ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Υλοποίηση εφαρμογής Federated Learning για την κυβερνοασφάλεια ασύρματων δικτύων 5G/6G, IoT και Edge Computing

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Άντρεα Μιχαηλίδου

Επιβλέπων: Ασκούνης Δημήτρης

Καθηγήτης Ε.Μ.Π.

Αθήνα, Ιούλιος, 2025

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Υλοποίηση εφαρμογής Federated Learning για την κυβερνοασφάλεια ασύρματων δικτύων 5G/6G, IoT και Edge Computing

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Άντρεα Μιχαηλίδου

Επιβλέπων: Ασκούνης Δημήτρης

Καθηγήτης Ε.Μ.Π.

Εγκρίθηκε από την τριμελή κριτική επιτροπή την 3η Ιουλίου 2025.

Ασκούνης Δημήτρης	Ψαρράς Ιωάννης	Μαρινάκης Βαγγέλης
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος, 2025

Άντρεα Μιχαηλίδου

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών, Ε.Μ.Π.

Περίληψη

Η συνεχώς αυξανόμενη χρήση έξυπνων συσκευών και η εξάπλωση των τεχνολογιών 5G/6G και του Διαδικτύου των Πραγμάτων (IoT) έχουν οδηγήσει σε ένα νέο περιβάλλον ψηφιακής επικοινωνίας, όπου η ασφάλεια των δεδομένων αποτελεί κρίσιμη πρόκληση. Η παρούσα διπλωματική εργασία εστιάζει στην ανάπτυξη ενός συστήματος ανίχνευσης επιθέσεων που αξιοποιεί την Ομοσπονδιακή Μάθηση, μια μέθοδο εκπαίδευσης αλγορίθμων χωρίς την αποστολή των δεδομένων σε κάποιο κεντρικό σημείο. Στη συγκεκριμένη προσέγγιση κάθε επιμέρους συμμετέχων, όπως μία συσκευή ή ένας κόμβος, εκπαιδεύει το δικό του μοντέλο τοπικά και στέλνει μόνο τις παραμέτρους της εκπαίδευσης, ενισχύοντας έτσι την προστασία των προσωπικών δεδομένων.

Για την υλοποίηση χρησιμοποιήθηκε το σύνολο δεδομένων N-BaIoT που περιλαμβάνει πραγματικά σενάρια επιθέσεων τύπου Mirai σε συσκευές IoT. Το σύστημα εκπαιδεύτηκε με νευρωνικό δίκτυο πλήρους συνδεσιμότητας, και εφαρμόστηκε μηχανισμός έγκαιρης διακοπής της εκπαίδευσης για την αποφυγή υπερπροσαρμογής. Η απόδοση του συστήματος αξιολογήθηκε με βάση δείκτες όπως η ακρίβεια και η ισορροπία μεταξύ ανίχνευσης και αποφυγής ψευδών θετικών αποτελεσμάτων. Το τελικό παγκόσμιο μοντέλο πέτυχε ακρίβεια 92,22% και F1-score 0,9206, επιδεικνύοντας σταθερή συμπεριφορά κατά τη διάρκεια των ομοσπονδιακών γύρων εκπαίδευσης. Για λόγους σύγκρισης, υλοποιήθηκε και μια κλασική μέθοδος κεντρικής εκπαίδευσης με τα ίδια δεδομένα αλλά συγκεντρωμένα, η οποία παρουσίασε ελαφρώς καλύτερη ακρίβεια 99,63%.

Η εργασία απευθύνεται σε όσους ασχολούνται με την ασφάλεια συστημάτων νέας γενιάς και αναζητούν μεθόδους που προστατεύουν τα δεδομένα κατά την εκπαίδευση μοντέλων τεχνητής νοημοσύνης. Τέλος, προτείνονται μελλοντικές επεκτάσεις και πιθανές βελτιώσεις του συστήματος.

Λέξεις - Κλειδία Ανίχνευση Δικτυακών Επιθέσεων, Ομοσπονδιακή Μάθηση, Διαδίκτυο των Πραγμάτων, Κυβερνοασφάλεια

Abstract

The rapid increase in the use of smart devices and the expansion of 5G/6G technologies and the Internet of Things (IoT) have created a new digital communication landscape, where data security poses a critical challenge. This thesis focuses on the implementation of a Federated Learning (FL) system for intrusion detection, a training method where algorithms are trained without sending data to a central server. In this approach, each individual participant, such as a device or node, trains its local model and only shares the updated model parameters, thereby enhancing data privacy.

For the implementation, the N-BaIoT dataset was used, which includes real-world Mirai attack scenarios targeting IoT devices. A fully connected neural network was employed, along with an early stopping mechanism to prevent overfitting. The system's performance was evaluated using metrics such as accuracy and the balance between true detections and false positives. The final global model achieved a test accuracy of 92.22% and an F1-score of 0.9206, demonstrating consistent performance throughout the federated training rounds. For comparison, a traditional Centralized Learning approach was also implemented using the same data in aggregated form, achieving slightly higher accuracy at 99.63%.

This work is intended for researchers and practitioners involved in the security of next-generation systems, offering privacy-preserving machine learning solutions. Finally, the thesis discusses possible future extensions.

Key words Intrusion Detection System, Federated Learning, Internet of Things, Cybersecurity

Ευχαριστίες

Θα ήθελα αρχικά να απευθύνω τις ευχαριστίες μου στον υποψήφιο Δρα Πάλμο Στέφανο για την επίβλεψη, καθοδήγησή και την εξαιρετική συνεργασία που είχαμε κατά την υλοποίηση αυτής της διπλωματικής εργασίας. Επίσης ευχαριστώ ιδιαίτερα τον Δρα Σωτήρη Πελέκη καθώς και την υποψήφια Δρα Αφροδίτη Μπλίκα για την βοήθεια τους. Τέλος θα ήθελα να ευχαριστήσω με όλη μου την καρδιά την οικογένεια μου, την γάτα μου και τους φίλους μου για την ψυχική και ηθική υποστήριξη που μου προσέφεραν.

Αθήνα, Ιούλιος 2025

Άντρεα Μιχαηλίδου

Περιεχόμενα

Ευρετήτιο Εικόνων

Ευρετήριο Πινάκων

Ακρωνύμια

IDS	Intrusion Detection System
IoT	Internet of Things
FL	Federated Learning
ML	Machine Learning
CL	Centralized Learning
DL	Deep Learning
DDoS	Distributed Denial of Service
DoS	Denial of Service
VFL	Vertical Federated Learning
FCNN	Fully Connected Neural Network
CNN	Convolutional Neural Network
DNN	Deep Neural Network
ANN	Artificial Neural Network
PIDS	Provenance-based Intrusion Detection
NIDS	Network-based Intrusion Detection System
HIDS	Host-based Intrusion Detection System
VR	Virtual Reality
AR	Augmented Reality
MEC	Mobile Edge Computing
SVR	Support Vector Machines
RNN	Recurrent Neural Networks
MLP	Multilayer Perceptron
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit

1. Εισαγωγή

Ζούμε στην ψηφιακή εποχή, όπου σχεδόν κάθε άτομο και συσκευή έχει πρόσβαση στο διαδίκτυο και η ασφάλεια των δικτύων καθίσταται κρίσιμος παράγοντας για την σωστή λειτουργία κάθε συστήματος. Η συνεχής αύξηση των χρηστών, των εφαρμογών και των διασυνδεδεμένων συσκευών έχει οδηγήσει σε μια εξίσου αυξανόμενη εμφάνιση κακόβουλης δικτυακής δραστηριότητας [1]. Επιθέσεις όπως Distributed Denial of Service (DDoS), η υποκλοπή δεδομένων και η μη εξουσιοδοτημένη πρόσβαση απειλούν την ιδιωτικότητα κρίσιμων πληροφοριών [2]. Οι επιθέσεις DDoS αποτελούν μια από τις μεγαλύτερες απειλές για την αξιοπιστία των δικτύων, αφού στοχεύουν στην υπερφόρτωση των πόρων ενός συστήματος, καθιστώντας απρόσιτες τις υπηρεσίες για τους νόμιμους χρήστες και προκαλώντας σημαντικές οικονομικές ζημιές. Σε αυτό το πλαίσιο, η έγκαιρη ανίχνευση και πρόληψη της κακόβουλης κίνησης αποτελεί βασικό στόχο της σύγχρονης ασφάλειας δικτύων.

Μια σύγχρονη προσέγγιση βασίζεται στη χρήση μεθόδων μηχανικής μάθησης (ML), οι οποίες εκπαιδεύουν μοντέλα να διαχωρίζουν τη δικτυακή κίνηση σε κανονική και κακόβουλη [3]. Η αποτελεσματικότητα αυτών των μοντέλων βελτιώνεται με όσο περισσότερα δεδομένα συγκεντρώνονται, ωστόσο λόγω της ανομοιογένειας της κυκλοφορίας ανάλογα με τον τύπο του περιβάλλοντος (π.χ. οικιακό, εταιρικό, βιομηχανικό) ή τη γεωγραφική τοποθεσία, απαιτείται η συγκέντρωση μεγάλου όγκου δεδομένων για την εκπαίδευση αξιόπιστων μοντέλων. Η προσέγγιση αυτή, εκτός από υπολογιστικά μεγάλη, εγείρει σημαντικά ζητήματα απορρήτου, καθώς τα δεδομένα δικτύου περιλαμβάνουν προσωπικές πληροφορίες.

Η Ομοσπονδιακή Μάθηση (FL) αντιμετωπίζει αυτά τα προβλήματα επιτρέποντας την τοπική εκπαίδευση των μοντέλων σε κάθε συμμετέχουσα οντότητα, τα αποτελέσματα της οποίας (οι ενημερωμένες παράμετροι) αποστέλλονται μόνο σε έναν κεντρικό server. Με αυτόν τον τρόπο, τα προσωπικά δεδομένα παραμένουν στις συσκευές ή στα τερματικά, διασφαλίζοντας την ιδιωτικότητα [2], [3].

1.1 Σκοπός Διπλωματικής εργασίας

Η παρούσα Διπλωματική Εργασία έχει ως σκοπό την μελέτη και εφαρμογή τεχνικών FL για την ανίχνευση επιθέσεων σε περιβάλλοντα κατανεμημένων συσκευών, όπως τα δίκτυα IoT. Στόχος είναι η διερεύνηση του κατά πόσο μια προσέγγιση FL μπορεί να ανταποκριθεί αποτελεσματικά στην ανάγκη για ασφάλεια, διατηρώντας ταυτόχρονα την ιδιωτικότητα των δεδομένων που παραμένουν τοπικά στις συσκευές.

Συγκεκριμένα, υλοποιείται ένα σύστημα όπου πολλαπλές συσκευές εκπαιδεύουν τοπικά ένα νευρωνικό δίκτυο, αποστέλλοντας μόνο τα μοντέλα τους σε έναν κεντρικό εξυπηρετητή για συγχώνευση. Για την εφαρμογή χρησιμοποιείται το εργαλείο Flower, ενώ ως σύνολο δεδομένων επιλέγεται το N-BaIoT, το οποίο περιλαμβάνει πραγματικά σενάρια επιθέσεων τύπου Mirai σε IoT συσκευές. Το μοντέλο βασίζεται σε Fully Connected Neural Network (FCNN) με έγκαιρη διακοπή εκπαίδευσης και αξιολογείται με βάση μετρικές όπως η ακρίβεια και το F1-score. Επιπλέον, για συγκριτικούς λόγους, αναπτύσσεται και ένα κεντρικοποιημένο μοντέλο που εξετάζεται με τα ίδια δεδομένα, ώστε να διαπιστωθεί η διαφορά στην απόδοση μεταξύ των δύο προσεγγίσεων.

Η εργασία στοχεύει να αποδείξει ότι μια προσέγγιση FL μπορεί να επιτύχει συγκρίσιμη απόδοση με παραδοσιακές μεθόδους, ενώ ταυτόχρονα παρέχει σημαντικά οφέλη ως προς την προστασία της ιδιωτικότητας και την κατανεμημένη υπολογιστική επεξεργασία.

1.2 Συνεισφορά της Διπλωματικής

Σε συνέχεια των υπαρχουσών μελετών στο πεδίο του FL και της κυβερνοασφάλειας, όπως παρουσιάζονται στο Κεφάλαιο 3, η παρούσα διπλωματική εργασία προτείνει και υλοποιεί έναν ολοκληρωμένο μηχανισμό ανίχνευσης επιθέσεων σε περιβάλλον IoT μέσω FL. Οι βασικές συνεισφορές της εργασίας συνοψίζονται ως εξής:

- Υλοποίηση συστήματος FL με χρήση του πλαισίου Flower και αξιολόγηση σε περιβάλλον ανίχνευσης επιθέσεων τύπου Mirai, χρησιμοποιώντας το ρεαλιστικό dataset N-BaIoT. Το dataset διαμοιράστηκε σε τρεις συμμετέχοντες, με ανεξάρτητη εκπαίδευση και εφαρμογή early stopping, κάτι που δεν παρουσιάζεται συχνά στις υπάρχουσες μελέτες.
- Προσαρμογή και επεξεργασία του N-BaIoT dataset, παρουσιάζοντας αναλυτικά τα στάδια προ επεξεργασίας των δεδομένων. Περιλαμβάνει δημιουργία custom υποσυνόλων δεδομένων και διαχωρισμό ανά συμμετέχοντα.
- Συγκριτική μελέτη παραδοσιακής κεντρικής μάθησης (CL) έναντι FL, αναδεικνύοντας τα πλεονεκτήματα του FL σε θέματα απορρήτου.
- Τοπική αξιολόγηση κάθε πελάτη πραγματοποιείται με τη χρήση του συνόλου δοκιμής, το οποίο περιλαμβάνει δείγματα επιθέσεων που δεν έχουν προηγουμένως παρουσιαστεί κατά τη φάση της εκπαίδευσης. Με αυτόν τον τρόπο, αξιολογείται η ικανότητα του μοντέλου να γενικεύει και να αναγνωρίζει άγνωστα μοτίβα κακόβουλης δραστηριότητας.

1.3 Δομή της διπλωματικής

Η συνέχεια της εργασίας οργανώνεται στα ακόλουθα Κεφάλαια και τη Βιβλιογραφία :

- **Κεφάλαιο 2:** Βιβλιογραφική ανασκόπηση της παρούσας διπλωματικής. Παρουσιάζονται και αναλύονται έννοιες σχετικές με το αντικείμενο μελέτης, όπως Iot, IDS, ML και FL.
- **Κεφάλαιο 3:** Εργασίες με συναφές αντικείμενο, εστιάζοντας κυρίως στη χρήση και αξιοποίηση ανοιχτών συνόλων δεδομένων.
- **Κεφάλαιο 4:** Μεθοδολογία, την επιλογή και την προ επεξεργασία των δεδομένων. Περιγραφή της πειραματικής διαδικασίας.
- **Κεφάλαιο 5:** Παρουσίαση και αξιολόγηση των αποτελεσμάτων. Σύγκριση απόδοσης FL με CL.
- **Κεφάλαιο 6:** Σχολιασμός και σύνοψη της διαδικασίας και των συμπερασμάτων που προκύψανε. Αναφορά σε ενδεχόμενες μελλοντικές επεκτάσεις της παρούσας υλοποίησης

2. Βιβλιογραφική Ανασκόπηση

2.1 Μεθοδολογία Έρευνας

Η μεθοδολογία έρευνας που ακολουθήθηκε για τη συλλογή της βιβλιογραφίας και τη συγγραφή της παρούσας εργασίας αναλύονται στην συνέχεια. Αρχικά, πραγματοποιήθηκε μια γενική μελέτη γύρω από το πεδίο του FL, με έμφαση στη συμβολή του στον τομέα της κυβερνοασφάλειας. Η μελέτη αυτή βοήθησε στην κατανόηση της λειτουργίας του FL, αλλά και στην ανάδειξη της αξίας του ως τεχνολογία προστασίας προσωπικών δεδομένων σε συστήματα ανίχνευσης εισβολών-IDS. Στη συνέχεια, πραγματοποιήθηκε μια ολοκληρωμένη ανάλυση εφαρμογών FL σε 5G δίκτυα, IoT αλλά και αναδυόμενα 6G συστήματα, καλύπτοντας συνολικά περίπου 44 επιστημονικές μελέτες. Παράλληλα, εξετάστηκαν με προσοχή διάφορες εργασίες με θέμα IDS, αντλώντας στοιχεία από 8 μελέτες. Επιπλέον, έγινε μελέτη για τα πιθανά Framewroks , Federated algorithms αλλά και μοντέλα χρησιμοποιώντας πληροφορίες από 24 άρθρα. Τέλος, πραγματοποιήθηκε ανασκόπηση 20 μελετών που περιλαμβάνουν θεωρητική ανάλυση καθώς και πειραματικά αποτελέσματα, με σκοπό την αξιολόγηση διαθέσιμων συνόλων δεδομένων 5G / IoT και τη διερεύνηση της καταλληλότητάς τους για χρήση στην παρούσα εργασία.

Πιο συγκεκριμένα, οι πηγές για την έρευνα μας, συγκεντρώθηκαν απο εργασίες και άρθρα από αξιόπιστα επιστημονικά περιοδικά, όπως τα IEEE Communications Magazine, IEEE Signal Processing Magazine, συνέδρια (π.χ. IEEE nternational Conference on Big Data Science and Engineering (BigDataSE)). Επιπλεόν, χρησιμοποιήθηκε το Google Scholar για την αναζήτηση εργασιών με χρήση συνδυασμών "λέξεων-κλειδιών", όπως "Federated Learning", "Machine Learning", "Intrution Detection System" ,"Cybersecurity", "5G", "6G", "IoT"

2.2 5G Δίκτυα και Internet of Things (IoT)

Το δίκτυο 5G αποτελεί θεμέλιο για τη διασύνδεση του IoT με τους ανθρώπους, καθιστώντας την ασφάλεια του απολύτως αναγκαία [4]. Στο πλαίσιο αυτό, ένα από τα κρίσιμα μέτρα για την ενίσχυση της ασφάλειας του 5G Iot είναι η υιοθέτηση Intrusion Detection Systems (IDS), τα οποία μπορούν να εντοπίζουν και να προλαμβάνουν κακόβουλες ενέργειες στο δίκτυο σε πραγματικό χρόνο. Η μηχανική μάθηση είναι ένα ισχυρό εργαλείο του IDS για τον έγκαιρο και ακριβή εντοπισμό κακόβουλης δραστηριότητας στο δίκτυο.

Ακολουθεί αναλυτική παρουσίαση των εννοιών αυτών, ώστε να καταστεί σαφής η σημασία τους.

2.2.1 5G Δίκτυα

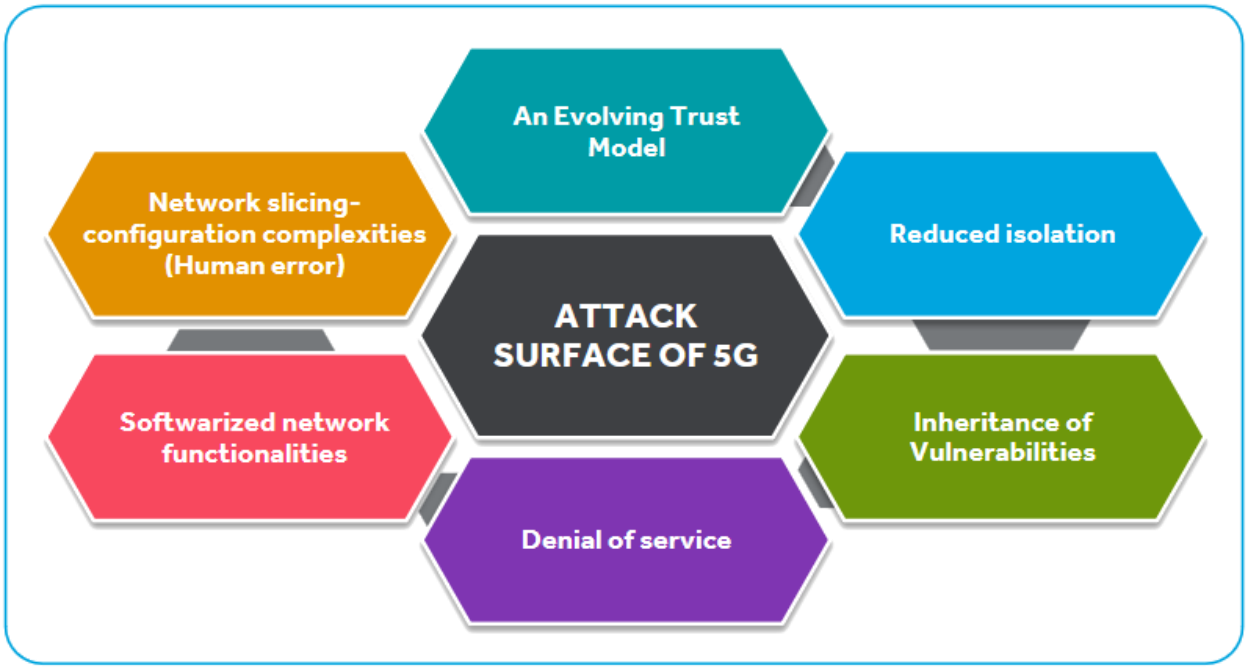
Η συνεχής εξέλιξη των τεχνολογιών επικοινωνίας οδήγησε στην ανάπτυξη και ενσωμάτωση των δικτύων πέμπτης γενιάς (5G), τα οποία προσφέρουν σημαντικές βελτιώσεις σε τομείς όπως η ταχύτητα μετάδοσης, η αξιοπιστία, η χωρητικότητα και η συνολική απόδοση των δικτύων. Τα δίκτυα 5G διαδραματίζουν κομβικό ρόλο στη διασύνδεση τεράστιου αριθμού συσκευών, ιδιαίτερα σε εφαρμογές του IoT, αλλά και σε κρίσιμες υποδομές, όπως είναι οι μονάδες υγειονομικής περίθαλψης [5] ή μεγάλα χρηματοοικονομικά συστήματα.

Ωστόσο, η αυξανόμενη εξάρτηση από τα 5G δίκτυα για τη διασύνδεση και λειτουργία τέτοιων ευαίσθητων συστημάτων φέρνει στην επιφάνεια σοβαρά ζητήματα ασφάλειας και ιδιωτικότητας [6]. Επιπλέον, καινοτόμες εφαρμογές που βασίζονται στο 5G ,όπως τα αυτόνομα οχήματα και οι λειτουργίες βιομηχανικού αυτοματισμού, απαιτούν εξαιρετικά χαμηλή καθυστέρηση και υψηλό επίπεδο αξιοπιστίας, στοιχεία που καθιστούν ακόμη πιο κρίσιμη την ανάγκη για ασφαλή και αποδοτική αρχιτεκτονική δικτύου [7]. Άλλες εφαρμογές του 5G που η ασφάλεια παίζει κομβικό ρόλο είναι είναι η απομακρυσμένη ρομποτική χειρουργική, η εικονική και επαυξημένη πραγματικότητα (VR/AR), τα δίκτυα έξυπνων πόλεων (smart cities), καθώς και η διαχείριση έξυπνων ενεργειακών συστημάτων (smart grids)[8], [9].

2.2.1.1 Ασφάλεια στο 5G

Η τεχνολογία 5G εισάγει σημαντικές προκλήσεις στον τομέα της ασφάλειας, κυρίως λόγω της αυξημένης πολυπλοκότητας του δικτύου και τον τεράστιο αριθμό χρηστών, ιδιαίτερα στο πλαίσιο του IoT. Η υποστήριξη νέων αρχιτεκτονικών, όπως το network slicing, αν και αυξάνει την αποδοτικότητα και ευελιξία, δημιουργεί επίσης νέες επιφάνειες επίθεσης που μπορούν να στοχοποιηθούν από κακόβουλους παράγοντες [10]. Ακόμη, ο μεγάλος αριθμός συσκευών IoT εντείνει τον κίνδυνο επιθέσεων τύπου DDoS και υποκλοπής δεδομένων.

Παρά τις βελτιώσεις που έχουν ενσωματωθεί για την ασφάλεια του 5G, όπως η υποστήριξη ισχυρότερης ταυτοποίησης χρηστών, η κρυπτογράφηση end-to-end και η καλύτερη διαχείριση ταυτότητας, υπάρχουν ακόμη ως κινδύνοι. Οι επιθέσεις σε επίπεδο πυρήνα δικτύου ή στα edge σημεία μπορούν να προκαλέσουν σοβαρές διαταραχές σε κρίσιμες εφαρμογές [11]. Συνεπώς, η ανάγκη για μηχανισμούς ανίχνευσης εισβολών, προσαρμοστική ασφάλεια και χρήση τεχνητής νοημοσύνης για την αντιμετώπιση των απειλών θεωρείται αναγκαία.



Εικόνα 1: 5G Attacks [1]

2.2.1.2 Το μέλλον του 5G και 6G

Καθώς τα δίκτυα επικοινωνιών 5G αναπτύσσονται παγκοσμίως, τόσο η βιομηχανία όσο και ο ακαδημαϊκός κόσμος έχουν αρχίσει να μετακινούνται απο το 5G και να διερευνούν τις επικοινωνίες 6G [12]. Σε αντίθεση με τις προηγούμενες γενιές, το 6G με πρωτοποριακές τεχνολογίες θα φέρει επανάσταση στην ανάπτυξη της ασύρματης επικοινωνίας από «connected thing» σε «connected intelligence» [13] . Το 6G αναμένεται να προσφέρει υπερυψηλές ταχύτητες, εξαιρετικά χαμηλή καθυστέρηση και μαζική συνδεσιμότητα, με εφαρμογές όπως το tactile Internet, τα αυτόνομα οχήματα, η επαυξημένη πραγματικότητα (XR) και τα έξυπνα δίκτυα ενέργειας.

Σε αυτό το νέο τοπίο, η ασφάλεια των δεδομένων και η διατήρηση της ιδιωτικότητας αποκτούν ακόμη μεγαλύτερη σημασία. Το FL αναμένεται να διαδραματίσει κεντρικό ρόλο στο 6G, καθώς επιτρέπει την εκπαίδευση μοντέλων τεχνητής νοημοσύνης χωρίς την ανάγκη μεταφοράς των ευαίσθητων δεδομένων . Η χρήση FL σε ένα περιβάλλον με μεγάλο αριθμό αποκεντρωμένων κόμβων (όπως αισθητήρες, οχήματα, drones) καθιστά εφικτή τη διατήρηση της ιδιωτικότητας των χρηστών σε πραγματικό χρόνο.

