατηγοριών, η τρισδιάστατη απεικόνιση δεδομένων και η διαδικασία εξαγωγής χαρακτηριστικών στα δεδομένα.

Οι επιστήμονες που διεξήγαγαν την έρευνα [184] είχαν ως βασικό στόχο την ανάπτυξη μιας ενοποιημένης προσέγγισης για την ανίχνευση εισβολών σε διαφορετικά περιβάλλοντα δικτύων, τόσο στα παραδοσιακά δίκτυα επιχειρήσεων (enterprise networks) όσο και στα ασύρματα δίκτυα (802.11 wireless networks). Αυτό που ήθελαν να επιτύχουν ήταν η δημιουργία ενός συστήματος που θα μπορούσε να συνδυάζει υψηλή ακρίβεια, γρήγορους χρόνους επεξεργασίας και χαμηλά ποσοστά ψευδών συναγερμών (False Positive Rate - FPR).

Η έρευνα βασίστηκε στην ιδέα ότι η ανίχνευση εισβολών αποτελεί μια πρόκληση που απαιτεί εξειδικευμένες προσεγγίσεις για κάθε τύπο δικτύου και κατηγορία επιθέσεων. Ωστόσο, οι επιστήμονες αναγνώρισαν την ανάγκη για ένα γενικευμένο μοντέλο που θα μπορεί να εφαρμοστεί σε διαφορετικά datasets και σενάρια δικτύωσης. Το σκεπτικό τους ήταν ότι η χρήση τεχνικών μηχανικής μάθησης (Machine Learning) και ειδικότερα της τεχνητής νοημοσύνης (AI) μπορεί να προσφέρει λύσεις με υψηλή ευελιξία και αποδοτικότητα.

Οι μέθοδοι που χρησιμοποιήθηκαν για την παραπάνω έρευνα είναι οι εξής:

1. **Υπο-Σύνολα Ταξινομητών (Sub-Ensembles):**
   Οι επιστήμονες υιοθέτησαν μια προσέγγιση βασισμένη σε σύνολα ταξινομητών (ensemble learning), όπου κάθε υπο-σύνολο αποτελούνταν από standard classifiers και boosted classifiers (π.χ. AdaBoost). Αυτό βοήθησε στην αντιμετώπιση της ποικιλομορφίας των επιθέσεων.

2. **Βαρύτητες (Weight Metrics):**
   Η χρήση βαρύνσεων όπως το True Positive Rate (TPR) και η ακρίβεια (accuracy) επέτρεψε τη βελτίωση της απόδοσης για συγκεκριμένες κατηγορίες δεδομένων. Για παράδειγμα, το TPR έδωσε μεγαλύτερη έμφαση στην ανίχνευση επιθέσεων με υψηλή σημασία, ενώ το accuracy χρησιμοποιήθηκε για γενική βελτιστοποίηση.

3. **Ενίσχυση (Boosting):**
   Εφάρμοσαν τεχνικές όπως το AdaBoost για την ενίσχυση της απόδοσης κάποιων υπο-ταξινομητών. Ωστόσο, διαπίστωσαν ότι η πλήρης ενίσχυση (full boosting) δεν απέδωσε τόσο καλά όσο η ημι-ενίσχυση (semi-boosting).

4. **Pruning**:
   Δοκιμάστηκε η μέθοδος pruning για την απομάκρυνση μη αποδοτικών υπο-ταξινομητών από το σύνολο, με στόχο τη μείωση της πολυπλοκότητας και την αύξηση της ακρίβειας.

5. **ROC και Σημαντικότητα (Statistical Significance Testing)**:
   Χρησιμοποίησαν το ROC AUC ως βασικό μέτρο απόδοσης, ενώ προχώρησαν σε στατιστικές δοκιμές για να εξετάσουν εάν οι διαφορές στην απόδοση μεταξύ των μοντέλων ήταν στατιστικά σημαντικές.

   Η μέθοδος που πρότειναν οι επιστήμονες ξεπέρασε τις υπάρχουσες μεθόδους, επιτυγχάνοντας χαμηλότερο FPR και καλύτερη ισορροπία μεταξύ ακρίβειας και χρόνου επεξεργασίας. Το μοντέλο έδειξε ότι μπορεί να γενικεύσει αποτελεσματικά τόσο για παραδοσιακά όσο και για ασύρματα δίκτυα, δημιουργώντας ένα ισχυρό θεμέλιο για μελλοντικές επεκτάσεις σε πιο εξειδικευμένες εφαρμογές, όπως τα SCADA συστήματα.

Οι επιστήμονες που ανέπτυξαν το **LIO-IDS** [185] είχαν ως στόχο να αντιμετωπίσουν τις προκλήσεις που σχετίζονται με τα παραδοσιακά Συστήματα Ανίχνευσης Δικτυακών Εισβολών (NIDS), τα οποία συχνά αδυνατούν να εντοπίσουν με ακρίβεια σπάνιες επιθέσεις και να διαχειριστούν ανισομερή δεδομένα. Το σκεπτικό τους βασίστηκε στην ανάγκη για πιο ευέλικτα και αποδοτικά εργαλεία που να μπορούν να προσαρμοστούν σε πραγματικά περιβάλλοντα με ποικιλία τύπων επιθέσεων και συνθηκών.

Το LIO-IDS στοχεύει στην ακριβή αναγνώριση εισβολών, διαχωρίζοντας με υψηλή ακρίβεια τη φυσιολογική από την κακόβουλη δικτυακή κίνηση. Παράλληλα, επιτυγχάνει πολυκατηγοριακή ταξινόμηση επιθέσεων, παρέχοντας στοχευμένες λύσεις ασφάλειας για κάθε τύπο επίθεσης. Η μέθοδος ενσωματώνει τεχνικές εξισορρόπησης δεδομένων, αντιμετωπίζοντας την ανισορροπία μεταξύ κατηγοριών και βελτιώνοντας την απόδοση στις σπάνιες επιθέσεις. Επιπλέον, εξασφαλίζει υψηλή αποδοτικότητα χρόνου, μειώνοντας τον χρόνο εκπαίδευσης και δοκιμής, καθιστώντας το σύστημα έτοιμο για εφαρμογή σε πραγματικά περιβάλλοντα.

Οι επιστήμονες βασίστηκαν σε τεχνικές τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης για την ανάπτυξη του συστήματος:

1. **Βαθιά Μάθηση (Deep Learning)**: Στο πρώτο επίπεδο, χρησιμοποιήθηκε ένας ταξινομητής LSTM (Long Short-Term Memory) για τον διαχωρισμό της φυσιολογικής από την κακόβουλη δικτυακή κίνηση. Τα δίκτυα LSTM είναι ιδιαίτερα αποτελεσματικά για τη διαχείριση ακολουθιακών δεδομένων, όπως η δικτυακή κίνηση.

2. **Μηχανική Μάθηση**: Στο δεύτερο επίπεδο, εφαρμόστηκαν οι μέθοδοι Random Forest και Bagging για την ταξινόμηση των κακόβουλων κινήσεων σε συγκεκριμένες κατηγορίες επιθέσεων. Αυτές οι μέθοδοι εξασφαλίζουν υψηλή ακρίβεια, αξιοποιώντας συνδυασμούς μοντέλων για καλύτερη γενίκευση.

3. **Τεχνικές Εξισορρόπησης Δεδομένων**: Χρησιμοποιήθηκαν οι **Random Oversampling**, **Borderline-SMOTE**, και **SVM-SMOTE** για την αντιμετώπιση της ανισορροπίας στα δεδομένα εκπαίδευσης. Αυτές οι τεχνικές διασφαλίζουν ότι τα σπάνια είδη επιθέσεων αντιπροσωπεύονται επαρκώς στο εκπαιδευτικό σύνολο.

4. **One-vs-One (OVO) Ταξινόμηση**: Για την ταξινόμηση πολλών κατηγοριών επιθέσεων, εφαρμόστηκε η τεχνική OVO, η οποία διαχωρίζει κάθε ζεύγος κατηγοριών και δημιουργεί πιο εξειδικευμένα μοντέλα για κάθε περίπτωση.

5. **Αξιολόγηση με Μετρικές**: Για την αξιολόγηση του συστήματος, υπολογίστηκαν μετρικές όπως η Ακρίβεια (Accuracy), η Ανάκληση (Recall), η Ακρίβεια (Precision), και η F1-score, καθώς και οι καμπύλες ROC και οι αντίστοιχες περιοχές (AUC). Αυτές οι μετρικές αξιολογούν την ικανότητα του συστήματος να εντοπίζει κακόβουλες δραστηριότητες με ακρίβεια και χαμηλό αριθμό ψευδών συναγερμών.

Το LIO-IDS αποτελεί ένα καινοτόμο σύστημα ανίχνευσης δικτυακών εισβολών που συνδυάζει τις δυνατότητες της βαθιάς μάθησης και της μηχανικής μάθησης για την επίτευξη υψηλής ακρίβειας και αποδοτικότητας. Η χρήση τεχνικών εξισορρόπησης δεδομένων και ο έξυπνος σχεδιασμός δύο επιπέδων καθιστούν το LIO-IDS ιδανικό για ανάπτυξη σε πραγματικά περιβάλλοντα, προσφέροντας αξιόπιστη προστασία απέναντι σε ποικιλία δικτυακών απειλών.

Οι επιστήμονες που εκπόνησαν την παρούσα έρευνα [186] επικεντρώθηκαν στην ανάπτυξη μεθόδων για την ανίχνευση κυβερνοεπιθέσεων σε βιομηχανικά συστήματα ελέγχου (ICS), τα οποία είναι κρίσιμα για τη λειτουργία πολλών υποδομών, όπως εργοστάσια παραγωγής ενέργειας και δίκτυα διανομής. Το κίνητρο για την εργασία τους προήλθε από την αυξανόμενη συχνότητα και πολυπλοκότητα των κυβερνοεπιθέσεων σε τέτοια συστήματα, καθώς και από την ανάγκη για λύσεις που να μπορούν να λειτουργούν σε πραγματικό χρόνο χωρίς να διακόπτεται η λειτουργία των συστημάτων.

Οι κύριοι στόχοι της έρευνας ήταν:

- Να αξιολογήσουν την αποτελεσματικότητα υπαρχόντων αλγορίθμων online μάθησης στην ανίχνευση κυβερνοεπιθέσεων.
- Να προτείνουν βελτιωμένες μεθόδους, όπως η ενσωμάτωση της cost-sensitive προσέγγισης, ώστε να βελτιωθεί η απόδοση στην αναγνώριση σπάνιων και σημαντικών επιθέσεων.
- Να συγκρίνουν τις επιδόσεις αυτών των μεθόδων σε δυαδικά και πολυκατηγορικά σύνολα δεδομένων, υπολογίζοντας μετρικές όπως η ευαισθησία, η εξειδίκευση και το συνολικό ποσοστό σφάλματος.

Για να επιτύχουν τους στόχους τους, οι ερευνητές χρησιμοποίησαν τις εξής μεθόδους τεχνητής νοημοσύνης και μηχανικής μάθησης:

1. **Online Learning Algorithms**: Οι αλγόριθμοι αυτοί επιτρέπουν την ενημέρωση των μοντέλων μάθησης σε πραγματικό χρόνο, χωρίς την ανάγκη πλήρους επανεκπαίδευσης. Ειδικότερα, αξιολόγησαν αλγορίθμους όπως:

- ALMA (Approximate Large Margin Algorithm)
- SCW (Soft Confidence-Weighted Learning)
- AROW (Adaptive Regularization of Weight Vectors)

2. **Cost-Sensitive Learning**: Εισήγαγαν την cost-sensitive προσέγγιση σε αλγορίθμους online μάθησης, προκειμένου να δώσουν μεγαλύτερη έμφαση στην ανίχνευση των σπάνιων θετικών δειγμάτων. Για παράδειγμα:
   - Δημιούργησαν έναν νέο αλγόριθμο, τον ARCSMC (Adaptive Regularized Cost-Sensitive Multiclass Classification), που εφαρμόζει cost-sensitive εκπαίδευση σε πολυκατηγορικά δεδομένα.
   - Ενσωμάτωσαν την cost-sensitive παραλλαγή στον AROW για τη βελτίωση της απόδοσης στις δυαδικές ταξινομήσεις.
3. **Αξιολόγηση Αλγορίθμων**: Εφάρμοσαν πειραματικές δοκιμές σε δυαδικά και πολυκατηγορικά σύνολα δεδομένων, υπολογίζοντας τις βασικές μετρικές (ευαισθησία, εξειδίκευση, ποσοστό σφάλματος). Επίσης, αξιολόγησαν την αποδοτικότητα των μεθόδων σε όρους χρόνου επεξεργασίας δεδομένων.
4. **Cost-Sensitive Metrics**: Για την πολυκατηγορική ταξινόμηση, τροποποίησαν τις μετρικές ώστε να υπολογίζονται για κάθε τάξη ξεχωριστά (ως θετική τάξη) και στη συνέχεια συνυπολόγισαν τις σταθμισμένες τιμές τους.

Η έρευνα κατέδειξε ότι οι cost-sensitive αλγόριθμοι online μάθησης υπερέχουν στην ανίχνευση σπάνιων επιθέσεων σε βιομηχανικά συστήματα. Ειδικά ο προτεινόμενος αλγόριθμος ARCSMC παρουσίασε εξαιρετική ισορροπία μεταξύ ευαισθησίας και εξειδίκευσης, καθιστώντας τον κατάλληλο για πρακτική εφαρμογή.

Οι επιστήμονες που πραγματοποίησαν την παρούσα έρευνα [188] είχαν ως βασικό στόχο την ανάπτυξη ενός καινοτόμου συστήματος ανίχνευσης εισβολών (IDS), ικανού να διαχειρίζεται σύγχρονα και πολύπλοκα δεδομένα κυβερνοασφάλειας. Το σκεπτικό τους βασίστηκε στη διαπίστωση ότι τα υπάρχοντα συστήματα ανίχνευσης εισβολών, αν και αποτελεσματικά, συχνά παρουσιάζουν περιορισμούς στην ακρίβεια, στη διαχείριση πολλαπλών τύπων επιθέσεων και στη δυνατότητα λειτουργίας σε πραγματικό χρόνο.

