ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΕΥΦΥΕΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ –WEB INTELLIGENCE

Εφαρμογή Μεθόδων Μηχανικής Μάθησης σε Δίκτυα 5G

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Νικολάου Μαρτίκα
Α.Μ.: 17/2018

Επιβλέπων : Κωνσταντίνος Διαμαντάρας
Καθηγητής – ΔΙ.ΠΑ.Ε.

Θεσσαλονίκη, Ιούλιος 2021

Η σελίδα αυτή είναι σκόπιμα λευκή.

ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΥΦΥΕΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ – WEB INTELLIGENCE

Εφαρμογή Μεθόδων Μηχανικής Μάθησης σε Δίκτυα 5G

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Νικολάου Μαρτίκα
Α.Μ.: 17/2018

Επιβλέπων : Κωνσταντίνος Διαμαντάρας
Καθηγητής, ΔΙ.ΠΑ.Ε.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την <Ημερομηνία> <έτος>.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Όνομα Επώνυμο	Όνομα Επώνυμο	Όνομα Επώνυμο
Καθηγητής	Καθηγητής	Καθηγητής

Θεσσαλονίκη, Ιούλιος 2021

(Υπογραφή)

.....

ΝΙΚΟΛΑΟΣ ΜΑΡΤΙΚΑΣ – Α.Μ.: 17/2018

Εφαρμογή Μεθόδων Μηχανικής Μάθησης σε Δίκτυα 5G

Περίληψη

Σκοπός αυτής της μεταπτυχιακής διπλωματικής εργασίας είναι η μελέτη των τεχνολογιών και της αρχιτεκτονικής των δικτύων 5G. Ώστε να προσδιοριστούν τα ζητήματα ελέγχου ταυτότητας και ασφάλειας των δικτύων που προκύπτουν και δεν μπορούν να επιλυθούν ή να προσφέρουν επαρκή ασφάλεια με τις γνωστές τεχνικές ασφαλείας των προηγούμενων δικτύων. Καθώς επίσης και η προσέγγιση προοπτικών ελέγχου ταυτότητας και ασφάλειας, οι οποίες βασίζονται σε τεχνικές της μηχανικής μάθησης. Μεθοδολογικά η εργασία αναπτύχθηκε μέσω της ανασκόπησης σε άρθρα και δημοσιεύσεις της ελληνικής και ξενόγλωσσης βιβλιογραφίας. Για την εξαγωγή ασφαλέστερων συμπερασμάτων πραγματοποιήθηκε σύγκριση των ευρημάτων από προηγούμενες μελέτες. Από την ολοκλήρωση αυτής της μελέτης συμπεραίνεται ότι τα δίκτυα 5G θα προσφέρουν σημαντικά οφέλη με τις υψηλές ταχύτητες και την χαμηλή καθυστέρηση του χρόνου. Ωστόσο, επειδή τα νέα δίκτυα 5G βασίζονται και περιλαμβάνουν ένα σύνολο νέων τεχνολογιών υλικού, αλλά και λογισμικού υπάρχουν πολλά ζητήματα ασφαλείας και ελέγχου ταυτότητας, τα οποία δεν μπορούν να αντιμετωπιστούν αποτελεσματικά με τις παραδοσιακές τεχνικές, αλλά απαιτούν τεχνικές που βασίζονται στη μηχανική μάθηση, η χρήση της οποία είναι απαραίτητη για την ασφάλεια και τον έλεγχο ταυτότητας στα δίκτυα 5G.

Λέξεις Κλειδιά: Δίκτυα 5ης γενιάς, ασφάλεια δικτύων, έλεγχος ταυτότητας, μηχανική μάθηση.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Abstract

The purpose of this master's thesis is to study the technologies and architecture of 5G networks. To identify emerging network security and security issues that may not be resolved or provide adequate security with known network security techniques. As well as the approach of identity and security control perspectives, which are based on machine learning techniques. The work was methodologically developed through the review of articles and publications literature. In order to extract safe conclusions, the findings from previous studies were compared. From the completion of this study it is concluded that 5G networks will offer significant benefits with high speeds and low time latency. However, because the new 5G networks are based on and include a set of new hardware and software technologies, there are many security and authentication issues that cannot be tackled effectively with classic techniques, but require techniques based on machine learning, the use of which is necessary for security and authentication on 5G networks.

Keywords: Machine Learning, 5G networks, network security, authentication.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Ευχαριστίες

Η ολοκλήρωση αυτής της πτυχιακής εργασίας υλοποιήθηκε με την υποστήριξη κάποιων ανθρώπων στους οποίους θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου.

Πρώτα από όλους, θα ήθελα να απευθύνω ιδιαίτερες ευχαριστίες στον επιβλέποντα καθηγητή μου, κύριο Κωνσταντίνο Διαμαντάρα, ο οποίος μου εμπιστεύτηκε ένα από τα θέματα μεταπτυχιακών εργασιών, αλλά και για το αμείωτο ενδιαφέρον καθ' όλη τη διάρκεια υλοποίησης της παρούσας μεταπτυχιακής εργασίας.

Συνεχίζοντας θα ευχαριστήσω τους γονείς μου, οι οποίοι με συνεχή συμπαράσταση, αγάπη και κατανόηση στήριξαν τις προσπάθειες μου καθ' όλη τη διάρκεια των σπουδών μου.

Πίνακας περιεχομένων

1 Εισαγωγή 1

1.1 Δίκτυα κινητής τηλεφωνίας και μηχανική μάθηση 1

1.2 Αντικείμενο διπλωματικής 2

1.2.1 Συνεισφορά 3

1.3 Οργάνωση κειμένου 4

2 Σχετικές εργασίες 5

2.1 Δίκτυα 5G και ασφάλεια δικτύων 5G 5

2.2 Ασφάλεια στα δίκτυα 5G με τη βοήθεια της μηχανικής μάθησης 6

3 Θεωρητικό υπόβαθρο 8

3.1 Μηχανική μάθηση (Machine Learning) 8

3.1.1 Ορισμός και χαρακτηριστικά μηχανικής μάθησης 8

3.1.2 Τύποι μηχανικής μάθησης 9

3.1.2.1 Μάθηση με επίβλεψη (supervised learning) 10

3.1.2.2 Μάθηση χωρίς επίβλεψη (Unsupervised learning) 11

3.1.2.3 Μάθηση με ημιεπίβλεψη (Semisupervised learning) 11

3.1.2.4 Μάθηση με ενίσχυση (Reinforcement learning) 11

3.1.2.5 Βαθιά μάθηση (Deep Learning) 12

3.1.2.6 Βαθιά μάθηση με ενίσχυση (Deep Reinforcement Learning) 13

3.1.2.7 Μεταφορά μάθησης (Transfer learning) 14

3.2 Δίκτυα κινητής τηλεφωνίας 5G 14

3.2.1 Βασικά χαρακτηριστικά 14

3.2.2 Βασικές τεχνολογίες 16

3.2.2.1 Massive MIMO (Multiple Input Multiple Output) 16

3.2.2.2 Υπολογιστικά Νέφη Κινητής (Mobile Cloud Computing) 17

3.2.2.3 Network Function Virtualization (NFV) 18

3.2.2.4 Software Defined Networking (SDN) 18
3.2.2.5 Millimeter Wave (mmWave) 19
3.2.2.6 Πολύ πυκνό και ετερογενές δίκτυο 20
3.2.2.7 Smallcells 21
3.2.2.8 Διαδίκτυο των Πραγμάτων (Internet of Things – IoT) 23
3.2.2.9 Επικοινωνία συσκευή με συσκευή με συσκευή (Device to Device - D2D) 23
3.2.2.10 Τεμαχισμός δικτύου (network slicing) 23
3.2.2.11 Multi-Access Edge Computing (MEC) 24
3.2.3 Αρχιτεκτονική 24
4 Ασφάλεια στα δίκτυα 5G 30
4.1 Τύποι επιθέσεων ασφαλείας 31
4.1.1 Παθητικές επιθέσεις 31
4.1.1.1 Eavesdropping (υποκλοπές) 31
4.1.1.1.1 Απελευθέρωση μηνύματος (Release of message) 31
4.1.1.1.2 Ανάλυση κυκλοφορίας (Traffic analysis) 32
4.1.1.1.3 Sniffing 32
4.1.1.1.4 Keyloggers (καταγραφείς πληκτρολόγησης) 32
4.1.2 Ενεργές επιθέσεις 32
4.1.2.1 Διακοπή (interruption) 33
4.1.2.1.1 Άρνηση υπηρεσίας (Denial of Service - DoS) 33
4.1.2.1.2 Κατανεμημένη Άρνηση Υπηρεσιών (Distributed Denial of Services - DDoS) 33
4.1.2.1.3 ΕγχύσειςSQL (SQL Injections) 34
4.1.2.2 Κατασκευή (Fabrication) 34
4.1.2.2.1 Επίθεση επανάληψης (replayattack) 34
4.1.2.2.2 Μεταμφίεση (masquerading) 35
4.2 Προκλήσεις ασφαλείας στις νέες τεχνολογίες 36
4.2.1 Προκλήσεις ασφαλείας στα Νέφη Κινητής (Mobile Clouds) 36

4.2.2 Προκλήσεις ασφαλείας σε SDN 37
4.2.3 Προκλήσεις ασφαλείας στα κανάλια επικοινωνίας 38
4.2.4 Προκλήσεις ασφαλείας σε NFV 39
4.2.5 Προκλήσεις σε παρόχους επικοινωνίας 40
4.2.6 Προκλήσεις ασφαλείας σε massive MIMO 41
4.3 Προκλήσεις ασφαλείας στο φυσικό επίπεδο 41
4.3.1 Προκλήσεις ελέγχου ταυτότητας (authentication) στο φυσικό επίπεδο 42
4.3.2 Προκλήσεις απόρρητου (privacy) στο φυσικό επίπεδο 45
4.4 Λύσεις ασφαλείας στις νέες τεχνολογίες 45
4.4.1 Λύσεις ασφαλείας σε massive MIMO 45
4.4.2 Λύσεις ασφαλείας σε SDN 47
4.4.3 Λύσεις ασφαλείας σε NFV 48
4.4.4 Λύσεις ασφαλείας σε νέφη κινητής 50
4.5 Έλεγχος ταυτότητας εξοπλισμού χρήστη και δικτύου 50
4.6 Μέθοδοι ελέγχου ταυτότητας (authentication) 51
4.6.1 Έλεγχος ταυτότητας με κωδικό πρόσβασης (password authentication) 51
4.6.2 Βιομετρικός έλεγχος ταυτότητας (biometrical authentication) 52
4.6.3 Έλεγχος ταυτότητας δύο παραγόντων (two-factor authentication - 2FA) 53
4.6.4 Έλεγχος ταυτότητας με token (τεκμήριο) 53
4.6.5 Έλεγχος ταυτότητας μοναδικής σύνδεσης (single sign-on (SSO)) 53
5 Μηχανική μάθηση και ασφάλεια δικτύων 5G 55
5.1 Μηχανική μάθηση και ασφάλεια στις νέες τεχνολογίες 56
5.1.1 Μηχανική μάθηση και ασφάλεια σε massive MIMO 57
5.1.2 Μηχανική μάθηση και ασφάλεια σε SDN 58
5.1.3 Μηχανική μάθηση και ασφάλεια σε NFV 59
5.2 Μηχανική μάθηση και έλεγχος ταυτότητας (authentication) 62
5.2.1 Βιομετρικός έλεγχος ταυτότητας με μηχανική μάθηση 63

5.2.2 Παραδείγματα μηχανικής μάθησης και ελέγχου ταυτότητας 65

6 Μελέτες περιπτώσεων 71

6.1 Μηχανική μάθηση για ανίχνευση ανωμαλιών σε δίκτυο 5G 71

6.1.1 Αρχιτεκτονική SDS για δίκτυα 5G 71

6.1.2 Ανίχνευση ανωμαλιών από Συνελικτικό Νευρωνικό Δίκτυο (CNN) 74

6.1.3 Υλοποίηση AutoML & NAS 76

6.1.3.1 Γενικές περιγραφές υποδείγματος με έναν ελεγκτή RNN 78

6.1.3.2 Εκπαίδευση με ενίσχυση 79

6.1.3.3 Συνδέσεις παράλειψης 81

6.1.3.4 Γενική αρχιτεκτονική επαναλαμβανόμενων κελιών 82

6.1.4 Σύνολα δεδομένων για την ανίχνευση ανωμαλιών 84

6.1.5 Προφίλ και δυνατότητες δικτύου 87

6.1.6 Χρόνος διαδρομής άφιξης και επιλογή δυνατοτήτων 89

6.1.7 Προ-επεξεργασία συνόλου δεδομένων 90

6.1.8 Αποτελέσματα ανίχνευσης ανωμαλίας 92

6.1.9 Συμπεράσματα περίπτωσης 97

6.2 Συνεχής έλεγχος ταυτότητας με δυναμική πληκτρολόγηση σε διαδικτυακές εφαρμογές 98

6.2.1 Μεθοδολογία 98

6.2.1.1 1ο βήμα: χρήση δυναμικής πληκτρολόγησης χρηστών 98

6.2.1.2 2ο βήμα: σχεδίαση υπηρεσίας συνεχούς ελέγχου ταυτότητας 99

6.2.1.3 3ο βήμα: εξαγωγή δεδομένων και προεπεξεργασία 108

6.2.1.4 4ο βήμα: εκπαίδευση και πρόβλεψη 109

6.2.1.5 5ο βήμα: συλλογή δεδομένων 113

6.2.2 Συλλογή δεδομένων και σύνολά δεδομένων 114

6.2.3 Αποτελέσματα 114

6.2.4 Συμπεράσματα 115

6.3 Έλεγχος ταυτότητας με δεδομένα ηλεκτροκαρδιογραφήματος 115

6.3.1 Μεθοδολογία 115

6.3.1.1 Μηχανική μάθηση και εξαγωγή χαρακτηριστικών χωρίς σταθερή βάση σύγκρισης 116