Παρόλο που το FL προσφέρει ισχυρές προοπτικές, βρίσκεται ακόμη σε πρώιμο στάδιο εφαρμογής σε περιβάλλοντα 6G και αντιμετωπίζει σημαντικές προκλήσεις [14]. Μία από αυτές είναι το υψηλό κόστος επικοινωνίας, το οποίο οφείλεται στους πολλαπλούς

[1] https://www.cyient.com/whitepaper/proactive-5g-securitytackling-risks-and-vulnerabilities-with-generative-ai

γύρους ανταλλαγής παραμέτρων μεταξύ των συμμετεχόντων μονάδων. Επιπλέον, η ασφάλεια του συστήματος απειλείται λόγω της ετερογένειας των συμμετεχόντων, γεγονός που μπορεί να οδηγήσει σε επιθέσεις τύπου poisoning ή backdoor. Τέλος, εγείρονται σοβαρές ανησυχίες για την προστασία της ιδιωτικότητας, καθώς ενδέχεται να προκύψουν διαρροές μέσω επιθέσεων inference ή membership inference, κατά τις οποίες γίνεται προσπάθεια εξαγωγής πληροφοριών από το μοντέλο.

Για την αντιμετώπιση των παραπάνω προκλήσεων, η ερευνητική κοινότητα προτείνει και εξετάζει καινοτόμες λύσεις όπως το προσαρμοστικό FL, η ασφαλής συσσωμάτωση παραμέτρων (secure aggregation), τα ομότιμα (peer-to-peer) FL δίκτυα καθώς και η ενσωμάτωση της τεχνολογίας blockchain για τη διασφάλιση της ακεραιότητας και της διαφάνειας της διαδικασίας εκπαίδευσης [14].

Συνολικά, το FL αποτελεί μια πολλά υποσχόμενη τεχνολογική κατεύθυνση για την ασφάλεια και την προστασία δεδομένων στα μελλοντικά ευφυή 6G δίκτυα.

2.2.2 Internet of Things (IoT)

Το IoT αναφέρεται σε ένα δίκτυο φυσικών αντικειμένων, όπως συσκευές, οχήματα ή ακόμη και ολόκληρα κτίρια, τα οποία είναι εφοδιασμένα με λογισμικό, αισθητήρες και δυνατότητες δικτύωσης [15]. Αυτό το δίκτυο επιτρέπει στα αντικείμενα να συλλέγουν, να επεξεργάζονται και να ανταλλάσσουν δεδομένα μεταξύ τους, δημιουργώντας μια έξυπνη σύνδεση μεταξύ του φυσικού και του ψηφιακού κόσμου. Οι συσκευές μπορούν να κάνουν το μεγαλύτερο μέρος της εργασίας χωρίς ανθρώπινη παρέμβαση, αν και οι άνθρωποι μπορούν να αλληλεπιδράσουν με αυτές, για παράδειγμα για να διαμορφώσουν, να δώσουν οδηγίες ή να έχουν πρόσβαση σε δεδομένα [16] . Βασικό χαρακτηριστικό του IoT είναι η δυνατότητα απομακρυσμένης παρακολούθησης και ελέγχου μέσω υφιστάμενων δικτυακών υποδομών, προσφέροντας σημαντικές βελτιώσεις σε απόδοση, ακρίβεια και οικονομική αποδοτικότητα σε πολλούς τομείς [15], [16].

2.2.2.1 Ιστορική Αναδρομή και Σημασία

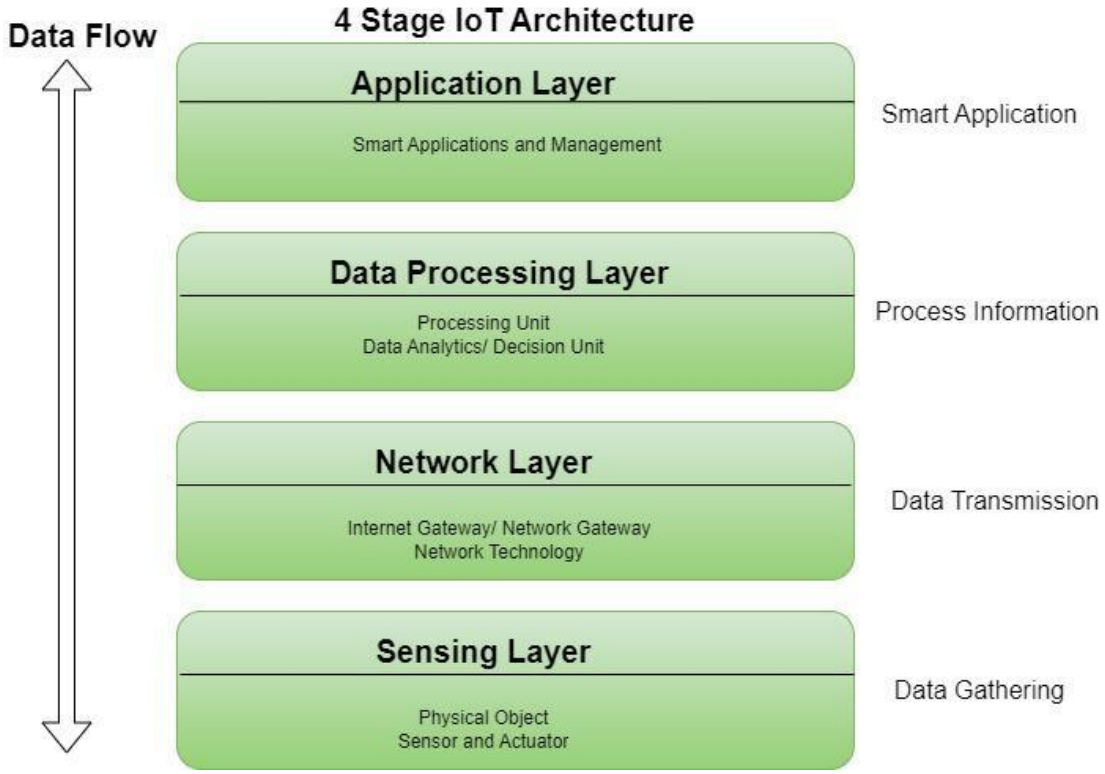
Ο όρος «Internet of Things» επινοήθηκε από τον Βρετανό τεχνολόγο Kevin Ashton το 1999 [15][17], με σκοπό να περιγράψει ένα σύστημα όπου οι αισθητήρες θα έφερναν τον φυσικό κόσμο απευθείας στον ψηφιακό, χωρίς την ανάγκη για συνεχή ανθρώπινη παρέμβαση. Μέχρι το 2005, με την ώθηση της International Telecommunications Unio (ITU) [2], το IoT άρχισε να αναπτύσσεται γρήγορα και υπήρξαν ολοένα και βαθύτερες έρευνες για αυτό και ευρύτερες εφαρμογές. Σήμερα, το IoT έχει εξελιχθεί με εφαρμογές που ποικίλλουν

[2] https://www.itu.int/pub/S-POL-IR.IT-2005

από ένα μικρό δίκτυο όπως ο αυτοματισμός κατοικιών έως μεγάλα δίκτυα όπως cloud για βιομηχανικές εφαρμογές [18].

2.2.2.2 Αρχιτεκτονική IoT

Μια απλή αρχιτεκτονική IoT αποτελείται από 4 βασικά επίπεδα το Επίπεδο Συσκευών, το Επίπεδο Δικτύου, Επίπεδο Επεξεργασίας και το Επίπεδο Εφαρμογών.[3] Το Επίπεδο Αισθητήρων (Sensing Layer) αποτελείται από αισθητήρες και actuators που τοποθετούνται στο περιβάλλον για τη συλλογή πληροφοριών σχετικά με τη θερμοκρασία, την υγρασία, το φως, τον ήχο και άλλες φυσικές παραμέτρους. Δεύτερο στη σειρά είναι το Επίπεδο Δικτύου (Network Layer) που είναι υπεύθυνο για την παροχή επικοινωνίας μεταξύ συσκευών στο σύστημα IoT, χρησιμοποιώντας πρωτόκολλα και τεχνολογίες. Μετά ακολουθεί το Επίπεδο Επεξεργασίας που αναφέρεται στα στοιχεία λογισμικού και υλικού που είναι υπεύθυνα για τη συλλογή, την ανάλυση και την διαχείριση δεδομένων από συσκευές IoT. Τέλος, το Επίπεδο Εφαρμογών (Application Layer) είναι το ανώτατο επίπεδο που αλληλεπιδρά άμεσα με τον τελικό χρήστη, δηλαδή είναι υπεύθυνο για την παροχή φιλικών προς το χρήστη διεπαφών και λειτουργιών που του επιτρέπουν να έχει πρόσβαση και να ελέγχει συσκευές IoT. Ένα παράδειγμα ροής είναι Αισθητήρας → Gateway → Cloud → Εφαρμογή → Χρήστης.



[3] https://www.geeksforgeeks.org/architecture-of-internet-of-things-iot/

Εικόνα 2: Stages of IoT Architecture[4]

2.2.2.3 Ασφάλεια στο IoT

Το IoT αντιμετωπίζει σημαντικές προκλήσεις ασφάλειας λόγω της ετερογενούς φύσης του, των περιορισμένων πόρων των συσκευών και της ευρείας κλίμακας ανάπτυξής του. Όπως αναφέρουν οι Suo et al. [19] , τα κύρια ζητήματα ασφάλειας στο IoT περιλαμβάνουν την εμπιστευτικότητα, την ακεραιότητα και την διαθεσιμότητα των δεδομένων, καθώς και την προστασία της ιδιωτικότητας.

Η εργασία αυτή επικεντρώνεται στην αντιμετώπιση επιθέσεων τύπου DDoS, οι οποίες αποτελούν μία από τις σοβαρότερες απειλές για τα συστήματα IoT. Οι Mishra και Pandya [20] υπογραμμίζουν ότι συχνά υπάρχει πρόβλημα ασφαλειας λόγω κόστους και ενεργειακών περιορισμών, εγχύσεις κακόβουλου κώδικα και υποκλοπές. Για την αντιμετώπιση αυτών των προκλήσεων, συστήματα ανίχνευσης εισβολών που αξιοποιούν τεχνικές ML και Deep Learning (DL) έχουν αναδειχθεί ως ελπιδοφόρες λύσεις, επιτρέποντας ανίχνευση και μετριασμό ανωμαλιών σε πραγματικό χρόνο. Καθώς η υιοθέτηση του IoT επεκτείνεται, η ανάπτυξη ανθεκτικών πλαισίων ασφαλείας θα είναι καθοριστική για την προστασία των δικτύων και την εμπιστοσύνη στις εφαρμογές IoT.

2.2.2.4 Προκλήσεις και Μελλοντικές Τάσεις

Παρόλο που οι τεχνολογίες IoT συνεχίζουν να εξελίσσονται, υπάρχουν ήδη πολλά υποσχόμενα πρωτότυπα και εμπορικές εφαρμογές, όμως η πλήρης ενσωμάτωσή τους στο σύγχρονο τεχνολογικό οικοσύστημα εξακολουθεί να αντιμετωπίζει σημαντικές προκλήσεις [17]. Η ασφάλεια παραμένει βασικό ζήτημα, καθώς η αυξημένη διασύνδεση εκατομμυρίων συσκευών δημιουργεί μεγαλύτερη επιφάνεια επιθέσεων. Παράλληλα, ζητήματα όπως η έλλειψη τυποποιημένων πρωτοκόλλων επικοινωνίας, η διαχείριση ενεργειακών πόρων σε χαμηλής ισχύος συσκευές και η δυσκολία αναβάθμισης λογισμικού σε απομακρυσμένα σημεία, δυσχεραίνουν τη μαζική υιοθέτηση. Σε επίπεδο δικτύου, το scalability αποτελεί κρίσιμο παράγοντα, καθώς οι παραδοσιακές αρχιτεκτονικές δεν επαρκούν για την υποστήριξη δισεκατομμυρίων συσκευών με χαμηλό latency και υψηλή αξιοπιστία.

Σε αυτό το πλαίσιο, μελλοντικές τάσεις περιλαμβάνουν την ενσωμάτωση τεχνητής νοημοσύνης απευθείας σε edge συσκευές (edge AI), την αξιοποίηση του FL για τη διατήρηση της ιδιωτικότητας, καθώς και την ενίσχυση της ασφάλειας μέσω αποκεντρωμένων τεχνολογιών, όπως το blockchain. Επίσης, η σταδιακή μετάβαση από τα 5G στα 6G δίκτυα, όπως αναλύθηκε σε προηγούμενη ενότητα, αναμένεται να υποστηρίξει τις αυξανόμενες

[4] Πηγή: https://www.geeksforgeeks.org/architecture-of-internet-of-things-iot/

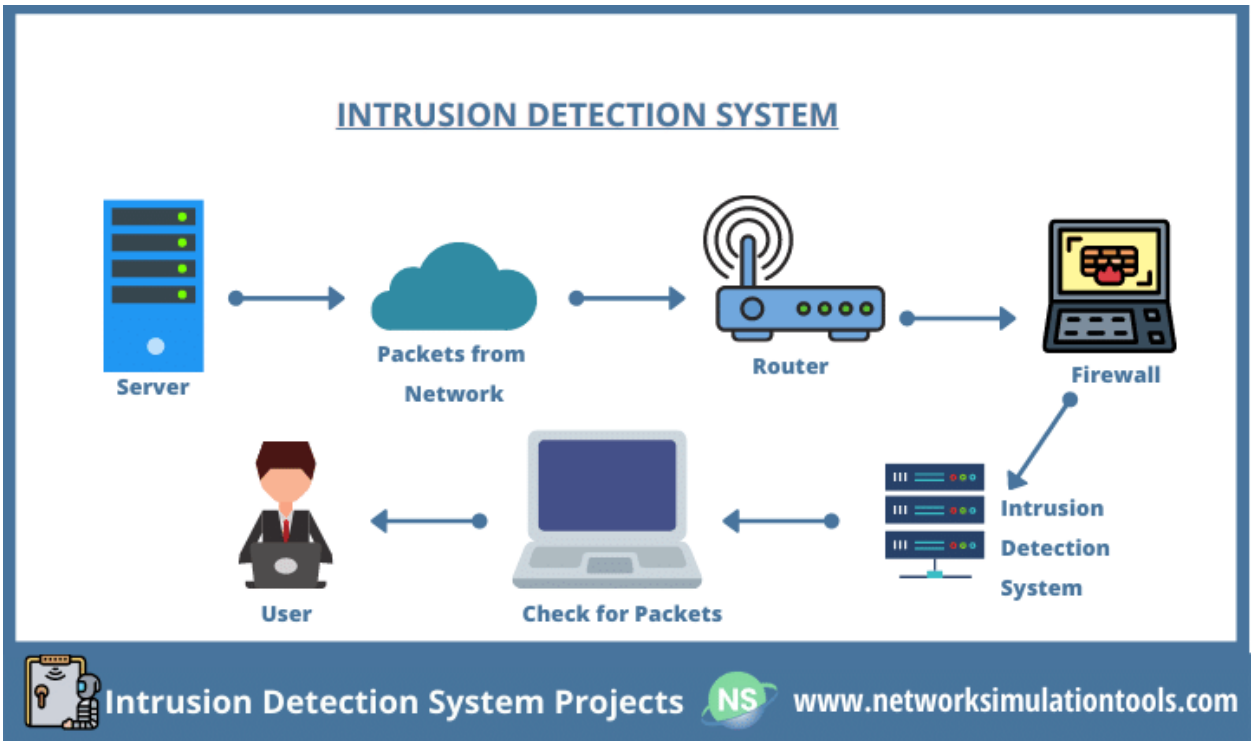
απαιτήσεις του IoT, επιτρέποντας ακόμη πιο εξελιγμένες εφαρμογές σε τομείς όπως η υγειονομική περίθαλψη, οι έξυπνες πόλεις και η αυτόνομη κινητικότητα.

2.3 Συστήματα Ανίχνευσης Εισβολής (IDS)

Το IDS αποτελεί κρίσιμο στοιχείο της αρχιτεκτονικής ασφαλείας, καθώς έχει ως στόχο τον εντοπισμό διαφόρων μορφών κακόβουλης δραστηριότητας. Επίσης το IDS παρακολουθεί διαρκώς την κυκλοφορία στο δίκτυο, εντοπίζοντας ύποπτες ενέργειες ή παραβιάσεις πολιτικών ασφαλείας, παρέχοντας στους διαχειριστές τη δυνατότητα έγκαιρης αναγνώρισης και αντίδρασης σε τρέχουσες απειλές [21].

Η ανίχνευση απειλών στα IDS βασίζεται κυρίως σε δύο προσεγγίσεις: την ανίχνευση βάσει υπογραφών και την ανίχνευση βάσει ανωμαλιών [22]. Η πρώτη, συγκρίνει τη δικτυακή κίνηση με μία βάση δεδομένων γνωστών υπογραφών επιθέσεων. Αν και προσφέρει υψηλή ακρίβεια στον εντοπισμό ήδη καταγεγραμμένων απειλών, αποτυγχάνει στην αναγνώριση νέων ή άγνωστων επιθέσεων και απαιτεί συνεχή ενημέρωση και υψηλό υπολογιστικό κόστος [23]. Αντιθέτως, τα IDS ανίχνευσης ανωμαλιών, που είναι η προσέγγιση που χρησιμοποιούμαι στην εργασία μας, δημιουργούν ένα πρότυπο φυσιολογικής συμπεριφοράς και εντοπίζουν αποκλίσεις από αυτό, καθιστώντας τα ικανά να ανιχνεύουν μη καταγεγραμμένες απειλές [24] . Ωστόσο, η προσέγγιση αυτή συνήθως συνοδεύεται από υψηλά ποσοστά ψευδώς θετικών αποτελεσμάτων, περιορίζοντας τη λειτουργικότητά της σε πραγματικές συνθήκες.

Επιπλέον, τα συστήματα IDS διακρίνονται και ως προς τη θέση τους στο δίκτυο σε συστήματα βάσει δικτύου (Network-based IDS - NIDS) και συστήματα βάσει κεντρικών υπολογιστών (Host-based IDS - HIDS) [25]. Συγκεκριμένα τα NIDS παρακολουθούν τη ροή δεδομένων σε πραγματικό χρόνο και είναι κατάλληλα για την επιτήρηση ευρέων δικτυακών υποδομών, καθώς αναλύουν την κυκλοφορία για πιθανές ανωμαλίες ή ενδείξεις εισβολής σε πολλαπλά σημεία του δικτύου. Αποτελούνται, συνήθως, από συνδυασμό υλικού (όπως αισθητήρες) και λογισμικού (κονσόλες διαχείρισης), επιτρέποντας την παρακολούθηση πακέτων δικτύου σε πραγματικό χρόνο. Από την άλλη πλευρά, τα HIDS είναι εγκατεστημένα σε μεμονωμένους υπολογιστές ή διακομιστές και εστιάζουν στην παρακολούθηση τοπικών συμβάντων, όπως μεταβολές σε αρχεία συστήματος, μητρώα ή ρυθμίσεις ασφαλείας. Αν και λειτουργούν σε επίπεδο συσκευής και όχι σε ολόκληρο το δίκτυο, τα HIDS μπορούν να προσφέρουν πιο λεπτομερή εικόνα, καθώς έχουν πρόσβαση και σε κρυπτογραφημένα δεδομένα, παρέχοντας έτσι δυνατότητες εντοπισμού επιθέσεων που δεν είναι ορατές από τα NIDS.



Εικόνα 3: Intrusion Detection System[5]

2.4 Μηχανική Μάθηση

Η Μηχανική Μάθηση [6] αποτελεί έναν από τους σημαντικότερους κλάδους της επιστήμης των υπολογιστών, με αντικείμενο τη σχεδίαση αλγορίθμων που επιτρέπουν σε ένα σύστημα να μαθαίνει από εμπειρικά δεδομένα. Μέσω της ανάλυσης μεγάλου όγκου δεδομένων, οι αλγόριθμοι μαθαίνουν να εντοπίζουν πρότυπα και συσχετίσεις, ώστε να μπορούν να πραγματοποιούν προβλέψεις, να ταξινομούν ή να ομαδοποιούν νέες πληροφορίες.

2.4.1 Βασικές Κατηγορίες Μηχανικής Μάθησης

Οι τεχνικές ML διακρίνονται σε τρεις βασικές κατηγορίες, ανάλογα με τον τρόπο εκπαίδευσης του μοντέλου:

- **Επιβλεπόμενη Μάθηση (Supervised Learning):** Στη συγκεκριμένη προσέγγιση, το σύστημα εκπαιδεύεται με τη χρήση ενός συνόλου δεδομένων όπου κάθε είσοδος συνοδεύεται από την αντίστοιχη ετικέτα (label). Στόχος του αλγορίθμου είναι να μάθει να προβλέπει τις εξόδους για νέες, άγνωστες εισόδους. Χαρακτηριστικοί

[5] Πηγή: https://networksimulationtools.com/intrusion-detection-system-projects/
[6] https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

αλγόριθμοι αυτής της κατηγορίας είναι τα Decision Trees και τα Support Vector Machines (SVM).

- **Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning):** Εδώ, τα δεδομένα εκπαίδευσης δεν φέρουν προκαθορισμένες ετικέτες. Οι αλγόριθμοι στοχεύουν στον εντοπισμό υποκείμενων δομών και συσχετίσεων, ταξινομώντας τα δεδομένα σε ομάδες με βάση κοινά χαρακτηριστικά. Τυπικές εφαρμογές περιλαμβάνουν clustering τεχνικές όπως το K-means.
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Σε αυτό το πλαίσιο, ο πράκτορας βρίσκεται μέσα σε ένα περιβάλλον και επιλέγει ενέργειες. Κάθε φορά που εκτελεί μια ενέργεια, το περιβάλλον του επιστρέφει μια ανταμοιβή (reward) ή μια ποινή (penalty). Στόχος του πράκτορα είναι να μάθει μια στρατηγική, δηλαδή ένα σύστημα κανόνων για το ποια ενέργεια να επιλέξει σε κάθε κατάσταση, έτσι ώστε μακροπρόθεσμα να μεγιστοποιεί τη σωρευτική ανταμοιβή. Μέσα από επαναλαμβανόμενες αλληλεπιδράσεις και αξιολόγηση των ανταμοιβών, ο πράκτορας ενημερώνει σταδιακά την πολιτική του, ενισχύοντας τις ενέργειες που οδηγούν σε θετικά αποτελέσματα.

2.4.2 Αλγόριθμοι Επιβλεπόμενης Μάθησης για IDS

Στο πεδίο της ανίχνευσης ανωμαλιών δικτυακής κίνησης, τα Τυχαία Δάση (Random Forests) αποτελούν μία από τις πλέον διαδεδομένες προσεγγίσεις [26]. Η μέθοδος αυτή συγκροτεί ένα σύνολο δέντρων αποφάσεων, κάθε ένα από τα οποία εκπαιδεύεται σε τυχαίο υποσύνολο τόσο των δειγμάτων όσο και των χαρακτηριστικών του συνόλου δεδομένων. Το τελικό αποτέλεσμα προκύπτει από τον συνδυασμό των επιμέρους προβλέψεων, μειώνοντας δραστικά τον κίνδυνο υπερπροσαρμογής και περιορίζοντας τα ψευδή θετικά σήματα. Στην πράξη, τα Τυχαία Δάση αξιοποιούν στατιστικά χαρακτηριστικά όπως ο όγκος πακέτων, τα inter-arrival times, αλλά και μεταδεδομένα (θύρες, πρωτόκολλα) για να διαχωρίσουν κανονικές ροές από ύποπτες ή κακόβουλες.

Οι SVM προσφέρουν ένα συμπληρωματικό εργαλείο, εστιάζοντας στην εύρεση της βέλτιστης υπερεπιφάνειας που διαχωρίζει τις δύο κλάσεις με το μέγιστο περιθώριο .Όταν η διάκριση δεν είναι γραμμική, η χρήση κατάλληλων πυρήνων επιτρέπει στους SVM να χειριστούν πολύπλοκες δομές στο χώρο των χαρακτηριστικών, ενισχύοντας την ευαισθησία τους σε δυσδιάκριτες απειλές. Παρά το γεγονός ότι απαιτούν εξειδικευμένο “tuning” υπερπαραμέτρων ,σταθεροποιούν τα ποσοστά ψευδών θετικών και συχνά αποδίδουν υψηλή ακρίβεια σε πραγματικά περιβάλλοντα [27].

Με την έλευση του DL, τα Νευρωνικά Δίκτυα, όπως τα MLP, τα CNN και τα Επαναληπτικά/ LSTM, έχουν αρχίσει να κυριαρχούν στην ανίχνευση ανωμαλιών. Αυτές οι

αρχιτεκτονικές μαθαίνουν αυτόματα σύνθετα μοτίβα απευθείας από ακατέργαστα δεδομένα χωρίς να απαιτείται χειροκίνητη επιλογή χαρακτηριστικών. Επιπλέον, οι autoencoders χρησιμοποιούνται συχνά σε μη επιβλεπόμενο πλαίσιο: το δίκτυο εκπαιδεύεται να ανακατασκευάζει κανονικά δείγματα κι αν παρουσιάσει μεγάλη απόκλιση ανακατασκευής, τότε η κίνηση χαρακτηρίζεται ως αναξιοπαθής ή κακόβουλη.

2.4.3 Κεντρική Μηχανική Μάθηση

Το CL αποτελεί την πλέον παραδοσιακή και διαδεδομένη προσέγγιση στην εκπαίδευση αλγορίθμων τεχνητής νοημοσύνης [28]. Όλα τα δεδομένα από διαφορετικές πηγές (όπως απομακρυσμένοι αισθητήρες, συσκευές IoT ή χρήστες) συγκεντρώνονται και αποθηκεύονται σε έναν κεντρικό διακομιστή. Η εκπαίδευση του μοντέλου πραγματοποιείται εξ ολοκλήρου στο κεντρικό αυτό σημείο, αξιοποιώντας το ενοποιημένο σύνολο δεδομένων.