Οι επιστήμονες εντόπισαν ότι οι παραδοσιακές μέθοδοι μηχανικής μάθησης, όπως οι γραμμικοί και μη γραμμικοί υποστηρικτικοί διανυσματικοί ταξινομητές (L-SVM, Q-SVM) ή οι κλασικοί μέθοδοι ανάλυσης δεδομένων (LDA, QDA), συχνά αποτυγχάνουν στην ανίχνευση πολύπλοκων επιθέσεων όπως οι Remote-to-Local (R2L). Παράλληλα, οι βαθιές αρχιτεκτονικές μάθησης, όπως το LSTM, έχουν υψηλές υπολογιστικές απαιτήσεις και περιορισμένη αποτελεσματικότητα σε περιπτώσεις ανίχνευσης πολλαπλών κατηγοριών επιθέσεων.

Με βάση αυτά, οι επιστήμονες πρότειναν μια προσέγγιση που συνδυάζει τη στατιστική ανάλυση για την επιλογή των πιο σημαντικών χαρακτηριστικών με έναν αυτόματο κωδικοποιητή (AE), μια μέθοδο βαθιάς μάθησης, που εστιάζει στην αποδοτική αναπαράσταση και ανάλυση των δεδομένων.

1. **Στατιστική Ανάλυση**:
   Χρησιμοποιήθηκε για την εξαγωγή των πιο σημαντικών χαρακτηριστικών του NSL-KDD dataset, ώστε να μειωθεί η διάσταση των δεδομένων και να βελτιωθεί η απόδοση των μοντέλων.

2. **Αυτόματος Κωδικοποιητής (AE)**:
   Ο AE αναπτύχθηκε ως η κύρια μέθοδος βαθιάς μάθησης. Πρόκειται για μια αρχιτεκτονική με ένα κρυφό επίπεδο και 50 μονάδες, που επιτρέπει την αποδοτική αναπαράσταση των δεδομένων και την ανίχνευση εισβολών με υψηλή ακρίβεια.

3. **Σύγκριση με Παραδοσιακούς Αλγορίθμους**:
   Οι κλασικοί αλγόριθμοι μηχανικής μάθησης, όπως οι MLP, L-SVM, Q-SVM, LDA και QDA, χρησιμοποιήθηκαν για σύγκριση επιδόσεων. Παράλληλα, το LSTM αξιοποιήθηκε ως σύγχρονη μέθοδος βαθιάς μάθησης.

4. **Μέτρηση Απόδοσης**:
   Χρησιμοποιήθηκαν μετρικές όπως η ακρίβεια, η F1 score, η καμπύλη ROC και το AUC, ώστε να αξιολογηθεί η αποτελεσματικότητα των μεθόδων τόσο για δυαδική όσο και για πολυταξινόμηση.
5. **Σύγκριση με Άλλες Έρευνες**:
   Το προτεινόμενο AE50 συγκρίθηκε με σύγχρονες μεθόδους της βιβλιογραφίας, επιδεικνύοντας υπεροχή σε ακρίβεια (87%) και χρόνο εκπαίδευσης (22.53s).

Με την ενσωμάτωση της στατιστικής ανάλυσης και τη σύγκριση με παραδοσιακές και σύγχρονες μεθόδους, η έρευνα ανέδειξε τη σημαντική δυνατότητα των βαθιών αρχιτεκτονικών μάθησης για ανίχνευση εισβολών σε πραγματικό χρόνο.

Η ανίχνευση εισβολών στα δίκτυα αποτελεί έναν από τους πιο κρίσιμους τομείς της κυβερνοασφάλειας, καθώς οι επιθέσεις γίνονται ολοένα και πιο εξελιγμένες και ποικίλες. Οι επιστήμονες που διεξήγαγαν την παρούσα έρευνα [189] επιδίωξαν να αναπτύξουν μια αποδοτική μέθοδο ανίχνευσης εισβολών, η οποία να είναι ικανή να εντοπίζει τόσο γνωστές όσο και άγνωστες επιθέσεις με υψηλή ακρίβεια και αξιοπιστία. Το βασικό σκεπτικό τους ήταν ότι ένα μοντέλο τεχνητής νοημοσύνης μπορεί να μάθει και να αναλύει τα δεδομένα από διαφορετικές οπτικές γωνίες, αξιοποιώντας διαφορετικούς τύπους χαρακτηριστικών και υπο-μοντέλων.

Οι στόχοι της έρευνας επικεντρώθηκαν στη βελτίωση της ακρίβειας ανίχνευσης εισβολών, μειώνοντας τα ψευδή θετικά αποτελέσματα και ελαχιστοποιώντας τον αριθμό των ψευδών συναγερμών. Επιπλέον, οι επιστήμονες στόχευσαν στην ανάπτυξη ενός συστήματος που να είναι ανθεκτικό σε ανισόρροπα δεδομένα, αντιμετωπίζοντας τις προκλήσεις που προκύπτουν από την άνιση κατανομή των κατηγοριών δεδομένων στις κυβερνοεπιθέσεις. Τέλος, το μοντέλο έπρεπε να είναι ευέλικτο και ικανό να προσαρμόζεται σε διάφορα σύνολα δεδομένων και τύπους εισβολών, επιτυγχάνοντας αυτοματοποίηση και προσαρμοστικότητα.

Οι επιστήμονες υιοθέτησαν μια πολυεπίπεδη προσέγγιση που συνδυάζει διαφορετικές τεχνικές μηχανικής μάθησης και βαθιάς μάθησης. Η μέθοδος βασίζεται σε ένα σύστημα πολλαπλών υπο-μοντέλων, το οποίο αξιοποιεί διαφορετικές τεχνικές AI για την εκμάθηση χαρακτηριστικών και την ταξινόμηση.

1. **Autoencoders**: Χρησιμοποιήθηκαν για την εκμάθηση χαρακτηριστικών από τα δεδομένα. Οι Autoencoders είναι μοντέλα βαθιάς μάθησης που εξειδικεύονται στη μείωση διαστάσεων και στην ανάλυση δομών στα δεδομένα. Βοήθησαν στην εξαγωγή σημαντικών πληροφοριών από μεγάλες ποσότητες δεδομένων.

2. **Focal Loss**: Αυτή η τεχνική χρησιμοποιήθηκε ως συνάρτηση κόστους κατά την εκπαίδευση του μοντέλου, ώστε να δοθεί μεγαλύτερη έμφαση στις πιο δύσκολες ταξινομήσεις. Με αυτόν τον τρόπο μειώθηκε το πρόβλημα που προκαλείται από ανισόρροπα δεδομένα.

3. **Clustering**: Εφαρμόστηκε ομαδοποίηση χαρακτηριστικών για την αναγνώριση πρότυπων στις επιθέσεις. Η τεχνική αυτή επέτρεψε την καλύτερη κατανόηση των δεδομένων και τη βελτίωση της ταξινόμησης.

4. **Συνδυασμός υπο-μοντέλων**: Η καινοτομία της προσέγγισης έγκειται στη χρήση πολλαπλών υπο-μοντέλων, καθένα από τα οποία μαθαίνει διαφορετικές πτυχές των δεδομένων. Ο συνδυασμός των αποτελεσμάτων αυτών των υπο-μοντέλων οδήγησε σε υψηλότερη ακρίβεια.

Η έρευνα απέδειξε ότι η ανίχνευση εισβολών μπορεί να ενισχυθεί σημαντικά μέσα από τη συνδυαστική χρήση μοντέλων τεχνητής νοημοσύνης, επιδεικνύοντας υψηλή αποτελεσματικότητα σε διαφορετικά και ανισόρροπα σύνολα δεδομένων. Αν και η μελέτη αναγνωρίζει περιορισμούς, ανοίγει τον δρόμο για την ανάπτυξη πιο ευέλικτων και αξιόπιστων συστημάτων ασφαλείας στο μέλλον.

Η ερευνητική ομάδα [190] επιδίωξε να αντιμετωπίσει τις προκλήσεις που αφορούν την ανίχνευση κυβερνοεπιθέσεων σε δικτυακά συστήματα, αναπτύσσοντας ένα προηγμένο πλαίσιο επίγνωσης κατάστασης βασισμένο σε μηχανική μάθηση (AI). Το πλαίσιο αυτό, γνωστό ως **ML-IDS** (Machine Learning Intrusion Detection System), στοχεύει στη βελτίωση της ανίχνευσης και προστασίας έναντι κυβερνοαπειλών μέσω της ανάλυσης δικτυακής κίνησης και ευπαθειών.

Η παραδοσιακή προσέγγιση ανίχνευσης επιθέσεων βασίζεται σε στατικά συστήματα που συγκρίνουν μοτίβα κίνησης δικτύου με γνωστές υπογραφές κακόβουλων ενεργειών. Ωστόσο, αυτή η μέθοδος παρουσιάζει περιορισμούς, όπως η αδυναμία εντοπισμού νέων ή εξελιγμένων επιθέσεων. Οι επιστήμονες, αναγνωρίζοντας αυτά τα κενά, αποφάσισαν να εστιάσουν σε τεχνικές μηχανικής μάθησης, οι οποίες μπορούν να ανιχνεύουν ανωμαλίες και να προβλέπουν επιθέσεις βάσει δυναμικής ανάλυσης μεγάλων συνόλων δεδομένων.

Οι κύριοι στόχοι της έρευνας επικεντρώνονται στη βελτίωση της ανίχνευσης κυβερνοεπιθέσεων μέσω της ανάπτυξης ενός συστήματος που θα μπορεί να ανιχνεύει επιθέσεις με υψηλή ακρίβεια, ακόμη και όταν αυτές παρουσιάζουν χαμηλό αποτύπωμα στο δίκτυο. Για τον σκοπό αυτό, χρησιμοποιήθηκε ένα εμπλουτισμένο σύνολο δεδομένων (enhanced dataset), το οποίο περιλαμβάνει όχι μόνο δεδομένα δικτύου, αλλά και πληροφορίες για τις ευπάθειες λογισμικού (CVE). Η έρευνα περιλαμβάνει επίσης τη σύγκριση της απόδοσης ενός "απλού" μοντέλου με ένα "ενισχυμένο" μοντέλο, αξιολογώντας την αποτελεσματικότητα της ανίχνευσης για διάφορους τύπους επιθέσεων. Επιπλέον, η ομάδα επικεντρώθηκε στην ανάπτυξη μεθόδων που μειώνουν τα ψευδώς θετικά και αρνητικά αποτελέσματα, προσφέροντας έτσι μια πιο αξιόπιστη διαδικασία ανίχνευσης και επιτρέποντας την κλιμάκωση του συστήματος για μεγαλύτερα δίκτυα.

Για να επιτύχουν τους στόχους τους, οι ερευνητές χρησιμοποίησαν τα εξής:

1. **Νευρωνικά Δίκτυα (Neural Networks)**:

   Η βασική τεχνολογία που εφαρμόστηκε ήταν ένα νευρωνικό δίκτυο για την ανάλυση της κίνησης στο δίκτυο. Το δίκτυο εκπαιδεύτηκε ώστε να αναγνωρίζει πέντε συγκεκριμένους τύπους επιθέσεων (π.χ. Remote File Inclusion, SSH Brute Force, Slow Loris DoS).

2. **Προσαρμοσμένα Datasets**:

   Δημιουργήθηκε ένα εμπλουτισμένο dataset που περιλαμβάνει δικτυακά δεδομένα (NetFlow) σε συνδυασμό με πληροφορίες για ευπάθειες συστημάτων (CVEs). Αυτό επέτρεψε την εκπαίδευση του συστήματος να αναγνωρίζει πρότυπα κακόβουλης δραστηριότητας που σχετίζονται με συγκεκριμένες ευπάθειες.

3. **Επεξεργασία Δεδομένων σε Πραγματικό Χρόνο**:

   Ο μηχανισμός ανίχνευσης συλλέγει και επεξεργάζεται την κίνηση δικτύου σε πραγματικό χρόνο. Χρησιμοποιώντας βιβλιοθήκες όπως το NumPy, η κίνηση μετατρέπεται σε μορφή κατάλληλη για το νευρωνικό δίκτυο.

4. **Συγκριτική Ανάλυση Μοντέλων**:

   Διεξήχθη σύγκριση μεταξύ του "απλού" και του "ενισχυμένου" μοντέλου, εξετάζοντας την ακρίβεια ανίχνευσης για κάθε τύπο επίθεσης.

5. **Οπτικοποίηση Αποτελεσμάτων**:

   Τα αποτελέσματα παρουσιάστηκαν μέσω πινάκων και γραφημάτων, δείχνοντας ότι το ενισχυμένο μοντέλο ξεπέρασε το απλό σε όλες τις περιπτώσεις, με μέση ακρίβεια ~87% σε σύγκριση με ~83% του απλού μοντέλου.

Η έρευνα κατέδειξε ότι η ενσωμάτωση προηγμένων μεθόδων μηχανικής μάθησης και προσαρμοσμένων datasets μπορεί να βελτιώσει σημαντικά την ανίχνευση κυβερνοεπιθέσεων. Παρά τα ενθαρρυντικά αποτελέσματα, οι επιστήμονες αναγνωρίζουν την ανάγκη περαιτέρω δοκιμών με περισσότερους τύπους επιθέσεων, βελτίωση της διαδικασίας δημιουργίας datasets, καθώς και δοκιμή νέων αλγορίθμων για μεγαλύτερη ακρίβεια και κλιμάκωση του συστήματος.

Η έρευνα [191] επικεντρώνεται στην ανάπτυξη ενός πολυεπίπεδου συστήματος ανίχνευσης κυβερνοεπιθέσεων για τα Βιομηχανικά Συστήματα Ελέγχου (Industrial Control Systems - ICS). Οι επιστήμονες ανέπτυξαν μια καινοτόμο προσέγγιση που συνδυάζει τεχνικές μηχανικής μάθησης και παρακολούθησης δεδομένων διαδικασίας, με στόχο την αντιμετώπιση των ολοένα και πιο εξελιγμένων απειλών που θέτουν σε κίνδυνο την ασφάλεια των βιομηχανικών υποδομών.