6.3.1.2 Βαθιά μάθηση και εξαγωγή χαρακτηριστικών χωρίς σταθερή βάση σύγκρισης 118

6.3.2 Συλλογή δεδομένων - προεπεξεργασία 118

6.3.3 Εξαγωγή χαρακτηριστικών - Μετασχηματισμοί 119

6.3.3.1 Διακριτός μετασχηματισμός Fourier 119

6.3.3.2 Διακριτός μετασχηματισμός συνημίτονου (cosine) 120

6.3.3.3 Αποσύνθεση – μετασχηματισμός Wavelet 120

6.3.4 Ταξινόμηση 120

6.3.4.1 Μηχανική μάθηση 120

6.3.4.2 Βαθιά μάθηση 121

6.3.5 Συμπεράσματα 121

7 Σύγκριση λύσεων ασφαλείας 123

7.1 Σύγκριση τυπικών μεθόδων και μεθόδων μηχανικής μάθησης για τις νέες τεχνολογίες 123

7.2 Σύγκριση παραδοσιακού και έξυπνου ελέγχου ταυτότητας 125

8 Επίλογος 129

8.1 Σύνοψη και συμπεράσματα 129

8.2 Μελλοντικές επεκτάσεις 131

9 Βιβλιογραφία 132

Παράρτημα: Μοντέλο CNN με αρχιτεκτονική NAS 144

1 Εισαγωγή

1.1 Δίκτυα κινητής τηλεφωνίας και μηχανική μάθηση

Η ραγδαία αύξηση των χρηστών και των ποικίλων δωρεάν εφαρμογών και υπηρεσιών είχε ως αποτέλεσμα να γίνει απαραίτητη η πρόσβαση στο διαδίκτυο, για πολλούς ανθρώπους, επιχειρήσεις και βιομηχανίες [1]. Τα δίκτυα κινητής επικοινωνίας συνδέουν μεγάλο μέρος του παγκόσμιου πληθυσμού και τα τελευταία χρόνια οι ασύρματες επικοινωνίες χρησιμοποιούνται όλο και περισσότερο από τους χρήστες, καθώς η χρήση των κινητών συσκευών και των έξυπνων κινητών συσκευών παρέχουν μια ευρεία ποικιλία υπηρεσιών [2]. Το διαδίκτυο έγινε κάτι απαραίτητο για την καθημερινότητα των ανθρώπων. Αυτό είχε ως αποτέλεσμα, οι χρήστες γίνονται όλο και πιο απαιτητικοί όσον αφορά την κάλυψη των αναγκών τους στο διαδίκτυο [1].

Τις τελευταίες τέσσερις δεκαετίες, οι τεχνολογίες δικτύων και τηλεπικοινωνιών έχουν εξελιχθεί από την 1η γενιά (1G) το 1979 στην 4η γενιά (4G) το 2009 και σήμερα στην 5η γένια (5G). Από την τεχνολογία δικτύου τηλεπικοινωνιών 1G έως την τεχνολογία 4G, σημειώθηκε σημαντική ανακάλυψη σε λειτουργίες και υπηρεσίες. Επιπλέον, η πρόοδος από τη μία γενιά δικτύων κινητής τηλεφωνίας στην άλλη συνοδεύτηκε από εξελίξεις στην τεχνολογία του υλικού [3]. Οι υπηρεσίες 4G, δεν μπορούσαν να ικανοποιήσουν τις αναμενόμενες απαιτήσεις των νέων επιστημονικών τεχνολογιών, όπως τα μη επανδρωμένα αεροσκάφη (Unmanned Aerial Vehicles - UAV), την εικονική πραγματικότητα, τα αυτόνομα οχήματα, τη ροή βίντεο υψηλής ανάλυσης, το gaming cloud (δηλαδή τα ηλεκτρονικά παιχνίδια σε νέφη) και πολλά άλλα [4]. Επομένως, αναπτύχθηκαν τα δίκτυα 5G με σκοπό την κάλυψη αυτών των αναγκών παρέχοντας πιο γρήγορες ταχύτητες και χαμηλότερους χρόνους καθυστέρησης.

Τα δίκτυα 5G ξεκίνησαν να συνδέουν υποδομές, οι οποίες απαιτούν μεγάλη ασφάλεια για την διασφάλιση της ίδιας της υποδομής, αλλά και για την κοινωνία γενικότερα. Για παράδειγμα,

μια παραβίαση ασφάλειας στα ηλεκτρονικά συστήματα τροφοδοσίας ηλεκτρικής ενέργειας, μπορεί να είναι καταστροφική για όλα τα ηλεκτρικά και ηλεκτρονικά συστήματα από τα οποία εξαρτάται η κοινωνία [5]. Επίσης, τα δίκτυα 5η γενιάς είναι απαραίτητα για την υποστήριξη διαφορετικών συσκευών, οι οποίες συνδέονται με ετερογενείς συσκευές και μηχανήματα (όπως, για παράδειγμα στο Διαδίκτυο των Πραγμάτων). Ωστόσο, αυτή η ετερογένεια συσκευών και μηχανημάτων αυξάνει άμεσα την ευπάθεια για διάφορες κακόβουλες επιθέσεις [6].

Η διαρκώς εξελισσόμενη αρχιτεκτονική δικτύου επικοινωνίας για την ενσωμάτωση διαφορετικού εύρους συσκευών με μοναδικές απαιτήσεις για διαφορετικές παραμέτρους δικτύου έχει οδηγήσει σε εξελιγμένες προκλήσεις για την ασφάλεια του δικτύου. Οι πρόσφατες εξελίξεις στα δίκτυα 5G διευκολύνουν την ανάπτυξη της επικοινωνίας παρέχοντας υψηλότερους ρυθμούς μετάδοσης και ταχύτητες δεδομένων. Αυτή η τεράστια αύξηση της μεταφοράς δεδομένων και των συνδεδεμένων συσκευών σημαίνει περισσότερες ευπάθειες, απειλές και επιθέσεις που έχουν ως αποτέλεσμα καταστροφικές ζημιές οικονομικά, κοινωνικά και για την ανθρωπότητα. Επομένως, ο έλεγχος και η ανάλυση τέτοιων μεγάλων δεδομένων για ύποπτες δραστηριότητες δεν μπορεί να επιτευχθεί μόνο με παραδοσιακές μεθόδους ασφαλείας.

Εν συντομία, ένα δίκτυο 4G είναι πολύ πιο περίπλοκο από ένα δίκτυο 2G λόγω του αυξημένου αριθμού σταθμών βάσης και χρηστών, αλλά και λόγω των βελτιώσεων στις τεχνολογίες ραδιοφωνικών σημάτων και δικτύων. Τα συστήματα κινητής τηλεφωνίας 5G είναι ακόμη πιο περίπλοκα καθώς χειρίζονται ένα ευρύτερο σύνολο σεναρίων που δεν αντιμετωπίζονται πλήρως από τα υπάρχοντα κυψελοειδή δίκτυα και επίσης χειρίζονται τεράστιους όγκους δεδομένων που παράγονται από το ίδιο το δίκτυο. Αυτή η πολυπλοκότητα αντιμετωπίζεται με την ανάπτυξη έξυπνων μεθόδων για την ανάλυση δεδομένων από τα δίκτυα 5G. Αυτές οι έξυπνες μέθοδοι, βασίζονται στην μηχανική μάθηση, μπορούν να συμβάλουν στη διαχείριση του δικτύου και να προβλέψουν τη μελλοντική συμπεριφορά του δικτύου και των χρηστών προκειμένου να λάβουν πιο έξυπνες αποφάσεις. Επιπλέον προσφέρουν νέες δυναμικές λύσεις στους τομείς της ασφάλειας, της ιδιωτικής ζωής και της ανίχνευσης απειλών σε συστήματα 5G.

1.2 Αντικείμενο διπλωματικής

Αντικείμενο της παρούσας εργασίας αποτελεί η εφαρμογή της μηχανικής μάθησης τα δίκτυα 5G. Οι πρόσφατες εξελίξεις στην τεχνολογία και την αρχιτεκτονική των δικτύων 5G έχουν αποδείξει την αξία τους, καθώς, η κατασκευή και η εφαρμογή τους έχει ξεκινήσει σε όλο τον κόσμο [7]. Τα δίκτυα κινητής τηλεφωνίας 5G είναι πολύ σημαντικά για την υποστήριξη

διαφορετικών εφαρμογών συνδέοντας ετερογενείς συσκευές και μηχανήματα [6]. Ο βασικός παράγοντας αύξησης της απόδοσης των δικτύων αυτών από την πρόσβαση στο κεντρικό δίκτυο είναι η λογισμοποίηση (softwareization), η νεφοποίηση (cloudification) και η εικονοποιήση (virtualization) του κλειδιού που επιτρέπει τις λειτουργίες του δικτύου [7].

Η ταχεία εξέλιξη φέρει κινδύνους, απειλές με τρωτά σημεία του συστήματος, γεγονός που αυξάνει άμεσα την ευπάθεια για διάφορες επιθέσεις πλαστογράφησης (spoofing attacks) [6][7]. Οι συμβατικές τεχνικές κρυπτογράφησης και ελέγχου ταυτότητας στο φυσικό επίπεδο αντιμετωπίζουν ορισμένες προκλήσεις σε ένα σύνθετο δυναμικό δίκτυο κινητής, όπως το 5G, με δυσκολίες στον προ-σχεδιασμό ενός ακριβούς μοντέλου ελέγχου ταυτότητας (authentication model), στην παροχή συνεχούς προστασίας και στην εκμάθηση διαφόρων χαρακτηριστικών [7]. Επομένως, είναι σημαντική η διασφάλιση της ασφάλειας από άκρο σε άκρο (end-to-end - E2E). Η μηχανική μάθηση (Machine Learning - ML) μπορεί να διαδραματίσει ζωτικό ρόλο στο σχεδιασμό, τη μοντελοποίηση και την αυτοματοποίηση αποτελεσματικών πρωτοκόλλων ασφαλείας απέναντι σε ένα ποικίλο και ευρύ φάσμα απειλών [6].

Με βάση τα παραπάνω, σκοπός αυτής της εργασίας είναι η μελέτη της τεχνολογίας και της αρχιτεκτονικής των δικτύων 5G, ο προσδιορισμός των ζητημάτων ελέγχου ταυτότητας (authentication) και της ασφάλειας των δικτύων που προκύπτουν και δεν μπορούν να επιλυθούν ή να προσφέρουν επαρκή ασφάλεια με τις γνωστές τεχνικές ασφαλείας των προηγούμενων δικτύων και η προσέγγιση προοπτικών ελέγχου ταυτότητας και ασφάλειας, οι οποίες βασίζονται σε τεχνικές της μηχανικής μάθησης.

1.2.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Μελετήθηκαν αναλυτικά τα δίκτυα κινητής τηλεφωνίας 5ης γενιάς και παρουσιάστηκαν, οι σχετικές αρχιτεκτονικές και το σύνολο των τεχνολογιών που περιλαμβάνονται σε αυτά.
2. Προσδιορίστηκαν τα ζητήματα ασφαλείας για κάθε μια από τις τεχνολογίες των δικτύων 5G, καθώς και για τον έλεγχο ταυτότητας.
3. Αναζητήθηκαν και παρουσιάστηκαν λύσεις για την αντιμετώπιση των ζητημάτων ασφαλείας που προκύπτουν στις νέες τεχνολογίες των δικτύων 5G με την χρήση της μηχανικής μάθησης.

4. Πραγματοποιήθηκε σύγκριση μεταξύ των παραδοσιακών τεχνικών και των τεχνικών μηχανικής μάθησης για την ασφάλεια των δικτύων και τον έλεγχο ταυτότητας στα δίκτυα 5G.
5. Διαπιστώθηκε η σημασία της μηχανικής μάθησης για την ασφάλεια των δικτύων 5G.

1.3 Οργάνωση κειμένου

Εδώ γίνεται μια σύντομη περιγραφή των κεφαλαίων της διπλωματικής. Στο κεφάλαιο 2 παρουσιάζονται οι εργασίες που σχετίζονται με το αντικείμενο της διπλωματικής. Στο κεφάλαιο 3 περιλαμβάνεται το θεωρητικό υπόβαθρο, όπου γίνεται περιγραφή της μηχανικής μάθησης και των δικτύων κινητής τηλεφωνίας 5G. Στο κεφάλαιο 4 προσεγγίζονται και δίνονται τυπικές λύσεις για τα ζητήματα ασφαλείας και ελέγχου ταυτότητας στα δίκτυα 5G. Στο κεφάλαιο 5 παρουσιάζονται οι λύσεις ασφαλείας και ελέγχου ταυτότητας που βασίζονται σε τεχνικές της μηχανικής μάθησης. Στο κεφάλαιο 6 παρουσιάζεται αναλυτικά μια μελέτη περίπτωσης για την ανίχνευση ανωμαλιών σε δίκτυα 5G με τη βοήθεια της μηχανικής μάθησης. Στο κεφάλαιο 7 γίνεται σύγκριση μεταξύ των τυπικών λύσεων ασφαλείας και ελέγχου ταυτότητας με τις λύσεις που βασίζονται στη μηχανική μάθηση. Το κεφάλαιο 8 αποτελεί τον επίλογο της διπλωματικής, όπου συνοψίζεται η διπλωματική, παρατίθενται τα συμπεράσματα και δίνονται προτάσεις για μελλοντική επέκταση.

2 *Σχετικές εργασίες*

Σε αυτό το κεφάλαιο παρουσιάζονται εργασίες που σχετίζονται με το αντικείμενο της παρούσας διπλωματικής, δηλαδή με τα δίκτυα 5G και τις προκλήσεις ασφαλείας, και με την ασφάλεια των δικτύων 5G με τη βοήθεια της μηχανικής μάθησης.

