ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«Έξυπνη Ασφάλεια: Προσεγγίσεις της Τεχνητής Νοημοσύνης στον Εντοπισμό Εισβολών και την Κατανομή Πόρων σε IoT Οικοσυστήματα»

Της φοιτήτριας
Ραφαηλία Ντόντογλου
Αρ. Μητρώου: 164701

Επιβλέπων
Δημήτριος Αμανατιάδης

Μάιος 2025

Έξυπνη Ασφάλεια: Προσεγγίσεις της Τεχνητής Νοημοσύνης στον Εντοπισμό Εισβολών και την Κατανομή Πόρων σε Οικοσυστήματα

24276

Ραφαηλία Ντόντογλου
Δημήτριος Αμανατιάδης

Ημερομηνία ανάληψης Π.Ε. 24 Οκτωβρίου 2024
Ημερομηνία περάτωσης Π.Ε. 30 Μαΐου 2025

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Πρόλογος

Ο αριθμός των συσκευών Internet of Things (IoT) ξεπερνάει πλέον τα 30 δισεκατομμύρια παγκοσμίως. Έτσι, η ανάγκη για ασφαλή και αποδοτική λειτουργία είναι επιτακτική. Η παρούσα πτυχιακή εργασία, "Έξυπνη Ασφάλεια: Προσεγγίσεις της Τεχνητής Νοημοσύνης στον Εντοπισμό Εισβολών και την Κατανομή Πόρων σε IoT Οικοσυστήματα", εστιάζει σε αυτές τις προκλήσεις, διερευνώντας τις δυνατότητες της Τεχνητής Νοημοσύνης τόσο στην ενίσχυση της ασφάλειας των δικτύων IoT, όσο και στη βελτιστοποίηση της χρήσης πόρων των συσκευών.

Οι παραδοσιακές μέθοδοι ασφάλειας δεν καλύπτουν πλήρως τα πολύπλοκα και δυναμικά IoT περιβάλλοντα, ενώ η ανίχνευση εισβολών με χρήση τεχνικών που βασίζονται σε Τεχνητή Νοημοσύνη μπορεί να προσαρμοστεί ώστε να αναγνωρίζει απειλές σε δίκτυα με διαφορετικά είδη συσκευών. Επιπρόσθετα, ο διαμοιρασμός πόρων βασιζόμενος σε Τεχνητή Νοημοσύνη είναι απαραίτητος για την επέκταση της λειτουργίας των συσκευών IoT, που τυπικά έχουν περιορισμένες υπολογιστικές δυνατότητες και μειωμένη αυτονομία ενέργειας.

Η μελέτη και η συγγραφή της εργασίας όχι μόνο συμβαδίζει με τις σπουδές μου και όσα έχω μάθει ως τώρα στα δίκτυα και την ασφάλεια, αλλά ενισχύει και τις γνώσεις και τις ικανότητές μου στην Τεχνητή Νοημοσύνη, την κυβερνοασφάλεια και τη διαχείριση δικτύου, τομείς που αποτελούν θεμέλιο για μία σύγχρονη επαγγελματική σταδιοδρομία στον χώρο της τεχνολογίας.

Περίληψη

Αυτή η έρευνα εστιάζει την ενσωμάτωση της Τεχνητής Νοημοσύνης (AI) στην ενίσχυση της ασφάλειας μέσα σε οικοσυστήματα Internet of Things (IoT). Εστιάζει σε δύο βασικές περιοχές: τον εντοπισμό εισβολών και το διαμοιρασμό πηγών.

Θα μελετήσουμε πως οι προσεγγίσεις με Τεχνητή Νοημοσύνη βοηθούν στην ανάπτυξη συστημάτων Intrusion Detection, που μπορούν να αναγνωρίσουν και να αποκριθούν σε απειλές σε πραγματικό χρόνο, με απώτερο στόχο τη βελτίωση της ακρίβειας και της απόδοσης του εντοπισμού απειλών, τη μείωση false positives και την βελτίωση χρόνων απόκρισης.

Επιπρόσθετα, αυτή η διπλωματική εργασία ερευνά μεθοδολογίες Τεχνητής Νοημοσύνης για δυναμικό διαμοιρασμό πηγών, εξασφαλίζοντας μέγιστη απόδοση και ασφάλεια σε δίκτυα IoT. Αναλύοντας μοτίβα χρήσης και τις απαιτήσεις συστήματος, αυτές οι προσεγγίσεις στοχεύουν στο να διαμοιράσουν αποδοτικά τις πηγές και να μειώσουν το ρίσκο υπερφόρτωσης.

Intelligent Safety: AI – Driven Approaches to Intrusion Detection and Resource Allocation in IoT Ecosystems

Rafailia Ntontoglou

Abstract

The rapid expansion of the Internet of Things (IoT) has revolutionized connectivity, influencing critical aspects of daily life. Millions of IoT devices handle data continuously, a process that significantly increases their vulnerability to cyberattacks. At the same time, they require efficient management of computational resources such as energy, processing power, and memory to maintain optimal performance. Ensuring both security and resource efficiency is essential for the reliable operation of IoT systems.

This thesis presents a literature review on the use of Artificial Intelligence (AI) for intrusion detection and dynamic resource allocation in IoT environments. It explores both traditional and modern techniques and evaluates their effectiveness in enhancing the security and efficiency of IoT networks.

Through detailed analysis, the study identifies existing research gaps and unresolved challenges. Finally, it highlights directions for future research, focusing on the development of adaptive, resilient, and energy-efficient solutions that can address the continuously evolving demands of IoT systems.

Ευχαριστίες

Θα ήθελα να εκφράσω τις ειλικρινείς και θερμές μου ευχαριστίες σε όλους όσους με στήριξαν σε όλη την πορεία ολοκλήρωσης αυτής της πτυχιακής εργασίας. Πρώτα απ' όλα, ευχαριστώ τον κύριο Δημήτριο Αμανατιάδη, για την πολύτιμη καθοδήγηση και υποστήριξή του. Οι γνώσεις και τα σχόλιά του ήταν καθοριστικά. Επίσης, θα ήθελα να ευχαριστήσω τον κύριο Περικλή Χατζημίσιο, που ανέλαβε την επίβλεψη της πτυχιακής και συνέβαλε στην ολοκλήρωσή της.

Θα ήθελα επίσης να ευχαριστήσω τους φίλους και συμφοιτητές μου, των οποίων η υποστήριξη και η συντροφικότητα έκαναν αυτή την εμπειρία πραγματικά αξέχαστη. Η ενθάρρυνση και οι συμβουλές τους με βοήθησαν να αντιμετωπίσω τις προκλήσεις και να συνεχίσω να έχω κίνητρο.

Τέλος, ένα ιδιαίτερο ευχαριστώ στην οικογένειά μου για την ακλόνητη πίστη σε εμένα και τη συνεχή υποστήριξή τους, η οποία ήταν το θεμέλιό μου σε κάθε βήμα αυτού του ταξιδιού.

Σας ευχαριστώ όλους για τη συμβολή σας σε αυτό το επίτευγμα.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Συντομογραφίες

IoT	Internet of Things
AI	Artificial Intelligence
IIoT	Industrial Internet of Things
DoS	Denial of Service
DDoS	Distributed Denial of Service
MitM	Man-in-the-Middle
BLE	Bluetooth Low Energy
APIs	Application Programming Interfaces
NFC	Near-Field Communication
LPWAN	Low-Power, Wide-Area Network
GSM	Global System for Mobile Communication
IDS	Intrusion Detection System
HIDS	Host-based Intrusion Detection System
NIDS	Network-based Intrusion Detection System
ML	Machine Learning
DL	Deep Learning
SVM	Support Vector Machines
FAR	False Alarm Rate
DT	Decision Tree
RF	Random Forest
ReLU	Rectified Linear Unit
ANNs	Artificial Neural Networks
CNNs	Convolutional Neural Networks
RNNs	Recurrent Neural Networks
BPTT	Backpropagation Through Time
LSTM	Long Short-Term Memory
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Oversampling Technique
GRUs	Gated Recurrent Units
AEs	Autoencoders

DAEs	Denoising Autoencoders
FCFS	First-Come-First-Served
RL	Reinforcement Learning
SARSA	State-Action-Reward-State-Action
DRL	Deep Reinforcement Learning
PPO2	Proximal Policy Optimization 2
FL	Federated Learning
MLP	Multi-Layer Perceptron
FRL	Federated Reinforcement Learning
DQN	Deep Q-Network

Κεφάλαιο 1ο: Από το Διαδίκτυο στο IoT

1.1 Εισαγωγή

Σε αυτό το κεφάλαιο παρουσιάζεται η εξέλιξη του Διαδικτύου (Internet) και η μετάβασή του στο Διαδίκτυο των Πραγμάτων (Internet of Things – IoT). Αρχικά, γίνεται μια σύντομη αναδρομή στη δημιουργία και την ανάπτυξη των τεχνολογιών δικτύωσης, περιγράφοντας πώς το Διαδίκτυο επεκτάθηκε πέρα από τους υπολογιστές και τα κινητά τηλέφωνα. Στη συνέχεια, εξετάζεται η έννοια του IoT και η παρουσίαση των προκλήσεων ασφάλειας που προκύπτουν σε τέτοια περιβάλλοντα. Έπειτα, γίνεται ανάλυση των περιορισμών των παραδοσιακών μεθόδων προστασίας και η ανάγκη αναζήτησης νέων προσεγγίσεων. Τέλος, εισάγεται η Τεχνητή Νοημοσύνη ως τεχνολογία που μπορεί να υποστηρίξει αποτελεσματικά την ανίχνευση εισβολών και την κατανομή πόρων σε κατανεμημένα IoT συστήματα.

1.2 IoT: Η εξέλιξη του Διαδικτύου

Η τεχνολογική εξέλιξη των τελευταίων δεκαετιών έχει αλλάξει ριζικά τον τρόπο με τον οποίο οι συσκευές επικοινωνούν και ανταλλάσσουν δεδομένα. Από τα πρώτα δικτυακά συστήματα, που είχαν σχεδιαστεί για τη σύνδεση υπολογιστών, η εξέλιξη της τεχνολογίας οδήγησε σε ένα περιβάλλον όπου η διασύνδεση επεκτείνεται σε κάθε μορφή ψηφιακής συσκευής.

Το Διαδίκτυο ξεκίνησε ως ένα δίκτυο επικοινωνίας μεταξύ απομακρυσμένων υπολογιστών, επιτρέποντας την ανταλλαγή δεδομένων σε παγκόσμια κλίμακα. Η εισαγωγή των πρωτοκόλλων TCP/IP και η ανάπτυξη του World Wide Web τη δεκαετία του 1990, συνέβαλαν καθοριστικά στη ευρεία διάδοση του Διαδικτύου και στη σταδιακή ενσωμάτωσή του στην καθημερινή ζωή [1]. Η επόμενη φάση της τεχνολογικής εξέλιξης ήρθε με την πρόοδο στις ασύρματες επικοινωνίες, τη μικροηλεκτρονική και τη μειωμένη ενεργειακή κατανάλωση των συσκευών, που επέτρεψαν την ανάπτυξη ολοένα και μικρότερων, φθηνότερων και αποδοτικότερων αισθητήρων [2]. Η εξάπλωση των δικτύων κινητής τηλεφωνίας, του Wi-Fi και των υποδομών cloud, σε συνδυασμό με την αυξημένη υπολογιστική ισχύ και τη διαθεσιμότητα μεγάλων ποσοτήτων δεδομένων, δημιούργησαν τις κατάλληλες συνθήκες για την εμφάνιση του Διαδικτύου των Πραγμάτων (Internet of Things – IoT) [3]. Το IoT αποτελεί ένα δίκτυο διασυνδεδεμένων φυσικών συσκευών που ενσωματώνουν αισθητήρες, λογισμικό και δυνατότητες επικοινωνίας, επιτρέποντας την ανταλλαγή δεδομένων μέσω του Διαδικτύου. Βασικός του στόχος είναι η παρακολούθηση, ο έλεγχος και η αυτοματοποίηση φυσικών διεργασιών με ελάχιστη ανθρώπινη παρέμβαση, μετατρέποντας το φυσικό περιβάλλον σε ένα συνεχώς εξελισσόμενο ψηφιακό οικοσύστημα [1].

1.3 Ζητήματα ασφάλειας στα σύγχρονα δίκτυα IoT

Κατά την τελευταία δεκαετία, τα οικοσυστήματα IoT έχουν εξελιχθεί από απλές, απομακρυσμένες εφαρμογές σε περίπλοκα, κατανεμημένα δίκτυα, που περιλαμβάνουν τεράστιο αριθμό συσκευών, διαφορετικών τύπων και κατασκευαστών. Οι συσκευές αυτές λειτουργούν με συνεχή συνδεσιμότητα, είναι διάσπαρτες σε φυσικούς και μη ελεγχόμενους χώρους και αλληλεπιδρούν δυναμικά με το περιβάλλον [3]. Η ετερογένεια του εξοπλισμού, η διάχυση στο φυσικό χώρο, η πολυπλοκότητα των ίδιων των συσκευών και ο τεράστιος όγκος δεδομένων που διακινούνται σε πραγματικό χρόνο, συνθέτουν ένα περιβάλλον όπου η ασφάλεια καθίσταται ιδιαίτερα δύσκολη υπόθεση. Η ανάγκη για προστασία των IoT υποδομών είναι πιο επιτακτική από ποτέ, καθώς κάθε συσκευή μπορεί να

αποτελέσει δυνητικό σημείο εισόδου για επιθέσεις, επηρεάζοντας όχι μόνο την ακεραιότητα του ίδιου του συστήματος, αλλά και των δεδομένων και των χρηστών που το αξιοποιούν [2].

1.4 Περιορισμοί των παραδοσιακών προσεγγίσεων ασφάλειας

Οι παραδοσιακές προσεγγίσεις ασφάλειας σχεδιάστηκαν με βάση υπολογιστικά περιβάλλοντα που διέθεταν σταθερή τοπολογία, σαφώς ορισμένα όρια, καθώς και επαρκή υπολογιστική ισχύ, αποθηκευτική ικανότητα και ενεργειακούς πόρους [4]. Τεχνικές όπως τα firewalls, η κρυπτογράφηση δεδομένων, οι μηχανισμοί ελέγχου ταυτότητας και τα συστήματα ανίχνευσης κακόβουλου λογισμικού αναπτύχθηκαν για να λειτουργούν αποτελεσματικά σε τέτοιες συνθήκες. Ωστόσο, στο πλαίσιο των IoT οικοσυστημάτων, οι ίδιες αυτές τεχνικές αποδεικνύονται ανεπαρκείς ή ακόμη και μη εφαρμόσιμες. Οι περισσότεροι κόμβοι του IoT διαθέτουν περιορισμένη υπολογιστική ισχύ, μνήμη και ενεργειακούς περιορισμούς, γεγονός που καθιστά δύσκολη την εφαρμογή πολύπλοκων και ενεργειακά απαιτητικών μηχανισμών ασφάλειας. Επιπλέον, πολλά συστήματα IoT βασίζονται σε απλοποιημένα ή προσαρμοσμένα πρωτόκολλα επικοινωνίας, που δεν υποστηρίζουν ενσωματωμένα χαρακτηριστικά ασφαλείας, αφήνοντας σημαντικές ευπάθειες ανοιχτές σε εκμετάλλευση. Η αδυναμία εφαρμογής των παραδοσιακών λύσεων, σε συνδυασμό με τις ιδιαιτερότητες των IoT συστημάτων, υποδεικνύει την ανάγκη για νέες μεθόδους ασφάλειας, ικανές να προσαρμόζονται σε πραγματικό χρόνο και να λειτουργούν αποδοτικά μέσα σε κατανεμημένα, ετερογενή περιβάλλοντα με περιορισμένους πόρους [3].

1.5 Η Τεχνητή Νοημοσύνη ως εργαλείο για την ασφάλεια και την κατανομή πόρων

Η συνεχής αύξηση της πολυπλοκότητας στα IoT οικοσυστήματα, σε συνδυασμό με τον ρυθμό των παραγόμενων δεδομένων και την ανάγκη για άμεση απόκριση, έχει αναδείξει την Τεχνητή Νοημοσύνη (Artificial Intelligence – AI) ως μία από τις πιο υποσχόμενες τεχνολογίες για την ενίσχυση της ασφάλειας και τη βελτιστοποίηση της λειτουργίας αυτών των συστημάτων. Η AI, αξιοποιώντας αλγόριθμους Μηχανικής και Βαθιάς Μάθησης, προσφέρει την ικανότητα εντοπισμού μοτίβων, πρόβλεψης απειλών και λήψης αποφάσεων σε πραγματικό χρόνο, χωρίς την ανάγκη ανθρώπινης παρέμβασης. Συνεπώς, θεωρείται ιδανική για περιβάλλοντα που μεταβάλλονται δυναμικά και απαιτούν "έξυπνη", αυτοματοποιημένη διαχείριση [5].

Η Τεχνητή Νοημοσύνη έχει αναδειχθεί σε βασικό εργαλείο για την ενίσχυση της ασφάλειας στα οικοσυστήματα του IoT, με ιδιαίτερη έμφαση στον εντοπισμό εισβολών. Σε περιβάλλοντα όπου παράγονται και μεταφέρονται τεράστιοι όγκοι δεδομένων σε πραγματικό χρόνο, η παρακολούθηση της κυκλοφορίας με στόχο τον εντοπισμό απειλών αποτελεί ιδιαίτερα απαιτητική διαδικασία. Τα παραδοσιακά συστήματα εντοπισμού απειλών παρουσιάζουν περιορισμένη αποτελεσματικότητα απέναντι σε σύνθετες ή άγνωστες μορφές κακόβουλης συμπεριφοράς, ιδιαίτερα σε συνδυασμό με την ολοένα αυξανόμενη πολυπλοκότητα και ποικιλομορφία των επιθέσεων. Αντίθετα, τεχνικές Μηχανικής και Βαθιάς Μάθησης επιτρέπουν την αυτόματη ανάλυση των χαρακτηριστικών της δικτυακής δραστηριότητας και την ανίχνευση παρεκκλίσεων από τη φυσιολογική λειτουργία, χωρίς να απαιτείται προηγούμενη γνώση του τύπου της επίθεσης [6]. Καθώς τα μοντέλα εκπαιδεύονται συνεχώς σε νέα δεδομένα και προσαρμόζονται σε εξελισσόμενες συνθήκες, η χρήση της AI συμβάλλει στη δημιουργία ευέλικτων και ανθεκτικών συστημάτων, που μπορούν να εντοπίζουν και να αποκρούουν απειλές σε πραγματικό χρόνο, ενισχύοντας την ασφάλεια του δικτύου συνολικά. Η συνεχής αυτή προσαρμογή δίνει στα συστήματα τη δυνατότητα να μαθαίνουν από την εμπειρία και να βελτιώνουν την ικανότητά τους στην ανίχνευση νέων, άγνωστων απειλών [4].

Πέρα από την ανίχνευση απειλών, η Τεχνητή Νοημοσύνη μπορεί να αξιοποιηθεί και για την κατανομή των περιορισμένων πόρων των IoT συστημάτων, όπως η ενέργεια, η υπολογιστική ισχύς και το εύρος ζώνης [7]. Σε τέτοια περιβάλλοντα, όπου εκατοντάδες ή χιλιάδες συσκευές λειτουργούν ταυτόχρονα και ανταγωνίζονται για πρόσβαση σε κοινά μέσα, είναι απαραίτητος ο καθορισμός μηχανισμών που αποφασίζουν την κατανομή των διαθέσιμων πόρων με τρόπο αποδοτικό και ευέλικτο. Οι αποφάσεις αυτές εξαρτώνται από την τρέχουσα κατάσταση του συστήματος και μεταβάλλονται συνεχώς, καθώς διαφοροποιούνται οι απαιτήσεις των εφαρμογών, η δραστηριότητα των συσκευών και ο συνολικός φόρτος του δικτύου. Η Τεχνητή Νοημοσύνη επιτρέπει τη δυναμική προσαρμογή αυτών των αποφάσεων, μέσα από την ανάλυση των μοτίβων λειτουργίας και την εκμάθηση στρατηγικών διαχείρισης που οδηγούν σε αποφάσεις για καλύτερη απόδοση του δικτύου. Σε αντίθεση με τους παραδοσιακούς, στατικούς αλγορίθμους που ακολουθούν προκαθορισμένους κανόνες χωρίς δυνατότητα προσαρμογής, τα συστήματα που βασίζονται σε Τεχνητή Νοημοσύνη μπορούν να ενσωματώνουν ανατροφοδότηση από το περιβάλλον και να εξελίσσουν τον τρόπο κατανομής πόρων, ανάλογα με τις πραγματικές συνθήκες και τις ανάγκες του δικτύου. Μέσα από αυτή τη συνεχή διαδικασία λήψης αποφάσεων, τα συστήματα μαθαίνουν από την εμπειρία και προσαρμόζουν σταδιακά τις πολιτικές κατανομής που εφαρμόζουν.

Σε αυτή την εργασία, θα μελετηθούν μοντέλα εντοπισμού εισβολών και κατανομής πόρων που βασίζονται σε Τεχνητή Νοημοσύνη και εφαρμόζονται σε περιβάλλοντα IoT. Θα εξεταστεί ο τρόπος λειτουργίας τους, οι τεχνολογίες που τα υποστηρίζουν, καθώς και οι προκλήσεις που προκύπτουν από την ενσωμάτωσή τους σε ετερογενή, κατανεμημένα και περιορισμένων πόρων οικοσυστήματα. Στόχος της εργασίας είναι να αναδειχθούν οι τρόποι με τους οποίους η AI μπορεί να αξιοποιηθεί πρακτικά για τη βελτίωση της ασφάλειας και της λειτουργικότητας στο IoT.

1.6 Επίλογος

Σε αυτό το κεφάλαιο πραγματοποιήθηκε μια συνοπτική ανασκόπηση της μετάβασης από το Διαδίκτυο στο IoT. Καταγράφηκε ο ρόλος της τεχνολογικής εξέλιξης στην επέκταση της συνδεσιμότητας πέρα από τους υπολογιστές, παρουσιάστηκαν οι βασικές έννοιες του IoT, καθώς και οι νέες απαιτήσεις που προκύπτουν σε επίπεδο ασφάλειας. Τονίστηκε η αδυναμία των παραδοσιακών μηχανισμών προστασίας να ανταποκριθούν στις ιδιαιτερότητες των IoT οικοσυστημάτων και εισήχθη η Τεχνητή Νοημοσύνη ως σύγχρονη προσέγγιση για την ανίχνευση απειλών και τη βελτιστοποίηση της κατανομής πόρων.

Κεφάλαιο 2ο: Βασικές Αρχές και Εφαρμογές του IoT

2.1 Εισαγωγή

Σε αυτό το κεφάλαιο εξετάζονται τα βασικά στοιχεία που συνθέτουν το περιβάλλον του IoT, με σκοπό να κατανοηθεί το πώς λειτουργεί και το που εφαρμόζεται. Αρχικά, δίνεται ένας σαφής ορισμός της έννοιας "Οικοσύστημα IoT" και έπειτα παρουσιάζονται οι βασικές εφαρμογές του IoT σε διάφορους τομείς της καθημερινότητας, όπως τα έξυπνα σπίτια, η υγειονομική περίθαλψη και η βιομηχανία. Ακολουθεί η αναλυτική περιγραφή των αρχιτεκτονικών που έχουν προταθεί, όπως το μοντέλο τριών ή τεσσάρων επιπέδων. Τέλος, εξετάζονται τα βασικά πρωτόκολλα επικοινωνίας που χρησιμοποιούνται στο IoT, με έμφαση στα χαρακτηριστικά και τις χρήσεις τους.

2.2 Ορισμός του Οικοσυστήματος IoT

Ένα οικοσύστημα IoT αποτελείται από ένα εκτεταμένο δίκτυο διασυνδεδεμένων έξυπνων συσκευών, αισθητήρων, λογισμικού και εφαρμογών, τα οποία επικοινωνούν μεταξύ τους και με το περιβάλλον, ανταλλάσσοντας δεδομένα μέσω ενσύρματων ή ασύρματων δικτύων [4]. Αυτά τα συστήματα επιτρέπουν την αυτοματοποιημένη συλλογή δεδομένων, τη λήψη αποφάσεων σε πραγματικό χρόνο και τη βελτιστοποίηση πόρων σε διάφορους τομείς [7]. Οι συσκευές διαθέτουν ενσωματωμένους αισθητήρες, λογισμικό και δυνατότητες συνδεσιμότητας, επιτρέποντας τη συνεχή παρακολούθηση, την ανάλυση και την ανταλλαγή δεδομένων [8].

2.3 Εφαρμογές του IoT

Το IoT έχει περάσει από το στάδιο της θεωρίας στην πράξη και χρησιμοποιείται πλέον σε πολλούς τομείς της καθημερινής ζωής. Το IoT επιτρέπει τη σύνδεση φυσικών αντικειμένων με το διαδίκτυο, δημιουργώντας περιβάλλοντα όπου οι συσκευές συλλέγουν δεδομένα και «επικοινωνούν» μεταξύ τους [8]. Μέσα από αυτή τη συνδεσιμότητα, προκύπτουν λύσεις που κάνουν διάφορες διαδικασίες πιο απλές, πιο ασφαλείς και πιο αποτελεσματικές. Συνολικά, οι εφαρμογές του IoT επηρεάζουν ουσιαστικά τόσο την καθημερινότητα των πολιτών όσο και τη λειτουργία κρίσιμων υποδομών, αναδεικνύοντας τη σημασία του ως θεμέλιο της σύγχρονης τεχνολογικής εξέλιξης.

2.3.1 Έξυπνα σπίτια

Η ιδέα της έξυπνης κατοικίας (smart home) καλύπτει ένα ευρύ φάσμα εφαρμογών, που συμβάλλουν στην ενίσχυση της παραγωγικότητας, της άνεσης και της ποιότητας ζωής των χρηστών. Ενισχύει την ασφάλεια χρηστών και κατοικιών, ενώ παράλληλα βελτιστοποιεί την κατανάλωση ενέργειας. Πολλές σύγχρονες οικιακές συσκευές μπορούν να συνδέονται στο Διαδίκτυο και να παρέχουν έξυπνες υπηρεσίες, για παράδειγμα θερμοστάτες, κλιματιστικά, συστήματα ελέγχου φωτισμού, κάμερες ασφαλείας και τηλεοράσεις. Μέσω της σύνδεσης σε ένα κεντρικό δίκτυο, συλλέγουν πληροφορίες από το περιβάλλον, παρέχουν στο χρήστη τη δυνατότητα απομακρυσμένης διαχείρισης και εξατομίκευσης του χώρου. Για παράδειγμα, ένα έξυπνο σύστημα ασφαλείας μπορεί να εντοπίζει πότε ο χρήστης αποχωρεί από την κατοικία (μέσω γεωεντοπισμού ή ανίχνευσης απουσίας κίνησης) και να ενεργοποιεί αυτόματα το σύστημα συναγερμού, να κατεβάζει τα ρολά και να απενεργοποιεί τα φώτα. Παράλληλα, μπορεί να αποστέλλει ειδοποιήσεις σε πραγματικό χρόνο σε περίπτωση ανίχνευσης κίνησης ή ασυνήθιστης δραστηριότητας [9].

2.3.2 Έξυπνες πόλεις

Οι έξυπνες πόλεις (smart cities) βασίζονται στο IoT για τη συλλογή και ανάλυση δεδομένων σε πραγματικό χρόνο, με σκοπό τη βελτίωση της ποιότητας ζωής των κατοίκων αλλά και τη βέλτιστη διαχείριση των πόρων. Το IoT παίζει σημαντικό ρόλο σε τομείς όπως η ρύθμιση της κυκλοφορίας (όπου τα συστήματα ελέγχου προσαρμόζουν τα φανάρια ανάλογα με την κίνηση), η διαχείριση των απορριμμάτων (με αισθητήρες που παρακολουθούν την πληρότητα των κάδων), ο έξυπνος φωτισμός (που μειώνει την κατανάλωση ενέργειας αναλόγως), αλλά και ο έλεγχος στάθμευσης (μέσω εφαρμογών που υποδεικνύουν διαθέσιμες θέσεις) [10][7].

2.3.3 Υγειονομική περίθαλψη

Στον τομέα της υγείας, το IoT επιτρέπει την απομακρυσμένη παρακολούθηση ασθενών και την παροχή εξατομικευμένης φροντίδας, μειώνοντας την ανάγκη για φυσική παρουσία σε ιατρικές δομές. Ιατρικές συσκευές όπως έξυπνα βραχιόλια, αισθητήρες καρδιακών παλμών ή μετρητές σακχάρου συλλέγουν δεδομένα σε πραγματικό χρόνο και τα μεταδίδουν στους επαγγελματίες υγείας. Παράλληλα, το IoT διευκολύνει τον εντοπισμό και τη διαχείριση ιατρικού εξοπλισμού, όπως αναπηρικά αμαξίδια ή συσκευές οξυγόνου, βελτιώνοντας τη λειτουργικότητα και την απόκριση του συστήματος υγείας [11].

2.3.4 Γεωργία

Η γεωργία αποτελεί έναν από τους βασικούς τομείς όπου το IoT βρίσκει πρακτική εφαρμογή, ειδικά στο πλαίσιο της «έξυπνης γεωργίας» (smart farming). Μέσω αισθητήρων που παρακολουθούν παραμέτρους όπως η υγρασία του εδάφους, η κατάσταση των φυτών και οι καιρικές συνθήκες, οι αγρότες έχουν πρόσβαση σε πραγματικά δεδομένα που τους βοηθούν να λαμβάνουν πιο ακριβείς και έγκαιρες αποφάσεις. Για παράδειγμα, αισθητήρες που ανιχνεύουν επίπεδα χλωροφύλλης ή σημάδια προσβολής από έντομα μπορούν να ειδοποιήσουν τον παραγωγό εγκαίρως, πριν επεκταθεί το πρόβλημα στην καλλιέργεια. Αυτό οδηγεί σε αύξηση της απόδοσης των καλλιεργειών, μείωση της σπατάλης νερού και φυτοφαρμάκων και συνολικά καλύτερη διαχείριση των διαθέσιμων πόρων [7].

2.3.5 Βιομηχανία

Το Industrial Internet of Things (IIoT) αναφέρεται στην ενσωμάτωση τεχνολογιών IoT σε βιομηχανικά περιβάλλοντα, με στόχο τη βελτίωση της αποδοτικότητας, τη μείωση του κόστους και την ενίσχυση της ασφάλειας. Οι επιχειρήσεις χρησιμοποιούν αισθητήρες και συστήματα παρακολούθησης για τη συλλογή και ανάλυση δεδομένων σε πραγματικό χρόνο, επιτηρώντας έτσι κρίσιμες παραμέτρους και εντοπίζοντας δυσλειτουργίες πριν εξελιχθούν σε σοβαρά προβλήματα. Αυτό τους επιτρέπει να βελτιώνουν διαδικασίες όπως ο ποιοτικός έλεγχος, η διαχείριση αποθέματος, η προσαρμογή της παραγωγής σε μεταβαλλόμενες ανάγκες και η προληπτική συντήρηση. Για παράδειγμα, μια γραμμή παραγωγής μπορεί να σταματήσει αυτόματα όταν εντοπιστεί απόκλιση στην ποιότητα των προϊόντων, ώστε να αποφευχθεί η συνέχιση της παραγωγής με ελαττωματικό υλικό. Το IIoT συμβάλλει έτσι σε ένα πιο σταθερό, ελεγχόμενο και αποδοτικό βιομηχανικό περιβάλλον [12].

2.3.6 Εφοδιαστική αλυσίδα και μεταφορές

Το IoT παίζει κρίσιμο ρόλο στον εκσυγχρονισμό της εφοδιαστικής αλυσίδας και του τομέα των μεταφορών. Μέσω αισθητήρων και συσκευών παρακολούθησης, παρέχεται η δυνατότητα εντοπισμού της θέσης και της κατάστασης των αποστολών σε πραγματικό χρόνο, η διαχείριση των αποθεμάτων, καθώς και η εξασφάλιση των κατάλληλων συνθηκών κατά τη μεταφορά ευπαθών προϊόντων.

Παράλληλα, το IoT υποστηρίζει τη βελτιστοποίηση των διαδρομών και τη διαχείριση στόλων οχημάτων, ενισχύοντας την αποδοτικότητα των μεταφορών και τη μείωση του κόστους. Χαρακτηριστικό παράδειγμα είναι η χρήση του IoT σε αυτόνομα οχήματα, τα οποία βασίζονται σε δεδομένα αισθητήρων για να κατανοούν το περιβάλλον τους και να λαμβάνουν αποφάσεις σε πραγματικό χρόνο. Συνολικά, οι εφαρμογές αυτές συμβάλλουν στη βελτίωση της εμπειρίας του πελάτη, την εξοικονόμηση πόρων και την ενίσχυση της ασφάλειας κατά τη μετακίνηση και τη διανομή [13].

2.3.7 Smart Energy Grids

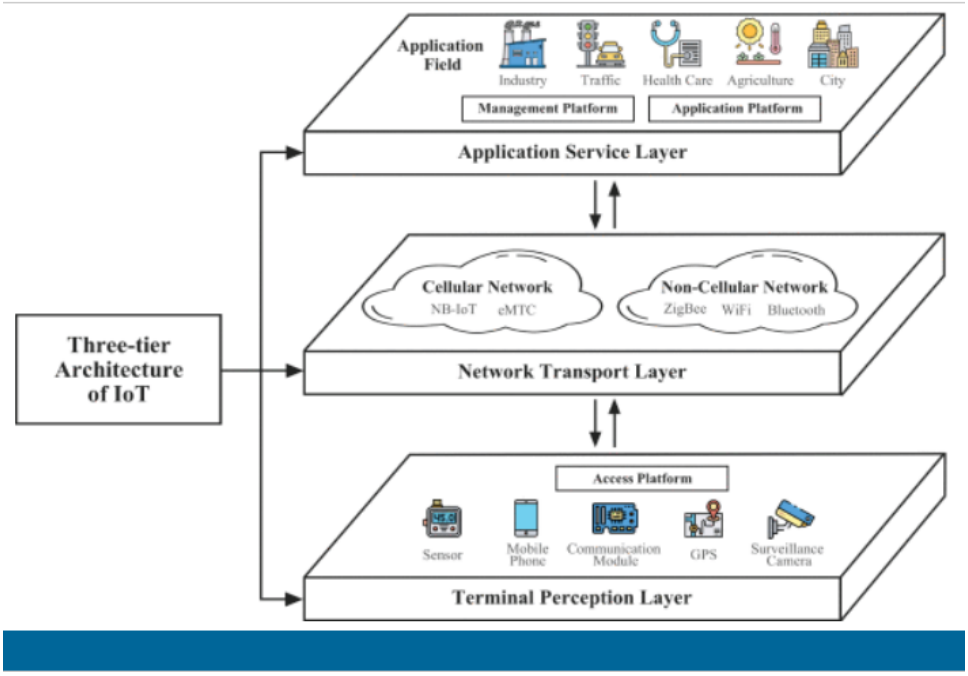
Τα έξυπνα δίκτυα ενέργειας είναι ηλεκτρικά δίκτυα που ενσωματώνουν τεχνολογίες IoT για τη συλλογή, μετάδοση και ανάλυση δεδομένων κατανάλωσης και παροχής ενέργειας σε πραγματικό χρόνο. Με τη χρήση αισθητήρων, έξυπνων μετρητών και συστημάτων αυτοματισμού, επιτρέπουν την καλύτερη ισορροπία μεταξύ προσφοράς και ζήτησης, τη βελτιστοποίηση της διανομής και τη μείωση των ενεργειακών απωλειών. Οι καταναλωτές μπορούν να παρακολουθούν και να ρυθμίζουν τη χρήση ενέργειας τους, ενώ οι πάροχοι έχουν τη δυνατότητα να εντοπίζουν άμεσα προβλήματα στο δίκτυο, όπως διακοπές ή υπερφορτώσεις. Συνολικά, τα smart grids συμβάλλουν στη δημιουργία ενός πιο ευέλικτου, σταθερού και βιώσιμου ενεργειακού συστήματος. Τέτοια δίκτυα εφαρμόζονται ήδη σε σύγχρονες αστικές περιοχές, σε βιομηχανικές εγκαταστάσεις υψηλής κατανάλωσης, καθώς και σε δίκτυα που συνδυάζουν παραδοσιακές και ανανεώσιμες πηγές ενέργειας, όπως φωτοβολταϊκά και ανεμογεννήτριες [14].

2.4 Αρχιτεκτονική IoT

Η αρχιτεκτονική IoT αποτελείται από φυσικά αντικείμενα, που ενσωματώνονται σε ένα δίκτυο επικοινωνίας, σχηματίζοντας ένα τεράστιο και ετερογενές οικοσύστημα συσκευών. Λόγω της πολυπλοκότητας και ποικιλομορφίας του IoT, είναι απαραίτητος ο σχεδιασμός μιας ευέλικτης αρχιτεκτονικής με διακριτά στρώματα (layers) [3]. Υπάρχουν διάφορες αρχιτεκτονικές και μοντέλα αναφοράς στη βιβλιογραφία, που προτάθηκαν από διάφορους οργανισμούς και συγγραφείς.

2.4.1 Αρχιτεκτονική Τριών Επιπέδων

Η Αρχιτεκτονική Τριών Επιπέδων (Three-Layer Architecture) Είναι το πιο βασικό και ευρέως χρησιμοποιούμενο μοντέλο αρχιτεκτονικής του IoT [15]. Αποτελείται από 3 επίπεδα, τα Perception, Network και Application Layer.



Σχήμα 2 1:Αρχιτεκτονική IoT τριών επιπέδων [5]

Στην παραπάνω εικόνα φαίνονται τα τρία στρώματα της αρχιτεκτονικής και η σχέση αλληλεπίδρασης μεταξύ τους. Το **Perception Layer** είναι το πρώτο επίπεδο του IoT συστήματος και αποτελεί την πηγή συλλογής δεδομένων. Συχνά αναφέρεται και ως "στρώμα συσκευών", καθώς περιλαμβάνει φυσικές συσκευές όπως αισθητήρες, κάμερες και μετρητές, που συλλέγουν και καταγράφουν δεδομένα από το περιβάλλον. Το στρώμα αυτό λειτουργεί ως σύνδεσμος ανάμεσα στον φυσικό και τον ψηφιακό κόσμο, αποστέλλοντας τις πληροφορίες προς στο Network Layer για περαιτέρω μεταφορά και επεξεργασία από τα ανώτερα επίπεδα. Η σύνδεση πραγματοποιείται μέσω ασύρματων ή ενσύρματων δικτύων [5]. Λόγω της άμεσης αλληλεπίδρασής του με το φυσικό περιβάλλον, το Perception Layer είναι ιδιαίτερα ευάλωτο σε επιθέσεις, όπως η φυσική παραβίαση (physical tampering) ή η εισαγωγή ψευδών δεδομένων (data injection). Η ασφάλεια σε αυτό το επίπεδο περιλαμβάνει μέτρα όπως η φυσική προστασία των συσκευών, ο έλεγχος ταυτότητας (authentication) για να διασφαλιστεί ότι μόνο έγκυρες συσκευές συμμετέχουν στο δίκτυο και, όπου το επιτρέπουν οι πόροι του υλικού, η χρήση ελαφριάς κρυπτογράφησης (lightweight encryption). Σε πολλές περιπτώσεις, η πλήρης κρυπτογράφηση και προστασία της ακεραιότητας εφαρμόζεται σε ανώτερα επίπεδα του συστήματος, για να αντισταθμίσει τους περιορισμούς των συσκευών πρώτου επιπέδου [3].

Το **Network Layer** είναι υπεύθυνο για τη μεταφορά δεδομένων από το Perception Layer στο Application Layer. Εξασφαλίζει τη συνδεσιμότητα μεταξύ συσκευών IoT, διακομιστών και άλλων στοιχείων του δικτύου. Περιλαμβάνει τόσο την ενσύρματη όσο και την ασύρματη δικτύωση, καθώς και τα ενδιάμεσα στοιχεία που διευκολύνουν τη μεταφορά δεδομένων, όπως δρομολογητές (routers), πύλες (gateways) και κόμβους μετάδοσης (transmission nodes). Χρησιμοποιεί non-cellular δίκτυα (όπως ZigBee, Bluetooth, Wi-Fi) αλλά και cellular δίκτυα (όπως NB-IoT, LTE), τα οποία επιλέγονται ανάλογα με τις ανάγκες σε κάλυψη, ταχύτητα και κατανάλωση ενέργειας [8].

Η ασφάλεια του Network Layer αποτελεί κρίσιμο ζήτημα, καθώς αυτό το επίπεδο είναι ευάλωτο σε επιθέσεις που στοχεύουν είτε στη διακοπή της επικοινωνίας, είτε στην παρακολούθηση και παραποίηση των δεδομένων που μεταδίδονται. Συχνές απειλές είναι οι επιθέσεις Denial of Service (DoS) και Distributed Denial of Service (DDoS) και οι επιθέσεις τύπου Man-in-the-Middle (MitM). Τα μέτρα

προστασίας περιλαμβάνουν τη χρήση πρωτοκόλλων ασφαλούς μετάδοσης (όπως TLS/DTLS), μηχανισμούς αυθεντικοποίησης και εξουσιοδότησης, διαχείριση κλειδιών για την κρυπτογράφηση, καθώς και συστήματα ανίχνευσης και αποτροπής εισβολών (IDS/IPS) για την παρακολούθηση της δραστηριότητας του δικτύου [15].

Το **Application Layer** αποτελεί το ανώτερο επίπεδο σε αυτή την αρχιτεκτονική του IoT και είναι υπεύθυνο για την παροχή υπηρεσιών στους τελικούς χρήστες. Εδώ γίνεται η επεξεργασία των δεδομένων που λαμβάνονται από το Network Layer και η ενσωμάτωσή τους σε λειτουργικές εφαρμογές, οι οποίες αλληλεπιδρούν με τον χρήστη μέσω γραφικών διεπαφών ή αυτόματων μηχανισμών. Παραδείγματα εφαρμογών που λειτουργούν σε αυτό το επίπεδο είναι λογισμικά για απομακρυσμένο έλεγχο, για ενεργειακή διαχείριση, ή έξυπνα οικιακά συστήματα ελέγχου [3].

Η ασφάλεια στο Application Layer επικεντρώνεται κυρίως στην προστασία των προσωπικών δεδομένων των χρηστών και στον περιορισμό της πρόσβασης μόνο σε εξουσιοδοτημένα άτομα. Για τον σκοπό αυτό, χρησιμοποιούνται μέθοδοι ελέγχου πρόσβασης, όπως η χρήση κωδικών πρόσβασης και η επαλήθευση ταυτότητας με περισσότερους από έναν τρόπους (π.χ. με επαλήθευση πολλαπλών παραγόντων). Επιπλέον, ιδιαίτερη σημασία δίνεται στην ασφαλή αποθήκευση των δεδομένων και στη σωστή διαχείριση της επικοινωνίας μεταξύ των εφαρμογών, ώστε να περιορίζεται ο κίνδυνος παραβίασης της ασφάλειας ή της εμπιστευτικότητας των δεδομένων [15].

Η αρχιτεκτονική τριών επιπέδων παρέχει ένα απλό αλλά ισχυρό μοντέλο για την κατανόηση και υλοποίηση του IoT, διαχωρίζοντας σαφώς τη συλλογή, τη μεταφορά και την αξιοποίηση των δεδομένων. Ωστόσο, για την κάλυψη πιο σύνθετων αναγκών και τη βελτίωση της ασφάλειας και της διαχείρισης, έχουν προταθεί εκτενέστερες αρχιτεκτονικές.

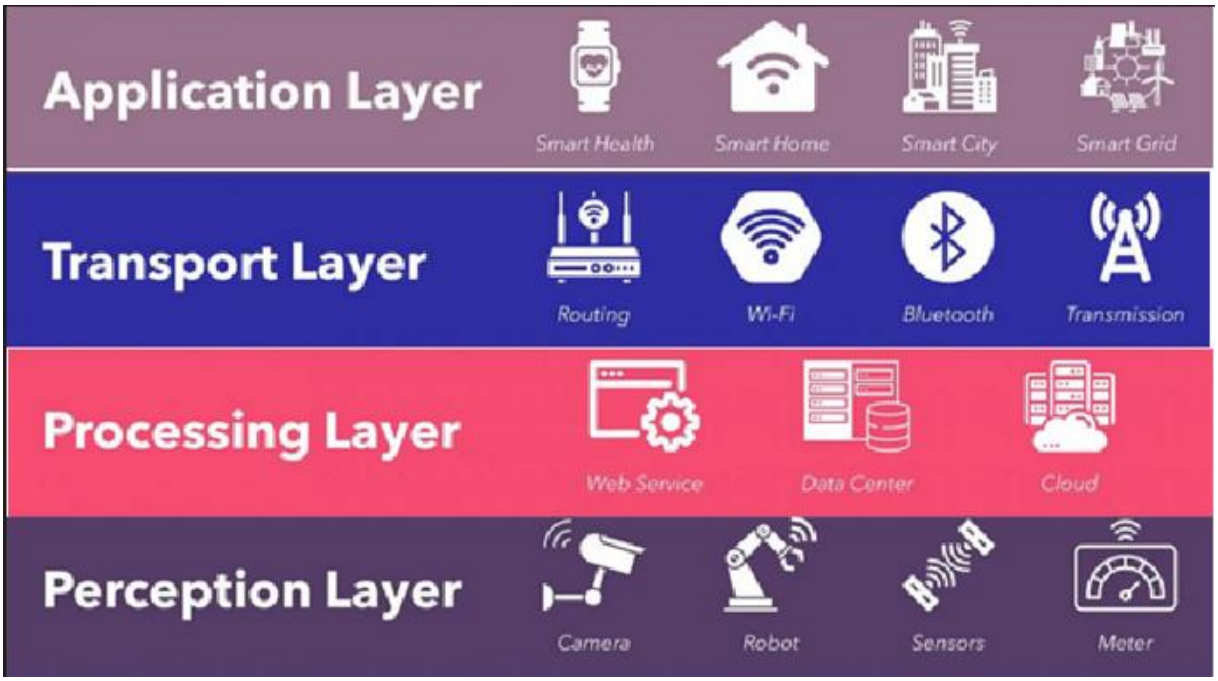
2.4.2 Αρχιτεκτονική Τεσσάρων Επιπέδων

Η Αρχιτεκτονική Τεσσάρων Επιπέδων (Four-Layer Architecture) επεκτείνει την Αρχιτεκτονική Τριών Επιπέδων, προσθέτοντας το Processing Layer, το οποίο τοποθετείται ανάμεσα στο Network Layer και το Application Layer [16]. Ο κύριος στόχος αυτού του επιπέδου είναι η τοπική επεξεργασία των δεδομένων που συλλέγονται από τις συσκευές IoT, πριν αυτά μεταδοθούν στο επόμενο επίπεδο.

Η προσθήκη του **Processing Layer** επιτρέπει την εφαρμογή τεχνολογιών όπως το Edge Computing, το οποίο διευκολύνει την επεξεργασία δεδομένων κοντά στην πηγή τους, μειώνοντας την ανάγκη αποστολής μεγάλου όγκου δεδομένων μέσω του δικτύου. Έτσι, η συνολική αποδοτικότητα του συστήματος αυξάνεται, καθώς περιορίζεται η κατανάλωση bandwidth και ελαττώνονται οι καθυστερήσεις στην επεξεργασία.

Το **Perception Layer**, το οποίο παραμένει ίδιο με την αρχιτεκτονική τριών επιπέδων, περιλαμβάνει αισθητήρες και φυσικές συσκευές που συλλέγουν δεδομένα από το περιβάλλον. Αυτά τα δεδομένα μεταφέρονται στο **Processing Layer**, το οποίο τα φιλτράρει και τα επεξεργάζεται, πριν τα στείλει στο **Network Layer** για περαιτέρω ανάλυση ή αποθήκευση στο **Application Layer**. Το Processing Layer περιλαμβάνει υπηρεσίες Web (Web Services), υποδομές αποθήκευσης και επεξεργασίας όπως Data Centers, καθώς και Cloud πλατφόρμες, οι οποίες παρέχουν την υπολογιστική ισχύ που απαιτείται για την ανάλυση των δεδομένων σε πραγματικό ή σχεδόν πραγματικό χρόνο. Αυτή η τοπική επεξεργασία μειώνει το φορτίο στο Network Layer, επιτρέποντας ταχύτερη, αποδοτικότερη και ασφαλέστερη επικοινωνία μεταξύ των επιπέδων. Επίσης, τα δεδομένα δεν χρειάζεται να μεταφέρονται σε απομακρυσμένα κέντρα δεδομένων (data centers) για επεξεργασία, κάτι που μειώνει την πιθανότητα

υποκλοπής δεδομένων κατά τη μεταφορά μέσω του δικτύου, ενισχύοντας έτσι τη συνολική ασφάλεια του συστήματος IoT [16].



Σχήμα 2 2: Αρχιτεκτονική IoT τεσσάρων επιπέδων [17]

Ωστόσο, η προσθήκη του Processing Layer στην αρχιτεκτονική του IoT εισάγει νέες προκλήσεις ασφάλειας. Πρώτον, επειδή το επίπεδο αυτό ενδέχεται να επεξεργάζεται ή να αποθηκεύει ευαίσθητα δεδομένα σε τοπικές μονάδες κοντά στις IoT συσκευές, αυξάνεται ο κίνδυνος παραβίασης ή κλοπής αν δεν υπάρχουν κατάλληλα μέτρα προστασίας [17]. Για παράδειγμα, η απουσία ασφαλούς κρυπτογράφησης κατά τη μεταφορά ή προσωρινή αποθήκευση των δεδομένων μπορεί να τα καταστήσει ευάλωτα σε επιθέσεις τύπου man-in-the-middle. Δεύτερον, η περιορισμένη υπολογιστική ισχύς των τοπικών συσκευών δυσκολεύει την εφαρμογή σύνθετων μηχανισμών ελέγχου πρόσβασης και αυθεντικοποίησης. Παρόλο που τεχνολογίες όπως το Edge ή Fog Computing μειώνουν την εξάρτηση από απομακρυσμένους servers, η τοπική επεξεργασία απαιτεί αξιόπιστη και προσαρμοσμένη ασφάλεια, σχεδιασμένη με βάση τις δυνατότητες και τα όρια των IoT συστημάτων. Αυτές οι προκλήσεις καθιστούν το Processing Layer ιδιαίτερα ευάλωτο σε επιθέσεις και απαιτούν αυξημένη προσοχή στον σχεδιασμό πολιτικών ασφαλείας, ειδικά σχεδιασμένων για τις ανάγκες και ιδιαιτερότητες του IoT οικοσυστήματος.

2.4.3 Αρχιτεκτονική που Βασίζεται στο Edge Computing

Η αρχιτεκτονική που βασίζεται στο edge computing (Edge-Centric IoT Architecture) είναι μια σύγχρονη προσέγγιση, που εστιάζει στην αποκεντρωμένη επεξεργασία δεδομένων, δηλαδή στην επεξεργασία δεδομένων που πραγματοποιείται κοντά στις συσκευές, στο "άκρο" (edge) του δικτύου [20]. Χρησιμοποιεί το edge computing, μειώνοντας την ανάγκη για συνεχή επικοινωνία με το cloud και ενισχύοντας τη συνολική απόδοση του συστήματος. Οι συσκευές επεξεργάζονται τα δεδομένα τοπικά, είτε πλήρως είτε σε αρχικό βαθμό, πριν αυτά προωθηθούν σε ανώτερα επίπεδα.