Οι ερευνητές αξιοποίησαν τέσσερις διαφορετικούς αλγορίθμους για την ταξινόμηση των δεδομένων:

1. **K-Nearest Neighbors (KNN)**: Χρησιμοποιήθηκε για την ανάλυση της ομοιότητας των δεδομένων και την ανίχνευση ανωμαλιών. Παρουσίασε την υψηλότερη ακρίβεια στην ανίχνευση επιθέσεων.
2. **Αποφασιστικό Δέντρο (Decision Tree - DT)**: Παρά την απλότητα και το χαμηλό υπολογιστικό του κόστος, είχε τη χαμηλότερη απόδοση.
3. **Bagging**: Μια τεχνική που βελτιώνει την ακρίβεια των αποφασιστικών δέντρων δημιουργώντας πολλαπλά μοντέλα.
4. **Random Forest (RF)**: Χρησιμοποίησε πολλά αποφασιστικά δέντρα για καλύτερη απόδοση, παρουσιάζοντας μηδενικά ψευδώς θετικά.

Επιπλέον, χρησιμοποίησαν ένα μοντέλο παρακολούθησης δεδομένων **(AAKR)**. Το μοντέλο αυτό αξιοποιεί την ανάλυση υπολειμμάτων των αισθητήρων, συγκρίνοντας τις προβλεπόμενες τιμές με τις πραγματικές. Σε περίπτωση που το υπολειμματικό σφάλμα υπερβεί ένα προκαθορισμένο όριο, ανιχνεύεται πιθανή κυβερνοεπίθεση.

Τέλος, το σύστημα συνδύασε δεδομένα δικτύου, δεδομένα αισθητήρων και ελέγχους φυσικών διαδικασιών. Αυτή η ολοκληρωμένη προσέγγιση εξασφαλίζει ανίχνευση ανωμαλιών που δεν θα μπορούσαν να εντοπιστούν μόνο με μία πηγή δεδομένων.

Η πολυεπίπεδη προσέγγιση και οι σύγχρονες τεχνικές τεχνητής νοημοσύνης που χρησιμοποίησαν οι ερευνητές πέτυχαν να παρέχουν ένα ανθεκτικό και αξιόπιστο σύστημα ανίχνευσης κυβερνοεπιθέσεων. Το σύστημα απέδειξε την ικανότητά του να εντοπίζει έγκαιρα επιθέσεις, προστατεύοντας κρίσιμες υποδομές από πιθανές καταστροφές.

4.2 Σύνοψη εργαλείων/μεθόδων AI

Ο παρακάτω πίνακας συνοψίζει τα εργαλεία, τις μεθόδους και τους αλγορίθμους τεχνητής νοημοσύνης (AI) που χρησιμοποιούνται στη λειτουργία "Εντοπισμός" (Detect) του NIST Cybersecurity Framework, η οποία επικεντρώνεται στην έγκαιρη ανίχνευση περιστατικών κυβερνοασφάλειας. Περιλαμβάνει τεχνικές όπως βαθιά νευρωνικά δίκτυα (DNN, LSTM, RNN – CNN, ResNet 101), αλγόριθμοι μηχανικής μάθησης (SVM, Random Forest, XGBoost, Naïve Bayes), γενετικοί αλγόριθμοι και GANs (IGAN – IDS, EsPADA, DnRANN), καθώς και μεθόδους κανονικοποίησης και ταξινόμησης (One-Hot Encoding, Min-Max Scaling, PCA, DBSCAN). Αυτά τα εργαλεία επιτρέπουν στους οργανισμούς να ανιχνεύουν γρήγορα επιθέσεις, να περιορίζουν τις επιπτώσεις και να ενισχύουν την αποτελεσματικότητα των συστημάτων κυβερνοασφάλειας.

2.3.1 Σύστημα Ανίχνευσης Εισβολής
(1) SOMs (2) LSTM (3) RNN – CNN (4) Αρχιτεκτονική SGDIDS (5) SVM, AIS (6) DBN, MSC, Blockchain, CNN – ResNet 101 (7) DT (Decision Tree), Logistic Regression, MLP, Gaussian – Naïve Bayes, Random Forest, AdaBoost, Gradient Boosting (8) Μετρικές αποδοτικότητας: F – Score Efficiency, ROC – AUC Efficiency, Explained Variance Efficiency, Recall, Precision (9) 2 κύρια συστήματα: Edge – ENClf, Cloud – ENClf , 1D CAE, OCSVM (10) One – Hot Encoding, Min – Max Scaling (11) CAE ταξινομητής (12) DNN, NLP (13) Πλαίσιο FAIXID, μέθοδος EsPADA, DnRANN (14) IGAN – IDS (15) GAN (16) Bloom – filter, Kohonen Map, SMOTE, Random Under – Sampling, PCA, k – fold cross – validation (17) KNN, XGBoost , NB, BN, Bagging Trees, k – Means, DBSCAN, EM (18) C4.5 Decision Tree, Corr, ACC (19) LIO – IDS (20) Ταξινόμηση OVO (21) ALMA, SCW, AROW (22) AAKR

Πίνακας 4 : Εργαλεία και αλγόριθμοι τεχνητής νοημοσύνης για τη λειτουργία «Εντοπισμός» (Detect) του NIST Cybersecurity Framework, συμβάλλοντας στην έγκαιρη ανίχνευση και αντιμετώπιση κυβερνοαπειλών.

5. Ανταπόκριση

*Η λειτουργία **«Ανταπόκριση»** (**Respond)** του πλαισίου NIST αφορά την ανάπτυξη και εφαρμογή κατάλληλων δραστηριοτήτων για την αντιμετώπιση ενός ανιχνευμένου περιστατικού κυβερνοασφάλειας. Ο στόχος της λειτουργίας είναι να περιορίσει τον αντίκτυπο ενός περιστατικού και να βοηθήσει στην άμεση αντιμετώπισή του. Η υιοθέτηση αυτής της λειτουργίας διασφαλίζει ότι το πρόγραμμα κυβερνοασφάλειας παραμένει αποτελεσματικό και συνεχώς βελτιώνεται, ενισχύοντας την ανθεκτικότητα του οργανισμού απέναντι σε κυβερνοεπιθέσεις.*

5.1 Διαχείριση Δυναμικών Περιστατικών

Η διαχείριση δυναμικών περιστατικών (Dynamic Case Management) αξιοποιεί δεδομένα από προηγούμενες παραβιάσεις ασφαλείας για να καταγράφει διάφορα σενάρια επιθέσεων και να προτείνει ενέργειες αντιμετώπισης πριν εκδηλωθεί κάποιο περιστατικό. Αυτό συμβάλλει στην

καλύτερη προετοιμασία για την αντιμετώπιση συγκεκριμένων τύπων παραβιάσεων και στην οργανωμένη αποθήκευση γνώσης μετά το πέρας ενός περιστατικού. Η έρευνα στον τομέα αυτό εστιάζει στη δημιουργία αυτοματοποιημένων συστάσεων για απόκριση, βασιζόμενη στην ταύτιση ενός νέου περιστατικού με παρόμοια περιστατικά που έχουν καταγραφεί σε σύστημα διαχείρισης γνώσης. Με αυτόν τον τρόπο, το σύστημα επικαιροποιείται και εμπλουτίζεται μετά από κάθε περιστατικό.

Οι επιστήμονες που διεξήγαγαν αυτήν την έρευνα [192] επικεντρώθηκαν στη δημιουργία ενός πλαισίου που θα βελτίωνε την ασφάλεια σε εταιρικά δίκτυα και συστήματα μέσω της χρήσης μεθόδων Τεχνητής Νοημοσύνης (AI) και συγκεκριμένα της Μεθόδου Βασισμένης σε Περιπτώσεις (Case-Based Reasoning - CBR). Το σκεπτικό τους βασίστηκε στην ανάγκη για ταχύτερη και πιο αποτελεσματική απόκριση σε επιθέσεις, ιδιαίτερα σε περιπτώσεις όπου οι απειλές είναι επαναλαμβανόμενες ή παραλλαγές παλαιότερων επιθέσεων.

Οι κύριοι στόχοι της έρευνας επικεντρώθηκαν στην ενίσχυση της ασφάλειας των συστημάτων μέσω της ταχείας ανίχνευσης ανωμαλιών και της αποτελεσματικής αντίδρασης σε επιθέσεις. Συγκεκριμένα, η έρευνα στόχευε στην ταχύτερη ανίχνευση ύποπτων δραστηριοτήτων μέσω ενός κεντρικού συστήματος ανάλυσης καταγραφών (logs), μειώνοντας έτσι τον χρόνο απόκρισης. Επίσης, επιδιώχθηκε η μείωση των ψευδών θετικών, ώστε οι ειδοποιήσεις να είναι ακριβείς και να συνδέονται μόνο με πραγματικές απειλές. Παράλληλα, το πλαίσιο που αναπτύχθηκε προέβλεπε την προστασία από επιθέσεις μηδενικής ημέρας, χρησιμοποιώντας προληπτικές άμυνες για την αποτελεσματική αντιμετώπιση νέων και άγνωστων απειλών.

Οι επιστήμονες χρησιμοποίησαν τη Μέθοδο Βασισμένης σε Περιπτώσεις **(CBR)**, που επιτρέπει στο σύστημα να βρίσκει ομοιότητες μεταξύ νέων και παλαιών επιθέσεων, αξιοποιώντας τη γνώση από προηγούμενα περιστατικά. Οι μέθοδοι που ανέπτυξαν περιλαμβάνουν:

1. **Προσδιορισμός παραγόντων ομοιότητας**: Ορισμένοι παράγοντες, όπως το “ποιος” (π.χ. ταυτότητα επιτιθέμενου), το “πώς” (υπογραφή επίθεσης), το “πού” (σημείο επίθεσης) και το “πότε” (χρονική στιγμή), χρησιμοποιούνται για την αξιολόγηση της ομοιότητας μεταξύ περιστατικών.

2. **Υπολογισμός συνολικής ομοιότητας**: Χρησιμοποιούν μια συνάρτηση που λαμβάνει υπόψη τη βαθμονόμηση ομοιότητας σε διαφορετικές κατηγορίες, όπως κατηγορία ευπάθειας, λειτουργικό σύστημα, θύρα δικτύου και παραλλαγή υπογραφής επίθεσης.

3. **Ανάλυση καταγραφών**: Το σύστημα ανιχνεύει ύποπτες δραστηριότητες αναλύοντας καταγραφές (logs) από διάφορα συστήματα, όπως καταγραφές εντολών “su” για απόπειρες κλοπής δικαιωμάτων root. Η μέθοδος συνδυάζει παρατήρηση γεγονότων και στατιστική ανάλυση για την ανίχνευση ανωμαλιών.

4. **Δημιουργία «μαύρης λίστας»**: Μέσα από την ανάλυση καταγραφών, δημιουργείται μια βάση δεδομένων με IP διευθύνσεις επιτιθέμενων και λογαριασμούς που συχνά χρησιμοποιούνται σε επιθέσεις.

5. **Αυτοματοποιημένη απόκριση**: Το σύστημα παρέχει αυτόματες ειδοποιήσεις και συντονίζει την αντίδραση μέσω ενός κεντρικού συστήματος υποστήριξης αποφάσεων (DSS).

Το προτεινόμενο πλαίσιο καταφέρνει να εντοπίσει γρήγορα και με ακρίβεια ανωμαλίες, μειώνοντας το χρόνο απόκρισης και την εξάρτηση από ανθρώπινους εμπειρογνώμονες. Παράλληλα,

διευκολύνει την προληπτική άμυνα απέναντι σε επιθέσεις μηδενικής ημέρας και προσφέρει ένα αυτοματοποιημένο σύστημα αναφορών και διαχείρισης περιστατικών, το οποίο μπορεί να προσαρμοστεί σε εταιρικά περιβάλλοντα με διαφορετικές ανάγκες.

Οι επιστήμονες που ασχολήθηκαν με τη διαχείριση περιστατικών ασφάλειας δικτύων [193] βασίστηκαν στην ανάγκη γρήγορης και αποτελεσματικής αντίδρασης σε νέα περιστατικά ασφαλείας. Το βασικό τους σκεπτικό ήταν ότι η γνώση που έχει συγκεντρωθεί από προηγούμενα περιστατικά μπορεί να αξιοποιηθεί για να αντιμετωπιστούν με επιτυχία νέα περιστατικά, μειώνοντας τον χρόνο αντίδρασης και την πιθανότητα λαθών.

Η ασφάλεια δικτύων αποτελεί έναν από τους πιο κρίσιμους τομείς στον σύγχρονο ψηφιακό κόσμο, όπου οι κυβερνοεπιθέσεις εξελίσσονται συνεχώς. Οι επιστήμονες συνειδητοποίησαν ότι τα δεδομένα από παλαιότερα περιστατικά ασφάλειας δικτύων μπορούν να αποτελέσουν πολύτιμη πηγή πληροφοριών για την κατανόηση και αντιμετώπιση νέων επιθέσεων. Το κύριο πρόβλημα που ήθελαν να επιλύσουν ήταν η αποτελεσματική αναζήτηση και ανάκτηση σχετικών περιπτώσεων από μια βάση δεδομένων περιστατικών, ώστε να προσφέρουν στοχευμένες λύσεις για κάθε νέο πρόβλημα.

Οι επιστήμονες υιοθέτησαν προηγμένες μεθόδους Τεχνητής Νοημοσύνης (AI) είναι οι ακόλουθες:

1. **Χρήση Λογικής Περιγραφής (Description Logic)**

   Οι περιπτώσεις περιστατικών αναπαραστάθηκαν με τη βοήθεια της λογικής ALCO(D), η οποία παρέχει ισχυρά εργαλεία περιγραφής. Η συγκεκριμένη μέθοδος επιτρέπει τη σαφή και δομημένη παρουσίαση των χαρακτηριστικών κάθε περιστατικού, όπως ο τύπος του (π.χ. ιός, DDoS), ο χρόνος που συνέβη και η έκταση των επιπτώσεων.

2. **Υπολογισμός Ομοιότητας μέσω Αλγορίθμων**

   Ο πυρήνας της προσέγγισης ήταν ο υπολογισμός της ομοιότητας μεταξύ περιστατικών. Οι ερευνητές ανέπτυξαν έναν αλγόριθμο βασισμένο στους τελεστές εξειδίκευσης και τα γραφήματα εξειδίκευσης. Αυτοί οι αλγόριθμοι μετρούν πόσο κοντά είναι δύο περιστατικά, λαμβάνοντας υπόψη κοινά χαρακτηριστικά και διαφορές.