2.1 Δίκτυα 5G και ασφάλεια δικτύων 5G

Στο [5] παρουσιάζεται μια επισκόπηση των σημαντικότερων προκλήσεων ασφαλείας που εντοπίζονται στις νέες τεχνολογίες των δικτύων 5G, όπως τα υπολογιστικά νέφη (cloud computing), τα δίκτυα που βασίζονται σε λογισμικό (Network Defined Networking – SDN) και η εικονοποίηση λειτουργιών δικτύου (Network Function Virtualization - NFV), καθώς επίσης και στις αυξανόμενες ανησυχίες σχετικά με το απόρρητο των χρηστών στα δίκτυα 5G. Επιπλέον, παρουσιάζονται ορισμένες λύσεις ασφαλείας για αυτές τις προκλήσεις και μελλοντικές οδηγίες για την ασφάλεια των συστημάτων 5G.

Στο [8] διερευνήθηκαν οι πιθανές προκλήσεις ασφαλείας που προκαλούνται από την ενσωμάτωση των εννοιών και των τεχνολογιών της μηχανικής μάθησης στα δίκτυα 5G και προτάθηκαν πιθανές λύσεις και ερευνητικές κατευθύνσεις. Μέσα από μια ευρεία έρευνα σχετικά με απειλές και λύσεις ασφαλείας που προκαλούνται από την μηχανική μάθηση στο πλαίσιο των δικτύων 5G.

Στο [4] παρουσιάζεται η αρχιτεκτονική των δικτύων 5G και την ακολουθούμενη από την ασφάλεια που σχετίζεται με αυτά τα δίκτυα. Επίσης, μελέτησαν το 5G ως ένα ενεργειακά αποδοτικό δίκτυο, τους διάφορους τύπους αποτελεσματικών κεραιών που αναπτύχθηκαν για τα δίκτυα 5G και προηγμένες προδιαγραφές για εφαρμογές IoT μαζί με τις τεχνολογίες επικοινωνίας. Περιγράφεται επίσης η ευρύτερη χρήση των δικτύων 5G και οι μελλοντικές επιπτώσεις της στη ζωή των ανθρώπων.

Στο [9] αναλύθηκε ο έλεγχος ταυτότητας των δικτύων 5G βάσει των πρωτοκόλλων AKA. Οι συγγραφείς (σύμφωνα με τους ίδιους) παρείχαν το πρώτο επίσημο ολοκληρωμένο μοντέλο ενός πρωτοκόλλου από την οικογένεια AKA: 5G AKA. Εξήγαγαν επίσης ακριβείς απαιτήσεις από τα πρότυπα 3GPP που ορίζουν το 5G και εντόπισαν τους στόχους ασφαλείας που λείπουν. Χρησιμοποιώντας το εργαλείο επαλήθευσης πρωτοκόλλου ασφαλείας Tamarin, πραγματοποίησαν μια πλήρη, συστηματική, αξιολόγηση ασφάλειας του μοντέλου σε σχέση με τους στόχους ασφαλείας 5G. Η αυτοματοποιημένη ανάλυσή τους προσδιορίζει τις ελάχιστες παραδοχές ασφαλείας που απαιτούνται για κάθε στόχο ασφαλείας και διαπιστώθηκε ότι δεν επιτυγχάνονται ορισμένοι κρίσιμοι στόχοι ασφαλείας, εκτός από τις υποθέσεις που λείπουν από το πρότυπο.

2.2 Ασφάλεια στα δίκτυα 5G με τη βοήθεια της μηχανικής μάθησης

Στο [6] παρουσιάζονται προσεγγίσεις ελέγχου ταυτότητας (authentication) για εφαρμογές των δικτύων κινητής τηλεφωνίας 5G, οι οποίες βασίζονται σε τεχνικές μηχανικής μάθησης, χρησιμοποιώντας τα χαρακτηριστικά του φυσικού επίπεδου και εισαγωγή πληροφοριών για έλεγχο ταυτότητας και αποτελεσματικότερη ασφάλεια. Επίσης, παρουσιάζονται παραδείγματα μηχανικής μάθησης για έξυπνο σχεδιασμό ελέγχου ταυτότητας, συγκεκριμένα για παραμετρικούς και μη παραμετρικούς και με επίβλεψη και χωρίς επίβλεψη αλγορίθμους μάθησης καθώς και αλγορίθμους με ενίσχυση.

Στο [7] παρουσιάζονται εφαρμογές τεχνητής νοημοσύνης και μηχανικής μάθησης στην ασφάλεια, για την ασφάλεια του δικτύου 5G, τις επιπτώσεις τους και τις πιθανές ερευνητικές οδηγίες. Επίσης, συζητείται μια επισκόπηση των βασικών σημείων συλλογής δεδομένων στην αρχιτεκτονική 5G για ταξινόμηση απειλών και ανίχνευση ανωμαλιών.

Στο [10] προτάθηκε η ασφάλεια που καθορίζεται από λογισμικό (Software Defined Security – SDS) ως μέσο για την παροχή ενός συστήματος άμυνας δικτύου, το οποίο είναι αυτοματοποιημένο, ευέλικτο και κλιμακούμενο. Στο SDS χρησιμοποιήθηκαν οι τρέχουσες εξελίξεις της μηχανικής μάθησης για να σχεδιαστεί ένα συνελικτικό νευρωνικό δίκτυο (Convolutional Neural Network - CNN) χρησιμοποιώντας την νευρωνική αρχιτεκτονική αναζήτησης (Neural Architecture Search – NAS) για την ανίχνευση της ανώμαλης κίνησης στο δίκτυο. Το SDS μπορεί να εφαρμοστεί σε ένα σύστημα ανίχνευσης εισβολής για να δημιουργήσει μια πιο προληπτική και end-to-end άμυνα για ένα δίκτυο 5G. Για να ελεγχθεί αυτή η υπόθεση, συλλέχθηκαν και αναλύθηκαν φυσιολογικές και ανώμαλες ροές δικτύου από προσομοιωμένο περιβάλλον με CNN. Σύμφωνα με τους συγγραφείς, τα αποτελέσματα αυτής

της μεθόδου είναι πολλά υποσχόμενα, καθώς το μοντέλο εντόπισε καλοήθη κυκλοφορία με ποσοστό ακρίβειας 100% και ανώμαλη κίνηση με ποσοστό ανίχνευσης 96,4%. Αυτό δείχνει την αποτελεσματικότητα της ανάλυσης ροής δικτύου για μια ποικιλία κοινών κακόβουλων επιθέσεων και παρέχει επίσης μια βιώσιμη επιλογή για τον εντοπισμό κρυπτογραφημένης κακόβουλης κίνησης δικτύου.

Στο [11] πραγματοποιήθηκε μια βιβλιογραφική μελέτη σχετικά με τα ζητήματα ασφαλείας σε εφαρμογές του διαδικτύου των πραγμάτων (Internet of Things - IoT) με έμφαση στο βιομηχανικό IoT, καθώς και με τη μηχανική μάθηση και την αναλυτική των μεγάλων δεδομένων που αποτελούν δυο σημαντικές τεχνολογίες για την ανάλυση και την ασφάλεια του IoT. Επίσης, μελέτησαν και υλοποίησαν την δική τους μελέτη περίπτωσης, η οποία περιλαμβάνει λεπτομέρειες από πραγματικές δοκιμές που δημιουργήθηκαν για την διεξαγωγή κυβερνοεπιθέσεων και για το σχεδιασμό ενός συστήματος εντοπισμού εισβολών (Intrusion Detection System - IDS). Έπειτα ανέπτυξαν επιθέσεις έγχυσης backdoor, command injection και Structured Query Language (SQL) στο σύστημα και απέδειξαν πώς ένα σύστημα ανίχνευσης ανωμαλιών που βασίζεται σε μηχανική μάθηση μπορεί να αποδώσει καλά στον εντοπισμό αυτών των επιθέσεων. Η απόδοση αξιολογήθηκε μέσω αντιπροσωπευτικών μετρήσεων για να εξαχθούν συμπεράσματα σχετικά με την αποτελεσματικότητα των μεθόδων.

3 *Θεωρητικό υπόβαθρο*

Σε αυτό το κεφάλαιο παρουσιάζεται η προαπαιτούμενη θεωρία σχετικά με τη μηχανική μάθηση και τα δίκτυα κινητής τηλεφωνίας της 5ης γενιάς (5G). Η μελέτη αυτής της θεωρίας είναι απαραίτητη ώστε να μπορεί ο αναγνώστης να κατανοήσει όσα θα αναλυθούν στο κύριο μέρος της εργασίας σχετικά με την ασφάλεια και τον έλεγχο ταυτότητας στα δίκτυα 5G, καθώς και πως μπορεί η μηχανική μάθηση να συμβάλει στα ζητήματα ασφαλείας που προκύπτουν.

3.1 Μηχανική μάθηση (Machine Learning)

Στην ενότητα αυτή γίνεται μία προσέγγιση της έννοιας της μηχανικής μάθησης, δηλαδή τί είναι η μηχανική μάθηση, ποια είναι τα βασικά χαρακτηριστικά της και ποιος ο τρόπος με τον οποίο λειτουργεί. Επίσης, σύντομη αναφορά γίνεται στους διαφορετικούς τύπους στους οποίους διακρίνεται η μηχανική μάθηση.

3.1.1 Ορισμός και χαρακτηριστικά μηχανικής μάθησης

Τις τελευταίες δεκαετίες, η Μηχανική Μάθηση (Machine Learning) έχει γίνει ένας από τους βασικούς άξονες της τεχνολογίας της πληροφορικής. Με τη διαρκώς αυξανόμενη ποσότητα δεδομένων να είναι διαθέσιμη, υπάρχει ένας σημαντικός λόγος να υποστηρίζεται ότι η έξυπνη ανάλυση δεδομένων θα αποτελέσει όλο και περισσότερο ένα απαραίτητο συστατικό για την τεχνολογική πρόοδο. Η μηχανική μάθηση μπορεί να εμφανιστεί σε πολλές μορφές [12].

Η μηχανική μάθηση λειτουργεί σε τρία βασικά στάδια, την προ-επεξεργασία, την μάθηση και την αξιολόγηση των δεδομένων. Η προ-επεξεργασία δεδομένων προετοιμάζει τις ακατέργαστες πληροφορίες (δηλαδή τα δεδομένα) στη σωστή μορφή για τα επακόλουθα βήματα μάθησης. Συνήθως, τα ακατέργαστά δεδομένα είναι αδόμητα. Το βήμα της προ-

επεξεργασίας μετατρέπει αυτές τις πληροφορίες σε ένα πλαίσιο (frame) που μπορεί να χρησιμοποιηθεί ως συμβολή στη μάθηση μέσω του καθαρισμού των πληροφοριών, της εξαγωγής, της αλλαγής και του συνδυασμού. Το στάδιο της μάθησης επιλέγει τους μαθησιακούς υπολογισμούς και τις παραμέτρους για την παραγωγή επιθυμητών αποδόσεων χρησιμοποιώντας τα δεδομένα εισόδου [13].

Ορισμένες τεχνικές μάθησης, ιδιαίτερα η αυθεντική μάθηση (authentic learning), μπορούν επίσης να χρησιμοποιηθούν για την προ-επεξεργασία πληροφοριών. Η αξιολόγηση γίνεται μετά από την απόφαση για την εκτέλεση των εκπαιδευμένων μοντέλων. Για παράδειγμα, η αξιολόγηση εκτέλεσης ενός ταξινομητή περιλαμβάνει προσδιορισμό του συνόλου των δεδομένων, μέτρηση της εκτέλεσης, εκτίμηση του σφάλματος και πραγματικές δοκιμές [13].

Με πιο απλά λόγια, το βασικό επίπεδο της μηχανικής μάθησης αναφέρεται σε κάθε τύπο προγράμματος (λογισμικού) υπολογιστή που μπορεί να μάθει από μόνο του χωρίς να χρειάζεται να προγραμματιστεί ρητά από έναν άνθρωπο. Μερικά παραδείγματα της μηχανικής μάθησης που συναντάμε στην καθημερινότητα μας αποτελούν τα φίλτρα της ανεπιθύμητης αλληλογραφίας, οι ανιχνευτές απάτης (fraud detector) και οι προτάσεις για τα προϊόντα [14].

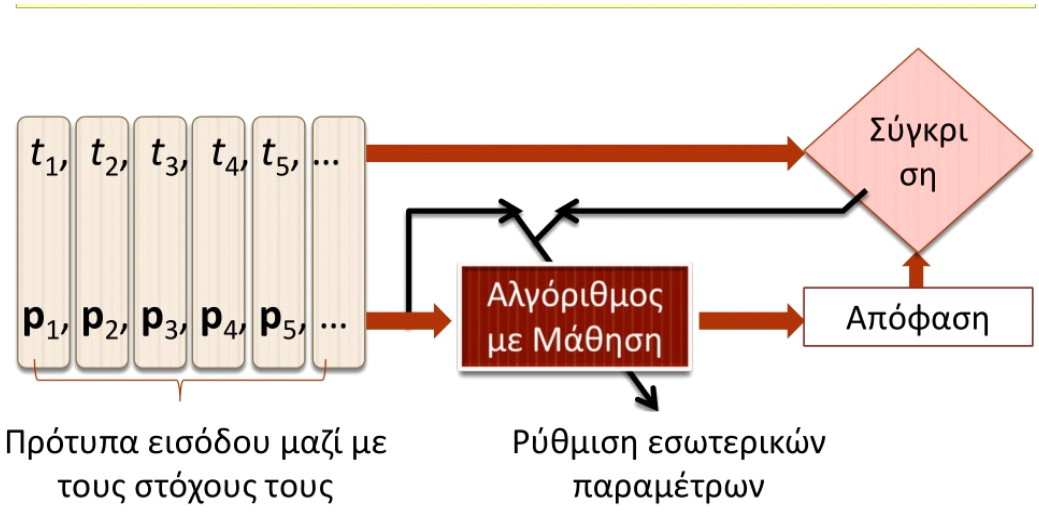
Επομένως, η μηχανική μάθηση μπορεί να οριστεί ως ένα σύνολο αλγορίθμων που μπορούν να μαθαίνουν και να κάνουν προβλέψεις σχετικά με τα δεδομένα. Αυτοί οι αλγόριθμοι δημιουργούν ένα μοντέλο από εισόδους προκειμένου να κάνουν προβλέψεις αντί να ακολουθούν αυστηρά στατικές οδηγίες προγράμματος. Η μηχανική μάθηση ορίζει την ανάγκη αυτόματης επεξεργασίας μεγάλου συνόλου δεδομένων προκειμένου να δημιουργηθούν χρήσιμες γνώσεις και να ληφθούν οι κατάλληλες αποφάσεις. Εν συντομία, η μηχανική μάθηση μπορεί να βοηθήσει στη βελτίωση της απόδοσης ενός συγκεκριμένου συνόλου εργασιών μέσω της δημιουργίας ενός μοντέλου που μπορεί να βρει μοτίβα χρησιμοποιώντας αλγόριθμους μάθησης.