Στα πλεονεκτήματα αυτής της αρχιτεκτονικής, περιλαμβάνονται οι μειωμένες καθυστερήσεις, η ταχύτερη ανταπόκριση, καθώς και η ενίσχυση της ασφάλειας και ιδιωτικότητας, αφού τα δεδομένα δεν αποστέλλονται σε απομακρυσμένα συστήματα. Είναι ιδανική για εφαρμογές IoT που απαιτούν υψηλή

απόδοση σε πραγματικό χρόνο, όπως η βιομηχανία (IIoT), τα έξυπνα οχήματα και οι υπηρεσίες υγειονομικής περίθαλψης. Η αρχιτεκτονική αυτή περιλαμβάνει τα εξής επίπεδα:

Το **Επίπεδο Συσκευών (Device Layer)** είναι το χαμηλότερο επίπεδο της αρχιτεκτονικής και περιλαμβάνει τις φυσικές συσκευές που συλλέγουν δεδομένα από το περιβάλλον, όπως αισθητήρες, κάμερες, μετρητές και ενεργοποιητές. Αποτελεί το σημείο όπου το IoT σύστημα αποκτά πρόσβαση στον φυσικό κόσμο, καταγράφοντας κρίσιμες πληροφορίες όπως θερμοκρασία, πίεση, φωτεινότητα ή κίνηση. Οι συσκευές αυτές διασυνδέονται με τα επόμενα επίπεδα μέσω δικτυακών τεχνολογιών μικρής εμβέλειας και χαμηλής κατανάλωσης, όπως Bluetooth Low Energy (BLE), ZigBee ή LoRa. Χάρη στην άμεση διασύνδεσή του με το Edge Layer, το Device Layer συμβάλλει στη μείωση της καθυστέρησης στη ροή των δεδομένων, επιτρέποντας την ταχύτερη αντίδραση του συστήματος. Αυτό είναι ιδιαίτερα σημαντικό σε εφαρμογές που απαιτούν λήψη αποφάσεων σε πραγματικό χρόνο, όπως η έγκυρη αποφυγή σύγκρουσης σε αυτόνομα οχήματα ή η άμεση ενεργοποίηση συναγερμών σε συστήματα ασφάλειας. Η φύση των συσκευών τις καθιστά ευάλωτες σε επιθέσεις φυσικής παραβίασης ή αλλοίωσης των αισθητήρων, ενώ η απουσία μηχανισμών αυθεντικοποίησης μπορεί να επιτρέψει την είσοδο πλαστών συσκευών στο δίκτυο. Για τον περιορισμό αυτών των κινδύνων, εφαρμόζονται μέτρα όπως η φυσική προστασία, ο βασικός έλεγχος ταυτότητας και, όπου είναι δυνατό, ελαφριά κρυπτογράφηση [22].

Το **Επίπεδο Edge (Edge Layer)** λειτουργεί ως ενδιάμεσο επίπεδο μεταξύ των συσκευών και του cloud, αναλαμβάνοντας την επεξεργασία των δεδομένων όσο το δυνατόν πιο κοντά στην πηγή τους. Περιλαμβάνει υπολογιστικούς πόρους όπως τοπικούς servers, micro data centers, έξυπνες πύλες (smart gateways) και συσκευές edge με ενσωματωμένες δυνατότητες επεξεργασίας και ανάλυσης. Οι υποδομές αυτές αξιοποιούν τεχνολογίες όπως Edge AI, Docker containers και ελαφριά μοντέλα Μηχανικής Μάθησης (lightweight ML models), επιτρέποντας την τοπική επεξεργασία, το φιλτράρισμα, ή και τη λήψη αποφάσεων σε πραγματικό χρόνο, χωρίς να απαιτείται άμεση σύνδεση με το cloud. Με αυτόν τον τρόπο, μειώνεται ο όγκος των δεδομένων που πρέπει να μεταφερθούν, ελαττώνεται η κατανάλωση bandwidth και ενισχύεται η απόδοση του συστήματος. Η αποκεντρωμένη φύση του Edge Layer ενισχύει την ιδιωτικότητα των δεδομένων, μειώνοντας την ανάγκη συνεχούς μεταφοράς προς το cloud. Ωστόσο, δημιουργεί και νέες απαιτήσεις ως προς την ασφάλεια των τοπικών κόμβων, οι οποίοι αποτελούν κρίσιμα σημεία επεξεργασίας και ενδέχεται να στοχοποιηθούν από επιθέσεις [20].

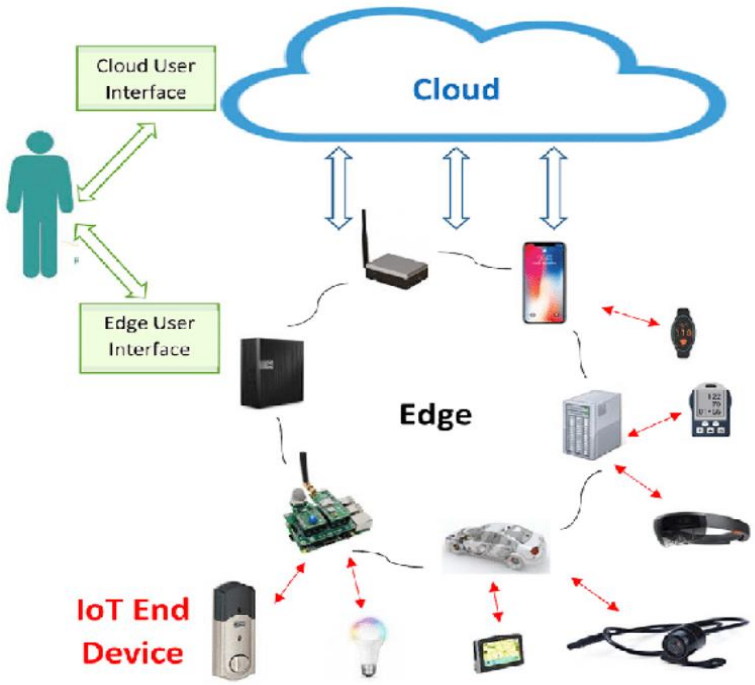
Το Edge Layer περιλαμβάνει συσκευές που εκτελούν τοπική επεξεργασία των δεδομένων, πριν αυτά αποσταλούν στο cloud. Τα δεδομένα αυτά χρησιμοποιούνται για τη λήψη άμεσων αποφάσεων, χωρίς να απαιτείται έγκριση από το cloud. Αυτό το χαρακτηριστικό καθιστά το συγκεκριμένο επίπεδο ιδιαίτερα κρίσιμο για την ασφάλεια, καθώς οποιαδήποτε αλλοίωση ή απώλεια των δεδομένων μπορεί να επηρεάσει άμεσα τη συμπεριφορά του συστήματος. Για παράδειγμα, σε εφαρμογές όπως τα αυτόνομα οχήματα, η βιομηχανία ή τα συστήματα ασφαλείας, η αξιοπιστία αυτής της επεξεργασίας είναι κρίσιμη. Αλλοίωση ή καθυστέρηση των δεδομένων σε αυτό το στάδιο μπορεί να οδηγήσει σε ανεπιθύμητες ενέργειες, όπως λανθασμένο φρενάρισμα, ενεργοποίηση συναγερμού χωρίς λόγο ή αποτυχία εντοπισμού μιας βλάβης σε μηχανήματα παραγωγής.

Συνήθεις επιθέσεις στο Edge Layer είναι η εισαγωγή κακόβουλου λογισμικού μέσω αδύναμων σημείων πρόσβασης, η παρακολούθηση ή αλλοίωση των δεδομένων κατά τη μεταφορά (μέσω επίθεσης MitM), καθώς και η αντικατάσταση μιας πραγματικής συσκευής edge με πλαστή συσκευή (spoofing). Για την προστασία αυτού του επιπέδου εφαρμόζονται πρωτόκολλα ασφαλούς επικοινωνίας, έλεγχος εξουσιοδοτημένης πρόσβασης, περιορισμός των διαχειριστικών λειτουργιών, καθώς και συστηματική

εγκατάσταση ενημερώσεων ασφαλείας στις συσκευές του επιπέδου, ώστε να διορθώνονται γνωστές ευπάθειες.

Το **Επίπεδο Cloud (Cloud Layer)** αποτελεί το ανώτερο επίπεδο της αρχιτεκτονικής και είναι υπεύθυνο για τη μακροχρόνια αποθήκευση, την κεντρική επεξεργασία και τη διαχείριση εφαρμογών. Περιλαμβάνει υποδομές όπως δημόσια ή ιδιωτικά cloud περιβάλλοντα, μεγάλα data centers, καθώς και υπηρεσίες cloud-based analytics. Το επίπεδο αυτό επιτρέπει την ολοκληρωμένη ανάλυση μεγάλων ποσοτήτων δεδομένων, καθώς και την εκπαίδευση ή τον συντονισμό αλγορίθμων που μπορούν να εφαρμοστούν αργότερα τοπικά στο Edge Layer. Αν και το Cloud Layer δεν προσφέρει άμεση απόκριση, η συμβολή του είναι κρίσιμη για την ευρύτερη εικόνα, τη στρατηγική λήψη αποφάσεων και την αποτελεσματική διαχείριση του IoT οικοσυστήματος συνολικά. Ενδεικτικές τεχνολογίες που αξιοποιούνται σε αυτό το επίπεδο είναι το Amazon Web Services (AWS IoT) [22] και το Microsoft Azure IoT Hub [23].

Το Cloud Layer διαχειρίζεται μεγάλο όγκο δεδομένων και λειτουργεί ως κεντρικός χώρος αποθήκευσης και ανάλυσης. Επειδή οι cloud εφαρμογές συνδέονται με άλλες υπηρεσίες και βάσεις δεδομένων μέσω διασυνδέσεων (Application Programming Interfaces – APIs), υπάρχουν αυξημένοι κίνδυνοι για επιθέσεις τύπου SQL injection, όπου ο εισβολέας εισάγει εντολές σε ερωτήματα βάσης για να αποκτήσει πρόσβαση σε ευαίσθητες πληροφορίες [21]. Επιπλέον, αν τα APIs δεν είναι κατάλληλα ασφαλισμένα, μπορεί να επιτραπεί η μη εξουσιοδοτημένη πρόσβαση σε δεδομένα ή λειτουργίες του συστήματος. Για την αντιμετώπιση αυτών των απειλών, εφαρμόζονται πρακτικές όπως κρυπτογράφηση κατά την αποθήκευση των δεδομένων, έλεγχος δικαιωμάτων πρόσβασης ανά χρήστη ή εφαρμογή και παρακολούθηση της δραστηριότητας με στόχο την ανίχνευση ύποπτης συμπεριφοράς [23].



Σχήμα 2 3: Αρχιτεκτονική IoT Edge Computing [21]

2.5 Πρωτόκολλα IoT

Τα πρωτόκολλα του IoT αποτελούν βασικό στοιχείο της επικοινωνίας και λειτουργίας των συσκευών του. Είναι υπεύθυνα για τη μετάδοση δεδομένων μεταξύ των συσκευών, δικτύων και εφαρμογών, εξασφαλίζοντας την ομαλή διασύνδεση και συνεργασία τους. Λόγω της μεγάλης ποικιλίας συσκευών και των διαφορών στις απαιτήσεις κάθε περιβάλλοντος, έχουν αναπτυχθεί πρωτόκολλα που καλύπτουν τόσο γενικές όσο εξειδικευμένες ανάγκες. Στην ενότητα αυτή παρουσιάζονται βασικά πρωτόκολλα του IoT, ταξινομημένα με βάση τη συχνότητα χρήσης και το πεδίο εφαρμογής τους.

2.5.1 Πρωτόκολλα Τοπικής Ασύρματης Επικοινωνίας

Η τοπική ασύρματη επικοινωνία στο IoT επιτρέπει τη σύνδεση συσκευών που βρίσκονται σε μικρή απόσταση μεταξύ τους, όπως αισθητήρες και οικιακές συσκευές. Τα πρωτόκολλα αυτής της κατηγορίας δίνουν έμφαση στη χαμηλή κατανάλωση ενέργειας και στη σταθερή μετάδοση δεδομένων σε περιορισμένα γεωγραφικά όρια.

- **Wi-Fi.** Είναι ασύρματη τεχνολογία επικοινωνίας, που λειτουργεί στο φάσμα συχνοτήτων 2.4 GHz και 5 GHz. Χρησιμοποιείται σε έξυπνα σπίτια (π.χ. έξυπνους θερμοστάτες και κάμερες παρακολούθησης) και σε βιομηχανικές συνθήκες όπου απαιτείται αξιόπιστη σύνδεση υψηλού εύρους ζώνης. Γενικά καταναλώνει περισσότερη ενέργεια, σε σχέση με άλλα πρωτόκολλα IoT [18].
- **Bluetooth.** Είναι ασύρματο πρωτόκολλο μικρής εμβέλειας, που συνήθως χρησιμοποιείται σε καθημερινές εφαρμογές ανταλλαγής δεδομένων ανάμεσα σε συσκευές όπως κινητά, ακουστικά και υπολογιστές. Στο IoT, το Bluetooth χρησιμοποιείται σε εφαρμογές που απαιτούν συνεχή ροή δεδομένων. Ωστόσο, η κατανάλωση ενέργειας είναι σχετικά υψηλή, ειδικά σε σύγκριση με το BLE [15].
- **Bluetooth Low Energy (BLE).** Είναι μία ενεργειακά αποδοτικότερη έκδοση του Bluetooth. Σχεδιάστηκε για φορητές συσκευές που λειτουργούν με μπαταρία και δεν απαιτούν συνεχή μετάδοση δεδομένων, όπως fitness trackers, smartwatches, αισθητήρες θερμοκρασίας ή συσκευές υγείας. Καταναλώνει ελάχιστη ενέργεια, μεταδίδοντας μικρά πακέτα δεδομένων σε αραιά χρονικά διαστήματα. Είναι ιδανικό για εφαρμογές μικρής εμβέλειας και μεγάλης διάρκειας λειτουργίας. Σε αντίθεση με το Bluetooth, το BLE δεν υποστηρίζει συνεχή ροή ήχου ή βίντεο, λόγω του περιορισμένου throughput και γι' αυτό δεν χρησιμοποιείται σε εφαρμογές όπως IP κάμερες [15]. Παράδειγμα χρήσης του BLE σε IoT εφαρμογές είναι τα Apple AirTags, που βασίζονται σε Bluetooth Low Energy για την αποστολή σήματος εντοπισμού προς το δίκτυο Find My της Apple [19].
- **ZigBee.** Ασύρματο πρωτόκολλο μικρής γεωγραφικής απόστασης, που υποστηρίζει χαμηλή κατανάλωση ενέργειας. Χρησιμοποιείται συχνά σε έξυπνα σπίτια, βιομηχανίες και συσκευές υγειονομικής περίθαλψης. Υποστηρίζει mesh networking, επιτρέποντας στις συσκευές να επικοινωνούν σε εκτεταμένες περιοχές με χαμηλή ενέργεια [15].
- **Z-Wave.** Το Z-Wave είναι πρωτόκολλο χαμηλής κατανάλωσης ενέργειας, σχεδιασμένο για κοντινές αποστάσεις και συχνά χρησιμοποιούμενο σε εφαρμογές έξυπνου σπιτιού. Υποστηρίζει mesh δικτύωση και προσφέρει εμβέλεια έως 100 μέτρα, με υψηλή αξιοπιστία και χαμηλές παρεμβολές. Χρησιμοποιείται σε συστήματα αυτοματισμού όπως φωτισμός, θερμοστάτες, αισθητήρες κίνησης ή επαφής (σε πόρτες και παράθυρα), καθώς και σε συναγερμούς και κάμερες ασφαλείας [12].
- **Thread.** Το Thread είναι ασύρματο πρωτόκολλο βασισμένο στο IPv6. Υποστηρίζει mesh δίκτυα χωρίς την ανάγκη κεντρικού hub, προσφέροντας υψηλή αξιοπιστία, ασφάλεια και ευελιξία σε μικρές αποστάσεις. Χρησιμοποιείται σε εφαρμογές όπως έξυπνος φωτισμός, θερμοστάτες και λοιπούς αυτοματισμούς κατοικίας, διευκολύνοντας την επικοινωνία πολλών συσκευών με χαμηλή κατανάλωση ενέργειας [19].
- **WirelessHART.** Το WirelessHART είναι ασύρματη επέκταση του βιομηχανικού πρωτοκόλλου HART και έχει σχεδιαστεί για αξιόπιστη και ασφαλή επικοινωνία σε πραγματικό χρόνο.

Χρησιμοποιείται σε βιομηχανικά περιβάλλοντα με υψηλές απαιτήσεις αξιοπιστίας, παρεμβολών και ηλεκτρικού θορύβου. Προσφέρει σταθερό χρόνο απόκρισης και ανοχή σε παρεμβολές, διασφαλίζοντας αδιάκοπη λειτουργία ακόμα και υπό δύσκολες συνθήκες [12].

- **ISA100.11a.** Το ISA100.11a είναι ασύρματο πρωτόκολλο μικρής απόστασης, σχεδιασμένο για εφαρμογές βιομηχανικού αυτοματισμού και παρακολούθησης. Παρέχει αξιόπιστη επικοινωνία με χαμηλή καθυστέρηση και υψηλή ανοχή σε παρεμβολές. Ενδείκνυται για περιβάλλοντα με αυστηρές απαιτήσεις αξιοπιστίας, όπως εργοστάσια παραγωγής, καθώς προσφέρει σταθερή λειτουργία και συνεχή ροή δεδομένων ακόμη και σε συνθήκες έντονου βιομηχανικού θορύβου [12].
- **HaLOW (Wi-Fi HaLow).** Το Wi-Fi HaLow (IEEE 802.11ah) είναι επέκταση του κλασικού Wi-Fi, σχεδιασμένη για IoT εφαρμογές που απαιτούν χαμηλή κατανάλωση ενέργειας. Παρέχει καλύτερη διείσδυση σε τοίχους και εμπόδια, καθιστώντας το ιδανικό για χρήση σε έξυπνα κτίρια, κάμερες ασφαλείας, γεωργικές εγκαταστάσεις και περιβάλλοντα με πολλές φυσικές παρεμβολές [19].
- **NFC (Near-Field Communication).** Το NFC είναι πρωτόκολλο ασύρματης επικοινωνίας πολύ μικρής εμβέλειας (μερικά εκατοστά), σχεδιασμένο για γρήγορες και ασφαλείς ανταλλαγές δεδομένων. Χρησιμοποιείται σε ανέπαφες πληρωμές, έλεγχο πρόσβασης, ταυτοποίηση συσκευών και μεταφορά δεδομένων μεταξύ δύο κοντινών συσκευών. Υποστηρίζει ενεργή ή παθητική λειτουργία, ανάλογα με το αν η ανταλλαγή γίνεται και από τις δύο συσκευές ή μόνο από τη μία [15].

2.5.2 Πρωτόκολλα Χαμηλής Κατανάλωσης και Ευρείας Περιοχής

Η ανάγκη διασύνδεσης IoT συσκευών σε μεγάλες αποστάσεις με χαμηλή κατανάλωση ενέργειας καλύπτεται από τα πρωτόκολλα χαμηλής κατανάλωσης και ευρείας περιοχής (Low-Power, Wide-Area Network – LPWAN). Αυτά επιτρέπουν τη μετάδοση δεδομένων σε μεγάλες αποστάσεις, διατηρώντας παράλληλα υψηλή απόδοση και χαμηλό ενεργειακό αποτύπωμα.

- **LoRaWAN.** Το LoRaWAN είναι ασύρματο πρωτόκολλο χαμηλής κατανάλωσης, σχεδιασμένο για επικοινωνία σε μεγάλες αποστάσεις. Χρησιμοποιείται σε εφαρμογές όπου απαιτούνται σποραδικές μεταδόσεις μικρών ποσοτήτων δεδομένων, όπως μετεωρολογικοί σταθμοί, αισθητήρες εδάφους ή έξυπνοι κάδοι απορριμμάτων. Εφαρμόζεται σε έξυπνες πόλεις, γεωργία και βιομηχανία, ιδίως σε απομακρυσμένες περιοχές χωρίς εύκολη πρόσβαση σε ενέργεια ή σταθερή σύνδεση. Υποστηρίζει αρχιτεκτονική αστέρα (star topology), αλλά προσφέρει χαμηλό ρυθμό μετάδοσης, γεγονός που το καθιστά ακατάλληλο για εφαρμογές με απαιτήσεις σε συνεχή ή υψηλής συχνότητας επικοινωνία [15].
- **6LoWPAN (IPv6 over Low-Power Wireless Personal Area Networks).** Το 6LoWPAN είναι πρωτόκολλο που επιτρέπει τη μετάδοση πακέτων IPv6 μέσω ασύρματων δικτύων χαμηλής ισχύος και περιορισμένου εύρους ζώνης. Υποστηρίζει συμπίεση επικεφαλίδων και τεμαχισμό πακέτων, διευκολύνοντας την αποστολή IP δεδομένων σε δίκτυα περιορισμένων πόρων. Δεν αναλαμβάνει τη φυσική επικοινωνία, αλλά λειτουργεί ως ενδιάμεσο επίπεδο και χρησιμοποιείται σε συνδυασμό με πρωτόκολλα όπως το ZigBee ή το Thread, τα οποία καλύπτουν τη φυσική και τη συνδεσμική στρώση [18].
- **SigFox.** Το SigFox είναι πρωτόκολλο LPWAN, σχεδιασμένο για απλές IoT εφαρμογές που απαιτούν εξαιρετικά χαμηλή κατανάλωση ενέργειας και μεταδίδουν μικρές ποσότητες δεδομένων. Υποστηρίζει μονής ή περιορισμένης διπλής κατεύθυνσης επικοινωνία και χρησιμοποιείται σε περιπτώσεις όπου δεν απαιτείται συνεχής σύνδεση, όπως έξυπνοι κάδοι απορριμμάτων, συστήματα εντοπισμού αντικειμένων (trackers) και αισθητήρες περιβάλλοντος [15].
- **NWave.** Το NWave είναι πρωτόκολλο που προσφέρει εξαιρετικά χαμηλή κατανάλωση ενέργειας και μεγάλη εμβέλεια επικοινωνίας. Αν και λιγότερο διαδεδομένο από πρωτόκολλα όπως το LoRaWAN ή το SigFox, παρέχει αξιόπιστη σύνδεση για εφαρμογές που δεν απαιτούν συχνή μετάδοση δεδομένων, όπως έξυπνοι μετρητές κατανάλωσης και περιβαλλοντικοί αισθητήρες [18].

2.5.3 Κυψελοειδή Πρωτόκολλα

Τα κυψελοειδή πρωτόκολλα (cellular protocols) χρησιμοποιούνται για αξιόπιστη σύνδεση IoT συσκευών μέσω υφιστάμενων υποδομών κινητής τηλεφωνίας. Παρέχουν ευρεία κάλυψη, δυνατότητα μετάδοσης μεγάλων όγκων δεδομένων και σταθερή, συνεχής επικοινωνία.

- **NB-IoT (Narrowband IoT).** Το NB-IoT είναι τεχνολογία σχεδιασμένη για συσκευές IoT χαμηλής κατανάλωσης που μεταδίδουν μικρές ποσότητες δεδομένων. Βασίζεται σε υπάρχουσες υποδομές κινητής τηλεφωνίας και επιτρέπει την εξυπηρέτηση μεγάλου αριθμού συσκευών ανά κεραία. Παρέχει μεγάλη κάλυψη, ακόμα και σε υπόγειους ή απομακρυσμένους χώρους και χρησιμοποιείται κυρίως σε στατικές εφαρμογές, όπως έξυπνοι μετρητές, αισθητήρες στάθμευσης ή παρακολούθησης περιβάλλοντος. Η περιοδική, χαμηλής έντασης επικοινωνία το καθιστά ιδανικό για συστήματα που απαιτούν μεγάλη αυτονομία μπαταρίας και χαμηλό κόστος συντήρησης [18].
- **LTE Cat-M.** Το LTE Cat-M είναι πρωτόκολλο LPWAN που βασίζεται σε δίκτυα 4G και έχει σχεδιαστεί για συσκευές IoT που απαιτούν κινητικότητα και σταθερή συνδεσιμότητα. Προσφέρει χαμηλή κατανάλωση ενέργειας, καλή ταχύτητα μετάδοσης και χαμηλό κόστος. Χρησιμοποιείται σε εφαρμογές όπως GPS trackers οχημάτων, φορητούς μετρητές και αισθητήρες, όπου η συσκευή αλλάζει θέση αλλά πρέπει να παραμένει συνεχώς συνδεδεμένη [18].
- **LTE-A (LTE-Advanced).** Το LTE-A είναι εξέλιξη του πρωτοκόλλου LTE και προσφέρει υψηλότερες ταχύτητες, χαμηλότερη καθυστέρηση και αυξημένη αξιοπιστία. Είναι κατάλληλο για IoT εφαρμογές που απαιτούν μετάδοση δεδομένων σε πραγματικό χρόνο και μεγάλο εύρος ζώνης, όπως βιομηχανικά περιβάλλοντα, συστήματα παρακολούθησης βίντεο (video surveillance) και επικοινωνία σε έξυπνα οχήματα [19].
- **GSM (Global System for Mobile Communication).** Το GSM είναι παλαιότερο πρωτόκολλο κινητής τηλεφωνίας, που εξακολουθεί να χρησιμοποιείται σε ορισμένες IoT εφαρμογές λόγω της παγκόσμιας κάλυψης του. Αν και καταναλώνει περισσότερη ενέργεια και προσφέρει χαμηλότερη ταχύτητα σε σχέση με πιο σύγχρονες τεχνολογίες, παραμένει χρήσιμο σε περιπτώσεις όπου απαιτείται βασική συνδεσιμότητα χωρίς υψηλές απαιτήσεις, όπως σε συστήματα παρακολούθησης ή ειδοποίησης σε απομακρυσμένες περιοχές [18].

2.6 Επίλογος

Σε αυτό το κεφάλαιο αναλύθηκε το οικοσύστημα του IoT, με στόχο την κατανόηση της δομής και της βασικής του λειτουργίας. Αφού ορίστηκε η έννοια του οικοσυστήματος, παρουσιάστηκαν οι εφαρμογές του σε διαφορετικούς τομείς της καθημερινής ζωής και της βιομηχανίας. Έμφαση δόθηκε στις αρχιτεκτονικές που έχουν αναπτυχθεί για την υλοποίηση των IoT συστημάτων, καθώς και στα βασικά πρωτόκολλα επικοινωνίας που επιτρέπουν την ανταλλαγή δεδομένων μεταξύ διάφορων ειδών συσκευών.

Κεφάλαιο 3ο: Ασφάλεια στο IoT

3.1 Εισαγωγή

Αυτό το κεφάλαιο επικεντρώνεται στο θέμα της ασφάλειας των πληροφοριακών συστημάτων, τόσο σε γενικό επίπεδο όσο και στο πλαίσιο του οικοσυστήματος του IoT. Αρχικά, ορίζεται η έννοια της ψηφιακής ασφάλειας και παρουσιάζονται οι βασικοί της στόχοι. Στη συνέχεια, εξετάζεται το πως μεταφράζονται αυτές οι αρχές στον χώρο του IoT, καθώς και οι νέες προκλήσεις που εισάγονται λόγω της φύσης και της ποικιλομορφίας των συσκευών που περιλαμβάνει. Έπειτα, γίνεται ανάλυση των συνηθέστερων μορφών επιθέσεων σε δίκτυα IoT. Το κεφάλαιο ολοκληρώνεται με μία σύντομη αναφορά στα παραδοσιακά συστήματα εντοπισμού εισβολών και τους λόγους που αποτρέπουν τη χρήση τους σε οικοσυστήματα IoT.

3.2 Ψηφιακή ασφάλεια: Ορισμός και θεμελιώδεις αρχές

Η ασφάλεια στα υπολογιστικά συστήματα ορίζεται ως η διαδικασία προστασίας ψηφιακών πόρων, όπως υπολογιστικά συστήματα, δίκτυα, λογισμικό και δεδομένα, από κακόβουλες ενέργειες ή επιθέσεις [24]. Οι επιθέσεις αυτές συνήθως στοχεύουν σε μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητες πληροφορίες, αλλοίωση ή διαγραφή δεδομένων, καθώς και στην παρεμπόδιση της ομαλής λειτουργίας των συστημάτων. Η εφαρμογή αποδοτικών μέτρων ψηφιακής ασφάλειας είναι ιδιαίτερα απαιτητική, εξαιτίας του μεγάλου αριθμού συνδεδεμένων συσκευών και της συνεχούς εξέλιξης των τεχνικών που χρησιμοποιούν οι εγκληματίες του ψηφιακού κόσμου. Οι βασικοί στόχοι της ψηφιακής ασφάλειας περιλαμβάνουν:

- **Εμπιστευτικότητα (Confidentiality).** Η διασφάλιση ότι οι ευαίσθητες πληροφορίες είναι προσβάσιμες μόνο από εξουσιοδοτημένα άτομα.
- **Ακεραιότητα (Integrity).** Η διατήρηση της ακρίβειας και πληρότητας των δεδομένων.
- **Διαθεσιμότητα (Availability).** Η εξασφάλιση ότι οι εξουσιοδοτημένοι χρήστες έχουν συνεχή και αξιόπιστη πρόσβαση σε δεδομένα, συστήματα και υπολογιστικούς πόρους, όταν αυτό απαιτείται.

3.3 Προκλήσεις ασφάλειας στο IoT

Η ασφάλεια στο IoT αναφέρεται στην προστασία των συσκευών, των δικτύων και των δεδομένων που μεταδίδονται. Βασικοί στόχοι της ασφάλειας είναι η διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των δεδομένων και υπηρεσιών [5]. Ωστόσο, η φύση του IoT, με τη μεγάλη ποικιλία συσκευών, την περιορισμένη υπολογιστική ισχύ και τους μεγάλους όγκους δεδομένων δημιουργεί νέες προκλήσεις που σχετίζονται με την ασφάλεια των οικοσυστημάτων.

Περιορισμένοι πόροι των συσκευών. Πολλές συσκευές IoT διαθέτουν ελάχιστους υπολογιστικούς πόρους, γεγονός που περιορίζει τη δυνατότητα εφαρμογής ισχυρών μέτρων ασφάλειας. Οι παραδοσιακές λύσεις ασφάλειας, όπως τα πρωτόκολλα κρυπτογράφησης ή οι σύνθετοι μηχανισμοί αυθεντικοποίησης, απαιτούν επεξεργαστική ισχύ και μνήμη που συχνά υπερβαίνουν τις δυνατότητες αυτών των συσκευών. Ως αποτέλεσμα, η ασφάλεια στηρίζεται σε ελαφρύτερες λύσεις που δεν προσφέρουν το ίδιο επίπεδο προστασίας [25].

Μέγεθος των δικτύων IoT. Τα δίκτυα IoT χαρακτηρίζονται από τεράστιο αριθμό συνδεδεμένων συσκευών. Η επιφάνεια επίθεσης (attack surface) είναι εκτεταμένη, αυξάνοντας τις πιθανότητες

παραβίασης. Επιπλέον, η συνεχόμενη εποπτεία του δικτύου είναι πιο δύσκολη λόγω της διασποράς και του πλήθους των συσκευών, επιτρέποντας σε επιθέσεις να παραμένουν κρυφές για μεγαλύτερο χρονικό διάστημα. Τέλος, σε περίπτωση εντοπισμού επίθεσης, η απόκριση μπορεί να καθυστερήσει λόγω φυσικών αποστάσεων ή καθυστερήσεων στη μετάδοση [25].

Πλήθος δεδομένων. Τα δίκτυα IoT παράγουν τεράστιες ποσότητες δεδομένων, συχνά σε συνεχή ροή. Η ανάλυσή τους επιβαρύνει σημαντικά τα συστήματα ανίχνευσης εισβολών, καθώς απαιτεί χρόνο και μεγάλη υπολογιστική ισχύ. Όταν οι υποδομές δεν μπορούν να διαχειριστούν τον αυξημένο φόρτο, η ανίχνευση ανωμαλιών και η απόκριση σε αυτές καθυστερεί ή αποτυγχάνει [3].

Ανομοιογένεια συσκευών. Τα οικοσυστήματα IoT περιλαμβάνουν ποικιλία ετερογενών συσκευών, που διαφέρουν ως προς το υλικό, τα πρωτόκολλα επικοινωνίας, το λειτουργικό σύστημα, τη μνήμη και την επεξεργαστική ισχύ. Αυτό καθιστά ιδιαίτερα δύσκολη την εφαρμογή ενιαίων μηχανισμών προστασίας, όπως πρότυπα αυθεντικοποίησης, κρυπτογράφησης ή ενημερώσεων λογισμικού, καθώς ο κάθε τύπος συσκευής έχει διαφορετικές δυνατότητες και περιορισμούς [1].

Φυσική προσβασιμότητα. Πολλές συσκευές IoT βρίσκονται εγκατεστημένες σε τοποθεσίες με εύκολη φυσική πρόσβαση, όπως δημόσιοι χώροι, εργοστάσια και ανοιχτά περιβάλλοντα. Η απουσία φυσικής προστασίας επιτρέπει σε κακόβουλους χρήστες να επέμβουν απευθείας στο υλικό ή στο λογισμικό των συσκευών, τροποποιώντας τη λειτουργία τους ή αποκτώντας μη εξουσιοδοτημένη πρόσβαση σε δεδομένα. Σε πολλές περιπτώσεις, τέτοιες επεμβάσεις μπορούν να πραγματοποιηθούν χωρίς ιδιαίτερο τεχνικό εξοπλισμό ή εξειδικευμένες γνώσεις [3].

Ευαίσθητα δεδομένα. Τα δίκτυα IoT συλλέγουν και επεξεργάζονται δεδομένα που είναι συχνά ευαίσθητα ή προσωπικά, όπως τοποθεσίες, βιομετρικά στοιχεία ή πληροφορίες υγείας. Η μετάδοση αυτών των δεδομένων ενδέχεται να θέσει σε κίνδυνο την ιδιωτικότητα των χρηστών, καθώς βασίζεται σε ασύρματα πρωτόκολλα επικοινωνίας, όπως Wi-Fi και Bluetooth, τα οποία δεν προσφέρουν επαρκή προστασία των δεδομένων. Όταν δεν εφαρμόζονται κατάλληλες τεχνικές κρυπτογράφησης και αυθεντικοποίησης, τα πρωτόκολλα αυτά καθίστανται ευάλωτα σε επιθέσεις υποκλοπής, παρεμβολών ή spoofing [25].

Εξάρτηση επιπέδων. Ένα δίκτυο IoT συνδυάζει συσκευές, υποδομές δικτύου και εφαρμογές σε ένα σύνθετο και πολυεπίπεδο σύστημα. Οι επιθέσεις σε ένα επίπεδο μπορούν να επηρεάσουν και τα υπόλοιπα. Για παράδειγμα, μια DoS/DDoS επίθεση στο Perception Layer μπορεί να προκαλέσει διακοπή λειτουργιών στα ανώτερα επίπεδα, με συνέπειες που επηρεάζουν ολόκληρο το δίκτυο και τις υπηρεσίες που παρέχει [3].

Έλλειψη καθολικών προτύπων ασφάλειας. Μέχρι σήμερα, δεν υπάρχουν ενιαία και καθολικά αποδεκτά πρότυπα ασφάλειας για το IoT [3]. Κάθε κατασκευαστής μπορεί να εφαρμόζει διαφορετικές λύσεις, πρωτόκολλα, επίπεδα προστασίας και τρόπους διαχείρισης. Αυτή η έλλειψη κοινών προδιαγραφών δυσκολεύει την εφαρμογή μηχανισμών ασφάλειας και οδηγεί σε κενά που μπορούν να εκμεταλλευτούν οι επιτιθέμενοι.

3.4 Είδη Επιθέσεων

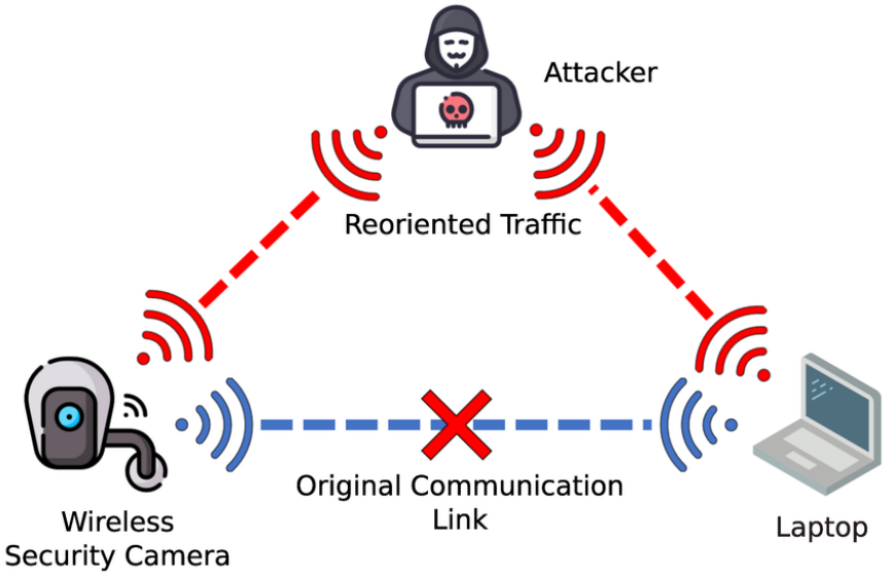
Η ανοιχτή και κατανεμημένη φύση των IoT οικοσυστημάτων τα καθιστά ιδιαίτερα ευάλωτα σε διαφορετικά είδη επιθέσεων. Οι απειλές ποικίλλουν ως προς το στόχο, την τεχνική προσέγγιση και τις επιπτώσεις τους, επηρεάζοντας τα δεδομένα, τη διαθεσιμότητα ή τη λειτουργία των συστημάτων. Ανάλογα με το σκοπό της επίθεσης, διακρίνονται τέσσερις βασικές κατηγορίες επιθέσεων.

3.4.1 Επιθέσεις Παραβίασης Δεδομένων & Απορρήτου

Η επεξεργασία και η ανταλλαγή ευαίσθητων δεδομένων αποτελεί σημαντικό μέρος της καθημερινής λειτουργίας των συσκευών IoT. Τα δεδομένα αυτά είναι ευάλωτα σε επιθέσεις που στοχεύουν στην παρακολούθηση, την υποκλοπή, την αλλοίωση ή την καταστροφή τους.

Παραβιάσεις Δεδομένων (Data Breaches). Τα ευαίσθητα δεδομένα που συλλέγονται από IoT συσκευές, όπως προσωπικές πληροφορίες ή δεδομένα υγείας, μπορούν να υποκλαπούν μέσω μη εξουσιοδοτημένης πρόσβασης, θέτοντας σε κίνδυνο το απόρρητο των χρηστών. Ένα χαρακτηριστικό παράδειγμα καταγράφηκε το 2015, όταν το VTech Hack εξέθεσε προσωπικά δεδομένα εκατομμυρίων ενηλίκων και παιδιών, περιλαμβάνοντας στοιχεία όπως ονόματα, διευθύνσεις, ημερομηνίες γέννησης, διευθύνσεις IP και ιστορικό λήψεων. Η επίθεση εκμεταλλεύτηκε ευπάθεια τύπου SQL injection για να αποκτήσει πρόσβαση στη βάση δεδομένων της πλατφόρμας Learning Lodge και ανέδειξε σοβαρές ελλείψεις στην ασφάλεια των συστημάτων της VTech, όπως η χρήση μη ασφαλών αλγορίθμων κατακερματισμού (MD5) για την αποθήκευση κωδικών πρόσβασης, καθώς και η απουσία ασφαλών πρωτοκόλλων μετάδοσης [26].

Επιθέσεις Man-in-the-Middle (MiTM Attacks). Σε επιθέσεις τύπου Man-in-the-Middle, ένας κακόβουλος χρήστης παρεμβάλλεται στην επικοινωνία μεταξύ δύο IoT συσκευών ή μεταξύ μιας συσκευής και ενός server, αναλαμβάνοντας τον ρόλο του «ενδιάμεσου». Η πρόσβαση μπορεί να αποκτηθεί μέσω μη ασφαλών ασύρματων δικτύων, όπως ανοιχτά Wi-Fi, ή μέσω παραβίασης κάποιου router που μεσολαβεί στην επικοινωνία. Ο επιτιθέμενος έχει τη δυνατότητα να παρακολουθεί την ανταλλαγή δεδομένων ή ακόμα και να τα τροποποιεί, χωρίς να γίνεται αντιληπτός από τις δύο πλευρές. Μπορεί να υποκλέψει στοιχεία ταυτοποίησης (credentials), να αλλοιώσει εντολές, ή να ανακατευθύνει την επικοινωνία προς κακόβουλους servers [27].



Σχήμα 3.1: Επίθεση Man-in-the-Middle [28]

Eavesdropping. Είναι επίθεση όπου ο επιτιθέμενος υποκλέπτει δεδομένα που μεταδίδονται σε ένα μη ασφαλές δίκτυο, χωρίς να αλληλεπιδρά με την επικοινωνία. Η πρόσβαση επιτυγχάνεται μέσω παρακολούθησης του καναλιού επικοινωνίας, ιδιαίτερα όταν δεν χρησιμοποιούνται ισχυροί μηχανισμοί κρυπτογράφησης. Ο επιτιθέμενος καταγράφει τα πακέτα δεδομένων και αποκτά πρόσβαση σε ευαίσθητες πληροφορίες. Πρόκειται για παθητική επίθεση που δεν τροποποιεί την επικοινωνία και

μπορεί να πραγματοποιηθεί χωρίς να γίνει αντιληπτή, παραβιάζοντας την εμπιστευτικότητα των δεδομένων [29].

Επιθέσεις Επαναμετάδοσης (Replay Attacks). Ο επιτιθέμενος καταγράφει πακέτα δεδομένων που ανταλλάσσονται μεταξύ δύο συσκευών και τα αναμεταδίδει αυτούσια σε αργότερη χρονική στιγμή. Στόχος είναι να αναγκάσει το σύστημα να εκτελέσει μία ενέργεια που έχει ήδη πραγματοποιηθεί, χωρίς την έγκριση των συσκευών που επικοινωνούν. Οι επιθέσεις επαναμετάδοσης εκμεταλλεύονται την απουσία μηχανισμών χρονικής επικύρωσης, όπως χρονικές σφραγίδες (timestamps) ή μοναδικά αναγνωριστικά μηνυμάτων. Είναι ιδιαίτερα επικίνδυνες σε περιβάλλοντα όπου η επανάληψη εντολών μπορεί να προκαλέσει παραβίαση φυσικής ασφάλειας ή ανεπιθύμητη ενεργοποίηση λειτουργιών, όπως στο άνοιγμα έξυπνων κλειδαριών ή στον έλεγχο βιομηχανικών μηχανημάτων [27].

Αντιστροφή Μοντέλου (Model Inversion Attacks). Αποτελεί μια προηγμένη απειλή που στοχεύει σε μοντέλα Τεχνητής Νοημοσύνης και επιτρέπει στους επιτιθέμενους να ανακτήσουν ευαίσθητα δεδομένα από το μοντέλο κατά την εκπαίδευσή του. Μέσω στοχευμένων ερωτημάτων προς το σύστημα, οι επιτιθέμενοι μπορούν να αναστρέψουν τη διαδικασία πρόβλεψης και να ανακατασκευάσουν τα αρχικά δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση, αποκτώντας ιδιωτικές πληροφορίες. Η απειλή είναι ιδιαίτερα σοβαρή όταν πρόκειται για εφαρμογές που βασίζονται σε βιομετρικά δεδομένα, όπως έξυπνες κάμερες αναγνώρισης προσώπου ή φωνητικοί βοηθοί. Μέσω τέτοιων εφαρμογών, ένας επιτιθέμενος μπορεί να ανακατασκευάσει εικόνες προσώπων ή δείγματα φωνής των χρηστών, παραβιάζοντας την ιδιωτικότητα και δημιουργώντας σοβαρούς κινδύνους ασφάλειας [30].

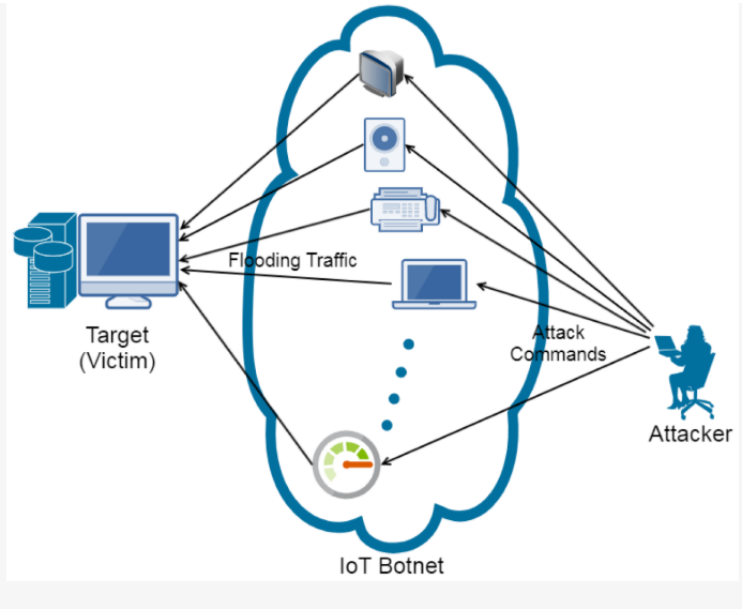
3.4.2 Επιθέσεις Διακοπής Υπηρεσιών

Η ομαλή λειτουργία και η διαθεσιμότητα των υπηρεσιών αποτελούν βασικές και κρίσιμες απαιτήσεις στα IoT περιβάλλοντα. Οι επιθέσεις που στοχεύουν στη διακοπή των υπηρεσιών επηρεάζουν την αξιοπιστία τους, είτε μέσω υπερφόρτωσης, είτε μέσω κακόβουλου λογισμικού.

Επιθέσεις Denial-of-Service (DoS Attacks). Ονομάζονται επιθέσεις άρνησης υπηρεσίας και αποτελούν έναν από τους πιο διαδεδομένους και καταστροφικούς τύπους επιθέσεων στο IoT. Εκμεταλλεύονται τους περιορισμένους πόρους των IoT συσκευών, με στόχο να καταστήσουν μία συσκευή, ένα δίκτυο ή μία υπηρεσία μη διαθέσιμα μέσω υπερφόρτωσης. Οι επιτιθέμενοι στέλνουν επανειλημμένα μεγάλο όγκο αιτημάτων ή πακέτων δεδομένων, με σκοπό να εξαντλήσουν την υπολογιστική ισχύ, τη μνήμη ή το bandwidth της συσκευής-στόχου. Οι IoT συσκευές αδυνατούν να διαχειριστούν τον αυξημένο όγκο της κίνησης και καθίστανται ιδιαίτερα ευάλωτες. Αυτό οδηγεί σε καθυστερήσεις ή πλήρη διακοπή λειτουργίας για τους κανονικούς χρήστες [27].

Επιθέσεις Distributed Denial-of-Service (DDoS Attacks). Αποτελούν παραλλαγή των DoS επιθέσεων, στις οποίες η κίνηση προέρχεται από πολλαπλές συσκευές ταυτόχρονα. Οι επιτιθέμενοι χρησιμοποιούν ένα δίκτυο μολυσμένων συσκευών, γνωστό ως botnet, το οποίο συχνά περιλαμβάνει IoT συσκευές που έχουν παραβιαστεί μέσω αδύναμων κωδικών ή μη ενημερωμένου firmware. Η επίθεση πραγματοποιείται από πολλές γεωγραφικά διασκορπισμένες πηγές, γεγονός που καθιστά δύσκολη την ανίχνευσή της. Η δυσκολία εντοπισμού οφείλεται επίσης στο γεγονός ότι η κακόβουλη κίνηση μπορεί να μοιάζει με κανονική δραστηριότητα [29]. Χαρακτηριστικό παράδειγμα αποτελεί το Mirai Botnet (2016), το οποίο εκμεταλλεύτηκε χιλιάδες κάμερες και routers με αδύναμα ή default credentials για να εξαπολύσει επίθεση DDoS στον πάροχο DNS Dyn. Η επίθεση προκάλεσε εκτεταμένες διακοπές λειτουργίας σε δημοφιλείς υπηρεσίες, όπως το Twitter, το Netflix και το Reddit, σε ΗΠΑ και Ευρώπη [31]. Το Mirai ήταν από τα πρώτα botnets που στόχευσε αποκλειστικά σε IoT συσκευές και διαδόθηκε

ταχύτατα, αναδεικνύοντας τη μαζική αδυναμία ασφάλειας που χαρακτηρίζει πολλές IoT συσκευές και οικοσυστήματα [33].



Σχήμα 3.2: Επίθεση DDoS [34]

Επιθέσεις Κακόβουλου Λογισμικού (Malware Attacks). Είναι ένας από τους συνηθέστερους και πιο επικίνδυνους τύπους επιθέσεων στο IoT. Το κακόβουλο λογισμικό (malware) είναι κάθε είδους λογισμικό που έχει σχεδιαστεί για να προκαλέσει βλάβη ή να επιτρέψει μη εξουσιοδοτημένη πρόσβαση σε συσκευές και συστήματα. Η εγκατάστασή του γίνεται μέσω ευπαθειών στο firmware, μη ασφαλών ενημερώσεων ή προσβάσιμων θυρών και υπηρεσιών. Μόλις ενεργοποιηθεί, μπορεί να υποκλέψει δεδομένα, να αναλάβει τον έλεγχο της συσκευής ή να την εντάξει σε botnet. Αυτά τα botnets χρησιμοποιούνται συχνά για την εξαπόλυση επιθέσεων DDoS ή για την παρακολούθηση των χρηστών [29]. Στο IoT, η διάδοσή του είναι ιδιαίτερα αποτελεσματική, καθώς πολλές συσκευές διαθέτουν ελάχιστη προστασία, χρησιμοποιούν προεπιλεγμένα credentials και δεν εφαρμόζουν βασικούς μηχανισμούς ενημέρωσης ή ελέγχου πρόσβασης [33].

Firmware Exploits. Οι επιθέσεις αυτού του τύπου αποτελούν μία υποκατηγορία των Malware Attacks και στοχεύουν αποκλειστικά στο firmware, δηλαδή στο χαμηλού επιπέδου λογισμικό που είναι ενσωματωμένο στο υλικό (hardware) μιας συσκευής. Πολλές IoT συσκευές βασίζουν τη λειτουργία τους αποκλειστικά στο firmware και δεν έχουν πρόσβαση σε πιο σύνθετα επίπεδα ασφάλειας. Επιπρόσθετα, σε πολλές περιπτώσεις δεν υποστηρίζονται τακτικές ή αυτόματες ενημερώσεις ασφαλείας, με αποτέλεσμα να παραμένουν εκτεθειμένες σε ευπάθειες. Οι επιτιθέμενοι μπορούν να εκμεταλλευτούν αυτές τις ευπάθειες για να αποκτήσουν πλήρη έλεγχο της συσκευής [32]. Ένα γνωστό παράδειγμα firmware επίθεσης που στόχευε IoT συσκευές είναι το BrickerBot (2017). Αφού αποκτούσε πρόσβαση στις συσκευές μέσω Telnet, εκτελούσε εντολές που κατέστρεφαν το σύστημα αρχείων και κρίσιμα στοιχεία του λειτουργικού συστήματος της συσκευής. Ως αποτέλεσμα, η συσκευή δεν μπορούσε να αποκατασταθεί χωρίς φυσική πρόσβαση και επαναπρογραμματισμό του firmware. Επηρεάστηκαν χιλιάδες συσκευές όπως οικιακές κάμερες και routers, αναδεικνύοντας τις ευπάθειες πολλών IoT συσκευών απέναντι σε τέτοιες επιθέσεις [33].