Η διαδικασία περιλαμβάνει αρχικά τη μεταφορά των δεδομένων στο κέντρο, όπου εφαρμόζονται τεχνικές προεπεξεργασίας, όπως καθαρισμός, κανονικοποίηση και επιλογή χαρακτηριστικών. Ακολούθως, το επιλεγμένο μοντέλο εκπαιδεύεται συνολικά στο σύνολο των δεδομένων και αξιολογείται ως προς την ακρίβειά του μέσω διαδικασιών επικύρωσης και ρύθμισης υπερπαραμέτρων.

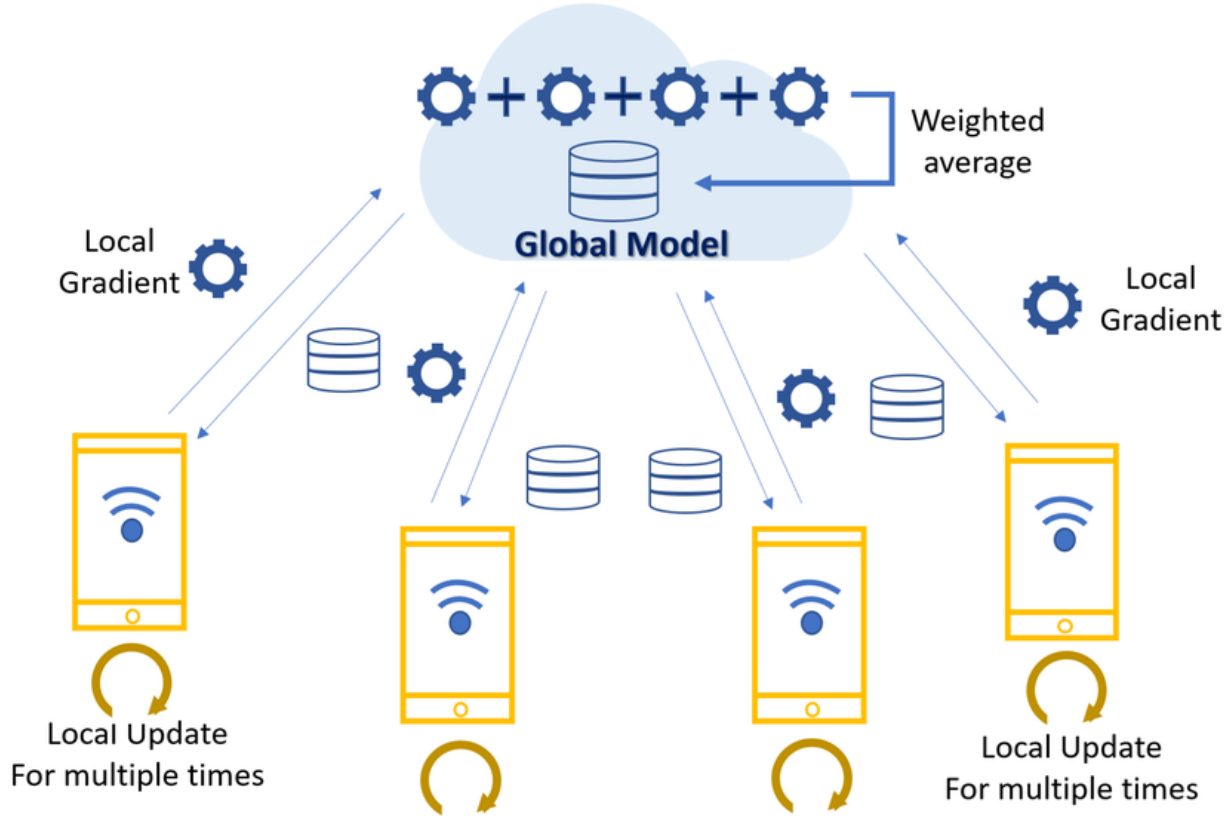
Η προσέγγιση αυτή παρουσιάζει ορισμένα σημαντικά πλεονεκτήματα. Η συγκέντρωση μεγάλου όγκου δεδομένων επιτρέπει την εκπαίδευση ισχυρών και ακριβών μοντέλων, ενώ παράλληλα απλοποιείται η διαχείριση του συστήματος, καθώς όλες οι λειτουργίες εκτελούνται εντός ενός ενιαίου υπολογιστικού περιβάλλοντος. Επιπλέον, διευκολύνεται η ανάπτυξη και η αναπαραγωγιμότητα των πειραμάτων.

Ωστόσο, το CL συνοδεύεται και από σοβαρούς περιορισμούς, κυρίως ως προς την προστασία της ιδιωτικότητας. Η ανάγκη μεταφοράς ευαίσθητων δεδομένων σε έναν κεντρικό κόμβο δημιουργεί προκλήσεις συμμόρφωσης, καθώς και ανησυχίες για πιθανές παραβιάσεις ασφαλείας. Παράλληλα, η μεταφορά μεγάλων ποσοτήτων δεδομένων αυξάνει το υπολογιστικό και δικτυακό κόστος, ενώ καθιστά το κεντρικό σύστημα εν δυνάμει σημείο αποτυχίας (single point of failure).

Η προσέγγιση CL, αν και εξακολουθεί να χρησιμοποιείται ευρέως, παρουσιάζει δυσκολίες εφαρμογής σε περιβάλλοντα με έντονη κατανομή δεδομένων, όπως IoT ή τα δίκτυα 5G. Σε τέτοιες περιπτώσεις, εναλλακτικές αποκεντρωμένες μέθοδοι, όπως το FL, προσφέρουν λύσεις που συνδυάζουν απόδοση με σεβασμό στην ιδιωτικότητα. Στην επόμενη ενότητα, πραγματοποιείται η επεξήγηση του FL και συγκριτική αποτίμηση μεταξύ CL και FL στο πλαίσιο της ανίχνευσης κυβερνοαπειλών.

2.5 Ομοσπονδιακή Μάθηση (FL)

Μία από τις πιο σημαντικές τεχνολογίες στον τομέα της ιδιωτικότητας στην πληροφορική είναι το FL, ένα κατανεμημένο μοντέλο Μηχανικής Μάθησης. Η διαδικασία ξεκινά με έναν κεντρικό server που στέλνει ένα αρχικοποιημένο global μοντέλο στους πελάτες (clients), οι οποίοι εκπαιδεύουν τοπικά μοντέλα με βάση τα τοπικά τους δεδομένα. Στη συνέχεια, οι πελάτες επιστρέφουν τις ενημερωμένες παραμέτρους των μοντέλων αντί να αποστείλουν τα ίδια τα δεδομένα. Η διαδικασία αυτή επαναλαμβάνεται σε διαδοχικούς γύρους (rounds) μέχρι το global μοντέλο να επιτύχει ικανοποιητική ακρίβεια [29] . Το FL εφαρμόζεται σε πολλούς τομείς, όπως τα έξυπνα κινητά, το διαδίκτυο των πραγμάτων (IoT), τη βιομηχανία και την υγειονομική περίθαλψη. Ενδεικτικά, χρησιμοποιείται για την πρόβλεψη κειμένου σε πληκτρολόγια (Google Gboard), την ανάλυση ιατρικών δεδομένων, καθώς και για τη βελτίωση έξυπνων συστημάτων σε σπίτια, ρομπότ και οχήματα, διασφαλίζοντας πάντα την ιδιωτικότητα των χρηστών [30].

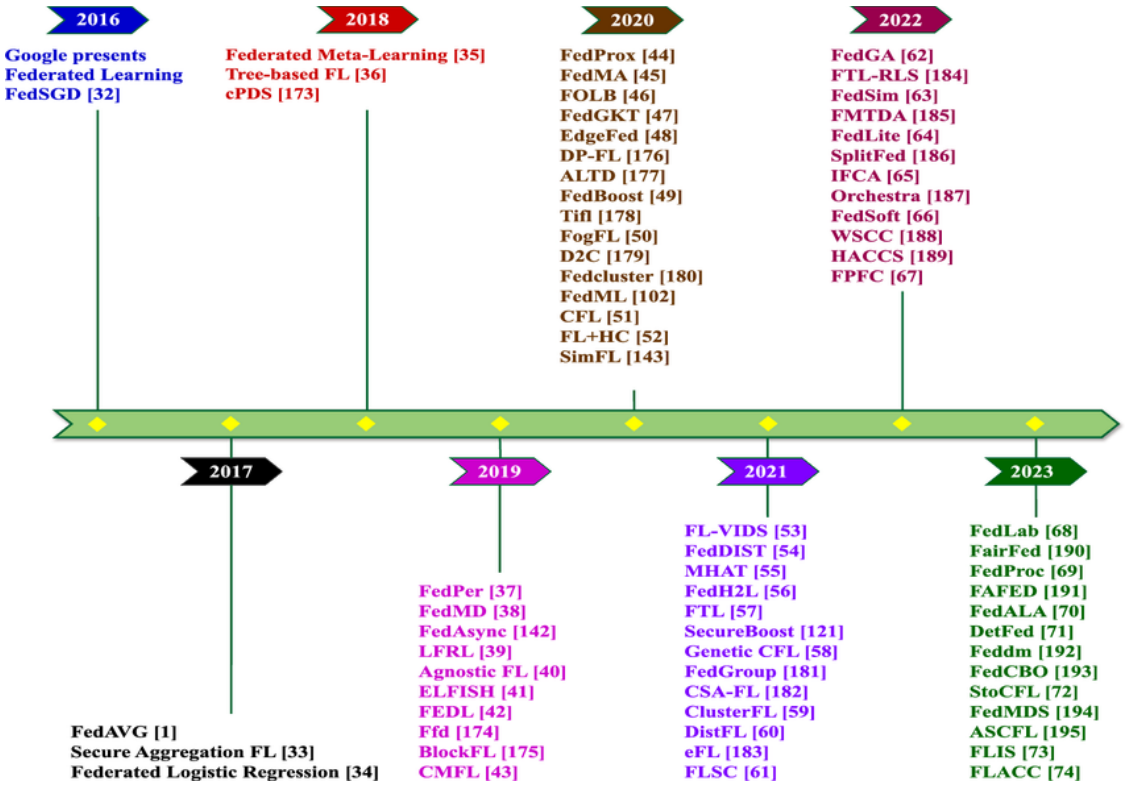


Εικόνα 4: Illustration of FL framework proposed by Google[7]

[7] https://www.researchgate.net/figure/Illustration-of-FL-framework-proposed-by-Google_fig3_373322254

2.5.1 Ιστορική Αναδρομή

Η έννοια του FL προτάθηκε για πρώτη φορά από την Google το 2016, κυρίως για να κάνει τους χρήστες κινητών τηλεφώνων Android να ενημερώνουν τα μοντέλα τους τοπικά χωρίς να αποκαλύπτουν προσωπικά δεδομένα [31]. Αυτό το νέο βήμα της μηχανικής μάθησης σηματοδότησε μια σημαντική πρόοδο στον τομέα, ιδιαίτερα σε ζητήματα που σχετίζονται με το απόρρητο και την ασφάλεια των δεδομένων. Έκτοτε, η τεχνολογία έχει εξελιχθεί ώστε να καλύπτει προβλέψεις emoji [32], ανθρώπινη συμπεριφορά [33], και πρόβλεψη διαδρομής [34] στο κομμάτι της κινητής συσκευής.



Εικόνα 5: Timeline evolution of federated learning[8]

2.5.2 Κατηγορίες FL

Σύμφωνα με τους Fan Y et al. [4], η Ομόσπονδη Μάθηση μπορεί να χωριστεί σε τρεις κατηγορίες:

1. **Horizontal Federated Learning**, η οποία εφαρμόζεται στα σενάρια όπου δύο σύνολα δεδομένων είναι διαφορετικά σε δείγματα και ίδια στον χώρο χαρακτηριστικών. Για παράδειγμα, δύο νοσοκομεία που συλλέγουν τις ίδιες πληροφορίες για τους ασθενείς (όπως ηλικία, φύλο, εξετάσεις) αλλά για διαφορετικούς ανθρώπους. Τα δεδομένα

[8] Πηγή: https://www.researchgate.net/figure/Timeline-evolution-of-federated-learning_fig1_385709639

είναι «οριζόντια κατανεμημένα» και η εκπαίδευση γίνεται τοπικά χωρίς ανταλλαγή προσωπικών πληροφοριών [35], [36].

2. **Vertical Federated Learning,** η οποία εφαρμόζεται στα σενάρια όπου δύο σύνολα δεδομένων είναι ίδια σε δείγματα αλλά διαφορετικά στον χώρο χαρακτηριστικών. Για παράδειγμα, ένα χρηματοπιστωτικό ίδρυμα και μια ασφαλιστική εταιρεία μπορεί να έχουν κοινούς πελάτες αλλά να συλλέγουν διαφορετικού τύπου δεδομένα (π.χ. οικονομικά και υγειονομικά). Το VFL επιτρέπει την από κοινού εκπαίδευση μοντέλων χωρίς ανταλλαγή πλήρων προφίλ πελατών [37].
3. **Federated Tranfer Learning**, η οποία εφαρμόζεται στα σενάρια όπου δύο σύνολα δεδομένων είναι διαφορετικά όχι μόνο σε δείγματα αλλά και στον χώρο χαρακτηριστικών [4].

2.5.3 Ομοσπονδιακοί Αλγόριθμοι

Μία από τις πιο διαδεδομένες και βασικές προσεγγίσεις του FL είναι ο αλγόριθμος **Federated Averaging (FedAvg)**, ο οποίος χρησιμοποιήθηκε και στο πλαίσιο της παρούσας εργασίας. Ο FedAvg αποτελεί μια επέκταση του κλασικού αλγορίθμου στοχαστικής καθόδου κλίσης (SGD) σε κατανεμημένο περιβάλλον [38]. Κάθε πελάτης εκπαιδεύει τοπικά το μοντέλο του για μερικούς γύρους χρησιμοποιώντας τα δικά του δεδομένα και στη συνέχεια αποστέλλει τις ενημερωμένες παραμέτρους σε έναν κεντρικό server. Ο server συγκεντρώνει, μέσω μέσου όρου, τις ενημερώσεις και δημιουργεί ένα νέο global μοντέλο, το οποίο στη συνέχεια διανέμεται εκ νέου σε όλους τους πελάτες για να συνεχιστεί η εκπαίδευση [31].

Algorithm 1: FederatedAveraging

The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:

initialize w_0

for each round t = 1, 2, ... **do**

$m \leftarrow \max(C \cdot K, 1)$

$S_t \leftarrow$ (random set of m clients)

for each client $k \in S_t$ **in parallel do**

$w^k_{t+1} \leftarrow \text{ClientUpdate}(k, w_t)$

$w_{t+1} \leftarrow \Sigma^K_{k=1} (n_k/n)\, w^k_{k+1}$

ClientUpdate(k, w): // Run on client k

$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)

for each local epoch i from 1 to E **do**

for batch b ∈ $\mathcal{B}$ **do**

w ← w - η∇ℓ(w; b)

return w to server

Ένας ακόμη πολύ γνωστός αλγόριθμος είναι ο FedProx που ενσωματώνει έναν επιπλέον όρο κανονικοποίησης, σε αντίθεση με το FedAvg, για τη διαχείριση δεδομένων που δεν ανήκουν στο IID. Αυτός ο όρος κανονικοποίησης εξισορροπεί την ανταλλαγή μεταξύ παγκόσμιων και τοπικών πληροφοριών στις ενημερώσεις του μοντέλου [39].

Algorithm 2: FedProx

Input: K, T, μ, γ, w^0, N, p_k, k = 1, ⋯, N

for t = 0, ⋯, T - 1 **do**

Server selects a subset S_t of K devices at random (each device k is chosen with probability p_k)

Server sends w^t to all chosen devices

Each chosen device k ∈ S_t finds a w_k^{t+1} which is a γ_k^t-inexact minimizer of:

$w_k^{t+1} \approx \arg\min_w h_k(w; w^t) = F_k(w) + \mu/2 \, ||w - w^t||^2$

Each device k ∈ S_t sends w_k^{t+1} back to the server

Server aggregates the w's as $w^{t+1} = 1/K \sum_{k \in S_t} w_k^{t+1}$

end for

Εκτός από τον FedProx έχουν προταθεί και άλλες τεχνικές για να βελτιώσουν την απόδοση του FL. Στο πλαίσιο αυτό, η μελέτη από Reddi et al. [40] παρουσιάζει τους FedAdam, FedAdagrad και FedYogi. Συγκεκριμένα, ο FedAdam δίνει «βάρος» τόσο στις προηγούμενες όσο και στις πιο πρόσφατες ενημερώσεις των μοντέλων. Ο FedAdagrad δίνει περισσότερη σημασία στους πελάτες που έχουν μεγάλες διαφορές από το κεντρικό μοντέλο, ενώ ο FedYogi προσπαθεί να διατηρήσει σταθερότητα στη μάθηση, λαμβάνοντας υπόψη και την κατεύθυνση της διαφοράς. Αυτές οι μέθοδοι είναι χρήσιμες ειδικά όταν τα δεδομένα είναι άνισα κατανεμημένα ή ανομοιογενή.

Algorithm: FedAdagrad, FedYogi, and FedAdam

Initialization: x_0, $v_{-1} > \tau^2$, decay parameters $\beta_1, \beta_2 \in [0, 1)$
for t = 0, ⋯, T - 1 **do**
  Sample subset $\mathcal{S}$ of clients
  $x^t_{i,0} = x_t$
  for each client $i \in \mathcal{S}$ **in parallel do**
    for k = 0, ⋯, K - 1 **do**
      Compute an unbiased estimate $g_{it,k}$ of $\nabla F_i(x_{it,k})$
      $x^t_{i,k+1} = x^t_{i,k} - \eta g^t_{i,k}$
  $\Delta^t_i = x^t_{i,K} - x_t$
  $\Delta_t = \beta_1 \Delta_{t-1} + (1 - \beta_1)(1/|\mathcal{S}|) \sum_{i \in \mathcal{S}} \Delta^t_i)$
  $v_t = v_{t-1} + \Delta_t^2$ (**FedAdagrad**)
  $v_t = v_{t-1} - (1 - \beta_2)\Delta_t^2 \, \text{sign}(v_{t-1} - \Delta_t^2)$ (**FedYogi**)
  $v_t = \beta_2 v_{t-1} + (1 - \beta_2)\Delta_t^2$ (**FedAdam**)
  $x_{t+1} = x_t + \eta(\Delta_t / \sqrt{(v_t + \tau)})$
end for

2.5.4 FL VS CL

Η κύρια διαφορά ανάμεσα σε FL και CL βρίσκεται στον τρόπο συλλογής και επεξεργασίας των δεδομένων. Στο CL, όλα τα δεδομένα μεταφέρονται σε έναν κεντρικό server για την εκπαίδευση του μοντέλου, γεγονός που δημιουργεί κινδύνους για την ιδιωτικότητα και την ασφάλεια, ιδιαίτερα όταν τα δεδομένα είναι ευαίσθητα, όπως ιατρικές ή προσωπικές πληροφορίες. Αντίθετα, στο FL, η εκπαίδευση γίνεται τοπικά στις συσκευές των χρηστών και μόνο οι ενημερώσεις των μοντέλων αποστέλλονται στον κεντρικό server, μειώνοντας έτσι την ανάγκη για μεταφορά προσωπικών δεδομένων. Αυτό καθιστά την FL πιο φιλική προς την προστασία προσωπικών δεδομένων και κατάλληλη για εφαρμογές σε περιβάλλοντα όπως το IoT και η υγειονομική περίθαλψη [41], [42].

Επιπλέον, υπάρχουν πρακτικές διαφορές ανάμεσα στα δύο μοντέλα που σχετίζονται με την απόδοση και την αποδοτικότητα. Στο CL η εκπαίδευση του μοντέλου μπορεί να φτάσει υψηλότερα επίπεδα ακρίβειας, καθώς αξιοποιεί το πλήρες και ενοποιημένο σύνολο δεδομένων. Αντίθετα, το FL μπορεί να υποφέρει από μειωμένη απόδοση λόγω της ανομοιομορφίας των δεδομένων (non-IID) και των περιορισμένων τοπικών δεδομένων σε κάθε πελάτη. Επίσης, το FL απαιτεί μεγαλύτερους υπολογιστικούς πόρους στις τοπικές

συσκευές, γεγονός που μπορεί να είναι περιοριστικό για συσκευές με χαμηλή ισχύ. Τέλος, παρουσιάζει αυξημένο επικοινωνιακό φορτίο, καθώς απαιτείται συνεχής συγχρονισμός και ανταλλαγή παραμέτρων μεταξύ των πελατών και του server, κάτι που ενδέχεται να επηρεάσει την ταχύτητα και την αποδοτικότητα της εκπαίδευσης [41], [42].

Table 1: Σύγκριση FL με CL

Κριτήριο	Federated Learning	Centralized Learning
Τοποθεσία Δεδομένων	Τα δεδομένα παραμένουν στις τοπικές συσκευές των πελατών	Όλα τα δεδομένα συλλέγονται και αποθηκεύονται σε έναν κεντρικό server
Ιδιωτικότητα	Υψηλή προστασία ιδιωτικότητας	Υψηλός κίνδυνος παραβίασης ιδιωτικότητας
Ακρίβεια Μοντέλου	Πιθανώς μικρότερη λόγω μη ομοιογενών δεδομένων (non-IID)	Συνήθως υψηλότερη λόγω πρόσβασης σε όλα τα δεδομένα
Υπολογιστικό Κόστος	Κατανεμημένο στις συσκευές – απαιτεί ισχυρές τοπικές υποδομές	Συγκεντρωμένο στον server
Επικοινωνιακό Φορτίο	Αυξημένο – συνεχής ανταλλαγή παραμέτρων	Μικρότερο – η εκπαίδευση γίνεται στον server
Ανθεκτικότητα σε Αποσυνδέσεις	Πιο ανεκτικό σε προσωρινές αποσυνδέσεις πελατών	Εξαρτάται πλήρως από τη διαθεσιμότητα του server
Καταλληλότητα για IoT/Edge	Ιδανικό	Όχι κατάλληλο λόγω της ανάγκης μεταφοράς όλων των δεδομένων

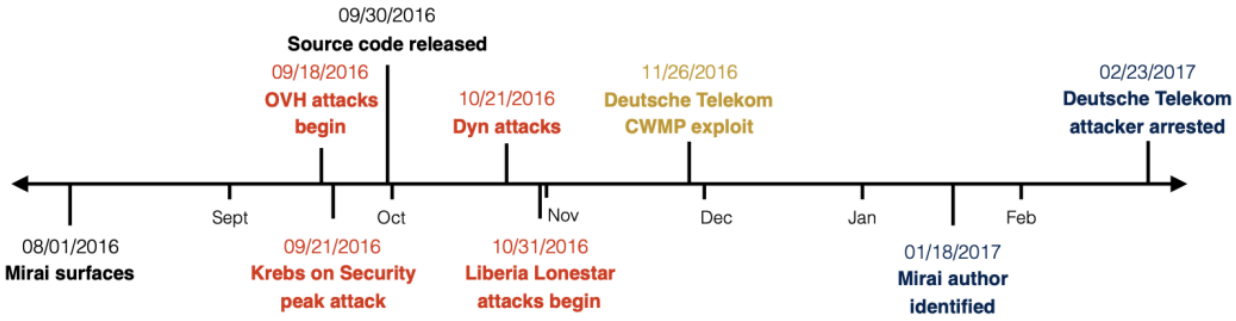
2.6 Περιγραφή Κυβερνοεπιθέσεων

Μια από τις σημαντικότερες απειλές για τα σύγχρονα δικτυακά συστήματα, ειδικά σε περιβάλλοντα όπως το IoT και τα 5G δίκτυα, είναι οι επιθέσεις τύπου DDoS. Οι επιθέσεις αυτές στοχεύουν στην εξάντληση των πόρων ενός υπολογιστικού συστήματος, όπως του επεξεργαστή, με σκοπό να καταστήσει την υπηρεσία μη διαθέσιμη [43] . Στο πλαίσιο της παρούσας εργασίας, χρησιμοποιείται το N-BaIoT dataset που περιλαμβάνει επιθέσεις τύπου Mirai και Bashlite, οι οποίες είναι γνωστές για την εκμετάλλευση ευάλωτων IoT συσκευών με στόχο τη δημιουργία botnets [44]. Εμείς στο πειραματικό κομμάτι αξιοποιήσαμε μόνο τις Mirai επιθέσεις.

2.6.1 Mirai Botnet

Αρχικά για να κατανοήσουμε το Mirai Botnet πρέπει να κατανοήσουμε τι είναι το botnet. Το κακόβουλο botnet [45] είναι ένα δίκτυο παραβιασμένων υπολογιστών που ονομάζονται «Bots» και βρίσκονται υπό τον τηλεχειρισμό ενός ανθρώπινου χειριστή που ονομάζεται «Botmaster». Ο όρος «Bot» προέρχεται από τη λέξη «Robot» και όπως τα ρομπότ, τα bots έχουν σχεδιαστεί για να εκτελούν ορισμένες προκαθορισμένες λειτουργίες με αυτοματοποιημένο τρόπο. Με άλλα λόγια, τα μεμονωμένα bots είναι προγράμματα λογισμικού που εκτελούνται σε έναν κεντρικό υπολογιστή, επιτρέποντας στον botmaster να ελέγχει τις ενέργειες του κεντρικού υπολογιστή από απόσταση.

Ακολούθως το Mirai [46] αποτελεί μια από τις πιο γνωστές οικογένειες κακόβουλου λογισμικού τύπου botnet, σχεδιασμένη να μολύνει συσκευές του IoT και να τις χρησιμοποιεί για την εκτέλεση επιθέσεων DDoS. Πρόκειται για κακόβουλο λογισμικό τύπου «σκουλήκι», καθώς εξαπλώνεται αυτόνομα από συσκευή σε συσκευή, αξιοποιώντας ευπάθειες όπως προεπιλεγμένοι ή ασθενείς κωδικοί πρόσβασης. Οι πρώτες αναφορές για τη δραστηριότητά του καταγράφηκαν στα τέλη Αυγούστου του 2016, ωστόσο η ευρεία δημοσιότητα ήρθε τον Σεπτέμβριο της ίδιας χρονιάς, όταν το Mirai χρησιμοποιήθηκε σε μαζικές επιθέσεις εναντίον γνωστών στόχων, όπως ο ιστότοπος ασφαλείας KrebsOnSecurity και ο πάροχος υπηρεσιών cloud OVH. Η κυκλοφορία του κώδικα επέτρεψε την ευρεία παραμετροποίηση και επαναχρησιμοποίηση του κακόβουλου λογισμικού, οδηγώντας στη δημιουργία πολυάριθμων παραλλαγών.