3. **Ανάπτυξη Γραφημάτων Εξειδίκευσης (Refinement Graphs)**

   Τα γραφήματα εξειδίκευσης επιτρέπουν την αναπαράσταση των περιστατικών με διαφορετικά επίπεδα γενικότητας ή εξειδίκευσης. Ένας τελεστής "προς τα κάτω" (down refinement) δημιουργεί πιο ειδικά περιστατικά, ενώ ένας τελεστής "προς τα πάνω" (up refinement) δημιουργεί πιο γενικά.

4. **Μέτρηση Σχετικής Ομοιότητας**

   Για τη μέτρηση της ομοιότητας, οι επιστήμονες χρησιμοποίησαν τον κοινό υποβιβαστή (Least Common Subsumer - LCS), δηλαδή την πιο εξειδικευμένη έννοια που περιλαμβάνει κοινά στοιχεία μεταξύ δύο περιστατικών. Η ομοιότητα υπολογίστηκε ως το ποσοστό των κοινών πληροφοριών σε σχέση με το σύνολο των πληροφοριών και των δύο περιστατικών.

5. **Πειραματική Αξιολόγηση**

   Η μέθοδος δοκιμάστηκε σε πραγματικά δεδομένα από περισσότερα από 20 περιστατικά που συλλέχθηκαν τα τελευταία τρία χρόνια. Η πειραματική ανάλυση έδειξε ότι το σύστημα μπορούσε

να διαφοροποιήσει αποτελεσματικά περιστατικά όπως οι ιοί, το malware και οι επιθέσεις DDoS, προσφέροντας στοχευμένες λύσεις.

Η έρευνα κατέδειξε ότι η ενσωμάτωση προηγμένων τεχνικών AI μπορεί να βελτιώσει σημαντικά τη διαχείριση περιστατικών ασφάλειας δικτύων. Με την αποτελεσματική αναζήτηση παρόμοιων περιστατικών, οι οργανισμοί μπορούν να μειώσουν τον χρόνο αντίδρασης και να εφαρμόσουν βέλτιστες πρακτικές για την επίλυση των προβλημάτων.

Η έρευνα [194] που παρουσιάζεται έχει ως στόχο την ανάπτυξη ενός αυτόνομου συστήματος για την επιλογή κατάλληλων «playbooks» σε περιστατικά κυβερνοασφάλειας. Πρόκειται για μια καινοτόμο προσέγγιση που επιχειρεί να ξεπεράσει τους περιορισμούς των παραδοσιακών μεθόδων, οι οποίες βασίζονται σε κανόνες, ανθρώπινη γνώση ή επαναλαμβανόμενες επεμβάσεις. Το σκεπτικό των επιστημόνων ήταν να αξιοποιήσουν τη δύναμη της τεχνητής νοημοσύνης (AI) για να δημιουργήσουν αυτό το σύστημα που μαθαίνει από τα υπάρχοντα δεδομένα, αυτόματα αναγνωρίζει μοτίβα και προτείνει τις καταλληλότερες δράσεις, χωρίς την ανάγκη συνεχούς ανθρώπινης παρέμβασης.

Η μελέτη είχε ως κύριους στόχους την αυτοματοποίηση της διαδικασίας επιλογής «playbooks» ώστε να μειωθεί η ανάγκη για ανθρώπινη συμμετοχή, την εξαγωγή γνώσης από προηγούμενα περιστατικά μέσω της ανάλυσης ιστορικών δεδομένων για την καλύτερη κατανόησή τους, την ανάπτυξη ενός ευέλικτου και επεκτάσιμου συστήματος που μπορεί να προσαρμόζεται σε νέα δεδομένα και τύπους περιστατικών χωρίς επανεκπαίδευση, καθώς και τη βελτίωση της ακρίβειας και της αποτελεσματικότητας συγκριτικά με τις παραδοσιακές μεθόδους.

Οι επιστήμονες χρησιμοποίησαν μια συνδυαστική προσέγγιση με τεχνικές μηχανικής μάθησης και βαθιάς μάθησης για να πετύχουν τους στόχους τους. Οι βασικές μέθοδοι περιλαμβάνουν:

1. **Εκμάθηση Μετρικών (Metric Learning)**

   Η εκμάθηση μετρικών επιτρέπει στο σύστημα να κατανοεί τη «σχετικότητα» μεταξύ διαφορετικών περιστατικών. Με τη χρήση ενός νευρωνικού δικτύου, τα περιστατικά χαρτογραφήθηκαν σε έναν διανυσματικό χώρο, όπου περιστατικά με παρόμοια χαρακτηριστικά (και αντίστοιχα «playbooks») βρίσκονται κοντά μεταξύ τους. Αυτό επέτρεψε τη δημιουργία ενός συστήματος που μαθαίνει να αναγνωρίζει τη δομική ομοιότητα των περιστατικών.

2. **Πολυ-Ετικετική Ταξινόμηση (Multi-Label Classification)**

   Επειδή ένα περιστατικό μπορεί να σχετίζεται με περισσότερα από ένα «playbooks», χρησιμοποιήθηκε η μέθοδος της πολυ-ετικετικής ταξινόμησης. Με αυτόν τον τρόπο, το σύστημα είναι σε θέση να αναθέσει πολλαπλές πιθανές ενέργειες σε κάθε περιστατικό, διασφαλίζοντας μεγαλύτερη ευελιξία.

3. **Ανάλυση Περιστατικών με «Case-Based Reasoning»**

   Η προσέγγιση «case-based reasoning» χρησιμοποιήθηκε για την αντιστοίχιση νέων περιστατικών με παλαιότερα. Το σύστημα αναλύει ένα νέο περιστατικό και το συγκρίνει με τα ήδη γνωστά περιστατικά που έχουν αποθηκευτεί στον διανυσματικό χώρο. Με βάση τα κοντινότερα περιστατικά, προτείνει τα πιο κατάλληλα «playbooks».

4. **Χρήση Εξωτερικών Βάσεων Γνώσεων (π.χ., MITRE ATT&CK)**

   Το σύστημα αντλεί πληροφορίες από εξωτερικές πηγές, όπως η βάση MITRE ATT&CK, για να εμπλουτίσει τα δεδομένα και να προσφέρει λεπτομερέστερες αναλύσεις.

Η χρήση τεχνητής νοημοσύνης και ειδικά τεχνικών όπως η εκμάθηση μετρικών και η πολυ-ετικετική ταξινόμηση επέτρεψε στους επιστήμονες να δημιουργήσουν ένα καινοτόμο, έξυπνο σύστημα για την αντιμετώπιση περιστατικών κυβερνοασφάλειας. Το σύστημα αυτό υπερτερεί των παραδοσιακών προσεγγίσεων, προσφέροντας μεγαλύτερη ακρίβεια, ευελιξία και αποτελεσματικότητα, ενώ ταυτόχρονα μειώνει την ανάγκη για συνεχή ανθρώπινη παρέμβαση.

Οι επιστήμονες Liu Ping, Yu Haifeng και Ma Guoqing πραγματοποίησαν μια σημαντική έρευνα [195] με στόχο την ανάπτυξη ενός συστήματος υποστήριξης αποφάσεων για την αντιμετώπιση περιστατικών ασφάλειας στον κυβερνοχώρο. Η έρευνά τους βασίζεται στην αξιοποίηση της Λογικής Βάσης Περιστατικών (Case-Based Reasoning - CBR) και της Οντολογίας, με σκοπό τη βελτίωση της ταχύτητας και της αποτελεσματικότητας στην απόκριση σε περιστατικά ασφάλειας. Το βασικό σκεπτικό πίσω από την έρευνα ήταν ότι τα περιστατικά ασφάλειας συχνά παρουσιάζουν ομοιότητες. Για παράδειγμα, μια επίθεση DoS ή η εκμετάλλευση μιας συγκεκριμένης ευπάθειας μπορεί να έχει αντιμετωπιστεί με επιτυχία στο παρελθόν, και η εμπειρία αυτή μπορεί να χρησιμοποιηθεί για την ταχύτερη αντιμετώπιση παρόμοιων περιστατικών στο μέλλον.

Η έρευνα στόχευσε στην ανάπτυξη ενός συστήματος λήψης αποφάσεων που αξιοποιεί τη μέθοδο της Λογικής Βάσης Περιστατικών **(CBR)** και την οντολογία για την τυποποίηση των περιστατικών, με στόχο τη βελτίωση της ταχύτητας και της ακρίβειας στην απόκριση περιστατικών ασφάλειας. Επιδίωξε να μειώσει τον χρόνο αναζήτησης κατάλληλων λύσεων, διευκολύνοντας παράλληλα την κοινή χρήση και την επαναχρησιμοποίηση γνώσης μεταξύ ειδικών στον τομέα της ασφάλειας πληροφοριών. Επιπλέον, επιδίωξε να παρέχει ένα ευέλικτο στρατηγικό πλαίσιο που να μπορεί να προσαρμοστεί στις ανάγκες διαφορετικών περιβαλλόντων, ενισχύοντας την αποτελεσματική διαχείριση και αντιμετώπιση περιστατικών ασφάλειας.

Οι επιστήμονες χρησιμοποίησαν δύο βασικές τεχνολογίες Τεχνητής Νοημοσύνης:

1. **Λογική Βάση Περιστατικών (Case-Based Reasoning - CBR)**

   Η μέθοδος CBR λειτουργεί με βάση έναν κύκλο 4R:

   - Ανάκτηση (Retrieve): Το σύστημα αναζητά την καλύτερη περίπτωση από τη βάση περιστατικών που ταιριάζει με το νέο περιστατικό.
   - Επαναχρησιμοποίηση (Reuse): Η λύση του προηγούμενου περιστατικού προσαρμόζεται στις ανάγκες του νέου περιστατικού.
   - Αξιολόγηση (Review): Η εφαρμογή της λύσης ελέγχεται για την αποτελεσματικότητά της.
   - Διατήρηση (Retain): Το νέο περιστατικό και η λύση του προστίθενται στη βάση περιστατικών για μελλοντική χρήση.

Η μέθοδος CBR επέτρεψε στο σύστημα να αξιοποιεί γρήγορα προηγούμενες εμπειρίες για την αντιμετώπιση νέων περιστατικών.

2. **Οντολογία**

   Η οντολογία χρησιμοποιήθηκε για τη δομημένη περιγραφή των περιστατικών. Συγκεκριμένα:

   - Τυποποιήθηκαν οι ιδιότητες των περιστατικών (π.χ. αναγνωριστικό, τεχνολογία διείσδυσης, ευπάθεια, χρόνος).

- Δημιουργήθηκε ένα μοντέλο που συνδέει τα περιστατικά με τις ενδεδειγμένες λύσεις.

Με τη χρήση οντολογίας, το σύστημα έγινε ικανό να κατανοεί και να συγκρίνει με ακρίβεια τα χαρακτηριστικά των περιστατικών, βελτιώνοντας την απόδοση της μεθόδου CBR.

Η έρευνα κατάφερε να συνδυάσει την εμπειρία απόκρισης περιστατικών με την τυποποίηση της γνώσης, δημιουργώντας ένα ισχυρό εργαλείο υποστήριξης αποφάσεων. Με τη συνεργασία των μεθόδων CBR και οντολογίας, επιτεύχθηκε ταχύτερη λήψη αποφάσεων, καθώς η αναζήτηση λύσεων βασίστηκε σε παλαιότερες επιτυχείς εμπειρίες. Παράλληλα, αυξήθηκε η επαναχρησιμοποίηση της γνώσης, διευκολύνοντας την κοινή χρήση πληροφοριών μεταξύ ειδικών. Το αποτέλεσμα ήταν η ανάπτυξη ενός προσαρμοστικού συστήματος, ικανού να ανταποκριθεί αποτελεσματικά σε διαφορετικά περιστατικά ασφάλειας, παρέχοντας ευελιξία και αξιοπιστία στη διαχείριση περιστατικών.

5.2 Αυτόματος Χαρακτηρισμός Περιστατικών

Ο αυτόματος χαρακτηρισμός περιστατικών (Automatic Incident Characterization) αναφέρεται στις διαδικασίες που χρησιμοποιούνται για να καθοριστεί η κατηγορία ενός περιστατικού, με βάση το σχέδιο αντιμετώπισης. Στο πλαίσιο αυτό, αξιολογείται η σοβαρότητα του περιστατικού και η σύνδεσή του με άλλα περιστατικά, ώστε να γίνεται αυτόματα η ιεράρχηση προτεραιοτήτων για περαιτέρω ανάλυση και διαχείριση.

Η ερευνητική ομάδα ξεκίνησε τη μελέτη της [196] με σκοπό την αντιμετώπιση της αυξανόμενης πολυπλοκότητας και ποσότητας δεδομένων που σχετίζονται με κυβερνοσυμβάντα. Στο επίκεντρο της έρευνας βρέθηκε η ανάγκη ανάπτυξης αυτοματοποιημένων μεθόδων που θα επιτρέπουν την κατηγοριοποίηση της σοβαρότητας των κυβερνοσυμβάντων, ώστε να διευκολυνθεί η άμεση και στοχευμένη απόκριση σε περιστατικά κυβερνοασφάλειας.

1. **Κατηγοριοποίηση δεδομένων σε ομάδες (G1, G2, G3):**
    - G1: Συμβάντα χωρίς σήμανση σοβαρότητας, που απαιτούν μη επιτηρούμενη μάθηση.
    - G2: Συμβάντα με έγκυρη σήμανση, που δεν απαιτούν ανάλυση.
    - G3: Συμβάντα με μερική σήμανση, που απαιτούν επιτηρούμενη/ημι-επιτηρούμενη μάθηση. Η ανάλυση επικεντρώθηκε στα συμβάντα G3.
2. **Επιλογή Χαρακτηριστικών:**
    - Από τα 113 χαρακτηριστικά, επιλέχθηκαν τα πιο σημαντικά βάσει της συνάρτησης αμοιβαίας πληροφορίας (MIF), που μετρά την εξάρτηση μεταξύ χαρακτηριστικών και στόχου.
3. **Χρήση Αλγορίθμων Μηχανικής Μάθησης:**
    - Επιτηρούμενη μάθηση: Χρησιμοποιήθηκαν αλγόριθμοι όπως **Random Forest (RF), Gradient Boosting (GB), Decision Tree (DT), AdaBoost (AB), και Multi-Layer Perceptron (MLP)**.
    - Ημι-επιτηρούμενη μάθηση: Χρησιμοποιήθηκε ο **αλγόριθμος K-means (KM)**.
4. **Αξιολόγηση Μοντέλων:**

- Ακρίβεια (Accuracy): Βασικός δείκτης αξιολόγησης της συνολικής απόδοσης.
- Matthews Correlation Coefficient (MCC): Κατάλληλο για μη ισορροπημένα δεδομένα.
- Μήτρα σύγχυσης (Confusion Matrix): Ανάλυση της απόδοσης των μοντέλων στις διαφορετικές κατηγορίες.
- Καμπύλες μάθησης (Learning Curves): Ανάλυση της γενίκευσης και αποφυγής υπερπροσαρμογής.
- Μακρο/Μικρο μέσοι όροι: Υπολογισμός μέσων όρων δεικτών όπως ευαισθησία, ειδικότητα, F1-score.