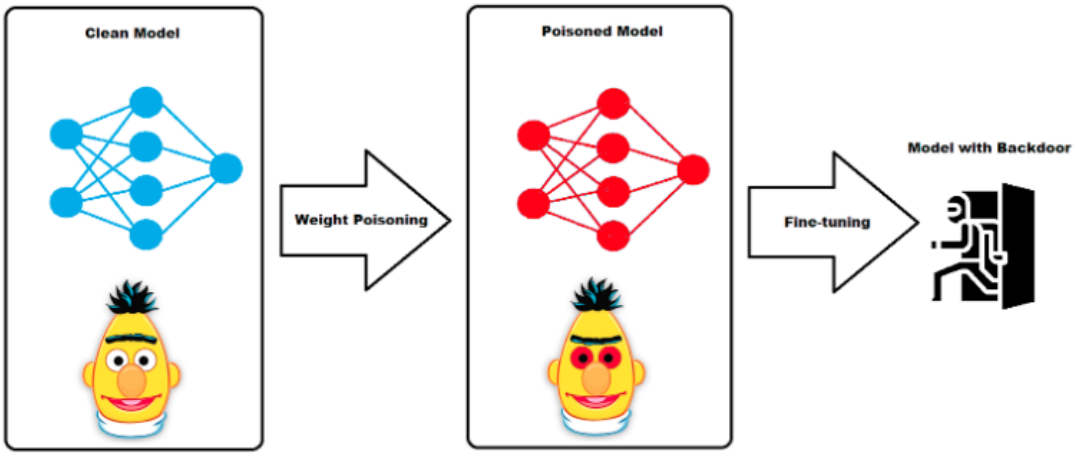
Ransomware. Οι επιθέσεις Ransomware είναι μία εξέλιξη των επιθέσεων Malware. Ο επιτιθέμενος εγκαθιστά κακόβουλο λογισμικό σε μία IoT συσκευή και κρυπτογραφεί τα δεδομένα της ή μπλοκάρει τη λειτουργικότητά της, απαιτώντας την καταβολή λύτρων για την αποκατάσταση της πρόσβασης [35].

Οι επιθέσεις αυτού του τύπου στο IoT δεν θέτουν μόνο οικονομικούς κινδύνους, αλλά μπορεί να έχουν και άμεσες επιπτώσεις . Ένα χαρακτηριστικό παράδειγμα είναι το Ekans Ransomware (2020), που στόχευσε περιβάλλοντα IIoT [36]. Η εισβολή στο σύστημα πραγματοποιούνταν μέσω ευπαθειών, όπως ανοιχτές θύρες ή μη ασφαλείς ρυθμίσεις απομακρυσμένης πρόσβασης. Το Ekans κρυπτογραφούσε αρχεία προσθέτοντας τυχαίους χαρακτήρες στις επεκτάσεις των αρχείων και διαγράφοντας τα αντίγραφα ασφαλείας, καθιστώντας την ανάκτησή τους αδύνατη. Επηρέασε εταιρίες όπως η Honda, προκαλώντας διακοπές στην παραγωγή και σημαντικές οικονομικές απώλειες. Η στοχευμένη φύση του το καθιστά ιδιαίτερα επικίνδυνο για περιβάλλοντα IoT.

3.4.3 Επιθέσεις Παραποίησης Λειτουργίας

Είναι επιθέσεις που στοχεύουν στον επηρεασμό της κανονικής λειτουργίας του συστήματος IoT, ιδιαίτερα σε εκείνα που βασίζονται σε τεχνικές τεχνητής νοημοσύνης. Συνήθως εστιάζουν στην αλλοίωση των εισόδων ή της διαδικασίας εκμάθησης, με σκοπό την παραπλάνηση του μοντέλου και την παραγωγή εσφαλμένων ή μη ασφαλών συμπερασμάτων.

Μόλυνση Δεδομένων (Data Poisoning Attacks). Τέτοιες επιθέσεις στοχεύουν στην παραποίηση της διαδικασίας εκπαίδευσης ενός μοντέλου τεχνητής νοημοσύνης, με σκοπό τη μείωση της ακρίβειας ή την πρόκληση σφαλμάτων κατά τη φάση της πρόβλεψης. Οι επιτιθέμενοι εισάγουν σκόπιμα παραποιημένα δείγματα στο σύνολο εκπαίδευσης, οδηγώντας το μοντέλο στην εκμάθηση λανθασμένων προτύπων. Αυτό επιτυγχάνεται είτε μέσω άμεσης πρόσβασης σε μη ασφαλισμένες συσκευές ή κόμβους του συστήματος, είτε έμμεσα, μέσω τρίτων πηγών δεδομένων, ιδίως όταν το σύστημα χρησιμοποιεί για την εκπαίδευσή του δεδομένα που προέρχονται από ανοιχτές πηγές (όπως εταιρίες, πλατφόρμες, APIs ή χρήστες). Στο περιβάλλον του IoT, τέτοιες επιθέσεις μπορεί να επιτρέψουν σε κακόβουλες ενέργειες να περάσουν απαρατήρητες, χαρακτηρίζοντάς τες ως φυσιολογικές. Για παράδειγμα, ένας επιτιθέμενος θα μπορούσε να εισάγει δεδομένα που αποκρύπτουν τη μη φυσιολογική συμπεριφορά αισθητήρων, ώστε να παρακάμπτεται η ανίχνευση εισβολών. Οι επιπτώσεις των επιθέσεων Data Poisoning είναι ιδιαίτερα κρίσιμες σε έξυπνες υποδομές και βιομηχανικά συστήματα IoT, καθώς ενδέχεται να οδηγήσουν σε λανθασμένες αποφάσεις με πραγματικές συνέπειες στη λειτουργικότητα και την ασφάλεια του συστήματος [39].



Σχήμα 3.3: Επίθεση Data Poisoning [37]

Adversarial Attacks. Οι συγκεκριμένες επιθέσεις στοχεύουν άμεσα σε συστήματα που βασίζονται σε τεχνητή νοημοσύνη. Ο επιτιθέμενος εισάγει μικρές αλλά στρατηγικά σχεδιασμένες τροποποιήσεις στα

δεδομένα εισόδου, με στόχο να επηρεάσει την έξοδο του μοντέλου και να οδηγήσει σε λανθασμένες αποφάσεις. Στο IoT, τέτοιες επιθέσεις μπορούν να οδηγήσουν σε παράκαμψη των μηχανισμών ασφαλείας και σε μη εξουσιοδοτημένη πρόσβαση σε συσκευές ή δίκτυα. Σε αντίθεση με τις επιθέσεις Data Poisoning, που στοχεύουν στη φάση εκπαίδευσης του μοντέλου, οι επιθέσεις Adversarial πραγματοποιούνται κυρίως κατά τη φάση πρόβλεψης του μοντέλου, επηρεάζοντας τη λειτουργία ενός ήδη εκπαιδευμένου συστήματος. Οι επιτιθέμενοι κατασκευάζουν κακόβουλα δείγματα εισόδου, που εξωτερικά μοιάζουν κανονικά αλλά έχουν σχεδιαστεί ώστε να προκαλούν λανθασμένες εξόδους. Για παράδειγμα, ένα σύστημα εντοπισμού εισβολών (Intrusion Detection System – IDS) ενδέχεται να αποτύχει να αναγνωρίσει κακόβουλη δραστηριότητα, εάν τα πακέτα δεδομένων έχουν παραποιηθεί με τέτοιο τρόπο ώστε να μοιάζουν με κανονική δικτυακή κίνηση. Η αντιμετώπιση των adversarial επιθέσεων απαιτεί την εφαρμογή προηγμένων τεχνικών άμυνας, όπως μεθόδους ανίχνευσης ανωμαλιών και την ανάπτυξη ανθεκτικών μοντέλων που βασίζονται σε μηχανική μάθηση [38].

Παραποίηση Δεδομένων Αισθητήρων (Sensor Spoofing Attacks). Οι συσκευές IoT βασίζονται σε δεδομένα αισθητήρων για τη λήψη κρίσιμων αποφάσεων σε πραγματικό χρόνο. Οι επιθέσεις Sensor Spoofing εκμεταλλεύονται ευπάθειες στο φυσικό ή στο λογισμικό των αισθητήρων IoT και παραποιούν τα δεδομένα που συλλέγονται, χωρίς να παραβιάζουν άμεσα το σύστημα. Ως αποτέλεσμα, μπορεί να προκληθεί λανθασμένη ενεργοποίηση ή απενεργοποίηση συστημάτων ή λειτουργιών με βάση μετρήσεις που δεν ανταποκρίνονται στις πραγματικές καταστάσεις. Για παράδειγμα, στις έξυπνες πόλεις, κακόβουλοι χρήστες μπορεί να αλλοιώσουν τις ενδείξεις θερμοκρασίας, ρύπανσης ή κυκλοφορίας, προκαλώντας εσφαλμένες αποφάσεις, όπως σφάλματα στη λειτουργία των φαναριών ή ενεργοποίηση συστημάτων συναγερμού. Σε βιομηχανικά περιβάλλοντα IoT, η παραποίηση των αισθητήρων θερμοκρασίας ή πίεσης μπορεί να οδηγήσει σε σοβαρές βλάβες ή ακόμη και ατυχήματα. Οι επιθέσεις αυτού του τύπου είναι ιδιαίτερα δύσκολες στην ανίχνευση, καθώς δεν παραβιάζουν άμεσα το σύστημα, αλλά αλλοιώνουν τα δεδομένα που λαμβάνει μέσω των μετρήσεων [40].

3.4.4 Επιθέσεις Φυσικής Πρόσβασης

Λόγω της έκθεσης τους στο φυσικό περιβάλλον, πολλές IoT συσκευές είναι ευάλωτες σε επιθέσεις φυσικής πρόσβασης, όπως κλοπή, βανδαλισμό ή μη εξουσιοδοτημένη πρόσβαση. Οι επιθέσεις αυτές δεν απαιτούν πάντα ψηφιακές τεχνικές, αλλά προϋποθέτουν άμεση φυσική πρόσβαση στη συσκευή. Ο επιτιθέμενος μπορεί να εξαγάγει ευαίσθητα δεδομένα απευθείας από το υλικό, να τροποποιήσει ή να αντικαταστήσει εξαρτήματα, ή να εγκαταστήσει κακόβουλο λογισμικό μέσω φυσικής σύνδεσης, όπως USB ή σειριακή θύρα. Η καταστροφή ή απώλεια της συσκευής μπορεί να προκαλέσει σοβαρά οικονομικά κόστη λόγω της αντικατάστασης. Ένα χαρακτηριστικό παράδειγμα είναι η παραβίαση έξυπνων κλειδαριών ή καμερών ασφαλείας, όπου ένας επιτιθέμενος αποκτά πρόσβαση και εκμεταλλεύεται την έλλειψη προστασίας έναντι φυσικής παραβίασης ή την απουσία μηχανισμών ανίχνευσης παρέμβασης [27].

3.5 Παραδοσιακά Συστήματα Εντοπισμού Εισβολών

Τα συστήματα εντοπισμού εισβολών αποτελούν έναν από τους βασικούς μηχανισμούς ασφάλειας σε δικτυακά περιβάλλοντα, καθώς έχουν σχεδιαστεί για να αναγνωρίζουν ύποπτη δραστηριότητα ή κακόβουλη κίνηση εντός ενός συστήματος. Λειτουργούν ως μηχανισμοί παρακολούθησης που αναλύουν τα εισερχόμενα δεδομένα με στόχο την ανίχνευση επιθέσεων ή αποκλίσεων από τη φυσιολογική συμπεριφορά [41].

Τα IDS βασίζονται είτε σε δεδομένα από τη λειτουργία του ίδιου του συστήματος είτε από βάσεις γνωστών υπογραφών επίθεσης (signatures), αξιοποιώντας τα για την αναγνώριση ανεπιθύμητων ενεργειών. Ανάλογα με τον τρόπο υλοποίησης, τα IDS διακρίνονται σε Host-based IDS (HIDS) και Network-based IDS (NIDS).

Τα HIDS παρακολουθούν τη δραστηριότητα σε επίπεδο συσκευής, ανιχνεύοντας ενέργειες όπως προσπάθειες πρόσβασης σε αρχεία συστήματος ή αλλαγές σε κρίσιμες ρυθμίσεις ασφάλειας. Αντίθετα, τα NIDS εστιάζουν στην ανάλυση της κυκλοφορίας δεδομένων σε επίπεδο δικτύου, αναλύοντας τη ροή των πακέτων, με στόχο τον εντοπισμό αποκλίσεων από την κανονική κίνηση και την αναγνώριση πιθανών απειλών. Η κάθε προσέγγιση έχει πλεονεκτήματα και περιορισμούς. Τα HIDS παρέχουν λεπτομερή παρακολούθηση σε επίπεδο συσκευής, αλλά έχουν τοπική εμβέλεια. Τα NIDS παρέχουν ευρύτερη κάλυψη σε κατανεμημένα περιβάλλοντα, χωρίς όμως να εντοπίζουν αλλαγές εντός των ίδιων των συσκευών [42].

Στο πλαίσιο του IoT, η εφαρμογή των IDS είναι απαραίτητη, καθώς πολλά συστήματα βασίζονται στη συλλογή και επεξεργασία δεδομένων σε πραγματικό χρόνο. Η παραβίαση της εμπιστευτικότητας ή της ακεραιότητας των δεδομένων μπορεί να οδηγήσει σε σοβαρές συνέπειες, τόσο για την ιδιωτικότητα των χρηστών όσο και για τη λειτουργικότητα ολόκληρου του IoT οικοσυστήματος.

Ωστόσο, η εφαρμογή των παραδοσιακών IDS σε περιβάλλοντα IoT παρουσιάζει προκλήσεις. Οι IoT συσκευές συχνά διαθέτουν περιορισμένους πόρους επεξεργασίας και αποθήκευσης, περιορίζοντας τη δυνατότητα ενσωμάτωσης πολύπλοκων μηχανισμών ασφάλειας [43]. Επιπλέον, η ετερογένεια των συσκευών, των πρωτοκόλλων επικοινωνίας και των αρχιτεκτονικών καθιστά δύσκολη την ανάπτυξη καθολικής λύσης. Αν και έχουν αναπτυχθεί πολλές διαφορετικές προσεγγίσεις IDS για το IoT, δεν υπάρχει ακόμη ενιαία λύση που να καλύπτει τις δυναμικές ανάγκες των διαφορετικών οικοσυστημάτων και η επιλογή της κατάλληλης τεχνικής παραμένει ανοιχτό ερευνητικό πεδίο [42].

3.5 Επίλογος

Σε αυτό το κεφάλαιο ορίστηκε η έννοια της ψηφιακής ασφάλειας και αναλύθηκαν οι θεμελιώδεις αρχές της, που αποτελούν τη βάση για την προστασία των πληροφοριακών συστημάτων. Δόθηκε έμφαση στο πώς η ασφάλεια αποκτά ιδιαίτερη σημασία στο περιβάλλον του IoT, λόγω της μεγάλης ετερογένειας, της περιορισμένης υπολογιστικής ισχύος και άλλων ιδιαίτερων χαρακτηριστικών του. Αναλύθηκαν οι βασικές κατηγορίες επιθέσεων που απειλούν την εμπιστευτικότητα, ακεραιότητα και λειτουργικότητα των συστημάτων IoT. Τέλος, έγινε σύντομη αναφορά στα παραδοσιακά συστήματα εντοπισμού εισβολών και επισημάνθηκαν οι προκλήσεις που δυσκολεύουν την αποτελεσματική εφαρμογή τους στο IoT.

Κεφάλαιο 4ο: Τεχνητή Νοημοσύνη στην Ασφάλεια του IoT

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο εξετάζεται ο ρόλος της Τεχνητής Νοημοσύνης στο πλαίσιο του IoT, με έμφαση στις βασικές τεχνικές που χρησιμοποιούνται για τον εντοπισμό εισβολών στα οικοσυστήματα IoT. Στη συνέχεια, παρουσιάζονται συνοπτικά τα πλεονεκτήματα και οι προκλήσεις που προκύπτουν από την ενσωμάτωσή της σε τέτοια περιβάλλοντα.

4.2 Ο ρόλος της Τεχνητής Νοημοσύνης στο IoT

Καθώς τα IoT οικοσυστήματα εξελίσσονται και επεκτείνονται, γίνονται όλο και πιο σύνθετα. Αντίστοιχα, αυξάνονται οι απαιτήσεις ασφάλειας. Οι παραδοσιακές μέθοδοι συχνά αποτυγχάνουν να ανταποκριθούν αποτελεσματικά σε συνθήκες πραγματικού χρόνου ή να προσαρμοστούν στις δυναμικές απαιτήσεις αυτών των συστημάτων. Η Τεχνητή Νοημοσύνη προσφέρει νέες προσεγγίσεις, επιτρέποντας την ανάλυση μεγάλου όγκου δεδομένων, την αναγνώριση προτύπων και τη λήψη αποφάσεων με ελάχιστη ή καθόλου ανθρώπινη παρέμβαση [1].

Η Τεχνητή Νοημοσύνη είναι ο κλάδος της επιστήμης των υπολογιστών που στοχεύει στην ανάπτυξη συστημάτων με δυνατότητα αντίληψης, μάθησης και λήψης αποφάσεων, μιμούμενων την ανθρώπινη νοημοσύνη. Περιλαμβάνει ένα ευρύ φάσμα τεχνικών και εφαρμογών και εφαρμόζεται σε πολλούς τομείς, με στόχο τη βελτιστοποίηση διαδικασιών και τη λήψη αποφάσεων. Στο πλαίσιο των IoT συστημάτων, η AI χρησιμοποιείται για την ανίχνευση μη φυσιολογικής δραστηριότητας στη δικτυακή συμπεριφορά (anomaly detection), για εντοπισμό απειλών (intrusion detection), για την κατανομή πόρων (resource allocation), αλλά και για την αυτοματοποίηση της απόκρισης, στοχεύοντας στη μείωση της ανθρώπινης παρέμβασης. Μέσω τεχνικών παρέχει προσαρμοστικές λύσεις που ενισχύουν την ασφάλεια και την απόδοση των δικτύων IoT, αντιμετωπίζοντας τους περιορισμούς των παραδοσιακών μηχανισμών εντοπισμού εισβολών.

4.3 Τεχνικές Τεχνητής Νοημοσύνης

Η Τεχνητή Νοημοσύνη περιλαμβάνει ένα ευρύ φάσμα τεχνικών που χρησιμοποιούνται για διαφορετικούς στόχους, όπως η ταξινόμηση ή η αναγνώριση προτύπων. Στο πλαίσιο της ασφάλειας των IoT οικοσυστημάτων, έχουν ιδιαίτερη σημασία δύο βασικές προσεγγίσεις [3].

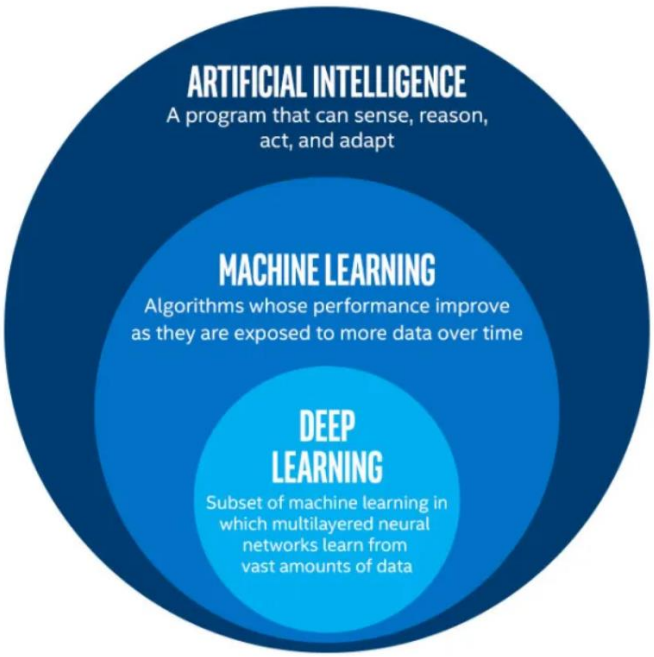
4.3.1 Μηχανική Μάθηση

Οι τεχνικές που βασίζονται σε Μηχανική Μάθηση (Machine Learning – ML) έχουν αποτελέσει αντικείμενο μελέτης στον τομέα του εντοπισμού εισβολών από τη δεκαετία του 1990. Οι πρώτες προσεγγίσεις χρησιμοποίησαν αλγορίθμους όπως τα δέντρα απόφασης (Decision Trees) και τα μοντέλα υποστήριξης διανυσμάτων (Support Vector Machines – SVMs), με σκοπό την ταξινόμηση της δικτυακής κίνησης σε "φυσιολογική" ή "κακόβουλη". Αν και οι μέθοδοι αυτές παρουσίασαν βελτιωμένα ποσοστά εντοπισμού σε σύγκριση με τις παραδοσιακές προσεγγίσεις, αντιμετώπιζαν δυσκολίες σε περιπτώσεις όπου τα δεδομένα ήταν πολύ υψηλών διαστάσεων ή τα σύνολα ήταν μη ισορροπημένα. Για την αντιμετώπιση αυτών των προβλημάτων, προτάθηκαν τεχνικές προεπεξεργασίας, επιλογής χαρακτηριστικών και στρατηγικές συνδυασμού μοντέλων (ensemble). Παρά τις εξελίξεις

αυτές, οι παραδοσιακές μέθοδοι Μηχανικής Μάθησης εξακολουθούν να δυσκολεύονται να αποτυπώσουν τα πολύπλοκα πρότυπα και τις σχέσεις που εντοπίζονται στη δικτυακή κίνηση. Αυτό οδήγησε σε διερεύνηση πιο εξελιγμένων μεθόδων, όπως αλγορίθμους που βασίζονται σε Βαθιά Μάθηση, οι οποίοι έχουν αποδείξει υψηλότερη ακρίβεια εντοπισμού και καλύτερη ικανότητα εκμάθησης ιεραρχικών αναπαραστάσεων σε δεδομένα υψηλής διάστασης [44].

4.3.2 Βαθιά Μάθηση

Η Βαθιά Μάθηση (Deep Learning – DL) βασίζεται σε τεχνητά νευρωνικά δίκτυα πολλαπλών επιπέδων και χρησιμοποιείται για την επεξεργασία πιο σύνθετων δεδομένων, όπως μεγαλύτερα και πολυδιάστατα σύνολα δεδομένων, εικόνες και ήχους [3]. Οι αλγόριθμοι Βαθιάς Μάθησης έχουν προσελκύσει έντονο ερευνητικό ενδιαφέρον τα τελευταία χρόνια λόγω της δυνατότητάς τους να ενισχύσουν την αποτελεσματικότητα των συστημάτων εντοπισμού εισβολών. Αποδεικνύονται ιδιαίτερα χρήσιμοι στον τομέα της κυβερνοασφάλειας, κυρίως λόγω της ικανότητάς τους να διαχειρίζονται και να αναλύουν μεγάλα και πολύπλοκα σύνολα δεδομένων, όπως αυτά που παράγονται από τα συστήματα IoT. Σε αντίθεση με τις παραδοσιακές μεθόδους, δεν απαιτεί πάντα χειροκίνητη επιλογή χαρακτηριστικών, καθώς μπορεί να μαθαίνει αυτόματα αναπαραστάσεις από τα δεδομένα εισόδου. Είναι κατάλληλη για διαχείριση μεγάλου όγκου μη δομημένων και ετερογενών δεδομένων, που συναντώνται συχνά σε περιβάλλοντα IoT. Επιπλέον, οι τεχνικές Βαθιάς Μάθησης έχουν τη δυνατότητα να γενικεύουν καλύτερα σε νέα, άγνωστα μοτίβα συμπεριφοράς, εντοπίζοντας άγνωστες επιθέσεις (zero-day attacks) [44]. Ωστόσο, η υψηλή υπολογιστική απαίτηση της Βαθιάς Μάθησης καθιστά την υλοποίησή της προκλητική σε συγκεκριμένα IoT περιβάλλοντα. Συνολικά, όμως, η Βαθιά Μάθηση συμπληρώνει τη Μηχανική Μάθηση στην ανάπτυξη πιο ευέλικτων και αποδοτικών λύσεων ασφάλειας, ικανών να ανταποκριθούν στην αυξημένη πολυπλοκότητα και το μεγάλο μέγεθος των σύγχρονων συστημάτων IoT.



Σχήμα 4.1: Σχέση Τεχνητής Νοημοσύνης, Μηχανικής και Βαθιάς Μάθησης [45]

4.4 Πλεονεκτήματα και Περιορισμοί της Χρήσης Τεχνητής Νοημοσύνης στο IoT

Η ενσωμάτωση τεχνικών Τεχνητής Νοημοσύνης στα οικοσυστήματα IoT προσφέρει σημαντικά πλεονεκτήματα. Η ικανότητα των αλγορίθμων να εντοπίζουν σύνθετα πρότυπα σε μεγάλους όγκους δεδομένων με αυξημένη ακρίβεια σε σύγκριση με τις παραδοσιακές μεθόδους, ενισχύει τον εντοπισμό εισβολών [3]. Επιπλέον, η προσαρμοστικότητα των μοντέλων επιτρέπει τη λειτουργία τους σε δυναμικά και μεταβαλλόμενα περιβάλλοντα, όπως το IoT, με την ικανότητα να “μαθαίνουν” από νέα δεδομένα. Η αυτοματοποίηση βασικών διαδικασιών μειώνει την ανάγκη ανθρώπινης παρέμβασης, επιταχύνει τον εντοπισμό απειλών και επιτρέπει την άμεση απόκριση [5].

Ωστόσο, η εφαρμογή της Τεχνητής Νοημοσύνης στο IoT συνοδεύεται από τεχνικές προκλήσεις. Ένα βασικό ζήτημα είναι οι περιορισμένοι υπολογιστικοί πόροι των IoT συσκευών, οι οποίοι αποτελούν πρόκληση για την εκτέλεση απαιτητικών μοντέλων, ειδικά αυτών που βασίζονται στη Βαθιά Μάθηση. Παράλληλα, η εκπαίδευση τέτοιων μοντέλων απαιτεί συνήθως μεγάλο όγκο δεδομένων, γεγονός που καθίσταται δύσκολο σε κάποια περιβάλλοντα IoT. Επιπλέον, η ανάγκη για συνεχή ενημέρωση ή επανεκπαίδευση των μοντέλων, επηρεάζει σημαντικά την αξιοπιστία των συστημάτων [44]. Τέλος, η ερμηνεία των αποφάσεων αποτελεί σημαντικό πρόβλημα, καθώς πολλά μοντέλα λειτουργούν ως "μαύρα κουτιά", καθιστώντας τα δύσκολα, ιδιαίτερα σε κρίσιμες εφαρμογές. Ωστόσο, η ανάπτυξη και υλοποίηση μοντέλων Βαθιάς Μάθησης μπορεί να υποστηριχθεί αποτελεσματικά μέσω edge ή cloud computing, όπου η επεξεργασία γίνεται πλησιέστερα στη συσκευή ή σε απομακρυσμένους πόρους, αντίστοιχα. Η μεταφορά της υπολογιστικής επιβάρυνσης εκτός της IoT συσκευής επιτρέπει την υλοποίηση μοντέλων με μεγαλύτερες απαιτήσεις σε μνήμη και επεξεργαστική ισχύς, διατηρώντας την απόκριση σε αποδεκτά επίπεδα.

4.5 Επίλογος

Σε αυτό το κεφάλαιο παρουσιάστηκε ο τρόπος με τον οποίο η Τεχνητή Νοημοσύνη μπορεί να εφαρμοστεί στο IoT για τον αποδοτικό εντοπισμό εισβολών. Αναλύθηκαν οι δύο βασικές κατηγορίες τεχνικών, η Μηχανική Μάθηση και η Βαθιά Μάθηση και έγινε σύντομη αναφορά στα πλεονεκτήματα και τους περιορισμούς της χρήσης της σε περιβάλλοντα IoT. Η αξιοποίηση της Τεχνητής Νοημοσύνης στο IoT παρουσιάζει σημαντικές δυνατότητες, καθώς μπορεί να ενισχύσει την ασφάλεια και την αποδοτικότητα, αλλά απαιτεί προσεκτικό σχεδιασμό και εφαρμογή, λαμβάνοντας υπόψη τόσο τις τεχνικές απαιτήσεις όσο και τους περιορισμούς του εκάστοτε περιβάλλοντος.

Κεφάλαιο 5ο: Εντοπισμός Εισβολών σε συστήματα IoT

5.1 Εισαγωγή

Το παρόν κεφάλαιο εστιάζει στις μεθόδους εντοπισμού εισβολών και στο ρόλο τους στο πλαίσιο του IoT. Αρχικά παρουσιάζονται οι βασικές κατηγορίες IDS, με στόχο την ταξινόμηση των διαφορετικών προσεγγίσεων. Στη συνέχεια, αναλύονται οι μέθοδοι μάθησης που εφαρμόζονται για ενίσχυση της ακρίβειας και αποδοτικότητας των IDS. Τέλος, παρουσιάζονται οι βασικές μετρικές αξιολόγησης των συστημάτων εντοπισμού εισβολών, με στόχο την ποσοτική εκτίμηση της αποτελεσματικότητάς τους σε IoT περιβάλλοντα.

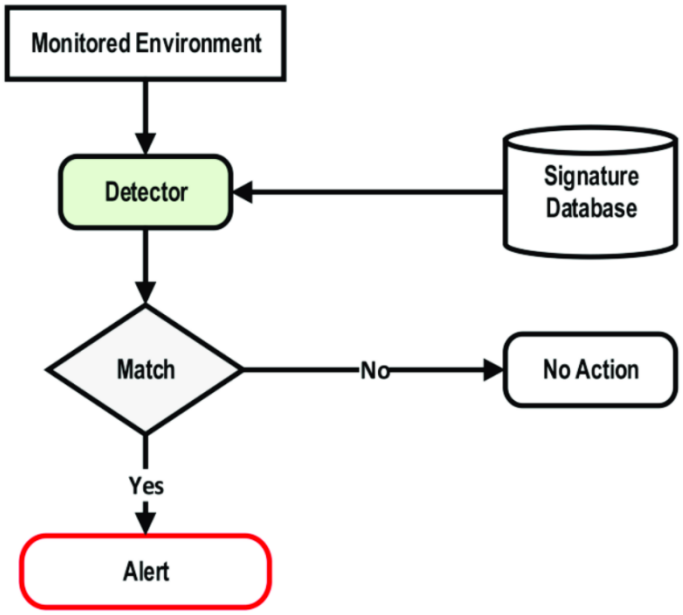
5.2 Κατηγορίες Συστημάτων Εντοπισμού Εισβολών

Διάφορα μοντέλα εντοπισμού εισβολών αξιοποιούν τεχνικές Τεχνητής Νοημοσύνης για την ενίσχυση της ασφάλειας στα δίκτυα IoT, διευκολύνοντας την αναγνώριση και κατηγοριοποίηση κακόβουλης δραστηριότητας. Ανάλογα με τον τρόπο ανίχνευσης των απειλών, οι τεχνικές διαφοροποιούνται σε signature-based, anomaly-based και υβριδικές προσεγγίσεις [46].

5.2.1 Signature Based IDS

Τα Signature-Based IDS λειτουργούν εντοπίζοντας γνωστά μοτίβα επιθέσεων, τα οποία έχουν καταγραφεί σε βάσεις δεδομένων υπογραφών. Οι υπογραφές (signatures) είναι προκαθορισμένοι κανόνες που περιγράφουν συγκεκριμένες μορφές κακόβουλης δραστηριότητες [47]. Μπορούν να περιλαμβάνουν χαρακτηριστικές ακολουθίες bytes, σειρές εντολών ή αιτημάτων, καθώς και αποκλίσεις σε επίπεδο πρωτοκόλλου (όπως ασυνήθιστα TCP flags ή αυξημένο πλήθος αιτημάτων σε μικρό χρονικό διάστημα). Το IDS συγκρίνει την εισερχόμενη κυκλοφορία του δικτύου με αυτές τις αποθηκευμένες υπογραφές, ανιχνεύοντας άμεσα γνωστές απειλές.

Τα Signature-Based IDS είναι ιδιαίτερα αποδοτικά στον εντοπισμό γνωστών επιθέσεων, προσφέροντας υψηλή ακρίβεια, χαμηλό ποσοστό ψευδών θετικών (false positives) και ταχύτητα, καθώς η αναγνώριση βασίζεται σε απλή αντιστοίχιση. Ωστόσο, δεν μπορούν να εντοπίσουν νέες ή άγνωστες επιθέσεις (zero-day attacks) που δεν περιλαμβάνονται σε βάσεις δεδομένων υπογραφών, ενώ η αποτελεσματικότητά τους εξαρτάται από το πόσο συχνά και έγκαιρα ενημερώνεται η βάση. Αυτό μπορεί να αυξήσει το υπολογιστικό κόστος και να καθυστερήσει την απόκριση σε νέες απειλές [46].



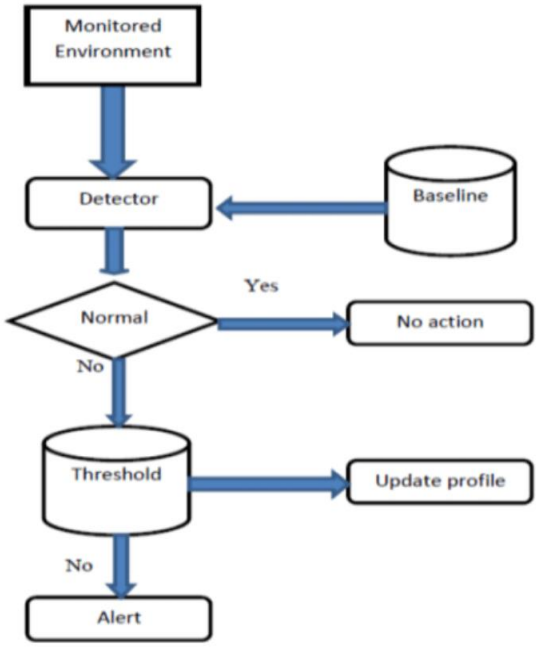
Σχήμα 5.1: Αρχιτεκτονική Signature-Based συστήματος [50]

5.2.2 Anomaly Based IDS

Τα Anomaly-Based IDS εντοπίζουν απειλές μέσω της ανίχνευσης αποκλίσεων από τη φυσιολογική συμπεριφορά του συστήματος. Σε αντίθεση με τα Signature-Based συστήματα, δεν βασίζονται σε γνωστά μοτίβα επιθέσεων, αλλά συγκρίνουν τη δραστηριότητα του δικτύου ή των συσκευών με ένα μοντέλο φυσιολογικής συμπεριφοράς (baseline), το οποίο δημιουργούν κατά την εκπαίδευση. Το baseline αυτό διαμορφώνεται βάσει παραμέτρων όπως ο ρυθμός μετάδοσης δεδομένων, ο αριθμός αιτημάτων, ή η χρήση συγκεκριμένων πρωτοκόλλων και θυρών [48].

Το baseline δημιουργείται με τη βοήθεια αλγορίθμων Μηχανικής Μάθησης, οι οποίοι αναλύουν ιστορικά δεδομένα δικτυακής κίνησης για να προσδιορίσουν ποια συμπεριφορά θεωρείται ως φυσιολογική και ποια ενδέχεται να είναι κακόβουλη. Κατά τη λειτουργία του συστήματος, η δικτυακή κίνηση συγκρίνεται με το baseline και αν εντοπιστεί απόκλιση που υπερβαίνει ένα προκαθορισμένο κατώφλι (threshold), η κίνηση καταγράφεται ως πιθανή εισβολή. Επειδή η ανίχνευση δεν βασίζεται σε προκαθορισμένες υπογραφές αλλά σε αποκλίσεις από τη φυσιολογική κίνηση, τα Anomaly-Based IDS μπορούν να εντοπίζουν νέες ή άγνωστες επιθέσεις, όπως οι zero-day.

Παρά την ικανότητά τους να εντοπίζουν νέες απειλές, εμφανίζουν συχνά υψηλά ποσοστά false positives, καθώς πολλές αποκλίσεις από τη φυσιολογική λειτουργία δεν είναι απαραίτητα κακόβουλες. Το πρόβλημα εντείνεται όταν το threshold δεν έχει ρυθμιστεί κατάλληλα ή όταν το σύστημα λειτουργεί σε πολύπλοκα και συνεχώς μεταβαλλόμενα περιβάλλοντα. Επιπλέον, η ανάλυση σε πραγματικό χρόνο απαιτεί σημαντικούς υπολογιστικούς πόρους σε σχέση με άλλες τεχνικές [47].



Σχήμα 5.2: Αρχιτεκτονική Anomaly-Based συστήματος [46]

5.2.3 Hybrid IDS

Τα Hybrid IDS συνδυάζουν χαρακτηριστικά των Signature-Based και Anomaly-Based συστημάτων, με στόχο την ανίχνευση γνωστών όσο και άγνωστων απειλών. Συγκεκριμένα, η ανίχνευση γνωστών επιθέσεων γίνεται μέσω της σύγκρισης της δικτυακής δραστηριότητας με βάσεις δεδομένων υπογραφών, ενώ οι άγνωστες απειλές εντοπίζονται μέσω της ανάλυσης αποκλίσεων από τη φυσιολογική συμπεριφορά του συστήματος. Ο συνδυασμός αυτών των δύο προσεγγίσεων επιτρέπει στα Hybrid IDS να καλύπτουν ευρύτερο φάσμα επιθέσεων σε σχέση με την κάθε τεχνική μεμονωμένη [50].

Ωστόσο, η ταυτόχρονη υλοποίηση και λειτουργία δύο διαφορετικών μηχανισμών ανίχνευσης αυξάνει σημαντικά τις απαιτήσεις σε υπολογιστικούς πόρους, όπως επεξεργαστική ισχύ, μνήμη και κατανάλωση ενέργειας. Αυτό καθιστά τα Hybrid IDS πιο απαιτητικά, ιδίως όταν εφαρμόζονται σε περιβάλλοντα περιορισμένων πόρων, όπως τα δίκτυα IoT.

5.3 Προσεγγίσεις Μάθησης για Εντοπισμό Εισβολών

Οι τεχνικές τεχνητής νοημοσύνης που εφαρμόζονται σε συστήματα εντοπισμού εισβολών βασίζονται κυρίως σε δύο προσεγγίσεις μάθησης: την επιβλεπόμενη (supervised learning) και τη μη επιβλεπόμενη (unsupervised learning) [3][41]. Η επιλογή μεταξύ των δύο προσεγγίσεων επηρεάζει σημαντικά τη φύση των αλγορίθμων, τον τρόπο εκπαίδευσής τους, καθώς και το είδος των δεδομένων που απαιτούνται.

5.3.1 Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση βασίζεται στην εκπαίδευση μοντέλων με επισημασμένα δεδομένα (labeled data), δηλαδή δεδομένα που περιλαμβάνουν πληροφορίες για το αν η συμπεριφορά είναι φυσιολογική ή κακόβουλη. Ο αλγόριθμος μαθαίνει να συσχετίζει τις εισόδους με τις αντίστοιχες εξόδους, ώστε να μπορεί να ταξινομεί με ακρίβεια νέα, άγνωστα δεδομένα. Στην ανίχνευση εισβολών, τα μοντέλα εκπαιδεύονται με επισημασμένα παραδείγματα δικτυακής δραστηριότητας και στη συνέχεια ταξινομούν νέα δεδομένα ως κανονικά ή κακόβουλα [47].

Το σύνολο των επισημασμένων δεδομένων χωρίζεται σε σύνολο εκπαίδευσης (training set) και σύνολο δοκιμής (test set), ώστε η απόδοση του μοντέλου να αξιολογείται σε δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση. Η εκπαίδευση πραγματοποιείται αποκλειστικά στο training set, ενώ η αξιολόγηση στο test set. Ο διαχωρισμός αυτός συμβάλλει στην αποφυγή της υπερπροσαρμογής στα δεδομένα εκπαίδευσης (overfitting) και εξασφαλίζει ότι το μοντέλο μπορεί να γενικεύει σωστά σε νέα δεδομένα.

5.3.2 Μη Επιβλεπόμενη Μάθηση

Τα μοντέλα μη επιβλεπόμενης μάθησης εκπαιδεύονται με μη επισημασμένα δεδομένα (unlabeled data), δηλαδή χωρίς προκαθορισμένες ετικέτες ή κατηγορίες. Οι αλγόριθμοι αναλύουν τα χαρακτηριστικά των δεδομένων και προσπαθούν να εντοπίσουν μοτίβα, ομοιότητες και αποκλίσεις. Οι κατηγορίες προκύπτουν δυναμικά, κατά τη διάρκεια της εκπαίδευσης.

Η μη επιβλεπόμενη μάθηση χρησιμοποιείται όταν δεν υπάρχουν διαθέσιμα επισημασμένα δεδομένα. Έχει τη δυνατότητα να ανιχνεύει νέα και μη καταγεγραμμένα μοτίβα, που μπορεί να σχετίζονται με νέες ή μη καταγεγραμμένες επιθέσεις [51].

5.4 Μετρικές Αξιολόγησης

Η αξιολόγηση της απόδοσης ενός συστήματος ανίχνευσης εισβολών είναι εξαιρετικά σημαντική για την ανάπτυξη και την εφαρμογή του. Βασίζεται σε μετρικές που προκύπτουν από τη σύγκριση των προβλέψεων του μοντέλου με τις πραγματικές κατηγορίες των δεδομένων [1] [3].

- **Συνολική Ακρίβεια (Accuracy).** Ποσοστό των σωστών προβλέψεων ως προς το σύνολο των δειγμάτων. Χρησιμοποιείται ευρέως και δείχνει πόσα δείγματα ταξινομήθηκαν στη σωστή κατηγορία.
- **Ανάκληση (Recall).** Ποσοστό των επιθέσεων που εντοπίστηκαν σωστά από το σύστημα. Αποτελεί σημαντική μετρική για την ανίχνευση εισβολών, καθώς δείχνει την ικανότητα του μοντέλου να αναγνωρίζει τις πραγματικές απειλές.
- **Ακρίβεια θετικών προβλέψεων (Precision).** Ποσοστό των περιπτώσεων που το σύστημα χαρακτήρισε ως επιθέσεις και ήταν όντως επιθέσεις. Υψηλή ακρίβεια θετικών σημαίνει ότι το μοντέλο δεν παράγει πολλά false positives.
- **F1-score.** Συνδυάζει την ακρίβεια θετικών προβλέψεων και την ανάκληση σε μία μόνο τιμή, εκφράζοντας την ισορροπία μεταξύ της ικανότητας εντοπισμού απειλών και της ακρίβειας των θετικών προβλέψεων.
- **Ποσοστό ψευδώς θετικών (False Alarm Rate – FAR).** Ποσοστό των δειγμάτων φυσιολογικής κίνησης που αναγνωρίστηκαν εσφαλμένα ως επιθέσεις. Η μείωσή του είναι σημαντική για την αποφυγή false positives.

5.5 Επίλογος

Σε αυτό το κεφάλαιο εξετάστηκαν οι κύριες κατηγορίες συστημάτων εντοπισμού εισβολών, καθώς και οι σύγχρονες προσεγγίσεις μάθησης. Η επιλογή κατάλληλης τεχνικής εντοπισμού εξαρτάται σε μεγάλο βαθμό από τις ιδιαιτερότητες του εκάστοτε IoT οικοσυστήματος και τα χαρακτηριστικά της δικτυακής κυκλοφορίας. Τέλος, αναλύθηκαν οι μετρικές αξιολόγησης που προσφέρουν ένα πρακτικό πλαίσιο για τη σύγκριση των λύσεων.

Κεφάλαιο 6ο: Τεχνικές Μηχανικής Μάθησης

6.1 Εισαγωγή

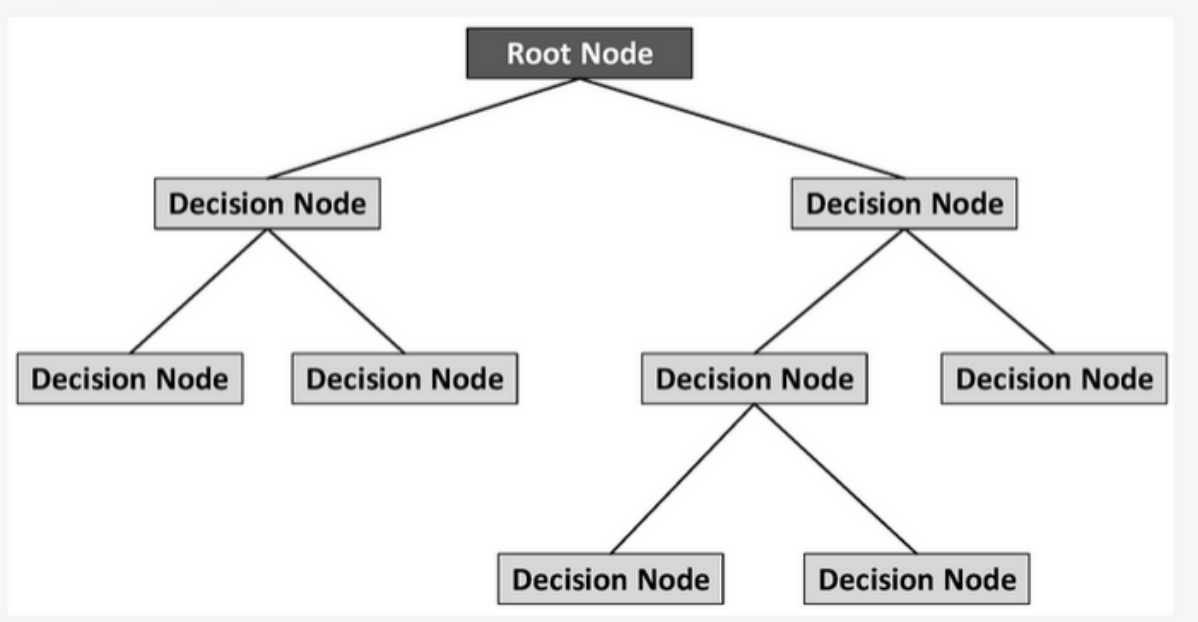
Οι τεχνικές μηχανικής μάθησης περιλαμβάνουν διάφορους αλγορίθμους που χρησιμοποιούνται για την ταξινόμηση της κίνησης σε κακόβουλη ή φυσιολογική. Κάθε αλγόριθμος έχει διαφορετικά χαρακτηριστικά ως προς την πολυπλοκότητα, την ακρίβεια, το υπολογιστικό κόστος και την ικανότητα γενίκευσης. Σε αυτό το κεφάλαιο θα παρουσιαστούν επιλεγμένες τεχνικές Μηχανικής Μάθησης που εφαρμόζονται στον εντοπισμό εισβολών σε περιβάλλοντα IoT. Για κάθε τεχνική αναλύεται ο τρόπος εκπαίδευσης και λειτουργίας της και παρουσιάζεται η εφαρμογή της μέσα από ενδεικτική επιστημονική έρευνα. Τέλος, συνοψίζονται τα βασικά πλεονεκτήματα και οι περιορισμοί κάθε μεθόδου, με στόχο την αξιολόγηση της καταλληλότητάς τους για χρήση σε διαφορετικού τύπου IoT συστήματα.

6.2 Decision Trees

Τα Δέντρα Απόφασης (Decision Trees – DTs) είναι ένας αλγόριθμος ταξινόμησης που βασίζεται σε επιβλεπόμενη μάθηση και χρησιμοποιείται για την ταξινόμηση ενός δείγματος με βάση τα χαρακτηριστικά του [52]. Η λειτουργία τους στηρίζεται στη σταδιακή διαίρεση των δεδομένων, επιλέγοντας κάθε φορά το χαρακτηριστικό που ξεχωρίζει καλύτερα τις διαφορετικές κατηγορίες. Η αναπαράσταση του μοντέλου έχει τη μορφή δενδροειδούς δομής, επιτρέποντας την εύκολη ερμηνεία και ανάλυση των αποφάσεων του συστήματος. Ο αλγόριθμος εκπαιδεύεται σε τέσσερα βήματα.

1. **Δημιουργία δέντρου απόφασης.** Η δημιουργία ενός δέντρου απόφασης ξεκινά από τον κόμβο ρίζας (root node), που περιλαμβάνει όλα τα δεδομένα εκπαίδευσης [53]. Κάθε δείγμα διαθέτει ένα σύνολο χαρακτηριστικών (π.χ. πρωτόκολλο, μέγεθος πακέτου) και μία ετικέτα που δηλώνει αν η κίνηση είναι φυσιολογική ή κακόβουλη. Στόχος του αλγορίθμου είναι να κατασκευάσει μία δομή που ταξινομεί νέα δείγματα με βάση αυτά τα χαρακτηριστικά.
2. **Επιλογή χαρακτηριστικού διαχωρισμού.** Ο αλγόριθμος αξιολογεί ποιο χαρακτηριστικό διαχωρίζει αποτελεσματικότερα τα δείγματα ανάλογα με τις κατηγορίες τους. Η επιλογή του χαρακτηριστικού γίνεται με βάση το κατά πόσο μπορεί να χωρίσει τα δείγματα σε ομάδες που περιέχουν δείγματα από την ίδια κατηγορία.
3. **Διαχωρισμός.** Με βάση το επιλεγμένο χαρακτηριστικό, τα δεδομένα χωρίζονται σε υποομάδες ανάλογα με τις τιμές τους. Κάθε υποομάδα οδηγεί στη δημιουργία ενός νέου κόμβου. Με αυτόν τον τρόπο, τα αρχικά δεδομένα διανέμονται στα επόμενα επίπεδα του δέντρου.
4. **Επανάληψη διαδικασίας.** Για κάθε νέο κόμβο, η ίδια διαδικασία επαναλαμβάνεται: επιλέγεται νέο χαρακτηριστικό, γίνεται διαχωρισμός και δημιουργούνται νέοι κόμβοι. Η ανάπτυξη συνεχίζεται είτε μέχρι τα δείγματα σε ένα κόμβο ανήκουν στην ίδια κατηγορία, είτε μέχρι να εξαντληθούν τα διαθέσιμα χαρακτηριστικά, είτε μέχρι να φτάσει στο προκαθορισμένο βάθος. Σε κάθε περίπτωση, ο κόμβος μετατρέπεται σε φύλλο (leaf node), που αντιστοιχεί σε κατηγορία εξόδου.

Το μοντέλο που προκύπτει μπορεί να χρησιμοποιηθεί για την ταξινόμηση νέων δεδομένων. Κάθε νέο δείγμα ξεκινά από τη ρίζα και κατευθύνεται μέσα στο δέντρο, ακολουθώντας τους κόμβους ανάλογα με τις τιμές των χαρακτηριστικών του μέχρι να φτάσει σε φύλλο που ορίζει την τελική του κατηγορία [52].



Σχήμα 6.1: Σχεδιάγραμμα δέντρου απόφασης για πρόβλημα ταξινόμησης [52]

Σε πρόσφατη μελέτη, προτάθηκε ένα σύστημα ανίχνευσης εισβολών βασισμένο σε Δέντρα Απόφασης, το οποίο συνδυάστηκε με τεχνικές βελτίωσης της ποιότητας των δεδομένων μέσω επεξεργασίας χαρακτηριστικών (feature engineering) και καθαρισμού (data cleaning) [54]. Η προσέγγιση αυτή πέτυχε ακρίβεια (accuracy) 99,8% στη συνολική ταξινόμηση των δειγμάτων. Επιπλέον, παρουσίασε ποσοστό εντοπισμού επιθέσεων (detection rate) 99,6%, γεγονός που αναδεικνύει την αποτελεσματικότητά τους στην αναγνώριση κακόβουλης δραστηριότητας. Η μελέτη επισημαίνει ότι τα Δέντρα Απόφασης έχουν χαμηλό υπολογιστικό κόστος και τονίζει ότι η απόδοσή τους επηρεάζεται σημαντικά από την επιλογή κατάλληλων χαρακτηριστικών εισόδου, γεγονός που αναδεικνύει τη σημασία της προεπεξεργασίας δεδομένων σε περιβάλλοντα αυξημένης πολυπλοκότητας όπως το IoT.