Εικόνα 6 Mirai Botnet Timeline

Στο πλαίσιο της παρούσας εργασίας, οι τύποι επιθέσεων που αναπαριστώνται στο dataset N-BaIoT, το οποίο αξιοποιείται για την εκπαίδευση και αξιολόγηση του προτεινόμενου συστήματος ανίχνευσης, είναι ACK,SYN και UPD.

Συγκεκριμένα οι επιθέσεις Mirai ACK βασίζονται στην αποστολή πακέτων TCP ACK (Acknowledgment), τα οποία χρησιμοποιούνται μαζικά για να προκαλέσουν συμφόρηση ή υπερφόρτωση στο δίκτυο [47]. Στο πλαίσιο της τυπικής TCP επικοινωνίας, τα πακέτα ACK χρησιμοποιούνται για την επιβεβαίωση της επιτυχούς παραλαβής δεδομένων. Ωστόσο, στην περίπτωση αυτής της επίθεσης, ο εισβολέας αποστέλλει πολλά πακέτα ACK χωρίς να έχει προηγηθεί έγκυρη σύνδεση, γεγονός που τα καθιστά ανώφελα και ανεπιθύμητα για τον διακομιστή-στόχο. Η επίθεση στοχεύει στην υπερφόρτωση του διακομιστή επεξεργαζόμενος πολλά μη έγκυρα πακέτα ACK, εξαντλώντας έτσι τους πόρους του, όπως ο χρόνος της CPU και η μνήμη, κάτι που μπορεί να οδηγήσει σε μείωση της απόδοσης ή σε πλήρη διακοπή της υπηρεσίας προς τους κανονικούς χρήστες. Επιπλέον, η φύση των πακέτων ACK καθιστά την επίθεση δύσκολα ανιχνεύσιμη, καθώς η κίνηση μοιάζει με έγκυρη TCP επικοινωνία. Η μαζικότητα των πακέτων σε συνδυασμό με την απουσία έγκυρης TCP σύνδεσης είναι οι παράγοντες που καθιστούν την επίθεση ιδιαίτερα αποτελεσματική και καταστροφική σε περιβάλλοντα χωρίς ισχυρούς μηχανισμούς φιλτραρίσματος ή ανίχνευσης ανωμαλιών.

Αντίστοιχα, οι επιθέσεις Mirai SYN στέλνουν μεγάλο αριθμό αιτημάτων σύνδεσης, πακέτα SYN, χωρίς να ολοκληρώνει το πρωτόκολλο τριών βημάτων (three-way handshake) [47] ,[48]. Υπό κανονικές συνθήκες, μια σύνδεση TCP δημιουργείται μέσω μιας διαδικασίας τριών βημάτων όπου ο πελάτης στέλνει ένα πακέτο SYN στον διακομιστή. Στη συνέχεια, ο διακομιστής απαντά με ένα SYN-ACK και ο πελάτης ολοκληρώνει τη διαδικασία στέλνοντας ένα πακέτο ACK. Ωστόσο, σε μια επίθεση SYN, ο εισβολέας στέλνει πολλά πακέτα SYN , συχνά χρησιμοποιώντας πλαστογραφημένες διευθύνσεις IP, αλλά δεν απαντά στα πακέτα SYN-ACK που λαμβάνει ο διακομιστής σε αντάλλαγμα. Αυτό αναγκάζει τον διακομιστή να διατηρεί μισάνοιχτες συνδέσεις, σπαταλώντας τους πόρους του στη συντήρησή τους. Ως

αποτέλεσμα, ο διακομιστής υπερφορτώνεται και δεν μπορεί να επεξεργαστεί νέα νόμιμα αιτήματα σύνδεσης, με αποτέλεσμα την άρνηση υπηρεσίας για κανονικούς χρήστες.

Τέλος, οι επιθέσεις Mirai UDP κάνουν χρήση του πρωτοκόλλου UDP, αποστέλλοντας τεράστιο αριθμό πακέτων προς τυχαίους ή στοχευμένους προορισμούς, προκαλώντας συμφόρηση, ιδιαίτερα όταν δεν υπάρχουν μηχανισμοί ελέγχου ή φιλτραρίσματος [43], [47]. Το UDP είναι ένα πρωτόκολλο μετάδοσης δεδομένων που δεν ελέγχει την παράδοση πακέτων και δεν δημιουργεί σύνδεση πριν από την αποστολή δεδομένων. Στην περίπτωση του Mirai UDP, οι εισβολείς στέλνουν πολλά πακέτα UDP σε τυχαίες θύρες στον διακομιστή-στόχο ή τη συσκευή δικτύου. Όταν ο διακομιστής λαμβάνει αυτά τα πακέτα, προσπαθεί να τα επεξεργαστεί, συμπεριλαμβανομένου του ελέγχου των εισερχόμενων δεδομένων και της προσπάθειας να απαντήσει σε αιτήματα, εάν είναι απαραίτητο. Ως αποτέλεσμα, ο διακομιστής ξοδεύει και πάλι πόρους για την επεξεργασία και την απάντηση σε μεγάλους όγκους πλαστογραφημένων αιτημάτων.

2.6.2 Bashlite Botnet

Το Bashlite, γνωστό και με διάφορα ψευδώνυμα όπως Gafgyt, QBot, Lizkebab, Torlus και LizardStresser, αποτελεί ένα από τα παλαιότερα και πιο διαδεδομένα botnets κακόβουλης δραστηριότητας σε IoT συσκευές [49]. Το όνομά του προέρχεται από την εκμετάλλευση της ευπάθειας Shellshock (γνωστή και ως Bash Bug), η οποία του επιτρέπει την απομακρυσμένη εκτέλεση εντολών σε συστήματα βασισμένα στο UNIX, μέσω της χρήσης της κονσόλας Bash.

Η πρώτη εμφάνιση του Bashlite καταγράφεται το 2014, ενώ ήδη από το 2016 είχε μολύνει περισσότερες από 1 εκατομμύριο συσκευές παγκοσμίως. Η κατάσταση επιδεινώθηκε περαιτέρω όταν ο πηγαίος του κώδικας έγινε δημόσιος το 2017, οδηγώντας σε μεγάλη αύξηση παραλλαγών και επιθέσεων. Το λογισμικό στοχεύει κυρίως ευάλωτους δρομολογητές και IoT συσκευές και μόλις επιτευχθεί πρόσβαση, η συσκευή εντάσσεται αυτόματα σε ένα botnet, επικοινωνώντας με τον Command and Control (C2) server του επιτιθέμενου.[50]

Οι πιο συνηθισμένοι τύποι επιθέσεων που εκτελεί το Bashlite περιλαμβάνουν υπερφόρτωση του δικτύου με πακέτα δεδομένων, γνωστή ως flooding. Συγκεκριμένα, κατά τη διάρκεια επιθέσεων TCP ή UDP flooding, αποστέλλεται τεράστιος αριθμός πακέτων TCP ή UDP σε έναν διακομιστή-στόχο σε πολύ σύντομο χρονικό διάστημα, με στόχο να εξαντληθούν οι υπολογιστικοί πόροι του συστήματος και να διακοπεί η παροχή των υπηρεσιών του. Επιπλέον, το Bashlite εκμεταλλεύεται τις TCP σημαίες (flags), προκαλώντας σύγχυση στους μηχανισμούς διαχείρισης συνδέσεων του διακομιστή. Μία ακόμη τακτική που χρησιμοποιείται είναι η δημιουργία ανοικτών TCP συνδέσεων χωρίς ολοκλήρωση όπου

η σύνδεση ξεκινά αλλά δεν ολοκληρώνεται ποτέ. Αυτό έχει ως αποτέλεσμα ο διακομιστής να παραμένει σε αναμονή και να καταναλώνει πόρους για μεγάλο αριθμό μη εξυπηρετούμενων συνδέσεων.

Οι πιο πρόσφατες παραλλαγές του Bashlite έχουν εξελιχθεί περαιτέρω, ενσωματώνοντας τεχνικές από άλλα botnets όπως το Mirai. Αυτό περιλαμβάνει και πιο προηγμένες μεθόδους επίθεσης, όπως HTTP flooding, όπου υπερφορτώνονται web servers με ψεύτικα αιτήματα HTTP, και UDP flooding με ειδικά τροποποιημένα πακέτα, καθιστώντας τις επιθέσεις ακόμα πιο δύσκολες στον εντοπισμό και την αντιμετώπιση.

Η συγκεκριμένη κατηγορία επιθέσεων, αν και δεν αξιοποιήθηκε στην παρούσα πειραματική διαδικασία, συμπεριλαμβάνεται στο dataset και αποτελεί αξιόλογο πεδίο μελλοντικής μελέτης.

3. Σχετικές Εργασίες

Σε αυτό το κομμάτι της διπλωματικής εργασίας γίνεται μια ανασκόπηση σχετικών εργασιών που αξιοποιούν ανοιχτά σύνολα δεδομένων στον τομέα της ασφάλειας δικτύων 5G και του FL. Συγκεκριμένα, εξετάζονται τα χαρακτηριστικά αυτών των datasets, τα πειράματα που έχουν πραγματοποιηθεί βάσει αυτών, καθώς και οι μέθοδοι και τα μοντέλα που εφαρμόστηκαν.

3.1 Χρήση του dataset 5G-NIDD σε FL εφαρμογές

Αρχικά εξετάζουμε το dataset 5G-NIDD [9], που όπως περιγράφεται στο άρθρο "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network" [51], αποτελεί μια πλήρως επισημασμένη συλλογή δεδομένων που δημιουργήθηκε σε λειτουργικό 5G δοκιμαστικό δίκτυο. Περιλαμβάνει δεδομένα σε μορφές network capture (pcapng), NetFlow και encoded data (CSV) και περιέχει τόσο κανονική όσο και κακόβουλη κυκλοφορία, όπως UDPFlood, HTTPFlood, SlowrateDoS, TCPConnectScan, SYNScan, UDPScan, SYNFlood και ICMPFlood. Το dataset, το οποίο περιλαμβάνει και την επέκταση 5G-SliciNdd, χρησιμοποιείται ευρέως για την ανάπτυξη και αξιολόγηση συστημάτων ανίχνευσης εισβολών (IDS) σε δίκτυα 5G.

Η εργασία των I. Makris et al [6] προτείνει τέτοιο σύστημα για την προστασία των 5G δικτύων από κυβερνοεπιθέσεις, αξιοποιώντας το προαναφερόμενο dataset. Η μελέτη εξετάζει διάφορες στρατηγικές aggregation, όπως FedAvg, FedProx και FedAdam, για την εκπαίδευση τοπικών MLP μοντέλων, με τη χρήση προεπεξεργασίας μέσω SMOTE για την αντιμετώπιση του class imbalance. Τα αποτελέσματα έδειξαν ότι το FedAvg πέτυχε ακρίβεια 97.89% και F1-score 97.85%.

Παρομοίως, ο G. Singh et al. [38] αξιοποιεί το dataset 5G-NIDD για την ανάπτυξη ενός Fed-IDS, το οποίο στοχεύει στην αντιμετώπιση των προκλήσεων του class imbalance και των μη ανεξάρτητων και ομοιόμορφα κατανεμημένων δεδομένων (non-IID) στα 5G δίκτυα. Το σύστημα χρησιμοποιεί balanced sampling, feature space augmentation και knowledge distillation για την εκπαίδευση ενός Deep Neural Network (DNN) σε κατανεμημένους πελάτες. Η αξιολόγηση του συστήματος έδειξε έως και 1.9% υψηλότερη ακρίβεια σε σχέση με τα FedAvg και FedProx, καθώς και ταχύτερη σύγκλιση.

[9] https://ieee-dataport.org/documents/5g-nidd-comprehensive-network-intrusion-detection-dataset-generated-over-5g-wireless

3.2 FL για IoT επιθέσεις με βάση το N-BaIoT

Εκτός του 5G-NIDD το Detection of IoT Botnet Attacks (N-BaIoT)[10] παρέχει δεδομένα πραγματικού κόσμου από εννέα IoT συσκευές, τόσο υπό κανονικές συνθήκες όσο και κατά τη διάρκεια επιθέσεων botnet, όπως Mirai και Bashlite. Περιλαμβάνοντας δικτυακή κυκλοφορία και στατιστικά χαρακτηριστικά, αποτελεί έναν πολύτιμο πόρο για την ανάπτυξη συστημάτων ανίχνευσης εισβολών και τη μελέτη της ασφάλειας δικτύων IoT.

Στο paper από Rey et al. [52],που επιλέγει να αξιοποιήσει το dataset N-BaIoT ,με τη μέθοδο federated averaging και την κατανομή δεδομένων μεταξύ συσκευών, οι συσκευές εκπαιδεύουν federated μοντέλα χωρίς να μοιράζονται δεδομένα. Τα federated μοντέλα, που περιλαμβάνουν multi-layer perceptrons και autoencoders, επιτυγχάνουν ακρίβεια συγκρίσιμη με τα centralized μοντέλα.

Σε μια πρόσφατη μελέτη, οι Do et al.[35] αξιολογούν εμπειρικά την αποδοτικότητα του FL για την ανίχνευση κακόβουλης κυκλοφορίας σε IoT δίκτυα, χρησιμοποιώντας το N-BaIoT. Η προσέγγιση βασίζεται σε horizontal FL με κατανομή των δεδομένων σε διαφορετικούς πελάτες, όπου κάθε πελάτης εκπαιδεύει τοπικά deep learning μοντέλα (CNN, LSTM, GRU). Η μελέτη αναδεικνύει την υπεροχή του CNN μοντέλου σε σχέση με τα υπόλοιπα, τόσο στο κεντρικοποιημένο όσο και στο κατανεμημένο σενάριο, παρουσιάζοντας ακρίβεια έως και 90.83%. Επιπλέον, η εργασία διερευνά την επίδραση της επιλεκτικής χρήσης χαρακτηριστικών βάσει χρονικών παραθύρων (π.χ. L0.01, L1, L5), καταδεικνύοντας ότι ορισμένα υποσύνολα χαρακτηριστικών μπορούν να προσφέρουν απόδοση συγκρίσιμη με τη χρήση ολόκληρου του dataset.

3.3 Εφαρμογές του Bot-IoT dataset

Στο ίδιο πλαίσιο, το dataset Bot-IoT [11] παρέχει ένα ολοκληρωμένο dataset σχεδιασμένο για τη μελέτη επιθέσεων botnet σε περιβάλλοντα IoT. Προσομοιώνει τόσο κανονική όσο και κακόβουλη κυκλοφορία, περιλαμβάνοντας ένα ευρύ φάσμα επιθέσεων, όπως Denial of Service (DoS), DDoS, reconnaissance και κλοπή πληροφοριών.

Συγκεκριμένα στο paper από Caldas Filho et al. [53], μελετάται η ανίχνευση και αντιμετώπιση επιθέσεων botnet σε δίκτυα IoT, με τη χρήση του συνόλου δεδομένων Bot-IoT. Σε αντίθεση με παραδοσιακές μεθόδους ανίχνευσης βασισμένες σε υπογραφές, οι συγγραφείς προτείνουν μια end-to-end προσέγγιση όπου το μοντέλο εκπαιδεύεται απευθείας σε στατιστικά χαρακτηριστικά της κυκλοφορίας. Αξιοσημείωτο είναι ότι η προτεινόμενη DL αρχιτεκτονική επιτυγχάνει ανίχνευση με ακρίβεια 99.99%, γεγονός που

[10] https://archive.ics.uci.edu/dataset/442/detection+of+iot+botnet+attacks+n+baiot
[11] https://research.unsw.edu.au/projects/bot-iot-dataset

υποδεικνύει τις δυνατότητες των νευρωνικών δικτύων στη μοντελοποίηση περίπλοκης, μη γραμμικής συμπεριφοράς κακόβουλης δραστηριότητας.

Η μελέτη των Torre et al. [54] προσφέρει ένα προηγμένο IDS για IoT εφαρμόζοντας horizontal FL σε συνδυασμό με μια 1-D CNN αρχιτεκτονική. Σε αντίθεση με πιο απλές FL προσεγγίσεις, το σύστημα ενσωματώνει τρεις τεχνολογίες προστασίας: Differential Privacy για προσθήκη θορύβου, Diffie-Hellman για ασφαλή ανταλλαγή κλειδιών, και Homomorphic Encryption για προστατευμένη εκτέλεση μαθηματικών πράξεων πάνω στα κρυπτογραφημένα δεδομένα. Η αξιολόγηση πραγματοποιήθηκε σε δημόσια IoT σύνολα (TON-IoT, IoT-23, Bot-IoT, CIC-IoT 2023) και έδειξε υψηλή μέση επίδοση (accuracy: 97.31 %, precision: 95.59 %, recall: 92.43 %, F1: 92.69 %). Συγκριτικά, η προσθήκη αυτών των privacy layers αποδεικνύει ότι είναι εφικτό να συνδυαστεί υψηλή αντιδραστικότητα στην ανίχνευση με robust ιδιωτικότητα, επιλύοντας θεμελιώδη ζητήματα για την υιοθέτηση FL σε πραγματικά, ευαίσθητα IoT περιβάλλοντα.

3.4 FL με χρήση του TON_IoT dataset

Συμπληρωματικά το TON_IoT [12], που αναπτύχθηκε από το Πανεπιστήμιο της Νέας Νότιας Ουαλίας (UNSW), περιλαμβάνει κανονική και κακόβουλη κυκλοφορία, με επιθέσεις όπως DDoS, data exfiltration και ransomware, καθώς και δεδομένα όπως δικτυακή κυκλοφορία και logs.

Αξιοποιώντας το ToN_IoT, το paper [55] χρησιμοποιεί ένα πλαίσιο FL για την ανίχνευση κυβερνοεπιθέσεων, συνδυάζοντας decision trees και support vector machines, εστιάζοντας σε περιβάλλοντα με περιορισμένους υπολογιστικούς πόρους. Η εργασία δίνει ιδιαίτερη έμφαση στη φάση της προεπεξεργασίας, εφαρμόζοντας κανονικοποίηση και επιλογή χαρακτηριστικών, κάτι που βελτιώνει τόσο την απόδοση όσο και την αποδοτικότητα των μοντέλων. Το πειραματικό αποτέλεσμα, με ακρίβεια που φτάνει το 98.7%, αναδεικνύει τη δυνατότητα υλοποίησης αποτελεσματικών, ελαφριών και privacy-aware IDS συστημάτων σε αποκεντρωμένα IoT περιβάλλοντα. Σε αντίθεση με πιο σύνθετες DL λύσεις, η μελέτη δείχνει ότι ακόμη και συμβατικά ML μοντέλα μπορούν να λειτουργήσουν αποτελεσματικά στο πλαίσιο του FL, εφόσον υποστηρίζονται από κατάλληλη επιλογή χαρακτηριστικών και σχεδιασμό ροής δεδομένων.

Στη μελέτη των Lazzarini et al.[56] εξετάζεται εμπειρικά η αποτελεσματικότητα του horizontal FL με δοκιμές σε δύο διαφορετικά datasets, TON-IoT και CICIDS2017, και εφαρμογή ενός ANN με χρήση του FedAvg κατά τη συνάθροιση των τοπικών ενημερώσεων. Η έρευνα αναδεικνύει ότι το FL σύστημα επιτυγχάνει σύγκριση με την κεντρικοποιημένη προσέγγιση όσον αφορά τις μετρικές ακρίβειας, precision, recall και F1-score,

[12] https://research.unsw.edu.au/projects/toniot-datasets

επιβεβαιώνοντας ότι το FL μπορεί να αποτελέσει έγκυρη εναλλακτική για IDS σε δίκτυα IoT . Επιπλέον, μελετήθηκαν και άλλοι μέθοδοι aggregation (FedAvgM, FedAdam, FedAdagrad), με τα αποτελέσματα να δείχνουν ότι ο FedAvg και ο FedAvgM υπερέχουν σε σχέση με τους προσαρμοστικούς αλγορίθμους στο συγκεκριμένο πλαίσιο.

3.5 FL εργασίες με άλλα Iot Datasets

Πλήθος πρόσφατων ερευνητικών εργασιών έχει αξιοποιήσει γνωστά IoT-related datasets για την ανάπτυξη και αξιολόγηση FL μοντέλων ανίχνευσης εισβολών. Ένα από τα πλέον πρόσφατα είναι το το IoT Dataset 2023, από το Canadian Institute for Cybersecurity (CIC), είναι ένα dataset που μελετά τη συμπεριφορά IoT συσκευών σε πραγματικά δικτυακά περιβάλλοντα. Περιλαμβάνει κανονική και κακόβουλη κυκλοφορία, υποστηρίζοντας την ανάπτυξη συστημάτων ανίχνευσης εισβολών (IDS) και παρέχοντας πολύτιμες πληροφορίες για την ασφάλεια IoT και 5G. Με την χρήση αυτου του dataset οι Abbas S et al.[57] προτείνει ένα federated learning με αλγόριθμο federated averaging για τη διαδικασία aggregation, ενώ το μοντέλο βασίζεται σε αρχιτεκτονική DNN. Η προσέγγιση επιτυγχάνει υψηλή ακρίβεια 99.99%, αποδεικνύοντας την αποτελεσματικότητά της στην ανίχνευση κυβερνοαπειλών.

Με στόχο την περαιτέρω ανάλυση της ασφάλειας του IoT, το CSE-CIC-IDS2018 παρέχει κανονική και κακόβουλη κυκλοφορία, καλύπτοντας επιθέσεις όπως DoS/DDoS, brute force και web attacks. Στο paper από Hamdi [58] ,χρησιμοποιώντας αυτό το dataset, αξιολογεί διάφορες μεθοδολογίες aggregation, όπως οι FedAvg, FedProx, FedOpt, FedAdam, FedYogi και FedAdagrad, για την αποτελεσματικότητά τους. Η προσέγγιση χρησιμοποιεί Convolutional neural network (CNN) και μεταξύ των μεθόδων που δοκιμάστηκαν, το FedAvg επιτυγχάνει την υψηλότερη ακρίβεια, ξεπερνώντας το 98%.

Συνεχίζοντας την έρευνα μας η εργασία απο Fan et al. [4] αξιοποιεί το CICIDS2017 dataset για την αξιολόγηση της απόδοσης του federated transfer learning μοντέλου, που χρησιμοποιεί Federated Averaging. Το μοντέλο είναι βασισμένο σε CNN και επιτυγχάνει AvgAccuracy=91.93% σε όλους τους πελάτες. Το συγκεκριμένο dataset περιέχει κρίσιμες πληροφορίες για κανονική κυκλοφορία αλλα και για επιθέσεις, όπως DDoS, brute force, infiltration, web attacks και botnets. Το dataset παρέχει επίσης ρεαλιστικά δεδομένα δικτυακής κυκλοφορίας, όπως packet captures (PCAPs), χαρακτηριστικά ροής (flow-based features) και στατιστικές μετρήσεις, υποστηρίζοντας την ανάπτυξη και αξιολόγηση μοντέλων μηχανικής και βαθιάς μάθησης για την ανίχνευση εισβολών χρησιμοποιείται από σε 5G IoT.

Μελετήσαμε επίσης το NSL-KDD dataset που αποτελείτε από δεδομένα δικτυακής κυκλοφορίας με ετικέτες που περιγράφουν κανονικές και κακόβουλες δραστηριότητες, κατηγοριοποιημένα σε τύπους επιθέσεων όπως Denial of Service (DoS), Probe, User-to-

Root (U2R) και Remote-to-Local (R2L). Η ισορροπημένη φύση του και η εκτενής κάλυψη προτύπων επιθέσεων το καθιστούν σημαντικό για την ανάπτυξη και αξιολόγηση πλαισίων IDS και τεχνικών μηχανικής μάθησης στον τομέα της κυβερνοασφάλειας. Οι Mirzaee et al.[59] εξετάζουν την ανίχνευση κυβερνοεπιθέσεων σε έξυπνα δίκτυα μέτρησης (smart grid), αξιοποιώντας το σύνολο δεδομένων NSL-KDD. Στην εργασία προτείνεται ένα Federated Intrusion Detection System (FIDS), το οποίο βασίζεται σε αρχιτεκτονική CNN και χρησιμοποιεί τον αλγόριθμο Federated Averaging για την ομοσπονδιακή συνένωση των μοντέλων. Το προτεινόμενο σύστημα επιτυγχάνει ακρίβεια ανίχνευσης 99,5%.

Συνεχίζοντας, η εργασία από Idrissi et al. [60],χρησιμοποιώντας τα datasets USTC-TFC2016, CIC-IDS2017 και CSE-CIC-IDS2018, τα οποία κατανέμονται μεταξύ των πελατών, εφαρμόζει τον αλγόριθμο FedProx για τη συνένωση των τοπικών μοντέλων. Χρησιμοποιούνται autoencoders, όπως Variational και Adversarial Autoencoders. Η προσέγγιση επιτυγχάνει ακρίβεια 97.2% στο USTC-TFC2016, 96.8% στο CIC-IDS2017 και 95.5% στο CSE-CIC-IDS2018. Μελετώντας το USTC-TFC2016[13] dataset, διαπιστώσαμε ότι είναι ένα real-world dataset που αναπτύχθηκε από το Πανεπιστήμιο Επιστήμης και Τεχνολογίας της Κίνας. Περιλαμβάνει κανονική και κακόβουλη κυκλοφορία σε μορφή pcap, με κάλυψη επιθέσεων όπως DoS και DdoS. Διαθέτει 14 χαρακτηριστικά, καθιστώντας το κατάλληλο για έρευνα στην ανίχνευση εισβολών.

Η εργασία από Almeida et al. [61] χρησιμοποιεί τα datasets NSL-KDD και UNSW-NB15. Οι συγγραφείς εφαρμόζουν τον αλγόριθμο Federated Averaging (FedAvg) και αρχιτεκτονική βασιζόμενη σε CNN. Το σύστημα επιτυγχάνει υψηλή ακρίβεια ανίχνευσης, με 98% στο NSL-KDD και 97% στο UNSW-NB15. Το UNSW NB-15 [14], που έρχεται να συμπληρώσει την έρευνα μας για τα IoT datasets, δημιουργήθηκε χρησιμοποιώντας το εργαλείο IXIA PerfectStorm, προσομοιώνοντας ένα ρεαλιστικό σύγχρονο δικτυακό περιβάλλον. Περιλαμβάνει εννέα κατηγορίες επιθέσεων, όπως DoS, Fuzzers, Backdoors και Exploits, παρέχοντας μια ολοκληρωμένη αναπαράσταση σύγχρονων απειλών δικτύου. Το dataset περιέχει ακατέργαστη κυκλοφορία δικτύου σε μορφή pcap καθώς και προεπεξεργασμένα flows με 49 εξαγόμενα χαρακτηριστικά ανά ροή.