Με τη χρήση προηγμένων τεχνικών τεχνητής νοημοσύνης και μαθηματικών εργαλείων, οι επιστήμονες ανέπτυξαν αξιόπιστα και αποδοτικά μοντέλα για την κατηγοριοποίηση της σοβαρότητας των κυβερνοσυμβάντων.

5.3 Σύνοψη εργαλείων/μεθόδων AI

Ο παρακάτω πίνακας συνοψίζει τα εργαλεία, τις μεθόδους και τους αλγορίθμους τεχνητής νοημοσύνης (AI) που υποστηρίζουν τη λειτουργία "Ανταπόκριση" (Respond) του NIST Cybersecurity Framework, η οποία επικεντρώνεται στην άμεση και αποτελεσματική αντιμετώπιση περιστατικών κυβερνοασφάλειας. Περιλαμβάνει τεχνικές όπως Dynamic Case Management (CRB, Multi-Label Classification) για την οργάνωση και διαχείριση περιστατικών, καθώς και αλγορίθμους μηχανικής μάθησης (Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Multi-Layer Perceptron, K-means) για τον αυτόματο χαρακτηρισμό συμβάντων. Αυτές οι μέθοδοι συμβάλλουν στη γρήγορη ανάλυση και ταξινόμηση των περιστατικών, επιτρέποντας στους οργανισμούς να μειώνουν τον αντίκτυπο των κυβερνοεπιθέσεων και να βελτιώνουν τη συνολική τους ασφάλεια.

2.4.1	Διαχείριση Δυναμικών Περιστατικών	(1) CRB (2) Multi – Label Classification
2.4.2	Αυτόματος Χαρακτηρισμός Περιστατικών	(1) Random Forest (RF) (2) Gradient – Boosting (GB) (3) Decision Tree (DT) (4) AdaBoost (AB) (5) Multi – Layer Perceptron (MLP) (6) K – means (KM)

Πίνακας 5 : Αλγόριθμοι και τεχνικές τεχνητής νοημοσύνης που υποστηρίζουν τη λειτουργία «Ανταπόκριση» (Respond) του NIST Cybersecurity Framework, διευκολύνοντας την ταχεία ανάλυση και διαχείριση περιστατικών κυβερνοασφάλειας.

6. Συμπεράσματα

Η πλειονότητα των εργαλείων και αλγορίθμων που αναφέρονται στο πλαίσιο του NIST Cybersecurity Framework βασίζεται σε τεχνικές Machine Learning (ML) και Deep Learning (DL). Αυτές οι τεχνικές είναι ιδιαίτερα χρήσιμες για την ανάλυση δεδομένων, την ανίχνευση ανωμαλιών, την ταξινόμηση ευπαθειών και τη βελτιστοποίηση λύσεων. Το Learning αποτελεί τον πυρήνα των περισσότερων εργαλείων που αξιοποιούνται για την ανίχνευση τρωτοτήτων, την κατηγοριοποίηση απειλών και τη βελτιστοποίηση διαδικασιών ασφαλείας.

Στον τομέα του εντοπισμού ευπαθειών (Automated Vulnerability Detection), οι πιο διαδεδομένες μέθοδοι περιλαμβάνουν αλγορίθμους όπως BERT, CNN, NeuFuzz και εργαλεία όπως Skyfire, DeepSmith και smart fuzzing. Αυτές οι τεχνικές εκμεταλλεύονται δεδομένα κώδικα και εισόδους για να αποκαλύψουν πιθανά σφάλματα ή αδυναμίες στα συστήματα. Επιπλέον, μέθοδοι όπως Random Forest (RF), SVM, και Logistic Regression χρησιμοποιούνται ευρέως για την κατηγοριοποίηση και την πρόβλεψη πιθανών τρωτοτήτων με βάση χαρακτηριστικά δεδομένων (Vulnerability Assessment & Priorization). Αντίστοιχα, για την πρόληψη διαρροής δεδομένων (Data Leakage Prevention), αλγόριθμοι όπως Autoencoders, Isolation Forest, DBSCAN, και άλλοι, παίζουν καθοριστικό ρόλο.

Παράλληλα, σημαντικό ρόλο διαδραματίζουν οι τεχνικές Reasoning (Λογική Συλλογιστική) και Planning (Σχεδιασμός). Εργαλεία όπως το PDDL, Max-SAT, και μεθοδολογίες όπως το Model Driven Security (MDS) επιτρέπουν τον σχεδιασμό πολιτικών ασφαλείας με βάση πολλαπλά κριτήρια (Automated Access Control). Παράλληλα, συστήματα όπως το Crown Jewel Analysis (CJA-RL) βασίζονται στη λογική εκτίμηση για την προστασία κρίσιμων πόρων, ενώ η χρήση εργαλείων όπως το Fast-Downward και το Attribute-Based Access Control (ABAC) επιτρέπει την αυτοματοποιημένη διαχείριση πρόσβασης.

Η Επεξεργασία Φυσικής Γλώσσας (NLP) και οι τεχνικές transformers διαδραματίζουν επίσης βασικό ρόλο στη συλλογή και ανάλυση δεδομένων. Εργαλεία όπως Word2Vec, TF-IDF, και BERT χρησιμοποιούνται για την ανάλυση κειμένων που περιέχουν πληροφορίες ευπαθειών ή απειλών (Vulnerability Exploration). Οι τεχνικές αυτές διευκολύνουν την κατανόηση των δεδομένων, επιτρέποντας την εξαγωγή χρήσιμων πληροφοριών και την αυτόματη κατηγοριοποίηση.

Ενδιαφέρον παρουσιάζει η ευρεία χρήση γενετικών αλγορίθμων και μεθόδων ενισχυτικής μάθησης για τη βελτιστοποίηση λύσεων. Αλγόριθμοι όπως οι Pareto Q-learning, Pareto SARSA, Pareto TD(0) και οι γενετικοί αλγόριθμοι NSGA-II και SPEA2 εφαρμόζονται σε πολυαντικειμενικές βελτιστοποιήσεις (Automated Configuration Management). Αυτές οι μέθοδοι είναι ιδιαίτερα χρήσιμες σε σενάρια όπου πρέπει να βρεθούν οι καλύτερες δυνατές λύσεις με βάση πολλαπλά κριτήρια, κάτι που είναι κρίσιμο στην κυβερνοασφάλεια.

Ένα ακόμα σημαντικό σημείο είναι η χρήση τεχνικών αξιολόγησης και μέτρων απόδοσης. Οι μετρικές όπως Precision, Recall, F1-score, ROC-AUC curves και άλλες χρησιμοποιούνται ευρέως για την αξιολόγηση της αποτελεσματικότητας των μοντέλων ασφαλείας. Συστήματα όπως το SAMM (Self-Adaptive Mathematical Morphology) και μέθοδοι όπως το Adaptive Boosting (AdaBoost) βελτιώνουν την προσαρμοστικότητα και απόδοση των μοντέλων, ειδικά σε δυναμικά περιβάλλοντα όπου οι απειλές εξελίσσονται (AI – Supported Device Authentication).

Ο τομέας των Intrusion Detection Systems (IDS) αποτελεί ένα χαρακτηριστικό παράδειγμα εφαρμογής των παραπάνω εργαλείων. Εργαλεία όπως τα LSTM, RNN-CNN, GAN, και συστήματα όπως το IGAN-IDS αξιοποιούνται για την ανίχνευση επιθέσεων. Μέθοδοι όπως το SMOTE, η PCA,

και η k-fold cross-validation βελτιώνουν την αποτελεσματικότητα της ανίχνευσης, ενώ συστήματα όπως το Edge-ENClf και το Cloud-ENClf ενσωματώνουν σύγχρονες τεχνικές ταξινόμησης για την προστασία των υποδομών (Intrusion Detection System – IDS).

Συνοψίζοντας, τα εργαλεία και οι μέθοδοι που περιγράφονται υπογραμμίζουν τη σημασία του Machine Learning (ML) ως κυρίαρχης τεχνολογίας στην κυβερνοασφάλεια. Οι εφαρμογές του εκτείνονται σε κρίσιμους τομείς, όπως η ανίχνευση ευπαθειών, η ανάλυση κινδύνων, η ανίχνευση εισβολών και η πρόληψη διαρροών δεδομένων. Μέσω προηγμένων τεχνικών, όπως το Deep Learning, οι πολυαντικειμενικές βελτιστοποιήσεις και οι μέθοδοι ενισχυτικής μάθησης, οι οργανισμοί μπορούν να ενισχύσουν την ασφάλεια των συστημάτων τους, αποκτώντας τη δυνατότητα να ανταποκριθούν αποτελεσματικά σε σύγχρονες προκλήσεις. Επιπλέον, η συλλογιστική και ο σχεδιασμός, που επιτρέπουν την αυτοματοποίηση της λήψης αποφάσεων, αποτελούν εξίσου σημαντικές διαστάσεις που ενισχύουν τη συνολική στρατηγική προστασίας.

Η χρήση της τεχνητής νοημοσύνης (AI) στην προστασία και διαχείριση πληροφοριών δεν είναι απλώς σημαντική, αλλά κρίσιμη για την αποτελεσματική αντιμετώπιση κυβερνοαπειλών. Ωστόσο, η ανάπτυξη και χρήση αυτής της τεχνολογίας πρέπει να καθοδηγείται αυστηρά από ηθικές αρχές και κανονιστικές απαιτήσεις. Οι εφαρμογές AI, όπως η ανίχνευση ευπαθειών, η ανάλυση απειλών και η διαχείριση πρόσβασης, έχουν αποδείξει ότι μπορούν να βελτιώσουν σημαντικά την ασφάλεια. Αν και η πλήρης εξάλειψη των κινδύνων δεν είναι εφικτή, ο συνδυασμός προληπτικών τεχνικών, δεοντολογίας και ανθρώπινης παρέμβασης μπορεί να επιτύχει την ανάπτυξη αξιόπιστου και αποτελεσματικού AI [200].

7. Πηγές – Βιβλιογραφία

[1] Kaur, R., Gabrijelčič, D. and Klobučar, T. (2023). Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions. *Information Fusion*, [online] 97(101804), p.101804. doi:https://doi.org/10.1016/j.inffus.2023.101804.

[2] *Watkins, C.J.C.H. (1989).* Learning from Delayed Rewards *(PDF) (Ph.D. thesis).* University of Cambridge. EThOS uk.bl.ethos.330022. https://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf

[3] Janikow, C. Z and Michalewicz, Z. (1991) *Http://www.cs.umsl.edu/~janikow/publications/1991/GABin/text.pdf | Ghostarchive* (no date). https://ghostarchive.org/archive/GIfUX.

[4] Tozer, B., Mazzuchi, T. and Sarkani, S. (2015) *Optimizing attack surface and configuration diversity using multi-objective reinforcement learning*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7424300.

[5] García-Hernández, L.E. *et al.* (2020) 'Multi-objective configuration of a secured distributed cloud data storage,' in *Communications in computer and information science*, pp. 78–93. https://doi.org/10.1007/978-3-030-41005-6_6.

[6] Ray, P. Pratim (2018) *An Introduction to Dew Computing: Definition, Concept and Implications*. https://ieeexplore.ieee.org/document/8114187.

[7] Hayes, C.F. *et al.* (2022) 'A practical guide to multi-objective reinforcement learning and planning,' *Autonomous Agents and Multi-Agent Systems*, 36(1). https://doi.org/10.1007/s10458-022-09552-y.

[8] Sharifi, M., Fink, E.F. and Carbonell, J. G (2010) *Learning of personalized security settings*. https://ieeexplore.ieee.org/document/5642461.

[9] Bringhenti, D. *et al.* (2019) *Towards a fully automated and optimized network security functions orchestration*. https://ieeexplore.ieee.org/document/8888130.

[10] Battula, L. Rao (2014) *Network Security Function Virtualization(NSFV) towards Cloud computing with NFV Over Openflow infrastructure: Challenges and novel approaches*. https://ieeexplore.ieee.org/abstract/document/6968453?casa_token=_l_Di7k7F1cAAAAA:bmBELymDuKkfCiqJ3dQP1Z22Lbrm51OTiTdYgi8rjq8qK8iIH7N8Rfq1waEr793ANcncwHX4.https://www.etsi.org/deliver/etsi_gs/NFV/001_099/001/01.01.01_60/gs_NFV001v010101p.pdf

[11] Bringhenti, D., Sisto, R. and Valenza, F. (2023) *A demonstration of VEREFOO: an automated framework for virtual firewall configuration*. https://ieeexplore.ieee.org/abstract/document/10175442.

[12] Blanchette, J.C., Böhme, S. and Paulson, L.C. (2013b) 'Extending Sledgehammer with SMT Solvers,' *Journal of Automated Reasoning*, 51(1), pp. 109–128. https://doi.org/10.1007/s10817-013-9278-5.

[13] Varela-Vaca, Á.J. *et al.* (2020) 'AMADEUS,' *ACM* [Preprint]. https://doi.org/10.1145/3382025.3414952.

[14] Varela-Vaca, Á.J. *et al.* (2019) 'CybERSPL: A framework for the verification of cybersecurity policy compliance of system configurations using software product lines,' *Applied Sciences*, 9(24), p. 5364. https://doi.org/10.3390/app9245364.

[15] *Ritz, B. et al. (2023) Solving Multi-Configuration Problems: A Performance Analysis with Choco Solver. https://arxiv.org/abs/2310.02658.*

[16] Liu, Yang *et al.* (2015) *Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents*. https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/liu .

[17] Aitchison, R. (2006) 'An introduction to DNS,' in *Apress eBooks*. Apress, pp. 3–19. https://doi.org/10.1007/978-1-4302-0050-5_1.