Τα Δέντρα Απόφασης είναι ένας απλός αλγόριθμος επιβλεπόμενης μάθησης. Ένα βασικό τους πλεονέκτημα είναι το ότι απαιτούν χαμηλή υπολογιστική ισχύ και είναι κατάλληλα για χρήση σε συσκευές ή δίκτυα με περιορισμένους πόρους. Η ερμηνεία των αποφάσεών τους είναι εύκολη, κάτι που είναι ιδιαίτερα χρήσιμο σε συστήματα όπου απαιτείται κατανόηση του τρόπου λειτουργίας του μοντέλου [52] [53].

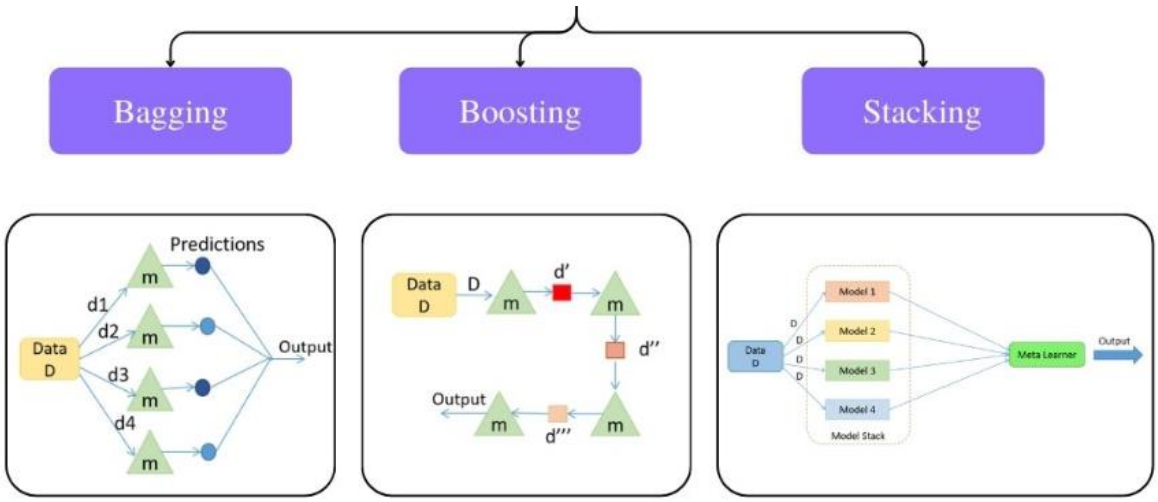
Ωστόσο, παρουσιάζουν ορισμένα μειονεκτήματα. Είναι ευαίσθητα στο overfitting, ειδικά όταν δεν εφαρμόζονται τεχνικές προεπεξεργασίας δεδομένων, όπως η επιλογή χαρακτηριστικών. Σε τέτοιες περιπτώσεις, το μοντέλο μπορεί να αποδίδει καλά στα δεδομένα εκπαίδευσης, αλλά όχι σε νέα, άγνωστα δείγματα. Είναι επίσης ευαίσθητα στον θόρυβο, καθώς μικρές μεταβολές στις τιμές των χαρακτηριστικών μπορεί να οδηγήσουν σε διαφορετικές διαδρομές ταξινόμησης, μεταβάλλοντας τη δομή του δέντρου [52].

Γενικά, τα Δέντρα Απόφασης προσφέρουν έναν ελαφρύ μηχανισμό ταξινόμησης για εφαρμογές ανίχνευσης εισβολών στο IoT, με ικανοποιητικά αποτελέσματα, όταν χρησιμοποιούνται μετά από κατάλληλη προεπεξεργασία των δεδομένων.

6.3 Ensemble Learning

Το Ensemble Learning είναι ο συνδυασμός πολλαπλών μοντέλων, με σκοπό τη βελτίωση της ακρίβειας και της αξιοπιστίας της τελικής πρόβλεψης. Κανένας αλγόριθμος ταξινόμησης που βασίζεται σε μηχανική μάθηση δεν αποτελεί ενιαία λύση για όλα τα προβλήματα [60]. Αντί να βασίζεται σε ένα μόνο ταξινομητή, το Ensemble Learning αξιοποιεί ένα σύνολο από απλά μοντέλα, των οποίων οι προβλέψεις συνδυάζονται με κατάλληλο τρόπο για την παραγωγή του τελικού αποτελέσματος.

Οι πιο συνηθισμένες τεχνικές ensemble είναι το Bagging (όπου κάθε μοντέλο εκπαιδεύεται σε διαφορετικό υποσύνολο των δεδομένων), το Boosting (όπου κάθε νέο μοντέλο επικεντρώνεται στα λάθη των προηγούμενων) και το Stacking (όπου οι προβλέψεις διαφόρων μοντέλων χρησιμοποιούνται ως είσοδοι για ένα τελικό μετα-μοντέλο). Τα ensemble συστήματα έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά σε προβλήματα ανίχνευσης εισβολών, καθώς προσφέρουν αυξημένη γενίκευση και μειωμένη ευαισθησία σε θόρυβο ή μεμονωμένα σφάλματα.



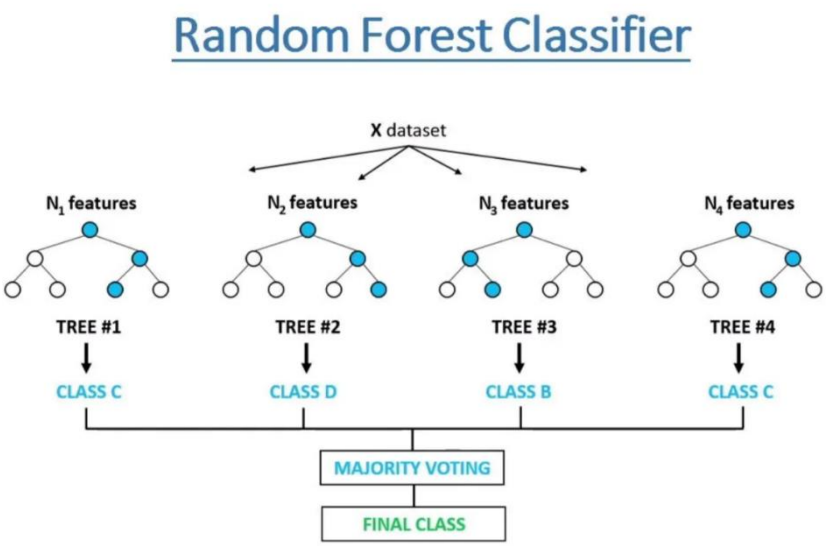
Σχήμα 6.2: Τεχνικές Ensemble Learning [57]

6.3.1 Προσέγγιση Bagging

Η προσέγγιση Bagging (Bootstrap Aggregating) έχει ως βασική ιδέα την ανεξάρτητη εκπαίδευση πολλών "αδύναμων" μοντέλων (weak learners) σε διαφορετικά υποσύνολα των δεδομένων εκπαίδευσης. Η τελική απόφαση προκύπτει με πλειοψηφία (majority voting). Η πιο χαρακτηριστική υλοποίηση του Bagging είναι ο αλγόριθμος RF, που συνδυάζει προβλέψεις πολλών δέντρων απόφασης για την ενίσχυση της συνολικής απόδοσης του συστήματος [59].

6.3.1.1 Random Forest

Ο Random Forest (RF) είναι ένας αλγόριθμος επιβλεπόμενης μηχανικής μάθησης και η πιο διαδεδομένη εφαρμογή της τεχνικής Bagging. Εκπαιδεύει πολλά Δέντρα Απόφασης σε τυχαία υποσύνολα των δεδομένων και συνδυάζει τις προβλέψεις τους μέσω ψηφοφορίας. Η προσέγγιση αυτή κάνει το μοντέλο πιο ανθεκτικό στα σφάλματα και λιγότερο επιρρεπές στο overfitting [3].



Σχήμα 6.3: Απεικόνιση της λειτουργίας του RF [56]

Η εκπαίδευσή του ολοκληρώνεται σε τέσσερα βασικά βήματα.

1. **Επιλογή δεδομένων από το σύνολο.** Αρχικά, από το σύνολο των διαθέσιμων επισημασμένων δεδομένων, δημιουργούνται νέα υποσύνολα μέσω τυχαίας δειγματοληψίας με επανατοποθέτηση (bootstrap sampling). Δηλαδή, κάθε δείγμα μπορεί να εμφανιστεί περισσότερες από μία φορές στο ίδιο υποσύνολο ή να μην συμπεριληφθεί καθόλου. Κάθε υποσύνολο χρησιμοποιείται για την εκπαίδευση ενός ξεχωριστού Δέντρου Απόφασης.
2. **Δημιουργία του πρώτου δέντρου.** Κάθε δέντρο εκπαιδεύεται με διαδοχικές διακλαδώσεις που διαχωρίζουν τα δεδομένα σε όλο και μικρότερες ομάδες. Σε διακλάδωση, επιλέγεται τυχαία ένα υποσύνολο χαρακτηριστικών. Από αυτό, επιλέγεται το χαρακτηριστικό που επιτυγχάνει τον καλύτερο διαχωρισμό των δειγμάτων. Τα δείγματα χωρίζονται σε δύο ομάδες, με βάση μία τιμή - όριο (threshold).
3. **Επανάληψη διαδικασίας.** Η διαδικασία συνεχίζεται μέχρι όλα τα δείγματα σε έναν κόμβο ανήκουν στην ίδια κατηγορία ή έως ότου ικανοποιηθεί κάποιο κριτήριο τερματισμού, όπως ο μικρός αριθμός δειγμάτων ή το μέγιστο επιτρεπόμενο βάθος του δέντρου.
4. **Συνδυασμός δέντρων για τελική πρόβλεψη.** Μετά την ολοκλήρωση της εκπαίδευσης, το μοντέλο αποτελείται από πολλά δέντρα απόφασης που λειτουργούν ανεξάρτητα κατά την πρόβλεψη. Κάθε νέο δείγμα περνά από όλα τα δέντρα και το καθένα καταλήγει σε μία απόφαση σχετικά με την κατηγορία στην οποία ανήκει. Στη συνέχεια εφαρμόζεται ψηφοφορία πλειοψηφίας των προβλέψεων των δέντρων και το δείγμα ταξινομείται ανάλογα με το αποτέλεσμα. Η διαδικασία επαναλαμβάνεται ανεξάρτητα για κάθε δείγμα [56].

Μια πρόσφατη μελέτη εξετάζει την εφαρμογή του RF για την ανίχνευση εισβολών σε δίκτυα IoT [58]. Οι ερευνητές εφάρμοσαν τεχνικές επιλογής χαρακτηριστικών και ρύθμισης υπερπαραμέτρων όπως ο αριθμός των δέντρων και το μέγιστο βάθος τους, με στόχο τη βελτίωση της ακρίβειας και της γενίκευσης του μοντέλου. Το βελτιστοποιημένο μοντέλο RF πέτυχε ακρίβεια 99,39%. Ωστόσο, αποδείχτηκε πως η απόδοση του μοντέλου εξαρτάται σε μεγάλο βαθμό από την ποιότητα των χαρακτηριστικών που χρησιμοποιούνται, γεγονός που καθιστά κρίσιμη την προεπεξεργασία των δεδομένων. Επιπλέον, η αξιολόγηση βασίστηκε αποκλειστικά σε ελεγχόμενα πειράματα με συγκεκριμένα σύνολα δεδομένων, αποδεικνύοντας ότι απαιτείται περαιτέρω έρευνα σε πραγματικά περιβάλλοντα IoT, όπου οι συνθήκες μεταβάλλονται και οι επιθέσεις εξελίσσονται συνεχώς.

Ο αλγόριθμος RF προσφέρει υψηλή ακρίβεια ταξινόμησης, ιδιαίτερα όταν προηγείται σωστή επιλογή χαρακτηριστικών. Είναι ανθεκτικός σε θορυβώδη δεδομένα και μπορεί να διαχειριστεί δεδομένα με

μεγάλο αριθμό μεταβλητών, όπως αυτά που συναντώνται συχνά στα δίκτυα IoT και περιλαμβάνουν τιμές όπως αριθμός πακέτων, διάρκεια σύνδεσης και πρωτόκολλο [60]. Τέλος, λόγω της παράλληλης εκπαίδευσης των δέντρων απόφασης, ο συνολικός χρόνος εκπαίδευσης του μοντέλου μειώνεται [59].

Παρά τα πλεονεκτήματά του, ο RF παρουσιάζει ορισμένους περιορισμούς. Η εκπαίδευση και η διαδικασία πρόβλεψης απαιτούν σημαντικούς υπολογιστικούς πόρους, ειδικά σε περιπτώσεις μεγάλου όγκου δεδομένων, όπως συμβαίνει συχνά στα δίκτυα IoT. Επιπλέον, χρειάζεται κατάλληλη επιλογή χαρακτηριστικών ώστε το μοντέλο να αποφύγει το overfitting [58]. Τέλος, λόγω της πολυπλοκότητας και του αριθμού δέντρων, η ερμηνεία των αποφάσεών του είναι περιορισμένη.

Ο RF είναι μια αξιόπιστη επιλογή για ανίχνευση εισβολών, αφού συνδυάζει σταθερότητα, ακρίβεια και ανθεκτικότητα. Είναι κατάλληλος για τα πολύπλοκα και δυναμικά περιβάλλοντα IoT, όπου απαιτείται αποτελεσματική διαχείριση μεγάλου όγκου δεδομένων με πολλαπλά χαρακτηριστικά.

6.3.2 Προσέγγιση Boosting

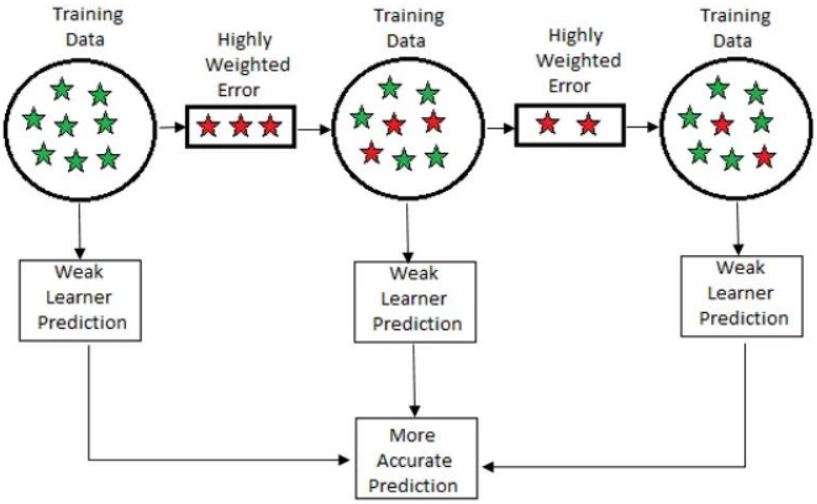
Το Boosting αποτελεί τεχνική ensemble learning που βασίζεται στη διαδοχική εκπαίδευση πολλών weak learners, με κάθε επόμενο μοντέλο να εστιάζει σε σφάλματα του προηγούμενου. Κατά τη διάρκεια της εκπαίδευσης, δίνεται μεγαλύτερη βαρύτητα στα δείγματα που ταξινομήθηκαν εσφαλμένα, ενισχύοντας την ικανότητα του συστήματος να χειρίζεται δύσκολες ή ακραίες περιπτώσεις. Το αποτέλεσμα είναι ένα συνδυασμένο μοντέλο με αυξημένη ακρίβεια, το οποίο μπορεί να γενικεύει καλύτερα σε δεδομένα υψηλής πολυπλοκότητας.

6.3.2.1 AdaBoost

Το AdaBoost (Adaptive Boosting) είναι μία από τις πρώτες και πιο διαδεδομένες μεθόδους Boosting. Βασίζεται στο συνδυασμό αδύναμων μοντέλων, συνήθως Δέντρων Απόφασης. Η εκπαίδευση πραγματοποιείται διαδοχικά, με κάθε μοντέλο να δίνει μεγαλύτερη έμφαση στα δείγματα που ταξινομήθηκαν λανθασμένα από τα προηγούμενα [61].

Αρχικά, κάθε δείγμα εκπαίδευσης συνοδεύεται από ένα βάρος (weight), το οποίο είναι ίσο για όλα τα δείγματα. Σε κάθε βήμα, ένα απλό δέντρο απόφασης εκπαιδεύεται και αξιολογείται η απόδοσή του. Τα δείγματα που ταξινομήθηκαν λανθασμένα αποκτούν μεγαλύτερο βάρος, ενώ εκείνα που ταξινομήθηκαν σωστά μειωμένο. Τα βάρη κανονικοποιούνται ώστε το συνολικό τους άθροισμα να είναι ίσο με 1.

Κάθε νέο δέντρο απόφασης εστιάζει στη διόρθωση των σφαλμάτων των προηγούμενων, δηλαδή στα δείγματα που προκάλεσαν σφάλμα. Η διαδικασία επαναλαμβάνεται για προκαθορισμένο αριθμό επαναλήψεων. Η τελική πρόβλεψη προκύπτει από σταθμισμένο άθροισμα των προβλέψεων όλων των μοντέλων (weighted voting), όπου το κάθε δέντρο συνεισφέρει ανάλογα με την ακρίβεια που πέτυχε κατά την εκπαίδευσή του.



Σχήμα 6.4: Εκπαίδευση μοντέλου AdaBoost [64]

Μία πρόσφατη μελέτη παρουσιάζει το Ada-IDS, ένα σύστημα ανίχνευσης εισβολών σε συστήματα IoT βασισμένο αποκλειστικά στον αλγόριθμο AdaBoost [63]. Εστιάζει στην ανίχνευση επιθέσεων που εκμεταλλεύονται τις ευπάθειες του πρωτοκόλλου ICMPv6. Το μοντέλο εκπαιδεύτηκε σε επισημασμένα δεδομένα που προήλθαν από προσομοιωμένο περιβάλλον. Η έρευνα έδειξε ότι το AdaBoost είναι ικανό να διαχειρίζεται με ακρίβεια εξειδικευμένες επιθέσεις σε περιβάλλοντα περιορισμένων πόρων, επιτυγχάνοντας ακρίβεια 99,6% και μηδενικό ποσοστό ψευδών συναγερμών. Ωστόσο, η αξιολόγηση περιορίστηκε σε ελεγχόμενες συνθήκες και συγκεκριμένο τύπο επιθέσεων, υποδεικνύοντας ότι είναι απαραίτητη περαιτέρω μελέτη σχετικά με την απόδοση του μοντέλου σε ετερογενή και συνεχώς μεταβαλλόμενα περιβάλλοντα IoT.

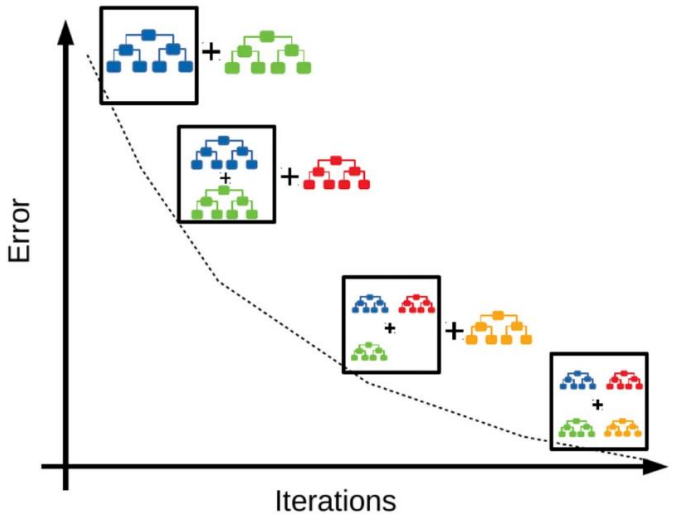
Το AdaBoost παρουσιάζει πλεονεκτήματα που το καθιστούν κατάλληλη επιλογή για εντοπισμό εισβολών σε δίκτυα IoT. Η εκπαίδευσή του είναι απλή και δεν απαιτεί περίπλοκη παραμετροποίηση. Καθώς κάθε νέο μοντέλο επικεντρώνεται στα δείγματα που ταξινομήθηκαν λανθασμένα από τα προηγούμενα, το σύστημα ενισχύει σταδιακά την ικανότητά του να ξεχωρίζει δύσκολες περιπτώσεις, επιτυγχάνοντας υψηλή ακρίβεια και καλύτερη γενίκευση [61].

Ωστόσο, η απόδοσή του μπορεί να μειωθεί όταν τα δεδομένα περιέχουν θόρυβο, καθώς το μοντέλο τείνει να δίνει υπερβολική έμφαση σε αυτά. Επίσης, το AdaBoost δεν περιλαμβάνει διαδικασία επιλογής χαρακτηριστικών, δηλαδή δεν εντοπίζει αυτόματα ποια χαρακτηριστικά συμβάλλουν ουσιαστικά στην εκπαίδευση. Έτσι, αν δεν έχει προηγηθεί προεπεξεργασία, το μοντέλο μπορεί να δώσει βάρος σε πληροφορίες που δεν σχετίζονται με τη διάκριση φυσιολογικής και κακόβουλης δραστηριότητας, με αποτέλεσμα να μειώνεται η ακρίβεια της πρόβλεψης. Τέλος, αν και τα επιμέρους μοντέλα είναι ασθενή και ελαφριά, ο συνολικός χρόνος εκπαίδευσης είναι αυξημένος, καθώς η εκπαίδευσή τους γίνεται διαδοχικά [61].

Συνοψίζοντας, το AdaBoost αποτελεί μια απλή αλλά ισχυρή τεχνική ταξινόμησης, η οποία, παρά τους περιορισμούς της, μπορεί να προσφέρει υψηλή απόδοση σε εφαρμογές ανίχνευσης εισβολών. Ωστόσο, παρά την αξιοπιστία του, δεν χρησιμοποιείται συχνά ως αυτόνομη τεχνική σε σύγχρονες εφαρμογές, καθώς έχουν προταθεί νεότερες μέθοδοι που βασίζονται στην ίδια λογική, αλλά επιτυγχάνουν μεγαλύτερη ευελιξία και καλύτερη απόδοση σε πιο απαιτητικά και μεταβαλλόμενα περιβάλλοντα, όπως τα οικοσυστήματα IoT.

6.3.2.2 Gradient Boosting

Το Gradient Boosting έχει αναδειχθεί ως αποτελεσματική μέθοδος Boosting για προβλήματα ταξινόμησης [66]. Αντί να δίνει μεγαλύτερο βάρος σε δείγματα που ταξινομήθηκαν λάθος, εκπαιδεύει διαδοχικά μοντέλα για να διορθώνουν τα σφάλματα των προηγούμενων, με στόχο την ελαχιστοποίηση της συνάρτησης απώλειας (loss function). Η συνάρτηση απώλειας υπολογίζει πόσο αποκλίνουν οι προβλέψεις του μοντέλου από τις πραγματικές τιμές. Για τη μείωσή της, χρησιμοποιείται ο αλγόριθμος Gradient Descent. Κάθε επόμενο μοντέλο εκπαιδεύεται πάνω στα αποτελέσματα της προηγούμενης πρόβλεψης. Το μοντέλο βελτιώνεται σταδιακά, αθροίζοντας το αποτέλεσμα των προβλέψεων και λαμβάνοντας υπόψη ένα συντελεστή μάθησης (learning rate). Με αυτό τον τρόπο, κάθε νέο μοντέλο εκπαιδεύεται για να διορθώνει τα λάθη των προηγούμενων βάσει αριθμητικά υπολογισμένων αποκλίσεων [65].



Σχήμα 6.5: Σταδιακή μείωση του σφάλματος στο Gradient Boosting [67]

Το Gradient Boosting έχει αποδειχθεί αποτελεσματικό στην ανίχνευση εισβολών σε περιβάλλοντα IoT, καθώς μπορεί να εντοπίζει λεπτές αποκλίσεις από τη φυσιολογική δικτυακή δραστηριότητα. Εστιάζει στις πιο δύσκολες περιπτώσεις, δηλαδή σε δείγματα που αποκλίνουν ελαφρώς από το φυσιολογικό αλλά ενδέχεται να υποδηλώνουν επίθεση [66]. Μια πρόσφατη μελέτη εξετάζει την απόδοση του Gradient Boosting στην ανίχνευση επιθέσεων σε περιβάλλον IoT [68]. Ο αλγόριθμος ανταποκρίνεται αποτελεσματικά σε επιθέσεις που εμφανίζουν έντονες αποκλίσεις από τη φυσιολογική κυκλοφορία, όπως οι επιθέσεις DoS και οι επιθέσεις port scanning, με απόδοση που φτάνει το 94%. Αντίθετα, η απόδοσή του υποχωρεί αισθητά σε περιπτώσεις όπου η κακόβουλη δραστηριότητα μοιάζει έντονα με τη συνηθισμένη, με την απόδοση να φτάνει το 60% σε ορισμένες περιπτώσεις. Η μελέτη επισημαίνει ότι, για να επιτευχθεί καλή απόδοση, κρίσιμο ρόλο παίζει η σωστή προεπεξεργασία των δεδομένων, με διαδικασίες όπως η κανονικοποίηση και η εξισορρόπηση του συνόλου εκπαίδευσης. Επιπλέον, λόγω του υπολογιστικού του κόστους και της μειωμένης αποτελεσματικότητας σε λιγότερο διακριτές επιθέσεις, τονίζεται πως η χρήση του σε IoT περιβάλλοντα δεν είναι πάντα πρακτική και απαιτεί αξιολόγηση με βάση τις ανάγκες του εκάστοτε συστήματος.

Το Gradient Boosting πετυχαίνει γενικά υψηλή ακρίβεια και έχει αποδειχθεί αποτελεσματικό σε προβλήματα ταξινόμησης. Επιπλέον, είναι ανθεκτικό σε δεδομένα με θόρυβο, καθώς η εκπαίδευση βασίζεται σε διαδοχικές διορθώσεις που περιορίζουν την επίδραση μεμονωμένων σφαλμάτων [65].

Παρόλα αυτά, η διαδικασία εκπαίδευσης είναι υπολογιστικά απαιτητική και πιο αργή συγκριτικά με άλλους αλγορίθμους [68]. Επίσης, η απόδοσή του εξαρτάται από τη σωστή προεπεξεργασία των δεδομένων αλλά και από την επιλογή των κατάλληλων παραμέτρων, όπως το learning rate και ο αριθμός μοντέλων που θα εκπαιδευτούν. Η λανθασμένη επιλογή αυτών των παραμέτρων μπορεί να οδηγήσει σε overfitting, καθώς ελέγχουν πόσο διορθώνεται η πρόβλεψη σε κάθε στάδιο. Τέλος, η ερμηνεία των αποφάσεών του είναι περιορισμένη, κάτι που δυσκολεύει την κατανόηση της εσωτερικής λειτουργίας του [65].

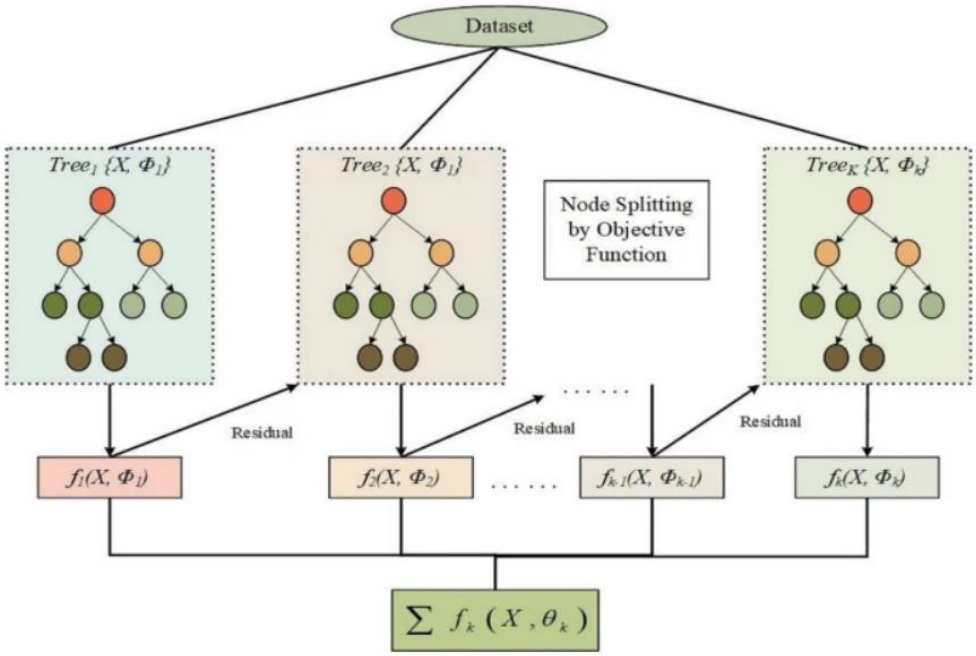
Παρά το σχετικά αυξημένο υπολογιστικό του κόστος, το Gradient Boosting παραμένει ισχυρή επιλογή για εφαρμογές εντοπισμού εισβολών, προσφέροντας καλή γενίκευση και σταθερή ακρίβεια σε καλά ορισμένα προβλήματα. Ωστόσο, στην πράξη χρησιμοποιούνται βελτιωμένες υλοποιήσεις του, που προσφέρουν καλύτερη απόδοση, υψηλότερη ταχύτητα και περισσότερη ευελιξία.

6.3.2.3 XGBoost

Το XGBoost (Extreme Gradient Boosting) αποτελεί μία βελτιστοποιημένη εκδοχή του Gradient Boosting, σχεδιασμένη να επιτυγχάνει μεγαλύτερη ακρίβεια με μειωμένη πολυπλοκότητα. Βασίζεται στη σταδιακή βελτίωση μέσω της διαδοχικής εκπαίδευσης μικρών δέντρων απόφασης, με τη διαφορά ότι κατά την εκπαίδευσή του εφαρμόζονται τεχνικές που εξασφαλίζουν ότι η βελτίωση γίνεται ελεγχόμενα, περιορίζοντας απότομες αλλαγές. Συγκεκριμένα, εφαρμόζεται κανονικοποίηση (regularization), που έχει ως στόχο την αποτροπή του overfitting του μοντέλου και συρρίκνωση (shrinkage), που επιτυγχάνεται μέσω της εφαρμογής του learning rate και ελέγχει την επίδραση κάθε δέντρου απόφασης στην τελική πρόβλεψη. Η εκπαίδευση ολοκληρώνεται σε 7 βήματα [69] [70].

1. **Αρχικοποίηση με μια ενιαία πρόβλεψη.** Η διαδικασία εκπαίδευσης ξεκινά με τον ορισμό μία σταθερής τιμής για τα δείγματα, που προκύπτει από το ποσοστό των δειγμάτων που έχουν χαρακτηριστεί ως “κακόβουλα” στο σύνολο εκπαίδευσης. Ο αλγόριθμος θεωρεί αρχικά ότι όλα τα δείγματα έχουν ίση πιθανότητα να ανήκουν σε αυτή την κατηγορία.
2. **Υπολογισμός του σφάλματος πρόβλεψης.** Εκπαιδεύεται το πρώτο δέντρο και για κάθε δείγμα, υπολογίζεται το αρχικό σφάλμα (residual). Είναι η αριθμητική διαφορά μεταξύ της αρχικής πρόβλεψης και της πραγματικής ετικέτας για το κάθε δείγμα. Οι τιμές θα αποτελέσουν τους στόχους για την εκπαίδευση του επόμενου μοντέλου.
3. **Εκπαίδευση δέντρου απόφασης για την πρόβλεψη των residuals.** Ο αλγόριθμος εκπαιδεύει ένα δέντρο απόφασης. Το νέο δέντρο δεν εκπαιδεύεται για να προβλέψει την τελική κατηγορία κάθε δείγματος, αλλά για να παράγει μια αριθμητική τιμή που προσεγγίζει όσο το δυνατόν περισσότερο την τιμή του residual. Η εκπαίδευση του νέου δέντρου έχει στόχο την ελαχιστοποίηση της τιμής της συνάρτησης απώλειας. Στο XGBoost, η συνάρτηση απώλειας περιλαμβάνει επιπλέον ποινές (penalties) τύπου L1 και L2, που περιορίζουν την επίδραση ορισμένων παραμέτρων του μοντέλου, ώστε να αποφεύγεται το overfitting.
4. **Διόρθωση προβλέψεων.** Η έξοδος του δέντρου προστίθεται στην αρχική πρόβλεψη, μετά από πολλαπλασιασμό με το learning rate, το οποίο ελέγχει το μέγεθος της κάθε διόρθωσης, ώστε η εκπαίδευση να εξελίσσεται σταδιακά. Αν το learning rate είναι πολύ μεγάλο, υπάρχει κίνδυνος overfitting. Αντίθετα, αν είναι πολύ μικρό, το μοντέλο απαιτεί περισσότερα δέντρα για να επιτύχει ικανοποιητική ακρίβεια, αυξάνοντας την υπολογιστική πολυπλοκότητα και το χρόνο εκπαίδευσης.
5. **Υπολογισμός νέων residuals.** Οι νέες προβλέψεις συγκρίνονται με τις πραγματικές ετικέτες και υπολογίζονται τα νέα residuals.
6. **Εκπαίδευση νέου δέντρου απόφασης.** Με βάση τα νέα residuals, εκπαιδεύεται ένα δεύτερο δέντρο απόφασης. Η διαδικασία επαναλαμβάνεται για συγκεκριμένο αριθμό επαναλήψεων, με κάθε δέντρο να εκπαιδεύεται πάνω στα residuals του προηγούμενου σταδίου, ώστε να βελτιωθεί σταδιακά η συνολική εκτίμηση του μοντέλου.

7. **Συνδυασμός προβλέψεων.** Η τελική πρόβλεψη του XGBoost προκύπτει από το άθροισμα όλων των διορθώσεων των δέντρων.



Σχήμα 6.6: Αρχιτεκτονική του XGBoost [71]

Μία πρόσφατη μελέτη αξιολόγησε την απόδοση του XGBoost στην ανίχνευση εισβολών του IoT [72]. Η αξιολόγηση του μοντέλου πραγματοποιήθηκε τόσο σε διαχωρισμό μεταξύ φυσιολογικής και κακόβουλης δραστηριότητας, όσο και σε κατηγοριοποίηση των επιθέσεων ανά τύπο (π.χ. DoS, Probe) αναδεικνύοντας την ευελιξία του XGBoost. Πέτυχε ιδιαίτερα υψηλή ακρίβεια, φτάνοντας το 98.8%. Ωστόσο, η αξιολόγηση περιορίστηκε σε ένα μόνο σύνολο δεδομένων, γεγονός που περιορίζει τη δυνατότητα γενίκευσης των συμπερασμάτων. Επιπλέον, η μελέτη δεν εστιάζει στην πρακτική εφαρμογή του μοντέλου σε πραγματικά IoT δίκτυα ή στην υλοποίησή του σε συσκευές με περιορισμένους πόρους. Οι συγγραφείς υπογραμμίζουν την ανάγκη για αξιολόγηση σε ρεαλιστικές συνθήκες λειτουργίας.

Το XGBoost αποτελεί μία αποδοτική επιλογή για ανίχνευση εισβολών σε περιβάλλοντα IoT, αφού συνδυάζει υψηλή ακρίβεια πρόβλεψης με ενσωματωμένους μηχανισμούς που μπορούν να μειώσουν την πολυπλοκότητα και να αποτρέψουν το overfitting. Είναι ιδιαίτερα αποτελεσματικό σε μη συμμετρικά σύνολα δεδομένων, δηλαδή σε αυτά που παρουσιάζουν ανισορροπία μεταξύ φυσιολογικών και κακόβουλων δειγμάτων. Επίσης, μπορεί να διαχειριστεί ελλιπή δεδομένα, χωρίς να απαιτεί προεπεξεργασία. Επιπλέον, επιτρέπει την αξιολόγηση της σημασίας των χαρακτηριστικών εισόδου, βάσει του πόσο συμβάλλουν στη διάκριση μεταξύ κακόβουλης και φυσιολογικής κίνησης, διευκολύνοντας την κατανόηση των χρήσιμων χαρακτηριστικών σε περιβάλλοντα IoT [70].

Ένα από τα βασικά του μειονεκτήματα είναι η αυξημένη υπολογιστική απαίτηση. Η εκπαίδευση του μοντέλου απαιτεί σημαντικούς πόρους μνήμης και επεξεργασίας, κάτι που περιορίζει τη δυνατότητα εφαρμογής του σε IoT οικοσυστήματα περιορισμένων πόρων. Επιπλέον, η απόδοση και ο χρόνος εκπαίδευσης του XGBoost επηρεάζεται από την επιλογή των παραμέτρων, όπως το learning rate [70].

Παρά την αυξημένη υπολογιστική του απαίτηση, το XGBoost ενσωματώνει τεχνικές που ελέγχουν την πολυπλοκότητα και μειώνουν σημαντικά το overfitting, διατηρώντας υψηλή ακρίβεια. Χάρη στη

δυνατότητά του να εστιάζει σταδιακά στα πιο δύσκολα δείγματα, αποτελεί καλή επιλογή για δυναμικά περιβάλλοντα με σύνθετα δεδομένα.

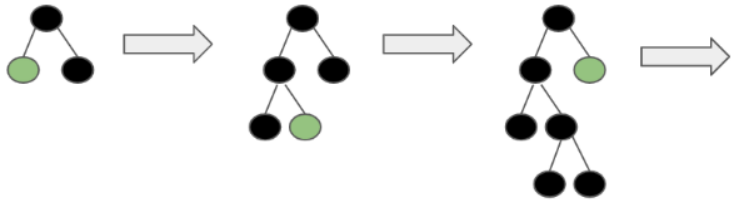
6.3.2.4 LightGBM

Ο LightGBM (Light Gradient Boosting Machine) είναι αλγόριθμος boosting που αναπτύχθηκε από τη Microsoft. Βασίζεται στη διαδοχική εκπαίδευση δέντρων απόφασης και έχει σχεδιαστεί για υψηλή απόδοση σε μεγάλα σύνολα δεδομένων, με γρηγορότερη και αποδοτικότερη εκπαίδευση [73]. Ακολουθεί παρόμοια λογική με το XGBoost, αλλά ενσωματώνει τεχνικές που επιταχύνουν τη διαδικασία εκπαίδευσης και μειώνουν σημαντικά την πολυπλοκότητα του μοντέλου [62].

Συγκεκριμένα, ο αλγόριθμος ενσωματώνει τεχνικές για μείωση της υπολογιστικής πολυπλοκότητας. Πρώτον, δεν επεξεργάζεται τα χαρακτηριστικά δεδομένων όπως είναι, αλλά τα χωρίζει σε διαστήματα τιμών (bins). Κάθε bin καλύπτει ένα εύρος συνεχόμενων τιμών του χαρακτηριστικού και ο μέγιστος αριθμός bins ανά χαρακτηριστικό είναι 255. Αφού οριστούν τα bins, η τιμή κάθε δείγματος αντικαθίσταται από έναν ακέραιο αριθμό που δηλώνει το bin στο οποίο ανήκει. Η χρήση των bins μειώνει τις απαιτήσεις σε μνήμη και υπολογιστική ισχύ [73]. Στη συνέχεια, για κάθε χαρακτηριστικό, κατασκευάζεται ένα ιστόγραμμα (histogram), δηλαδή ένας πίνακας όπου κάθε θέση αντιστοιχεί σε ένα bin. Για κάθε bin, υπολογίζεται η πρώτη παράγωγος (gradient) και η δεύτερη παράγωγος (hessian) της συνάρτησης απώλειας. Το gradient εκφράζει πόσο διαφέρει η τιμή πρόβλεψης από την πραγματική, ενώ το hessian περιγράφει το ρυθμό μεταβολής της συνάρτησης απώλειας. Η χρήση histograms κατά τη δημιουργία του δέντρου βοηθά τον αλγόριθμο να βρει άμεσα τα καλύτερα σημεία διαχωρισμού, αντί να εξετάζει όλες τις πιθανές τιμές. Αυτό μειώνει σημαντικά την υπολογιστική πολυπλοκότητα και τη χρήση της μνήμης, ειδικά αν πρόκειται για σύνολα δεδομένων με πολλά χαρακτηριστικά [62].

Για να μειώσει το χρόνο εκπαίδευσης του μοντέλου, το LightGBM χρησιμοποιεί την τεχνική Gradient-based One-Side Sampling (GOSS), για μείωση του αριθμού των δειγμάτων που χρησιμοποιούνται. Αντί να χρησιμοποιεί όλα τα δεδομένα για την εκπαίδευση κάθε δέντρου, το GOSS εστιάζει περισσότερο στα δείγματα με υψηλότερες τιμές gradient, δηλαδή αυτά για τα οποία το μοντέλο έχει μεγαλύτερο σφάλμα πρόβλεψης, καθώς παρέχουν πιο κρίσιμες πληροφορίες για τη βελτίωσή του. Από τα υπόλοιπα δείγματα, επιλέγεται τυχαία ένα υποσύνολο ώστε να διατηρηθεί η ποικιλία του συνόλου εκπαίδευσης [73].

Το LightGBM αναπτύσσει τα δέντρα του με στρατηγική leaf-wise, αντί για depth-wise όπως οι άλλοι αλγόριθμοι boosting. Υπολογίζει ποιο από τα υπάρχοντα φύλλα θα αποφέρει τη μεγαλύτερη μείωση της συνάρτησης απώλειας αν διαχωριστεί, όπως φαίνεται στο παρακάτω σχήμα. Με αυτό τον τρόπο, το δέντρο αναπτύσσεται ασύμμετρα, με έμφαση στα σημεία που μειώνουν πιο αποτελεσματικά το σφάλμα. Επιτυγχάνεται γρηγορότερη σύγκλιση και υψηλότερη ακρίβεια με μικρότερο αριθμό δέντρων. Για να αποφευχθεί η υπερβολική περιπλοκότητα και ο κίνδυνος overfitting, εφαρμόζονται περιορισμοί μέσω ρύθμισης υπερπαραμέτρων, όπως το μέγιστο βάθος του δέντρου και ο ελάχιστος αριθμός δειγμάτων ανά φύλλο [74].



Σχήμα 6.7: Leaf-wise ανάπτυξη δέντρου στο LightGBM. [75]

Το LightGBM υποστηρίζει και τη βελτιστοποίηση της χρήσης μνήμης. Συγκεκριμένα, με την τεχνική Exclusive Feature Bundling (EFB) μειώνει σημαντικά το πλήθος των χαρακτηριστικών, ομαδοποιώντας αυτά που δεν εμφανίζουν ταυτόχρονα μη μηδενικές τιμές στο ίδιο δείγμα. Με αυτό τον τρόπο, μειώνονται οι διαστάσεις των δεδομένων και βελτιώνεται η απόδοση της δημιουργίας δέντρων, ειδικά αν πρόκειται για σύνολα δεδομένων που συνδυάζουν μεγάλο αριθμό κατηγορικών και αριθμητικών χαρακτηριστικών [73] [74].

Σύμφωνα με μελέτη, το LightGBM μπορεί να χρησιμοποιηθεί για ανίχνευση εισβολών σε συστήματα IoT για να ταξινομεί με ακρίβεια τη δικτυακή δραστηριότητα σε φυσιολογική και κακόβουλη [76]. Στόχος της μελέτης ήταν η μείωση του υπολογιστικού κόστους με διατήρηση της υψηλής ακρίβειας, ώστε το σύστημα να είναι κατάλληλο για περιορισμένων πόρων περιβάλλοντα IoT. Το μοντέλο πέτυχε τελική ακρίβεια 99% , με πολύ χαμηλά ποσοστά false positives. Επιπλέον, έγιναν πειράματα σε πραγματικό IoT περιβάλλον βασισμένο σε Raspberry Pi και πέτυχε ακρίβεια 98,87% με μικρό μέσο χρόνο επεξεργασίας πακέτων, γεγονός που απέδειξε τη δυνατότητα ανίχνευσης εισβολών σχεδόν σε πραγματικό χρόνο. Το σύστημα πέτυχε πολύ καλή ανίχνευση επιθέσεων DoS, DDoS αλλά και δραστηριότητες botnet C&C. Αυτό αποδόθηκε στο ότι τέτοιες επιθέσεις δημιουργούν ξεκάθαρα, μαζικά μοτίβα στη δικτυακή κίνηση, όπως μεγαλύτερη διάρκεια σύνδεσης και υψηλή συχνότητα μικρών αιτημάτων. Ωστόσο, παρατηρήθηκαν περιορισμοί στην ανίχνευση κάποιων τύπων επιθέσεων (όπως Reconnaissance και Password Cracking), η συμπεριφορά των οποίων έμοιαζε περισσότερο με κανονική κίνηση. Τέλος, αν και η φάση της πρόβλεψης είναι γρήγορη και χαμηλού υπολογιστικού κόστους, η φάση της εκπαίδευσης παρέμεινε υπολογιστικά απαιτητική.

Ο αλγόριθμος LightGBM παρουσιάζει σημαντικά πλεονεκτήματα στη χρήση σε προβλήματα ανίχνευσης εισβολών σε περιβάλλοντα IoT. Επιτυγχάνει υψηλή ακρίβεια στην ανίχνευση κακόβουλης κίνησης, αφού μπορεί να εντοπίζει πολύπλοκες μη γραμμικές σχέσεις ανάμεσα στα χαρακτηριστικά των δεδομένων. Μπορεί να διαχειριστεί αποτελεσματικά διάφορα σύνολα δεδομένων, ακόμη και σε περιπτώσεις που περιλαμβάνουν μεγάλο αριθμό χαρακτηριστικών, ελλιπή δεδομένα ή ακραίες τιμές (outliers) [73]. Ένα ακόμη βασικό πλεονέκτημα είναι η ταχύτητα εκπαίδευσης που προσφέρει σε σύγκριση με άλλα boosting μοντέλα, χάρη στη χρήση ιστογραμμάτων για τη διαχείριση χαρακτηριστικών, την ανάπτυξη των δέντρων με leaf-wise στρατηγική και την εφαρμογή των τεχνικών GOSS και EFB [74]. Επιπλέον, το LightGBM απαιτεί λιγότερη προεπεξεργασία δεδομένων σε σχέση με άλλους αλγορίθμους, καθώς οι ενσωματωμένες τεχνικές του επιτρέπουν τη μείωση του υπολογιστικού κόστους χωρίς την ανάγκη επεξεργασίας χαρακτηριστικών [62].

Παρά τα πλεονεκτήματά του, το LightGBM παρουσιάζει και ορισμένα μειονεκτήματα. Αν και είναι γενικά πιο αποδοτικό από άλλες boosting μεθόδους, παραμένει υπολογιστικά ακριβό σε σύγκριση με πιο απλούς αλγορίθμους [75]. Επιπλέον, απαιτεί προσεκτική επιλογή υπερπαραμέτρων ώστε να διατηρηθεί η απόδοση και να αποφευχθεί το overfitting. Για παράδειγμα, η επιλογή πολύ μικρού leaf

size αυξάνει την πολυπλοκότητα, ενώ η επιλογή πολύ μικρού learning rate οδηγεί στη δημιουργία περισσότερων δέντρων και στην αύξηση του υπολογιστικού κόστους. Τέλος, η ερμηνεία των αποφάσεων και αποτελεσμάτων του είναι δύσκολη.

Συνολικά, το LightGBM αποτελεί μια ισχυρή επιλογή για προβλήματα ανίχνευσης εισβολών σε IoT περιβάλλοντα, συνδυάζοντας υψηλή ακρίβεια, γρήγορη εκπαίδευση και ικανότητα χειρισμού πολύπλοκων δεδομένων. Η αποτελεσματικότητά του στην ανίχνευση σύνθετων και δυναμικών απειλών, σε συνδυασμό με την προσαρμοστικότητά του σε ποικίλα σενάρια, το καθιστούν έναν από τους πιο κατάλληλους αλγορίθμους για εφαρμογές ασφάλειας σε σύγχρονα IoT οικοσυστήματα.

6.3.3 Προσέγγιση Stacking

Το Stacking (Stacked Generalization) αποτελεί μία προσέγγιση ensemble learning, στην οποία συνδυάζονται διαφορετικοί αλγόριθμοι (base learners) και τα αποτελέσματά τους χρησιμοποιούνται ως είσοδοι για την εκπαίδευση ενός τελικού μοντέλου (meta-learner) [60]. Η προσέγγιση αυτή επιτρέπει την αξιοποίηση των πλεονεκτημάτων κάθε επιμέρους μοντέλου, προσφέροντας μεγαλύτερη ευελιξία και υψηλότερη ακρίβεια ταξινόμησης. Ωστόσο, η εφαρμογή του stacking δεν είναι ευρέως διαδεδομένη στην ανίχνευση εισβολών σε περιβάλλοντα IoT και εμφανίζεται σπάνια στη βιβλιογραφία. Αυτό οφείλεται κυρίως στις αυξημένες υπολογιστικές απαιτήσεις που συνεπάγεται η εκπαίδευση και η αξιολόγηση πολλών μοντέλων [77].

6.4 Σύγκριση τεχνικών μηχανικής μάθησης

Μετά την ανάλυση των κυριότερων τεχνικών μηχανικής μάθησης που έχουν εφαρμοστεί στην ανίχνευση εισβολών σε περιβάλλοντα IoT, στον παρακάτω πίνακα παρουσιάζονται οι σημαντικότεροι παράμετροι αξιολόγησης, για άμεση σύγκριση μεταξύ των μοντέλων.

Πίνακας 6.1: Σύγκριση τεχνικών μηχανικής μάθησης

Μοντέλο	**Υπολογιστικό κόστος**	**Απαιτήσεις δεδομένων**	**Ερμηνευσιμότητα**	**Προσαρμοστικότητα**	**Πραγματική εφαρμογή**
Decision Trees	Χαμηλό	Απαιτεί προεπεξεργασία	Υψηλή	Χαμηλή	Ναι
Random Forest	Μέτριο	Εξαρτάται από την ποιότητα	Χαμηλή	Μέτρια	Σε προσομοίωση
AdaBoost	Μέτριο	Δεν χρειάζεται μεγάλη ποσότητα	Μέτρια	Μέτρια	Ναι
Gradient Boosting	Υψηλό	Απαιτεί προεπεξεργασία	Χαμηλή	Μέτρια	Ναι
XGBoost	Υψηλό	Απαιτεί προεπεξεργασία	Χαμηλή	Υψηλή	Σε προσομοίωση
LightGBM	Μέτριο	Μπορεί να διαχειριστεί ελλείψεις	Χαμηλή	Υψηλή	Ναι

6.5 Επίλογος

Το παρόν κεφάλαιο παρουσίασε επιλεγμένες τεχνικές Μηχανικής Μάθησης που εφαρμόζονται για τον εντοπισμό εισβολών σε περιβάλλοντα IoT. Για κάθε τεχνική αναλύθηκε η βασική λειτουργία της, η χρήση της στο πλαίσιο της ασφάλειας, καθώς και η επίδοσή της σύμφωνα με σχετικές επιστημονικές μελέτες. Μέσω της συγκριτικής παρουσίασης των πλεονεκτημάτων και των περιορισμών κάθε μεθόδου, αναδείχθηκαν τα κριτήρια επιλογής κατάλληλων αλγορίθμων με βάση τις απαιτήσεις, τα δεδομένα και τους πόρους κάθε IoT συστήματος.