3.1.2 Τύποι μηχανικής μάθησης

Η μηχανική μάθηση μπορεί να διαχωριστεί σε μάθηση με επίβλεψη (supervised), χωρίς επίβλεψη (unsupervised) και με ημιεπίβλεψη (semi-supervised). Η μάθηση με ημιεπίβλεψη συνδυάζει προσεγγίσεις με και χωρίς επίβλεψη [14][15]. Επίσης, άλλους τύπους μάθησης αποτελούν η μάθηση με ενίσχυση (reinforcement learning), η βαθιά μάθηση (deep learning), η βαθιά μάθηση με ενίσχυση (deep reinforcement learning) [16] και η μεταφορά μάθησης (transfer learning).

3.1.2.1 Μάθηση με επίβλεψη (supervised learning)

Στη μάθηση με επίβλεψη, ο χρήστης εκπαιδεύει το πρόγραμμα για να δημιουργήσει μια απάντηση βασισμένη σε ένα γνωστό και επισημασμένο σύνολο δεδομένων [14]. Μία εφαρμογή λαμβάνει πρότυπα εισόδου καθώς και τις κλάσεις που ανήκουν (στόχοι), σκοπεύει στο να διδάξει στο σύστημα ένα κανόνα για να αντιστοιχήσει το κάθε πρότυπο με την αντίστοιχη κλάση του.



Εικόνα 1: Μάθηση με επίβλεψη [17]

Με άλλα λόγια, η μάθηση με επίβλεψη είναι μια τεχνική μηχανικής μάθησης που δίνει δεδομένα εκπαίδευσης που περιέχουν τη «σωστή απάντηση» για κάθε παράδειγμα, προκειμένου να αναπτυχθεί ένα μοντέλο πρόβλεψης. Δηλαδή, στην μάθηση με επίβλεψη υπάρχουν οι ερωτήσεις και οι απαντήσεις τους, τίθεται μία ερώτηση στο εκπαιδευόμενο σύστημα, αυτό με τη σειρά του δίνει μία απάντηση, έπειτα συγκρίνεται η απάντηση του με την σωστή απάντηση, αν είναι λάθος γίνεται μία τροποποίηση των παραμέτρων του συστήματος μέχρι αυτό να βρει την σωστή απάντηση [17].

Αυτός ο τύπος μηχανικής μάθησης χρησιμοποιείται περισσότερο για προβλήματα ταξινόμησης (classification), πρόγνωσης (prediction) και διερμηνείας (interpretation) [15]. Στη μάθηση με επίβλεψη οι αλγόριθμοι μηχανικής μάθησης που ασχολούνται περισσότερο με την ταξινόμηση (classification) είναι οι γραμμικοί ταξινομητές (linear classifiers), η αλγοριθμική παλινδρόμηση (logistic regression), ο ταξινομητής Naïve Bayes (Classifier Naive Bayes), ο μηχανισμός perceptron (προσομοίωση ανθρώπινου εγκεφάλου), οι διανυσματικές υποστηρικτικές μηχανές (Support Vector Machines - SVM), οι τετραγωνικοί ταξινομητές (quadratic classifiers), η ομαδοποίηση (K-Means Clustering), η ενίσχυση (boosting), τα δέντρα απόφασης (decision tress), τα τυχαία δάση (random forest - RF), τα

νευρωνικά δίκτυα (neural networks), τα δίκτυα Bayesian (Bayesian networks) και άλλα [14][18].

3.1.2.2 Μάθηση χωρίς επίβλεψη (Unsupervised learning)

Στην μη επιτηρούμενη μάθηση χωρίς επίβλεψη, οι αλγόριθμοι δημιουργούν απαντήσεις σε άγνωστα και μη επισημασμένα δεδομένα. Οι επιστήμονες δεδομένων συνήθως χρησιμοποιούν μη επιτηρούμενες τεχνικές για την ανεύρεση προτύπων σε νέα σύνολα δεδομένων. Συχνά, στη μηχανική μάθηση χωρίς επίβλεψη χρησιμοποιούνται αλγόριθμοι ομαδοποίησης (clustering), όπως ο αλγόριθμος K-means [14], καθώς και αλγόριθμοι ανάλυσης συσχετισμών (association analysis) [15].

3.1.2.3 Μάθηση με ημιεπίβλεψη (Semisupervised learning)

Στον πραγματικό κόσμο τα δεδομένα είναι ένα μείγμα δομημένων και μη δομημένων δεδομένων. Τα δομημένα δεδομένα αντιμετωπίζονται αποτελεσματικά από αλγόριθμους μάθησης με επίβλεψη, ενώ η μάθηση χωρίς επίβλεψη λειτουργεί αποτελεσματικά σε μη δομημένα δεδομένα. Στην περίπτωση ενός συνδυασμού δεδομένων δομημένων και αδόμητων, η μάθηση με ημιεπίβλεψη είναι η λύση. Η μάθηση με ημιεπίβλεψη μπορεί να χρησιμοποιηθεί σε αρκετές εφαρμογές σε πραγματικό χρόνο, όπως η επεξεργασία της φυσικής γλώσσας, η ταξινόμηση του περιεχομένου ιστού, η αναγνώριση ομιλίας, το φιλτράρισμα ανεπιθύμητων μηνυμάτων, η παρακολούθηση βίντεο και άλλα [19].

3.1.2.4 Μάθηση με ενίσχυση (Reinforcement learning)

Η μηχανική μάθηση με ενίσχυση αναφέρεται στη διαδικασία κατά την οποία ένας αλγόριθμος μαθαίνει μια στρατηγική με ενέργειες αλληλοεπιδρώντας με το περιβάλλον. Αυτός ο τύπος μηχανικής μάθησης χρησιμοποιείται κατά κύριο λόγο σε προβλήματα σχεδιασμού (planning), παραδείγματος χάριν σε ρομπότ για τον έλεγχο της κίνησης ή σε εργοστάσια για τη βελτιστοποίηση των εργασιών [15]. Στην πραγματικότητα, ο αλγόριθμος μάθησης με ενίσχυση διαθέτει μόνο μια συνάρτηση επιβράβευσης/κόστους (reward / cost function), η οποία δείχνει στον αλγόριθμο πότε τα πάει καλά και πότε δεν τα πάει καλά [20]. Συγκεκριμένα, η μηχανή κατά την εκπαίδευση της επιβραβεύεται όταν κάνει μία σωστή ενέργεια και τιμωρείται όταν κάνει μία λάθος. Έτσι, μαθαίνει με τις κινήσεις και σύμφωνα με το πρόβλημα ακολουθεί τις διαδρομές με τις περισσότερες επιβραβεύσεις και τις λιγότερες ή καθόλου τιμωρίες. Αυτός ο τύπος μάθησης χρησιμοποιείται σε προβλήματα περίπλοκων χώρων όπως λαβύρινθοι [21][22].

Η μάθηση με ενίσχυση δεν χρειάζεται προηγούμενη γνώση, υπάρχει ένας πράκτορας (agent) που αλληλοεπιδρά με το περιβάλλον και μαθαίνει από τα παραδείγματα, μαθαίνει με την εξερεύνηση του περιβάλλοντος και την εκμετάλλευση της γνώσης [20]. Ουσιαστικά, στο πλαίσιο της μάθησης με ενίσχυση, ο πράκτορας είναι μια οντότητα ή αλλιώς ένα σύστημα, το οποίο μαθαίνει από τις αποφάσεις και λαμβάνει αποφάσεις, ενώ οτιδήποτε άλλο εκτός του πράκτορα στη μάθηση με ενίσχυση ονομάζεται περιβάλλον (environment) με το οποίο αλληλοεπιδρά ο πράκτορας [23]. Ο πράκτορας εξαρτάται από τη δική του εμπειρία και αποκτά την κατάλληλη αξιολόγηση του περιβάλλοντος και αναθεωρεί τη δική του στρατηγική δράσης για να προσαρμοστεί στο περιβάλλον [20].

Στο πλαίσιο των σύγχρονων δικτύων, όπως για παράδειγμα σε ένα δίκτυο του Διαδικτύου των Πραγμάτων (Internet of Things – IoT). Περισσότερες πληροφορίες για το Διαδίκτυο των Πραγμάτων δίνονται στο ακόλουθο κεφάλαιο), τα οποία μπορεί να λειτουργούν πιο αυτόνομα (δηλαδή, χωρίς την ανθρώπινη συμμετοχή για τη λήψη και επεξεργασία των πληροφοριών ή την εκτέλεση εργασιών) και σε πιο απομακρυσμένες περιοχές, οι οντότητες που περιλαμβάνονται σε ένα τέτοιο δίκτυο, θα πρέπει να λαμβάνουν αποφάσεις ώστε να μεγιστοποιηθεί η απόδοση του δικτύου, εντός ενός περιβάλλοντος που χαρακτηρίζεται από αβεβαιότητα. Η μάθηση με ενίσχυση μπορεί να χρησιμοποιηθεί αποτελεσματικά για να επιτρέψει στις οντότητες ενός τέτοιου δικτύου να αποκτήσουν καλύτερη πολιτική, όπως για παράδειγμα για αποφάσεις ή δράσεις με δεδομένο ότι πρόκειται για δίκτυα και χώρους μικρής κλίμακας. Παρόλα αυτά, αυτή η μέθοδος μάθησης δεν μπορεί να εφαρμοστεί σε δίκτυα μεγάλης κλίμακας, καθώς θα πρέπει να διερευνήσει και να αποκτήσει γνώση ενός ολόκληρου συστήματος και κατά συνέπεια χρειάζεται πολύς χρόνος για να επιτευχθεί η βέλτιστη πολιτική. Κατά συνέπεια, οι εφαρμογές της μάθησης ενίσχυσης είναι πολύ λίγες στην πράξη [16].

3.1.2.5 Βαθιά μάθηση (Deep Learning)

Η βαθιά μάθηση έχει προταθεί προκειμένου να αντιμετωπιστεί το πρόβλημα των δικτύων μεγάλης κλίμακας [16]. Η βαθιά μάθηση είναι ένα υποπεδίο της μηχανικής μάθησης και σε αντίθεση με τα κλασικά εργαλεία μηχανικής μάθησης, τα οποία βασίζονται κατά κύριο λόγο σε χαρακτηριστικά που έχουν οριστεί από ειδικούς του τομέα, οι αλγόριθμοι βαθιάς μάθησης εξάγουν ιεραρχικά γνώσεις από ανεπεξέργαστα δεδομένα μέσω πολλαπλών επιπέδων μη γραμμικών μονάδων επεξεργασίας, προκειμένου να προβλέψουν ή να αναλάβουν ενέργειες σύμφωνα με κάποιο στόχο. Επομένως, η βαθιά μάθηση είναι η μελέτη τεχνητών νευρικών δικτύων και συναφών αλγορίθμων μηχανικής μάθησης που περιέχουν περισσότερα από ένα κρυμμένα επίπεδα. Οι πιο συνηθισμένοι αλγόριθμοι βαθιάς μάθησης βρίσκουν εφαρμογή σε

μοντέλα τα οποία αποτελούν τα Νευρωνικά Δίκτυα (Neural Networks) που έχουν επαρκή αριθμό κρυφών επιπέδων (συνήθως περισσότερα από ένα) [24].

Ο βασικός στόχος των βαθιών νευρωνικών δικτύων είναι η προσέγγιση πολύπλοκων λειτουργιών μέσω μιας σύνθεσης απλών και προκαθορισμένων λειτουργιών μονάδων (ή νευρώνων). Μια τέτοια αντικειμενική συνάρτηση μπορεί να είναι σχεδόν οποιουδήποτε τύπου, όπως μια χαρτογράφηση μεταξύ εικόνων και των ετικετών της κατηγορίας τους (ταξινόμηση), υπολογισμός μελλοντικών τιμών μετοχών βάσει ιστορικών τιμών (παλινδρόμηση), ή ακόμη και απόφαση της επόμενης βέλτιστης κίνησης σκακιού δεδομένης της τρέχουσας κατάστασης του πίνακα (έλεγχος) [24]. Διευκρινίζεται ότι με τον όρο αντικειμενική συνάρτηση (objective function) εννοείται η έκφραση που δείχνει τη σχέση μεταξύ των μεταβλητών του προβλήματος και του σκοπού του συστήματος. Μια αντικειμενική συνάρτηση μπορεί να έχει δύο κατευθύνσεις τη μεγιστοποίηση ή την ελαχιστοποίηση μιας συγκεκριμένης λειτουργίας του συστήματος.

Η αρχιτεκτονική του νευρωνικού δικτύου μοιάζει με τη διαδικασία αντίληψης σε έναν εγκέφαλο, όπου ένα συγκεκριμένο σύνολο μονάδων ενεργοποιείται δεδομένου του τρέχοντος περιβάλλοντος, επηρεάζοντας την έξοδο του μοντέλου νευρωνικού δικτύου [24].