[18] Butler, K.B. *et al.* (2010) *A survey of BGP security Issues and solutions*. https://ieeexplore.ieee.org/abstract/document/5357585.

[19] Zhan, Z., Xu, M. and Xu, S. (2015) 'A Characterization of Cybersecurity Posture from Network Telescope Data,' in *Lecture notes in computer science*, pp. 105–126. https://doi.org/10.1007/978-3-319-27998-5_7.

[20] Shumway, R.H. and Stoffer, D.S. (2017) 'ARIMA models,' in *Springer texts in statistics*, pp. 75–163. https://doi.org/10.1007/978-3-319-52452-8_3.

[21] Bollerslev, Tim *(1986). "Generalized Autoregressive Conditional Heteroskedasticity".* Journal of Econometrics. ***31*** *(3): 307–327.* CiteSeerX 10.1.1.468.2892. doi:10.1016/0304-4076(86)90063-1. S2CID 8797625.https://www.semanticscholar.org/paper/Generalized-autoregressive-conditional-Bollerslev/584c7954ebb89d6155fa50e5bcf44098fb881faa

[22] Müller, M. (2007) 'Dynamic time warping,' in *Springer eBooks*. Springer, Berlin, Heidelberg, pp. 69–84. https://doi.org/10.1007/978-3-540-74048-3_4.

[23] Mushtaq, R. (2011) 'Augmented Dickey Fuller test,' *SSRN Electronic Journal* [Preprint]. https://doi.org/10.2139/ssrn.1911068.

[24] Gourisetti, S. Nikhil Gupta *et al.* (2017) *Multi-scenario use case based demonstration of Buildings Cybersecurity Framework webtool*. https://ieeexplore.ieee.org/document/8285240.

[25] Gourisetti, S. Nikhil Gupta *et al.* (2017b) *Multi-scenario use case based demonstration of Buildings Cybersecurity Framework webtool*. https://ieeexplore.ieee.org/abstract/document/8285240.

[26] Stepanov, L.V., Koltsov, A.S. and Parinov, A.V. (2021) 'Evaluating the cybersecurity of an enterprise based on a genetic algorithm,' in *Lecture notes in electrical engineering*, pp. 580–590. https://doi.org/10.1007/978-3-030-71119-1_57.

[27] Ablon, L., Bogart, A., and RAND Corporation (2017) *Zero Days, Thousands of Nights: the life and times of Zero-Day vulnerabilities and their exploits*, *RAND Corporation*. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1751/RAND_RR1751.pdf.

[28] Nembhard, F. and Carvalho, M. (2019) *The impact of interface design on the usability of code analyzers*. https://ieeexplore.ieee.org/document/9020339.

[29] Kim, T.K. (2015) 'T test as a parametric statistic,' *Korean Journal of Anesthesiology*, 68(6), p. 540. https://doi.org/10.4097/kjae.2015.68.6.540.

[30] Shaw, R.G. and Mitchell-Olds, T. (1993) 'Anova for Unbalanced Data: An Overview,' *Ecology*, 74(6), pp. 1638–1645. https://doi.org/10.2307/1939922.

[31] Liu, S. *et al.* (2019) *A Novel Modified Robust Model-Free Adaptive Control Method for a Class of Nonlinear Systems with Time Delay*. https://ieeexplore.ieee.org/document/8908835.

[32] Feng, Z. *et al.* (2019) 'MFAC and parameter optimization for a class of models in HVAC,' in *Advances in intelligent systems and computing*, pp. 249–260. https://doi.org/10.1007/978-981-13-6733-5_23.

[33] Zhu, Y. and Hou, Z. (2012) *Enhanced model free adaptive control by integrating with lazy learning*. https://ieeexplore.ieee.org/abstract/document/6244325.

[34] Liu, S. *et al.* (2019b) *A Novel Modified Robust Model-Free Adaptive Control Method for a Class of Nonlinear Systems with Time Delay*. https://ieeexplore.ieee.org/abstract/document/8908835.

[35] Jeon, S. *et al.* (no date) *AutoVAS: An automated vulnerability analysis system with a deep learning approach*, *Computers & Security*, p. 102308.

[36] Szpor, G. and Gryszczyńska, A. (2022) 'Hacking in the (Cyber)space,' *bazawiedzy.uksw.edu.pl* [Preprint]. https://doi.org/10.57599/gisoj.2022.2.1.141.

[37] Witte, H.B.D.R., Gregory A. (2021) *The National Vulnerability Database (NVD): Overview*. https://www.nist.gov/publications/national-vulnerability-database-nvd-overview.

[38] Debnath, B. K, Lilja, D. J and Mokbel, M. F (2008) *SARD: A statistical approach for ranking database tuning parameters*. https://ieeexplore.ieee.org/abstract/document/4498279.

[39] Ma, L. and Zhang, Y. (2015) *Using Word2Vec to process big text data*. https://ieeexplore.ieee.org/abstract/document/7364114.

[40] Sturman, D. J and Zeltzer, D. (1994) *A survey of glove-based input*. https://ieeexplore.ieee.org/abstract/document/250916.

[41] Mojumder, P. *et al.* (2020) 'A study of FastText Word embedding Effects in document classification in Bangla language,' in *Springer eBooks*, pp. 441–453. https://doi.org/10.1007/978-3-030-52856-0_35.

[42] Berezovskiy, V. *et al.* (2023) 'Machine learning code snippets semantic classification,' *PeerJ Computer Science*, 9, p. e1654. https://doi.org/10.7717/peerj-cs.1654.

[43] Balarin, F. *et al.* (1997) *Hardware-Software Co-Design of embedded systems*. https://books.google.gr/books?hl=el&lr=&id=_pzqBwAAQBAJ&oi=fnd&pg=PR9&dq=%CE%BC%CE%AD%CE%B8%CE%BF%CE%B4%CE%BF%CF%82+embedding+in+software&ots=urFlPpOnKa&sig=EGppHgjVW9_UDmuecGjHvC_reME&redir_esc=y#v=onepage&q&f=false.

[44] Weiser, M. (1984) *Program slicing*. https://ieeexplore.ieee.org/abstract/document/5010248.

[45] Li, Z. *et al.* (2018) 'VulDeePecker: A Deep Learning-Based System for Vulnerability Detection,' *Arxiv* [Preprint]. https://doi.org/10.14722/ndss.2018.23158.

[46] Li, Z. *et al.* (2022) *SYSEVR: a framework for using deep learning to detect software vulnerabilities*. https://ieeexplore.ieee.org/abstract/document/9321538.

[47] Huff, P. *et al.* (2021) 'A Recommender System for Tracking Vulnerabilities,' *Acm*, pp. 1–7. https://doi.org/10.1145/3465481.3470039.

[48] Witte, H.B.D.R., Gregory A. (2021b) *The National Vulnerability Database (NVD): Overview*. https://www.nist.gov/publications/national-vulnerability-database-nvd-overview.

[49] Cheikes, B.A. *et al.* (2011) *Common Platform Enumeration: naming specification*, *NIST Interagency Report 7695*. report. National Institute of Standards and Technology, p. 49. https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7695.pdf.

[50] Turchioe, M.R. *et al.* (2021) 'Systematic review of current natural language processing methods and applications in cardiology,' *Heart*, 108(12), pp. 909–916. https://doi.org/10.1136/heartjnl-2021-319769.

[51] Rahutomo, F. *et al.* (2012) *Semantic Cosine similarity*, *Conference Paper*. https://www.researchgate.net/publication/262525676.

[52] *Yggdrasil — Early Detection of Cybernetic Vulnerabilities from Twitter* (2021). https://ieeexplore.ieee.org/document/9481044.

[53] Alaparthi, S. and Mishra, M. (2020) *Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey*. https://arxiv.org/abs/2007.01127.

[54] Suthaharan, S. (2015) 'Support Vector machine,' in *Integrated series on information systems/Integrated series in information systems*, pp. 207–235. https://doi.org/10.1007/978-1-4899-7641-3_9.

[55] Iorga, D. *et al.* (2022) *A Survey of Convolutional Neural Networks: Analysis, applications, and Prospects*. https://ieeexplore.ieee.org/abstract/document/9451544.

[56] Saha, T. *et al.* (2022) *SHARKS: Smart hacking approaches for RISK scanning in Internet-of-Things and Cyber-Physical systems based on machine learning*. https://ieeexplore.ieee.org/abstract/document/9319511.

[57] Wang, B. *et al.* (2019) *Cybersecurity Enhancement of power trading within the networked microgrids based on blockchain and directed acyclic graph approach*. https://ieeexplore.ieee.org/abstract/document/8725596.

[58] Farsi, M., Barbosa, M. and Ratcliff, K. (1999) 'An overview of Controller Area Network,' *Computing & Control Engineering Journal*, 10(3), pp. 113–120. https://doi.org/10.1049/cce:19990304.

[59] *NeuFuZz: efficient fuzzing with deep neural network* (2019). https://ieeexplore.ieee.org/abstract/document/8672949.

[60] Black, P.E. and National Institute of Standards and Technology (2017) *SARD: Thousands of reference programs for software assurance*. journal-article. https://www.researchgate.net/publication/322129343.

[61] Cosentino, V., Luis, J. and Cabot, J. (2016) 'Findings from GitHub,' *Acm* [Preprint]. https://doi.org/10.1145/2901739.2901776.

[62] Wikipedia contributors (2024a) *Common vulnerabilities and exposures*. https://en.wikipedia.org/wiki/Common_Vulnerabilities_and_Exposures.

[63] Zhang, G. *et al.* (2018) *PTFuZZ: Guided fuzzing with processor trace feedback*. https://ieeexplore.ieee.org/abstract/document/8399803.

[64] *Grey-box fuzzing with deep reinforcement learning and process trace back* (2021). https://ieeexplore.ieee.org/abstract/document/9512987?casa_token=NOdigqJ8jf8AAAAA:PZgXBWTf-tDDxRfGCwG8E0H3RzkWaAUSxbqJ87IgP3hdoS_APjqgK7ofUNvPLQZf9Al63IPS.

[65] Wang, J. *et al.* (2017) *SkyFire: Data-Driven Seed Generation for fuzzing*. https://ieeexplore.ieee.org/document/7958599.

[66] Kim, S.J. and Kim, T. (2020) 'Design methodology and computational fluid analysis for the printed circuit steam generator (PCSG),' *Journal of Mechanical Science and Technology*, 34(12), pp. 5303–5314. https://doi.org/10.1007/s12206-020-1131-2.

[67] Owen, A. (2018) *Monte Carlo Sampling*. https://artowen.su.domains/mc/Ch-intro.pdf.

[68] Godefroid, P., Peleg, H. and Singh, R. (2017) *Learn&Fuzz: Machine learning for input fuzzing*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8115618.

[69] I, P. and Paulose, J. (2019) 'Prediction of Answer Keywords using Char-RNN,' *International Journal of Electrical and Computer Engineering (IJECE)*, 9(3), p. 2164. https://doi.org/10.11591/ijece.v9i3.pp2164-2176.

[70] Cummins, C. *et al.* (2018) 'Compiler fuzzing through deep learning,' *Acm* [Preprint]. https://doi.org/10.1145/3213846.3213848.

[71] Cummins, C. *et al.* (2018b) *DeepSmith: Compiler Fuzzing through Deep Learning DeepSmith*. journal-article. https://chriscummins.cc/pub/2018-acaces.pdf.

[72] Graves, A. (2012) 'Long Short-Term memory,' in *Studies in computational intelligence*, pp. 37–45. https://doi.org/10.1007/978-3-642-24797-2_4.

[73] Munshi, A. *et al.* (2012) *OpenCL Programming Guide*. https://books.google.gr/books?hl=el&lr=&id=M-Sve_KItQwC&oi=fnd&pg=PR5&dq=OpenCL&ots=cMPrTl2fld&sig=TYcmd6UBNX7WQ4PCLixF3UtMySg&redir_esc=y#v=onepage&q=OpenCL&f=false.

[74] Ctuning (2016) *GitHub - ctuning/ck-clsmith: Collective Knowledge extension to crowdsource bug detection in OpenCL compilers using CLSmith tool from Imperial College London*. https://github.com/ctuning/ck-clsmith.

[75] Wang, Z. *et al.* (2021) *An empirical study of solidity language features*. https://ieeexplore.ieee.org/abstract/document/9742076?casa_token=vnXg4Pu8I7QAAAAA:pNcU4KfLC_aaC9cHS_dnbOflFnaYwiiIacmsrlHBu--LLX8J6nsii3mjwgXBp-BYSIU7Yb_OcYVI.

[76] Xu, H. *et al.* (2020) *DSmith: Compiler Fuzzing through Generative Deep Learning Model with Attention*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9206911.

[77] *GCC, the GNU Compiler Collection - GNU Project* (2022). https://gcc.gnu.org/.

[78] Chen, Y. *et al.* (2019) *Learning-Guided network fuzzing for testing Cyber-Physical system defences*. https://ieeexplore.ieee.org/document/8952193.

[79] Nelson, A.L., a *et al.* (2008) *Fitness functions in evolutionary robotics: A survey and analysis*, *Robotics and Autonomous Systems*, pp. 345–370. https://doi.org/10.1016/j.robot.2008.09.009.

[80] Goh, J. *et al.* (2017) 'A dataset to support research in the design of secure water treatment systems,' in *Lecture notes in computer science*, pp. 88–99. https://doi.org/10.1007/978-3-319-71368-7_8.

[81] Walski, T.M., Chase, D.V. and Savic, D.A. (no date) *Water distribution modeling*. https://ecommons.udayton.edu/cee_fac_pub/17/.

[82] She, D. *et al.* (2019) *IEEE Xplore Full-Text PDF*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8835342.

[83] Yang, S. *et al.* (2022) *A Gradient-Guided evolutionary approach to training deep neural networks* (2022). https://ieeexplore.ieee.org/abstract/document/9369973.

[84] Youngdale, E. (no date) *The ELF object file format by dissection*. https://picture.iczhiku.com/resource/eetop/wykEsIQUTozfYxNX.pdf.

[85] Youngdale, E. (no date) *The ELF object file format by dissection*. https://picture.iczhiku.com/resource/eetop/wykEsIQUTozfYxNX.pdf.

[86] Liu, X. *et al.* (2019) *DeepFuzz: Automatic generation of Syntax Valid C programs for fuzz testing*, *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*.