[13] https://github.com/yungshenglu/USTC-TFC2016/tree/master
[14] https://research.unsw.edu.au/projects/unsw-nb15-dataset

3.6 5G datasets σε ML εργασίες

Σε αυτή την υποενότητα, παρουσιάζονται πειραματικές μελέτες μηχανικής μάθησης ML που βασίζονται σε γνωστά 5G datasets, τα οποία χρησιμοποιούνται για την ανίχνευση επιθέσεων και την αξιολόγηση μοντέλων. Αν και η αξιοποίηση του FL σε συνδυασμό με 5G δεδομένα παραμένει περιορισμένη, τα εξεταζόμενα 5G datasets παρουσιάζουν σημαντικές δυνατότητες για μελλοντική εφαρμογή σε ομοσπονδιακά περιβάλλοντα, συμβάλλοντας στην ενίσχυση της ασφάλειας σε έξυπνα και κατανεμημένα δίκτυα.

Ένα σημαντικό dataset που συμβάλλει στην έρευνα για την ασφάλεια δικτύων 5G είναι το 5GAD, το οποίο παρέχει ένα προσομοιωμένο περιβάλλον που αναπαριστά την κυκλοφορία δικτύων 5G, περιλαμβάνοντας τόσο κανονική όσο και κακόβουλη κυκλοφορία, όπως reconnaissance, network reconfiguration και denial of service. Τα δεδομένα διατίθενται σε διάφορες μορφές, όπως pcapng και log files.

Με βάση αυτό το dataset, το άρθρο "Payload-based 5G Attack Detection" [62] προτείνει ένα σύστημα ανίχνευσης επιθέσεων στο 5G core network, εστιάζοντα στο Packet Forwarding Control Protocol (PFCP). Το προτεινόμενο σύστημα, βασισμένο σε CNN, αναλύει τα PFCP payloads για την ανίχνευση κακόβουλων δραστηριοτήτων, εφαρμόζοντας προ επεξεργασία μέσω feature extraction και normalization. Το σύστημα επιτυγχάνει υψηλή ακρίβεια στην ανίχνευση στην ανίχνευση PFCP-based επιθέσεων.

Παράλληλα, οι Bakar et al. [63] προτείνουν ένα άλλο σύστημα ανίχνευσης, επίσης βασισμένο σε CNN, εστιάζοντας αυτή τη φορά σε GTP κυκλοφορία, δηλαδή σε διαφορετικό επίπεδο του 5G πρωτοκόλλου. Το σύστημα εκπαιδεύεται με δεδομένα που συλλέγονται από P4 switches και αξιοποιεί το dataset 5GDAD, το οποίο αποτελεί διαφορετική συλλογή από το 5GAD. Η μελέτη επιτυγχάνει υψηλή απόδοση στην ανίχνευση DDoS επιθέσεων στο επίπεδο GTP, υποδεικνύοντας ότι η ανάλυση payloads σε διαφορετικά layers του 5G core μπορεί να προσφέρει αποτελεσματικά σχήματα προστασίας.

Συμπληρωματικά,το DoS/DDoS Attack Dataset for 5G Network Slicing [15] είναι ένα προσομοιωμένο dataset που έχει σχεδιαστεί για την ανάλυση επιθέσεων DoS και DDoS σε περιβάλλοντα 5G network slicing. Το dataset διατίθεται σε δύο μορφές: αρχεία τύπου pcap, τα οποία περιλαμβάνουν την ακατέργαστη δικτυακή κυκλοφορία, και αρχεία CSV, στα οποία έχουν εξαχθεί συνολικά 84 χαρακτηριστικά. Τα δεδομένα καλύπτουν μια ποικιλία επιθέσεων, όπως UDP flooding και επιθέσεις που βασίζονται στο πρωτόκολλο TCP (π.χ. sync, push, fin, rst). Το σύνολο δεδομένων χρησιμοποιήθηκε για την αξιολόγηση του μοντέλου SliceSecure

[15] https://ieee-dataport.org/documents/dosddos-attack-dataset-5g-network-slicing

που μπόρεσε να ταξινομήσει την καλοήθη κίνηση και τις επιθέσεις DoS/DDoS με ποσοστό επιτυχίας 99,99% [64].

Στο πλαίσιο του πειράματος [65], οι ερευνητές για την ανίχνευση εισβολών, εφάρμοσαν ένα Provenance-based Intrusion Detection System (PIDS), το οποίο εκπαιδεύτηκε χρησιμοποιώντας τα δεδομένα του 5GProvGen. Τα αποτελέσματα του πειράματος έδειξαν ότι το PIDS ήταν ικανό να εντοπίζει αποτελεσματικά κακόβουλες ενέργειες στο 5G core network. Το dataset 5GProvGen [16] είναι ένα προσομοιωμένο dataset, που περιλαμβάνει τόσο κανονική όσο και κακόβουλη κυκλοφορία, που επικεντρώνεται στα 5G core networks, παρέχοντας raw W3C provenance graph logs.

Σε συνέχεια της ανάλυσης εργασιών το SPEC5G [17] dataset που παρουσιάζεται από τους I. Karim et al. [66] περιλαμβάνει πραγματικά δεδομένα κειμένου για την ανάλυση και επικύρωση πρωτοκόλλων 5G, όπως VoWiFi, 5G-AKA και Cellular IoT. Εστιάζει στην ασφάλεια των πρωτοκόλλων και μπορεί να χρησιμοποιηθεί για event classification, παρέχοντας πολύτιμα δεδομένα για την ενίσχυση της αξιοπιστίας και της ασφάλειας στο 5G. Στο πείραμα οι ερευνητές αξιοποίησαν το SPEC5G σε δύο καθορισμένα tasks: την ταξινόμηση προτάσεων ασφάλειας (security sentence classification) και την σύνοψη παραγράφων πρωτοκόλλων (paragraph summarization). Για την εκπαίδευση χρησιμοποιήθηκαν προεκπαιδευμένα μοντέλα φυσικής γλώσσας όπως BERT, RoBERTa και XLNet, τα οποία στη συνέχεια επαναπροσαρμόστηκαν πάνω στο SPEC5G. Τα αποτελέσματα έδειξαν ότι οι παραλλαγές BERT5G και RoBERTa5G, που προεκπαιδεύτηκαν στο SPEC5G, υπερείχαν σημαντικά των βασικών εκδόσεων των μοντέλων.

Τέλος, παρουσιάζονται δύο datasets που αφορούν δίκτυα 5G και παρότι δεν έχουν αξιοποιηθεί ακόμη σε πειράματα, θα μπορούσαν να αποτελέσουν χρήσιμη βάση για μελλοντική έρευνα τόσο σε σενάρια Federated Learning όσο και σε γενικότερα πειράματα μηχανικής μάθησης.Το πρώτο είναι το 5GC-PFCP, όπως περιγράφεται στο συγκεκριμένο άρθρο [67], είναι ένα προσομοιωμένο dataset που εστιάζει στο PFCP των 5G Core networks. Περιλαμβάνει κανονική και κακόβουλη κυκλοφορία, με επιθέσεις όπως Session Deletion Flood DoS και Session Modification Flood attacks και διατίθεται σε μορφές CSV και pcap. Το δεύτερο είναι το 5G Traffic Datasets που παρέχει δεδομένα πραγματικής κυκλοφορίας δικτύων, προσφέροντας πολύτιμες πληροφορίες για τη συμπεριφορά των 5G περιβαλλόντων. Διατίθενται σε μορφές CSV και pcap και περιλαμβάνουν αποκλειστικά κανονική κυκλοφορία, γεγονός που τα καθιστά κατάλληλα για βασική μοντελοποίηση και ανίχνευση ανωμαλιών σε εφαρμογές σχετικές με 5G και IoT. Οι Choi et al. [68] παρουσιάζουν

[16] https://ieee-dataport.org/documents/5gprovgen
[17] https://github.com/Imtiazkarimik23/SPEC5G

το εν λόγω σύνολο δεδομένων ως σχεδιασμένο ώστε να αντανακλά τις πραγματικές ροές δεδομένων σε περιβάλλοντα 5G.

4. Μεθολογία

Στην ενότητα αυτή περιγράφεται αναλυτικά η μεθοδολογική προσέγγιση που ακολουθήθηκε για την υλοποίηση της εργασίας. Αρχικά, παρουσιάζεται το επιλεγμένο dataset καθώς και τα βήματα προ επεξεργασίας που εφαρμόστηκαν. Στη συνέχεια, αναλύεται η διαδικασία υλοποίησης του πειράματος FL με στόχο την ανίχνευση κακόβουλης δραστηριότητας σε περιβάλλοντα IoT και 5G.

4.1 Δεδομένα και Προεπεξεργασία

4.1.1 Σύνολο δεδομένων (Dataset)

Στην παρούσα έρευνα, επιλέχθηκε το dataset NBaIoT [18](Network-based Anomaly Detection in IoT) [69] ως η βασική πηγή δεδομένων για την ανάλυση και την ανίχνευση ανωμαλιών σε δίκτυα IoT. Η επιλογή αυτού του dataset βασίστηκε σε μια σειρά από κριτήρια που το καθιστούν ιδανικό για την μελέτη της ανίχνευσης επιθέσεων και ανωμαλιών σε περιβάλλοντα IoT. Το συγκεκριμένο dataset περιέχει πραγματικά δεδομένα κίνησης, που συλλέγονται από εννέα εμπορικές συσκευές IoT που είχαν μολυνθεί από κακόβουλο λογισμικό Bashlite και Mirai. Το dataset N-BaIoT περιλαμβάνει κεφαλίδες χαρακτηριστικών που περιγράφουν διάφορες πτυχές της κίνησης του δικτύου. Μεταξύ των χαρακτηριστικών που σχετίζονται με τη συγκέντρωση ροών περιλαμβάνονται το H, το οποίο παρέχει στατιστικά στοιχεία που συνοψίζουν την πρόσφατη επισκεψιμότητα από τον κεντρικό υπολογιστή (IP), το HH, αναφέρεται στην επισκεψιμότητα από τον κεντρικό υπολογιστή στον κεντρικό υπολογιστή προορισμού, και το HpHp, που καλύπτει την πρόσφατη κίνηση από τον κεντρικό υπολογιστή και τη θύρα (IP + port) προς τον κεντρικό υπολογιστή και τη θύρα προορισμού. Το HH_jit συνοψίζει το jitter της κίνησης μεταξύ των δύο κεντρικών υπολογιστών και το MI ("Source MAC-IP") συνοψίζει την πρόσφατη κίνηση από τον κεντρικό υπολογιστή, λαμβάνοντας υπόψη τόσο τη διεύθυνση IP όσο και τη φυσική διεύθυνση (MAC). Η αποσύνθεση του χρονικού πλαισίου καθορίζεται από τον παράγοντα Λάμδα (λ), με τις τιμές L5, L3, L1 κ.ά. να προσφέρουν διαφορετικά επίπεδα λεπτομέρειας στην ανάλυση της κίνησης.

Το N-BaIoT παρέχει επίσης στατιστικά στοιχεία που εξάγονται από τη ροή πακέτων, τα οποία είναι κρίσιμα για την ανάλυση της κίνησης. Μεταξύ αυτών περιλαμβάνονται το βάρος ροής (Weight), που αντιπροσωπεύει τον αριθμό των στοιχείων που παρατηρήθηκαν στην πρόσφατη ιστορία της ροής, ο μέσος όρος (Mean), που δηλώνει τη μέση τιμή της ροής και η τυπική απόκλιση (Std), που συμβολίζει τη διασπορά των τιμών γύρω από τον μέσο όρο. Επιπρόσθετα, η ακτίνα (Radius) αντιπροσωπεύει την τετραγωνική ρίζα του αθροίσματος των διακυμάνσεων δύο ρευμάτων,ενώ το μέγεθος (Magnitude), που είναι το τετραγωνικό

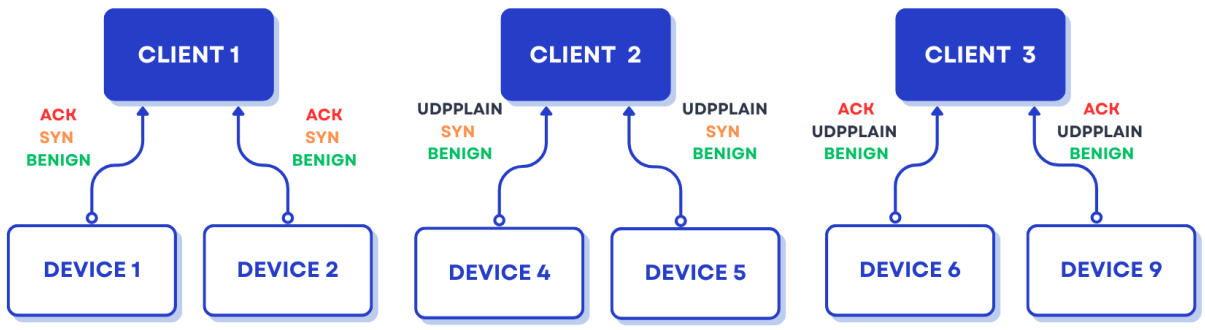
[18] https://archive.ics.uci.edu/dataset/442/detection+of+iot+botnet+attacks+n+baiot

άθροισμα της ρίζας των μέσων τιμών δύο ρευμάτων. Η συνδιακύμανση (Cov) παρέχει μια κατά προσέγγιση συνδιακύμανση μεταξύ δύο ρευμάτων, και ο συντελεστής συσχέτισης (Pc)συμβολίζει την κατά προσέγγιση συνδιακύμανση μεταξύ δύο ρευμάτων.

Αυτά τα χαρακτηριστικά είναι ιδιαίτερα χρήσιμα για την ανάπτυξη και αξιολόγηση συστημάτων ανίχνευσης ανωμαλιών και εισβολών σε δίκτυα IoT. Η ποικιλία των χαρακτηριστικών και η λεπτομερής περιγραφή της κίνησης καθιστούν το N-BaIoT ένα ιδανικό εργαλείο για την εφαρμογή τεχνικών μηχανικής μάθησης και στατιστικής ανάλυσης. Συγκεκριμένα το dataset έχει 115 χαρακτηριστικά με δέκα διαφορετικές επιθέσεις (5 για κάθε μία από τις Mirai και Gafgyt) [70] και περιλαμβάνει 7062606 περιπτώσεις. Για το Gafgyt, οι τύποι επίθεσης περιλαμβάνουν combo, junk, scan, TCP και UDP, ενώ το Mirai εισάγει κλάσεις όπως ack, scan, syn, UDP και UDP plain. Σημαντικό να σημειωθεί ότι οι συσκευές 3 και 7 έχουν εμφανή απουσία δεδομένων που σχετίζονται με την Mirai, αφού η Mirai απέφυγε να μολύνει αυτές τις συσκευές [70].

4.1.2 Προετοιμασία συνόλου δεδομένων

Στην παρούσα εργασία αξιοποιείται το ήδη διαθέσιμο σύνολο δεδομένων N-BaIoT, το οποίο παρουσιάστηκε πιο πάνω. Για τις ανάγκες της μελέτης, επιλέχθηκαν συγκεκριμένα υποσύνολα του dataset που αφορούν τρεις τύπους επιθέσεων από το botnet Mirai, τις mirai.tcp, mirai.udpplain και mirai.syn. Το αρχικό σύνολο δεδομένων οργανώθηκε σε 3 διακριτούς πελάτες, καθένας εκ των οποίων περιλαμβάνει δεδομένα από 2 διαφορετικές συσκευές, προκειμένου να αναπαρασταθεί ένα πιο ρεαλιστικό σενάριο κατανεμημένης εκπαίδευσης. Επιπλέον, η κατανομή των επιθέσεων σχεδιάστηκε έτσι ώστε κάθε τύπος επίθεσης να μοιράζεται μεταξύ 2 από τους 3 πελάτες, με στόχο την ανάλυση της επίδρασης και της ανίχνευσης των επιθέσεων σε ένα κατανεμημένο περιβάλλον. Έτσι στο τέλος θα έχουμε 2 benign και 2/3 attack traffic στο training set του κάθε πελάτη.



Εικόνα 7: Κατανομή επιθέσεων

Εξετάσαμε το σύνολο δεδομένων N-BaIoT με επιθέσεις από τα botnets Gafgyt (Bashtile) και Mirai. Διαπιστώσαμε ότι τα δεδομένα που σχετίζονται με το Gafgyt παρουσίαζαν πλήθος ελλιπών στηλών (columns με μηδενικές τιμές), περιττές επαναλήψεις γραμμών και ασυνεπή δομή, καθιστώντας τα ακατάλληλα για χρήση στην εργασία μας. Αντιθέτως, τα δεδομένα που αφορούσαν το Mirai, μετά από κατάλληλη προ επεξεργασία, παρουσίασαν επαρκή συνοχή και πληρότητα, γεγονός που τα καθιστά κατάλληλα για τα πειράματά μας.

Στη συνέχεια, παρουσιάζονται τα βήματα που ακολουθήθηκαν για να έρθει το dataset σε μια μορφή κατάλληλη για ανάλυση και επίτευξη των παραπάνω στόχων. Ο αρχικός μας κώδικας βασίστηκε στο [19] ενώ στη συνέχεια προστέθηκαν επιπλέον βήματα επεξεργασίας για την καλύτερη προσαρμογή του dataset στις απαιτήσεις του πειράματος.

4.1.2.1 Γενικά βήματα προ επεξεργασίας

Στην αρχή τα δεδομένα οργανώθηκαν σε αρχεία με τη μορφή device.attack.type.csv για επιθέσεις και device.benign.csv για κανονική κίνηση. Τα devices που χρησιμοποιήθηκαν είναι τα εξής:

Device 1 -> Danmini_Doorbell

Device 2 -> Ecobee_Thermostat

Device 4 -> Philips_B120N10_Baby_Monitor

Device 5 -> Provision_PT_737E_Security_Camera

Device 6 -> Provision_PT_838_Security_Camera

Device 9 -> SimpleHome_XCS7_1003_WHT_Security_Camera.

Για την προετοιμασία των δεδομένων, αρχικά εφαρμόστηκε data labeling, κατά την οποία η κανονική κίνηση (normal) επισημάνθηκε με την τιμή 0, ενώ οι διαφορετικοί τύποι επιθέσεων έλαβαν διακριτές ετικέτες: mirai.ack → 1, mirai.udpplain → 2 και mirai.syn → 3. Στη συνέχεια, πραγματοποιήθηκε καθαρισμός των δεδομένων , περιλαμβάνοντας τη διαγραφή διπλότυπων εγγραφών (τα raw data έχουν πάνω από 4,5 εκατομμύρια διπλότυπες εγγραφές), καθώς και την αντιμετώπιση των ελλιπών τιμών με τη χρήση του μέσου όρου για συνεχείς μεταβλητές πχ “H_iat” και της επικρατούσας τιμής για κατηγορηματικές ή διακριτές αριθμητικές στήλες “H_pkt”. Τέλος, εφαρμόστηκε τυχαία αναδιάταξη (shuffle) τόσο πριν όσο και μετά τη διάσπαση του συνόλου δεδομένων, ώστε να αποφευχθεί η εκμάθηση ψευδών μοτίβων από το μοντέλο.

[19] https://github.com/husseinalygit/N-BaIoT-reloaded

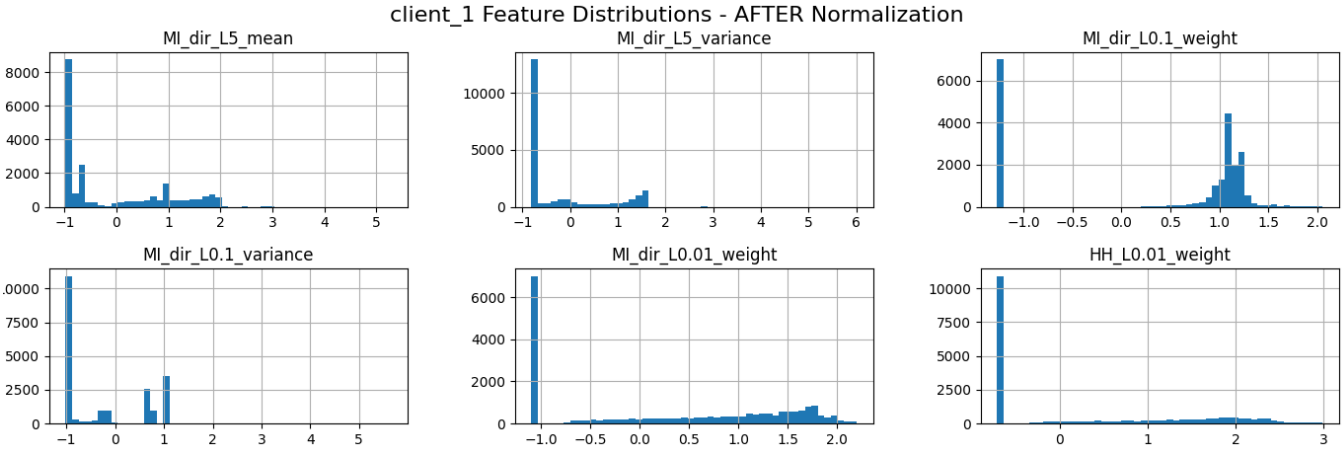
4.1.2.2 Κανονικοποίηση/Τυποποίηση

Για την κανονικοποίηση των δεδομένων, αρχικά εφαρμόστηκε η μέθοδος Min-Max Normalization, η οποία μετασχηματίζει τις τιμές των χαρακτηριστικών σε ένα εύρος [0, 1]. Η προσέγγιση αυτή κρίθηκε απαραίτητη, καθώς το σύνολο δεδομένων N-BaIoT περιλαμβάνει χαρακτηριστικά με σημαντικά διαφορετικές κλίμακες τιμών, όπως το H_iat (τιμές από 0 έως 1000) και το MI_dir (τιμές από 0 έως 1). Μέσω της καθολικής κανονικοποίησης, όλα τα χαρακτηριστικά αποκτούν το ίδιο εύρος, εξασφαλίζοντας ότι κανένα δεν υπερισχύει δυσανάλογα στη διαδικασία μάθησης.

Ωστόσο, διαπιστώθηκε ότι η μέθοδος αυτή δεν ήταν κατάλληλη για το συγκεκριμένο σύνολο δεδομένων, καθώς οι τιμές παρουσίαζαν εξαιρετικά μεγάλο δυναμικό εύρος, κυμαινόμενο από περίπου 10^{-4} έως και 10^{13}. Ως εκ τούτου, επιλέχθηκε τελικά η μέθοδος Z-score (StandardScaler), η οποία πραγματοποιεί τυποποίηση των δεδομένων, μετασχηματίζοντας τις τιμές ώστε να έχουν μέσο όρο ίσο με μηδέν και τυπική απόκλιση ίση με ένα. Η τυποποίηση βασίζεται στον υπολογισμό του z-score, σύμφωνα με τον τύπο [71]:

$$z = \frac{\chi - \mu}{\sigma}$$

όπου x είναι η αρχική τιμή, μ-ο μέσος όρος και σ-η τυπική απόκλιση. Η μέθοδος αυτή είναι ιδιαίτερα χρήσιμη όταν τα χαρακτηριστικά παρουσιάζουν μεγάλες αποκλίσεις και όταν τα μοντέλα μηχανικής μάθησης προτιμούν είσοδο με δεδομένα που ακολουθούν περίπου κανονική κατανομή.



Εικόνα 8: Feature Distribution After Normalazation

Στην Εικόνα 3 παρουσιάζονται οι κατανομές επιλεγμένων χαρακτηριστικών (features) μετά την κανονικοποίηση των δεδομένων. Όπως είναι αναμενόμενο, οι τιμές των

χαρακτηριστικών συγκεντρώνονται γύρω από το μηδέν, καθώς εφαρμόστηκε κανονικοποίηση με χρήση ZeroScaler.

4.1.2.3 Διαχωρισμός Δεδομένων

Για κάθε τύπο επίθεσης, τα δεδομένα χωρίστηκαν σε τμήματα των 5000 γραμμών. Σε κάθε πελάτη ανατίθενται δύο τέτοια τμήματα, το καθένα προερχόμενο από διαφορετική επίθεση. Αυτή η προσέγγιση διασφαλίζει την ομοιόμορφη κατανομή των δεδομένων μεταξύ των πελάτων, γεγονός που καθιστά εφικτή την ανίχνευση επιθέσεων σε ένα κατανεμημένο υπολογιστικό περιβάλλον.

Αναφορικά με την κανονική κίνηση, ακολουθήθηκε η ίδια διαδικασία επιλογής, με τη λήψη ενός ισομεγέθους τμήματος από κάθε συσκευή. Καθώς κάθε πελάτη σχετίζεται με δύο συσκευές, καταλήγει να διαχειρίζεται συνολικά δύο τμήματα benign δεδομένων διατηρώντας την ισορροπία μεταξύ των κατηγοριών και των συσκευών που μελετώνται.

Το πλήρες σύνολο δεδομένων χωρίστηκε σε τρεις διακριτές υποκατηγορίες: 70% για εκπαίδευση (training), 15% για επικύρωση (validation) και 15% για τελική δοκιμή (testing). Το validation set αποτελεί ανεξάρτητο υποσύνολο από το training set και χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου κατά τη διάρκεια της εκπαίδευσης, προκειμένου να αποφευχθεί η υπερπροσαρμογή (overfitting), δηλαδή το μοντέλο να γίνει πολύ καλό στην ταξινόμηση των δειγμάτων στο σύνολο εκπαίδευσης, αλλά δεν μπορεί να γενικεύσει και να κάνει ακριβείς ταξινομήσεις σε δεδομένα που δεν έχει ξαναδεί [72]. Τα δεδομένα εκπαίδευσης αξιοποιούνται για την εκμάθηση κρυφών προτύπων, ενώ τα δεδομένα δοκιμής προσφέρουν μια ανεπηρέαστη τελική μέτρηση της ακρίβειας του μοντέλου σε άγνωστα δείγματα.