Από τα πιο δημοφιλή νευρωνικά δίκτυα, αποτελούν τα **Συνελικτικά Νευρωνικά Δίκτυα** (Convolutional Neural Network - CNN). Η ονομασία **Συνελικτικά** προέρχεται από τα μαθηματικά και τη γραμμική άλγεβρα και συγκεκριμένα από την λειτουργία μεταξύ των πινάκων που ονομάζεται συνέλιξη. Τα CNN έχουν πολύ υψηλή απόδοση σε προβλήματα της μηχανικής μάθησης, ιδίως σε εφαρμογές που ασχολούνται με δεδομένα εικόνας (όπως η ταξινόμηση ενός μεγάλου συνόλου δεδομένων (dataset) εικόνων (ImageNet)), την επεξεργασία φυσικής γλώσσας (natural language processing - NLP) [25].

3.1.2.6 Βαθιά μάθηση με ενίσχυση (Deep Reinforcement Learning)

Η βαθιά μάθηση μπορεί να ξεπεράσει τους περιορισμούς της μάθησης με ενίσχυση. Ο στόχος αυτού του τύπου μάθησης είναι η δημιουργία αυτόνομων συστημάτων με υψηλότερη κατανόηση του οπτικού κόσμου. Αυτός ο συνδυαστικός τύπος μάθησης συνδυάζει πραγματικά τεχνικές βαθιάς μάθησης (βαθιά νευρωνικά δίκτυα) με αλγόριθμους μάθησης με ενίσχυση (Q-learning, Actor-kritik). Ο Q-learning είναι ένα αλγόριθμος της μάθησης με ενίσχυση που δεν απαιτεί κάποιο μοντέλο για να μάθει το περιβάλλον και χρησιμοποιείται για να μάθει την ποιότητα των ενεργειών και να πει στον πράκτορα ποια ενέργεια πρέπει να κάνει και ποια όχι, ενώ παράλληλα μπορεί να χειριστεί προβλήματα με στοχαστικές μεταβάσεις και ανταμοιβές, χωρίς να απαιτείται προσαρμογή. Ο αλγόριθμος Q-learning σε

συνδυασμό με τα βαθιά νευρωνικά δίκτυα δίνει το λεγόμενο Deep Q-learning, το οποίο χρησιμοποιείται στον τεμαχισμό δικτύου 5G [16].

3.1.2.7 Μεταφορά μάθησης (Transfer learning)

Η μεταφορά μάθησης είναι μια έννοια που στοχεύει στη χρήση της μάθησης και της εμπειρίας από μια παλιά εργασία για την ταχύτερη επίλυση νέων προβλημάτων ή με καλύτερες λύσεις. Οι παραδοσιακές τεχνικές μηχανικής μάθησης προσπαθούν να μάθουν κάθε εργασία από το μηδέν, ενώ οι τεχνικές μεταφοράς μάθησης προσπαθούν να μεταφέρουν τη γνώση από ορισμένες προηγούμενες εργασίες σε μια εργασία στόχου, όπου η εργασία στόχου έχει λιγότερα δεδομένα εκπαίδευσης υψηλής ποιότητας [26].

3.2 Δίκτυα κινητής τηλεφωνίας 5G

Στην ενότητα αυτή παρουσιάζονται τα βασικά χαρακτηριστικά των δικτύων 5G, οι βασικές τεχνολογίες που συμπεριλαμβάνονται στα δίκτυα 5G, καθώς επίσης προσεγγίζεται η αρχιτεκτονική των δικτύων 5G, όπως έχει προταθεί από διάφορους οργανισμούς.

3.2.1 Βασικά χαρακτηριστικά

Πρόσφατα, παγκόσμιες εταιρείες έχουν αναλάβει το προβάδισμα στο διαγωνισμό για την νέα κυτταρική τεχνολογία (cellular technology) 5ης γενιάς (5G), η οποία θεωρείται ότι είναι η πιο σημαντική πηγή εσόδων του μέλλοντος [4]. Τα δίκτυα 5G έχουν πολύ μεγαλύτερες ταχύτητες μετάδοσης δεδομένων από τα τρέχοντα δίκτυα 4G. Ειδικότερα, το δίκτυο 5G έχει τη δυνατότητα να παρέχει ρυθμούς μετάδοσης δεδομένων έως και 10 Gbps, που είναι 10 έως 100 φορές υψηλότεροι από τα δίκτυα 4G και 4G-LTE. Τα δίκτυα 5G αναμένεται να ξεπεράσουν τα υπερ-ευρυζωνικά (ultra-broadband) δίκτυα, δηλαδή τα δίκτυα οπτικών ινών, και να συνδυάσουν υπάρχουσες τεχνολογίες όπως το Διαδίκτυο των Πραγμάτων (Internet of Things -IoT) [5][27], τα νέφη (clouds), τα μεγάλα δεδομένα (big data), την τεχνητή νοημοσύνη (Artificial Intelligence - AI) και το blockchain για να υποστηρίξουν τη δημιουργία καινοτόμων υπηρεσιών. Εκτός από τη βελτίωση της ταχύτητας, ένα άλλο σημαντικό χαρακτηριστικό του 5G είναι η χαμηλότερη καθυστέρηση. Στην πραγματικότητα, στην εποχή των δικτύων 5G, ο χρόνος καθυστέρησης θα είναι μικρότερος από ένα χιλιοστό του δευτερολέπτου (ms), ο οποίος είναι σχεδόν ίσος με τον χρόνο μηδενικής απόκρισης δεδομένων στον πραγματικό κόσμο [27].

Η ένωση Next Generation Mobile Networks (NGMN) περιγράφει μια σειρά από αναδυόμενες περιπτώσεις χρήσης 5G που εστιάζουν σε [28]:

- Ευρυζωνική πρόσβαση σε πυκνές περιοχές: Παρέχει ευρυζωνική πρόσβαση με εύρος ζώνης έως και 10 Gbps και διαθεσιμότητα υπηρεσιών σε πυκνοκατοικημένες περιοχές, όπως για παράδειγμα σε πυκνά αστικά κέντρα ή εκδηλώσεις, στάδια ή υπαίθρια φεστιβάλ.
- Ευρυζωνική πρόσβαση παντού: Στοχεύει σε ελάχιστο εύρος ζώνης τουλάχιστον 50Mbps, διασφαλίζοντας μια συνδεδεμένη παγκόσμια κοινωνία, μέσω διαδικτύου υψηλής ταχύτητας.
- Υψηλή κινητικότητα χρηστών: παρέχει ευρυζωνική υποστήριξη για χρήστες κινητών συσκευών σε πολύ γρήγορα κινούμενα οχήματα όπως τρένα υψηλής ταχύτητας.
- Μαζικό Διαδίκτυο των Πραγμάτων (Massive IoT): Παρέχει ευρυζωνική πρόσβαση για εξαιρετικά πυκνά δίκτυα αισθητήρων και ενεργοποιητών.
- Εξαιρετική επικοινωνία σε πραγματικό χρόνο: Παρέχει συνδεσιμότητα με εξαιρετικά μικρή καθυστέρηση, κάτι που είναι απαραίτητο στις διαδραστικές επικοινωνίες. Έτσι, παρέχεται εξαιρετικά αξιόπιστη επικοινωνία, αυξάνεται η διαθεσιμότητα σύνδεσης στο δίκτυο κάτι που είναι απαραίτητο σε διαδραστικές επικοινωνίες και την αυτόνομη οδήγηση.
- Γραμμή επικοινωνίας: Υποστηρίζει τη συνδεσιμότητα σε περίπτωση φυσικών καταστροφών και καταστάσεων έκτακτης ανάγκης. Για αυτές τις περιπτώσεις χρήσης απαιτείται πολύ υψηλό επίπεδο διαθεσιμότητας και ικανότητα διατήρησης της κυκλοφορίας.

Η ανάπτυξη των δικτύων 5G έγκειται στην παροχή πολύ υψηλών ποσοστών δεδομένων και υψηλότερης κάλυψης μέσω της ανάπτυξης πυκνών σταθμών βάσης (dense base stations) με αυξημένη χωρητικότητα, σημαντικά καλύτερη ποιότητα υπηρεσίας (Quality of Service - QoS) και εξαιρετικά χαμηλή καθυστέρηση [5]. Επίσης, με βάση το εξαιρετικό εύρος ζώνης των δικτύων 5G, τη συνδεσιμότητα, την κάλυψη (σχεδόν 100%) και την ικανότητα σύνδεσης συσκευών, μπορεί να δημιουργηθεί ένα οικοσύστημα, όπου «έξυπνα δίκτυα» (“smart networks”) μπορούν να χρησιμοποιηθούν για μεγάλες ιατρικές συσκευές και να παρέχουν διαδραστικότητα σε πραγματικό χρόνο [29]. Το ίδιο ισχύει και για άλλες έξυπνες εφαρμογές.

Σύμφωνα με τον Τομέα Ραδιοεπικοινωνιών ITU (ITU Radio Communication- ITU-R) M.2083, το σύστημα 5G μπορεί να υποστηρίξει διάφορες περιπτώσεις χρήσης, οι οποίες χωρίζονται σε τρεις ευρείες κατηγορίες. α) Βελτιωμένες κινητές ευρυζωνικές συνδέσεις (Enhanced Mobile Broadband – eMBB), β) Επικοινωνίες τύπου μαζικής μηχανής (Massive Machine Type Communications - mMTC) και γ) Εξαιρετικά αξιόπιστες και χαμηλής καθυστέρησης επικοινωνίες (Ultra-Reliable and Low Latency Communications - URLLC).

Συγκεκριμένα, το eMBB αναφέρεται σε ευρυζωνικές υπηρεσίες κινητής τηλεφωνίας μεγάλου όγκου που έχουν υψηλές απαιτήσεις για εύρος ζώνης, όπως βίντεο υψηλής ευκρίνειας, Εικονική Πραγματικότητα (Virtual Reality - VR) και Επαυξημένη Πραγματικότητα (Augmented Reality - AR). Το mMTC αναφέρεται σε υπηρεσίες IoT μεγάλης κλίμακας, υπηρεσίες που περιλαμβάνουν υψηλές απαιτήσεις για πυκνότητα σύνδεσης, όπως η έξυπνη πόλη και η έξυπνη γεωργία και το URLLC αναφέρεται σε υπηρεσίες ευαίσθητες σε καθυστέρηση, όπως υποβοηθούμενη και αυτοματοποιημένη οδήγηση, βιομηχανικός αυτοματισμός και η απομακρυσμένη διαχείριση [30].

3.2.2 Βασικές τεχνολογίες

Ως μια νέα γενιά τεχνολογίας κινητής επικοινωνίας, το 5G έχει μια πολύ διαφορετική δομή δικτύου, δυνατότητες δικτύου και απαιτήσεις από τις προηγούμενες γενιές, και ενσωματώνεται ένας μεγάλος αριθμός τεχνολογιών.

Το 5G χρειάζεται νέες τεχνολογίες τόσο στο Δίκτυο Ραδιοπρόσβασης (Radio Access Network - RAN) όσο και στον πυρήνα του δικτύου κινητής (Core Network - CN), προκειμένου να αντιμετωπίσει τις απαιτητικές απαιτήσεις του 5G. Το δίκτυο Ραδιοπρόσβασης (RAN) χρειάζεται νέες τεχνολογίες φυσικού επιπέδου όπως Massive Multiple-Input Multiple-Output (Massive MIMO), Non-Orthogonal Multiple Access (NOMA), Full-Duplex (FD), millimeter Wave (mmWave) επικοινωνία, επικοινωνία συσκευή με συσκευή (Device-to-Device - D2D) και επικοινωνία με ορατό φως (visible light). Επιπλέον, το Cloud Radio Access Network (C-RAN), δηλαδή το Νέφος Δικτύου Ραδιοπρόσβασης, με τη βοήθεια της τεχνολογίας των υπολογιστικών νεφών (Cloud Computing) αποτελεί μια πολλά υποσχόμενη και οικονομικά αποδοτική αρχιτεκτονική δικτύου κινητής τηλεφωνίας για τη βελτίωση του φάσματος και της ενεργειακής απόδοσης των δικτύων 5G. Επιπλέον, το 5G χρησιμοποιεί προηγμένες τεχνολογίες, όπως δικτύωση που καθορίζεται από λογισμικό (Software Defined Networking - SDN), εικονικοποίηση λειτουργιών δικτύου (Network Function Virtualization- NFV), υπερσυγκέντρωση (ultra-densification), Διαδίκτυο των Πραγμάτων (Internet of Things – IoT), αυτο-οργάνωση δικτύου και τεμαχισμό δικτύου (network slicing). Όλες αυτές οι τεχνολογίες περιγράφονται στη συνέχεια.

3.2.2.1 Massive MIMO (Multiple Input Multiple Output)

Το μαζικό MIMO (massive Multiple Input Multiple Output) είναι μια προηγμένη καινοτομία που προέρχεται από την αναβάθμιση της τρέχουσας τεχνολογίας MIMO. Ο θεμελιώδης στόχος της αναβάθμισης σε μαζικό MIMO είναι να διαχωριστούν τα πλεονεκτήματα από το

παραδοσιακό MIMO της προηγούμενης γενιάς, ώστε να επιτευχθεί μεγαλύτερο εύρος επεκτείνοντας την απόδοση, την αποτελεσματικότητα του φάσματος και την ενεργειακή επάρκεια [4]. Το μαζικό MIMO προτείνεται ως μια από τις πιο βασικές λύσεις για τα δίκτυα 5G. Το μαζικό MIMO χρησιμοποιεί μεγάλες συστοιχίες κεραίας στους σταθμούς βάσης για ταυτόχρονη εξυπηρέτηση πολλών διαφορετικών αυτόνομων τερματικών, για να αυξήσει περισσότερο την χωρητικότητα και την απόδοση του συστήματος. Προκειμένου το μαζικό MIMO να εξυπηρετεί πολλούς χρήστες στους πόρους χρονικής συχνότητας χρησιμοποιεί χωρική πολυπλεξία (spatial multiplexing) και διπλή διαίρεση χρόνου (Time Division Duplexing - TDD) [31].