[87] Zhou, S. *et al.* (2021) 'Autonomous penetration testing based on improved deep Q-Network,' *Applied Sciences*, 11(19), p. 8823. https://doi.org/10.3390/app11198823.

[88] Kaelbling, L.P., Littman, M.L. and Moore, A.W. (1996) 'Reinforcement Learning: a survey,' *Journal of Artificial Intelligence Research*, 4, pp. 237–285. https://doi.org/10.1613/jair.301.

[89] Βλάχου, A. and Εθνικό Μετσόβιο Πολυτεχνείο (2023) *Μαρκοβιανές αλυσίδες, Μαρκοβιανές διαδικασίες απόφασης και προσομοίωση*, *Εθνικό Μετσόβιο Πολυτεχνείο*. https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/59096/Alexandra_Vlachou_Thesis_2023.pdf?sequence=1.

[90] Roderick, M., MacGlashan, J. and Tellex, S. (2017) *Implementing the deep Q-Network*. https://arxiv.org/abs/1711.07478.

[91] AdelNasim (2023) *Adel Nasim platform*. https://adelnasim.com/.

[92] Dorri, A., Kanhere, S. S and Jurdak, R. (2018) *Multi-Agent Systems: a survey*. https://ieeexplore.ieee.org/abstract/document/8352646.

[93] Pateria, S. *et al.* (2021) 'Hierarchical reinforcement learning,' *ACM Computing Surveys*, 54(5), pp. 1–35. https://doi.org/10.1145/3453160.

[94] Gangupantulu, R. *et al.* (2021) *Crown Jewels Analysis using Reinforcement Learning with Attack Graphs* (2021). https://ieeexplore.ieee.org/document/9659947.

[95] Tayouri, D. *et al.* (2023) *A survey of MULVAL extensions and their attack scenarios coverage* (2023). https://ieeexplore.ieee.org/abstract/document/10070747.

[96] Neal, C. *et al.* (2021) *Reinforcement learning based penetration testing of a microgrid control algorithm* (2021). https://ieeexplore.ieee.org/document/9376126.

[97] Babaeizadeh, M. *et al.* (2016) *Reinforcement Learning through Asynchronous Advantage Actor-Critic on a GPU*. https://arxiv.org/abs/1611.06256.

[98] Russo, E.R., a,c *et al.* (2019) *Summarizing vulnerabilities' descriptions to support experts during vulnerability assessment activities*, *The Journal of Systems and Software*, pp. 84–99. https://doi.org/10.1016/j.jss.2019.06.001.

[99] De Marneffe, M.-C. *et al.* (2008) *The Stanford typed dependencies representation*, *Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pp. 1–8. https://aclanthology.org/W08-1301.pdf.

[100] Zhang, Y. *et al.* (2024) 'BASIC: BayesNet Structure Learning for Computational Scalable Neural Image Compression,' in *Lecture notes in computer science*, pp. 269–285. https://doi.org/10.1007/978-3-031-72698-9_16.

[101] Aota, M. *et al.* (2020) *Automation of Vulnerability Classification from its Description using Machine Learning*. https://ieeexplore.ieee.org/document/9219568.

[102] Neuhaus, S. and Zimmermann, T. (2010) *Security Trend Analysis with CVE Topic Models*. https://ieeexplore.ieee.org/abstract/document/5635130?casa_token=caoauhqzcoYAAAAA:brsBKmDmXOZ146JifBkY593TddNrlz7ca7o8N3JfgVJMOHlW-MK6fZwuCnp_6STop8hDmWqn.

[103] Fattahi, J. *et al.* (2024) 'Cyberbullying detection using Bag-of-Words, TF-IDF, parallel CNNs and BILSTM neural networks,' in *Frontiers in artificial intelligence and applications*. https://doi.org/10.3233/faia240357.

[104] Farhana, N. *et al.* (2022) *Evaluation of Boruta algorithm in DDoS detection*, *Egyptian Informatics Journal*. journal-article, pp. 27–42. https://doi.org/10.1016/j.eij.2022.10.005.

[105] Vanamala, M., Yuan, X. and Roy, K. (2020) *Topic Modeling and Classification of Common Vulnerabilities and Exposures Database*. https://ieeexplore.ieee.org/document/9183814.

[106] Jelodar, H. *et al.* (2018) 'Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,' *Multimedia Tools and Applications*, 78(11), pp. 15169–15211. https://doi.org/10.1007/s11042-018-6894-4.

[107] Bakirtzis, G. *et al.* (2020) *Data-Driven Vulnerability exploration for design phase system analysis*. https://ieeexplore.ieee.org/abstract/document/8850328.

[108] Friedenthal, S., Moore, A. and Steiner, R. (2015) *A Practical Guide to SysML*. https://books.google.gr/books?id=Ze60AwAAQBAJ&printsec=frontcover&hl=el#v=onepage&q&f=false.

[109] Tarnowski, I. and Wroclaw Centre for Networking and Supercomputing, Wroclaw University of Science and Technology (2017) 'How to use cyber kill chain model to build cybersecurity?,' *Case Study* [Preprint]. https://tnc17.geant.org/getfile/tnc17_paper_TNC17-IreneuszTarnowski-HowToUseCyberKillChainModelToBuildCybersecurity_-En.pdf.

[110] Kuppa, A., Aouad, L. and Le-Khac, N.-A. (2021) 'Linking CVE's to MITRE ATT&CK Techniques,' *ACM* [Preprint]. https://doi.org/10.1145/3465481.3465758.

[111] Forman, G. (2008) 'BNS feature scaling,' *ACM* [Preprint]. https://doi.org/10.1145/1458082.1458119.

[112] Chatterjee, S. and Thekdi, S. (2019). An iterative learning and inference approach to managing dynamic cyber-vulnerabilities of complex systems. *Reliability Engineering & System Safety*, p.106664. doi:https://doi.org/10.1016/j.ress.2019.106664.

[113] Eddy, S.R. (1996). Hidden Markov models. *Current Opinion in Structural Biology*, [online] 6(3), pp.361–365. doi:https://doi.org/10.1016/s0959-440x(96)80056-x.

[114] Jiang, Y. and Atif, Y. (2021). A selective ensemble model for cognitive cybersecurity analysis. *Journal of Network and Computer Applications*, 193, p.103210. doi:https://doi.org/10.1016/j.jnca.2021.103210.

[115] Samtani, S., Yu, S., Zhu, H., Patton, M., Matherly, J. and Chen, H. (2018). Identifying SCADA Systems and Their Vulnerabilities on the Internet of Things: A Text-Mining Approach. *IEEE Intelligent Systems*, 33(2), pp.63–73. doi:https://doi.org/10.1109/mis.2018.111145022.

[116] Rodrigues, A., Best, T. and Ravi Pendse (2011). SCADA security device. doi:https://doi.org/10.1145/2179298.2179325.

[117] Google Books. (2024). *Nessus Network Auditing*. [online] Available at: https://books.google.gr/books?hl=el&lr=&id=3OiclLcGdTgC&oi=fnd&pg=PP1&dq=Nessus&ots=YtKeYMjTXv&sig=JFVlZQsYdN7aXNbjREEh3mv3jro&redir_esc=y#v=onepage&q=Nessus&f=false [Accessed 13 Nov. 2024].

[118] Brown, J., Saha, T. and Jha, N.K. (2021). GRAVITAS: Graphical Reticulated Attack Vectors for Internet-of-Things Aggregate Security. *IEEE Transactions on Emerging Topics in Computing*, pp.1–1. doi:https://doi.org/10.1109/tetc.2021.3082525.

[119] DREXLER, H. (1956). GRAVITAS. *Aevum*, [online] 30(4), pp.291–306. doi:https://doi.org/10.2307/20858938.

[120] Gao, P., Shao, F., Liu, X., Xiao, X., Qin, Z., Xu, F., Mittal, P., Kulkarni, S.R. and Song, D. (2021). *Enabling Efficient Cyber Threat Hunting With Cyber Threat Intelligence*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICDE51399.2021.00024.

[121] Hafeez, A., Topolovec, K. and Awad, S. (2019). *ECU Fingerprinting through Parametric Signal Modeling and Artificial Neural Networks for In-vehicle Security against Spoofing Attacks*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICENCO48310.2019.9027298.

[122] Corrigan, S. (2002). *Application Report Introduction to the Controller Area Network (CAN)*. [online] Available at: https://masters.donntu.ru/2005/fvti/trofunenko/library/sloa101.pdf.

[123] Baldini, G., Giuliani, R., Gemo, M. and Dimc, F. (2021). On the application of sensor authentication with intrinsic physical features to vehicle security. *Computers & Electrical Engineering*, [online] 91, p.107053. doi:https://doi.org/10.1016/j.compeleceng.2021.107053.

[124] Cui, Y., Bai, F., Yan, R., Saha, T., Ko, R.K.L. and Liu, Y. (2021). Source Authentication of Distribution Synchrophasors for Cybersecurity of Microgrids. *IEEE Transactions on Smart Grid*, pp.1–1. doi:https://doi.org/10.1109/tsg.2021.3089041.

[125] Ćurić, V., Landström, A., Thurley, M.J. and Luengo Hendriks, C.L. (2014). Adaptive mathematical morphology – A survey of the field. *Pattern Recognition Letters*, [online] 47, pp.18–28. doi:https://doi.org/10.1016/j.patrec.2014.02.022.

[126] Benedetti, M. and Mori, M. (2019). On the use of Max-SAT and PDDL in RBAC maintenance. *Cybersecurity*, 2(1). doi:https://doi.org/10.1186/s42400-019-0036-9.

[127] Sandhu, R.S. (1998). *Role-based Access Control11Portions of this chapter have been published earlier in Sandhu et al. (1996), Sandhu (1996), Sandhu and Bhamidipati (1997), Sandhu et al. (1997)*

and Sandhu and Feinstein (1994). [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/S0065245808602065.

[128] Hansen, P. and Jaumard, B. (1990). Algorithms for the maximum satisfiability problem. *Computing*, 44(4), pp.279–303. doi:https://doi.org/10.1007/bf02241270.

[129] Patrik Haslum, Nir Lipovetzky, Magazzeni, D. and Muise, C. (2019). *An Introduction to the Planning Domain Definition Language*. *Synthesis lectures on artificial intelligence and machine learning*. Morgan & Claypool Publishers. doi:https://doi.org/10.1007/978-3-031-01584-7.

[130] Patrik Haslum, Nir Lipovetzky, Magazzeni, D. and Muise, C. (2019). *An Introduction to the Planning Domain Definition Language*. *Synthesis lectures on artificial intelligence and machine learning*. Morgan & Claypool Publishers. doi:https://doi.org/10.1007/978-3-031-01584-7.

[131] Lee, D. and Seung, H.S. (2000). *Algorithms for Non-negative Matrix Factorization*. [online] Neural Information Processing Systems. Available at: https://proceedings.neurips.cc/paper_files/paper/2000/hash/f9d1152547c0bde01830b7e8bd60024c-Abstract.html.

[132] SAPUTRA, D.M., SAPUTRA, D. and OSWARI, L.D. (2020). Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method. *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*. doi:https://doi.org/10.2991/aisr.k.200424.051.

[133] Leander, B., Čaušević, A., Hansson, H. and Lindström, T. (2020). Access Control for Smart Manufacturing Systems. *Communications in computer and information science*, pp.463–476. doi:https://doi.org/10.1007/978-3-030-59155-7_33.

[134] Hu, V.C., Kuhn, D.R. and Ferraiolo, D.F. (2015). Attribute-Based Access Control. *Computer*, 48(2), pp.85–88. doi:https://doi.org/10.1109/mc.2015.33.

[135] Basin, D., Clavel, M. and Egea, M. (2011). A decade of model-driven security. doi:https://doi.org/10.1145/1998441.1998443.

[136] Lazouski, A., Martinelli, F. and Mori, P. (2010). Usage control in computer security: A survey. *Computer Science Review*, 4(2), pp.81–99. doi:https://doi.org/10.1016/j.cosrev.2010.02.002.

[137] Le, D.C. and Zincir-Heywood, N. (2021). Anomaly Detection for Insider Threats Using Unsupervised Ensembles. *IEEE Transactions on Network and Service Management*, pp.1–1. doi:https://doi.org/10.1109/tnsm.2021.3071928.

[138] S, A., D, S. and G, P. (2023). Malicious insider threat detection using variation of sampling methods for anomaly detection in cloud environment. *Computers and Electrical Engineering*, 105, p.108519. doi:https://doi.org/10.1016/j.compeleceng.2022.108519.

[139] Kim, J., Park, M., Kim, H., Cho, S. and Kang, P. (2019). Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms. *Applied Sciences*, 9(19), p.4018. doi:https://doi.org/10.3390/app9194018.

[140] Al-Shehari, T. and Alsowail, R.A. (2021). An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques. *Entropy*, 23(10), p.1258. doi:https://doi.org/10.3390/e23101258.

[141] Fernandez, A., Garcia, S., Herrera, F. and Chawla, N.V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, 61, pp.863–905. doi:https://doi.org/10.1613/jair.1.11192.

[142] Narkhede, S. (2018). *Understanding AUC - ROC Curve*. [online] Medium. Available at: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

[143] Khudran Alzhrani, Rudd, E.M., Boult, T.E. and C. Edward Chow (2016). Automated big text security classification. doi:https://doi.org/10.1109/isi.2016.7745451.

[144] Kim, D., Seo, D., Cho, S. and Kang, P. (2019). Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. *Information Sciences*, 477, pp.15–29. doi:https://doi.org/10.1016/j.ins.2018.10.006.

[145] Yongyan, G., Jiayong, L., Wenwu, T. and Cheng, H. eds., (2024). *Scopus preview - Scopus - Welcome to Scopus*. [online] Scopus.com. Available at: https://www.scopus.com/record/display.uri?eid=2-s2.0-85098740671&origin=inward [Accessed 9 Dec. 2024].

[146] Huiling, L., Jun, W., Hansong, X., Gaolei, L. and Mohsen, G. eds., (2024). *Scopus preview - Scopus - Welcome to Scopus*. [online] Scopus.com. Available at: https://www.scopus.com/record/display.uri?eid=2-s2.0-85120567200&origin=inward [Accessed 9 Dec. 2024].

[147] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*, 58(1), pp.82–115.