4.1.2.4 Επιλογή χαρακτηριστικών (Feature Selection)

Ξεκινάμε με ένα σύνολο 115 χαρακτηριστικών, τα οποία προκύπτουν από τον υπολογισμό των ίδιων 23 μετρικών σε πέντε χρονικά παράθυρα, ώστε να αποτυπώνεται το στιγμιότυπο κάθε πακέτου [70]. Στο πλαίσιο της επιλογής χαρακτηριστικών, εφαρμόστηκε το κατώφλι διασποράς (threshold variance), μια υπερπαράμετρος που επιτρέπει την απομάκρυνση των χαρακτηριστικών με χαμηλή διακύμανση[20]. Τα χαρακτηριστικά αυτά θεωρούνται ανεπαρκώς διακριτικά και συνεπώς μη χρήσιμα για το μοντέλο, καθώς δεν συνεισφέρουν ουσιαστικά στη διαδικασία μάθησης.

[20] https://medium.com/@venky.jishu2021/feature-selection-in-machine-learnig-using-variance-threshold-a120366eec2f

Για τον σκοπό αυτό, εφαρμόστηκε διαδοχική διαδικασία επιλογής χαρακτηριστικών βάσει κατωφλίου διασποράς. Το κατώφλι ορίστηκε σε διαφορετικά επίπεδα και παρακάτω παρουσιάζονται αναλυτικά τα χαρακτηριστικά που αφαιρέθηκαν για κάθε περίπτωση:

- **Threshold=0.0001:** Αφαιρέθηκαν συνολικά 33 χαρακτηριστικά, κυρίως μετρικές τύπου covariance, pcc, radius και weight, στις μικρές χρονικές κλίμακες (π.χ. L0.01, L0.1, L1). Αυτές οι μετρικές παρουσίαζαν σχεδόν μηδενική διακύμανση και ήταν πρακτικά σταθερές σε όλα τα δείγματα. Συγκεκριμένα αφαιρέθηκαν:

  HH_L5_covariance, HH_L5_pcc, HH_L3_covariance, HH_L3_pcc

  HH_L1_covariance, HH_L1_pcc, HH_L0.1_covariance, HH_L0.1_pcc,

  HH_L0.01_covariance, HH_L0.01_pcc, HpHp_L5_weight, HpHp_L5_std,

  HpHp_L5_radius, HpHp_L5_covariance, HpHp_L5_pcc, HpHp_L3_weight,

  HpHp_L3_std, HpHp_L3_radius, HpHp_L3_covariance, HpHp_L3_pcc,

  HpHp_L1_weight, HpHp_L1_std, HpHp_L1_radius, HpHp_L1_covariance,

  HpHp_L1_pcc, HpHp_L0.1_std, HpHp_L0.1_radius, HpHp_L0.1_covariance,

  HpHp_L0.1_pcc, HpHp_L0.01_std, HpHp_L0.01_radius, HpHp_L0.01_pcc,

  HpHp_L0.01_covariance

- **Threshold=0.001:** Αφαιρέθηκαν επιπλέον 4 χαρακτηριστικά, όπως το που επίσης δεν παρουσίαζαν επαρκή διαφοροποίηση μεταξύ των δειγμάτων. Ενδεικτικά, μεταξύ αυτών περιλαμβάνονται:

  HpHp_L0.1_weight, HpHp_L0.01_weight, MI_dir_L0.01_mean, H_L0.01_mean

- **Threshold=0.01:** Απομακρύνθηκαν 18 ακόμη χαρακτηριστικά. Πολλά από αυτά προέρχονταν από μεταβλητές σχετικές με τη διακύμανση jitter, τη μέση εντροπία και την κατεύθυνση πληροφορία, χαρακτηριστικά τα οποία κρίθηκαν επίσης μη επαρκώς διακριτικά. Αναλύτικά αφαιρέθηκαν:

  HH_L5_weight, HH_L5_std, HH_L5_radius, HH_L3_weight, HH_L3_std,

  HH_L3_radius, HH_L1_weight, HH_L1_std, HH_L0.1_std, HH_L0.01_std

  HH_jit_L5_weight, HH_jit_L5_variance, HH_jit_L3_weight, HH_jit_L1_weight

  MI_dir_L5_weight, MI_dir_L0.1_mean, H_L5_weight, H_L0.1_mean

- **Threshold=0.1:** Σε αυτό το επίπεδο, δεν αφαιρέθηκε κανένα επιπλέον χαρακτηριστικό.

- **Threshold=0.5:** Σε αυτό το αυστηρότερο επίπεδο, απορρίφθηκαν 30 χαρακτηριστικά με περιορισμένη συμβολή, που σχετίζονται με βάρη (weights), τυπικές αποκλίσεις (std), ακτίνες (radius) και στατιστικά από τις χαμηλότερες χρονικές κλίμακες. Συγκεκριμένα αφαιρέθηκαν:

  HH_L5_weight, HH_L3_weight, HH_L1_weight, HH_L0.1_weight, HH_L5_std,

  HH_L3_std, HH_L1_std, HH_L0.1_std, HH_L0.01_std, HH_L5_radius, HH_L3_radius

  HH_jit_L5_weight, HH_jit_L3_weight, HH_jit_L1_weight, HH_jit_L0.1_weight,

  HH_jit_L5_variance, MI_dir_L5_weight, MI_dir_L3_weight, MI_dir_L1_weight

  MI_dir_L0.1_mean, MI_dir_L1_mean, MI_dir_L0.01_mean, MI_dir_L0.01_variance

  H_L5_weight, H_L3_weight, H_L1_weight, H_L0.1_mean, H_L1_mean,

  H_L0.01_mean, H_L0.01_variance

Για το variance threshold αρχικά υλοποιήθηκε κώδικας με στόχο την αυτόματη αναγνώριση των χαρακτηριστικών που πρέπει να αφαιρεθούν από κάθε αρχείο δεδομένων. Στη συνέχεια, αναπτύχθηκε πρόσθετος κώδικας για τη χειροκίνητη (manual) αφαίρεση των ίδιων χαρακτηριστικών από όλα τα σύνολα δεδομένων. Με αυτόν τον τρόπο διασφαλίζεται ότι όλες οι στήλες που κρίθηκαν μη χρήσιμες απορρίπτονται με συνέπεια και ότι όλα τα datasets διαθέτουν κοινό και ομοιογενές σύνολο χαρακτηριστικών για την επόμενη φάση της ανάλυσης. Η εφαρμογή του threshold αυτού οδήγησε στην αφαίρεση συνολικά 85 χαρακτηριστικών, διατηρώντας τα πιο διακριτικά για την εκπαίδευση του ταξινομητή.

Στην συνέχεια ελέγχουμε την συσχέτιση χαρακηριστικών (feature correlation) για να εντοπίσουμε τυχόν περιττά χαρακτηριστικά που δε συμβάλλουν στη διαδικασία εκμάθησης του μοντέλου. Για κάθε πιθανό ζεύγος χαρακτηριστικών, υπολογίζουμε το συντελεστή συσχέτισης Pearson [73] και εάν αυτός ξεπερνάει ένα προκαθορισμένο κατώφλι (το οποίο έχουμε θέσει ίσο με 0.95) τότε επιλέγεται τυχαία ένα από τα δύο χαρακτηριστικά του ζεύγους αυτού και διαγράφεται από το σύνολο δεδομένων.

Correlation > 0.95: Αφαιρέθηκαν συνολικά 40 χαρακτηριστικά, κυρίως μετρικές τύπου variance, mean και radius. Τα χαρακτηριστικά αυτά παρουσίαζαν πολύ υψηλή συσχέτιση μεταξύ τους, γεγονός που υποδηλώνει πλεονασμό πληροφορίας. Η αφαίρεσή τους στόχευσε στη μείωση της πολυδιάστατης πλεοναστικότητας και στη βελτίωση της γενίκευσης του μοντέλου. Συγκεκριμένα, αφαιρέθηκαν:

MI_dir_L3_mean, MI_dir_L3_variance, MI_dir_L1_variance,H_L5_mean, H_L5_variance

H_L3_mean, H_L3_variance, H_L1_variance, H_L0.1_weight, H_L0.1_variance,H_L0.01_weight

HH_L5_mean, HH_L5_magnitude, HH_L3_mean, HH_L3_magnitude, HH_L1_mean,

HH_L1_magnitude, HH_L1_radius, HH_L0.1_mean, HH_L0.1_magnitude, HH_L0.1_radius

HH_L0.01_mean, HH_L0.01_magnitude, HH_L0.01_radius, HH_jit_L5_mean, HH_jit_L3_mean

HH_jit_L1_mean, HH_jit_L1_variance, HH_jit_L0.1_mean, HH_jit_L0.01_weight,

HH_jit_L0.01_mean, HpHp_L5_mean, HpHp_L5_magnitude, HpHp_L3_mean,

HpHp_L3_magnitude, HpHp_L1_mean, HpHp_L1_magnitude, HpHp_L0.1_mean,

HpHp_L0.1_magnitude, HpHp_L0.01_mean, HpHp_L0.01_magnitude

4.2 Διαδικασία Υλοποιήσης

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, υλοποιήθηκαν και αξιολογήθηκαν δύο διαφορετικές προσεγγίσεις για την ανίχνευση εισβολών. Η πρώτη βασίζεται σε FL αξιοποιώντας τη δυνατότητα τοπικής εκπαίδευσης χωρίς ανταλλαγή ευαίσθητων δεδομένων. Η δεύτερη προσέγγιση υλοποιεί ένα παραδοσιακό μοντέλο CL στο οποίο τα δεδομένα συγκεντρώνονται σε έναν κεντρικό server για την εκπαίδευση του μοντέλου.

Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε είναι η Python (έκδοση 3.11.11). Η Python είναι μια γλώσσα υψηλού επιπέδου και γενικής χρήσης, η οποία έχει σχεδιαστεί ώστε να προσφέρει ευανάγνωστο και κατανοητό κώδικα. Αυτά τα χαρακτηριστικά την καθιστούν ιδιαίτερα δημοφιλή στον τομέα της Μηχανικής Μάθησης. Παράλληλα, έχει πολλές έτοιμες βιβλιοθήκες τις οποίες αξιοποιήσαμε όπως, TensorFlow, PyTorch, NumPy (1.26.0), Scikit-learn (sklearn) και Pandas (2.0.3), οι οποίες διευκολύνουν την ανάπτυξη, εκπαίδευση και αξιολόγηση μοντέλων μηχανικής μάθησης.

Το dataset N-BaIoT υπέστη προ επεξεργασία, όπως περιγράφηκε στην προηγούμενη υποενότητα. Η τελική του μορφή περιλαμβάνει ένα global test set καθώς και τρία αρχεία για τους πελάτες, όπου το καθένα περιέχει ένα training set και ένα validation set. Υπενθυμίζεται ότι κάθε πελάτης αντιστοιχεί σε δεδομένα από δύο διαφορετικές συσκευές και περιλαμβάνει δύο από τους τρεις τύπους επιθέσεων (attacks).

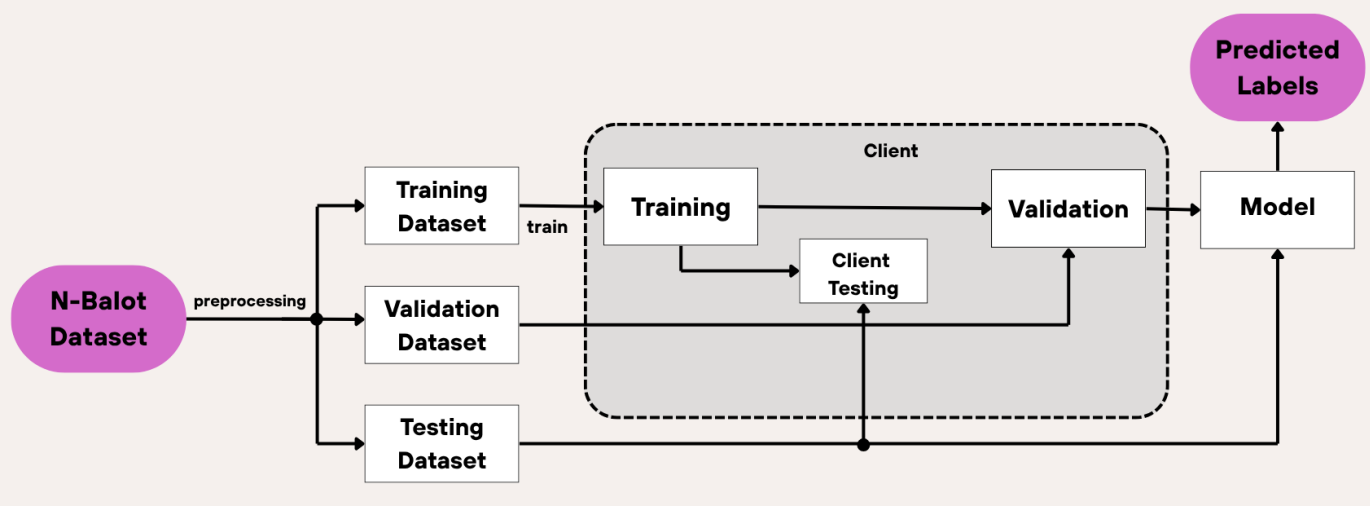
Client_1"devices": [1, 2], "attacks": mirai.ack, mirai.syn

Client_2 "devices": [4, 5], "attacks": mirai.syn, mirai.udpplain

Client_3 "devices": [6, 9], "attacks": mirai.ack, mirai.udpplain

4.2.1 Υλοποιήση Federated Learning

Η ροή της διαδικασίας εκπαίδευσης και αξιολόγησης που ακολουθήθηκε στην παρούσα εργασία παρουσιάζεται στην παρακάτω εικόνα και εξηγείται στην συνέχεια. Σαν πρώτο βήμα το N-Balot dataset υφίσταται προ επεξεργασία για να προκύψουν τα τελικά σύνολα δεδομένων με χαρακτηριστικά και ετικέτες (Features and Labels). Αυτά στη συνέχεια διαχωρίζονται σε τρία διακριτά σύνολα: Training Dataset, Validation Dataset και Testing Dataset. Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκμάθηση του μοντέλου που γίνεται σε τοπικό επίπεδο στους πελάτες, ενώ το σύνολο επικύρωσης χρησιμεύει για την εκτίμηση της απόδοσης του μοντέλου ώστε να επιλεγεί η καλύτερη εκδοχή του. Τέλος, το σύνολο δοκιμής χρησιμοποιείται για την αξιολόγηση μοντέλου αλλά και την τοπική αξιολόγηση των πελατών σε δεδομένα που δεν έχουν ξανά δει. Η έξοδος του τελικού εκπαιδευμένου μοντέλου είναι οι προβλεπόμενες ετικέτες (Predicted Labels).



Εικόνα 9: Διάγραμμα ροής της Ομοσπονδιακής Μάθησης

Για να πετύχουμε αυτό αρχικά δημιουργήσαμε ένα flwr app ,καθοδηγούμενοι από [21], με την χρήση του framework Flower (1.15.2), μέσα σε ένα Virtual environment. Το Flower app ορίζει ένα FL environment στο οποίο πολλαπλοί πελάτες εκπαιδεύουν ένα κοινόχρηστο μοντέλο τοπικά, ενώ ένας κεντρικός server συγκεντρώνει τις ενημερώσεις τους, χωρίς να μοιράζεται raw δεδομένα. Το framework παρέχει ενσωματωμένα εργαλεία για επικοινωνία client-server και model aggregation (Federated Averaging), καθιστώντας το ιδανικό για βάση της εργασίας μας.

Στην συνέχεια για να προσαρμόσουμε το σύστημα στις δικές μας ανάγκες χρησιμοποιήσαμε ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο (Fully Connected Neural Network – FCNN) με μία κρυφή στρώση, το οποίο υλοποιήθηκε σε PyTorch. Η επιλογή ενός απλού FCNN έγινε με σκοπό τη διατήρηση της υπολογιστικής αποδοτικότητας κατά την τοπική εκπαίδευση στους πελάτες, ενώ παρέχει επαρκή ικανότητα μάθησης για το συγκεκριμένο πρόβλημα. Η αρχιτεκτονική του μοντέλου περιλαμβάνει μία είσοδο με μέγεθος ίσο με τον αριθμό των χαρακτηριστικών του dataset (input size), μία κρυφή στρώση με 64 νευρώνες (hidden size) και μία έξοδο που αντιστοιχεί στις κλάσεις ταξινόμησης (output size). Αυτές οι κλάσεις είναι η "κανονική" κίνηση που επισημαίνεται με την ετικέτα 0, ενώ οι διαφορετικοί τύποι επιθέσεων έχουν τις εξής ετικέτες: mirai.ack → 1, mirai.udpplain → 2 και mirai.syn → 3. Μεταξύ των επιπέδων εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU για την εισαγωγή μη γραμμικότητας. Η συγκεκριμένη αρχιτεκτονική χρησιμοποιείται τόσο στους πελάτες όσο και στον server, εξασφαλίζοντας ομοιομορφία στην εκπαίδευση.

4.2.1.1 Server

Ο server της εφαρμογής FL έχει υλοποιηθεί χρησιμοποιώντας το πλαίσιο Flower, με βασική στρατηγική εκπαίδευσης τη Federated Averaging (FedAvg). Ο ρόλος του server είναι να διαχειρίζεται το παγκόσμιο μοντέλο δηλαδή να διατηρεί και να ενημερώνει τις παραμέτρους του με βάση τις τοπικές ενημερώσεις που στέλνουν οι πελάτες μετά την τοπική εκπαίδευσή τους. Παράλληλα, συγχρονίζει τους πελάτες, ελέγχει την εκτέλεση των γύρων εκπαίδευσης (training rounds) και αξιολογεί τη συνολική απόδοση του συστήματος. Στις επόμενες παραγράφους αναλύονται κάποιες από τις βασικές λειτουργίες του.

Κατά την αρχικοποίηση, ο server φορτώνει ένα κεντρικό σύνολο δοκιμής (global test set) μέσω της συνάρτησης load_data, και ορίζει τις αρχικές παραμέτρους του νευρωνικού δικτύου (input size, hidden layer, output classes). Η συνάρτηση get_evaluate_fn ορίζει τη διαδικασία αξιολόγησης του παγκόσμιου μοντέλου σε κάθε γύρο εκπαίδευσης, επιστρέφοντας μετρικές όπως accuracy, precision, recall και F1-score. Παράλληλα, γίνεται

[21] https://github.com/adap/flower/tree/main/examples/flower-simulation-step-by-step-pytorch

χρήση της weighted_average για τον υπολογισμό των σταθμισμένων μέσων όρων των αξιολογήσεων που λαμβάνονται από τους πελάτες.

Επιπλέον, εφαρμόζεται μια απλή δυναμική προσαρμογή του ρυθμού εκμάθησης (learning rate decay) μέσα από τη συνάρτηση on_fit_config, όπου μειώνεται η τιμή του learning rate μετά από έναν αριθμό γύρων για πιο σταθερή εκπαίδευση. Ξεκινάει με ένα υψηλό ρυθμό εκμάθησης (0,01) ,που βοηθά το μοντέλο να μαθαίνει γρήγορα στους πρώτους γύρους, και μετά τον γύρο 2 χαμηλώνει σε (0,005) που βελτιώνει το μοντέλο για να αποφευχθεί overfitting.

Η στρατηγική FedAvg καθορίζεται ώστε να συμμετέχουν το 100% των διαθέσιμων πελατών τόσο στην εκπαίδευση όσο και στην αξιολόγηση.Τέλος, όλα τα παραπάνω ενσωματώνονται στο αντικείμενο ServerApp, το οποίο ενεργοποιεί τον server και τον καθιστά έτοιμο να δεχθεί συνδέσεις από τους πελάτες.

4.2.1.2 Client

Η πλευρά του πελάτη στην υλοποίηση του FL ακολουθεί την κλάση NumPyClient, η οποία ορίζει τις βασικές λειτουργίες που απαιτούνται από κάθε συμμετέχοντα πελάτη σε ένα ομοσπονδιακό περιβάλλον εκπαίδευσης.

Κατά την αρχικοποίηση κάθε πελάτη, φορτώνεται ένα τοπικό υποσύνολο δεδομένων το οποίο περιλαμβάνει σετ εκπαίδευσης, επικύρωσης και δοκιμής (training, validation και test set αντίστοιχα). Ο πελάτης υλοποιεί ένα νευρωνικό δίκτυο το οποίο εκπαιδεύεται τοπικά σε πολλαπλές εποχές (local epochs) με βάση τις παραμέτρους του εκάστοτε γύρου που αποστέλλει ο server.

Η μέθοδος fit() αναλαμβάνει την τοπική εκπαίδευση του μοντέλου, ενημερώνοντας τα βάρη του δικτύου και υπολογίζοντας την απώλεια εκπαίδευσης (training loss). Στο τέλος κάθε γύρου, το εκπαιδευμένο μοντέλο αξιολογείται τοπικά με το σετ δοκιμής μέσω της test_client_model(), παρέχοντας ενδείξεις για την απόδοσή του στον εκάστοτε πελάτη, αν και αυτά τα αποτελέσματα δεν αποστέλλονται στον server. Η τοπική αξιολόγηση προσφέρει μία πιο πλήρη εικόνα της απόδοσης του μοντέλου σε ένα ανεξάρτητο test set. Αυτό το test set περιλαμβάνει δείγματα που δεν έχουν χρησιμοποιηθεί κατά την τοπική εκπαίδευση του πελάτη και περιέχει τύπους επιθέσεων που ο συγκεκριμένος πελάτης δεν έχει ξαναδεί. Η χρήση του επιτρέπει την αξιολόγηση της γενίκευσης του μοντέλου, δηλαδή το κατά πόσο μπορεί να αναγνωρίσει και να εντοπίσει άγνωστες επιθέσεις, αξιοποιώντας τη γνώση που έχει αποκτήσει μέσω του συνεργατικού FL. Έτσι, η τοπική αξιολόγηση με αυτό το test set παρέχει κρίσιμες ενδείξεις για την ικανότητα κάθε πελάτη να ανταποκρίνεται σε πραγματικά σενάρια, όπου οι απειλές είναι μεταβαλλόμενες και μη προβλέψιμες. Υπολογίζονται και καταγράφονται οι μετρικές accuracy, precision, recall και F1-score, δίνοντας μας σημαντική

πληροφόρηση για την ποιότητα του μοντέλου σε κάθε μεμονωμένο πελάτη. Κατά την τοπική εκπαίδευση, χρησιμοποιείται ο βελτιστοποιητής Adam, ο οποίος προσφέρει ταχεία σύγκλιση μέσω της προσαρμογής του learning rate για κάθε παράμετρο, ενώ ως συνάρτηση απώλειας επιλέγεται η CrossEntropyLoss, κατάλληλη για προβλήματα πολυταξινόμησης όπως το παρόν. Επιπλέον, εφαρμόζεται τεχνική early stopping κατά την τοπική εκπαίδευση των πελατών, με στόχο την αποτροπή του overfitting και την εξοικονόμηση υπολογιστικών πόρων. Ο μηχανισμός αυτός παρακολουθεί τη συνάρτηση απώλειας στο validation loss και διακόπτει την εκπαίδευση εάν δεν παρατηρηθεί βελτίωση για τρεις συνεχόμενες εποχές (patience = 3). Έτσι, η διαδικασία τερματίζεται πρόωρα διατηρώντας τα βάρη του καλύτερου μοντέλου.

Αντίθετα, η μέθοδος evaluate() εκτελεί αξιολόγηση στο σετ επικύρωσης (validation set) και επιστρέφει στον server μετρικές όπως accuracy και F1-score, οι οποίες αξιοποιούνται για την ομοσπονδιακή σύγκλιση του παγκόσμιου μοντέλου. Η αξιολόγηση αυτή εξασφαλίζει ότι ο server λαμβάνει στατιστικά στοιχεία από όλους τους πελάτες και μπορεί να υπολογίσει σταθμισμένους μέσους όρους.

Μετά την ολοκλήρωση της τοπικής εκπαίδευσης και της αξιολόγησης, ο πελάτης επιστρέφει στο server τα ενημερωμένα βάρη του μοντέλου μέσω της ClientApp. Η ClientApp αποτελεί το λογισμικό μέρος κάθε πελάτη που διαχειρίζεται την εκπαίδευση με τα τοπικά δεδομένα και τη μεταφορά των παραμέτρων πίσω στον server. Αυτός ο κύκλος επιτρέπει τη συνεχή ενημέρωση του παγκόσμιου μοντέλου με βάση τις τοπικές συνεισφορές όλων των πελατών.

4.2.2 Υλοποιήση Centralized Learning

Για να αξιολογηθεί η αποτελεσματικότητα του συστήματος FL και να πραγματοποιηθεί δίκαιη σύγκριση, υλοποιήθηκε ένα σενάριο κεντρικής εκπαίδευσης (centralized training). Σε αυτό το σενάριο, τα τοπικά δεδομένα των πελατών συνενώνονται σε ένα ενιαίο dataset, πάνω στο οποίο εκπαιδεύεται το μοντέλο με κλασική supervised μάθηση. Η υλοποίηση πραγματοποιήθηκε σε PyTorch, χρησιμοποιώντας το ίδιο μοντέλο (FCNN) με εκείνο που χρησιμοποιείται στο ομοσπονδιακό περιβάλλον, ώστε η σύγκριση να είναι άμεση και αντικειμενική.

Συγκεκριμένα, φορτώθηκαν τα δεδομένα εκπαίδευσης και επικύρωσης από όλους τους πελάτες (client_1, client_2, client_3), τα οποία συνενώθηκαν με τη χρήση της ConcatDataset. Το ενιαίο σύνολο εκπαίδευσης χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου για 15 εποχές, ενώ το validation set βοήθησε στην αξιολόγηση της απόδοσης μετά από κάθε epoch. Έγινε προσθήκη μηχανισμού early stopping και σε αυτό το σύστημα, βασισμένο στη συνάρτηση απώλειας στο validation loss και patience = 3. Έτσι αποφεύγεται και εδώ το overfitting και η διαδικασία εκπαίδευσης τερματίζεται νωρίτερα, διατηρώντας το

μοντέλο που επιτεύχθηκε στην καλύτερη εποχή. Το μοντέλο εκπαιδεύτηκε με χρήση της συνάρτησης απώλειας CrossEntropy και του βελτιστοποιητή Adam.