3.2.2.2 Υπολογιστικά Νέφη Κινητής (Mobile Cloud Computing)

Για την παροχή των απαραίτητων υπηρεσιών που προβλέπονται από το 5G, απαιτούνται νέες τεχνολογίες δικτύωσης, ανάπτυξης υπηρεσιών, αποθήκευσης και επεξεργασίας δεδομένων. Τα υπολογιστικά νέφη (cloud computing) παρέχουν έναν αποτελεσματικό τρόπο για τους χρήστες να διατηρούν δεδομένα, υπηρεσίες και εφαρμογές χωρίς οι τελικοί χρήστες να χρειάζεται να διαθέτουν την σχετική υποδομή για την επίτευξη αυτών των σκοπών. Επομένως, τα νέφη φέρνουν νέα τεχνολογικά διακριτά συστήματα μέσω των οποίων μπορούν να αναπτυχθούν πολλαπλές υπηρεσίες για να επιτευχθεί υψηλότερος βαθμός ευελιξίας και διαθεσιμότητας με λιγότερες δαπάνες κεφαλαίου (Capital Expenditures - CapEx) και λιγότερα λειτουργικά έξοδα (Operational Expenses - OpEx) [5].

Στα δίκτυα 5G, η αποθήκευση σε νέφη (cloud storage) θα είναι αναπόσπαστο μέρος. Τα υπολογιστικά νέφη κινητής (mobile cloud computing) θα είναι η κύρια μέθοδος υπολογισμού δεδομένων. Αυτή η τεχνολογία παρέχει τους πόρους υπολογισμών για χρήστες κινητών συσκευών, παρέχει τη δυνατότητα αποθήκευσης και επεξεργασίας δεδομένων εκτός των κινητών συσκευών. Όλο και περισσότεροι χρήστες χρησιμοποιούν διακομιστές αποθήκευσης νέφους και τα υπολογιστικά νέφη κινητής τους βοηθά προσφέροντας διαχείριση δεδομένων στους διακομιστές νέφους. Τα υπολογιστικά νέφη κινητής απαιτούν σταθερή σύνδεση με το διακομιστή νέφους και αυτό θέτει υψηλές απαιτήσεις. Η μεγάλη ποσότητα δεδομένων που παράγεται από τις κινητές συσκευές και τα δίκτυα επιβάλλει τη χρήση αναλυτικών δεδομένων [31][32], καθώς τα υπολογιστικά νέφη παράγουν μεγάλους όγκους δεδομένων, τα οποία αποτελούν τα μεγάλα δεδομένα (big data) [33], δηλαδή τη χρήση της αναλυτικής των μεγάλων δεδομένων. Η αναλυτική των μεγάλων δεδομένων αναφέρεται στη διαδικασία συλλογής, οργάνωσης και ανάλυσης μεγάλων συνόλων δεδομένων, ώστε να ανακαλυφθούν διαφορετικά μοτίβα και χρήσιμες πληροφορίες [34]. Μία από τις σημαντικότερες περιπτώσεις χρήσης υπολογιστικού νέφους στο 5G είναι η μαζική επικοινωνία τύπου μηχανής (mMTC).

Αυτή η περίπτωση χρήσης χαρακτηρίζεται από τεράστιο αριθμό συνδεδεμένων συσκευών, σχηματίζοντας ένα δίκτυο δισεκατομμυρίων αισθητήρων και ενεργοποιητών, υποστηρίζοντας έναν τεράστιο αριθμό συσκευών χαμηλού κόστους και ενέργειας. Επομένως, υπογραμμίζεται η ανάγκη για τη χρήση υπολογιστικών νεφών σε δυο διαφορετικές πτυχές. Η πρώτη αφορά την ανάγκη ανταλλαγής δεδομένων σε πραγματικό χρόνο μεταξύ συσκευών και η δεύτερη αναφέρεται στην ανάγκη των συσκευών που διαθέτουν χαμηλή αποθήκευση και για το λόγο αυτό απαιτείται να διαθέτουν εικονική χωρητικότητα αποθήκευσης σε νέφος [35].

3.2.2.3 Network Function Virtualization (NFV)

Η τεχνολογία Network Function Virtualization (NFV), δηλαδή η εικονικοποίηση δικτυακών λειτουργιών, διαχωρίζει τις λειτουργίες δικτύου από το εξειδικευμένο υλικό υποδομής (όπως για παράδειγμα τον δρομολογητή (router)), ώστε να μπορούν να εκτελούνται σε λογισμικό ως τυποποιημένο υλικό, παρέχοντας έτσι οφέλη από την άποψη της επεκτασιμότητας και της ευελιξίας. Το NFV μειώνει τις δαπάνες κεφαλαίου που απαιτείται για την αγορά συσκευών υλικού και μειώνει τα έξοδα λειτουργίας συνδυάζοντας πόρους για λειτουργίες εικονικού δικτύου που εκτελούνται σε κεντρικό συγκεντρωτικό διακομιστή [31]. Σύμφωνα με το Ευρωπαϊκό Ινστιτούτο Προτύπων Τηλεπικοινωνιών (European Telecommunications Standards Institute - ETSI), η συνολική αρχιτεκτονική του NFV αποτελείται από τέσσερα βασικά στοιχεία: 1) την υποδομή (infrastructure) NFV (NFVI), 2) τις λειτουργίες εικονικού δικτύου, 3) τις εποπτικές αρχές (hypervisor[1]) και 4) τη διαχείριση (management) και ενορχήστρωση (orchestration) NFV (NFV MANO) [31].

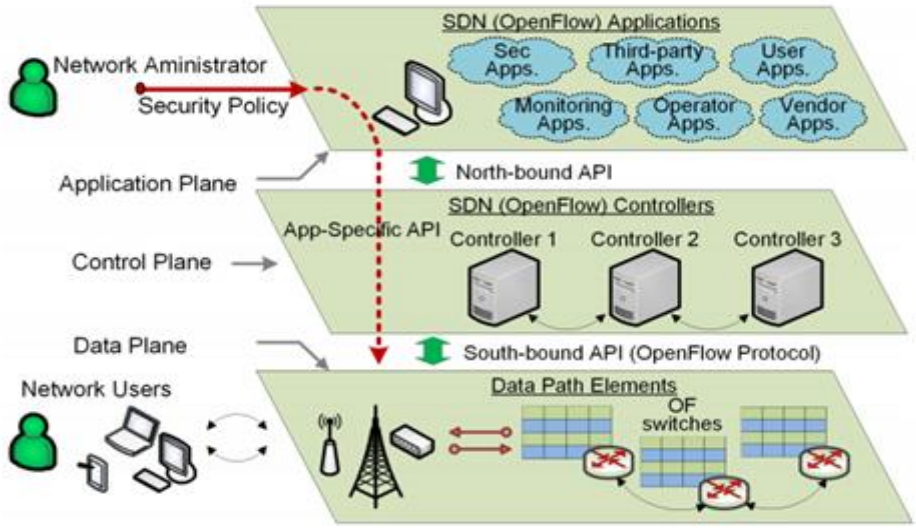
3.2.2.4 Software Defined Networking (SDN)

Το δίκτυο 5G είναι ένα πολύπλοκο σύστημα και για το λόγο αυτό η σχεδίαση του απαιτεί μια ευέλικτη αρχιτεκτονική. Αυτό μπορεί να πραγματοποιηθεί μέσω της τεχνολογίας SDN [31]. Η τεχνολογία Software Defined Networking (SDN), δηλαδή η τεχνολογία όπου το δίκτυο καθορίζεται από το λογισμικό, επιτρέπει την εκμάθηση της λειτουργίας του δικτύου διαχωρίζοντας τα επίπεδα ελέγχου δικτύου και τη προώθηση των δεδομένων [5]. Πιο συγκεκριμένα, οι στόχοι της τεχνολογίας SDN είναι να αποσυνδέσουν το επίπεδο δεδομένων από το επίπεδο ελέγχου και να προτείνουν νέες λειτουργίες ελέγχου δικτύου [31].

Αναλυτικότερα, το SDN διαχωρίζει το επίπεδο ελέγχου δικτύου από το επίπεδο προώθησης (ή αλλιώς το επίπεδο δεδομένων ή το επίπεδο χρήστη) και συγκεντρώνει τον έλεγχο δικτύου

[1] ένα πρόγραμμα που χρησιμοποιείται για την εκτέλεση και τη διαχείριση ενός ή περισσότερων εικονικών μηχανών σε έναν υπολογιστή.

σε πλατφόρμες λογισμικού που βασίζονται στον έλεγχο δικτύου. Οι λειτουργίες του λογισμικού για τον έλεγχο του δικτύου συγκεντρώνονται λογικά βάσει του τρόπου με τον οποίο αλληλοεπιδρούν με τις συσκευές προώθησης μέσω προγραμματιζόμενων εφαρμογών (API). Αυτό επιτυγχάνει απλότητα στον έλεγχο, τη διαχείριση και τη λειτουργία του δικτύου και επιταχύνει την ανάπτυξη καινοτόμων δυνατοτήτων του δικτύου. Η αρχιτεκτονική SDN έχει τρία λειτουργικά επίπεδα με διασυνδέσεις μεταξύ των επιπέδων, όπως φαίνεται στην ακόλουθη εικόνα. Το OpenFlow είναι η πρώτη εφαρμογή του SDN που ακολουθεί την αρχιτεκτονική τριών επιπέδων του SDN με εφαρμογές OpenFlow, ελεγκτές OpenFlow και διακόπτες OpenFlow [36].



Εικόνα 2: Αρχιτεκτονική SDN [36]

Οι τεχνολογίες SDN και NFV, που αλληλοσυμπληρώνονται, βελτιώνουν την ελαστικότητα του δικτύου, απλοποιούν τον έλεγχο και τη διαχείριση του δικτύου, ξεπερνούν το εμπόδιο συγκεκριμένων ιδιοκτησιακών λύσεων προμηθευτών και, ως εκ τούτου, θεωρούνται εξαιρετικά σημαντικά για τα μελλοντικά δίκτυα [5].

3.2.2.5 Millimeter Wave (mmWave)

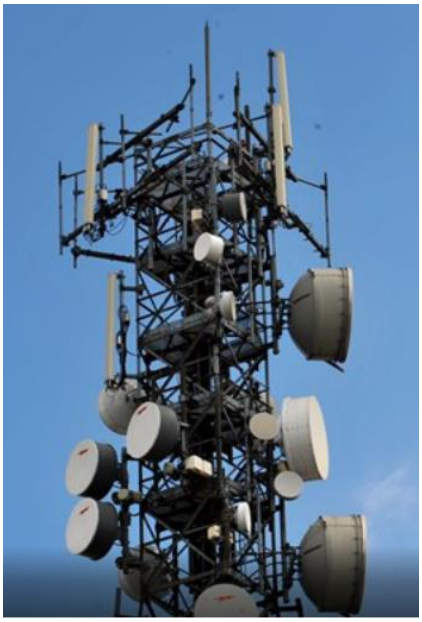
Το mmWave αναφέρεται στο φάσμα ραδιοσυχνοτήτων μεταξύ 30 GHz και 300GHz. Το mmWave έχει πολύ μικρό μήκος κύματος που κυμαίνεται από 10 mm στα 30 GHz και μειώνεται σε 1mm στα 300GHz που είναι αχρησιμοποίητο, οπότε ο στόχος του mmWave είναι να αυξήσει σημαντικά το διαθέσιμο εύρος ζώνης.

Οι χαμηλότερες συχνότητες συσσωρεύονται περισσότερο με σήματα τηλεόρασης και ραδιοφώνου, καθώς και με τρέχοντα δίκτυα 4G LTE. Το mmWaves μπορεί να μεταφέρει

δεδομένα ακόμα πιο γρήγορα, όμως σε μικρότερη απόσταση. Τα mmWaves μπορούν να χρησιμοποιηθούν για την παροχή υπηρεσιών επικοινωνίας πολλαπλών gigabit, όπως τηλεόραση υψηλής ευκρίνειας (HDTV) και βίντεο Ultra-High Definition (UHDV), καθώς επιτρέπουν υψηλότερους ρυθμούς δεδομένων έως και 10 Gbps [37]. Εν συντομία, οι ζώνες χαμηλότερης συχνότητας καλύπτουν μεγαλύτερες αποστάσεις, αλλά προσφέρουν χαμηλότερες ταχύτητες δεδομένων, ενώ οι ζώνες υψηλής συχνότητας καλύπτουν πολύ μικρότερες περιοχές αλλά μπορούν να μεταφέρουν πολύ περισσότερα δεδομένα, δηλαδή παρέχουν μεγαλύτερο εύρος ζώνης δικτύου, χαμηλότερη καθυστέρηση και πολύ μεγαλύτερη πυκνότητα σύνδεσης.

3.2.2.6 Πολύ πυκνό και ετερογενές δίκτυο

Η πολύ πυκνή δικτύωση (ultra-dense networking) είναι μια πολλά υποσχόμενη τεχνολογία για την αντιμετώπιση των απαιτήσεων της ταχύτερης μεταφοράς δεδομένων σε επικοινωνίες 5G και για τη βελτίωση της φασματικής απόδοσης του δικτύου και της απόδοσης του συστήματος. Η εξαιρετικά πυκνή ετερογενής δικτύωση αναφέρεται κυρίως στην πολύ υψηλή πυκνότητα των κυψελοειδών δικτύων, συμπεριλαμβανομένης τόσο της πυκνότητας των κινητών συσκευών όσο και της πυκνότητας του σταθμού βάσης, όπου η πυκνότητα των σταθμών βάσης μπορεί να υπερβαίνει εκείνη των κινητών συσκευών [24]. Ως σταθμοί βάσης ορίζονται ως οι σταθερές κεραίες που χρησιμοποιούνται σε ένα δίκτυο κινητής τηλεφωνίας, και συγκεκριμένα στις κυψελωτές επικοινωνίες. Ουσιαστικά, κάθε σταθμός βάσης αποτελείται από κεραίες και ηλεκτρονικό εξοπλισμό. Η τοποθέτηση τους γίνεται σε ψηλά σημεία (π.χ. σε ταράτσα πολυκατοικίας) και οι κεραίες τροφοδοτούνται με σήματα μέσω καλωδίων, έπειτα, τα σήματα εκπέμπονται από τις κεραίες σε μορφή ραδιοκυμάτων σε όλη την περιοχή που περιλαμβάνει το σταθμό βάσης [38]. Στην ακόλουθη φωτογραφία δίνεται μια εικόνα ενός συνηθισμένου σταθμού βάσης κυψελωτού δικτύου κινητής τηλεφωνίας.