Κεφάλαιο 7ο Τεχνικές Βαθιάς Μάθησης

7.1 Εισαγωγή

Το κεφάλαιο αυτό παρουσιάζει τεχνικές βαθιάς μάθησης που εφαρμόζονται στον εντοπισμό εισβολών σε περιβάλλοντα IoT. Για κάθε τεχνική περιγράφεται ο τρόπος λειτουργίας της και η εφαρμογή της στο IoT. Αναλύονται επίσης τα πλεονεκτήματα και μειονεκτήματά τους. Η επιλογή κάθε τεχνικής εξαρτάται από τις απαιτήσεις του εκάστοτε συστήματος, τα διαθέσιμα δεδομένα και τους υπολογιστικούς πόρους του περιβάλλοντος IoT.

7.2 Βαθιά Μάθηση

Η Βαθιά Μάθηση αποτελεί υποκατηγορία της Μηχανικής Μάθησης και βασίζεται στη χρήση τεχνικών βαθιών νευρωνικών δικτύων (deep neural networks). Τα δίκτυα αυτά περιλαμβάνουν πολλά διαδοχικά επίπεδα επεξεργασίας, μέσω των οποίων εξάγονται αναπαραστάσεις χαρακτηριστικών από τα δεδομένα εισόδου. Οι τεχνικές βαθιάς μάθησης έχουν αποδειχθεί ιδιαίτερα αποτελεσματικές στον εντοπισμό εισβολών σε περιβάλλοντα IoT, καθώς μπορούν να επεξεργάζονται μεγάλους όγκους δεδομένων, να εντοπίζουν μη γραμμικές συσχετίσεις, να προσαρμόζονται σε δυναμικά περιβάλλοντα και να υποστηρίζουν τη λήψη αποφάσεων ενίσχυσης της ασφάλειας. Παράλληλα, είναι κατάλληλες για αναγνώριση πολύπλοκων ή σπάνιων μορφών επιθέσεων που συχνά δεν ανιχνεύονται από παραδοσιακές προσεγγίσεις, ενώ μπορούν να εντοπίζουν και άγνωστες επιθέσεις, όπως zero-day [80].

Οι τεχνητοί νευρώνες αποτελούνται από **βάρη (weights)** και **μεροληπτικούς όρους (biases)**. Κάθε είσοδος του νευρώνα πολλαπλασιάζεται με ένα βάρος, το οποίο ρυθμίζεται κατά την εκπαίδευση ώστε να επηρεάζει την έξοδο του νευρώνα, αναλόγως της σημασίας της αντίστοιχης εισόδου. Ο όρος bias προστίθεται στη γραμμική συνάρτηση πριν την ενεργοποίηση και επιτρέπει στο μοντέλο να μετατοπίζει την έξοδο, βελτιώνοντας έτσι την ικανότητα μάθησης, ακόμη και όταν όλες οι είσοδοι είναι μηδενικές.

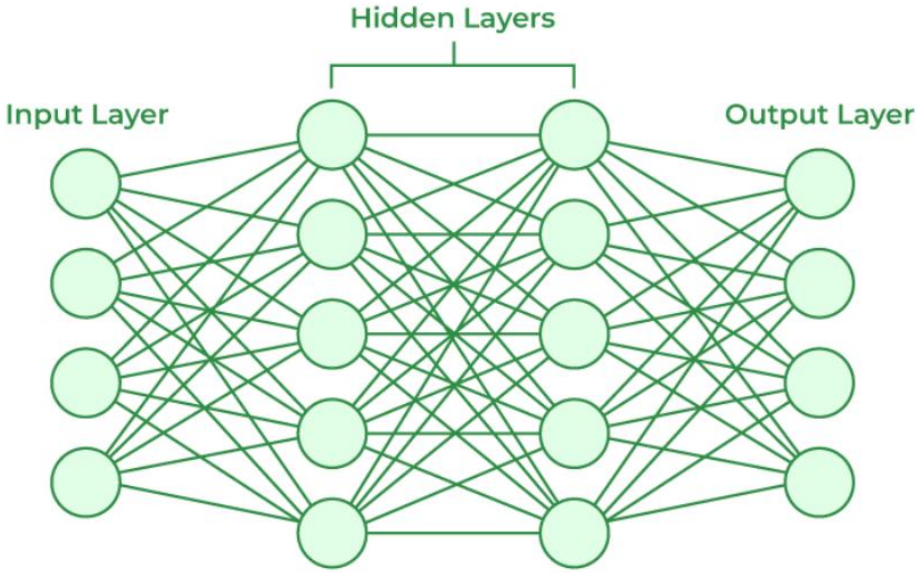
Για τον ορισμό της τελικής εξόδου, χρησιμοποιούνται συναρτήσεις ενεργοποίησης (activation functions), οι οποίες καθορίζουν τη συμπεριφορά του νευρώνα. Κάποια παραδείγματα συναρτήσεων που χρησιμοποιούνται στη βαθιά μάθηση είναι:

- **ReLU (Rectified Linear Unit.)** Είναι η πιο διαδεδομένη σε σύγχρονα δίκτυα βαθιάς μάθησης. Εισάγει μη γραμμικότητα και διατηρεί τις θετικές τιμές αμετάβλητες, ενώ μηδενίζει τις αρνητικές. Είναι απλή και υπολογιστικά αποδοτική.
- **Sigmoid.** Χρησιμοποιείται όταν απαιτείται πιθανολογική ερμηνεία της εξόδου, όπως σε προβλήματα δυαδικής ταξινόμησης. Επιστρέφει τιμές μεταξύ 0 και 1.
- **Tanh (υπερβολική εφαπτομένη).** Επιστρέφει τιμές στο διάστημα (-1, 1) και έχει μηδενικό μέσο όρο, γεγονός που συνήθως επιταχύνει τη σύγκλιση κατά την εκπαίδευση.

7.3 Artificial Neural Networks

Τα Artificial Neural Networks (ANNs) είναι μία από τις βασικές τεχνικές επιβλεπόμενης βαθιάς μάθησης και αποτελούνται από τεχνητούς νευρώνες που οργανώνονται σε διαδοχικά επίπεδα. Κάθε νευρώνας λαμβάνει σήματα από τους νευρώνες του προηγούμενου επιπέδου, τα επεξεργάζεται μέσω ενός μαθηματικού μετασχηματισμού και αποστέλλει το αποτέλεσμά στους επόμενους νευρώνες [78]. Τα ANNs είναι κατάλληλα για προβλήματα ταξινόμησης και εντοπισμού εισβολών, καθώς έχουν τη δυνατότητα να αναγνωρίζουν πολύπλοκες, μη γραμμικές σχέσεις μεταξύ των δεδομένων. Στην

ασφάλεια του IoT, χρησιμοποιούνται για την ανάλυση μεγάλου όγκου δεδομένων και την αναγνώριση προτύπων επιθέσεων με μεγαλύτερη ακρίβεια σε σχέση με τους παραδοσιακούς αλγορίθμους Μηχανικής Μάθησης [79].



Σχήμα 7.1: Αρχιτεκτονική μοντέλου ANN [78]

Όπως φαίνεται στο σχήμα 7.1, ένα ANN αποτελείται από ένα επίπεδο εισόδου (input layer), ένα ή περισσότερα κρυφά επίπεδα (hidden layers) και ένα επίπεδο εξόδου (output layer). Οι νευρώνες κάθε επιπέδου είναι πλήρως συνδεδεμένοι με τους νευρώνες του επόμενου. Η εκπαίδευση ενός ANN περιλαμβάνει οκτώ στάδια:

1. **Εισαγωγή δεδομένων.** Τα δεδομένα εκπαίδευσης εισάγονται στους νευρώνες του επιπέδου εισόδου. Κάθε χαρακτηριστικό των δεδομένων αντιστοιχεί σε ένα νευρώνα. Κάθε σύνδεση μεταξύ δύο νευρώνων συνοδεύεται από μία τιμή βάρους (weight), που καθορίζει πόσο επηρεάζει η έξοδος ενός νευρώνα την ενεργοποίηση του επόμενου. Αρχικά, οι τιμές των βαρών αρχικοποιούνται τυχαία.
2. **Υπολογισμός εισόδων.** Κάθε τιμή εισόδου πολλαπλασιάζεται με το αντίστοιχο βάρος της σύνδεσής της προς κάθε νευρώνα του πρώτου κρυφού επιπέδου. Για κάθε νευρώνα, υπολογίζεται το σταθμισμένο άθροισμα των εισερχόμενων τιμών, δηλαδή το άθροισμα των εισόδων του που έχουν πολλαπλασιαστεί με τα βάρη.
3. **Ενεργοποίηση νευρώνων.** Στο σταθμισμένο άθροισμα κάθε νευρώνα προστίθεται μία σταθερά (bias), η οποία προσφέρει ευελιξία στο μοντέλο επιτρέποντας την ενεργοποίηση ακόμη και όταν οι είσοδοι έχουν χαμηλή ή μηδενική τιμή. Το αποτέλεσμα περνά από μία συνάρτηση ενεργοποίησης (activation function), που καθορίζει αν και σε ποιο βαθμό θα ενεργοποιηθεί ο νευρώνας. Η ενεργοποίηση ενός νευρώνα σημαίνει ότι η έξοδός του επηρεάζει την επεξεργασία των επόμενων επιπέδων και συμμετέχει ενεργά στη διαδικασία μάθησης. Οι πιο συνηθισμένες συναρτήσεις ενεργοποίησης είναι η ReLU, η sigmoid και η tanh. Η χρήση συνάρτησης ενεργοποίησης επιτρέπει στο δίκτυο να αναγνωρίζει μη γραμμικές σχέσεις στα δεδομένα.
4. **Μεταφορά εξόδου (forward propagation).** Η έξοδος κάθε νευρώνα μεταφέρεται ως είσοδος στους νευρώνες του επόμενου κρυφού επιπέδου. Η διαδικασία αυτή επαναλαμβάνεται διαδοχικά σε όλα τα κρυφά επίπεδα του δικτύου, μέχρι να φτάσει στο επίπεδο εξόδου. Σε αυτό το στάδιο δεν γίνεται καμία αλλαγή στα weights ή τα biases του μοντέλου.
5. **Επίπεδο εξόδου.** Στο επίπεδο εξόδου χρησιμοποιείται συνήθως ένας νευρώνας, ο οποίος δέχεται εισόδους από όλους τους νευρώνες του τελευταίου κρυφού επιπέδου. Η έξοδος υπολογίζεται μέσω μιας συνάρτησης ενεργοποίησης sigmoid, η οποία παράγει τιμή μεταξύ 0

και 1. Η τιμή αυτή ερμηνεύεται ως πιθανότητα το δείγμα να ανήκει στην κατηγορία της κακόβουλης κίνησης. Η τελική ταξινόμηση γίνεται βάσει προκαθορισμένου κατωφλιού.

6. **Υπολογισμός σφάλματος.** Η τελική έξοδος του μοντέλου συγκρίνεται με την πραγματική ετικέτα του δείγματος ώστε να υπολογιστεί το σφάλμα (error) με τη χρήση συνάρτησης απώλειας (loss function). Συνήθως χρησιμοποιείται η binary cross-entropy, η οποία εκφράζει την απόκλιση της πρόβλεψης από την επιθυμητή τιμή. Όσο μικρότερη είναι η τιμή της συνάρτησης απώλειας, τόσο καλύτερη θεωρείται η απόδοση του μοντέλου για το συγκεκριμένο δείγμα.
7. **Backpropagation.** Αφού υπολογιστεί το σφάλμα στο επίπεδο εξόδου, το δίκτυο υπολογίζει πως τα βάρη του τελευταίου κρυφού επιπέδου συνέβαλαν στο συνολικό σφάλμα. Αυτή η διαδικασία ονομάζεται Backpropagation και ξεκινά από το επίπεδο εξόδου, προχωρώντας προς τα πίσω. Σε κάθε βήμα, υπολογίζονται οι παράγωγοι της συνάρτησης κόστους (gradients) ως προς τα βάρη, ώστε να εκτιμηθεί η επίδραση του κάθε βάρους στο συνολικό σφάλμα.
8. **Ενημέρωση βαρών.** Οι τιμές των gradients χρησιμοποιούνται για την προσαρμογή των βαρών με στόχο τη μείωση του σφάλματος στις επόμενες προβλέψεις, μέσω του αλγορίθμου Gradient Descent. Ο αλγόριθμος ενημερώνει τα βάρη προς την κατεύθυνση που μειώνει περισσότερο το σφάλμα, ανάλογα με το μέγεθος της παραγώγου, λαμβάνοντας υπόψη και τον συντελεστή μάθησης (learning rate). Η διαδικασία επαναλαμβάνεται για όλα τα δείγματα του συνόλου εκπαίδευσης και συνεχίζεται μέχρι το σφάλμα να φτάσει σε ένα αποδεκτό επίπεδο ή να σταθεροποιηθεί.

Σύμφωνα με σχετική μελέτη, τα ANNs είναι κατάλληλα για εντοπισμό εισβολών σε περιβάλλοντα IoT [81]. Το σύστημα εκπαιδεύεται offline με επισημασμένα δεδομένα του δικτύου και χρησιμοποιείται για ταξινόμηση και εντοπισμό εισβολών σε πραγματικό χρόνο. Το μοντέλο επιτυγχάνει ακρίβεια εντοπισμού έως 93% και διατηρεί χαμηλό ποσοστό false positives (3,3%). Αναφέρεται ότι η υπολογιστική επιβάρυνση είναι χαμηλή (0,8% σε μνήμη και 0,05% σε CPU). Ένας βασικός περιορισμός της προσέγγισης είναι ότι η απόδοση του συστήματος εξαρτάται άμεσα από την ποιότητα του συνόλου εκπαίδευσης, καθώς τα μη επισημασμένα δεδομένα μπορεί να αυξήσουν τα false positives. Τέλος, η μελέτη εστιάζει σε συγκεκριμένα χαρακτηριστικά εισόδου και δεν αξιολογεί την απόδοση σε πιο σύνθετες επιθέσεις. Αποδεικνύεται η δυνατότητα εφαρμογής των ANNs σε περιβάλλον IoT σε πραγματικό χρόνο, υπό την προϋπόθεση ότι τα δεδομένα είναι επισημασμένα και καλής ποιότητας.

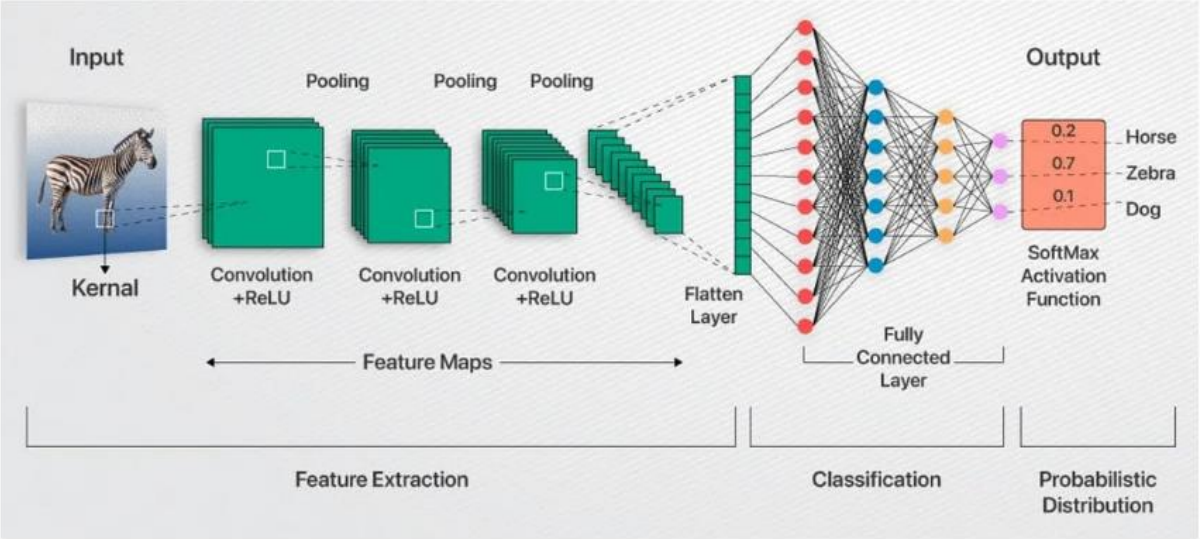
Τα ANNs επιτυγχάνουν υψηλή ακρίβεια στην ταξινόμηση γνωστών επιθέσεων [79]. Αναγνωρίζουν πολύπλοκες, μη γραμμικές σχέσεις ακόμη και σε πολυδιάστατα και ετερογενή δεδομένα, όπως αυτά που παράγονται σε οικοσυστήματα IoT. Πετυχαίνουν καλή γενίκευση και μπορούν να εντοπίζουν κρυφά μοτίβα που δεν είναι άμεσα εμφανή στα δεδομένα. Μπορούν επίσης να εντοπίζουν με αποτελεσματικότητα παραλλαγές γνωστών επιθέσεων, καθώς κοινά χαρακτηριστικά ακόμη και όταν υπάρχουν μικρές αποκλίσεις. Τέλος, υποστηρίζουν παράλληλη επεξεργασία, αυξάνοντας την αποδοτικότητά τους, εφόσον υπάρχουν επαρκείς υπολογιστικοί πόροι.

Ωστόσο, έχουν υψηλές απαιτήσεις σε υπολογιστικούς πόρους, τόσο κατά την εκπαίδευση όσο και κατά τη φάση πρόβλεψης, κάτι που αποτελεί εμπόδιο στην εφαρμογή τους σε περιβάλλοντα περιορισμένων πόρων, όπως το IoT. Απαιτούν τη χρήση μεγάλων και ποιοτικών συνόλων δεδομένων κατά την εκπαίδευσή τους, για να πετύχουν υψηλή απόδοση. Αν το σύνολο δεδομένων είναι μικρό ή μη ισορροπημένο, αυξάνεται ο κίνδυνος overfitting. Τέλος, τα ANNs θεωρούνται black box μοντέλα, δηλαδή η ερμηνεία των αποτελεσμάτων τους είναι δύσκολη.

Συμπερασματικά, τα ANNs παραμένουν μια από τις βασικές επιλογές για εντοπισμό εισβολών σε IoT συστήματα, προσφέροντας υψηλή απόδοση και ικανότητα χειρισμού σύνθετων, δυναμικών και ετερογενών δεδομένων.

7.4 Convolutional Neural Networks

Τα Convolutional Neural Networks (CNNs) αποτελούν μία κατηγορία επιβλεπόμενων τεχνικών βαθιάς μάθησης και εξειδικευμένο είδος ANNs, σχεδιασμένο για την επεξεργασία δεδομένων με χωρική ή χρονική δομή, όπως εικόνες ή σειρές δεδομένων [83]. Αν και αναπτύχθηκαν αρχικά για την ανάλυση εικόνας, έχουν εφαρμοστεί και σε προβλήματα εντοπισμού εισβολών, κυρίως όταν τα δεδομένα μπορούν να αναπαρασταθούν ως πίνακες.



Σχήμα 7.2: Διάγραμμα λειτουργίας των CNNs. [82]

Όπως φαίνεται στο παραπάνω σχήμα, η αρχιτεκτονική των CNNs περιλαμβάνει διαδοχικά επίπεδα που συνεργάζονται για την εξαγωγή και επεξεργασία σημαντικών χαρακτηριστικά από τα δεδομένα εισόδου, με σκοπό να τα ταξινομήσουν [84]. Περιλαμβάνουν συνελικτικά επίπεδα (convolutional layers) που εφαρμόζουν φίλτρα ή πυρήνες (kernels) για την αυτόματη ανίχνευση των σημαντικότερων χαρακτηριστικών των δεδομένων, για τη δημιουργία πινάκων χαρακτηριστικών (feature maps). Ενδιάμεσα, ενσωματώνονται επίπεδα συγκέντρωσης (pooling layers) που μειώνουν τις διαστάσεις των δεδομένων, διατηρώντας τις σημαντικότερες πληροφορίες. Τα χαρακτηριστικά που προκύπτουν από την επεξεργασία, προωθούνται σε ένα ή περισσότερα πλήρως συνδεδεμένα επίπεδα (fully connected layers), τα οποία παράγουν την τελική πρόβλεψη. Η εκπαίδευσή των CNNs γίνεται με backpropagation, όπως στα κλασσικά ANNs.

Η εφαρμογή των CNNs στον εντοπισμό εισβολών σε δίκτυα IoT είναι περιορισμένη στη βιβλιογραφία. Οι περισσότερες μελέτες βασίζονται σε offline σύνολα δεδομένων και πειραματικές υλοποιήσεις, χωρίς αξιολόγηση σε ροή δεδομένων ή σε πραγματικά IoT περιβάλλοντα. Παρ' όλα αυτά, τα CNNs χρησιμοποιούνται συχνά για την εξαγωγή χαρακτηριστικών από δεδομένα υψηλής διαστασιμότητας, λειτουργώντας ως στάδιο προεπεξεργασίας πριν από την τελική ταξινόμηση με διαφορετικό μοντέλο [84].

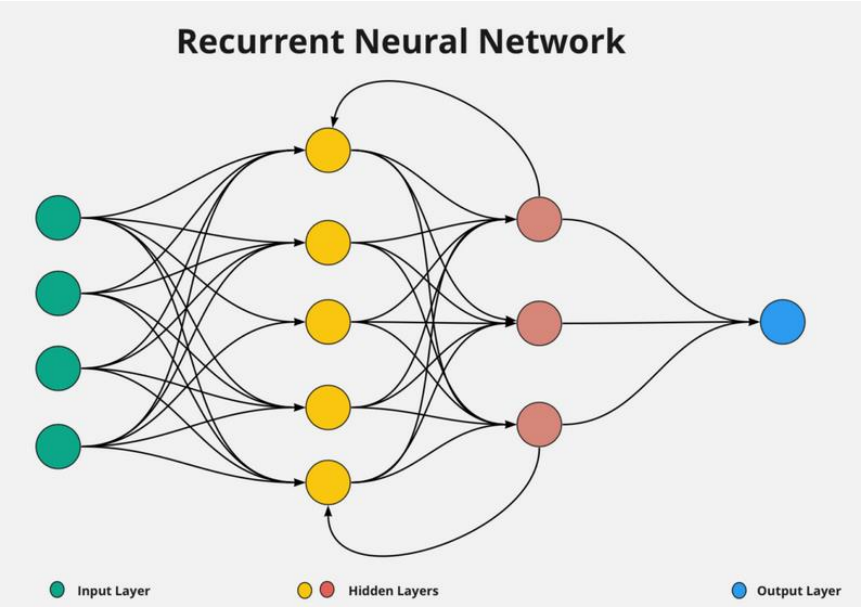
Μια χαρακτηριστική περίπτωση είναι η χρήση Temporal CNN (TCNN), που προτείνεται για ταξινόμηση πέντε τύπων δραστηριότητας, πετυχαίνοντας ακρίβεια έως 99,7% [85]. Ωστόσο, η αξιολόγηση γίνεται αποκλειστικά εντός συγκεκριμένου συνόλου δεδομένων και δεν εξετάζεται η ικανότητα γενίκευσης ή η χρήση σε πραγματικό σύστημα IoT. Σε άλλη μελέτη, έχουν εφαρμοστεί τα 1D-CNNs σε ακολουθίες χαρακτηριστικών από πακέτα δικτύου, με στόχο την ανίχνευση επιθέσεων [86]. Τα μοντέλα αυτά παρουσιάζουν μειωμένη υπολογιστική πολυπλοκότητα, αλλά η εκπαίδευσή τους γίνεται αποκλειστικά με επισημασμένα και προεπεξεργασμένα δεδομένα. Παραμένει ασαφές το κατά

πόσο μπορούν να ενσωματωθούν σε πραγματικά IoT περιβάλλοντα, όπου υπάρχουν καθυστερήσεις, αστάθεια και περιορισμένοι υπολογιστικοί πόροι.

Τα CNNs προσφέρουν σημαντικά πλεονεκτήματα, ως προς την εξαγωγή τοπικών χαρακτηριστικών και τη μείωση της διαστασιμότητας των δεδομένων. Πετυχαίνουν υψηλή ακρίβεια σε προβλήματα ταξινόμησης όταν τα δεδομένα έχουν χωρική ή χρονική δομή. Ωστόσο, η χρήση τους στον εντοπισμό εισβολών σε περιβάλλοντα IoT είναι περιορισμένη, λόγω της ανάγκης για προκαθορισμένη δισδιάστατη αναπαράσταση των δεδομένων, της απαίτησης για επισημασμένα σύνολα και της μειωμένης προσαρμοστικότητας σε δυναμικά περιβάλλοντα. Αν και παρουσιάζουν ερευνητικό ενδιαφέρον και αξιοποιούνται σε συνδυασμό με άλλες προσεγγίσεις, δεν αποτελούν κύρια επιλογή για την υλοποίηση IoT Intrusion Detection Systems.

7.5 Recurrent Neural Networks

Τα Recurrent Neural Networks (RNNs) αποτελούν προσέγγιση επιβλεπόμενης μάθησης που βασίζεται σε νευρωνικά δίκτυα. Είναι σχεδιασμένα για την επεξεργασία ακολουθιακών δεδομένων, όπου η χρονική σειρά εμφάνισης των τιμών έχει σημασία. Σε αντίθεση με τα κλασσικά νευρωνικά δίκτυα, τα RNNs ενσωματώνουν μηχανισμό μνήμης, που τους επιτρέπει να λαμβάνουν υπόψη τόσο την τρέχουσα είσοδο όσο και προηγούμενες καταστάσεις. Με αυτό τον τρόπο, μπορούν να εντοπίζουν χρονικές εξαρτήσεις ανάμεσα στα δεδομένα και να προσαρμόζουν την έξοδό τους με βάση ιστορικές πληροφορίες [87].



Σχήμα 7.3: Διάγραμμα λειτουργίας των RNNs [88]

Όπως φαίνεται στο παραπάνω σχήμα, η βασική ιδιαιτερότητα των RNNs είναι ότι, κατά την επεξεργασία κάθε εισόδου, διατηρούν εσωτερικές καταστάσεις μέσω αναδρομικών σχέσεων (feedback loops), λαμβάνοντας υπόψη πληροφορίες από προηγούμενα σημεία της ακολουθίας. Η εσωτερική κατάσταση του δικτύου ενημερώνεται διαδοχικά και μεταφέρεται σε επόμενα χρονικά βήματα. Η έξοδος κάθε νευρώνα σε ένα χρονικό βήμα επιστρέφεται και χρησιμοποιείται ως είσοδος του δικτύου στο επόμενο βήμα, δημιουργώντας έναν κυκλικό μηχανισμό μνήμης. Το RNN διαμορφώνει έτσι την τρέχουσα έξοδό του όχι μόνο βάσει του νέου δεδομένου, αλλά και βάσει της ιστορίας που έχει καταγραφεί έως εκείνη τη στιγμή. Η εκπαίδευση των RNNs γίνεται μέσω της μεθόδου Backpropagation

Through Time (BPTT), η οποία επεκτείνει τον κλασικό αλγόριθμο backpropagation ώστε να καλύπτει όλα τα χρονικά βήματα της ακολουθίας. Με αυτό τον τρόπο, οι παράμετροι του δικτύου προσαρμόζονται με βάση την επίδραση των σφαλμάτων που προκύπτουν τόσο από το παρόν, όσο και από προηγούμενα βήματα [87].

Η χρήση των RNNs στην ανίχνευση εισβολών σε περιβάλλοντα IoT είναι περιορισμένη. Υπάρχουν ωστόσο ορισμένες μελέτες που τα αξιοποιούν, με θετικά αποτελέσματα υπό συγκεκριμένες συνθήκες. Σε μία από αυτές, προτείνεται σύστημα βασισμένο σε RNN, το οποίο επιτυγχάνει ακρίβεια 92,13% στη συνολική ταξινόμηση επιθέσεων [88]. Ωστόσο, το σύστημα αξιολογείται μόνο σε offline δεδομένα και όχι σε πραγματικό χρόνο. Σε πιο πρόσφατη μελέτη, προτείνεται ένα σύστημα ανίχνευσης εισβολών βασισμένο σε RNNs, που επιτυγχάνει ποσοστά ανίχνευσης έως 98,2% σε συγκεκριμένο σύνολο δεδομένων [89]. Αν και τα αποτελέσματα είναι θετικά, η αξιολόγηση γίνεται σε πειραματικό στάδιο και δεν εξετάζεται η εφαρμογή του συστήματος σε ροή δεδομένων ή σε πραγματικές συνθήκες λειτουργίας.

Παρά τον περιορισμένο αριθμό εφαρμογών, τα RNNs εμφανίζουν πλεονεκτήματα. Η δυνατότητά τους να ανιχνεύουν χρονικά εξαρτημένα πρότυπα τα καθιστά κατάλληλα για την επεξεργασία ακολουθιακών δεδομένων που εμφανίζουν χρονική εξάρτηση. Είναι απλούστερα σε υλοποίηση σε σχέση με άλλες, πιο εξελιγμένες τεχνικές, γεγονός που επιτρέπει γρηγορότερη υλοποίηση [88].

Ωστόσο, η περιορισμένη αξιοποίησή τους στο IoT αποδίδεται στο ότι δεν είναι κατάλληλα για εφαρμογή σε συσκευές IoT που διαθέτουν περιορισμένους πόρους, όταν απαιτείται εντοπισμός εισβολών σε πραγματικό χρόνο. Ένα ακόμη μειονέκτημα είναι η τάση τους να εμφανίζουν προβλήματα σταθερότητας κατά την εκπαίδευση, όπως το vanishing και το exploding gradient. Αυτά σχετίζονται με τον τρόπο που μεταδίδονται τα gradients κατά τη διαδικασία BPTT. Στην περίπτωση του vanishing gradient, τα gradients μειώνονται εκθετικά όσο μεταδίδονται προς τα πίσω, οδηγώντας σε αδυναμία εκπαίδευσης των αρχικών επιπέδων του δικτύου. Στο exploding gradient, οι τιμές των gradients αυξάνονται υπερβολικά, προκαλώντας αριθμητική αστάθεια. Αυτά τα προβλήματα αντιμετωπίζονται πιο αποτελεσματικά από εξελιγμένες αρχιτεκτονικές που αναπτύσσονται παρακάτω.

7.6 Long Short-Term Memory

Τα Long Short-Term Memory (LSTM) είναι μια βελτιωμένη εκδοχή των RNNs, σχεδιασμένη με σκοπό να ξεπεράσει τα προβλήματα σταθερότητας που αντιμετωπίζουν τα RNNs, όπως το vanishing και το exploding gradient [90]. Η αρχιτεκτονική τους επιτρέπει καλύτερη ικανότητα εκμάθησης ακολουθιών από τα δεδομένα και αποθήκευση σημαντικών πληροφοριών για μεγαλύτερα χρονικά διαστήματα. Τα LSTM βασίζονται σε επιβλεπόμενη μάθηση και χρησιμοποιούνται ευρέως στην ασφάλεια του IoT, καθώς μπορούν να αναλύσουν τη δικτυακή κίνηση σε μεγαλύτερο βάθος χρόνου, ανιχνεύοντας αργά εξελισσόμενες ή σταδιακά μεταβαλλόμενες επιθέσεις [91].

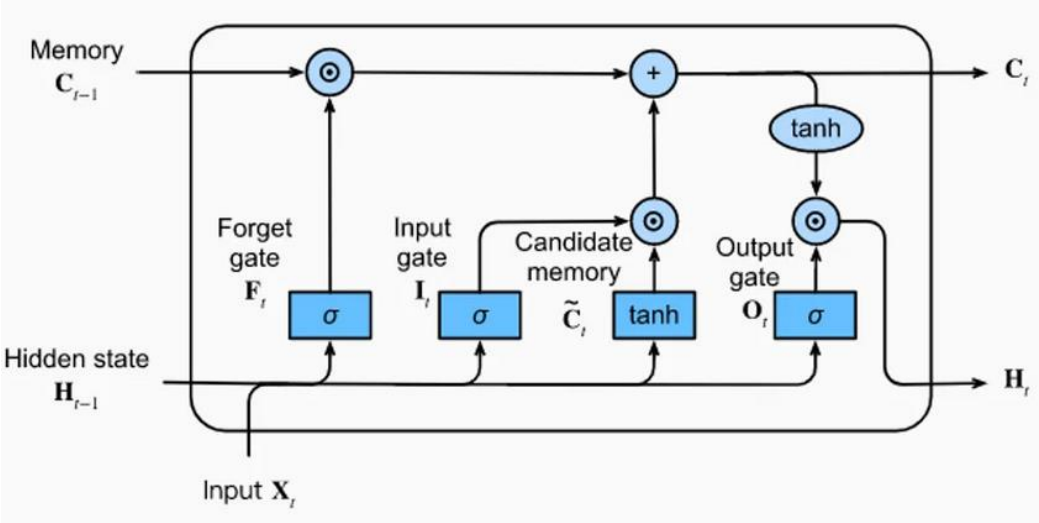
Τα LSTM διαθέτουν έναν εσωτερικό αποθηκευτικό μηχανισμό, το κύτταρο μνήμης (memory cell) και τρεις βασικές πυλες ελέγχου (gates): την Input Gate, την Forget Gate και την Output Gate [90]. Το κύτταρο μνήμης είναι ο βασικός αποθηκευτικός χώρος του δικτύου και διατηρεί πληροφορίες σε βάθος χρόνου. Οι πύλες ελέγχουν τη ροή της πληροφορίας προς και από το κύτταρο μνήμης, έτσι ώστε το δίκτυο να διατηρεί μόνο τις κρίσιμες πληροφορίες.

Η Forget Gate αποφασίζει ποια πληροφορία δεν είναι αρκετά σημαντική και πρέπει να διαγραφεί από την υπάρχουσα κατάσταση μνήμης. Η Input Gate επιλέγει ποια νέα πληροφορία θα προστεθεί στο κύτταρο. Η Output Gate καθορίζει ποια τμήματα της αποθηκευμένης πληροφορίας θα χρησιμοποιηθούν

ως έξοδος και θα μεταφερθούν στο επόμενο χρονικό βήμα. Η εκπαίδευσή ενός LSTM μοντέλου εκτελείται σε επτά βήματα [90] [92].

1. **Είσοδος δεδομένων.** Τα επισημασμένα δεδομένα εισάγονται στο μοντέλο σε μορφή διανύσματος, ως ακολουθίες στις οποίες κάθε στοιχείο αντιστοιχεί σε μια χρονική στιγμή. Στο IoT, οι είσοδοι είναι επεξεργασμένες ροές χαρακτηριστικών, που περιλαμβάνουν πληροφορίες όπως το πρωτόκολλο που χρησιμοποιείται, ο αριθμός των bytes που μεταδίδονται, διάρκεια σύνδεσης και άλλα χαρακτηριστικά.
2. **Υπολογισμός της Forget Gate.** Η Forget Gate καθορίζει ποιες πληροφορίες της τρέχουσας κατάστασης του κυττάρου μνήμης θα διαγραφούν. Σε κάθε χρονικό βήμα, λαμβάνει ως είσοδο το διάνυσμα εισόδου και την έξοδο του μοντέλου από το προηγούμενο χρονικό βήμα. Στο πρώτο χρονικό βήμα όπου δεν υπάρχει προηγούμενη έξοδος, το μοντέλο χρησιμοποιεί μηδενικό διάνυσμα μαζί με το διάνυσμα εισόδου. Το αποτέλεσμα της επεξεργασίας της Forget Gate είναι ένα διάνυσμα τιμών μεταξύ 0 και 1, με κάθε τιμή να αντιστοιχεί σε στοιχείο της κατάστασης μνήμης. Τιμή ίση με 1 σημαίνει ότι το αντίστοιχο στοιχείο θα διατηρηθεί πλήρως, ενώ τιμή ίση με 0 σημαίνει ότι θα διαγραφεί.
3. **Υπολογισμός της Input Gate.** Η Input Gate είναι υπεύθυνη για την εισαγωγή νέας πληροφορίας στο κύτταρο μνήμης. Χρησιμοποιεί ως είσοδο το διάνυσμα της τρέχουσας χρονικής στιγμής και την έξοδο του μοντέλου από το προηγούμενο βήμα, για να κάνει δύο υπολογισμούς: πρώτα εφαρμόζει τη συνάρτηση sigmoid και για κάθε θέση της μνήμης και προκύπτει ένα διάνυσμα μεταξύ 0 και 1, με κάθε τιμή να αντιστοιχεί σε μία θέση της κατάστασης μνήμης και να καθορίζει ποια θα ενημερωθεί. Στη συνέχεια, εφαρμόζει τη συνάρτηση tanh για να υπολογίσει τις νέες υποψήφιες τιμές που θα προστεθούν στο κύτταρο μνήμης, στις θέσεις που θα ενεργοποιηθούν. Με αυτό τον τρόπο, το μοντέλο δημιουργεί τις νέες πληροφορίες για αποθήκευση βάσει της εισόδου και ελέγχει αν και πού θα την ενσωματώσει.
4. **Ενημέρωση του κυττάρου μνήμης.** Με βάση τους υπολογισμούς που έγιναν από τη Forget Gate και την Input Gate, ενημερώνεται το περιεχόμενο του κυττάρου μνήμης. Το προηγούμενο περιεχόμενο της μνήμης πολλαπλασιάζεται με τις τιμές που προέκυψαν από τη Forget Gate, ώστε να διατηρηθούν μόνο οι πληροφορίες που κρίθηκαν σημαντικές. Στη συνέχεια προστίθενται οι νέες τιμές που υπολόγισε στη δεύτερη φάση η Input Gate, οι οποίες έχουν φιλτραριστεί από το αποτέλεσμα της πρώτης φάσης (sigmoid). Ως αποτέλεσμα, το κύτταρο μνήμης περιέχει τις πληροφορίες που διατηρήθηκαν από το παρελθόν αλλά και τις νέες πληροφορίες που κρίθηκαν σημαντικές για αποθήκευση.
5. **Υπολογισμός της Output Gate.** Η Output Gate καθορίζει τι θα παραχθεί ως έξοδος από το LSTM στην τρέχουσα χρονική στιγμή. Αρχικά εφαρμόζει τη συνάρτηση sigmoid στις εισόδους, ώστε να υπολογιστεί ποιο τμήμα του κυττάρου μνήμης θα επηρεάσει την έξοδο. Το περιεχόμενο του κυττάρου μνήμης περνά στη συνέχεια από συνάρτηση tanh, ώστε οι τιμές του να περιοριστούν στο διάστημα [-1, 1]. Η τελική έξοδος προκύπτει από τον πολλαπλασιασμό των δύο αποτελεσμάτων, ώστε να εξαχθεί μόνο το μέρος της μνήμης που έχει επιλεγεί από την Output Gate.
6. **Πρόβλεψη και έξοδος του μοντέλου.** Τα παραπάνω βήματα επαναλαμβάνονται για κάθε χρονικό σημείο της εισόδου. Ως αποτέλεσμα, το μοντέλο έχει δημιουργήσει μία εσωτερική αναπαράσταση της ακολουθίας, που διατηρεί ενσωματωμένη μόνο τη σημαντική πληροφορία των δεδομένων στο κύτταρο μνήμης. Η τελευταία έξοδος του LSTM προωθείται σε ένα πλήρως συνδεδεμένο επίπεδο (dense layer) και εισάγεται σε μία συνάρτηση ενεργοποίησης, συνήθως sigmoid. Το αποτέλεσμα είναι μια τιμή μεταξύ 0 και 1. Αν αυτή η πιθανότητα υπερβαίνει ένα προκαθορισμένο κατώφλι, η είσοδος ταξινομείται ως κακόβουλη.
7. **Εκπαίδευση μέσω BPTT.** Η τελική έξοδος του μοντέλου συγκρίνεται με την πραγματική ετικέτα του δείγματος μέσω συνάρτησης απώλειας, συνήθως binary cross-entropy. Προκύπτει το σφάλμα πρόβλεψης, το οποίο χρησιμοποιείται για την προσαρμογή των παραμέτρων του μοντέλου. Διαδίδεται προς τα πίσω μέσω του αλγορίθμου BPTT, που ενημερώνει τις παραμέτρους του LSTM σε κάθε χρονικό βήμα, ανάλογα με το πόσο συνέβαλαν στην τελική έξοδο. Υπολογίζονται τα gradients του σφάλματος ως προς τα βάρη και τα bias κάθε πύλης (Forget, Input, Output). Στη συνέχεια, οι παράγωγοι χρησιμοποιούνται για την ενημέρωση των

παραμέτρων μέσω κάποιου αλγορίθμου βελτιστοποίησης, όπως ο Adam. Με την επανάληψη αυτής της διαδικασίας σε πολλά παραδείγματα, το μοντέλο βελτιώνει σταδιακά την ικανότητά του να προβλέπει σωστά την κατηγορία μιας νέας ακολουθίας.



Σχήμα 7.4: Αρχιτεκτονική του LSTM [92]

Ένα από τα βασικά πλεονεκτήματα των LSTM σε σχέση με τα κλασικά RNNs είναι η ικανότητά τους να διατηρούν σταθερές τις τιμές των gradients κατά τη διαδικασία BPTT, περιορίζοντας τα προβλήματα vanishing και exploding gradients. Αυτό επιτυγχάνεται χάρη στην αρχιτεκτονική του κυττάρου μνήμης και των gates. Συγκεκριμένα, η Forget Gate ελέγχει αυστηρά την ποσότητα πληροφορίας που διατηρείται από προηγούμενες χρονικές στιγμές, ενώ η Input Gate περιορίζει την ανεξέλεγκτη προσθήκη νέων τιμών. Ο κίνδυνος αριθμητικής αστάθειας κατά την εκπαίδευση μειώνεται και η μάθηση γίνεται πιο σταθερή. Με αυτό τον τρόπο, ενισχύεται η ικανότητα εντοπισμού μακροπρόθεσμων εξαρτήσεων στη χρονική ακολουθία των δεδομένων [91].

Η αποτελεσματικότητα των LSTM στην ανίχνευση εισβολών σε περιβάλλοντα IoT επιβεβαιώνεται στη βιβλιογραφία. Ωστόσο, οι περισσότερες μελέτες προϋποθέτουν ότι προηγείται κατάλληλη προεπεξεργασία των δεδομένων. Σε μελέτη του 2021, εξετάστηκαν μοντέλα LSTM σε συνδυασμό με τεχνικές μείωσης διαστάσεων, συγκεκριμένα Principal Component Analysis (PCA) και Mutual Information, με στόχο τη βελτίωση της ακρίβειας και τη μείωση του χρόνου εκπαίδευσης [93]. Η χρήση της PCA απέδωσε τα καλύτερα αποτελέσματα, ακόμη και με δύο μόνο χαρακτηριστικά εισόδου. Συγκεκριμένα, πέτυχε ακρίβεια 99.49% στην ταξινόμηση κακόβουλης και φυσιολογικής κίνησης, ενώ ο χρόνος εκπαίδευσης μειώθηκε κατά 95% σε σύγκριση με τη χρήση του πλήρους συνόλου χαρακτηριστικών. Σε άλλη μελέτη, προτάθηκε μοντέλο βασισμένο αποκλειστικά σε LSTM, με εφαρμογή της τεχνικής Synthetic Minority Oversampling Technique (SMOTE) [94], για την εξισορρόπηση του συνόλου δεδομένων. Η προσέγγιση πέτυχε ακρίβεια 99.56%, αποδεικνύοντας ότι η ανίχνευση σπάνιων επιθέσεων είναι εφικτή. Και οι δύο μελέτες επιβεβαιώνουν ότι τα LSTM αποτελούν ισχυρό εργαλείο ανίχνευσης εισβολών σε δίκτυα IoT, αρκεί να έχει προηγηθεί προσεκτική επιλογή χαρακτηριστικών και αντιμετώπιση της ανισορροπίας των δεδομένων στο σύνολο εκπαίδευσης.

Τα LSTM παρουσιάζουν σημαντικά πλεονεκτήματα στην ανίχνευση επιθέσεων σε δίκτυα IoT, κυρίως λόγω της ικανότητάς τους να αναγνωρίζουν χρονικές εξαρτήσεις στα δεδομένα. Η δυνατότητα διατήρησης πληροφορίας από προηγούμενα χρονικά βήματα τα καθιστά κατάλληλα για τον εντοπισμό επιθέσεων που εξελίσσονται σταδιακά, όπως DoS, DDoS και scanning. Επιπλέον, σε αντίθεση με κλασικούς αλγορίθμους που απαιτούν χειροκίνητη εξαγωγή χαρακτηριστικών, τα LSTM μαθαίνουν

αυτόματα τα σημαντικότερα χαρακτηριστικά των δεδομένων. Η χρήση του κυττάρου μνήμης και των gates συμβάλλει στη σταθερότητα της εκπαίδευσης, περιορίζοντας τα vanishing και exploding gradients [92]. Τέλος, επιτυγχάνουν υψηλά ποσοστά ακρίβειας σε πραγματικό χρόνο, ιδιαίτερα όταν συνδυάζονται με τεχνικές μείωσης διαστάσεων ή επιλογής χαρακτηριστικών [90].

Ωστόσο, παρά την υψηλή απόδοσή τους, τα LSTM είναι υπολογιστικά απαιτητικά, ειδικά όταν επεξεργάζεται δεδομένα υψηλής διαστατικότητας χωρίς προεπεξεργασία. Ο χρόνος εκπαίδευσής τους είναι αυξημένος, ενώ απαιτούν μεγάλο όγκο δεδομένων για να αποδώσουν σωστά και να αποφύγουν το overfitting. Τέλος, είναι ευαίσθητα σε μη ισορροπημένα σύνολα δεδομένων, καθώς τείνουν να εστιάζουν στις κυρίαρχες κλάσεις και να αγνοούν τις σπάνιες επιθέσεις [92].

Συνολικά, τα LSTM αποτελούν ισχυρή επιλογή για την ανίχνευση εισβολών σε περιβάλλοντα IoT, καθώς μπορούν να επεξεργάζονται χρονικές ακολουθίες και να εντοπίζουν πρότυπα που εξελίσσονται με διαδοχικά βήματα.

7.7 Gated Recurrent Units

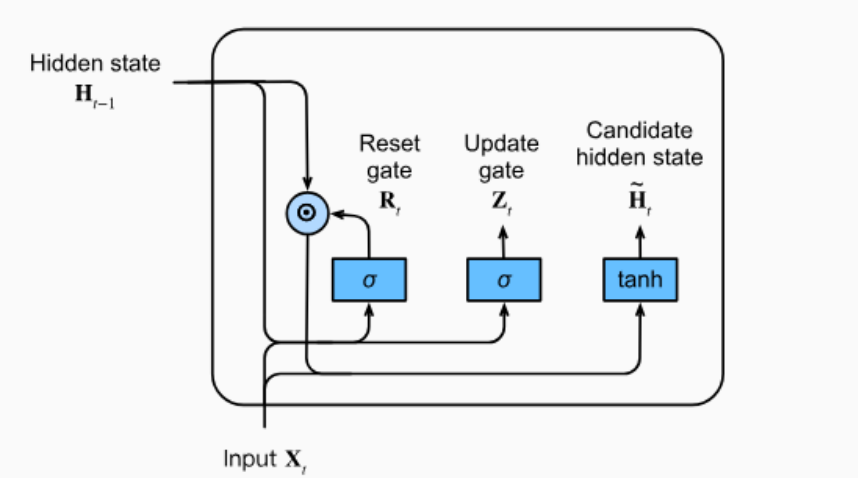
Τα Gated Recurrent Units (GRUs) αποτελούν παραλλαγή των LSTM, σχεδιασμένη με στόχο να μειώσει την υπολογιστική πολυπλοκότητα, διατηρώντας παράλληλα την ικανότητα εκμάθησης μακροπρόθεσμων εξαρτήσεων σε ακολουθιακά δεδομένα. Η βασική διαφορά με τα LSTMs είναι ότι έχουν απλούστερη αρχιτεκτονική, με λιγότερες gates και μικρότερο αριθμό παραμέτρων. Είναι πιο ελαφριά υπολογιστικά σε περιβάλλοντα με περιορισμένους πόρους, όπως τα IoT συστήματα [95]. Η απλούστερη αρχιτεκτονική τους δεν συνεπάγεται απαραίτητα χαμηλότερη απόδοση, καθώς σε πολλές περιπτώσεις τα GRUs επιτυγχάνουν ισοδύναμα ή καλύτερα αποτελέσματα στον εντοπισμό εισβολών, σε σχέση με τα LSTM [96].

Τα GRUs δεν διαθέτουν κύτταρο μνήμης, αλλά ένα διάνυσμα που ονομάζεται hidden state και διατηρεί πληροφορία από προηγούμενα χρονικά βήματα της ακολουθίας. Η ροή της πληροφορίας στο hidden state ελέγχεται από τις πύλες Reset και Update. Η Reset Gate καθορίζει αν και σε ποιον βαθμό η πληροφορία από το προηγούμενο χρονικό βήμα θα χρησιμοποιηθεί στον υπολογισμό της νέας κατάστασης. Η Update Gate ελέγχει αν η νέα πληροφορία θα αντικαταστήσει την προηγούμενη ή αν η προηγούμενη πληροφορία θα διατηρηθεί. Η διαδικασία εκπαίδευσης ολοκληρώνεται σε 7 βήματα [97]:

1. **Είσοδος δεδομένων.** Το μοντέλο δέχεται δύο εισόδους: το διάνυσμα εισόδου της τρέχουσας χρονικής στιγμής και το διάνυσμα hidden state από το προηγούμενο χρονικό βήμα. Στο πρώτο χρονικό βήμα όπου δεν υπάρχει προηγούμενη έξοδος, το μοντέλο χρησιμοποιεί μηδενικό διάνυσμα.
2. **Υπολογισμός Reset Gate.** Η reset gate καθορίζει πόση πληροφορία από το προηγούμενο hidden state θα χρησιμοποιηθεί στον υπολογισμό της νέας κατάστασης. Το διάνυσμα εισόδου και το προηγούμενο hidden state πολλαπλασιάζονται με συγκεκριμένα βάρη, προστίθεται μία σταθερά bias και στη συνέχεια εφαρμόζεται μια συνάρτηση sigmoid. Το αποτέλεσμα είναι ένα νέο διάνυσμα με τιμές μεταξύ 0 και 1. Κάθε τιμή καθορίζει σε ποιο βαθμό θα χρησιμοποιηθεί η πληροφορία στην αντίστοιχη θέση του προηγούμενου hidden state.
3. **Υπολογισμός Update Gate.** Η update gate καθορίζει πόση από την προηγούμενη πληροφορία θα διατηρηθεί και πόση νέα πληροφορία θα προστεθεί. Υπολογίζεται με παρόμοιο τρόπο με τη reset gate, χρησιμοποιώντας διαφορετικά βάρη και bias. Το διάνυσμα εισόδου και το προηγούμενο hidden state πολλαπλασιάζονται με βάρη, προστίθεται bias και εφαρμόζεται η συνάρτηση sigmoid. Το αποτέλεσμα είναι ένα διάνυσμα με τιμές από 0 έως 1, που ορίζουν την αναλογία διατήρησης και ενημέρωσης της μνήμης.
4. **Υπολογισμός υποψήφιας νέας κατάστασης.** Το προηγούμενο hidden state πολλαπλασιάζεται στοιχειομετρικά με τη reset gate, ώστε να ελεγχθεί ποιες πληροφορίες θα περάσουν στο επόμενο

στάδιο. Το νέο διάνυσμα που προκύπτει πολλαπλασιάζεται με έναν πίνακα βαρών, ενώ το διάνυσμα εισόδου πολλαπλασιάζεται με διαφορετικό πίνακα βαρών. Τα δύο αποτελέσματα προστίθενται με ένα bias, πριν εφαρμοστεί η συνάρτηση tanh, ώστε οι τιμές να περιοριστούν στο διάστημα (-1,1). Το αποτέλεσμα αποτελεί τη νέα υποψήφια κατάσταση που ενσωματώνει στοιχεία της τρέχουσας εισόδου και της φιλτραρισμένης παλαιότερης κατάστασης.