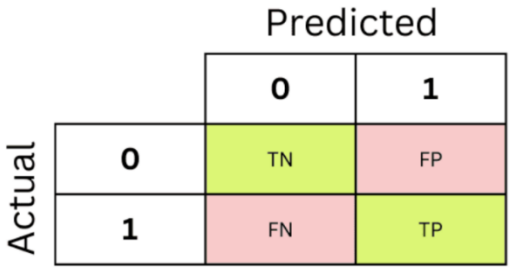
Τέλος, το εκπαιδευμένο μοντέλο αξιολογήθηκε χρησιμοποιώντας το κοινό παγκόσμιο σύνολο δοκιμής (global test set), ώστε να διασφαλιστεί η συγκρισιμότητα με τα αποτελέσματα του FL, και καταγράφηκαν μετρικές όπως ακρίβεια, precision, recall και F1-score.

4.3 Μετρικές Αξιολόγησης

Για την αξιολόγηση των μοντέλων ταξινόμησης χρησιμοποιούνται τέσσερις ευρέως διαδεδομένες μετρικές στον χώρο της Μηχανικής Μάθησης. Αναλυτικά, οι μετρικές αυτές είναι οι ακόλουθες [22]:

Accuracy, η οποία ορίζεται ως ο λόγος των σωστά ταξινομημένων δειγμάτων προς το συνολικό πλήθος των δειγμάτων. Η μετρική αυτή αποτυπώνει τη συνολική ακρίβεια του μοντέλου και αποτελεί έναν βασικό δείκτη της γενικής του απόδοσης

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$



Precision, μετρά την ακρίβεια των θετικών προβλέψεων. Απαντά στην ερώτηση: «Από όλα τα στοιχεία που το μοντέλο χαρακτήρισε ως θετικά, πόσα ήταν στην πραγματικότητα θετικά;»

[22] https://medium.com/@piyushkashyap045/understanding-precision-recall-and-f1-score-metrics-ea219b908093

$$Precision = \frac{TP}{TP + FP}$$

Recall ,μετρά την ικανότητα του μοντέλου να βρίσκει όλα τα θετικά παραδείγματα. Απαντά στην ερώτηση «Από όλα τα πραγματικά θετικά, πόσα αναγνώρισε σωστά το μοντέλο;»

$$Recall = \frac{TP}{TP + FN}$$

F1 Score, ο αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης. Εξισορροπεί τις δύο μετρήσεις σε έναν μόνο αριθμό, καθιστώντας τον ιδιαίτερα χρήσιμο όταν η ακρίβεια και η ανάκληση βρίσκονται σε αντιστάθμιση.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

5. Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα της εκπαίδευσης και αξιολόγησης των μοντέλων τόσο στο FL όσο και στο κεντρικοποιημένο CL σενάριο. Στόχος είναι η αποτίμηση της απόδοσης των δύο προσεγγίσεων με βάση τις καθιερωμένες μετρικές που αναφέρθηκαν παραπάνω. Μέσα από την σύγκριση των μοντέλων, αναδεικνύονται τα πλεονεκτήματα και οι περιορισμοί της κάθε μεθόδου, ιδίως σε ό,τι αφορά την ποιότητα της ταξινόμησης και την αποδοτικότητα σε κατανεμημένα περιβάλλοντα εκπαίδευσης. Παράλληλα, εξετάζεται η ικανότητα γενίκευσης του κάθε μοντέλου σε άγνωστα δεδομένα, μέσω της αξιολόγησης σε global test set.

5.1 Απόδοση στο Federated Learning

Η παρούσα υποενότητα εστιάζει στην αξιολόγηση της FL διαδικασίας, παρουσιάζοντας τα αποτελέσματα που προέκυψαν από τη συνεργασία των επιμέρους πελατών. Η εκπαιδευτική διαδικασία υλοποιήθηκε για 10 ομοσπονδιακούς γύρους, num_server_rounds = 10, ενώ κάθε πελάτης εκπαίδευσε το τοπικό του μοντέλο για έως και 30 εποχές, local_epochs = 30.

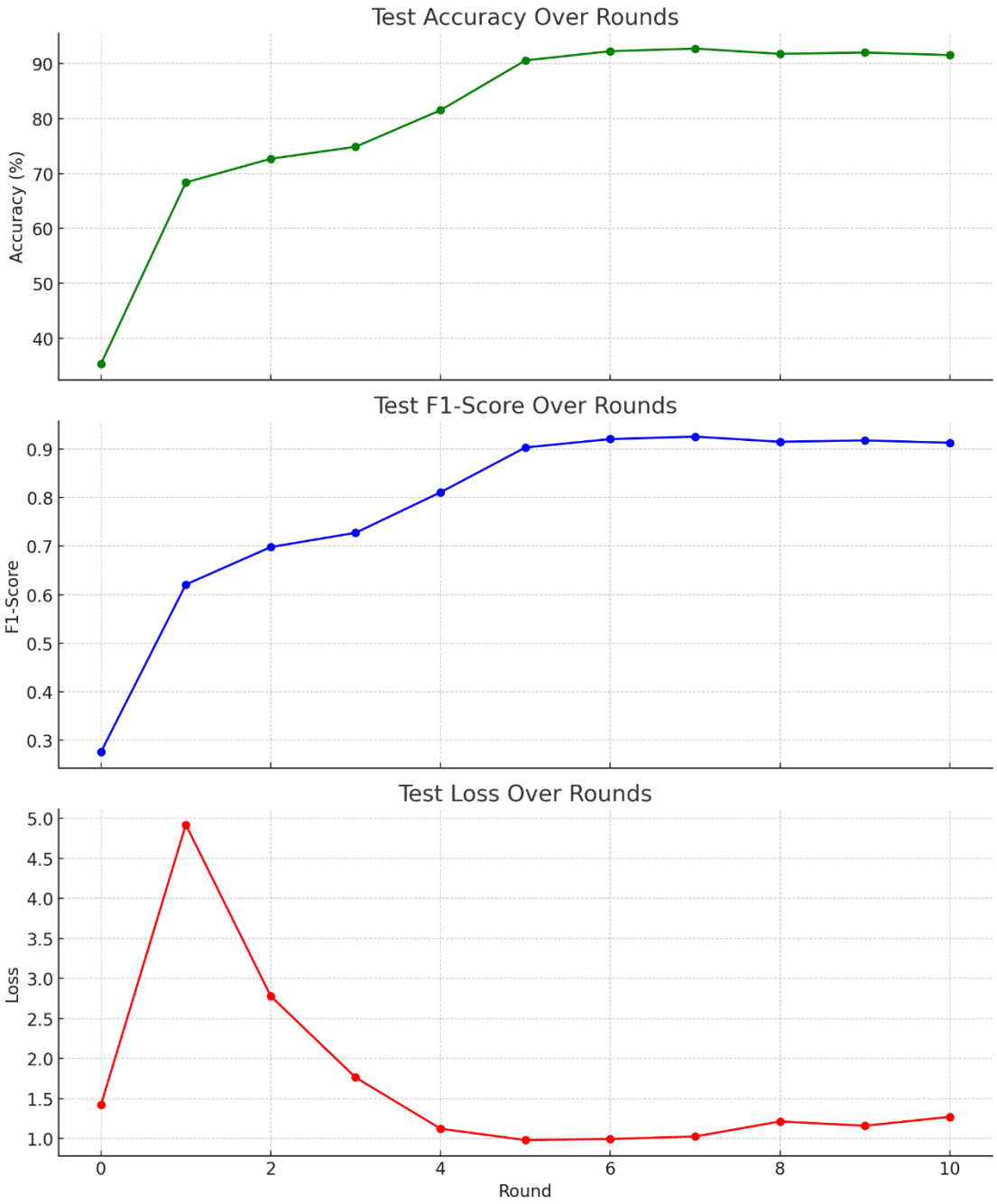
Ακολουθούν αναλυτικά τα αποτελέσματα της διαδικασίας, τόσο σε επίπεδο παγκόσμιου μοντέλου, όσο και επιμέρους αξιολογήσεις που καταγράφηκαν κατά τη διάρκεια των γύρων εκπαίδευσης.

Table 2: Απόδοση FL per round στο aggregated μοντέλο

Round	Test Loss	Test Accuracy	Test F1-Score
0	1.4212	35.39%	0.2758
1	4.9202	68.37%	0.6209
2	2.7801	72.67%	0.6981
3	1.7642	74.85%	0.7275
4	1.1243	81.49%	0.8109
5	0.9822	90.54%	0.9033
6	0.9957	92.22%	0.9206

Round	Test Loss	Test Accuracy	Test F1-Score
7	1.0269	92.69%	0.9256
8	1.2153	91.73%	0.9150
9	1.1611	91.97%	0.9178
10	1.2729	91.51%	0.9130

Παρόλο που η εκπαίδευση πραγματοποιήθηκε για 10 rounds, το μοντέλο πέτυχε την υψηλότερη απόδοσή του στο κεντρικό σύνολο δοκιμών στον γύρο 6, με ακρίβεια δοκιμής 92,22% και σταθμισμένη βαθμολογία F1 0,9206. Οι επόμενοι γύροι δεν οδήγησαν σε σημαντικές βελτιώσεις και σε ορισμένες περιπτώσεις εισήγαγαν μικρές υποβαθμίσεις, υποδηλώνοντας ότι το μοντέλο είχε ουσιαστικά συγκλίνει.



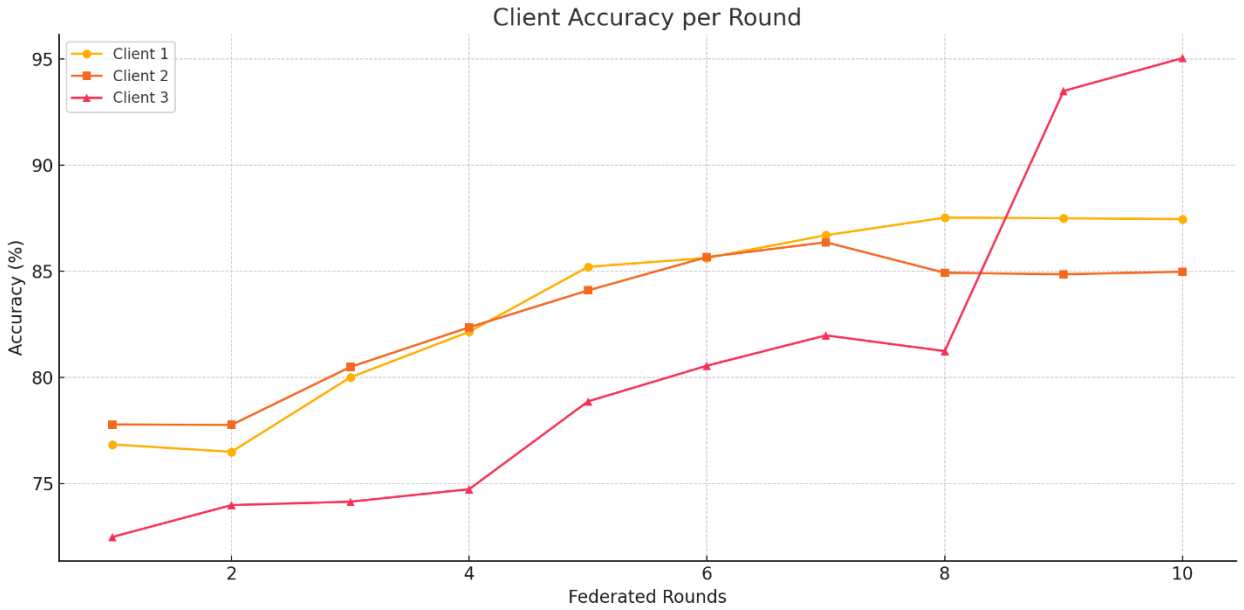
Εικόνα 10: Διάγραμμα απόδοσης FL per round στο aggregated μοντέλο

Παρατηρούμε και στο διάγραμμα ότι η ακρίβεια του παγκόσμιου μοντέλου αυξάνεται σταθερά κατά τη διάρκεια των rounds, φτάνοντας σε μέγιστο ποσοστό 92.69% στο 7ο round, ξεκινώντας από 35.39% στον 0 round. Παρόμοια πορεία ακολουθεί και ο δείκτης F1-Score, ο οποίος κορυφώνεται στο 0.9256, υποδεικνύοντας ισορροπία μεταξύ precision και recall. Η απώλεια στο test σετ μειώνεται γενικά με την πρόοδο των rounds, αν και παρατηρείτε μικρή αύξηση στους τελευταίους γύρους.

Table 3: Test Evaluation Results για τους Clients

Rounds	Epochs 1	Accuracy 1	Epochs 2	Accuracy 2	Epochs 3	Accuracy 3
1	4	76.84%	6	77.78%	24	72.48%
2	4	76.49%	7	77.76%	15	73.98%
3	4	80.00%	4	80.49%	10	74.14%
4	4	82.15%	4	82.36%	6	74.73%
5	4	85.21%	4	84.10%	5	78.87%
6	4	85.63%	4	85.67%	4	80.55%
7	4	86.70%	4	86.37%	6	81.98%
8	4	87.53%	4	84.93%	9	81.24%
9	4	87.50%	4	84.86%	8	93.50%
10	4	87.46%	4	84.98%	6	95.05%

Ο παραπάνω πίνακας παρουσιάζει την απόδοση (accuracy) των τριών πελατών ανά γύρο εκπαίδευσης στο πλαίσιο του FL. Όπως φαίνεται, κάθε πελάτης εκπαιδεύεται τοπικά για διαφορετικό αριθμό εποχών (epochs) σε κάθε γύρο, λόγω της εφαρμογής του μηχανισμού early stopping, ο οποίος σταματά την εκπαίδευση νωρίτερα όταν δεν παρατηρείται βελτίωση στο validation loss. Η διαφοροποίηση αυτή επιτρέπει στους πελάτες να αποφεύγουν το υπερεκπαίδευση και να προσαρμόζουν τη διάρκεια εκπαίδευσης με βάση τα δικά τους τοπικά δεδομένα, οδηγώντας σε αποδοτικότερη και πιο προσαρμοστική διαδικασία εκπαίδευσης.



Εικόνα 11: Client Accuracy per Round

Βάσει των αποτελεσμάτων του table 4 και της αντίστοιχης γραφικής παράστασης στην Εικόνα 10, παρατηρείται ότι η απόδοση των τριών πελατών βελτιώνεται σταδιακά με την πρόοδο των ομοσπονδιακών γύρων. Οι πελάτες 1 και 2 παρουσιάζουν σταθερή και προβλέψιμη αύξηση της ακρίβειας, ξεκινώντας από περίπου 76–77% και φτάνοντας σταθερά έως και το 87.5% στον γύρο 10. Αυτή η ομαλή ανοδική πορεία καταδεικνύει ότι το κοινό μοντέλο μαθαίνει αποτελεσματικά και προσαρμόζεται στα δεδομένα των πελατών. Ο πελάτης 3, από την άλλη, ξεκινά με χαμηλότερη απόδοση περίπου 72.5% αλλά παρουσιάζει απότομη βελτίωση από τον 5ο round και μετά, φτάνοντας στο 95.05% στον τελευταίο γύρο. Συνολικά, τα αποτελέσματα αναδεικνύουν τη συνεπή συνεισφορά όλων των πελατών στη βελτίωση του παγκόσμιου μοντέλου και την αποτελεσματικότητα της ομοσπονδιακής εκπαίδευσης.

5.2 Απόδοση στο Centralized Learning

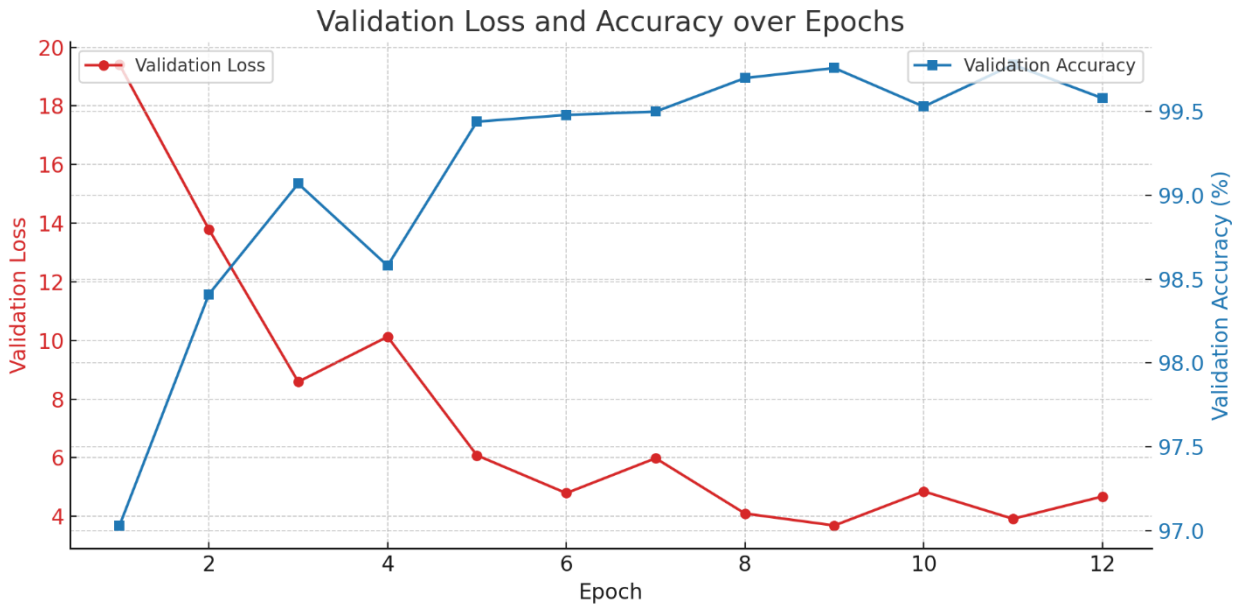
Εδώ παρουσιάζουμε τα αποτελέσματα που προέκυψαν από το Centralized μοντέλο, το οποίο εκπαιδεύτηκε χρησιμοποιώντας δεδομένα από όλους τους πελάτες συγκεντρωμένα σε ένα ενιαίο σύνολο. Στόχος αυτής της υλοποίησης ήταν να αποτελέσει σημείο αναφοράς για τη σύγκριση με την απόδοση του Ομοσπονδιακού Μοντέλου.

Table 4: Validation values over epochs

Epoch	Train Loss	Validation Loss	Validation Accuracy
1	171.4426	19.4051	97.03%
2	76.3621	13.7890	98.41%
3	53.0184	8.5895	99.07%
4	41.4776	10.1262	98.58%
5	31.0814	6.0809	99.44%
6	24.9746	4.7925	99.48%
7	22.0183	5.9829	99.50%

Epoch	Train Loss	Validation Loss	Validation Accuracy
8	19.6074	4.0969	99.70%
9	17.0532	3.6889	99.76%
10	21.7000	4.8527	99.53%
11	17.8204	3.9160	99.78%
12	19.1890	4.6774	99.58%

Κατά την εκπαίδευση του κεντρικοποιημένου μοντέλου εφαρμόστηκε μηχανισμός early stopping, ο οποίος ενεργοποιήθηκε στο epoch 12, καθώς η συνάρτηση απώλειας επικύρωσης δεν παρουσίασε βελτίωση για τρεις συνεχόμενες εποχές. Το καλύτερο μοντέλο καταγράφηκε στο epoch 9, όπου επιτεύχθηκε το χαμηλότερο validation loss, γεγονός που το καθιστά ως το βέλτιστο σημείο εκπαίδευσης. Σε αυτό το σημείο, η ακρίβεια επικύρωσης (validation accuracy) έφτασε το 99.76%, ενώ η τελική ακρίβεια στο σύνολο δοκιμής (test accuracy) ήταν 99.63%, αποδεικνύοντας την εξαιρετική γενίκευση του μοντέλου. Επιπλέον, η απώλεια στο test set ήταν ιδιαίτερα χαμηλή στο 0.0181, ενισχύοντας την εικόνα υψηλής απόδοσης. Οι τιμές του validation loss δείχνουν ότι δεν υπάρχει overfitting.



Εικόνα 12: Validation Loss and Accuracy over Epochs

Στο παραπάνω γράφημα απεικονίζονται το validation loss και το validation accuracy καθ' όλη τη διάρκεια των 12 εποχών εκπαίδευσης. Παρατηρείται ότι η απώλεια επικύρωσης μειώνεται γενικά με την πρόοδο της εκπαίδευσης, παρουσιάζοντας όμως μικρές διακυμάνσεις, όπως για παράδειγμα μια απότομη αύξηση στις εποχές 4 και 10. Αντίθετα, η ακρίβεια επικύρωσης βελτιώνεται σταθερά και φαίνεται να σταθεροποιείται γύρω στην εποχή 9, κάτι που ευθυγραμμίζεται με την ενεργοποίηση του μηχανισμού early stopping. Η πορεία αυτή επιβεβαιώνει την απόφαση για πρόωρη διακοπή της εκπαίδευσης, καθώς τα περιθώρια βελτίωσης μετά το 9ο epoch είναι ελάχιστα και σε ορισμένες περιπτώσεις σημειώνεται και μικρή μείωση στην απόδοση.

Table 5: Classification Report on Test Set

Attack	Precision	Recall	F1-Score
0	0.9996	0.9998	0.9997
1	0.9900	0.9937	0.9918
2	0.9940	0.9903	0.9922

Attack	Precision	Recall	F1-Score
3	1.0000	0.9997	0.9998

Αναλύοντας τα αποτελέσματα του ταξινομητή ανά επίθεση, παρατηρείται ότι η επίθεση 3 (mirai.syn) επιτυγχάνει τέλειο precision ίσο με 1.0000, γεγονός που δείχνει πως όλα τα δείγματα που ταξινομήθηκαν ως επίθεση 3 ήταν σωστά. Αντίθετα, η επίθεση 1 (mirai.ack) παρουσιάζει τη χαμηλότερη ακρίβεια (0.9900), αν και εξακολουθεί να κινείται σε πολύ υψηλό επίπεδο. Από την πλευρά του recall, η επίθεση 0 (normal) ξεχωρίζει με σχεδόν τέλεια τιμή 0.9998, υποδηλώνοντας ότι σχεδόν όλα τα πραγματικά δείγματα αυτής της επίθεσης αναγνωρίστηκαν σωστά. Τέλος, όλοι τα F1-scores είναι μεγαλύτερα από 0.99, γεγονός που επιβεβαιώνει την εξαιρετική ισορροπία ανάμεσα στο precision και recall και αποδεικνύει τη συνολικά πολύ υψηλή απόδοση του μοντέλου.

Table 6: Other Metrics for CL

Metric	Value
Accuracy	99.63%
Macro Avg	0.9959
Weighted Avg	0.9963
Total Support	13,500

5.3 Σύγκριση Centralized vs Federated

Η συγκριτική μελέτη ανάμεσα στο μοντέλο FL και στο CL αποκαλύπτει ενδιαφέροντα ευρήματα ως προς την αποτελεσματικότητα και τις επιδόσεις τους. Το παγκόσμιο μοντέλο του FL, το οποίο εκπαιδεύτηκε μέσω της συνεργασίας τριών clients σε 10 ομοσπονδιακούς γύρους, πέτυχε τη μέγιστη ακρίβεια 92.69% στον γύρο 7 αλλά την καλύτερη απόδοση στον γύρο 6 με ακρίβεια 92.22% και test loss 0.9957. Αντίθετα το CL μοντέλο που εκπαιδεύτηκε σε ενιαίο σύνολο δεδομένων, εμφάνισε σημαντικά υψηλότερη τελική απόδοση, με Test Accuracy 99.63%, F1-score > 0.99 σε όλες τις κλάσεις, και εξαιρετικά χαμηλό test loss (0.0181).

Παρόλο που το CL υπερέχει σε ακρίβεια και γενίκευση, το FL κατάφερε να επιτύχει υψηλή απόδοση χωρίς να συγκεντρώνει τα δεδομένα σε έναν κεντρικό server, διατηρώντας την ιδιωτικότητα και μειώνοντας τους κινδύνους διαρροής ευαίσθητων πληροφοριών. Επιπλέον, η συμπεριφορά του FL ήταν συνεπής και σταθερή, με όλους τους πελάτες να παρουσιάζουν προοδευτική βελτίωση, κάτι που αναδεικνύει τη λειτουργικότητα του ομοσπονδιακού σχήματος.

Συμπερασματικά, ενώ το CL υπερτερεί σε απόλυτους αριθμούς απόδοσης, το FL προσφέρει έναν ρεαλιστικό και αποτελεσματικό συμβιβασμό για κατανεμημένα συστήματα με περιορισμούς στην ιδιωτικότητα και στην υπολογιστική ισχύ, διατηρώντας ικανοποιητική ακρίβεια που σε πολλές εφαρμογές μπορεί να θεωρηθεί επαρκής.

6. Συμπεράσματα και Μελλοντική Εργασία

6.1 Σχολιασμός

Ένα από τα πλέον καθοριστικά ευρήματα ήταν η ικανότητα των πελατών να εντοπίζουν επιθέσεις εκτός του δικού τους αρχικού dataset, κάτι που δεν είναι αυτονόητο σε ένα κατανεμημένο σενάριο με μη ομοιογενή δεδομένα. Αυτό υποδεικνύει ότι το παγκόσμιο μοντέλο κατάφερε να ενσωματώσει τη γνώση από όλους τους συμμετέχοντες πελάτες και να τη διανείμει ξανά προς τα πίσω. Είναι ένα σαφές παράδειγμα επιτυχημένης συνεργατικής μάθησης σε περιβάλλον χωρίς ανταλλαγή δεδομένων.

Αυτό το φαινόμενο έχει κρίσιμη σημασία σε πραγματικά περιβάλλοντα κυβερνοασφάλειας, όπου οι επιθέσεις είναι πολυμορφικές και δεν είναι πάντα εκ των προτέρων γνωστές σε όλους τους κόμβους του συστήματος. Η δυνατότητα ενός πελάτη να «μάθει» να ανιχνεύει τέτοιες απειλές από τη συλλογική εμπειρία του δικτύου, χωρίς να έχει δει τα αντίστοιχα δείγματα τοπικά, ενισχύει δραστικά την προστασία και την ανθεκτικότητα των συστημάτων.