[148] Shin, B. and Lowry, P.B. (2020). A review and theoretical explanation of the 'Cyberthreat-Intelligence (CTI) capability' that needs to be fostered in information security practitioners and how this can be accomplished. *Computers & Security*, 92, p.101761. doi:https://doi.org/10.1016/j.cose.2020.101761.

[149] Alghamdi, A.A. and Reger, G. (2020). Pattern Extraction for Behaviours of Multi-Stage Threats via Unsupervised Learning. *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, [online] pp.1–8. doi:https://doi.org/10.1109/cybersa49311.2020.9139697.

[150] Zeng, Z., Yang, Z., Huang, D. and Chung, C.-J. (2021). LICALITY – Likelihood and Criticality: Vulnerability Risk Prioritization Through Logical Reasoning and Deep Learning. *IEEE Transactions on Network and Service Management*, pp.1–1. doi:https://doi.org/10.1109/tnsm.2021.3133811.

[151] Yin, J., Tang, M., Cao, J. and Wang, H. (2020). Apply transfer learning to cybersecurity: Predicting exploitability of vulnerabilities by description. *Knowledge-Based Systems*, 210, p.106529. doi:https://doi.org/10.1016/j.knosys.2020.106529.

[152] Wang, T., Tang, M., Cao, J., Wang, H. and You, M. (2021). A real-time dynamic concept adaptive learning algorithm for exploitability prediction. 472, pp.252–265. doi:https://doi.org/10.1016/j.neucom.2021.01.144.

[153] Choufani, S., Shuman, C. and Weksberg, R. (2010). Beckwith-Wiedemann syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 154C(3), pp.343–354. doi:https://doi.org/10.1002/ajmg.c.30267.

[154] Bai, T. *et al.* (2021) "RDP-based Lateral Movement detection using Machine Learning," *Computer communications*, 165, pp. 9–19. Available at: https://doi.org/10.1016/j.comcom.2020.10.013.

[155] Afzaliseresht, N., Miao, Y., Michalska, S., Liu, Q. and Wang, H. (2020). From logs to Stories: Human-Centred Data Mining for Cyber Threat Intelligence. *IEEE Access*, 8, pp.19089–19099. doi:https://doi.org/10.1109/access.2020.2966760.

[156] de la Torre-Abaitua, G., Lago-Fernández, L.F. and Arroyo, D. (2021). A Compression-Based Method for Detecting Anomalies in Textual Data. *Entropy*, 23(5), p.618. doi:https://doi.org/10.3390/e23050618.

[157] Eljasik-Swoboda, T. and Demuth, W. (2020). Leveraging Clustering and Natural Language Processing to Overcome Variety Issues in Log Management. *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. [online] doi:https://doi.org/10.5220/0008856602810288.

[158] Dimitrios Sisiaridis and Olivier Markowitch (2018). Reducing Data Complexity in Feature Extraction and Feature Selection for Big Data Security Analytics. doi:https://doi.org/10.1109/icdis.2018.00014.

[159] Wang, Z. and Li, X. (2013). Intrusion Prevention System Design. *Lecture Notes in Electrical Engineering*, pp.375–382. doi:https://doi.org/10.1007/978-1-4471-4847-0_47.

[160] Freitas De Araujo-Filho, P., Pinheiro, A.J., Kaddoum, G., Campelo, D.R. and Soares, F.L. (2021). An Efficient Intrusion Prevention System for CAN: Hindering Cyber-Attacks With a Low-Cost Platform. *IEEE Access*, 9, pp.166855–166869. doi: https://doi.org/10.1109/access.2021.3136147.

[161] Constantinides, C., Shiaeles, S., Ghita, B. and Kolokotronis, N. (2019). *A Novel Online Incremental Learning Intrusion Prevention System*. [online] IEEE Xplore. doi:https://doi.org/10.1109/NTMS.2019.8763842.

[162] Furao, S., Ogura, T. and Hasegawa, O. (2007). An enhanced self-organizing incremental neural network for online unsupervised learning. *Neural Networks*, 20(8), pp.893–903. doi:https://doi.org/10.1016/j.neunet.2007.07.008.

[163] Wu, J.-C., Lu, S., Fuh, C.-S. and Liu, T.-L. (2021). One-class anomaly detection via novelty normalization. *Computer Vision and Image Understanding*, 210, p.103226. doi:https://doi.org/10.1016/j.cviu.2021.103226.

[164] Le, D.C., Nur Zincir-Heywood, A. and Heywood, M.I. (2016). Data analytics on network traffic flows for botnet behaviour detection. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. doi:https://doi.org/10.1109/ssci.2016.7850078.

[165] Saveetha, D. and Maragatham, G. (2022). Design of Blockchain enabled intrusion detection model for detecting security attacks using deep learning. *Pattern Recognition Letters*, 153, pp.24–28. doi:https://doi.org/10.1016/j.patrec.2021.11.023.

[166] Zhang, Y., Wang, L., Sun, W., Green II, R.C. and Alam, M. (2011). Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids. *IEEE Transactions on Smart Grid*, 2(4), pp.796–808. doi:https://doi.org/10.1109/tsg.2011.2159818.

[167] Nguyen, G.N., Viet, N.H.L., Elhoseny, M., Shankar, K., Gupta, B.B., A, A. and El-Latif, A. eds., (2021). Secure blockchain enabled Cyber-physical systems in healthcare using deep belief network with ResNet model. *Journal of Parallel and Distributed Computing*. [online] doi:https://doi.org/10.1016/j.jpdc.2021.03.011.

[168] Alhowaide, A., Alsmadi, I. and Tang, J. (2021). Ensemble Detection Model for IoT IDS. *Internet of Things*, p.100435. doi:https://doi.org/10.1016/j.iot.2021.100435.

[169] Binbusayyis, A. and Vaiyapuri, T. (2021). Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM. *Applied Intelligence*. doi:https://doi.org/10.1007/s10489-021-02205-9.

[170] Dutta, V., Michał Choraś, Kozik, R. and Pawlicki, M. (2020). Hybrid Model for Improving the Classification Effectiveness of Network Intrusion Detection. *Advances in intelligent systems and computing*, pp.405–414. doi:https://doi.org/10.1007/978-3-030-57805-3_38.

[171] Catillo, M., Pecchia, A. and Villano, U. (2022). AutoLog: Anomaly detection by deep autoencoding of system logs. *Expert Systems with Applications*, 191, p.116263. doi:https://doi.org/10.1016/j.eswa.2021.116263.

[172] Liu, H., Zhong, C., Alnusair, A. and Islam, S.R. (2021). FAIXID: A Framework for Enhancing AI Explainability of Intrusion Detection Results Using Data Cleaning Techniques. *Journal of Network and Systems Management*, 29(4). doi:https://doi.org/10.1007/s10922-021-09606-8.

[173] Maestre Vidal, J., Sotelo Monge, M.A. and Monterrubio, S.M.M. (2020). EsPADA: Enhanced Payload Analyzer for malware Detection robust against Adversarial threats. *Future Generation Computer Systems*, 104, pp.159–173. doi:https://doi.org/10.1016/j.future.2019.10.022.

[174] Latif, S., Huma, Z. e, Jamal, S.S., Ahmed, F., Ahmad, J., Zahid, A., Dashtipour, K., Umar Aftab, M., Ahmad, M. and Abbasi, Q.H. (2021). Intrusion Detection Framework for the Internet of Things using a Dense Random Neural Network. *IEEE Transactions on Industrial Informatics*, [online] pp.1–1. doi:https://doi.org/10.1109/TII.2021.3130248.

[175] Farooq, U., Marrakchi, Z. and Mehrez, H. (2012). FPGA Architectures: An Overview. *Tree-based Heterogeneous FPGA Architectures*, [online] pp.7–48. doi:https://doi.org/10.1007/978-1-4614-3594-5_2.

[176] Iwendi, C., Rehman, S.U., Javed, A.R., Khan, S. and Srivastava, G. (2021). Sustainable Security for the Internet of Things Using Artificial Intelligence Architectures. *ACM Transactions on Internet Technology*, 21(3), pp.1–22. doi:https://doi.org/10.1145/3448614.

[177] Petros Toupas, Dimitra Chamou, Giannoutakis, K.M., Anastasios Drosou and Dimitrios Tzovaras (2019). An Intrusion Detection System for Multi-class Classification Based on Deep Neural Networks. [online] doi:https://doi.org/10.1109/icmla.2019.00206.

[178] D'hooge, L., Verkerken, M., Wauters, T., Volckaert, B. and De Turck, F. (2021). Hierarchical feature block ranking for data-efficient intrusion detection modeling. *Computer Networks*, 201, p.108613. doi:https://doi.org/10.1016/j.comnet.2021.108613.

[179] Huang, S. and Lei, K. (2020). IGAN-IDS: An Imbalanced Generative Adversarial Network towards Intrusion Detection System in Ad-hoc Networks. *Ad Hoc Networks*, p.102177. doi:https://doi.org/10.1016/j.adhoc.2020.102177.

[180] Jagtap, S.S., V. S., S.S. and V., S. (2021). A hypergraph based Kohonen map for detecting intrusions over cyber–physical systems traffic. *Future Generation Computer Systems*, 119, pp.84–109. doi:https://doi.org/10.1016/j.future.2021.02.001.

[181] Liu, J., Kantarci, B. and Adams, C. (2020). Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset. *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning*. doi:https://doi.org/10.1145/3395352.3402621.

[182] Pawlicki, M., Kozik, R. and Michał Choraś (2019). Artificial Neural Network Hyperparameter Optimisation for Network Intrusion Detection. *Lecture notes in computer science*, pp.749–760. doi:https://doi.org/10.1007/978-3-030-26763-6_72.

[183] Shafiq, M., Tian, Z., Bashir, A.K., Du, X. and Guizani, M. (2020). IoT malicious traffic identification using wrapper-based feature selection mechanisms. *Computers & Security*, 94, p.101863. doi:https://doi.org/10.1016/j.cose.2020.101863.

[184] Mikhail, J.W., Fossaceca, J.M. and Iammartino, R. (2019). A Semi-Boosted Nested Model With Sensitivity-Based Weighted Binarization for Multi-Domain Network Intrusion Detection. 10(3), pp.1–27. doi:https://doi.org/10.1145/3313778.

[185] Gupta, N., Jindal, V. and Bedi, P. (2021). LIO-IDS: Handling class imbalance using LSTM and improved one-vs-one technique in intrusion detection system. *Computer Networks*, 192, p.108076. doi:https://doi.org/10.1016/j.comnet.2021.108076.

[186] Li, G., Shen, Y., Zhao, P., Lu, X., Liu, J., Liu, Y. and Hoi, S.C. H. (2019). Detecting cyberattacks in industrial control systems using online learning algorithms. *Neurocomputing*, 364, pp.338–348. doi:https://doi.org/10.1016/j.neucom.2019.07.031.

[187] Zhang, J., Xu, S., Hamad, K.I., Jasim, A.M. and Xing, Y. (2019). High retention rate NCA cathode powders from spray drying and flame assisted spray pyrolysis using glycerol as the solvent. *Powder Technology*, [online] 363, pp.1–6. doi:https://doi.org/10.1016/j.powtec.2019.12.057.

[188] Ieracitano, C., Adeel, A., Morabito, F.C. and Hussain, A. (2020). A novel statistical analysis and autoencoder driven intelligent intrusion detection approach. *Neurocomputing*, 387, pp.51–62. doi:https://doi.org/10.1016/j.neucom.2019.11.016.

[189] Liu, Q., Wang, D., Jia, Y., Luo, S. and Wang, C. (2021). A multi-task based deep learning approach for intrusion detection. *Knowledge-Based Systems*, p.107852. doi:https://doi.org/10.1016/j.knosys.2021.107852.

[190] Nikoloudakis, Y., Kefaloukos, I., Klados, S., Panagiotakis, S., Pallis, E., Skianis, C. and Markakis, E.K. (2021). Towards a Machine Learning Based Situational Awareness Framework for Cybersecurity: An SDN Implementation. *Sensors*, 21(14), p.4939. doi:https://doi.org/10.3390/s21144939.

[191] Zhang, F., Kodituwakku, H.A.D.E., Hines, J.W. and Coble, J. (2019). Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data. *IEEE Transactions on Industrial Informatics*, [online] 15(7), pp.4362–4369. doi:https://doi.org/10.1109/tii.2019.2891261.

[192] Kim, H.K., Im, K.H. and Park, S.C. (2010). DSS for computer security incident response applying CBR and collaborative response. *Expert Systems with Applications*, 37(1), pp.852–870. doi:https://doi.org/10.1016/j.eswa.2009.05.100.

[193] Jiang, F., Gu, T., Chang, L. and Xu, Z. (2014). Case Retrieval for Network Security Emergency Response Based on Description Logic. *Lecture notes in computer science*, pp.284–293. doi:https://doi.org/10.1007/978-3-662-44980-6_32.

[194] Kraeva, I. and Yakhyaeva, G. (2021). *Application of the Metric Learning for Security Incident Playbook Recommendation*. [online] IEEE Xplore. doi:https://doi.org/10.1109/EDM52169.2021.9507632.

[195] Ping, N.L., None Yu Haifeng and None Ma Guoqing (2010). An incident response decision support system based on CBR and ontology. [online] pp.V11-340. doi:https://doi.org/10.1109/iccasm.2010.5623194.

[196] Noemí DeCastro-García, Muñoz, L. and Fernández-Rodríguez, M. (2020). Machine learning for automatic assignment of the severity of cybersecurity events. 2(1). doi:https://doi.org/10.1002/cmm4.1072.

[197] NIST (2024). The NIST Cybersecurity Framework (CSF) 2.0. *The NIST Cybersecurity Framework (CSF) 2.0*, [online] 2.0(29). doi:https://doi.org/10.6028/nist.cswp.29.

[198] NIST (2023). NIST Drafts Major Update to Its Widely Used Cybersecurity Framework. *NIST*. [online] Available at: https://www.nist.gov/news-events/news/2023/08/nist-drafts-major-update-its-widely-used-cybersecurity-framework.

[199] Bresnahan, E. (2023). *NIST Cybersecurity Framework Core Explained*. [online] www.cybersaint.io. Available at: https://www.cybersaint.io/blog/nist-cybersecurity-framework-core-explained.

[200] Greaves, R. (2024). *Trends in InfoSec: Data Minimization, Autoclassification, and Ethical AI*. [online] InfoQ. Available at: https://www.infoq.com/presentations/trends-infosec/ [Accessed 21 Jan. 2025].