5. **Υπολογισμός τελικής νέας κατάστασης.** Για τον υπολογισμό του νέου hidden state, χρησιμοποιείται το προηγούμενο hidden state και η υποψήφια νέα κατάσταση, με βάση τις τιμές που προέκυψαν από την update gate. Αρχικά, κάθε τιμή της update gate πολλαπλασιάζεται με αντίστοιχη τιμή του προηγούμενου hidden state, ώστε να καθοριστεί ποιες πληροφορίες θα διατηρηθούν, ποιες θα μειωθούν και ποιες θα μηδενιστούν. Η κάθε τιμή της update gate αφαιρείται από το 1 και το αποτέλεσμα πολλαπλασιάζεται με την αντίστοιχη τιμή της υποψήφιας νέας κατάστασης. Με αυτόν τον τρόπο, καθορίζεται ποιες νέες πληροφορίες θα προστεθούν. Τέλος, τα δύο διανύσματα προστίθενται και το άθροισμα αποτελεί τη νέα τιμή του hidden state για το επόμενο χρονικό βήμα.
6. **Έξοδος του μοντέλου.** Καθώς το μοντέλο επεξεργάζεται διαδοχικά τα χρονικά βήματα της εισόδου, το hidden state ενημερώνεται δυναμικά, διατηρώντας μόνο τις σημαντικότερες πληροφορίες. Αφού ολοκληρωθεί η επεξεργασία όλων των χρονικών βημάτων, προκύπτει το τελικό hidden state, το οποίο περιέχει την πληροφορία που συγκεντρώθηκε μέχρι την τελευταία χρονική στιγμή. Το hidden state μεταφέρεται σε ένα πλήρως συνδεδεμένο νευρωνικό επίπεδο, όπου κάθε τιμή του συνδέεται με κάθε νευρώνα μέσω βαρών. Σε προβλήματα ταξινόμησης, χρησιμοποιείται ένας νευρώνας εξόδου με συνάρτηση ενεργοποίησης sigmoid, που επιστρέφει μία τιμή μεταξύ 0 και 1. Αυτή ερμηνεύεται ως πιθανότητα το δείγμα να υποδηλώνει κακόβουλη δραστηριότητα.
7. **Εκπαίδευση με BPTT.** Η έξοδος του μοντέλου συγκρίνεται με την πραγματική ετικέτα της εισόδου και υπολογίζεται το σφάλμα με χρήση της συνάρτησης κόστους binary cross-entropy. Το σφάλμα διαδίδεται προς τα πίσω σε όλα τα χρονικά βήματα με BPTT, ώστε να υπολογιστούν τα gradients ως προς τα βάρη και τα bias. Στη συνέχεια, οι τιμές αυτές χρησιμοποιούνται για την ενημέρωση των παραμέτρων του μοντέλου, μέσω αλγορίθμου βελτιστοποίησης, όπως ο Adam. Η διαδικασία επαναλαμβάνεται για κάθε δείγμα του συνόλου εκπαίδευσης, επιτρέποντας στο GRU να βελτιώνει σταδιακά την ικανότητά του να εντοπίζει κακόβουλα μοτίβα στη ροή των δεδομένων.



Σχήμα 7.5: Αρχιτεκτονική GRU [95]

Η εφαρμογή των GRUs σε συστήματα ανίχνευσης εισβολών σε περιβάλλοντα IoT έχει μελετηθεί, με στόχο την αυτόματη εξαγωγή χαρακτηριστικών και την ταξινόμηση της δικτυακής κίνησης [98]. Σε σχετική μελέτη, το σύστημα πέτυχε ακρίβεια 99.4%, με πολύ χαμηλά ποσοστά false positives (0,40%). Η απόδοσή του ήταν ιδιαίτερα υψηλή στον εντοπισμό επιθέσεων DoS και probing, αλλά χαμηλότερη σε περιπτώσεις μη εξουσιοδοτημένης πρόσβασης, λόγω περιορισμένης παρουσίας τέτοιων δειγμάτων στο σύνολο δεδομένων. Το αποτέλεσμα αυτό να δεικνύει την ευαισθησία του μοντέλου στην

κατανομή των δεδομένων και την περιορισμένη ικανότητα ανίχνευσης σπανιότερων αλλά σοβαρών επιθέσεων. Αν και η μελέτη επιβεβαιώνει ότι τα GRUs μπορούν να εφαρμοστούν αποτελεσματικά σε συστήματα ανίχνευσης εισβολών στο IoT, τα πειράματα βασίζονται αποκλειστικά σε συνθετικά δεδομένα και δεν περιλαμβάνουν αξιολόγηση σε πραγματικές συνθήκες λειτουργίας.

Τα GRUs διαθέτουν απλοποιημένη αρχιτεκτονική, διατηρώντας παράλληλα τη βασική ικανότητα επεξεργασίας ακολουθιών και αναγνώρισης χρονικών εξαρτήσεων στα δεδομένα. Λόγω του μικρότερου ρυθμού παραμέτρων, η εκπαίδευσή τους είναι γρηγορότερη και υπολογιστικά αποδοτικότερη, καθιστώντας τα κατάλληλα για περιβάλλοντα με περιορισμένους υπολογιστικούς πόρους, όπως το IoT.

Ωστόσο, αν και ανταποκρίνονται σε πολλά σενάρια, η απλούστερη δομή τους ενδέχεται να επηρεάσει την ικανότητά τους να διαχειριστούν πολύ μεγάλες ακολουθίες ή πολύπλοκες εξαρτήσεις, λόγω της μειωμένης δυνατότητας μνήμης τους. Επίσης, απαιτούν προσεκτική ρύθμιση των παραμέτρων για να επιτευχθεί η μέγιστη ακρίβεια και αν το σύνολο δεδομένων εκπαίδευσης είναι μικρό ή μη ισορροπημένο, μπορεί να γίνουν επιρρεπή στο overfitting. Τέλος, όπως και άλλα μοντέλα βαθιάς μάθησης, λειτουργούν ως black box, δηλαδή η ερμηνεία των αποφάσεών τους είναι δύσκολη.

Τα GRUs αποτελούν μια αποδοτική λύση για την ανάλυση χρονικών δεδομένων, ισορροπώντας ανάμεσα στην υπολογιστική αποδοτικότητα και την ακρίβεια. Παραμένουν αξιόπιστη εναλλακτική στα LSTM σε πλήθος εφαρμογών ανίχνευσης εισβολών, ιδιαίτερα όταν απαιτείται χαμηλή υπολογιστική πολυπλοκότητα.

7.8 Autoencoders

Τα Autoencoders (AEs) είναι μια κατηγορία νευρωνικών δικτύων που βασίζονται σε μη επιβλεπόμενη μάθηση και εκπαιδεύονται ώστε να αναπαριστούν τα δεδομένα με συμπιεσμένο τρόπο, διατηρώντας μόνο τις σημαντικές πληροφορίες [100]. Αποτελούνται από δύο βασικά μέρη: τον encoder και τον decoder. Ο encoder μετασχηματίζει τις αρχικές εισόδους σε έναν συμπαγή ενδιάμεσο χώρο (latent space ή bottleneck), ενώ ο decoder επιχειρεί να επαναφέρει τα δεδομένα στην αρχική τους μορφή από το latent space. Στόχος του δικτύου είναι να ελαχιστοποιήσει την απόκλιση μεταξύ αρχικής εισόδου και εξόδου, με ελάχιστη απώλεια πληροφορίας.

Η αρχιτεκτονική των Autoencoders τα καθιστά κατάλληλα για ανίχνευση αποκλίσεων από την κανονική κίνηση, σε IoT περιβάλλοντα όπου η κακόβουλη συμπεριφορά παρουσιάζει μικρές αλλά κρίσιμες αλλαγές. Ένα καλά εκπαιδευμένο μοντέλο μπορεί να ανακατασκευάζει με μεγάλη ακρίβεια τα φυσιολογικά δείγματα, ενώ εμφανίζει σημαντικό σφάλμα ανακατασκευής σε περιπτώσεις κακόβουλης δραστηριότητας.

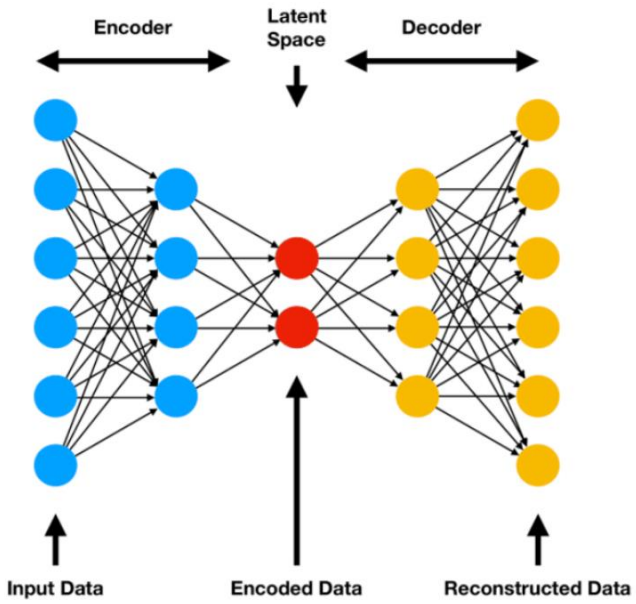
Η εκπαίδευση των Autoencoders ολοκληρώνεται σε 6 βήματα [100].

1. **Είσοδος δεδομένων.** Τα μη επισημασμένα δεδομένα εισάγονται στο επίπεδο εισόδου (input layer) σε μορφή διανύσματος. Κάθε νευρώνας του επιπέδου εισόδου αντιστοιχεί σε ένα χαρακτηριστικό των δεδομένων.
2. **Συμπίεση.** Ο encoder περιλαμβάνει ένα ή περισσότερα κρυφά επίπεδα και έχει στόχο να συμπιέσει την πληροφορία, αναπαριστώντας τα δεδομένα εισόδου με μειωμένες διαστάσεις. Τα δεδομένα μεταφέρονται στο πρώτο κρυφό επίπεδο του encoder. Οι νευρώνες του επιπέδου εισόδου συνδέονται με κάθε νευρώνα του πρώτου επιπέδου με weights που καθορίζουν πόσο σημαντικό είναι το κάθε χαρακτηριστικό. Κάθε νευρώνας πολλαπλασιάζει το σήμα εισόδου με το βάρος της σύνδεσης και οι πολλαπλασιασμένες τιμές αθροίζονται. Το αποτέλεσμα περνάει από μία συνάρτηση ενεργοποίησης, συνήθως ReLU ή Sigmoid. Η διαδικασία επαναλαμβάνεται για τα επόμενα κρυφά επίπεδα, με τον αριθμό των νευρώνων να μειώνεται σταδιακά. Καθώς η

πληροφορία διαδίδεται στον encoder, φιλτράρονται οι λιγότερο σημαντικές διαστάσεις και διατηρούνται μόνο τα βασικά χαρακτηριστικά.

3. **Συμπιεσμένος χώρος.** Μετά την επεξεργασία από τον encoder, τα δεδομένα εισόδου μετατρέπονται σε μία συμπαγή αναπαράσταση μειωμένης διάστασης, που ονομάζεται latent space. Στο σημείο αυτό, το μοντέλο έχει διατηρήσει τα πιο σημαντικά χαρακτηριστικά της εισόδου. Το αποτέλεσμα είναι ένα διάνυσμα (latent vector), που αποτελεί την έξοδο του encoder και την είσοδο του decoder.
4. **Αποσυμπίεση.** Το latent vector μεταφέρεται στον decoder, ο οποίος έχει στόχο να ανακατασκευάσει τα δεδομένα στην αρχική τους μορφή. Περιλαμβάνει ένα ή περισσότερα κρυφά επίπεδα, όπου κάθε επόμενο επίπεδο περιλαμβάνει περισσότερους νευρώνες. Ο decoder προσπαθεί να επαναφέρει τα δεδομένα στην αρχική τους διάσταση. Οι τιμές του latent vector πολλαπλασιάζονται με νέα βάρη, αθροίζονται και το αποτέλεσμα περνά από μία συνάρτηση ενεργοποίησης. Έτσι, οι νευρώνες αναπαράγουν σταδιακά τις κύριες πληροφορίες των αρχικών δεδομένων.
5. **Έξοδος.** Το επίπεδο εξόδου έχει ίδιο αριθμό νευρώνων με το επίπεδο εισόδου, ώστε να παράγει ένα διάνυσμα ίδιας διάστασης. Το τελικό αποτέλεσμα της αποκωδικοποίησης είναι η ανακατασκευή της αρχικής εισόδου.
6. **Υπολογισμός σφάλματος.** Η ανακατασκευασμένη έξοδος συγκρίνεται με την αρχική είσοδο μέσω μιας συνάρτησης απώλειας, συνήθως Mean Squared Error (MSE). Το σφάλμα μεταδίδεται προς τα πίσω με backpropagation και τα βάρη του encoder και του decoder ενημερώνονται με τη χρήση gradient descent ή κάποιας παραλλαγής του, όπως ο Adam. Η διαδικασία επαναλαμβάνεται για κάθε δείγμα, με στόχο τη συνεχή μείωση του σφάλματος ανακατασκευής.

Ένα νέο δείγμα χαρακτηρίζεται ως κακόβουλο αν το σφάλμα σε σχέση με την ανακατασκευή του υπερβαίνει ένα προκαθορισμένο κατώφλι. Αυτό το κατώφλι ορίζεται πειραματικά, με βάση την απόδοση σε άλλα, γνωστά φυσιολογικά δείγματα.



Σχήμα 7.6: Αρχιτεκτονική ενός Autoencoder [102]

Οι autoencoders έχουν αποδειχθεί εξαιρετικά αποτελεσματικοί στον εντοπισμό αποκλίσεων από την κανονική συμπεριφορά του δικτύου, λόγω της ικανότητάς τους να μαθαίνουν συμπιεσμένες αναπαραστάσεις των φυσιολογικών προτύπων κίνησης και να ανακατασκευάζουν τις εισόδους με ελάχιστο σφάλμα. Αυτό τούς καθιστά κατάλληλους για εφαρμογές εντοπισμού αποκλίσεων (anomaly detection), που αποτελούν θεωρητική βάση και για την ανίχνευση εισβολών σε δίκτυα IoT. Μπορούν

να ανιχνεύουν νέες επιθέσεις σε πραγματικό χρόνο και να εντοπίζουν τα πιο κρίσιμα χαρακτηριστικά, χωρίς να απαιτείται χειροκίνητη προεπεξεργασία δεδομένων ή επισημασμένα δεδομένα [101].

Ωστόσο, δεν υπάρχουν πολλές μελέτες στη βιβλιογραφία που να αξιοποιούν Autoencoders στην πράξη για εντοπισμό εισβολών σε περιβάλλοντα IoT, κυρίως επειδή η απόδοσή τους εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων εκπαίδευσης [104]. Οι περισσότερες μελέτες τους ενσωματώνουν ως μέρος πιο σύνθετων συστημάτων. Ενδεικτικά, σε μελέτη αξιοποιείται Autoencoder σε συνδυασμό με ensemble learning για την ταξινόμηση των ανωμαλιών σε συγκεκριμένους τύπους επιθέσεων, χρησιμοποιώντας επίπεδα GRU ως one-class μοντέλο για τον εντοπισμό αποκλίσεων από την φυσιολογική κίνηση [102].

7.9 Denoising Autoencoders

Μεγαλύτερη προσοχή στην έρευνα προσελκύουν διάφορες παραλλαγές των Autoencoders, με κυρίαρχη την κατηγορία των Denoising Autoencoders (DAEs) [103]. Η ικανότητά τους να ανακατασκευάζουν καθαρές αναπαραστάσεις από θορυβώδη δεδομένα τους καθιστά ιδιαίτερα κατάλληλους για ρεαλιστικά σενάρια. Οι DAEs αποτελούν παραλλαγή των κλασικών Autoencoders, σχεδιασμένη για την αντιμετώπιση δεδομένων με θόρυβο [104]. Σε αντίθεση με τους απλούς autoencoders, οι οποίοι εκπαιδεύονται ώστε να ανακατασκευάζουν την ίδια είσοδο, οι DAEs μαθαίνουν να ανακατασκευάζουν την καθαρή εκδοχή ενός σήματος που έχει αλλοιωθεί με τεχνητό θόρυβο. Στόχος είναι το μοντέλο να γίνει πιο ανθεκτικό και να γενικεύει καλύτερα, εντοπίζοντας μικρές αποκλίσεις στη δικτυακή κίνηση.

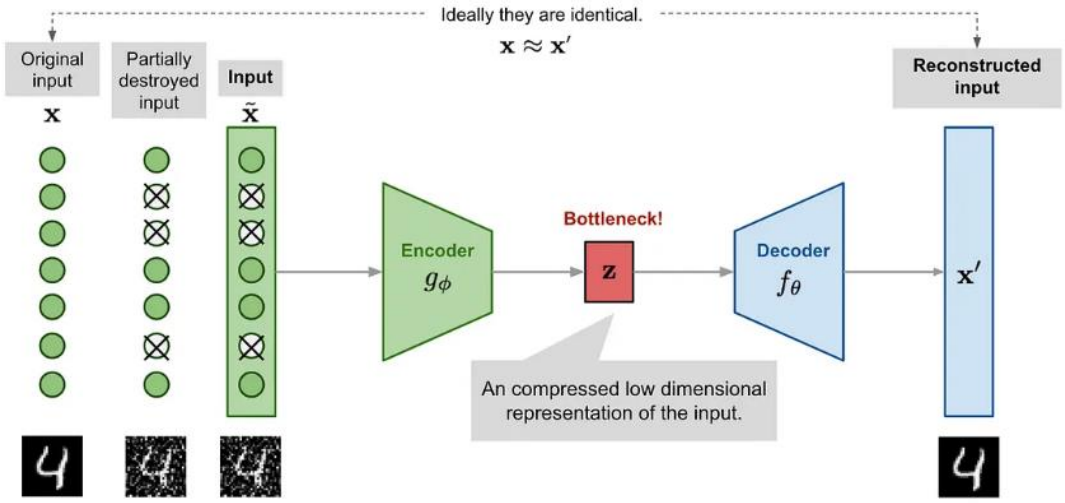
Η αρχιτεκτονική των DAEs είναι ίδια με αυτή των απλών autoencoders [103]. Η βασική τους διαφορά είναι το ότι, πριν την είσοδο στο δίκτυο, τα δεδομένα παραμορφώνονται με την προσθήκη ελεγχόμενου θορύβου. Αυτό μπορεί να γίνει με διάφορες τεχνικές:

- Gaussian noise: προσθήκη τυχαίου θορύβου σε κάθε τιμή εισόδου.
- Dropout noise: επιλεγμένα bits ή pixels μηδενίζονται, προσομοιώνοντας απώλεια πακέτων.
- Salt & Pepper Noise: ορισμένα bits ή pixels μετατρέπονται σε τυχαίες τιμές.
- Speckle Noise: εισάγεται θόρυβος που προσομοιώνει παρεμβολές.

Ο encoder μαθαίνει να αγνοεί μη κρίσιμες μεταβολές και αποκλίσεις και να διατηρεί τα ουσιώδη χαρακτηριστικά, διαμορφώνοντας ένα latent space που είναι λιγότερο ευαίσθητο σε θόρυβο ή αλλοιώσεις. Ο decoder ανακατασκευάζει την αρχική έκδοση των δεδομένων από τη συμπιεσμένη αναπαράσταση, προσπαθώντας να εξαλείψει τον επιπλέον θόρυβο. Δηλαδή, το μοντέλο μαθαίνει να εξάγει αναπαραστάσεις από θορυβώδη είσοδο, που οδηγούν σε καθαρή έξοδο.

Το σφάλμα υπολογίζεται μεταξύ της καθαρής αρχικής εισόδου και της εξόδου του μοντέλου. Επειδή η συνάρτηση απώλειας βασίζεται στην απόκλιση από την καθαρή είσοδο, δεν ενισχύει την είσοδο με θόρυβο, αλλά αναδεικνύει τις πληροφορίες που μένουν συνεπείς παρά την παραμόρφωση. Έτσι, κατά την εκπαίδευση, τα χαρακτηριστικά που σχετίζονται με το θόρυβο παραβλέπονται, καθώς δεν συνεισφέρουν στη βελτίωση της απόδοσης.

Τα βάρη ενημερώνονται μέσω backpropagation με στόχο την ελαχιστοποίηση της συνάρτησης απώλειας. Η διαδικασία επαναλαμβάνεται, ώστε το σύστημα να μάθει να εντοπίζει και να αφαιρεί αποτελεσματικά τον θόρυβο από τα δεδομένα [103]. Ένα νέο δείγμα χαρακτηρίζεται ως κακόβουλο αν το σφάλμα σε σχέση με την ανακατασκευή του υπερβαίνει ένα προκαθορισμένο κατώφλι.



Σχήμα 7.7: Αρχιτεκτονική Denoising Autoencoders [106].

Διάφορες μελέτες εξετάζουν την εφαρμογή των DAEs στον εντοπισμό εισβολών σε συστήματα IoT. Μία πρόσφατη εργασία [103], προτείνει ένα μοντέλο που εκπαιδεύεται σε παραμορφωμένα δείγματα και δείχνει οι DAEs μπορούν να αναγνωρίσουν τα σημαντικά χαρακτηριστικά των εισόδων, ακόμη και όταν αυτές έχουν αλλοιωθεί με θόρυβο. Η αφαίρεση θορύβου μέσω DAEs επιτυγχάνει υψηλή ακρίβεια που φτάνει το 99.6% χωρίς επίβλεψη. Τα αποτελέσματα έδειξαν ότι η χρήση των DAEs για συμπίεση των δεδομένων ενισχύει την απόδοση του μοντέλου, σε σύγκριση με άλλες μεθόδους. Ωστόσο, η αξιολόγηση έγινε σε offline περιβάλλον και όχι σε real-time, δυναμικό οικοσύστημα IoT. Τέλος, η γενίκευση του μοντέλου δεν εξετάζεται, καθώς τα πειράματα γίνονται σε ένα μόνο σύνολο δεδομένων.

Άλλη μελέτη παρουσιάζει ένα μη επιβλεπόμενο σύστημα ανίχνευσης εισβολών, που χρησιμοποιεί DAEs για εκμάθηση χαρακτηριστικών και συμπίεση των δεδομένων και ένα deep neural network για την ταξινόμηση των επιθέσεων [105]. Αποδείχτηκε ότι ο DAE μειώνει τη διάσταση των δεδομένων κατά 90%, διατηρώντας τη σημαντική πληροφορία και αφαιρώντας το θόρυβο. Το μοντέλο πέτυχε ακρίβεια εντοπισμού εισβολών που φτάνει το 99.62%. Επιπλέον, αναδείχθηκε η χρησιμότητα των DAEs στη συμπίεση και την αφαίρεση θορύβου πριν την ταξινόμηση. Συγκεκριμένα, η χρήση του DAE ως εργαλείο επιλογής χαρακτηριστικών βελτίωσε την απόδοση διάφορων ταξινομητών, όπως Random Forest και SVM. Ωστόσο, η απόδοση του συστήματος ήταν μειωμένη σε επιθέσεις με περιορισμένη παρουσία στο σύνολο εκπαίδευσης, όπως SQL injection. Επίσης, το σύστημα δεν αξιολογήθηκε σε real-time περιβάλλον IoT, καθώς τα πειράματα έγιναν σε offline συνθήκες. Τέλος, ως μελλοντική κατεύθυνση προτάθηκε η αυτοματοποίηση της ρύθμισης υπερπαραμέτρων με τεχνικές όπως meta-optimization.

Οι DAEs προσφέρουν σημαντικά πλεονεκτήματα στην ανίχνευση εισβολών σε περιβάλλοντα IoT. Εκπαιδεύονται με αλλοιωμένα δεδομένα και μαθαίνουν να ανακατασκευάζουν την αρχική είσοδο, αγνοώντας τις παρεμβολές. Έχουν καλή γενίκευση σε περιβάλλοντα όπως το IoT, όπου τα δεδομένα είναι μη επισημασμένα και ενδεχομένως παραμορφωμένα. Επιπλέον, δεν να απαιτούν επισημασμένα δεδομένα για εκπαίδευση, γεγονός που τους επιτρέπει να εντοπίζουν αποκλίσεις ακόμη και όταν δεν υπάρχουν γνωστά πρότυπα επιθέσεων. Αυτό τους καθιστά κατάλληλους και για τον εντοπισμό άγνωστων ή zero-day επιθέσεων [106].

Παρά τα πλεονεκτήματά τους, οι DAEs παρουσιάζουν ορισμένους περιορισμούς, όπως αυξημένες υπολογιστικές απαιτήσεις, λόγω της εισαγωγής και διαχείρισης τεχνητού θορύβου κατά την εκπαίδευση, ειδικά σε μεγάλα ή πολύπλοκα σύνολα δεδομένων. Επιπλέον, η επιλογή του είδους και της έντασης του θορύβου επηρεάζει άμεσα την απόδοση και απαιτεί προσεκτική ρύθμιση. Το ίδιο ισχύει και για το κατώφλι απόφασης (threshold) που καθορίζει αν ένα δείγμα θεωρείται κακόβουλο. Τέλος, οι DAEs είναι δύσκολοι στην ερμηνεία των αποτελεσμάτων (black box), ενώ σε ορισμένες περιπτώσεις ενδέχεται να απορρίψουν χρήσιμα δεδομένα που μοιάζουν με θόρυβο, οδηγώντας σε απώλεια πληροφορίας [104] [103].

Συνολικά, οι DAEs είναι ισχυρή λύση μη επιβλεπόμενης μάθησης για την ανίχνευση εισβολών σε συστήματα IoT, καθώς προσφέρουν ανθεκτικότητα και υψηλή απόδοση. Παρότι η χρήση τους απαιτεί προσεκτική παραμετροποίηση, μελέτες δείχνουν ότι μπορούν να συμβάλουν σημαντικά στην αναγνώριση επιθέσεων, ακόμη και σε πολύπλοκα ή θορυβώδη περιβάλλοντα.

7.10 Σύγκριση τεχνικών βαθιάς μάθησης

Μετά την ανάλυση των κυριότερων τεχνικών βαθιάς μάθησης που χρησιμοποιούνται στην ανίχνευση εισβολών σε περιβάλλοντα IoT, παρουσιάζονται οι σημαντικότεροι παράμετροι αξιολόγησης στον παρακάτω πίνακα.

Πίνακας 7.1: Σύγκριση τεχνικών βαθιάς μάθησης

Μοντέλο	**Υπολογιστικό κόστος**	**Απαιτήσεις δεδομένων**	**Ερμηνευσιμότητα**	**Προσαρμοστικότητα**	**Πραγματική εφαρμογή**
ANNs	Μέτριο	Επισημασμένα	Χαμηλή	Μέτρια	Ναι
CNNs	Υψηλό	Επισημασμένα	Χαμηλή	Μέτρια	Σε προσομοίωση
RNNs	Υψηλό	Επισημασμένα	Χαμηλή	Υψηλή	Σε προσομοίωση
LSTM	Υψηλό	Επισημασμένα	Χαμηλή	Υψηλή	Ναι
GRUs	Μέτριο	Επισημασμένα	Χαμηλή	Υψηλή	Ναι
Autoencoders	Μέτριο	Μη Επισημασμένα	Χαμηλή	Μέτρια	Οχι
Denoising Autoencoders	Υψηλό	Μη Επισημασμένα	Χαμηλή	Υψηλή	Ναι

7.11 Επίλογος

Σε αυτό το κεφάλαιο παρουσιάστηκαν τα βασικά μοντέλα βαθιάς μάθησης που χρησιμοποιούνται για τον εντοπισμό εισβολών σε οικοσυστήματα IoT. Για κάθε τεχνική που εξετάστηκε, αναλύθηκε ο τρόπος λειτουργίας της, η εφαρμογή της στο πλαίσιο του IoT, καθώς και τα βασικά πλεονεκτήματα και μειονεκτήματά της. Η επιλογή της κάθε τεχνικής εξαρτάται από τις απαιτήσεις και τους περιορισμούς του εκάστοτε περιβάλλοντος. Τα μοντέλα βαθιάς μάθησης προσφέρουν υψηλή ακρίβεια, δυνατότητα

εντοπισμού άγνωστων επιθέσεων και ανθεκτικότητα σε αλλοιωμένα δεδομένα. Ωστόσο, η εφαρμογή τους απαιτεί προσεκτικό σχεδιασμό και αξιολόγηση, ώστε να εξασφαλίζεται η αξιοπιστία και η αποδοτικότητά τους σε πραγματικά περιβάλλοντα IoT.

Κεφάλαιο 8ο Κατανομή Πόρων στο IoT

8.1 Εισαγωγή

Το κεφάλαιο αυτό εξετάζει τη διαχείριση πόρων σε οικοσυστήματα IoT, ένα κρίσιμο ζήτημα που προκύπτει λόγω των περιορισμένων υπολογιστικών και ενεργειακών πόρων των συσκευών. Αρχικά, ορίζεται η έννοια της κατανομής πόρων και αναδεικνύεται η σημασία της διατήρησης της αποδοτικότητας των συστημάτων. Στη συνέχεια, παρουσιάζονται σύγχρονες τεχνικές Τεχνητής Νοημοσύνης που εφαρμόζονται για τη βελτιστοποίηση της κατανομής πόρων, αρκετές από τις οποίες μπορούν να χρησιμοποιηθούν και για τον εντοπισμό εισβολών. Για κάθε τεχνική, περιγράφεται ο τρόπος λειτουργίας της, η εφαρμογή της σε συστήματα IoT, καθώς και τα βασικά πλεονεκτήματα και μειονεκτήματά της.

8.2 Κατανομή Πόρων στο IoT

Η κατανομή πόρων στο IoT αναφέρεται στη διαδικασία διαμοιρασμού υπολογιστικών, ενεργειακών και δικτυακών πόρων σε συσκευές ή εφαρμογές που λειτουργούν σε ένα κατανεμημένο περιβάλλον [107]. Οι πόροι αυτοί περιλαμβάνουν επεξεργαστική ισχύ (CPU), χώρο αποθήκευσης, εύρος ζώνης (bandwidth), ενέργεια (σε συσκευές με περιορισμένη διάρκεια μπαταρίας) και προτεραιότητα πρόσβασης σε κρίσιμες υπηρεσίες. Σε ένα περιβάλλον όπου χιλιάδες συσκευές επικοινωνούν ταυτόχρονα, η αποτελεσματική κατανομή αυτών των πόρων είναι απαραίτητη τόσο για την ομαλή λειτουργία του συστήματος και τη συνολική απόδοσή του, όσο και για την ασφάλειά του, καθώς η αναποτελεσματική κατανομή πόρων μπορεί να κάνει το σύστημα ευάλωτο σε επιθέσεις όπως οι DDoS.

Οι παραδοσιακές προσεγγίσεις κατανομής πόρων, όπως οι στατικοί κανόνες κατανομής ή αλγόριθμοι Round-Robin και First-Come-First-Served (FCFS), βασίζονται σε απλοποιημένα μοντέλα και δεν ανταποκρίνονται στις δυναμικές συνθήκες και στις απαιτήσεις των σύγχρονων περιβαλλόντων IoT. Οι συσκευές είναι ετερογενείς, με διαφορετικές δυνατότητες και ανάγκες, ενώ το δίκτυο λειτουργεί δυναμικά, καθώς νέες συσκευές προστίθενται ή αποσυνδέονται. Πολλές εφαρμογές λειτουργούν σε πραγματικό χρόνο και έχουν αυστηρούς χρονικούς περιορισμούς. Λόγω αυτών των χαρακτηριστικών, οι παραδοσιακοί αλγόριθμοι κατανομής πόρων δεν μπορούν να ανταποκριθούν επαρκώς. Επιπρόσθετα, συνήθως δεν ενσωματώνουν μηχανισμούς που λαμβάνουν υπόψη τις απειλές ασφάλειας, αφήνοντας τα συστήματα εκτεθειμένα [108].

Για την αντιμετώπιση αυτών των προκλήσεων, το ερευνητικό ενδιαφέρον έχει στραφεί στην εφαρμογή τεχνικών τεχνητής νοημοσύνης, που έχουν στόχο τόσο στη δυναμική και αποδοτική κατανομή των διαθέσιμων πόρων, όσο και στην ενίσχυση της ασφάλειας των συστημάτων [107]. Τέτοιες τεχνικές περιλαμβάνουν μεθόδους μη επιβλεπόμενης μάθησης, ενισχυτικής μάθησης (reinforcement learning) και νευρωνικών δικτύων, που επιτρέπουν στα συστήματα να μαθαίνουν από την εμπειρία και να προσαρμόζουν δυναμικά τις αποφάσεις κατανομής, με ελάχιστη ή μηδενική ανθρώπινη παρέμβαση. Παράλληλα, έχουν τη δυνατότητα να βελτιστοποιούν ταυτόχρονα πολλαπλά κριτήρια, όπως ο χρόνος απόκρισης και η ενεργειακή κατανάλωση. Χάρη στην αυτονομία και την ικανότητά τους να επεξεργάζονται μεγάλα σύνολα δεδομένων σε πραγματικό χρόνο, οι τεχνικές αυτές θεωρούνται κατάλληλες για εφαρμογή σε πολύπλοκα περιβάλλοντα IoT [108].

8.3 Τεχνητή νοημοσύνη για ανίχνευση εισβολών και κατανομή πόρων

Για τη διατήρηση της απόδοσης και της αξιοπιστίας στα σύγχρονα, δυναμικά δίκτυα IoT, απαιτείται ένας μηχανισμός λήψης αποφάσεων που μπορεί να προσαρμόζεται σε πραγματικό χρόνο, αξιοποιώντας τα δεδομένα του δικτύου. Δύο βασικοί τομείς που αναδεικνύουν αυτή την ανάγκη για δυναμική προσαρμογή είναι η ανίχνευση εισβολών (π.χ. το σύστημα καλείται να αποφασίσει αν θα επιτραπεί ή θα αποκλειστεί η επικοινωνία μιας συσκευής) και η κατανομή πόρων (π.χ. καθορίζεται σε ποια συσκευή θα κατανεμηθούν οι διαθέσιμοι πόροι, όπως το bandwidth). Αν και τα δύο προβλήματα διαφέρουν στον στόχο, παρουσιάζουν κοινά υπολογιστικά χαρακτηριστικά: απαιτούν λήψη αποφάσεων σε μεταβαλλόμενο περιβάλλον, με δεδομένα που είναι συχνά ελλιπή ή περιέχουν θόρυβο. Κάθε απόφαση βασίζεται στην τρέχουσα κατάσταση του συστήματος και επηρεάζει άμεσα την απόδοση και την ασφάλειά του [3][107].

Η τεχνητή νοημοσύνη μπορεί να καλύψει τις απαιτήσεις προσαρμοστικότητας και γενίκευσης μέσω αλγορίθμων που εκπαιδεύονται ώστε να λαμβάνουν αποφάσεις με βάση τα διαθέσιμα δεδομένα. Οι αλγόριθμοι αυτοί έχουν τη δυνατότητα να επεξεργάζονται πολύπλοκα δεδομένα μεγάλου όγκου, να εντοπίζουν μοτίβα και να μαθαίνουν από την εμπειρία. Αυτό επιτρέπει τη διαμόρφωση μοντέλων που δεν βασίζονται σε στατικούς κανόνες, αλλά προσαρμόζονται αυτόνομα σε διαφορετικές συνθήκες και περιορισμούς, όπως εκείνες που υπάρχουν στα περιβάλλοντα IoT. Για τον λόγο αυτό, συγκεκριμένοι αλγόριθμοι τεχνητής νοημοσύνης μπορούν να χρησιμοποιηθούν ευρέως τόσο για τον εντοπισμό εισβολών, όσο και για τη βελτιστοποίηση της διαχείρισης πόρων.

8.4 Reinforcement Learning

Το Reinforcement Learning (RL) είναι ένας τύπος μηχανικής μάθησης, όπου ο αλγόριθμος μαθαίνει να παίρνει αποφάσεις μέσω αλληλεπίδρασης με το περιβάλλον του, με στόχο τη βελτιστοποίηση μιας συνάρτησης απόδοσης (reward function) [109]. Το RL δεν ανήκει στις τεχνικές επιβλεπόμενης ή μη επιβλεπόμενης μάθησης, καθώς δεν βασίζεται σε επισημασμένα δεδομένα ή στην ανακάλυψη μοτίβων. Το περιβάλλον επηρεάζεται από τις ενέργειες ενός πράκτορα (agent), ο οποίος εκπαιδεύεται μέσω ανάδρασης (feedback) από το περιβάλλον. Μετά από κάθε ενέργεια, ο agent λαμβάνει feedback υπό τη μορφή ανταμοιβής ή ποινής. Αν η ενέργειά του οδηγεί σε θετικό αποτέλεσμα, η θετική ανταμοιβή αυξάνει την πιθανότητα να επιλεγεί ξανά η ίδια ενέργεια σε παρόμοιο πλαίσιο. Αν το αποτέλεσμα είναι αρνητικό, μειώνεται η πιθανότητα να επιλεγεί ξανά στο μέλλον. Η διαδικασία εκμάθησης είναι αλληλεπιδραστική και διαρκώς εξελισσόμενη. Το σύστημα μαθαίνει να βελτιώνει τις ενέργειές του, προσπαθώντας να μεγιστοποιήσει τη συνολική του ανταμοιβή μακροπρόθεσμα [110].

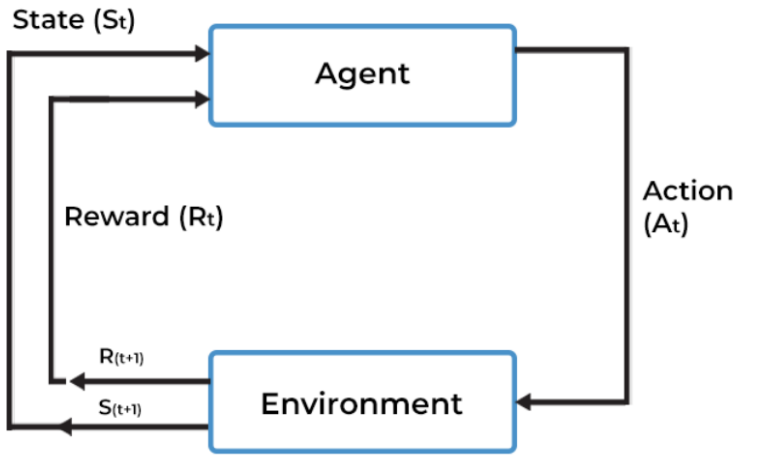
8.4.1 Βασικές αρχές και λειτουργία του RL

Αρχικά, ο agent παρατηρεί την τρέχουσα κατάσταση του περιβάλλοντος με σκοπό να επιλέξει την πρώτη ενέργεια που θα ακολουθήσει. Στα αρχικά στάδια της εκπαίδευσης, η επιλογή γίνεται μέσω εξερεύνησης (exploration), δηλαδή με δοκιμή ενεργειών που επιλέγονται τυχαία. Καθώς ο agent δεν διαθέτει ακόμη γνώση του περιβάλλοντος, η εξερεύνηση του επιτρέπει να ανακαλύψει στρατηγικές που οδηγούν σε θετικά αποτελέσματα [109].

Αφού εκτελέσει την ενέργεια, ο agent λαμβάνει μία ανταμοιβή (reward), δηλαδή μία αριθμητική τιμή που εκφράζει πόσο κατάλληλη ήταν η ενέργεια ως προς τον στόχο. Αν η ενέργεια θεωρείται αποτελεσματική, η ανταμοιβή είναι θετική. Αν οδηγεί σε αρνητικές συνέπειες, η ανταμοιβή είναι αρνητική (penalty).

Ο agent αξιοποιεί τις ανταμοιβές για να προσαρμόσει τη στρατηγική του (policy), η οποία καθορίζει ποιες ενέργειες είναι προτιμότερες σε κάθε περίπτωση. Η στρατηγική ενημερώνεται σταδιακά μέσω αλγορίθμων, όπως το Q-learning και το State-Action-Reward-State-Action (SARSA). Στο Q-learning ο agent διατηρεί εκτιμήσεις της αξίας κάθε ζεύγους κατάστασης - ενέργειας σε έναν πίνακα Q, ο οποίος ενημερώνεται μετά από κάθε αλληλεπίδραση με το περιβάλλον. Στο SARSA, η ενημέρωση της στρατηγικής γίνεται με βάση την αλληλουχία ενεργειών που εκτελέστηκαν. Σε κάθε περίπτωση, η πιθανότητα επιλογής μιας ενέργειας αυξάνεται αν στο παρελθόν απέδωσε υψηλή ανταμοιβή, ενώ μειώνεται αν είχε αρνητική επίδραση [110].

Η διαδικασία επαναλαμβάνεται συνεχώς, επιτρέποντας στον agent να βελτιώνει τη στρατηγική του μέσω της εμπειρίας. Όσο προχωρά η εκπαίδευση, ο agent επιλέγει ενέργειες που έχουν ήδη αποδώσει υψηλές ανταμοιβές (exploitation). Ωστόσο, συνεχίζει να κάνει εξερεύνηση, ώστε να διατηρεί την ικανότητα προσαρμογής σε πιθανές αλλαγές του περιβάλλοντος. Η στρατηγική συγκλίνει σταδιακά σε μία βέλτιστη πολιτική, που επιτρέπει στον agent να επιλέγει τις καταλληλότερες ενέργειες και να μεγιστοποιεί τη συνολική ανταμοιβή του.



Σχήμα 8.1: Βασική ροή λειτουργίας του RL [109]

8.4.2 RL στον εντοπισμό εισβολών στο IoT

Στον εντοπισμό εισβολών σε περιβάλλοντα IoT, ένας agent μπορεί να εκπαιδευτεί ώστε να εντοπίζει κακόβουλες δραστηριότητες με βάση τη συμπεριφορά των συσκευών στο δίκτυο, χωρίς την ανάγκη για επισημασμένα δεδομένα [111]. Ο agent παρακολουθεί συνεχώς τη ροή του δικτύου και ανάλογα με την κατάστασή του, επιλέγει μια ενέργεια όπως να επιτρέψει ή να απορρίψει μία εισερχόμενη επικοινωνία. Η ανταμοιβή ορίζεται με βάση την ορθότητα της απόφασης. Για παράδειγμα, αν μία απειλή εντοπιστεί, ο agent λαμβάνει θετική ανταμοιβή, ενώ αν επιτραπεί μια κακόβουλη επικοινωνία, η ανταμοιβή είναι αρνητική. Μέσω της εκπαίδευσης, ο agent θα μάθει να βελτιώνει τις αποφάσεις του ώστε να αναγνωρίζει τις επιθέσεις.

Αν και η προσέγγιση RL είναι θεωρητικά εφαρμόσιμη στον εντοπισμό εισβολών σε περιβάλλοντα IoT, η χρήση της στη βιβλιογραφία είναι περιορισμένη. Οι περισσότερες έρευνες επικεντρώνονται σε παραλλαγές της μεθόδου, όπως το Deep Reinforcement Learning, καθώς στην πράξη το RL παρουσιάζει δυσκολίες στη διαχείριση περιβαλλόντων μεγάλης πολυπλοκότητας και δεδομένων υψηλών διαστάσεων.

Η εφαρμογή του RL στον εντοπισμό εισβολών παρουσιάζει σημαντικά πλεονεκτήματα, που βασίζονται κυρίως στη δυνατότητα να λειτουργεί χωρίς επισημασμένα δεδομένα και να προσαρμόζεται δυναμικά σε νέες συνθήκες [110]. Μπορεί να εντοπίζει νέες και άγνωστες απειλές ανιχνεύοντας μοτίβα κακόβουλης συμπεριφοράς, ενώ η μάθηση είναι συνεχής, επιτρέποντας στο σύστημα να προσαρμόζεται καθώς το δίκτυο μεταβάλλεται [111].

Ωστόσο, παρουσιάζει κάποια μειονεκτήματα. Η διαδικασία εκπαίδευσης μπορεί να είναι υπολογιστικά απαιτητική, ειδικά όταν το περιβάλλον έχει πολλές επιλογές ενεργειών. Σε μεγάλα και σύνθετα δίκτυα, η σύγκλιση είναι αργή, καθώς ο agent χρειάζεται μεγάλο αριθμό επαναλήψεων μέχρι να καταλήξει στη βέλτιστη πολιτική. Τέλος, η απόδοση του συστήματος εξαρτάται άμεσα από το πως ορίζεται η ανταμοιβή. Αυτό μπορεί να είναι πρόβλημα στον εντοπισμό εισβολών, όπου κάποιες επιθέσεις δεν είναι άμεσα παρατηρήσιμες [111].

8.4.3 RL στην κατανομή πόρων στο IoT

Στην κατανομή πόρων, το RL εφαρμόζεται ώστε ο agent να μαθαίνει πώς να διαχειρίζεται δυναμικά τους διαθέσιμους πόρους του συστήματος IoT [112]. Με βάση την τρέχουσα κατάσταση του δικτύου, ο agent επιλέγει ενέργειες που κατανέμουν το bandwidth, την υπολογιστική ισχύ και άλλους πόρους μεταξύ συσκευών ή υπηρεσιών. Η ανταμοιβή καθορίζεται από την αποτελεσματικότητα της απόφασης: για παράδειγμα, αν υπάρξει μείωση της καθυστέρησης, η ανταμοιβή θα είναι θετική, ενώ αν προκληθεί συμφόρηση, θα είναι αρνητική. Μέσα από συνεχή αλληλεπίδραση με το περιβάλλον, ο agent βελτιώνει τη στρατηγική του, βελτιστοποιώντας σταδιακά τη χρήση των πόρων του συστήματος.

Η εφαρμογή του RL για βελτιστοποίηση της κατανομής πόρων σε IoT περιβάλλοντα έχει αξιολογηθεί σε πρόσφατη μελέτη, όπου πολλές συσκευές υποβάλλουν αιτήματα για την εκτέλεση εργασιών σε edge κόμβους [113]. Στη μελέτη έγιναν προσομοιώσεις με αυξανόμενο αριθμό αιτημάτων και δυναμικές συνθήκες. Τα αποτελέσματα έδειξαν ότι ο agent κατάφερε να μειώσει τη μέση καθυστέρηση απόκρισης και να αυξήσει το ποσοστό εργασιών που ολοκληρώθηκαν επιτυχώς εντός χρονικών ορίων, σε σύγκριση με στατικές μεθόδους κατανομής. Η απόδοση παρέμεινε σταθερή καθώς αυξανόταν ο φόρτος, ενώ τα στατικά συστήματα παρουσίασαν σημαντική επιβάρυνση. Ωστόσο, η πολυπλοκότητα του αλγορίθμου μεγαλώνει όσο αυξάνονται οι διαθέσιμοι κόμβοι και τα αιτήματα, οδηγώντας σε μεγαλύτερο χρόνο εκπαίδευσης. Επιπλέον, η μέθοδος προϋποθέτει ότι οι κόμβοι μπορούν να ταξινομηθούν σε ομάδες με παρόμοια χαρακτηριστικά, κάτι που περιορίζει την εφαρμογή της σε ετερογενή δίκτυα.

Η χρήση του RL στην κατανομή πόρων σε IoT δίκτυα προσφέρει σημαντικά πλεονεκτήματα, καθώς επιτρέπει στο σύστημα να προσαρμόζει δυναμικά τις αποφάσεις του με βάση την τρέχουσα κατάσταση του περιβάλλοντος [112]. Ο agent μπορεί να διαχειρίζεται υπολογιστικούς πόρους σε πραγματικό χρόνο και να προσαρμόζεται σε μεταβαλλόμενες συνθήκες φόρτου. Η συνολική απόδοση του συστήματος αυξάνεται και η χρήση των πόρων βελτιστοποιείται σταδιακά μέσω της εμπειρίας. Τέλος, το RL είναι κατάλληλο για κατανεμημένα περιβάλλοντα, αφού μπορεί να εφαρμοστεί τοπικά σε κόμβους χωρίς την ανάγκη κεντρικού ελέγχου, μειώνοντας τις καθυστερήσεις και αυξάνοντας την αυτονομία του συστήματος [113].

Παρά τα οφέλη του, το RL παρουσιάζει κάποια μειονεκτήματα στην κατανομή πόρων. Όσο αυξάνεται ο αριθμός των συσκευών και των διαθέσιμων επιλογών, η εκπαίδευση του agent γίνεται πιο αργή και υπολογιστικά απαιτητική. Επίσης, η ποιότητα της πολιτικής επηρεάζεται από τις μεταβλητές που περιγράφουν την κατάσταση του συστήματος, όπως η χρήση της CPU και το διαθέσιμο bandwidth. Τέλος, η αξιοπιστία της πολιτικής του RL εξαρτάται από την ικανότητα του συστήματος να μετρά και

να αξιολογεί την απόδοση των ενεργειών του agent, κάτι που ενδέχεται να μην είναι εφικτό σε κάποια περιβάλλοντα IoT.

8.5 Deep Reinforcement Learning

Το Deep Reinforcement Learning (DRL) αποτελεί συνδυασμό του παραδοσιακού RL με deep neural networks και χρησιμοποιείται σε δυναμικά συστήματα μεγάλης κλίμακας και υψηλής διαστατικότητας, όπως τα δίκτυα IoT [114]. Σε αντίθεση με το RL, που βασίζεται σε πίνακες για αποθήκευση των καταστάσεων και ενεργειών, το DRL χρησιμοποιεί βαθιά δίκτυα για να μάθει μία πολιτική που μεγιστοποιεί τη συνολική ανταμοιβή. Το νευρωνικό δίκτυο μαθαίνει να εκτιμά την αναμενόμενη συνολική ανταμοιβή για κάθε ενέργεια, ώστε ο agent να επιλέγει τη βέλτιστη σε κάθε κατάσταση. Μπορεί να επεξεργάζεται μεγάλα σύνολα δεδομένων, επιτρέποντας στον πράκτορα να λαμβάνει πιο αποτελεσματικές αποφάσεις, ακόμα και σε δυναμικά και πολύπλοκα περιβάλλοντα.

8.5.1 Βασικές αρχές και λειτουργία του DRL

Όπως και στο RL, η εκπαίδευση στο DRL βασίζεται στη συνεχή αλληλεπίδραση ενός agent με το περιβάλλον του. Ο agent παρατηρεί την τρέχουσα κατάσταση, επιλέγει μία ενέργεια και λαμβάνει μία ανταμοιβή που αντανακλά το αποτέλεσμα της ενέργειας που επέλεξε [116].

Αρχικά, ο agent παρατηρεί την κατάσταση του περιβάλλοντος και τη μετατρέπει σε διάνυσμα που εισάγεται στο νευρωνικό δίκτυο. Η έξοδος του νευρωνικού δικτύου είναι ένα διάνυσμα που περιέχει μία αριθμητική τιμή για κάθε δυνατή ενέργεια που μπορεί να επιλέξει ο agent. Οι τιμές δηλώνουν τις εκτιμώμενες ανταμοιβές για κάθε ενέργεια, αλλά αρχικά είναι τυχαίες, ως αποτέλεσμα των αρχικών, μη εκπαιδευμένων βαρών. Ο agent επιλέγει τυχαία μια ενέργεια και την εκτελεί στο περιβάλλον. Το περιβάλλον επιστρέφει την ανταμοιβή και τη νέα κατάσταση στην οποία βρέθηκε. Αυτά τα δεδομένα, δηλαδή η αρχική κατάσταση, η ενέργεια που επιλέχθηκε, η ανταμοιβή και η νέα κατάσταση, αποθηκεύονται ως μία εμπειρία σε ειδική μνήμη (replay buffer) που θα χρησιμοποιηθεί για την εκπαίδευση του νευρωνικού δικτύου.