Εικόνα 3: Τυπικός σταθμός βάσης κινητής τηλεφωνίας [39]

Η γενική αρχιτεκτονική του εξαιρετικά πυκνού δικτύου αποτελείται από picocells, femtocells, αποτελούμενα από hotspots, relays και microcell. Η περιοχή του microcell χωρίζεται σε μικρά κύτταρα (smallcells).

3.2.2.7 Smallcells

Τα smallcells (μικρά κύτταρα ή μικρές κυψέλες) μπορούν να έχουν έναν κεντρικό σταθμό βάσης ή μια απομακρυσμένη κεφαλίδα ραδιοφώνου (radio header), η οποία μπορεί να συνδεθεί ενσύρματα ή ασύρματα στον πυρήνα του δικτύου [4]. Στην ακόλουθη εικόνα δίνεται ένα παράδειγμα smallcells σε μια κολώνα.

Εικόνα 4: Smallcells σε κολώνα [40]]

Τα smallcells μπορεί να έχουν διαφορετικά μεγέθη, και ανάλογα με το μέγεθος τους μπορούν να ταξινομηθούν ως: Femtocells, Picocells και Microcells [4]. Το καθένα βασίζεται στην ικανότητα κάλυψης και στον αριθμό των μεμονωμένων χρηστών που μπορεί να υποστηρίξει. Αναλυτικότερα, για τα femtocells, η περιοχή κάλυψης είναι από 10 έως 50 μέτρα, λειτουργούν σε εσωτερικό χώρο (indoor), η μετάδοση ισχύος είναι 100mW και 20dBm, και μπορεί να υποστηρίξει από 8 έως 16 χρήστες. Επομένως, τα femtocells αποτελούν μικρούς σταθμούς βάσης για κινητές συσκευές σε οικιακούς και επιχειρησιακούς χώρους και χρησιμοποιούνται κυρίως για αποσυμφόρηση των δικτύων και την επέκταση κάλυψης εντός των κτιρίων. Τα picocell μπορούν να υποστηρίζουν περιοχή κάλυψης από 100 έως 250 μέτρα, λειτουργούν επίσης σε εσωτερικό χώρο, αλλά και σε εξωτερικό χώρο, η μετάδοση ισχύος είναι 250mW και 20dBm και μπορούν να υποστηρίξουν 32 έως και 64 χρήστες. Τα picocells είναι ιδανικά για εφαρμογές σε γραφεία, νοσοκομεία, εμπορικά συγκροτήματα, σχολεία και πανεπιστήμια (κυρίως μικρές επιχειρήσεις) για εκτεταμένη κάλυψη δικτύου και μεγάλη απόδοση δεδομένων. Τέλος, τα microcells μπορούν να καλύψουν μια περιοχή από 500 μέτρα έως και 2,5 χιλιόμετρα με μετάδοση ισχύος 2000-5000mW και 33-37dBm και μπορεί να υποστηρίξει ταυτόχρονα 200 χρήστες και λειτουργεί σε εξωτερικό χώρο. Τα microcells αποτελούν την τελευταία από τις τεχνολογίες small-cells και αποτελούν μια κυψέλη σε ένα δίκτυο κινητής τηλεφωνίας που εξυπηρετείται από σταθμό βάσης χαμηλής ισχύος και καλύπτει περιορισμένη περιοχή, όπως εμπορικά κέντρα, ξενοδοχεία, μοναδικούς χώρους εντός έξυπνων πόλεων ή κόμβων μεταφοράς [41].

3.2.2.8 Διαδίκτυο των Πραγμάτων (Internet of Things – IoT)

Ο αριθμός των συσκευών αναμένεται να αυξηθεί δραματικά. Οι συσκευές IoT διαδραματίζουν σημαντικό ρόλο στο σχεδιασμό των συστημάτων 5G και τα διάφορα μέρη του δικτύου 5G, καθώς σχεδιάζονται λαμβάνοντας υπόψη τις απαιτήσεις του IoT [31].

Το Διαδίκτυο των Πραγμάτων (Internet of Things – IoT) είναι μια συλλογή από «πράγματα», τα οποία είναι ενσωματωμένα με ηλεκτρονικά, λογισμικά, αισθητήρες, ενεργοποιητές και είναι συνδεδεμένα με τη χρήση του διαδικτύου για τη συλλογή και ανταλλαγή δεδομένων μεταξύ τους. Οι συσκευές IoT είναι εξοπλισμένες με αισθητήρες και ισχύ επεξεργασίας που τις επιτρέπουν να αναπτυχθούν σε πολλά περιβάλλοντα [42]. Η διαφορά μεταξύ του IoT και του παραδοσιακού διαδικτύου είναι η απουσία του ανθρώπινου ρόλου.

Ο όρος IoT χρησιμοποιείται για να γίνει αναφορά στον αναδυόμενο παγκόσμιο δίκτυο που διασυνδέει έξυπνα αντικείμενα μέσω εκτεταμένων τεχνολογιών διαδικτύου και στο σύνολο των υποστηρικτικών τεχνολογιών που απαιτούνται για την πραγματοποίηση ενός τέτοιου οράματος (όπως για παράδειγμα η τεχνολογία αναγνώρισης ραδιοσυχνοτήτων (RFIDs), οι αισθητήρες και οι ενεργοποιητές, οι συσκευές επικοινωνίας από μηχανή σε μηχανή και άλλα [43].

3.2.2.9 Επικοινωνία συσκευή με συσκευή με συσκευή (Device to Device - D2D)

Η επικοινωνία D2D είναι ένα νέο παράδειγμα στα κυψελοειδή δίκτυα που επιτρέπει στους Εξοπλισμούς Χρήστη (User Equipment – Ue-s) να επικοινωνούν σε κοντινή απόσταση χρησιμοποιώντας έναν απευθείας σύνδεσμο αντί να μετακινούν το ραδιοφωνικό τους σήμα σε όλη τη διαδρομή μέσω του σταθμού βάσης ή του κεντρικού δικτύου [44].

Η επικοινωνία D2D σε ένα κυψελοειδές δίκτυο έχει πολλά πλεονεκτήματα, όπως τον περιορισμό του φορτίου στο σταθμό βάσης, τη μείωση της καθυστέρησης μετάδοσης από άκρο σε άκρο (end-to-end), την αύξηση της απόδοσης του φάσματος και τη μείωση της ισχύος μετάδοσης τερματικού. Για παράδειγμα, σε περιπτώσεις καταστροφών όταν δεν υπάρχει υποδομή ασύρματης επικοινωνίας, το τερματικό μπορεί να χρησιμοποιήσει το D2D για να επιτύχει επικοινωνία από άκρο σε άκρο και ακόμη και πρόσβαση στο δίκτυο κινητής τηλεφωνίας [31].

3.2.2.10 Τεμαχισμός δικτύου (network slicing)

Ο τεμαχισμός δικτύου (network slicing) στοχεύει στη διαίρεση του φυσικού δικτύου του χειριστή σε πολλαπλά εικονικά δίκτυα. Κάθε δίκτυο έχει διαφορετικές απαιτήσεις υπηρεσίας. Ο τεμαχισμός δικτύου επιτρέπει τον διαχωρισμό πολλών λογικών δικτύων σε ένα ανεξάρτητο

φυσικό δίκτυο. Με αυτόν τον τρόπο αποφεύγεται η κατασκευή ενός αποκλειστικού φυσικού δικτύου για κάθε υπηρεσία με χαμηλότερο κόστος κατασκευής. Κάθε ένα από τα τεμαχισμένα δίκτυα περιλαμβάνει τη δική του πρόσβαση ραδιοσυχνοτήτων, μεταφορά και δίκτυο πυρήνα που προορίζεται να εξυπηρετήσει έναν μόνο τύπο συσκευής και να παρέχει ανεξάρτητες πηγές δικτύου, ώστε να μην υπάρχει παρέμβαση μεταξύ άλλων τεμαχισμένων δικτύων. Ο τεμαχισμός δικτύου παρέχει μια μεγάλη ώθηση για κρίσιμες υπηρεσίες IoT και πραγματοποιεί μια δυναμική κατανομή πόρων δικτύου για διάφορα σενάρια κυκλοφορίας [31].

3.2.2.11 Multi-Access Edge Computing (MEC)

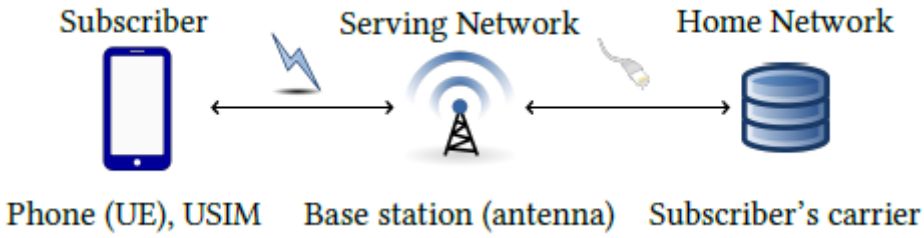
Η τεχνολογία MEC είναι σημαντική για τα δίκτυα 5G και αναμένεται να εξυπηρετήσει την επικοινωνία με χαμηλή καθυστέρηση. Η βασική της ιδέα είναι να μεταφέρει υπολογιστικούς πόρους, αποθήκευσης και δικτύωσης σε κοντινή απόσταση από τον εξοπλισμό του χρήστη (UE) με απώτερο στόχο τη μείωση της συμφόρησης του δικτύου κατά τη μεταφορά των δεδομένων από διαφορετικές συσκευές, και συνεπώς, να οδηγήσει σε γρηγορότερο χρόνο απόκρισης. Οι εφαρμογές MEC μπορούν να χρησιμοποιηθούν στην παράδοση περιεχομένου σε πραγματικό χρόνο, στην αναλυτική των μεγάλων δεδομένων, όπου η αποτελεσματικότητα του χρόνου είναι πολύ σημαντική για την ορθή λειτουργία ενός συστήματος. Στη γενική αρχιτεκτονική MEC, οι διακομιστές MEC παρέχουν υπολογιστικούς πόρους, χωρητικότητα αποθήκευσης, συνδεσιμότητα και πρόσβαση στην κυκλοφορία των χρηστών καθώς και στις πληροφορίες ραδιοφώνου και δικτύου [31].

Ουσιαστικά, η τεχνολογία MEC αποτελεί μια βασική περίπτωση χρήσης υπολογιστικού νέφους που έχει μεγάλη σημασία για δίκτυα κινητής τηλεφωνίας 5G. Η MEC επεκτείνει τις δυνατότητες της πληροφορικής και των υπολογιστικών νεφών στο RAN στην άκρη των δικτύων κινητής τηλεφωνίας. Αυτό παρέχει στους προγραμματιστές και στους παρόχους περιεχομένου άμεση πρόσβαση σε πληροφορίες ραδιοπρόσβασης σε πραγματικό χρόνο, προωθώντας έτσι εξαιρετικά χαμηλή καθυστέρηση και υψηλότερο εύρος ζώνης βελτιώνοντας τη συνολική εμπειρία των χρηστών [45].

3.2.3 Αρχιτεκτονική

Η βασική αρχιτεκτονική ενός κυψελοειδούς δικτύου περιλαμβάνει τρεις βασικές οντότητες: 1) Εξοπλισμός Χρήστη (User Equipment - UE), όπως για παράδειγμα κινητά τηλέφωνα ή συσκευές IoT, οι οποίες περιέχουν μια ενότητα ταυτότητας συνδρομητή, γνωστή ως USIM (Universal Subscriber Identity Module). Ουσιαστικά, ο συνδυασμός του UE και της USIM αποτελούν τον συνδρομητή. 2) Τα οικιακά δίκτυα (Home Networks - HN), τα οποία

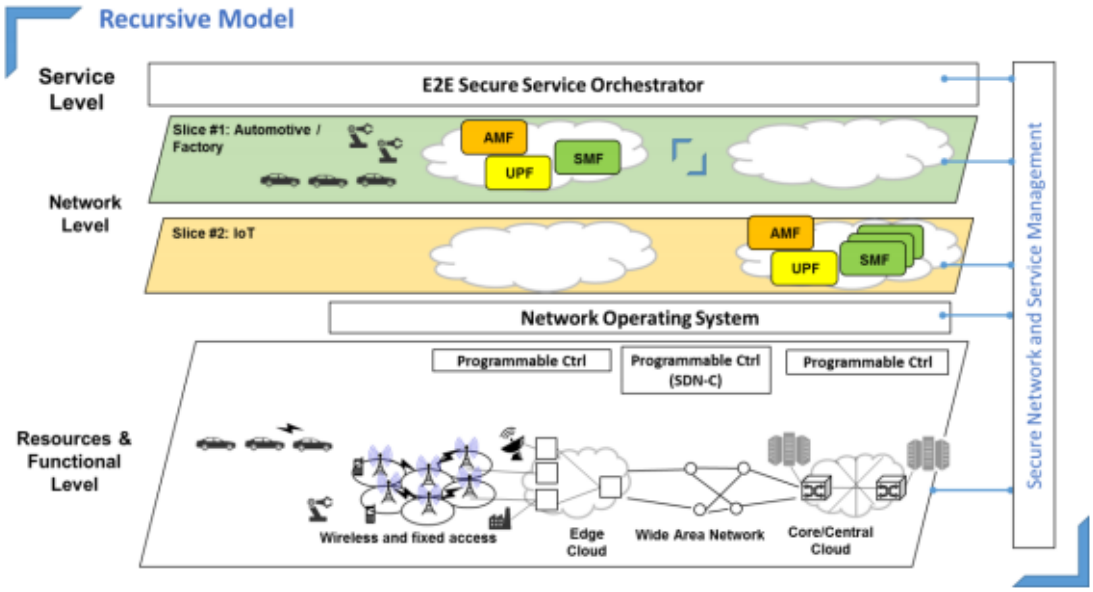
περιέχουν μια βάση δεδομένων των συνδρομητών τους και είναι υπεύθυνα για τον έλεγχο της ταυτότητάς τους. Ωστόσο, οι συνδρομητές μπορεί να βρίσκονται σε τοποθεσίες όπου το αντίστοιχο οικιακό τους δίκτυο (HN) δεν έχει σταθμό βάσης (δηλαδή, κεραίες που μπορούν να συνδέσουν UE στο δίκτυο), για παράδειγμα κατά την περιαγωγή (roaming). Επομένως, η αρχιτεκτονική έχει μια τρίτη οντότητα. 3) Τα δίκτυα εξυπηρέτησης (Serving Networks – SNs) στα οποία μπορούν να συνδέονται οι UE. Ένας SN παρέχει υπηρεσίες (π.χ. κλήση ή SMS) όταν τόσο το UE όσο και το SN έχουν ταυτοποιηθεί μεταξύ τους και έχουν δημιουργήσει ένα ασφαλές κανάλι με τη βοήθεια του HN του συνδρομητή. Στο ακόλουθο σχήμα δίνεται αυτή η γενική αρχιτεκτονική[9].