Επιπρόσθετα η απλότητα του μοντέλου επιτρέπει γρήγορη σύγκλιση και χαμηλό υπολογιστικό κόστος, καθιστώντας το κατάλληλο για ομοσπονδιακά περιβάλλοντα, ενώ παράλληλα αφήνει περιθώρια για μελλοντική επέκταση με επιπλέον κρυφά επίπεδα. Η γρήγορη σύγκλιση του μοντέλου, όπως παρατηρήθηκε στους πρώτους γύρους, υποδεικνύει ότι οι τοπικές ενημερώσεις ήταν επαρκώς αντιπροσωπευτικές για το παγκόσμιο μοντέλο. Παρά τις διαφορές στα δεδομένα ανά πελάτη, η χρήση στρατηγικών όπως το Federated Averaging, η δυναμική ρύθμιση του learning rate, αλλά και η προσεκτικά σχεδιασμένη αρχιτεκτονική δικτύου (FCNN), επέτρεψαν τη σταθερή και αποδοτική μάθηση.

Επίσης αξίζει να σημειωθεί ότι το μοντέλο πέτυχε απόδοση που πλησιάζει αυτή της κεντρικοποιημένης εκπαίδευσης, γεγονός που ενισχύει την εμπιστοσύνη στην πρακτική αξιοποίηση του FL χωρίς σημαντικές θυσίες σε απόδοση.

Η μη μεταφορά δεδομένων προς κεντρικό server αποτελεί σημαντικό όφελος του FL, ειδικά σε εφαρμογές όπου η προστασία προσωπικών ή ευαίσθητων πληροφοριών είναι καθοριστική. Επιπλέον, η εργασία ανέδειξε ότι το FL δεν προσφέρει μόνο προστασία της ιδιωτικότητας, αλλά και ενισχύει την ικανότητα κάθε πελάτη να εντοπίζει επιθέσεις που δεν περιλαμβάνονται στα δικά του τοπικά δεδομένα. Αυτό οφείλεται στο ότι κάθε συμμετέχων επωφελείται από τη συνδυασμένη γνώση των υπολοίπων. Γι' αυτό το λόγο, το FL μπορεί να αποδειχθεί χρήσιμο σε πραγματικές συνθήκες, όπου οι επιθέσεις είναι διαφορετικές για κάθε χρήστη και συνεχώς μεταβάλλονται.

6.2 Σύνοψη και Συμπεράσματα

Στην παρούσα εργασία μελετήθηκε και υλοποιήθηκε ένα πλήρες σύστημα FL με χρήση του framework Flower και της βιβλιοθήκης PyTorch. Στόχος ήταν η εκπαίδευση ενός κοινού μοντέλου σε περιβάλλον κατανεμημένης υπολογιστικής, χωρίς τη μεταφορά δεδομένων προς έναν κεντρικό server, διατηρώντας έτσι την ιδιωτικότητα και την τοπικότητα των δεδομένων.

Η αρχιτεκτονική που επιλέχθηκε βασίστηκε σε ένα FCNN με μία κρυφή στρώση, το οποίο εφαρμόστηκε ομοιόμορφα τόσο στους πελάτες όσο και στον server. Η διαδικασία εκπαίδευσης εκτελέστηκε για 10 ομοσπονδιακούς γύρους, με τοπική εκπαίδευση διάρκειας έως 30 εποχών, ενισχυμένη με early stopping για αποφυγή overfitting και βελτίωση της αποδοτικότητας.

Τα αποτελέσματα έδειξαν σημαντική βελτίωση της απόδοσης του παγκόσμιου μοντέλου από τον πρώτο γύρο (accuracy 35.39%, F1-score 0.27) έως τον 7ο γύρο (peak accuracy 92.69%, F1-score 0.9256), αποδεικνύοντας ότι η προσέγγιση συγκλίνει γρήγορα και αποτελεσματικά. Η ανάλυση των αποτελεσμάτων σε επίπεδο πελάτη ανέδειξε μια σταθερή βελτίωση της ακρίβειας σε τοπικό επίπεδο, επιβεβαιώνοντας την ικανότητα του μοντέλου να γενικεύει αποτελεσματικά και να εντοπίζει μοτίβα που δεν είχε προηγουμένως συναντήσει. Η εφαρμογή στρατηγικών όπως το Federated Averaging και η δυναμική προσαρμογή του learning rate συνέβαλαν στην ευστάθεια της εκπαίδευσης και στην αποτελεσματική σύγκλιση.

Συγκριτικά με την κεντρικοποιημένη εκπαίδευση, το FL μοντέλο παρουσίασε χαμηλότερη τελική απόδοση αλλά πρόσφερε αυξημένη ασφάλεια και προστασία ιδιωτικότητας. Η εργασία ανέδειξε τις δυνατότητες και τα πλεονεκτήματα της FL σε περιβάλλοντα με κατανεμημένα δεδομένα.

Συνολικά, η μελέτη αποδεικνύει ότι το FL αποτελεί μία πολλά υποσχόμενη προσέγγιση για ασφαλή και αποδοτική εκπαίδευση μοντέλων σε πραγματικά κατανεμημένα περιβάλλοντα. Οι παρατηρήσεις που καταγράφηκαν θέτουν γερές βάσεις για περαιτέρω ανάπτυξη και εξέλιξη του πεδίου.

6.3 Μελλοντική Εργασία

Η παρούσα εργασία έθεσε τα θεμέλια για την υλοποίηση και αξιολόγηση ενός συστήματος FL, βασισμένου στο framework Flower και προσαρμοσμένο σε σενάριο ανίχνευσης επιθέσεων μέσω κατανεμημένων πελατών. Ωστόσο, υπάρχουν πολλαπλές κατευθύνσεις στις οποίες μπορεί να επεκταθεί στο μέλλον η συγκεκριμένη μελέτη.

Θα μπορούσε να διερευνηθεί η χρήση πιο πολύπλοκων αρχιτεκτονικών νευρωνικών δικτύων, όπως CNNs ή Recurrent Neural Networks (RNNs), οι οποίες ενδέχεται να αποδώσουν καλύτερα.

Μία σημαντική προέκταση αφορά τη δοκιμή διαφορετικών αλγορίθμων εκπαίδευσης και aggregation, όπως οι FedProx, FedNova, FedAdam και Scaffold, οι οποίοι έχουν σχεδιαστεί για να αντιμετωπίζουν προκλήσεις όπως η ετερογένεια δεδομένων (non-IID) και η μη συγχρονισμένη συμμετοχή πελατών. Αυτοί οι αλγόριθμοι μπορεί να βελτιώσουν τη σταθερότητα και την απόδοση του παγκόσμιου μοντέλου σε πιο σύνθετα περιβάλλοντα.

Μια ακόμη ενδιαφέρουσα προέκταση αποτελεί η μετάβαση από το παραδοσιακό, κεντρικά ελεγχόμενο σχήμα σε ένα αποκεντρωμένο περιβάλλον FL, στο οποίο απουσιάζει εντελώς ο κεντρικός server. Σε αυτό το σενάριο, οι πελάτες ανταλλάσσουν απευθείας μεταξύ τους τα μοντέλα ή τις παραμέτρους τους, διατηρώντας τη γνώση που έχει αποκτηθεί τοπικά και προωθώντας την σταδιακά μέσα στο δίκτυο. Η αυθεντικότητα και αξιοπιστία των τοπικών μοντέλων σε μια τέτοια ρύθμιση μπορούν να διασφαλιστούν μέσω της τεχνολογίας blockchain [29], [74].

Βιβλιογραφία

[1] A. Karunamurthy, K. Vijayan, P. R. Kshirsagar, and K. T. Tan, "An optimal federated learning-based intrusion detection for IoT environment," *Sci Rep*, vol. 15, no. 1, pp. 1–19, Dec. 2025, doi: 10.1038/S41598-025-93501-8;SUBJMETA=166,639,987;KWRD=ELECTRICAL+AND+ELECTRONIC+ENGINEERING, ENGINEERING.

[2] S. A. Mahmud, N. Islam, Z. Islam, Z. Rahman, and S. T. Mehedi, "Privacy-Preserving Federated Learning-Based Intrusion Detection Technique for Cyber-Physical Systems," *Mathematics 2024, Vol. 12, Page 3194*, vol. 12, no. 20, p. 3194, Oct. 2024, doi: 10.3390/MATH12203194.

[3] B. Olanrewaju-George and B. Pranggono, "Federated learning-based intrusion detection system for the internet of things using unsupervised and supervised deep learning models," *Cyber Security and Applications*, vol. 3, p. 100068, Dec. 2025, doi: 10.1016/J.CSA.2024.100068.

[4] Y. Fan, Y. Li, M. Zhan, H. Cui, and Y. Zhang, "IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT," in *2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE)*, Dec. 2020, pp. 88–95. doi: 10.1109/BigDataSE50710.2020.00020.

[5] P. Radoglou-Grammatikis, P. Sarigiannidis, G. Efstathopoulos, T. Lagkas, G. Fragulis, and A. Sarigiannidis, "A Self-Learning Approach for Detecting Intrusions in Healthcare Systems," *IEEE International Conference on Communications*, Jun. 2021, doi: 10.1109/ICC42927.2021.9500354.

[6] I. Makris *et al.*, "Elevating 5G Network Security: A Profound Examination of Federated Learning Aggregation Strategies for Attack Detection," in *2023 IEEE Future Networks World Forum (FNWF)*, Nov. 2023, pp. 1–6. doi: 10.1109/FNWF58287.2023.10520474.

[7] D. S. Baggam and S. Rani, "Autonomous Systems for 5G Networks: A Comprehensive Analysis of Features Toward Generalization and Adaptability," *Current and Future Cellular Systems: Technologies, Applications, and Challenges*, pp. 107–137, Jan. 2025, doi: 10.1002/9781394256075.CH6.

[8] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 3, pp. 1657–1681, Jul. 2017, doi: 10.1109/COMST.2017.2705720.

[9] V. O. Nyangaresi, Z. A. Abduljabbar, M. A. Al Sibahee, I. Q. Abduljaleel, and E. W. Abood, "Towards Security and Privacy Preservation in 5G Networks," *2021 29th Telecommunications Forum, TELFOR 2021 - Proceedings*, 2021, doi: 10.1109/TELFOR52709.2021.9653385.

[10] A. Sukumar, A. Singh, A. Gupta, and M. Singh, "Enhancing Security and Privacy Implications in 5G Network Slicing," *2024 4th International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2024*, 2024, doi: 10.1109/ICAECT60202.2024.10469006.

[11] I. Ahmad, S. Shahabuddin, T. Kumar, J. Okwuibe, A. Gurtov, and M. Ylianttila, "Security for 5G and beyond," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3682–3722, Oct. 2019, doi: 10.1109/COMST.2019.2916180.

[12] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated Learning for 6G Communications: Challenges, Methods, and Future Directions," *China Communications*, vol. 17, no. 9, pp. 105–118, Jun. 2020, doi: 10.23919/JCC.2020.09.009.

[13] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, Aug. 2019, doi: 10.1109/MCOM.2019.1900271.

[14] Z. Yang, M. Chen, K. K. Wong, H. V. Poor, and S. Cui, "Federated Learning for 6G: Applications, Challenges, and Opportunities," *Engineering*, vol. 8, pp. 33–41, Jan. 2022, doi: 10.1016/J.ENG.2021.12.002.

[15] P. Gokhale, O. Bhat, and S. Bhat, "Introduction to IOT," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 5, no. 1, pp. 41–44, 2018.

[16] A. A. Laghari, K. Wu, R. A. Laghari, M. Ali, and A. A. Khan, "RETRACTED ARTICLE: A Review and State of Art of Internet of Things (IoT)," *Archives of Computational Methods in Engineering*, vol. 29, no. 3, pp. 1395–1413, 2022, doi: 10.1007/s11831-021-09622-6.

[17] M. Kusek, "Internet of Things: Today and tomorrow," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, pp. 335–338. doi: 10.23919/MIPRO.2018.8400064.

[18] H. D. Kotha and V. M. Gupta, "IoT application: a survey," *Int. J. Eng. Technol*, vol. 7, no. 2.7, pp. 891–896, 2018.

[19] H. Suo, J. Wan, C. Zou, and J. Liu, "Security in the Internet of Things: A Review," in *2012 International Conference on Computer Science and Electronics Engineering*, 2012, pp. 648–651. doi: 10.1109/ICCSEE.2012.373.

[20] N. Mishra and S. Pandya, "Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review," *IEEE Access*, vol. 9, pp. 59353–59377, May 2021, doi: 10.1109/ACCESS.2021.3073408.

[21] O. H. Abdulganiyu, T. Ait Tchakoucht, and Y. K. Saheed, "A systematic literature review for network intrusion detection system (IDS)," *Int J Inf Secur*, vol. 22, no. 5, pp. 1125–1162, 2023, doi: 10.1007/s10207-023-00682-2.

[22] Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, "An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 42210–42219, 2019, doi: 10.1109/ACCESS.2019.2904620.

[23] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Comput Secur*, vol. 28, no. 1–2, pp. 18–28, Feb. 2009, doi: 10.1016/J.COSE.2008.08.003.

[24] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey acm computing surveys vol. 41 (3) article 15," 2009.

[25] L. Ashiku and C. Dagli, "Network Intrusion Detection System using Deep Learning," *Procedia Comput Sci*, vol. 185, pp. 239–247, Jan. 2021, doi: 10.1016/J.PROCS.2021.05.025.

[26] P. A. A. Resende and A. C. Drummond, "A survey of random forest based methods for intrusion detection systems," *ACM Comput Surv*, vol. 51, no. 3, Jul. 2018, doi: 10.1145/3178582;PAGE:STRING:ARTICLE/CHAPTER.

[27] M. Mohammadi *et al.*, "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems," *Journal of Network and Computer Applications*, vol. 178, p. 102983, Mar. 2021, doi: 10.1016/J.JNCA.2021.102983.

[28] G. Drainakis, K. V. Katsaros, P. Pantazopoulos, V. Sourlas, and A. Amditis, "Federated vs. Centralized Machine Learning under Privacy-elastic Users: A Comparative Analysis," *2020 IEEE 19th International Symposium on Network Computing and Applications, NCA 2020*, Nov. 2020, doi: 10.1109/NCA51143.2020.9306745.

[29] S. Otoum, I. Al Ridhawi, and H. T. Mouftah, "Blockchain-Supported Federated Learning for Trustworthy Vehicular Networks," *Proceedings - IEEE Global*

Communications Conference, GLOBECOM, 2020, doi: 10.1109/GLOBECOM42002.2020.9322159.

[30] L. Li, Y. Fan, M. Tse, and K. Y. Lin, "A review of applications in federated learning," *Comput Ind Eng*, vol. 149, p. 106854, Nov. 2020, doi: 10.1016/J.CIE.2020.106854.

[31] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, A. Singh and J. Zhu, Eds., in Proceedings of Machine Learning Research, vol. 54. PMLR, May 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[32] S. Ramaswamy, R. Mathews, K. Rao, and F. Beaufays, "Federated Learning for Emoji Prediction in a Mobile Keyboard," *CoRR*, vol. abs/1906.04329, 2019, [Online]. Available: http://arxiv.org/abs/1906.04329

[33] C. Jobanputra, J. Bavishi, and N. Doshi, "Human Activity Recognition: A Survey," *Procedia Comput Sci*, vol. 155, pp. 698–703, Jan. 2019, doi: 10.1016/J.PROCS.2019.08.100.

[34] J. Feng, C. Rong, F. Sun, D. Guo, and Y. Li, "PMF: A privacy-preserving human mobility prediction framework via federated learning," *Proc ACM Interact Mob Wearable Ubiquitous Technol*, vol. 4, no. 1, Mar. 2020, doi: 10.1145/3381006,.

[35] P. H. Do, T. D. Le, V. Vishnevsky, A. Berezkin, and R. Kirichek, "A Horizontal Federated Learning Approach to IoT Malware Traffic Detection: An Empirical Evaluation with N-BaIoT Dataset," in *2024 26th International Conference on Advanced Communications Technology (ICACT)*, 2024, pp. 1494–1506. doi: 10.23919/ICACT60172.2024.10471929.

[36] A. Blika *et al.*, "Federated Learning For Enhanced Cybersecurity And Trustworthiness In 5G and 6G Networks: A Comprehensive Survey," *IEEE Open Journal of the Communications Society*, 2024, doi: 10.1109/OJCOMS.2024.3449563.

[37] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans Intell Syst Technol*, vol. 10, no. 2, Jan. 2019, doi: 10.1145/3298981;JOURNAL:JOURNAL:TIST;WGROUP:STRING:ACM.

[38] G. Singh, K. Sood, P. Rajalakshmi, D. D. N. Nguyen, and Y. Xiang, "Evaluating Federated Learning-Based Intrusion Detection Scheme for Next Generation

Networks," *IEEE Transactions on Network and Service Management*, vol. 21, no. 4, pp. 4816–4829, 2024, doi: 10.1109/TNSM.2024.3385385.

[39] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, Mar. 2020.

[40] S. J. Reddi *et al.*, "Adaptive Federated Optimization," *ICLR 2021 - 9th International Conference on Learning Representations*, Feb. 2020, Accessed: May 23, 2025. [Online]. Available: https://arxiv.org/pdf/2003.00295

[41] P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, Jun. 2021, doi: 10.1561/2200000083.

[42] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Process Mag*, vol. 37, no. 3, pp. 50–60, May 2020, doi: 10.1109/MSP.2020.2975749.

[43] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDOS) flooding attacks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 4, pp. 2046–2069, 2013, doi: 10.1109/SURV.2013.031413.00127.

[44] M. Antonakakis *et al.*, *Understanding the Mirai Botnet*. 2017. Accessed: May 27, 2025. [Online]. Available: http://ipads.se.sjtu.edu.cn/lib/exe/fetch.php?media=publications:vtz.pdf%0Ahttps://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/hua

[45] M. Feily, A. Shahrestani, and S. Ramadass, "A survey of botnet and botnet detection," *Proceedings - 2009 3rd International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2009*, pp. 268–273, 2009, doi: 10.1109/SECURWARE.2009.48.

[46] M. Antonakakis *et al.*, *Understanding the Mirai Botnet*. 2017. Accessed: Jun. 13, 2025. [Online]. Available: http://ipads.se.sjtu.edu.cn/lib/exe/fetch.php?media=publications:vtz.pdf%0Ahttps://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/hua

[47] D. Tymoshchuk, O. Yasniy, M. Mytnyk, N. Zagorodna, and V. Tymoshchuk, "Detection and classification of DDoS flooding attacks by machine learning method," Dec. 2024, Accessed: May 27, 2025. [Online]. Available: https://arxiv.org/pdf/2412.18990

[48] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of network-based defense mechanisms countering the DoS and DDoS problems," *ACM Comput Surv*, vol. 39, no. 1, p. 42, Apr. 2007, doi: 10.1145/1216370.1216373;PAGE:STRING:ARTICLE/CHAPTER.

[49] "BASHLITE Botnet - NHS England Digital." Accessed: Jun. 13, 2025. [Online]. Available: https://digital.nhs.uk/cyber-alerts/2018/cc-2557

[50] M. Freunek and A. Rombos, "Classification of cyber attacks on IoT and ubiquitous computing devices".

[51] S. Samarakoon *et al.*, "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network," 2022. [Online]. Available: https://arxiv.org/abs/2212.01298

[52] V. Rey, P. M. Sánchez Sánchez, A. Huertas Celdrán, and G. Bovet, "Federated learning for malware detection in IoT devices," *Computer Networks*, vol. 204, p. 108693, Feb. 2022, doi: 10.1016/J.COMNET.2021.108693.

[53] F. L. de Caldas Filho *et al.*, "Botnet Detection and Mitigation Model for IoT Networks Using Federated Learning," *Sensors*, vol. 23, no. 14, 2023, doi: 10.3390/s23146305.

[54] D. Torre, A. Chennamaneni, J. Y. Jo, G. Vyas, and B. Sabrsula, "Toward Enhancing Privacy Preservation of a Federated Learning CNN Intrusion Detection System in IoT: Method and Empirical Study," *ACM Transactions on Software Engineering and Methodology*, vol. 34, no. 2, p. 53, Feb. 2025, doi: 10.1145/3695998/ASSET/136FB913-BDF2-4F76-BCAC-7A5AB352571A/ASSETS/GRAPHIC/TOSEM-2024-0079-F04.JPG.

[55] F. Naeem, A. W. Malik, S. Abbas Khan, and F. Jabeen, "Enhancing Intrusion detection: Leveraging Federated Learning and Hybrid Machine Learning Algorithms On ToN_IoT Dataset," in *2023 International Conference on Frontiers of Information Technology (FIT)*, 2023, pp. 73–78. doi: 10.1109/FIT60620.2023.00023.

[56] R. Lazzarini, H. Tianfield, and V. Charissis, "Federated Learning for IoT Intrusion Detection," *AI 2023, Vol. 4, Pages 509-530*, vol. 4, no. 3, pp. 509–530, Jul. 2023, doi: 10.3390/AI4030028.

[57] S. Abbas *et al.*, "A Novel Federated Edge Learning Approach for Detecting Cyberattacks in IoT Infrastructures," *IEEE Access*, vol. 11, pp. 112189–112198, 2023, doi: 10.1109/ACCESS.2023.3318866.

[58] N. Hamdi, "Federated learning-based intrusion detection system for Internet of Things," *Int J Inf Secur*, vol. 22, no. 6, pp. 1937–1948, 2023, doi: 10.1007/s10207-023-00727-6.

[59] P. H. Mirzaee, M. Shojafar, Z. Pooranian, P. Asefy, H. Cruickshank, and R. Tafazolli, "FIDS: A Federated Intrusion Detection System for 5G Smart Metering Network," in *2021 17th International Conference on Mobility, Sensing and Networking (MSN)*, Dec. 2021, pp. 215–222. doi: 10.1109/MSN53354.2021.00044.

[60] M. J. Idrissi *et al.*, "Fed-ANIDS: Federated learning for anomaly-based network intrusion detection systems," *Expert Syst Appl*, vol. 234, p. 121000, Dec. 2023, doi: 10.1016/J.ESWA.2023.121000.

[61] L. Almeida, P. Rodrigues, R. Teixeira, M. Antunes, and R. L. Aguiar, "Privacy-Preserving Defense: Intrusion Detection in IoT using Federated Learning," in *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*, 2024, pp. 908–913. doi: 10.1109/MELECON56669.2024.10608461.

[62] R. Kale, K. W. Fok, and V. L. L. Thing, "Payload-based 5G Attack Detection," in *2023 9th International Conference on Computer and Communications (ICCC)*, 2023, pp. 1262–1266. doi: 10.1109/ICCC59590.2023.10507422.

[63] R. A. Bakar *et al.*, "5GDAD: A Deep Learning Approach for DDoS Attack Detection in 5G P4-based UPF," in *2024 IEEE 25th International Conference on High Performance Switching and Routing (HPSR)*, 2024, pp. 185–190. doi: 10.1109/HPSR62440.2024.10635995.

[64] M. S. Khan, B. Farzaneh, N. Shahriar, N. Saha, and R. Boutaba, "SliceSecure: Impact and Detection of DoS/DDoS Attacks on 5G Network Slices," in *2022 IEEE Future Networks World Forum (FNWF)*, 2022, pp. 639–642. doi: 10.1109/FNWF55208.2022.00117.

[65] A. Abouelkhair, K. Majdi, N. Limam, M. A. Salahuddin, and R. Boutaba, "5GProvGen: 5G Provenance Dataset Generation Framework," *Proceedings of the 2024 20th International Conference on Network and Service Management: AI-Powered Network and Service Management for Tomorrow's Digital World, CNSM 2024*, 2024, doi: 10.23919/CNSM62983.2024.10814296.

[66] I. Karim, K. S. Mubasshir, M. M. Rahman, and E. Bertino, “SPEC5G: A Dataset for 5G Cellular Network Protocol Analysis,” 2023. [Online]. Available: https://arxiv.org/abs/2301.09201

[67] G. Amponis *et al.*, “Threatening the 5G core via PFCP DoS attacks: the case of blocking UAV communications,” *EURASIP J Wirel Commun Netw*, vol. 2022, no. 1, p. 124, 2022, doi: 10.1186/s13638-022-02204-5.

[68] Y. H. Choi *et al.*, “ML-Based 5G Traffic Generation for Practical Simulations Using Open Datasets,” *IEEE Communications Magazine*, vol. 61, no. 9, pp. 130–136, Sep. 2023, doi: 10.1109/MCOM.001.2200679.

[69] Y. Meidan *et al.*, “N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders,” *IEEE Pervasive Comput*, vol. 17, no. 3, pp. 12–22, 2018, doi: 10.1109/MPRV.2018.03367731.

[70] M. Rawat, A. Singh Bedi, B. Singh, S. Gupta, G. Singal, and P. Kaur, “Ensemble-Based Botnet Attack Detection and Classification Using Machine Learning Algorithms on NBaIoT Dataset,” in *2024 IEEE Region 10 Symposium (TENSYMP)*, 2024, pp. 1–6. doi: 10.1109/TENSYMP61132.2024.10752221.

[71] Z. L. Thakker and S. H. Buch, “Effect of Feature Scaling Pre-processing Techniques on Machine Learning Algorithms to Predict Particulate Matter Concentration for Gandhinagar, Gujarat, India,” *Int. J. Sci. Res. Sci. Technol*, vol. 11, pp. 410–419, 2024.

[72] S. Pelekis *et al.*, “Adversarial machine learning: a review of methods, tools, and critical industry sectors,” *Artificial Intelligence Review 2025 58:8*, vol. 58, no. 8, pp. 1–87, May 2025, doi: 10.1007/S10462-025-11147-4.

[73] “Pearson’s Correlation Coefficient,” *Encyclopedia of Public Health*, pp. 1090–1091, 2008, doi: 10.1007/978-1-4020-5614-7_2569.

[74] A. Zainudin, M. A. P. Putra, R. N. Alief, R. Akter, D.-S. Kim, and J.-M. Lee, “Blockchain-Inspired Collaborative Cyber-Attacks Detection for Securing Metaverse,” *IEEE Internet Things J*, vol. 11, no. 10, pp. 18221–18236, May 2024, doi: 10.1109/JIOT.2024.3364247.