Η διαδικασία επαναλαμβάνεται μέχρι να συγκεντρωθεί επαρκής αριθμός εμπειριών. Ο agent ενεργεί και καταγράφει εμπειρίες αλληλεπίδρασης με το περιβάλλον, δημιουργώντας μία συλλογή από δεδομένα που περιγράφουν το αποτέλεσμα της κάθε ενέργειας.

Η εκπαίδευση του νευρωνικού δικτύου βασίζεται σε τυχαίες εμπειρίες από το παρελθόν. Για κάθε εμπειρία, το νευρωνικό δίκτυο δέχεται ως είσοδο μόνο την αρχική κατάσταση του περιβάλλοντος. Η έξοδός του είναι ένα διάνυσμα που περιέχει μία τιμή για κάθε δυνατή ενέργεια από αυτή την κατάσταση, η οποία εκφράζει την εκτίμηση της συνολικής αναμενόμενης ανταμοιβής. Από αυτές τις τιμές, επιλέγεται εκείνη που αντιστοιχεί στην ενέργεια που είχε πραγματοποιηθεί στην εμπειρία, ώστε να υπολογιστεί το σφάλμα σε σχέση με μία τιμή-στόχο (target value) [114].

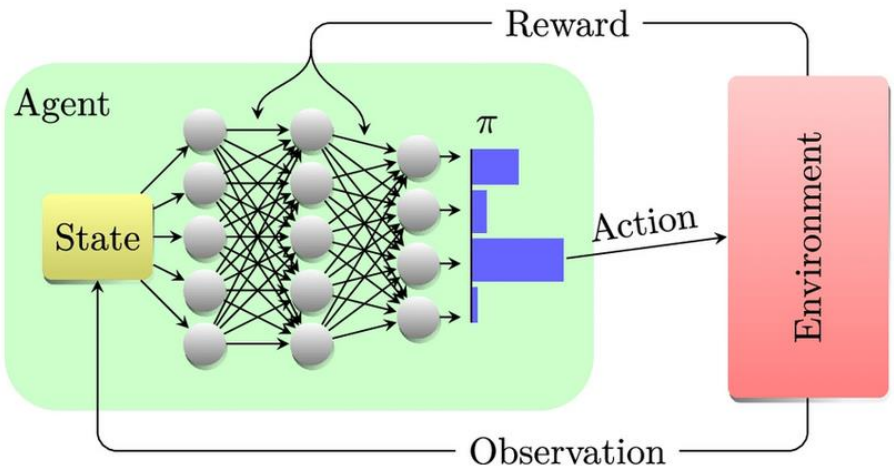
Για να υπολογιστεί η τιμή-στόχος, θα χρησιμοποιηθεί ένα δεύτερο νευρωνικό δίκτυο (target network), που έχει ίδια αρχιτεκτονική με το αρχικό αλλά παραμένει σταθερό για περισσότερα βήματα εκπαίδευσης. Για κάθε εμπειρία, το target network λαμβάνει ως είσοδο τη νέα κατάσταση του περιβάλλοντος, που προέκυψε μετά την εκτέλεση της ενέργειας. Στη συνέχεια, υπολογίζει εκτιμήσεις της συνολικής αναμενόμενης ανταμοιβής για όλες τις δυνατές ενέργειες από αυτή τη νέα κατάσταση. Επιλέγει τη μεγαλύτερη ανταμοιβή, που αποτελεί την καλύτερη εκτίμηση για τις μελλοντικές ανταμοιβές από τη νέα κατάσταση και μετά. Αυτή η τιμή αθροίζεται με την ανταμοιβή της ενέργειας

από την εμπειρία. Το αποτέλεσμα, δηλαδή η τιμή-στόχος, περιλαμβάνει την άμεση ανταμοιβή της ενέργειας που έχει ήδη συμβεί αλλά και την εκτίμηση για το μέλλον μέσω του target network.

Στη συνέχεια, η πρόβλεψη του πρώτου δικτύου συγκρίνεται με την τιμή-στόχο. Η διαφορά τους εκφράζει το σφάλμα του δικτύου για τη συγκεκριμένη εμπειρία. Όσο μεγαλύτερη είναι η απόκλιση, τόσο περισσότερο απέχει η εκτίμηση του δικτύου από αυτό που θα έπρεπε να έχει προβλέψει, με βάση την πραγματική κατάσταση του περιβάλλοντος μετά την ενέργεια του agent [117].

Το σφάλμα χρησιμοποιείται για να ενημερωθούν τα βάρη του κύριου δικτύου μέσω backpropagation. Οι παράμετροι και τα βάρη προσαρμόζονται, ώστε να ελαχιστοποιηθεί η συνάρτηση απώλειας. Η διαδικασία επαναλαμβάνεται για πολλές εμπειρίες και το μοντέλο γίνεται σταδιακά πιο αποτελεσματικό στην εκτίμηση της συνολικής αναμενόμενης ανταμοιβής. Η λειτουργία και εκπαίδευση του δικτύου συνεχίζεται, καθώς ο agent αλληλεπιδρά με το περιβάλλον και συγκεντρώνει νέες εμπειρίες. Σε κάθε χρονικό βήμα, ο agent είτε επιλέγει ενέργεια με βάση την τρέχουσα πρόβλεψη του δικτύου, είτε συνεχίζει την εξερεύνηση. Το δίκτυο εκπαιδεύεται και βελτιώνεται συνεχώς.

Το target network δεν ενημερώνεται συνεχώς. Αντίθετα, παραμένει σταθερό για ορισμένα χρονικά διαστήματα και στη συνέχεια ενημερώνεται μαζικά, αντιγράφοντας τα βάρη από το κύριο δίκτυο. Έτσι, τα δύο δίκτυα διατηρούνται συγχρονισμένα, αλλά όχι συνεχώς. Αυτό εξασφαλίζει ότι οι τιμές-στόχοι που χρησιμοποιούνται για την εκπαίδευση δεν αλλάζουν συνεχώς, με αποτέλεσμα η διαδικασία εκπαίδευσης να μην βασίζεται σε μεταβαλλόμενες προβλέψεις, προσφέροντας σταθερότητα [114].



Σχήμα 8.2: Δομή και λειτουργία του DRL [116]

8.5.2 DRL στον εντοπισμό εισβολών στο IoT

Το DRL χρησιμοποιείται στον εντοπισμό εισβολών IoT και βασίζεται στην ικανότητα του agent να διακρίνει, μέσω αλληλεπίδρασης με το περιβάλλον, ποια δικτυακή δραστηριότητα είναι φυσιολογική και ποια θα μπορούσε να υποδηλώνει επίθεση. Σε κάθε βήμα, ο agent λαμβάνει μία περιγραφή της τρέχουσας κατάστασης του δικτύου, όπως πληροφορίες για πρωτόκολλα, ports και διευθύνσεις IP και επιλέγει μία ενέργεια, ανάλογα με το αν το δείγμα αξιολογείται ως φυσιολογική ή ύποπτη δραστηριότητα [111]. Το περιβάλλον επιστρέφει μία ανταμοιβή που εξαρτάται από το αν έκανε σωστή εκτίμηση και το νευρωνικό δίκτυο προσαρμόζει τα βάρη του σταδιακά, ώστε να αυξάνει την αποτελεσματικότητα του agent. Με αυτό τον τρόπο, ο agent μαθαίνει να εντοπίζει με ακρίβεια επιθέσεις σε πραγματικό χρόνο, ακόμη και σε περιπτώσεις που δεν παρουσιάστηκαν κατά την αρχική φάση εκπαίδευσης [117].

Σε σχετική μελέτη, ερευνήθηκε η εφαρμογή του DRL στον εντοπισμό εισβολών σε περιβάλλον IIoT [119]. Συγκεκριμένα, χρησιμοποιήθηκε DRL με τον αλγόριθμο Proximal Policy Optimization 2 (PPO2), έναν policy gradient αλγόριθμο που επιτρέπει πιο γρήγορη και σταθερή σύγκλιση σε μεγαλύτερα και περίπλοκα σύνολα δεδομένων. Πριν την εκπαίδευση του agent, έγινε επιλογή χαρακτηριστικών με τον αλγόριθμο LightGBM, με σκοπό τη μείωση της διάστασης των δεδομένων και της υπολογιστικής επιβάρυνσης. Από τα 26 χαρακτηριστικά του αρχικού συνόλου, διατηρήθηκαν τα 12 σημαντικότερα, γεγονός που μείωσε την υπολογιστική πολυπλοκότητα χωρίς απώλεια ακρίβειας. Το τελικό μοντέλο πέτυχε ακρίβεια 99,09%, ξεπερνώντας άλλα μοντέλα όπως LSTM και CNN σε όλες τις μετρικές απόδοσης. Οι τιμές precision, recall και F1 ξεπέρασαν το 97% σε όλα τα είδη επιθέσεων. Η σύγκλιση ήταν γρήγορη και σταθερή και το σύστημα αποδείχτηκε κατάλληλο για εντοπισμό σε πραγματικό χρόνο, καθώς ο χρόνος της εκπαίδευσης παρέμεινε χαμηλός. Ωστόσο, η έρευνα δεν αξιολογεί την κατανάλωση πόρων, όπως τη χρήση μνήμης ή ενέργειας, ούτε εξετάζει την αποτελεσματικότητα του μοντέλου σε πραγματικές IoT συσκευές με περιορισμένους υπολογιστικούς πόρους. Τέλος, αναδεικνύεται η σημασία της επιλογής χαρακτηριστικών πριν την εκπαίδευση, με σκοπό την επίτευξη της μέγιστης απόδοσης.

Τα πλεονεκτήματα του DRL περιλαμβάνουν την υψηλή ακρίβεια στον εντοπισμό εισβολών και την αναγνώριση πολύπλοκων μοτίβων, λόγω της χρήσης νευρωνικών δικτύων. Δεν απαιτούνται επισημασμένα δεδομένα, καθώς το μοντέλο μαθαίνει από την εμπειρία και όχι από labels, γεγονός που επιτρέπει τον εντοπισμό άγνωστων επιθέσεων [117]. Τέλος, είναι ιδανικό για χρήση σε δυναμικά περιβάλλοντα, όπως το IoT, καθώς προσαρμόζεται όσο αλλάζουν οι συνθήκες του δικτύου. Η εκπαίδευσή του είναι συνεχής και η απόδοσή του βελτιώνεται όσο επεξεργάζεται νέα δεδομένα [111].

Ωστόσο, το DRL είναι υπολογιστικά απαιτητικό, λόγω της επαναλαμβανόμενης διαδικασίας εκπαίδευσης και της χρήσης νευρωνικού δικτύου σε κάθε βήμα απόφασης. Η επιλογή των κατάλληλων παραμέτρων που καθορίζουν την ικανότητα μάθησης του συστήματος, όπως το learning rate και ο ορισμός της συνάρτησης ανταμοιβής, αποτελεί σημαντική πρόκληση. Η ερμηνεία των αποτελεσμάτων του μπορεί να είναι δύσκολη, καθώς λειτουργούν ως συστήματα black box. Τέλος, η διαδικασία προεπεξεργασίας των δεδομένων και η επιλογή των σημαντικών χαρακτηριστικών επηρεάζει σημαντικά την απόδοση του μοντέλου [117].

8.5.3 DRL στην κατανομή πόρων στο IoT

Το DRL μπορεί να χρησιμοποιηθεί για δυναμική λήψη αποφάσεων σχετικά με την κατανομή πόρων σε πραγματικό χρόνο σε περιβάλλον IoT. Ένας agent εκπαιδεύεται ώστε, σε κάθε βήμα εκπαίδευσης, να λαμβάνει ως είσοδο πληροφορίες για την κατάσταση του συστήματος (π.χ. χρήση CPU, διαθέσιμο bandwidth, καθυστερήσεις). Με βάση αυτά τα δεδομένα, επιλέγει ενέργειες που έχουν στόχο τη βελτιστοποίηση της συνολικής απόδοσης, όπως κατανομή υπολογιστικής ισχύος ή εκχώρηση προτεραιότητας σε πακέτα. Ο agent λαμβάνει ανταμοιβές ανάλογα με το αν οι αποφάσεις του οδηγούν σε μείωση καθυστερήσεων, αύξηση ποσοστού επιτυχημένων εργασιών ή περιορισμό της κατανάλωσης ενέργειας. Μέσα από συνεχή αλληλεπίδραση με το περιβάλλον, το μοντέλο βελτιώνει σταδιακά τη στρατηγική του και προσαρμόζεται σε νέες συνθήκες λειτουργίας [118].

Η εφαρμογή του DRL στην κατανομή πόρων σε περιβάλλον IoT έχει μελετηθεί σε πρόσφατη έρευνα [120]. Εξετάζεται το κατά πόσο ένας DRL agent μπορεί να μάθει να διαχειρίζεται δυναμικά τους υπολογιστικούς πόρους ενός edge IoT συστήματος. Οι συγγραφείς προτείνουν μία παραλλαγή του DRL, η οποία χρησιμοποιεί ξεχωριστές μνήμες εμπειριών για κάθε συσκευή ή τύπο εργασίας (task), ώστε οι αλληλεπιδράσεις διαφορετικών συσκευών να μην επηρεάζουν την ποιότητα της εκπαίδευσης. Ο agent

λαμβάνει αποφάσεις σχετικά με την επιλογή edge server, την ανάθεση ή αποδέσμευση CPU cores και την κατανομή του χρόνου εξυπηρέτησης των tasks. Τα πειράματα πραγματοποιούνται σε προσομοιωμένο περιβάλλον με 20 tasks και 5 edge servers. Η προτεινόμενη μέθοδος συγκρίνεται με δύο παραδοσιακές μεθόδους: μια τυχαία και μία βασισμένη σε FCFS. Τα αποτελέσματα δείχνουν σημαντική βελτίωση στη χρήση πόρων (83,3% σε σχέση με το 50,8% του FCFS), στον συνολικό χρόνο εξυπηρέτησης (βελτίωση περίπου 25%) και στην τιμή του reward, αποδεικνύοντας ότι η στρατηγική του agent βελτίωσε την αποδοτικότητα του συστήματος. Αποδεικνύεται επίσης ότι ο DRL agent διατηρεί ομαλή συμπεριφορά και χαμηλή μέση καθυστέρηση, ακόμη κι όταν αυξάνεται το φορτίο. Ωστόσο, δεν αξιολογείται η υπολογιστική επιβάρυνση του μοντέλου και υπάρχει η προϋπόθεση ότι οι παράμετροι των tasks είναι γνωστοί εκ των προτέρων, κάτι που δεν είναι πάντα εφικτό σε δίκτυα IoT.

Το DRL υποστηρίζει την αυτόνομη λήψη αποφάσεων και την κατανομή πόρων, χωρίς να βασίζεται σε κανόνες και χωρίς να απαιτεί ανθρώπινη παρέμβαση. Μπορεί να προσαρμόζεται σε μεταβαλλόμενες συνθήκες φορτίου και αριθμού εργασιών, βελτιστοποιώντας ταυτόχρονα πολλαπλά κριτήρια. Δεν απαιτεί επισημασμένα δεδομένα για να λειτουργήσει, κάτι που το καθιστά πιο ανθεκτικό, ενώ μπορεί να εφαρμοστεί σε μεγάλα, πολύπλοκα και ετερογενή δίκτυα IoT, όπου δεν υπάρχει γνώση του ιδανικού τρόπου κατανομής. Όσο περισσότερη εμπειρία αποκτά ο agent, τόσο πιο αποδοτικές γίνονται οι αποφάσεις του [118].

Στα μειονεκτήματά του είναι το υψηλό υπολογιστικό κόστος, που απαιτείται τόσο κατά την εκπαίδευσή του όσο και για τη λειτουργία του. Η εκπαίδευσή του απαιτεί μεγάλη ποσότητα δεδομένων για να πετύχει καλή απόδοση, με αποτέλεσμα η συμπεριφορά του agent να είναι ασταθής στα αρχικά στάδια. Είναι ευαίσθητο στον ορισμό της συνάρτησης ανταμοιβής, αλλά και στην επιλογή των υπερπαραμέτρων που επηρεάζουν την τελική απόδοση. Τέλος, η ερμηνεία των αποφάσεών του μπορεί να είναι δύσκολη [116][118].

8.6 Federated Learning

Το Federated Learning (FL) είναι μία τεχνική μηχανικής μάθησης, η οποία επιτρέπει σε πολλαπλές συσκευές ή κόμβους ενός δικτύου να συνεργάζονται για την εκπαίδευση ενός μοντέλου, χωρίς να ανταλλάσσουν ακατέργαστα δεδομένα [121]. Κάθε συσκευή εκπαιδεύει τοπικά ένα μοντέλο χρησιμοποιώντας τα δικά της δεδομένα και αποστέλλει μόνο ενημερώσεις του μοντέλου (model updates) σε έναν κεντρικό server. Ο server συγκεντρώνει τις ενημερώσεις από όλους τους κόμβους και δημιουργεί ένα βελτιωμένο παγκόσμιο μοντέλο (global model), το οποίο επιστρέφει στους κόμβους.

Το FL ενισχύει την ασφάλεια των συστημάτων, καθώς τα δεδομένα δεν εγκαταλείπουν τις συσκευές, ενώ ταυτόχρονα μειώνει το κόστος επικοινωνίας περιορίζοντας τις απαιτήσεις μεταφοράς δεδομένων. Λόγω της αποκεντρωμένης εκπαίδευσής του, το FL είναι κατάλληλο για κατανεμημένα συστήματα όπως το IoT, όπου η ασφάλεια και οι περιορισμοί σε υπολογιστική ισχύ και bandwidth παίζουν καθοριστικό ρόλο [122].

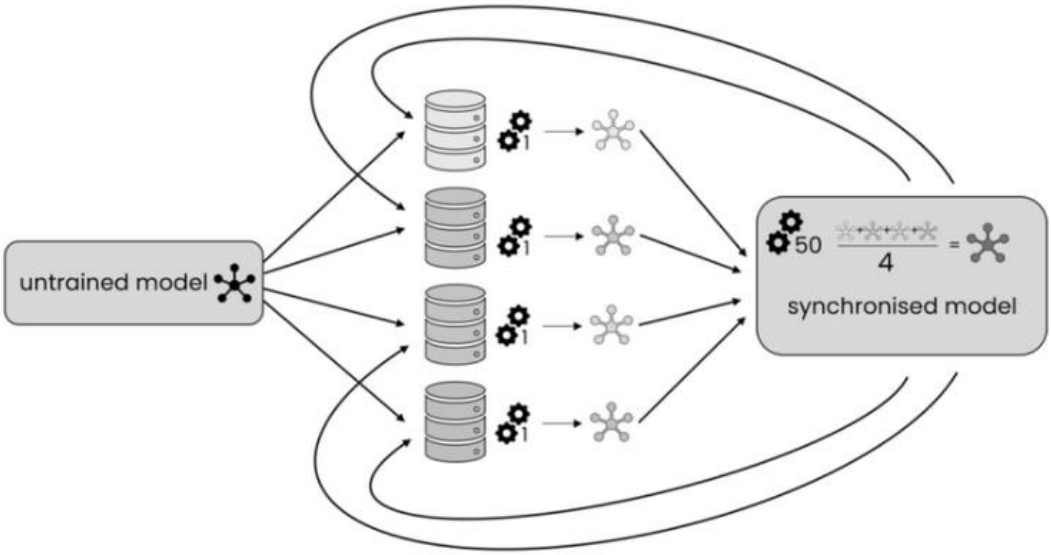
8.6.1 Βασικές αρχές και λειτουργία του FL

Η διαδικασία εκπαίδευσης περιλαμβάνει πέντε βήματα [121] [123]:

1. **Αρχικοποίηση και διαμοιρασμός του μοντέλου.** Ένας κεντρικός server δημιουργεί ένα αρχικό μοντέλο μηχανικής μάθησης, συνήθως Multi-Layer Perceptron (MLP) ή LSTM. Το MLP είναι η πιο συχνή επιλογή, λόγω του χαμηλού υπολογιστικού κόστους, ενώ το LSTM προσφέρει καλύτερη απόδοση σε ακολουθιακά δεδομένα. Το μοντέλο αρχικοποιείται με τυχαίες τιμές στις παραμέτρους του (weights και biases) και διανέμεται σε ένα υποσύνολο των διαθέσιμων

συσκευών. Η επιλογή τους γίνεται είτε τυχαία, είτε βάσει διαθεσιμότητας (συσκευές που είναι ενεργές κατά τον συγκεκριμένο γύρο εκπαίδευσης), είτε υπό προϋποθέσεις (όπως διαθεσιμότητα CPU, επάρκεια ενέργειας, σταθερότητα σύνδεσης και χαμηλή καθυστέρηση επικοινωνίας).

2. **Τοπική εκπαίδευση.** Κάθε συσκευή εκπαιδεύει το μοντέλο τοπικά, χρησιμοποιώντας μόνο τα δικά της δεδομένα (όπως καταγραφές δραστηριότητας δικτύου ή δεδομένα αισθητήρων). Η εκπαίδευση πραγματοποιείται για έναν προκαθορισμένο αριθμό εποχών (local epochs) και ενημερώνει τα weights και τα biases του μοντέλου.
3. **Αποστολή ενημερώσεων στον κεντρικό server.** Μετά την τοπική εκπαίδευση, κάθε συσκευή αποστέλλει μόνο τις ενημερώσεις του μοντέλου, όπως τις μεταβολές στα weights και τα biases. Τα ακατέργαστα δεδομένα παραμένουν τοπικά και δεν μεταφέρονται.
4. **Συγχώνευση ενημερώσεων.** Ο server συγκεντρώνει τις ενημερώσεις από τις συσκευές και δημιουργεί ένα νέο, βελτιωμένο global model. Συνήθως εφαρμόζεται ο αλγόριθμος Federated Averaging (FedAvg), ο οποίος υπολογίζει το μέσο όρο των τιμών των βαρών από όλα τα τοπικά μοντέλα. Ο server δεν συμμετέχει στην εκπαίδευση, αλλά εκτελεί μόνο τη διαδικασία συγχώνευσης των παραμέτρων.
5. **Αναδιανομή του βελτιωμένου μοντέλου.** Ο server αποστέλλει το global model πίσω στις ίδιες ή σε νέες συσκευές. Κάθε συσκευή αντικαθιστά το προηγούμενο μοντέλο με το νέο και ξεκινά νέο κύκλο τοπικής εκπαίδευσης.
6. **Τερματισμός εκπαίδευσης.** Η διαδικασία εκπαίδευσης και συγχώνευσης επαναλαμβάνεται για πολλαπλούς γύρους και τερματίζεται όταν το global model συγκλίνει, δηλαδή όταν οι αλλαγές στις παραμέτρους από γύρο σε γύρο είναι αμελητέες, ή όταν επιτευχθεί ένα προκαθορισμένο επίπεδο απόδοσης. Στη συνέχεια, το τελικό μοντέλο μπορεί να χρησιμοποιηθεί για την ανίχνευση εισβολών στο σύστημα.



Σχήμα 8.3: Εκπαίδευση του FL [123]

8.6.2 FL στον εντοπισμό εισβολών στο IoT

Το FL επιτρέπει στις συσκευές IoT να εντοπίσουν κακόβουλες δραστηριότητες τοπικά, χωρίς να απαιτείται μεταφορά δεδομένων σε κεντρικό server. Κάθε συσκευή παρακολουθεί σε πραγματικό χρόνο τη δραστηριότητα του δικτύου της και καταγράφει χαρακτηριστικά, εκπαιδεύοντας τοπικά ένα μοντέλο με δεδομένα από το δικό της περιβάλλον. Καθώς το μοντέλο εκπαιδεύεται, μαθαίνει να διαχωρίζει τη φυσιολογική κίνηση από την κακόβουλη, εντοπίζοντας μοτίβα στη συμπεριφορά του δικτύου. Οι ενημερώσεις του μοντέλου αποστέλλονται στον κεντρικό server, ο οποίος δημιουργεί ένα global model, ενσωματώνοντας τη γνώση από όλες τις συσκευές. Το global model διανέμεται στις συσκευές και τους

επιτρέπει να ανιχνεύουν επιθέσεις σε όλο το δίκτυο IoT, αξιοποιώντας τη συλλογική εμπειρία του συστήματος. Η διαδικασία επαναλαμβάνεται, βελτιώνοντας συνεχώς την ακρίβεια στην αναγνώριση νέων ή εξελιγμένων απειλών, με τις συσκευές IoT να λειτουργούν ανεξάρτητα [124].

Μία από τις πρώτες εφαρμογές του FL στον εντοπισμό εισβολών σε IoT συσκευές παρουσιάζεται σε μελέτη που παρουσιάζει το σύστημα DÏoT [125]. Το μοντέλο εκπαιδεύεται τοπικά στις συσκευές και μαθαίνει αυτόματα τη φυσιολογική συμπεριφορά κάθε τύπου, παρακολουθώντας τη δικτυακή δραστηριότητα και αναγνωρίζοντας αποκλίσεις που μπορεί να υποδηλώνουν εισβολές. Η μέθοδος αξιολογήθηκε σε πραγματικές συνθήκες με 33 συσκευές IoT και παρουσίασε ακρίβεια πάνω από 95% στην ανίχνευση επιθέσεων όπως spoofing, port scanning και DoS, χωρίς να καταγραφούν false positives. Ωστόσο, η μελέτη υποθέτει ότι τα αρχικά δεδομένα είναι καθαρά και δεν αξιολογεί την ύπαρξη επιθέσεων data poisoning. Σε άλλη εργασία, ερευνάται η ευπάθεια του FL σε επιθέσεις data poisoning [126]. Αποδεικνύεται ότι ένας επιτιθέμενος μπορεί να αποκτήσει πρόσβαση σε απλές IoT συσκευές και να εισάγει μικρές ποσότητες κακόβουλης κυκλοφορίας στο τοπικό σύνολο δεδομένων εκπαίδευσης. Καθώς η κυκλοφορία αυτή δεν ανιχνεύεται κακόβουλη, ενσωματώνεται στο μοντέλο και οδηγεί σταδιακά το global model να θεωρεί φυσιολογική την κακόβουλη συμπεριφορά. Σε πείραμα σε 46 συσκευές IoT (όπως Amazon Echo, IP κάμερες και έξυπνες πρίζες), διαπιστώθηκε ότι ακόμη και όταν μόνο το 20% των clients συμμετείχε στην επίθεση και μόνο το 35% τοπικών δεδομένων ήταν μολυσμένα, η επίθεση ήταν 100% πετυχημένη. Η μελέτη τονίζει την ανάγκη ανάπτυξης νέων τεχνικών άμυνας, καθώς οι υπάρχουσες δεν επαρκούν για την αντιμετώπιση τέτοιων στοχευμένων επιθέσεων.

Ένα σημαντικό πλεονέκτημα του FL είναι η διατήρηση της ιδιωτικότητας [121]. Λόγω του ότι τα δεδομένα παραμένουν εντός των συσκευών, μειώνεται σημαντικά ο κίνδυνος διαρροής και υποκλοπής κατά τη μετάδοση. Παρέχει βελτιωμένη απόδοση στην ανίχνευση εισβολών, καθώς το global model εκπαιδεύεται με δεδομένα που προέρχονται από διαφορετικές συσκευές, διευκολύνοντας τον εντοπισμό ασυνήθιστων μοτίβων και άγνωστων απειλών [124]. Η ανθεκτικότητα του συστήματος είναι αυξημένη, καθώς δεν υπάρχει ένα κεντρικό σημείο αποτυχίας. Η ανίχνευση μπορεί να γίνει σε πραγματικό χρόνο, ακόμη και σε συσκευές με περιορισμένους υπολογιστικούς πόρους. Τέλος, το FL είναι επεκτάσιμο, καθώς δεν απαιτείται συγκέντρωση μεγάλου όγκου δεδομένων σε έναν κεντρικό server.

Ωστόσο, οι υπολογιστικές απαιτήσεις που σχετίζονται με την τοπική εκπαίδευση είναι αυξημένες και ορισμένες συσκευές IoT ενδέχεται να μην μπορούν να συμμετέχουν λόγω των περιορισμένων πόρων τους [124]. Οι συνεχείς μεταδόσεις ενημερώσεων μεταξύ server και συσκευών ενδέχεται να αυξήσουν τη χρήση bandwidth. Επιπλέον, το FL έχει ευαισθησία σε επιθέσεις τύπου data poisoning, όπου ένας μολυσμένος client μπορεί να στείλει αλλοιωμένες παραμέτρους και να παραποιήσει το global model, ώστε να μην αναγνωρίζει συγκεκριμένες απειλές [126]. Τέλος, η απόδοση του μοντέλου μπορεί να μειωθεί σε ετερογενή περιβάλλοντα, καθώς τα δεδομένα των συσκευών είναι ανομοιόμορφα.

8.6.3 FL στην κατανομή πόρων στο IoT

Το FL μπορεί να συμβάλλει στη βελτιστοποίηση της κατανομής των πόρων σε συστήματα IoT, επιτρέποντας στις συσκευές να λαμβάνουν πιο αποδοτικές αποφάσεις, χωρίς να απαιτείται ανταλλαγή ευαίσθητων δεδομένων. Κάθε συσκευή εκπαιδεύει τοπικά το μοντέλο με βάση τα δικά της δεδομένα, όπως η χρήση επεξεργαστικής ισχύος, η κατανάλωση ενέργειας και το bandwidth. Μετά την τοπική εκπαίδευση, αποστέλλει ενημερώσεις στον server και το global model που επιστρέφεται μπορεί να αξιοποιηθεί για λήψη αποφάσεων σε πραγματικό χρόνο, όπως η εκχώρηση πόρων σε εργασίες, η εκχώρηση προτεραιότητας σε μεταδόσεις και η εξισορρόπηση φορτίου μεταξύ συσκευών [127]. Με αυτό τον τρόπο, η κατανομή προσαρμόζεται δυναμικά στις μεταβαλλόμενες συνθήκες λειτουργίας του

συστήματος, χωρίς συγκέντρωση δεδομένων σε κεντρικό κόμβο και λαμβάνοντας υπόψη τους περιορισμούς των IoT συσκευών.

Η εφαρμογή του FL στην κατανομή πόρων έχει μελετηθεί σε έρευνα [128], με στόχο τη μείωση της συνολικής κατανάλωσης ενέργειας και του χρόνου ολοκλήρωσης των εργασιών. Κάθε συσκευή εκπαιδεύει τοπικά ένα CNN, με τη loss function προσαρμοσμένη στους διαθέσιμους πόρους της, όπως η χρήση CPU και η κατανάλωση ενέργειας. Οι ενημερώσεις αποστέλλονται στον server, ο οποίος δεν εφαρμόζει απλό μέσο όρο όπως το FedAvg, αλλά υπολογίζει ξεχωριστό συντελεστή βάρους (weighted aggregation) για κάθε συσκευή, με βάση την ενεργειακή και υπολογιστική της αποδοτικότητα. Με αυτό τον τρόπο, ο αλγόριθμος αποφασίζει ποιο ποσοστό υπολογιστικού έργου αναλαμβάνει κάθε κόμβος, με βάση τη δυνατότητα του να το εκτελεί εντός χρονικού ορίου και χωρίς υπερβολική ενεργειακή κατανάλωση. Οι συσκευές με περιορισμένους πόρους ή υψηλή κατανάλωση λαμβάνουν μικρότερη συμμετοχή στη διαμόρφωση του global model, επιτρέποντας στην κατανομή να προσαρμόζεται στη φυσική κατάσταση κάθε συσκευής. Σε πειράματα με σταθερό αριθμό clients, το μοντέλο μείωσε την κατανάλωση ενέργειας κατά 18–22% και χρειάστηκε 25–30% λιγότερους γύρους για να σταθεροποιηθεί η απόδοση του συστήματος, σε σύγκριση με τη χρήση FedAvg. Η χρήση weighted aggregation επιτρέπει πιο ομοιόμορφη κατανομή υπολογιστικού φορτίου, βελτιώνοντας τη συνολική αξιοποίηση των πόρων, ακόμη και όταν οι συσκευές έχουν ανομοιόμορφα δεδομένα. Ωστόσο, η μελέτη δεν εξετάζει περιπτώσεις δυναμικής συμμετοχής ή αποτυχίας κόμβων, ούτε αξιολογεί παράγοντες επικοινωνίας, όπως καθυστερήσεις ή απώλειες.

Το FL παρουσιάζει σημαντικά πλεονεκτήματα στην κατανομή πόρων σε περιβάλλοντα IoT. Επιτρέπει αποκεντρωμένη λήψη αποφάσεων και προσαρμοστικότητα στις τοπικές συνθήκες και τους πόρους της κάθε συσκευής. Το global model αποτυπώνει πρότυπα κατανομής που οδηγούν σε βελτιστοποιημένη χρήση πόρων και αύξηση της απόδοσης του συστήματος [122]. Η ιδιωτικότητα διατηρείται, καθώς δεν ανταλλάσσονται ακατέργαστα δεδομένα.

Ωστόσο, το επικοινωνιακό κόστος και η χρήση bandwidth μπορεί να παραμείνουν υψηλά, λόγω των συχνών μεταδόσεων και των απαιτήσεων συντονισμού μεταξύ των συσκευών. Οι υπολογιστικές απαιτήσεις είναι επίσης αυξημένες, καθώς οι συσκευές πρέπει να εκτελέσουν τοπική εκπαίδευση [123]. Κάποιες συσκευές ενδέχεται να αποκλείονται από την εκπαίδευση λόγω των χαμηλών πόρων, επηρεάζοντας τη συνοχή του global model. Η ετερογένεια των συσκευών μπορεί να επηρεάσει την ικανότητα γενίκευσης. Τέλος, το FL δεν εγγυάται ότι το global model θα οδηγεί πάντα στη βέλτιστη κατανομή για κάθε συσκευή, καθώς η κάθε μία μπορεί να έχει διαφορετικές απαιτήσεις και δυνατότητες [127].

8.7 Federated Reinforcement Learning

Το Federated Reinforcement Learning (FRL) είναι μια προσέγγιση μηχανικής μάθησης που συνδυάζει τις βασικές αρχές του RL και του FL. Πολλαπλοί agents εκπαιδεύονται τοπικά στο δικό τους περιβάλλον, χρησιμοποιώντας αλγορίθμους RL και δεδομένα που δεν κοινοποιούνται. Μετά από κάθε γύρο εκπαίδευσης, αποστέλλουν τις ενημερώσεις των μοντέλων (όπως πίνακες Q ή βάρη νευρωνικών δικτύων) σε έναν κεντρικό server, ο οποίος επιστρέφει ένα global model που ενσωματώνει τη συλλογική εμπειρία όλων των agents [129].

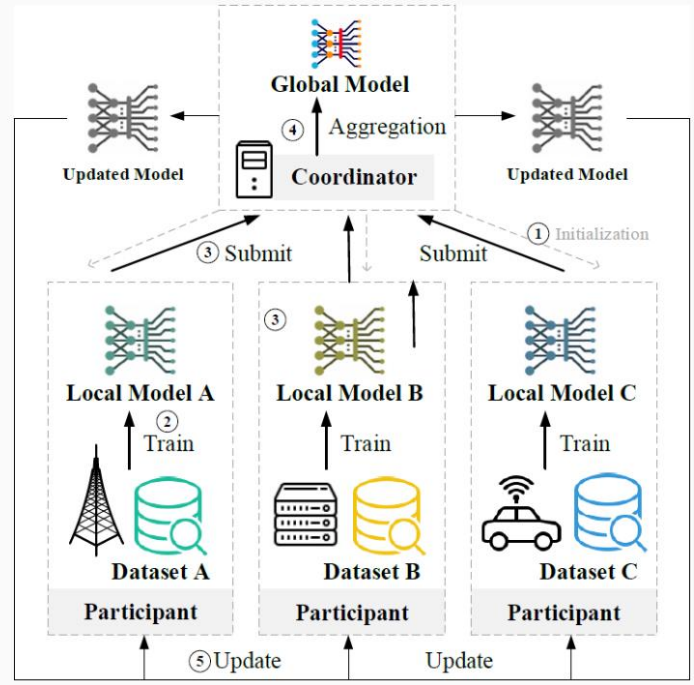
Το FRL είναι κατάλληλο τόσο για τον εντοπισμό εισβολών όσο και για τη δυναμική κατανομή πόρων σε περιβάλλοντα IoT. Ενισχύει την ιδιωτικότητα του συστήματος και βελτιώνει την ακρίβεια και τη γενίκευση των αποφάσεων, αξιοποιώντας τη συλλογική γνώση που αποκτούν οι agents από διαφορετικά περιβάλλοντα [130].

8.7.1 Βασικές Αρχές και Λειτουργία του FRL

Η λειτουργία του FRL ξεκινά με την αρχικοποίηση ενός μοντέλου, συνήθως MLP, λόγω του χαμηλού υπολογιστικού κόστους και της απλότητας στην υλοποίηση, που το καθιστούν ιδανικό για συσκευές με περιορισμένους πόρους. Σε άλλες υλοποιήσεις, μπορεί να χρησιμοποιηθεί CNN, RNN ή LSTM, ανάλογα με τις απαιτήσεις της εφαρμογής [129].

Το μοντέλο διαμοιράζεται σε όλους τους συμμετέχοντες agents, οι οποίοι μπορεί να είναι έξυπνες συσκευές (όπως θερμοστάτες και αισθητήρες) ή edge συσκευές (όπως routers και gateways). Κάθε agent αλληλεπιδρά με το δικό του περιβάλλον και συλλέγει εμπειρίες, που περιλαμβάνουν την τρέχουσα κατάσταση, την επιλεγμένη ενέργεια, την ανταμοιβή και τη νέα κατάσταση που προκύπτει μετά την εκτέλεση της ενέργειας. Η τοπική εκπαίδευση πραγματοποιείται με αλγορίθμους RL, όπως Q-learning ή SARSA. Σε πιο απαιτητικές εφαρμογές, μπορεί να χρησιμοποιούνται μέθοδοι DRL, που βασίζονται σε νευρωνικά δίκτυα.

Οι agents αποστέλλουν τις ενημερώσεις των παραμέτρων των μοντέλων τους στον κεντρικό server, ο οποίος τις συγχωνεύει μέσω του αλγορίθμου FedAvg. Προκύπτει ένα global model το οποίο αποστέλλεται ξανά στους agents. Οι agents ενημερώνουν τα τοπικά τους μοντέλα και ξεκινάει ο νέος γύρος εκπαίδευσης. Η διαδικασία επαναλαμβάνεται μέχρι η πολιτική να συγκλίνει ή να επιτευχθεί ένα προκαθορισμένο επίπεδο απόδοσης [131].



Σχήμα 8.4: Αρχιτεκτονική και λειτουργία του FRL [131]

8.7.2 FRL στον εντοπισμό εισβολών στο IoT

Το FRL μπορεί να εφαρμοστεί στον εντοπισμό εισβολών σε περιβάλλοντα IoT, με κάθε agent να εκπαιδεύει τοπικά ένα μοντέλο για την αναγνώριση κακόβουλης δραστηριότητας. Παρακολουθεί χαρακτηριστικά της δικτυακής κίνησης σε πραγματικό χρόνο και επιλέγει ενέργειες σύμφωνα με την τρέχουσα πολιτική του, δηλαδή ταξινομεί την κίνηση ως κακόβουλη ή φυσιολογική. Ανάλογα με το πόσο σωστή ήταν η ενέργειά του, λαμβάνει θετική ή αρνητική ανταμοιβή και ενημερώνει την πολιτική του με βάση τη νέα εμπειρία. Στη συνέχεια, οι agents αποστέλλουν τις τοπικά εκπαιδευμένες πολιτικές

στον κεντρικό server, ο οποίος συγχωνεύει τις παραμέτρους και δημιουργεί το global model. Αυτό διανέμεται ξανά στις συσκευές, οι οποίες το χρησιμοποιούν ως σημείο εκκίνησης για τον επόμενο γύρο εκπαίδευσης [132].

Η εφαρμογή του FRL έχει μελετηθεί σε διάφορες έρευνες, συχνά σε συνδυασμό με διάφορες τεχνικές ενίσχυσης της απόδοσης. Σε μία πρόσφατη μελέτη, προτείνεται ένα σύστημα ανίχνευσης εισβολών για συστήματα υγείας βασισμένα σε IoT, που χρησιμοποιεί FRL με Q-learning [133]. Κάθε συσκευή εκπαιδεύει τοπικά ένα μοντέλο και αποστέλλει τις ενημερώσεις στον server, ο οποίος τις συγχωνεύει με τη μέθοδο FedAvg. Το σύστημα πέτυχε ακρίβεια εντοπισμού 98,5% σε επιθέσεις όπως DDoS, PortScan και SQL Injection, μετά από 10 γύρους εκπαίδευσης. Ωστόσο, τονίζεται η σημασία της προεπεξεργασίας των δεδομένων, όπως η κανονικοποίηση και η επιλογή των κρίσιμων χαρακτηριστικών, ώστε να εξασφαλιστεί η σταθερότητα της μάθησης και η αποδοτικότητα του Q-learning αλγορίθμου.

Σε άλλη μελέτη, παρουσιάζεται ένα σύστημα ανίχνευσης εισβολών που βασίζεται σε Deep FRL, χρησιμοποιώντας νευρωνικό δίκτυο (Deep Q-Network – DQN) και μηχανισμό δυναμικής προσοχής (attention mechanism) [134]. Κάθε agent εκπαιδεύει ένα τοπικό μοντέλο, το οποίο αποστέλλει στον server. Η διαδικασία συγχώνευσης των μοντέλων γίνεται μέσω dynamic weighted aggregation, όπου κάθε agent αποκτά διαφορετική βαρύτητα, ανάλογα με την επίδοση του global model στα τοπικά δεδομένα. Η προσέγγιση πέτυχε ακρίβεια 98% και χαμηλό ποσοστό false positives. Η απόδοση παρέμεινε σταθερή όσο αυξάνονταν οι agents, επιβεβαιώνοντας την επεκτασιμότητα του συστήματος. Η μελέτη επισημαίνει ότι ο δυναμικός μηχανισμός προσοχής ενίσχυσε σημαντικά την απόδοση και ανθεκτικότητα του συστήματος, ενώ επέτρεψε καλύτερη προσαρμογή σε ετερογενή δεδομένα. Ωστόσο, η έρευνα επικεντρώνεται σε offline εκπαίδευση και όχι σε real-time εντοπισμό εισβολών, ενώ δεν λαμβάνει υπόψη την καθυστέρηση επικοινωνίας. Τέλος, δεν αξιολογείται η ανθεκτικότητα σε επιθέσεις data poisoning.

Το FRL παρουσιάζει σημαντικά πλεονεκτήματα στον εντοπισμό εισβολών σε περιβάλλοντα IoT [126]. Η ιδιωτικότητα διατηρείται, καθώς τα δεδομένα παραμένουν τοπικά. Κάθε συσκευή μαθαίνει από το δικό της περιβάλλον, αποκτώντας τη δυνατότητα να προσαρμόζεται σε τοπικές απειλές. Το κόστος μεταφοράς δεδομένων μειώνεται, αφού μεταδίδονται μόνο οι ενημερώσεις του μοντέλου και αποφεύγεται το bottleneck. Τέλος, το μοντέλο έχει τη δυνατότητα να εντοπίζει άγνωστες και zero-day επιθέσεις [132].

Ωστόσο, το FRL έχει αυξημένη υπολογιστική επιβάρυνση, ειδικά όταν χρησιμοποιούνται νευρωνικά δίκτυα για την τοπική εκπαίδευση (deep FRL). Μπορεί να γίνει ασταθές όταν τα δεδομένα και τα περιβάλλοντα των συσκευών είναι ετερογενή, οδηγώντας σε πιο αργή σύγκλιση. Απαιτείται μεγάλος αριθμός δεδομένων για την εκπαίδευση, με αποτέλεσμα το μοντέλο να ενδέχεται να είναι ασταθές στους πρώτους γύρους εκπαίδευσης. Τέλος, το FRL είναι ευάλωτο σε επιθέσεις τύπου data poisoning [131] [132].

8.7.3 FRL στην κατανομή πόρων στο IoT

Το FRL μπορεί να συμβάλλει στη βέλτιστη κατανομή πόρων σε δίκτυα IoT, με στόχους τη μείωση της ενεργειακής κατανάλωσης, τη μείωση της καθυστέρησης (ειδικά σε real-time εφαρμογές) και τη βελτιστοποίηση του throughput του δικτύου. Σε κάθε χρονική στιγμή, ο κάθε agent παρακολουθεί την κατάσταση των τοπικών πόρων του. Εκτελεί ενέργειες που επηρεάζουν την κατανομή πόρων και λαμβάνει ανταμοιβή: θετική όταν επιτυγχάνει χαμηλή καθυστέρηση ή αποδοτική χρήση ενέργειας και αρνητική για υπερκατανάλωση ή χαμηλή απόδοση. Η πολιτική του ενημερώνεται τοπικά, βάσει της

αλληλεπίδρασης με το περιβάλλον. Στη συνέχεια, οι ενημερώσεις του μοντέλου αποστέλλονται στον server, ο οποίος συγχωνεύει τα μοντέλα από όλες τις συσκευές και επιστρέφει ένα global model που θα χρησιμοποιηθεί για τον επόμενο γύρο εκπαίδευσης [126].

Παρά τη θεωρητική δυνατότητα εφαρμογής του FRL σε συστήματα IoT, οι μελέτες που εστιάζουν αποκλειστικά στη δυναμική κατανομή πόρων, όπως CPU, bandwidth ή μνήμη, σε επίπεδο συσκευών ή edge servers, είναι περιορισμένες. Οι περισσότερες έρευνες επικεντρώνονται σε task offloading (λήψη αποφάσεων σχετικά με το αν μια εργασία θα εκτελεστεί τοπικά ή θα μεταφερθεί σε άλλο κόμβο) [135] ή σε task scheduling (χρονοπρογραμματισμός εκτέλεσης εργασιών βάσει πόρων και προτεραιοτήτων) [136]. Επιπλέον, έχουν προταθεί προσεγγίσεις όπως το Concurrent Federated Reinforcement Learning (CFRL), το οποίο επιτρέπει την εκπαίδευση των agents χωρίς άμεση ανταλλαγή παραμέτρων, μειώνοντας έτσι το επικοινωνιακό κόστος [137].

Το FRL προσφέρει σημαντικά πλεονεκτήματα στην κατανομή πόρων σε περιβάλλοντα IoT [126]. Η αποκεντρωμένη λήψη αποφάσεων επιτρέπει σε κάθε agent να προσαρμόζεται στις τοπικές του συνθήκες και να βελτιστοποιεί τη χρήση των διαθέσιμων πόρων, ενώ η ενσωμάτωση των τοπικών πολιτικών στο global model βελτιστοποιεί την κατανομή σε όλο το δίκτυο. Το σύστημα μπορεί να προσαρμόζεται σε πραγματικό χρόνο, εντοπίζοντας αλλαγές στη δικτυακή κίνηση και τις απαιτήσεις. Μπορεί να εφαρμοστεί σε μεγάλα, δυναμικά περιβάλλοντα καθώς έχει επεκτασιμότητα [131].

Ωστόσο, ένα βασικό μειονέκτημα του FRL είναι η υψηλή υπολογιστική πολυπλοκότητα. Η τοπική εκπαίδευση μπορεί να είναι απαιτητική για κάποιες συσκευές IoT, ενώ ο χρόνος που απαιτείται για τη σύγκλιση σε μία αποδοτική πολιτική είναι σχετικά μεγάλος, ειδικά σε μεγάλα και μεταβαλλόμενα περιβάλλοντα. Η ετερογένεια μεταξύ των συσκευών επηρεάζει αρνητικά την ποιότητα του global model, καθώς οι συσκευές έχουν διαφορετικές δυνατότητες και ανάγκες. Τέλος, η αποστολή και συγχώνευση των μοντέλων απαιτεί συχνή επικοινωνία, επηρεάζοντας την κατανάλωση του bandwidth, ακόμη κι αν ανταλλάσσονται μόνο οι ενημερώσεις των μοντέλων.

8.8 Σύγκριση τεχνικών εντοπισμού εισβολών

Οι τεχνητές που παρουσιάστηκαν αποτελούν προηγμένες προσεγγίσεις στην ανίχνευση εισβολών, ιδιαίτερα κατάλληλες για δυναμικά και ετερογενή περιβάλλοντα όπως το συστήματα IoT. Για λόγους σύγκρισης, ο Πίνακας 8.1 συγκεντρώνει τα βασικά χαρακτηριστικά κάθε τεχνικής, εστιάζοντας σε συγκεκριμένες μετρικές.

Πίνακας 8.1: Σύγκριση τεχνικών για εντοπισμό εισβολών

Μοντέλο	**Υπολογιστικό κόστος**	**Χρόνος σύγκλισης**	**Ευελιξία σε δυναμικά δίκτυα**	**Δυσκολία παραμετροποίησης**	**Πραγματική εφαρμογή**
RL	Υψηλό	Αργός	Υψηλή	Υψηλή	Οχι
DRL	Πολύ υψηλό	Αργός	Πολύ υψηλή	Πολύ υψηλή	Ναι
FL	Υψηλό	Μέτριος	Υψηλή	Υψηλή	Ναι
FRL	Πολύ υψηλό	Αργός	Πολύ υψηλή	Πολύ υψηλή	Ναι

8.9 Σύγκριση τεχνικών κατανομής πόρων

Οι τεχνικές RL, DRL, FL και FRL έχουν αξιοποιηθεί και για την επίλυση προβλημάτων κατανομής πόρων σε δίκτυα IoT, που αξιοποιούν την προσαρμοστικότητά τους και την ικανότητά τους να λαμβάνουν αποφάσεις αυτόνομα. Ο Πίνακας 8.2 συνοψίζει τα βασικά λειτουργικά χαρακτηριστικά τους, δίνοντας έμφαση σε παραμέτρους που αποκτούν αξία στο πλαίσιο του IoT.

Πίνακας 8.2: Σύγκριση τεχνικών για κατανομή πόρων

Μοντέλο	**Υπολογιστικό κόστος**	**Υποστήριξη αποκέντρωσης**	**Ευελιξία σε μεταβολές**	**Ανάγκη συγχρονισμού**	**Πραγματική εφαρμογή**
RL	Υψηλό	Μερικώς	Υψηλή	Όχι	Όχι
DRL	Πολύ υψηλό	Μερικώς	Πολύ υψηλή	Όχι	Ναι
FL	Υψηλό	Ναι	Υψηλή	Ναι	Σε προσομοίωση
FRL	Πολύ υψηλό	Ναι	Πολύ υψηλή	Ναι	Όχι

8.10 Επίλογος

Σε αυτό το κεφάλαιο εξετάστηκε το πρόβλημα της διαχείρισης πόρων στα IoT οικοσυστήματα και ο τρόπος που ορισμένες σύγχρονες τεχνικές τεχνητής νοημοσύνης μπορούν να συμβάλλουν στην επίλυσή του. Ορίστηκε το πρόβλημα της κατανομής πόρων, αναφέροντας τους λόγους που καθιστούν δύσκολη την εφαρμογή στατικών τεχνικών σε δυναμικά και ετερογενή περιβάλλοντα. Παρουσιάστηκε πώς η τεχνητή νοημοσύνη μπορεί να αποτελέσει λύση για δυναμική λήψη αποφάσεων τόσο στην κατανομή πόρων, όσο και στον εντοπισμό εισβολών σε περιβάλλοντα IoT. Αναλύθηκαν μοντέλα που μπορούν να εφαρμοστούν και στους δύο τομείς, με έμφαση στον τρόπο λειτουργίας τους, την εφαρμογή τους, τα πλεονεκτήματα και τα μειονεκτήματά τους. Η δυνατότητα αξιοποίησής τους σε πολλαπλές λειτουργίες των IoT συστημάτων αναδεικνύει την ευελιξία και τη δυναμική της Τεχνητής Νοημοσύνης στο συγκεκριμένο πεδίο.

Κεφάλαιο 9ο Συμπεράσματα και μελλοντικές κατευθύνσεις

Η παρούσα εργασία μελέτησε τη χρήση τεχνικών τεχνητής νοημοσύνης για την ενίσχυση της ασφάλειας και της αποδοτικότητας σε περιβάλλοντα IoT, με έμφαση στην ανίχνευση εισβολών και την κατανομή πόρων. Αναλύθηκαν οι βασικές αρχές, τα πλεονεκτήματα και οι περιορισμοί κάθε κατηγορίας αλγορίθμων, εστιάζοντας στη μηχανική μάθηση, τη βαθιά μάθηση και τις αποκεντρωμένες τεχνικές ενισχυτικής μάθησης. Μέσα από τη σύγκρισή τους, αναδεικνύεται η σημασία της επιλογής της κατάλληλης τεχνικής, ανάλογα με το περιβάλλον λειτουργίας, τις απαιτήσεις ιδιωτικότητας και την υπολογιστική δυνατότητα των συσκευών.

Στην ανίχνευση εισβολών, τα μοντέλα βαθιάς μάθησης, και κυρίως τα LSTM και οι Denoising Autoencoders, εμφανίζουν ιδιαίτερα υψηλή απόδοση λόγω της ικανότητάς τους να αναγνωρίζουν πολύπλοκα και χρονικά εξαρτώμενα μοτίβα. Τα LSTM ξεχωρίζουν για την ακρίβεια και τη σταθερότητά τους σε real-time εφαρμογές [93], ενώ οι Denoising Autoencoders προσφέρουν σημαντική ανθεκτικότητα σε θόρυβο και τη δυνατότητα εντοπισμού άγνωστων επιθέσεων [103]. Η τεχνική του Federated Learning προσφέρει επίσης ενίσχυση της ιδιωτικότητας και αυτονομία, κάτι που την καθιστά κατάλληλη για περιβάλλοντα IoT, αν και συνοδεύεται από προκλήσεις όπως η υπολογιστική επιβάρυνση και η ευπάθεια σε data poisoning [121].

Οι τεχνικές μηχανικής μάθησης, όπως το XGBoost και το LightGBM, παραμένουν αποτελεσματικές επιλογές για προβλήματα ταξινόμησης, κυρίως λόγω της ανθεκτικότητάς τους στο overfitting και της ικανότητάς τους να διαχειρίζονται ανισόρροπα δεδομένα [69] [73]. Παρ' όλα αυτά, η εφαρμογή τους σε πραγματικά IoT περιβάλλοντα περιορίζεται από τις αυξημένες υπολογιστικές απαιτήσεις. Από τις τεχνικές αυτές, το LightGBM ξεχωρίζει ως μια από τις πιο ισορροπημένες προσεγγίσεις, προσφέροντας ταχύτερη εκπαίδευση, ενσωματωμένη προεπεξεργασία και μειωμένο υπολογιστικό φόρτο [74].

Στον τομέα της κατανομής πόρων, τεχνικές όπως το RL και το DRL παρέχουν δυνατότητα δυναμικής προσαρμογής σε μεταβαλλόμενες συνθήκες λειτουργίας [110]. Το DRL, αξιοποιώντας νευρωνικά δίκτυα, μπορεί να διαχειρίζεται σύνθετες καταστάσεις και να λαμβάνει αποφάσεις βελτιστοποίησης χωρίς την ανάγκη επισημασμένων δεδομένων [115]. Ωστόσο, η υψηλή υπολογιστική επιβάρυνση και η πολυπλοκότητα στον σχεδιασμό της πολιτικής ανταμοιβής δυσκολεύουν την πρακτική εφαρμογή του σε συσκευές περιορισμένων πόρων [114].

Αντίστοιχα, οι τεχνικές FL και FRL παρουσιάζουν σημαντικά πλεονεκτήματα, όπως η διατήρηση της ιδιωτικότητας, η δυνατότητα αποκεντρωμένης εκπαίδευσης και η επεκτασιμότητα σε μεγάλης κλίμακας IoT δίκτυα [121]. Το FRL, συγκεκριμένα, ενσωματώνει τα πλεονεκτήματα του FL και του DRL, επιτρέποντας δυναμική κατανομή των πόρων [129]. Όμως, η ανάγκη για συγχρονισμό μεταξύ των κόμβων και η πολυπλοκότητα της τοπικής εκπαίδευσης παραμένουν κρίσιμοι παράγοντες που επηρεάζουν τη σταθερότητα και τη συνολική απόδοση του συστήματος.

Παρά τις τεχνολογικές εξελίξεις, η μετάβαση από εργαστηριακές συνθήκες σε πραγματικές εφαρμογές παραμένει περιορισμένη. Η πλειονότητα των ερευνών εστιάζει στην προώθηση συγκεκριμένων τεχνικών, παρουσιάζοντας συχνά εξιδανικευμένα αποτελέσματα σε συνθετικά ή ιδανικά datasets, χωρίς επαρκή αξιολόγηση σε πραγματικά IoT περιβάλλοντα. Κάθε μοντέλο αξιολογείται με βάση τις ειδικές απαιτήσεις του σεναρίου εφαρμογής του, με αποτέλεσμα να μην υπάρχει μια καθολικά κατάλληλη λύση για όλα τα προβλήματα.

Στη βιβλιογραφία διακρίνονται σημαντικές αδυναμίες που παραμένουν ανοιχτές. Η πιο σημαντική είναι η ανάγκη για μείωση της υπολογιστικής πολυπλοκότητας, με σκοπό οι τεχνικές να μπορούν να ενσωματωθούν σε συσκευές IoT περιορισμένων πόρων. Επιπλέον, απαιτείται περαιτέρω ενίσχυση της real-time απόκρισης και της ακρίβειας, καθώς και αύξηση της ανθεκτικότητας σε εξελιγμένες επιθέσεις, όπως οι adversarial και data poisoning. Τέλος, η ενοποίηση της ανίχνευσης εισβολών και της κατανομής πόρων δεν συναντάται στη βιβλιογραφία και ελάχιστες μελέτες προτείνουν συστήματα που να μπορούν ταυτόχρονα να αξιολογούν την ασφάλεια και να διαχειρίζονται τους διαθέσιμους πόρους με βάση τη σοβαρότητα των απειλών.

Η μελλοντική έρευνα μπορεί να στραφεί προς κατευθύνσεις που ενισχύουν την αποδοτικότητα και την προσαρμοστικότητα των συστημάτων ασφάλειας στο IoT. Η ανάγκη για ανάλυση σε πραγματικό χρόνο και επεκτασιμότητα είναι έντονη και είναι σημαντικό να διερευνηθούν νέες υπολογιστικές προσεγγίσεις και αρχιτεκτονικές, που ξεπερνούν τους περιορισμούς των παραδοσιακών μοντέλων.

Μία κατεύθυνση μελλοντικής έρευνας είναι τα Adaptive AI Models [138], τα οποία προσαρμόζουν την υπολογιστική πολυπλοκότητα με βάση την κατάσταση του δικτύου και τα χαρακτηριστικά της εισόδου. Η προσαρμογή αυτή επιτυγχάνεται μέσω τεχνικών όπως η δυναμική ενεργοποίηση επιπέδων νευρώνων, η παράλειψη υπολογιστικών βημάτων και η ρύθμιση της ακρίβειας των υπολογισμών, ανάλογα με τη διαθεσιμότητα πόρων. Η λειτουργία τους βασίζεται συχνά σε μηχανισμούς όπως το conditional computation και το dynamic routing, που επιτρέπουν στο μοντέλο να αξιοποιεί μόνο ένα υποσύνολο του συνολικού δικτύου κατά την εκτέλεση. Σε περιβάλλοντα IoT, τα μοντέλα αυτά διατηρούν τη χρήση πόρων όσο το δυνατόν χαμηλότερη, αυξάνοντας προσωρινά την υπολογιστική πολυπλοκότητα μόνο όταν εντοπίζονται ύποπτα μοτίβα που υποδηλώνουν πιθανή εισβολή ή κρίσιμες αλλαγές στη ροή των δεδομένων. Με αυτόν τον τρόπο, επιτυγχάνεται εξοικονόμηση ενέργειας και υπολογιστικής ισχύος χωρίς απώλεια στην αποτελεσματικότητα της ανίχνευσης.

Για εφαρμογές IoT με αυστηρούς περιορισμούς σε ενέργεια και υπολογιστική ισχύ, τα Lightweight AI Models προσφέρουν μία πρακτική λύση [139]. Έχουν σχεδιαστεί με στόχο τη μόνιμα χαμηλή υπολογιστική επιβάρυνση, ώστε να είναι κατάλληλα για λειτουργία σε IoT συσκευές με περιορισμένους πόρους. Επιτυγχάνουν αποδοτικότητα μέσω τεχνικών όπως η συμπίεση μοντέλων (model compression), το pruning, η ποσοτικοποίηση (quantization) και η χρήση απλοποιημένων αρχιτεκτονικών. Επιπλέον, ενσωματώνουν μεθόδους όπως το online και continual learning, ώστε να επεξεργάζονται δεδομένα σε πραγματικό χρόνο, με συνεχή εκπαίδευση, μειώνοντας περαιτέρω το κόστος επικοινωνίας και αποθήκευσης. Η υποκατηγορία TinyML επιτρέπει την εκτέλεση μοντέλων σε εξαιρετικά περιορισμένο hardware, όπως μικροελεγκτές, διατηρώντας τη δυνατότητα για τοπική ανάλυση συμβάντων σε επίπεδο milliwatts, κάτι κρίσιμο για εφαρμογές intrusion detection και ενεργειακά αποδοτική διαχείριση πόρων στο IoT [140].

Μία εναλλακτική κατεύθυνση μελλοντικής έρευνας είναι τα Sparse Computing μοντέλα [141], τα οποία εστιάζουν στη μείωση της υπολογιστικής επιβάρυνσης, ενεργοποιώντας μόνο τα κρίσιμα τμήματα του δικτύου κατά την επεξεργασία. Η λογική τους βασίζεται στο ότι, σε πολλές περιπτώσεις, μόνο ένα υποσύνολο των νευρώνων και των συνδέσεών τους είναι απαραίτητο για να παραχθεί έξοδος με την επιθυμητή ακρίβεια. Τεχνικές όπως το weight pruning, η δυναμική αραίωση (dynamic sparsity) και η επιλογή ενεργών διαδρομών υπολογισμού εφαρμόζονται για να μειωθεί η χρήση μνήμης και ενέργειας χωρίς σημαντική απώλεια απόδοσης. Σε IoT περιβάλλοντα, αυτή η προσέγγιση επιτρέπει την υλοποίηση βαθιών μοντέλων ακόμη και σε συσκευές με εξαιρετικά περιορισμένους πόρους, βελτιώνοντας την ταχύτητα απόκρισης σε σενάρια ανίχνευσης εισβολών και διαχείρισης πόρων με ελάχιστο κόστος.

Σε περιβάλλοντα IoT όπου η υπολογιστική ισχύς και τα δεδομένα είναι περιορισμένα, το Transfer Learning αποτελεί μια πρακτική και αποδοτική προσέγγιση για την ανάπτυξη μοντέλων τεχνητής νοημοσύνης [142]. Η βασική ιδέα είναι η αξιοποίηση γνώσης που έχει αποκτηθεί από ένα προηγούμενο μοντέλο, εκπαιδευμένο σε μεγαλύτερο σύνολο δεδομένων, ώστε να εφαρμοστεί σε νέο, πιο εξειδικευμένο πρόβλημα. Με αυτόν τον τρόπο, μειώνεται δραστικά ο απαιτούμενος όγκος δεδομένων και το υπολογιστικό κόστος εκπαίδευσης, επιταχύνοντας την ανάπτυξη πιο απαιτητικών μοντέλων σε IoT συσκευές με περιορισμένους πόρους. Το transfer learning είναι ιδιαίτερα χρήσιμο για real-time εντοπισμό εισβολών σε edge συσκευές, όπου η άμεση μάθηση από περιορισμένο αριθμό δεδομένων είναι κρίσιμη. Επιπλέον, διευκολύνει την κατανομή πόρων, επιτρέποντας σε κόμβους του δικτύου να ανταλλάσσουν προϋπάρχουσα γνώση χωρίς να απαιτείται πλήρης εκπαίδευση σε κάθε σημείο του συστήματος.

Μία από τις καινοτόμες και υποσχόμενες προσεγγίσεις είναι το Neuromorhpic Computing [143]. Βασίζεται σε εξειδικευμένα κυκλώματα που εμπνέονται από τη λειτουργία του ανθρώπινου εγκεφάλου και υποστηρίζουν Spiking Neural Networks (SNNs), τα οποία ενεργοποιούνται μόνο όταν υπάρχει νέα πληροφορία (event-driven processing). Η προσέγγιση αυτή εξασφαλίζει μείωση της κατανάλωσης ενέργειας σε σύγκριση με τα παραδοσιακά δίκτυα που λειτουργούν συνεχώς. Επιπλέον, οι υπολογισμοί πραγματοποιούνται τοπικά με εξαιρετικά χαμηλή καθυστέρηση, επιτρέποντας την άμεση αναγνώριση και απόκριση σε εισβολές, χωρίς ανάγκη συνεχούς επικοινωνίας με κεντρικούς servers. Η ικανότητα αυτοπροσαρμογής στη διαθεσιμότητα πόρων, σε συνδυασμό με την υψηλή ενεργειακή απόδοση, καθιστούν τα neuromorphic chips ιδανικά για χρήση σε κόμβους IoT. Ενδεικτικά, το Loihi και το Loihi 2 της Intel, καθώς και το TrueNorth της IBM, αποτελούν εξειδικευμένα chips που υποστηρίζουν την αρχιτεκτονική SNN, προσφέροντας δυνατότητα μάθησης και απόκρισης με εξαιρετικά μικρό υπολογιστικό κόστος [144] [145].

Στο μέλλον, οι τεχνικές που αναλύθηκαν στις προηγούμενες ενότητες θα συνεχίσουν να εξελίσσονται και να αξιοποιούνται σε ακόμη πιο αποδοτικές παραλλαγές. Ο συνδυασμός τους σε υβριδικά μοντέλα, προσαρμοσμένα στις απαιτήσεις των IoT περιβαλλόντων, μπορεί να οδηγήσει σε σημαντική βελτίωση τόσο στα ποσοστά ανίχνευσης εισβολών όσο και στη δυναμική κατανομή πόρων. Επιπλέον, η ενσωμάτωσή τους με σύγχρονες κατευθύνσεις ενδέχεται να προσφέρει νέες δυνατότητες σε επίπεδο απόδοσης, ενεργειακής αποδοτικότητας και real-time απόκρισης. Αυτό αποδεικνύει ότι η μελλοντική εξέλιξη δεν βρίσκεται μόνο στη βελτίωση μεμονωμένων μοντέλων, αλλά στον συνδυασμό και την συνεχή προσαρμογή τους στο περιβάλλον λειτουργίας τους.

Βιβλιογραφία

[1] P. Illy, "AI-driven solutions for safeguarding IoT environments: an intrusion detection and prevention study," *Espace ETS*, 2023. [Online]. Available: https://espace.etsmtl.ca/id/eprint/3456/.

[2] E. Baccour, N. Mhaisen, A. A. Abdellatif, A. Erbad, A. Mohamed, and M. Hamdi, "Pervasive AI for IoT Applications: A Survey on Resource-Efficient Distributed Artificial Intelligence," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22077-22099, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9866918.

[3] J. Asharf, N. Moustafa, S. Nepal, A. Anwar, and Z. Baig, "A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions," *Electronics*, vol. 9, no. 7, p. 1177, 2020. doi: 10.3390/electronics9071177.

[4] D. Craigen, N. Diakun-Thibault, and R. Purse, "Defining cybersecurity," *Technology Innovation Management Review*, vol. 4, no. 10, pp. 13–21, Oct. 2014, doi: 10.22215/timreview/835

[5] H. Wu, H. Han, X. Wang, and S. Sun, "Research on Artificial Intelligence Enhancing Internet of Things Security: A Survey," *IEEE Access*, vol. 8, pp. 153826-153848, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9172062.

[6] G. F. E. and S. Sheeja, "Intrusion detection system and mitigation of threats in IoT networks using AI techniques: A review," *Engineering and Applied Science Research*, vol. 50, no. 3, pp. 56-72, 2023. doi: 10.14456/easr.2023.66.

[7] M. Rahman, T. Chippy, V. Dankan Gowda, K. Prasad, H. S. Sethi and P. V. Prasanth, "Optimizing Resource Allocation in Agriculture through IoT and AI," 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 1573-1578, doi: 10.1109/ICECA58529.2023.10394763.

[8] R. Woodhead, P. Stephenson, and D. Morrey, "Digital construction: From point solutions to IoT ecosystem," *Automation in Construction*, vol. 93, pp. 35–46, May 2018, doi: 10.1016/j.autcon.2018.05.004.

[9] C. Paul, A. Ganesh and C. Sunitha, "An overview of IoT based smart homes," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2018, pp. 43-46, doi: 10.1109/ICISC.2018.8398858.

[10] H. Arasteh et al., "Iot-based smart cities: A survey," 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), Florence, Italy, 2016, pp. 1-6, doi: 10.1109/EEEIC.2016.7555867.

[11] R. De Michele and M. Furini, "IoT Healthcare: Benefits, Issues and Challenges," in *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '19)*, Rhodes, Greece, Jun. 2019, doi: 10.1145/3342428.3342693

[12] D. K. Sah, M. Vahabi, and H. Fotouhi, "Federated learning at the edge in Industrial Internet of Things: A review," *ICT Express*, vol. 11, no. 1, pp. 36–44, Feb. 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210537925000071

[13] Tran-Dang, H., Krommenacker, N., Charpentier, P., & Kim, D. S. (2020). The Internet of Things for Logistics: Perspectives, Application Review, and Challenges. *IETE Technical Review*, *39*(1), 93–121. https://doi.org/10.1080/02564602.2020.1827308

[14] S. M. A. A. Abir, A. Anwar, J. Choi and A. S. M. Kayes, "IoT-Enabled Smart Energy Grid: Applications and Challenges," in IEEE Access, vol. 9, pp. 50961-50981, 2021, doi: 10.1109/ACCESS.2021.3067331.

[15] T. Domínguez-Bolaño, O. Campos, V. Barral, C. J. Escudero, and J. A. García-Naya, "An overview of IoT architectures, technologies, and existing open-source projects," *Internet of Things*, vol. 20, p. 100626, Nov. 2022, doi: 10.1016/j.iot.2022.100626.

[16] "Unpacking IoT Architecture: Layers and Components Explained," *Device Authority*. [Online]. Available: https://deviceauthority.com/unpacking-iot-architecture-layers-and-components-explained/

[17] A. Simmons, "Internet of Things (IoT) Architecture: Layers Explained," *Dgtl Infra*, Nov. 2022. [Online]. Available: https://dgtlinfra.com/internet-of-things-iot-architecture/

[18] J. Tournier, F. Lesueur, F. Le Mouël, L. Guyon, and H. Ben-Hassine, "A survey of IoT protocols and their security issues through the lens of a generic IoT stack," *Internet of Things*, vol. 14, p. 100264, Dec. 2021, doi: 10.1016/j.iot.2020.100264.

[19] A. Kondoro, I. B. Dhaou, H. Tenhunen, and N. Mvungi, "Real time performance analysis of secure IoT protocols for microgrid communication," *Future Generation Computer Systems*, vol. 115, pp. 102–115, Mar. 2021, doi: 10.1016/j.future.2020.09.031.

[20] P. Singh, J. J. P, A. Pankaj and R. Mitra, "Edge-Detect: Edge-Centric Network Intrusion Detection using Deep Neural Network," 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2021, pp. 1-6, doi: 10.1109/CCNC49032.2021.9369469.

[21] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing based designs for IoT security," *Internet of Things*, vol. 5, pp. 1–12, Sep. 2019, doi: 10.1016/j.iot.2019.100080.

[22] C. A. Marino, F. Chinelato, and M. Marufuzzaman, "AWS IoT analytics platform for microgrid operation management," *Computers & Industrial Engineering*, vol. 173, p. 108331, Aug. 2022, doi: 10.1016/j.cie.2022.108331.

[23] W. H. Halabi *et al*., "Viability of Azure IoT Hub for Processing High Velocity Large Scale IoT Data," in *Proceedings of the 2021 ACM Southeast Conference (ACM SE '21)*, Virtual Event, USA, Apr. 2021, pp. 78–85, doi: 10.1145/3447545.3451187.

[24] M. C. Osazuwa and O. Mitchell, "Confidentiality, Integrity, and Availability in Network Systems: A Review of Related Literature," Dec. 2023, doi: 10.5281/zenodo.10464076. [Online]. Available: https://www.researchgate.net/publication/377535526

[25] P. K. Sadhu, V. P. Yanambaka, and A. Abdelgawad, "Internet of Things: Security and Solutions Survey," *Sensors*, vol. 22, no. 19, p. 7433, Sep. 2022, doi: 10.3390/s22197433.

[26] A. Schwarz, "Cybersecurity Hall Of Shame: The Worst 10 Breaches Of All Time," *Forbes*, 2015. [Online]. Available: https://www.forbes.com/pictures/647f6bdfc7e02cebf1595934/vtech-2015/

[27] T. Sasi, A. H. Lashkari, R. Lu, P. Xiong, and S. Iqbal, "A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges," *Journal of Information Interaction in Context*, vol. 9, no. 3, Nov. 2024, doi: 10.1016/j.jiixd.2023.12.001

[28] E. Staddon, V. Loscri, and N. Mitton, "Attack Categorisation for IoT Applications in Critical Infrastructures, a Survey," *Applied Sciences*, vol. 11, no. 16, Art. no. 7228, Aug. 2021. doi: 10.3390/app11167228

[29] Y. R. Siwakoti, M. Bhurtel, D. B. Rawat, A. Oest and R. C. Johnson, "Advances in IoT Security: Vulnerabilities, Enabled Criminal Services, Attacks, and Countermeasures," in IEEE Internet of Things Journal, vol. 10, no. 13, pp. 11224-11239, 1 July1, 2023, doi: 10.1109/JIOT.2023.3252594.

[30] M. Ivezic and L. Ivezic, "How Model Inversion Attacks Compromise AI Systems," *Securing AI*, Feb. 2021. [Online]. Available: https://securing.ai/ai-security/model-inversion/

[31] J. Margolis, T. Oh, S. Jadhav, Y. H. Kim, and J.-N. Kim, "An In-Depth Analysis of the Mirai Botnet," *2017 International Conference on Software Security and Assurance (ICSSA)*, 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8392610.

[32] T. Bakhshi, B. Ghita, and I. Kuzminykh, "A Review of IoT Firmware Vulnerabilities and Auditing Techniques," *Sensors*, vol. 24, no. 2, p. 708, Jan. 2024, doi: 10.3390/s24020708

[33] M. Gelgi, Y. Guan, S. Arunachala, M. S. S. Rao, and N. Dragoni, "Systematic Literature Review of IoT Botnet DDOS Attacks and Evaluation of Detection Techniques," *Sensors*, vol. 24, no. 11, Art. no. 3571, May 2024. doi: 10.3390/s24113571

[34] S.-H. Lee, Y.-L. Shiue, C.-H. Cheng, Y.-H. Li, and Y.-F. Huang, "Detection and Prevention of DDoS Attacks on the IoT," *Applied Sciences*, vol. 12, no. 23, Art. no. 12407, Dec. 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/23/12407

[35] J. H. Park, S. K. Singh, M. M. Salim, A. E. Azzaoui, and J. H. Park, "Ransomware-based Cyber Attacks: A Comprehensive Survey," *Journal of Internet Technology*, vol. 23, no. 7, pp. 1907–1924, 2022. [Online]. Available: https://jit.ndhu.edu.tw/article/view/2821

[36] "Ekans Ransomware: Insights on OT Cyber Attacks," *Darktrace Blog*, Jun. 2020. [Online]. Available: https://www.darktrace.com/de/blog/what-the-ekans-ransomware-attack-reveals-about-the-future-of-ot-cyber-attacks.

[37] D. Shah, "Introduction to Training Data Poisoning: A Beginner's Guide," *Lakera Blog*, Mar. 2025. [Online]. Available: https://www.lakera.ai/blog/training-data-poisoning

[38] C. Liu, B. Chen, W. Shao, C. Zhang, K. K. L. Wong, and Y. Zhang, "Unraveling Attacks in Machine Learning-based IoT Ecosystems: A Survey and the Open Libraries Behind Them," *arXiv preprint*, Jan. 2024. [Online]. Available: https://arxiv.org/html/2401.11723v1/#S1

[39] F. A. Yerlikaya and Ş. Bahtiyar, "Data poisoning attacks against machine learning algorithms," *Expert Systems with Applications*, vol. 207, p. 118101, Dec. 2022, doi: 10.1016/j.eswa.2022.118101.

[40] F. Khan *et al*., "Development of a Model for Spoofing Attacks in Internet of Things," *Mathematics*, vol. 10, no. 19, p. 3686, Oct. 2022, doi: 10.3390/math10193686.

[41] L. Arnaboldi and C. Morisset, "A Review of Intrusion Detection Systems and Their Evaluation in the IoT," *arXiv preprint arXiv:2105.08096*, May 2021. [Online]. Available: https://arxiv.org/abs/2105.08096

[42] M. Ozkan-Okay, R. Samet, Ö. Aslan and D. Gupta, "A Comprehensive Systematic Literature Review on Intrusion Detection Systems," in IEEE Access, vol. 9, pp. 157727-157760, 2021, doi: 10.1109/ACCESS.2021.3129336.

[43] A. Khraisat and A. Alazab, "A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges," *Cybersecurity*, vol. 4, no. 1, pp. 1–27, Mar. 2021. [Online]. Available: https://link.springer.com/article/10.1186/s42400-021-00077-7

[44] M. L. Ali, K. Thakur, S. Schmeelk, J. Debello, and D. Dragos, "Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study," *Applied Sciences*, vol. 15, no. 4, p. 1903, Feb. 2025, doi: 10.3390/app15041903.

[45] S. Singh, "Cousins of Artificial Intelligence," *Medium*, Apr. 2, 2024. [Online]. Available: https://medium.com/data-science/cousins-of-artificial-intelligence-dda4edc27b55

[46] N. A. Azeez, T. M. Bada, S. Misra, A. Adewumi, C. Van der Vyver, and R. Ahuja, "Intrusion Detection and Prevention Systems: An Updated Review," in *Data Management, Analytics and Innovation*, Springer, 2021, pp. 589–608. doi: 10.1007/978-981-32-9949-8_48.

[47] J. Díaz-Verdejo, J. Muñoz-Calle, A. Estepa Alonso, R. Estepa Alonso, and G. Madinabeitia, "On the Detection Capabilities of Signature-Based Intrusion Detection Systems in the Context of Web Attacks," *Applied Sciences*, vol. 12, no. 2, p. 852, 2022. doi: 10.3390/app12020852.

[48] P. I. Radoglou-Grammatikis and P. G. Sarigiannidis, "An Anomaly-Based Intrusion Detection System for the Smart Grid Based on CART Decision Tree," *IEEE Journal of Photovoltaics*, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8635743

[49] Q. Jiao and L. Mhamdi, "Deep Learning based Intrusion Detection for IoT Networks," *2022 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/10449910

[50] A. Aldallal, "Signature-based IDS architecture," *ResearchGate*, 2021. [Online]. Available: https://www.researchgate.net/figure/Signature-based-IDS-architecture-13_fig1_356759205.

[51] H. Yang, S. Liang, J. Ni, H. Li, and X. S. Shen, "Secure and Efficient k NN Classification for Industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10945–10954, Nov. 2020. doi: 10.1109/JIOT.2020.2992349.

[52] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," in IEEE Access, vol. 12, pp. 86716-86727, 2024, doi: 10.1109/ACCESS.2024.3416838.

[53] M. Kumar, M. Hanumanthappa and T. V. S. Kumar, "Intrusion Detection System using decision tree algorithm," 2012 IEEE 14th International Conference on Communication Technology, Chengdu, China, 2012, pp. 629-634, doi: 10.1109/ICCT.2012.6511281.

[54] A. Guezzaz, S. Benkirane, M. Azrour, and S. Khurram, "A Reliable Network Intrusion Detection Approach Using Decision Tree with Enhanced Data Quality," *Security and Privacy*, vol. 2021, Article ID 1230593, pp. 1–15, Aug. 2021. doi: https://doi.org/10.1155/2021/1230593

[55] M. Mohammadi, T. A. Rashid, S. H. T. Karim, A. H. M. Aldalwie, Q. T. Tho, M. Bidaki, A. M. Rahmani, and M. Hosseinzadeh, "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems," *Journal of Network and Computer Applications*, vol. 181, p. 103060, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1084804521000102.

[56] A. Chauhan, "Random Forest Classifier and its Hyperparameters," *Medium*, 2022. [Online]. Available: https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6.

[57] G. Boesch, "Ensemble Learning: A Combined Prediction Model (2025 Guide)," *Viso.ai*, 2025. [Online]. Available: https://viso.ai/deep-learning/ensemble-learning/

[58] M. Z. Mahmud, S. Islam, S. R. Alve, and A. J. Pial, "Optimized IoT Intrusion Detection using Machine Learning Technique," *arXiv preprint arXiv:2412.02845*, Dec. 2024. [Online]. Available: https://arxiv.org/abs/2412.02845

[59] Y. Alotaibi and M. Ilyas, "Ensemble-learning framework for intrusion detection to enhance Internet of Things' devices security," *Sensors*, vol. 23, no. 12, p. 5568, Jun. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/12/5568.

[60] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *National Science Review*, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9893798.

[61] S. Divakar, R. Priyadarshini, R. K. Barik, and D. S. Roy, "An Intelligent Intrusion Detection Scheme Powered by Boosting Algorithm," *IEEE Transactions on Information Forensics and Security*, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9377076.

[62] A. Shahraki, M. Abbasi, and Ø. Haugen, "Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost," *Engineering Applications of Artificial Intelligence*, vol. 94, p. 103770, 2020. [Online]. Available: https://doi.org/10.1016/j.engappai.2020.103770.

[63] A. Arul Anitha and L. Arockiam, "Ada-IDS: AdaBoost Intrusion Detection System for ICMPv6 based Attacks in Internet of Things", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 12, No. 11, 2021. DOI: 10.14569/IJACSA.2021.0121156

[64] A. Chauhan, "Fully Explained Gradient Boosting Technique in Supervised Learning," *Towards AI*, 2023. [Online]. Available: https://pub.towardsai.net/fully-explained-gradient-boosting-technique-in-supervised-learning-d3e293ca70e1.

[65] M. Saied, S. Guirguis, and M. Madbouly, "Review of artificial intelligence for enhancing intrusion detection in the internet of things," *Elsevier*, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095219762301415X.

[66] P. Verma, S. Anwar, S. Khan and S. B. Mane, "Network Intrusion Detection Using Clustering and Gradient Boosting," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1-7, doi: 10.1109/ICCCNT.2018.8494186.

[67] "Understanding Gradient Boosting," *Medium*, [Online]. Available: https://medium.com/@hemashreekilari9/understanding-gradient-boosting-632939b98764. [Accessed: Apr. 16, 2025].

[68] M. Saied, S. Guirguis, and M. Madbouly, "A Comparative Study of Using Boosting-Based Machine Learning Algorithms for IoT Network Intrusion Detection," *Journal of Electrical Engineering & Technology*, vol. 18, no. 6, pp. 2945–2961, Nov. 2023. [Online]. Available: https://link.springer.com/article/10.1007/s44196-023-00355-x

[69] J. Choudhary, "Mastering XGBoost: A Technical Guide for Machine Learning Practitioners," *Medium*, Sep. 2023. [Online]. Available: https://medium.com/@jyotsna.a.choudhary/mastering-xgboost-a-technical-guide-for-intermediate-machine-learning-practitioners-f7ad167c6865

[70] S. S. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective intrusion detection system using XGBoost," *Information*, vol. 9, no. 7, p. 149, Jun. 2018. [Online]. Available: https://www.mdpi.com/2078-2489/9/7/149.

[71] S. Khan, S. Noor, T. Javed, A. Naseem, F. Aslam, S. A. AlQahtani, and N. Ahmad, "XGBoost-enhanced ensemble model using discriminative hybrid features for the prediction of sumoylation sites," *BioData Mining*, vol. 18, no. 1, pp. 1–17, Feb. 2025. [Online]. Available: https://biodatamining.biomedcentral.com/articles/10.1186/s13040-024-00415-8

[72] S. S. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective Intrusion Detection System Using XGBoost," *Information*, vol. 9, no. 7, Art. no. 149, Jun. 2018. doi: 10.3390/info9070149

[73] N. Khadka, "LightGBM Algorithm: The Key to Winning Machine Learning Competitions," *Data Aspirant*, [Online]. Available: https://dataaspirant.com/lightgbm-algorithm/.

[74] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *NeurIPS*, 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

[75] "What is Light GBM," *DataScience.eu*, [Online]. Available: https://datascience.eu/machine-learning/1-what-is-light-gbm/.

[76] G. Zhao, Y. Wang, and J. Wang, "Intrusion Detection Model of Internet of Things Based on LightGBM," *IEICE Transactions on Communications*, vol. E106-B, no. 8, pp. 672–680, Aug. 2023. doi: 10.1587/transcom.2022EBP3169

[77] "Stacking in Machine Learning," *GeeksforGeeks*, May 2019. [Online]. Available: https://www.geeksforgeeks.org/stacking-in-machine-learning/.

[78] "Artificial Neural Networks and Their Applications," *GeeksforGeeks*, 2023. [Online]. Available: https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/.

[79] E. Hodo, X. Bellekens, A. Hamilton, P.-L. Dubouilh, E. Iorkyase, and C. Tachtatzis, "Threat analysis of IoT networks using artificial neural network intrusion detection system," *IEEE Xplore*, 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7746067.

[80] K. A. Shukla, S. Ahamad, G. N. Rao, A. J. Al-Asadi, A. Gupta, and M. Kumbhkar, "Artificial Intelligence Assisted IoT Data Intrusion Detection," *IEEE Access*, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9711795.

[81] J. Pacheco, V. H. Benitez, L. C. Félix-Herrán and P. Satam, "Artificial Neural Networks-Based Intrusion Detection System for Internet of Things Fog Nodes," in IEEE Access, vol. 8, pp. 73907-73918, 2020, doi: 10.1109/ACCESS.2020.2988055.

[82] R. Singh, "Decoding CNNs: A Beginner's Guide to Convolutional Neural Networks and their Applications," *Medium*, Dec. 2024. [Online]. Available: https://ravjot03.medium.com/decoding-cnns-a-beginners-guide-to-convolutional-neural-networks-and-their-applications-1a8806cbf536

[83] D. Scherer, A. Müller, and S. Behnke, "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition," *Artificial Neural Networks – ICANN 2010*, 2010. doi: 10.1007/978-3-642-15825-4_10. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-15825-4_10.

[84] P. V. Huong, L. D. Thuan, L. T. Hong Van and D. V. Hung, "Intrusion Detection in IoT Systems Based on Deep Learning Using Convolutional Neural Network," 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), Hanoi, Vietnam, 2019, pp. 448-453, doi: 10.1109/NICS48868.2019.9023871.

[85] A. Aljumah, "IoT-based intrusion detection system using convolution neural networks," *PeerJ Computer Science*, vol. 7, Art. no. e721, Sep. 2021. [Online]. Available: https://peerj.com/articles/cs-721/

[86] M. Arsalan, M. Mubeen, M. Bilal and S. F. Abbasi, "1D-CNN-IDS: 1D CNN-based Intrusion Detection System for IIoT," 2024 29th International Conference on Automation and Computing (ICAC), Sunderland, United Kingdom, 2024, pp. 1-4, doi: 10.1109/ICAC61394.2024.10718772.

[87] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent Advances in Recurrent Neural Networks," *arXiv preprint*, 2018. doi: 10.48550/arXiv.1801.01078. [Online]. Available: https://arxiv.org/abs/1801.01078.

[88] M. Almiani, A. AbuGhazleh, A. Al-Rahayfeh, S. Atiewi, and A. Razaque, "Deep recurrent neural network for IoT intrusion detection system," *International Journal of Electronics and Communications*, vol. 114, p. 152971, May 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569190X19301625.

[89] H. Sharma, P. Kumar, and K. Sharma, "Recurrent Neural Network based Incremental model for Intrusion Detection System in IoT," *Scalable Computing: Practice and Experience*, vol. 25, no. 5, pp. 617–627, 2024. doi: 10.12694/scpe.v25i5.3004

[90] D. T. Shipmon, J. M. Gurevitch, P. M. Piselli, and S. T. Edwards, "Time Series Anomaly Detection; Detection of anomalous drops with limited features and sparse examples in noisy highly periodic data," *arXiv preprint*, 2017. doi: 10.48550/arXiv.1708.03665. [Online]. Available: https://arxiv.org/abs/1708.03665.

[91] I. Ullah and Q. H. Mahmoud, "Design and Development of RNN Anomaly Detection Model for IoT Networks," in *IEEE Access*, vol. 10, pp. 62722-62750, 2022, doi: 10.1109/ACCESS.2022.3176317.

[92] "10.1. Long Short-Term Memory (LSTM)," *Dive into Deep Learning*, 2022. [Online]. Available: https://d2l.ai/chapter_recurrent-modern/lstm.html.

[93] H. R. Sayegh, W. Dong, and A. M. Al-madani, "Enhanced Intrusion Detection with LSTM-Based Model, Feature Selection, and SMOTE for Imbalanced Data," *Applied Sciences*, vol. 14, no. 2, Art. no. 479, Jan. 2024. doi: 10.3390/app14020479

[94] Laghrissi, F., Douzi, S., Douzi, K. *et al.* Intrusion detection systems using long short-term memory (LSTM). *J Big Data* **8**, 65 (2021). https://doi.org/10.1186/s40537-021-00448-4

[95] "10.2. Gated Recurrent Units (GRU)," *Dive into Deep Learning*, 2022. [Online]. Available: https://d2l.ai/chapter_recurrent-modern/gru.html.

[96] P. N. K., S. Deepika, N. S. Ramya, N. Sharmila, K. Vaishnavi, and P. Bhavitha, "Intrusion Detection System Using Gated Recurrent Neural Network," *Journal of Computer Science Engineering and Software Testing (JCSEST)*, [Online]. Available: https://journalspub.com/wp-content/uploads/2024/05/1-8-Intrusion-Detection-System-Using-Gated-Recurrent-Neural-Network-1.pdf.

[97] S. Kostadinov, "Understanding GRU Networks," *Medium*, Dec. 2017. [Online]. Available: https://medium.com/data-science/understanding-gru-networks-2ef37df6c9be

[98] C. Xu, J. Shen, X. Du and F. Zhang, "An Intrusion Detection System Using a Deep Neural Network With Gated Recurrent Units," in IEEE Access, vol. 6, pp. 48697-48707, 2018, doi: 10.1109/ACCESS.2018.2867564.

[99] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[100] U. Michelucci, "An Introduction to Autoencoders," *arXiv preprint*, 2022. doi: 10.48550/arXiv.2201.03898. [Online]. Available: https://arxiv.org/abs/2201.03898.

[101] M. A. Alsoufi, M. M. Siraj, F. A. Ghaleb, and A. H. Abdulqader, "An Anomaly Intrusion Detection Systems in IoT Based on Autoencoder: A Review," in *Proceedings of the International Conference on Recent Advances in Intelligent Systems and Smart Applications (RAISSA 2024)*, May 2024, pp. 273–284. doi: 10.1007/978-3-031-59707-7_20

[102] W. Yao, L. Hu, Y. Hou and X. Li, "A Lightweight Intelligent Network Intrusion Detection System Using One-Class Autoencoder and Ensemble Learning for IoT," *Sensors*, vol. 23, no. 8, p. 4141, Apr. 2023. doi: 10.3390/s23084141

[103] F. S. Alrayes, M. Zakariah, S. U. Amin, Z. I. Khan, and M. Helal, "Intrusion Detection in IoT Systems Using Denoising Autoencoder," *IEEE Xplore*, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10658641.

[104] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," *ACM Digital Library*, 2008. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/1390156.1390294.

[105] I. O. Lopes, D. Zou, I. H. Abdulqadder, F. A. Ruambo, B. Yuan, and H. Jin, "Effective network intrusion detection via representation learning: A denoising autoencoder approach," *Computer Communications*, vol. 192, pp. 98-107, Jul. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366422002742.

[106] A. Keshavarz, "Image Denoising Using Autoencoders," *Medium*, Jun. 2023. [Online]. Available: https://medium.com/@a.keshavarz/image-denoising-using-autoencoders-improved-version-5f8a90019971

[107] J. H. Joloudari, R. Alizadehsani, I. Nodehi, and S. Mojrian, "Resource allocation optimization using artificial intelligence methods in various computing paradigms: A Review," *ResearchGate*, Mar. 2022. doi: 10.13140/RG.2.2.32857.39522

[108] Z. Ghanbari, N. J. Navimipour, M. Hosseinzadeh, and A. Darwesh, "Resource allocation mechanisms and approaches on the Internet of Things," *Cluster Computing*, vol. 22, pp. 11227–11256, Jan. 2019. [Online]. Available: https://link.springer.com/article/10.1007/s10586-019-02910-8

[109] V. Kanade, "What Is Reinforcement Learning? Working, Algorithms, and Uses," *Spiceworks*, Sep. 2022. [Online]. Available: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-reinforcement-learning/

[110] A. K. Shakya, G. Pillai, and S. Chakrabarty, "Reinforcement learning algorithms: A brief survey," *Expert Systems with Applications*, vol. 236, Art. no. 120495, Nov. 2023. doi: 10.1016/j.eswa.2023.120495

[111] S. Jamshidi, A. Nikanjam, K. W. Nafi, F. Khomh, and R. Rasta, "Application of deep reinforcement learning for intrusion detection in Internet of Things: A systematic review," *Internet of Things*, vol. 23, Art. no. 101531, May 2025. doi: 10.1016/j.iot.2025.101531

[112] R. G. Goriparthi, "Reinforcement Learning in IoT: Enhancing Smart Device Autonomy through AI," *ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/386142956_Reinforcement_Learning_in_IoT_Enhancing_Smart_Device_Autonomy_through_AI

[113] K. Gai and M. Qiu, "Optimal resource allocation using reinforcement learning for IoT," *Applied Soft Computing*, vol. 73, pp. 26–36, Aug. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494618302540

[114] Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc G. Bellemare and Joelle Pineau (2018), "An Introduction to Deep Reinforcement Learning", Foundations and Trends in Machine Learning: Vol. 11: No. 3-4, pp 219-354. http://dx.doi.org/10.1561/2200000071

[115] Mousavi, S.S., Schukat, M., Howley, E. (2018). Deep Reinforcement Learning: An Overview. In: Bi, Y., Kapoor, S., Bhatia, R. (eds) Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016. IntelliSys 2016. Lecture Notes in Networks and Systems, vol 16. Springer, Cham. https://doi.org/10.1007/978-3-319-56991-8_32

[116] N. D. Pozza, L. Buffoni, S. Martina, and F. Caruso, "Deep reinforcement learning scheme: A deep neural network learns the policy" *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/figure/Deep-reinforcement-learning-scheme-A-deep-neural-network-learns-the-policy_fig1_360910430.

[117] J. F. Cevallos M., A. Rizzardi, S. Sicari, and A. C. Porisini, "Deep Reinforcement Learning for intrusion detection in Internet of Things: Best practices, lessons learnt, and open challenge," *ScienceDirect*, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1389128623004619.

[118] P. Cheng, Y. Chen, M. Ding, Z. Chen, S. Liu and Y. -P. P. Chen, "Deep Reinforcement Learning for Online Resource Allocation in IoT Networks: Technology, Development, and Future Challenges," in IEEE Communications Magazine, vol. 61, no. 6, pp. 111-117, June 2023, doi: 10.1109/MCOM.001.2200526.

[119] S. Tharewal, M. W. Ashfaque, S. S. Banu, P. Uma, S. M. Hassen, and M. Shabaz, "Intrusion Detection System for Industrial Internet of Things Based on Deep Reinforcement Learning," *Computational Intelligence and Neuroscience*, vol. 2022, Art. no. 9023719, Mar. 2022. doi: 10.1155/2022/9023719

[120] X. Xiong, K. Zheng, L. Lei, and L. Hou, "Resource Allocation Based on Deep Reinforcement Learning in IoT Edge Computing," *IEEE Xplore*, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9060882.

[121] P. M. Mammen, "Federated Learning: Opportunities and Challenges," *arXiv preprint arXiv:2101.05428*, Jan. 2021. doi: 10.48550/arXiv.2101.05428

[122] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges," *IEEE Communications Surveys & Tutorials*, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9460016.

[123] M. Baumgartner, S. P. K. Veeranki, D. Hayn, and G. Schreier, "Federated Learning Scheme," *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/figure/Federated-learning-scheme-M3a-In-a-first-step-an-initial-untrained-model-was_fig4_373192668.

[124] R. Lazzarini, H. Tianfield, and V. Charissis, "Federated Learning for IoT Intrusion Detection," *MDPI Cryptography*, 2023. [Online]. Available: https://www.mdpi.com/2673-2688/4/3/28.

[125] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan and A. -R. Sadeghi, "DÏoT: A Federated Self-learning Anomaly Detection System for IoT," 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 2019, pp. 756-767, doi: 10.1109/ICDCS.2019.00080.

[126] T. D. Nguyen, P. Rieger, M. Miettinen, and A.-R. Sadeghi, "Poisoning Attacks on Federated Learning-based IoT Intrusion Detection System," *Proceedings of the NDSS Symposium*, [Online]. Available: https://www.ndss-symposium.org/wp-content/uploads/2020/04/diss2020-23003-paper.pdf.

[127] A. Imteaj, U. Thakker, S. Wang, J. Li and M. H. Amini, "A Survey on Federated Learning for Resource-Constrained IoT Devices," in IEEE Internet of Things Journal, vol. 9, no. 1, pp. 1-24, 1 Jan.1, 2022, doi: 10.1109/JIOT.2021.3095077.

[128] H. Chen, S. Huang, D. Zhang, M. Xiao, M. Skoglund, and H. V. Poor, "Federated Learning over Wireless IoT Networks with Optimized Communication and Resources," *arXiv preprint arXiv:2110.11775*, Oct. 2021. [Online]. Available: https://arxiv.org/pdf/2110.11775

[129] J. Qi, Q. Zhou, L. Lei, and K. Zheng, "Federated Reinforcement Learning: Techniques, Applications, and Open Challenges," *arXiv preprint arXiv:2108.11887*, Oct. 2021. [Online]. Available: https://arxiv.org/abs/2108.11887

[130] E. C. P. Neto, S. Sadeghi, X. Zhang, and S. Dadkhah, "Federated Reinforcement Learning in IoT: Applications, Opportunities and Open Challenges," *Applied Sciences*, vol. 13, no. 11, Art. no. 6497, May 2023. doi: 10.3390/app13116497

[131] J. Qi, Q. Zhou, and K. Zheng, "Federated reinforcement learning: techniques, applications, and open challenges," *Intelligence & Robotics*, vol. 1, no. 2, pp. 135–150, Oct. 2021. [Online]. Available: https://www.oaepublish.com/articles/ir.2021.02

[132] E. M. Campos, P. F. Saura, A. González-Vidal, J. L. Hernández-Ramos, J. B. Bernabé, G. Baldini, and A. Skarmeta, "Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges," *Computer Networks*, vol. 205, Art. no. 108693, Dec. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621005405

[133] S. Otoum, N. Guizani and H. Mouftah, "Federated Reinforcement Learning-Supported IDS for IoT-steered Healthcare Systems," ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500698.

[134] S. Vadigi, K. Sethi, D. Mohanty, S. P. Das, and P. Bera, "Federated reinforcement learning based intrusion detection system using dynamic attention mechanism," *Journal of Information Security and Applications*, vol. 75, Art. no. 103608, Nov. 2023. doi: 10.1016/j.jisa.2023.103608

[135] S. Mali, F. Zeng, D. Adhikari, I. Ullah, M. A. Al-Khasawneh, S. Alfarraj, and F. Alblehai, "Federated Reinforcement Learning-Based Dynamic Resource Allocation and Task Scheduling in Edge for IoT Applications," *Sensors*, vol. 25, no. 7, Art. no. 2197, Mar. 2025. doi: 10.3390/s25072197

[136] A. S. M. S. Sagar, A. Haider, and H. S. Kim, "A hierarchical adaptive federated reinforcement learning for efficient resource allocation and task scheduling in hierarchical IoT network," *Computer Communications*, vol. 213, Art. no. 107969, Jan. 2025. doi: 10.1016/j.comcom.2024.107969

[137] Z. Tianqing, W. Zhou, D. Ye, Z. Cheng and J. Li, "Resource Allocation in IoT Edge Computing via Concurrent Federated Reinforcement Learning," in IEEE Internet of Things Journal, vol. 9, no. 2, pp. 1414-1426, 15 Jan.15, 2022, doi: 10.1109/JIOT.2021.3086910.

[138] H. Shah and J. Patel, "Adaptive AI architectures: Integrating machine learning and self-healing capabilities," *International Business and Humanities Studies Symposium (IBHSS)*, Nov. 2024. [Online]. Available: https://ibhss.com/index.php/ibhss/article/view/34.

[139] I. Idrissi, M. Azizi, and O. Moussaoui, "A lightweight optimized deep learning-based host-intrusion detection system deployed on the edge for IoT," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, p. 17, 2022. doi: 10.12785/ijcds/110117.

[140] L. Dutta and S. Bharali, "TinyML Meets IoT: A Comprehensive Survey," *Internet of Things*, vol. 16, Art. no. 100432, Dec. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S2542660521001025

[141] A. Alnoman, S. Erkucuk, and A. Anpalagan, "Sparse code multiple access-based edge computing for IoT systems," *IEEE Xplore*, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8704942.

[142] S. Yılmaz, E. Aydogan and S. Sen, "A Transfer Learning Approach for Securing Resource-Constrained IoT Devices," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4405-4418, 2021, doi: 10.1109/TIFS.2021.3096029.

[143] S. Salehi, T. Sheaves, K. I. Gubbi, S. A. Beheshti, S. M. P. D., and S. Rafatirad, "Neuromorphic-enabled security for IoT," *IEEE Xplore*, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9842256.

[144] M. Davies, N. Srinivasa, T. H. Lin, G. Chinya, Y. Cao, S. Choday, G. Dimou, P. Joshi, N. Imam, E. Lekuch, D. R. Liu, A. Mathaikutty, S. McCoy, A. Paul, J. Tse, and I. V. T. Tho, "Loihi: A neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82-99, Jan./Feb. 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8259423.

[145] H. -P. Cheng, W. Wen, C. Wu, S. Li, H. H. Li and Y. Chen, "Understanding the design of IBM neurosynaptic system and its tradeoffs: A user perspective," *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*, Lausanne, Switzerland, 2017, pp. 139-144, doi: 10.23919/DATE.2017.7926972.