Εικόνα 5: Γενική αρχιτεκτονική κυψελοειδούς δικτύου [9]

Συνοπτικά, ο συνδρομητής χρησιμοποιεί το τηλέφωνό του (UE), εξοπλισμένο με USIM, για να επικοινωνεί με έναν σταθμό βάσης που λειτουργεί από το SN μέσω ενός μη ασφαλούς ασύρματου καναλιού. Ο SN επικοινωνεί με τον πάροχο του συνδρομητή (HN) στα δεξιά μέσω ενός πιστοποιημένου (ενσύρματου) καναλιού.

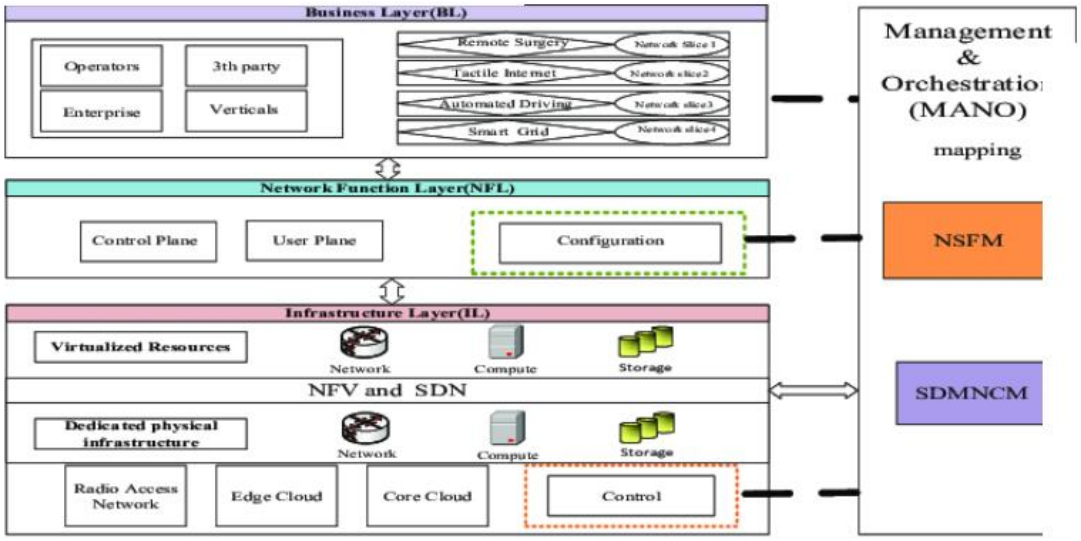
Η βασική αρχιτεκτονική E2E για τα δίκτυα 5G, δηλαδή η αρχιτεκτονική από άκρο σε άκρο (end-to-end), αποτυπώνεται όπως αναπτύχθηκε από τη Huawei στο ακόλουθο σχήμα.



Εικόνα 6: Αρχιτεκτονική από άκρο σε άκρο (E2E) σε δίκτυο 5G [4]

Τα δίκτυα 5G είναι ετερογενή, χρησιμοποιούν διαφορετικές ζώνες συχνοτήτων, έχουν διαφορετικά μεγέθη κυψελών και διαφορετικές τεχνολογίες ραδιοπρόσβασης (RAT).

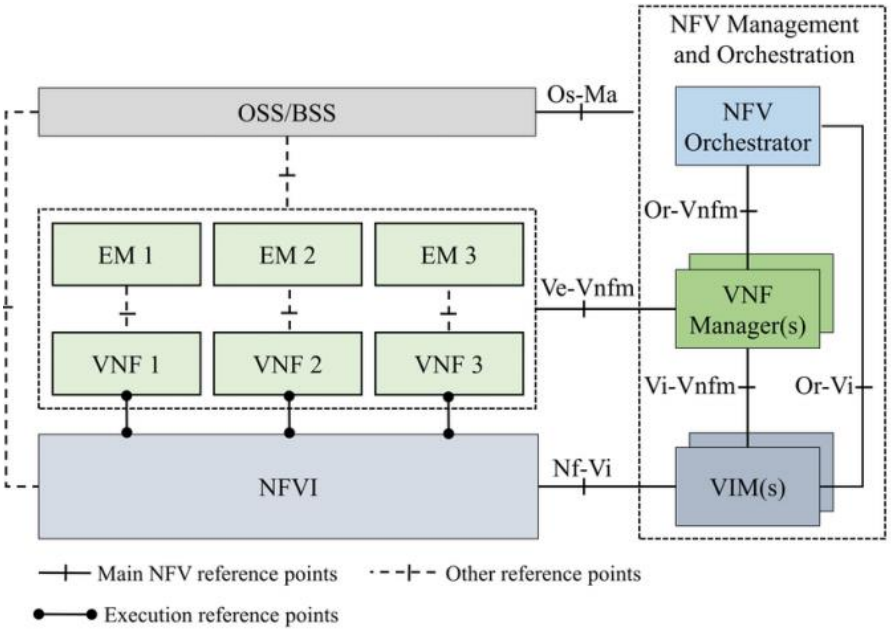
Για να καλυφτούν οι ανάγκες των δικτύων 5G σημαντική συμβολή παρέχει ο τεμαχισμός των δικτύων. Ένα τεμάχιο (slice) δικτύου μπορεί να διαχωριστεί σε δύο κατηγορίες. Η μια κατηγορία περιλαμβάνει το δίκτυο ραδιοπρόσβασης (RAN) στο υποδίκτυο του τεμαχίου του δικτύου (network slice subnet instance - NSSI) και η δεύτερη κατηγορία περιλαμβάνει τον πυρήνα του δικτύου (CN) στο NSSI. Ένα τεμαχισμένο δίκτυο μπορεί να παρέχει διάφορες υπηρεσίες ταυτόχρονα, και το δίκτυο και η υπηρεσία τεμαχισμού δικτύου είναι ανεξάρτητα. Το παρακάτω σχήμα δείχνει το διάγραμμα της 3GPP για τον τεμαχισμό του δικτύου [4].



Εικόνα 7: Αρχιτεκτονική τεμαχισμού δικτύου [4]

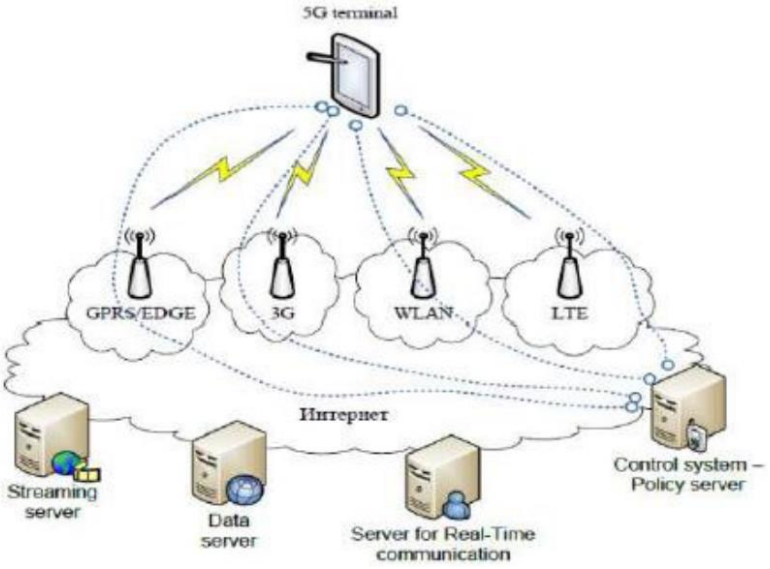
Ουσιαστικά, τα δύο σχήματα που αναφέρθηκαν για την προσέγγιση της αρχιτεκτονικής του δικτύου 5G έχουν τα ίδια επίπεδα αλλά επικεντρώνονται σε διαφορετικά χαρακτηριστικά.

Η διαχείριση και η ενορχήστρωση (Management and Orchestration - MANO) αποτελούν σημαντική συνιστώσα της αρχιτεκτονικής εικονικοποίησης λειτουργιών δικτύου (network functions virtualization - NFV), όπως επισημαίνει το Ευρωπαϊκό Ινστιτούτου Προτύπων Τηλεπικοινωνιών (European Telecommunications Standards Institute – ETSI). Η διαχείριση και η ενορχήστρωση (MANO) συντονίζει τους πόρους του δικτύου τη διαχείριση του κύκλου ζωής των λειτουργιών εικονικού δικτύου (VNFs), καθώς και για τις εφαρμογές που βασίζονται σε νέφη και υπηρεσίες δικτύου. Στο ακόλουθο σχήμα εμφανίζεται το πλαίσιο ETSI. Ωστόσο, αξίζει να αναφερθεί ότι πολλοί οργανισμοί ανοιχτού κώδικα έχουν αναπτύξει τα δικά τους πλαίσια NFV MANO [4].



Εικόνα 8: Αρχιτεκτονική ESTI NFV [4]

Ολοκληρώνοντας, στο ακόλουθο σχήμα απεικονίζεται το μοντέλο συστήματος που ενσωματώνει το σχεδιασμό της υποδομής δικτύου από τις προηγούμενες γενιές για τα δίκτυα 5G. Πρόκειται για ένα μοντέλο all-IP για διαλειτουργικότητα ασύρματου και κυψελοειδούς δικτύου. Η αρχιτεκτονική περιέχει έναν τερματικό υπολογιστή (ο οποίος διαδραματίζει καθοριστικό ρόλο στην τρέχουσα δομή) και μια σειρά από ανεξάρτητες τεχνολογίες ραδιοσυστήματος. Κάθε τεχνολογία ραδιοεπικοινωνίας είναι αισθητή σε κάθε τερματικό λόγω μιας υπερσύνδεσης IP με το εξωτερικό περιβάλλον του διαδικτύου. Ωστόσο, μέσω της τεχνολογίας ραδιοπρόσβασης (RAT) στο εσωτερικό του κινητού τερματικού πρέπει να υπάρχει ξεχωριστή διασύνδεση δικτύου. Για παράδειγμα, εάν υπάρχει ανάγκη επίτευξης τεσσάρων ξεχωριστών RAT, πρέπει να παρέχουμε τέσσερις διαφορετικές προσεγγίσεις σε παρόμοιες διεπαφές εντός του κινητού τερματικού και να ενεργοποιήσουμε όλες αυτές τις διεπαφές ταυτόχρονα, έτσι ώστε η αρχιτεκτονική να λειτουργεί σωστά [4].



Εικόνα 9: Λειτουργική αρχιτεκτονική [4]

Τόσο οι τεχνολογίες που χρησιμοποιούνται στα δίκτυα 5G όσο και η αρχιτεκτονική που προτείνεται υπόσχονται ένα δίκτυο με μεγάλο εύρος ζώνης, πολύ υψηλούς ρυθμούς ταχύτητας και πολύ μικρή καθυστέρηση στη μετάδοση των δεδομένων. Ωστόσο, όλες αυτές οι νέες τεχνολογίες που σχετίζονται με τα δίκτυα 5G δημιουργούν πολλές προκλήσεις σχετικά με την ασφάλεια και τον έλεγχο της ταυτότητας. Όλα αυτά τα ζητήματα και οι προκλήσεις ασφαλείας προσεγγίζονται στο επόμενο κεφάλαιο.

4 *Ασφάλεια στα δίκτυα 5G*

Το 5G αποτελεί ένα νέο σύστημα δικτύου κινητής που συνδέει σχεδόν όλες τις πτυχές της κοινωνίας, δηλαδή οχήματα, οικιακές συσκευές, υγειονομική περίθαλψη, βιομηχανία, επιχειρήσεις και πολλά άλλα στο δίκτυο. Ωστόσο, αυτή η εξέλιξη, εισαγάγει μια νέα σειρά απειλών και τρωτών σημείων ασφαλείας που αποτελούν σημαντική πρόκληση για τα δίκτυα. Για παράδειγμα, στο 5G μπορεί να συνδεθεί ένα δίκτυο τροφοδοσίας ηλεκτρικού ρεύματος. Η παραβίαση ενός τέτοιου κρίσιμου συστήματος μπορεί να έχει μεγάλες καταστροφικές συνέπειες τόσο για την ίδια την υποδομή, αλλά και για την κοινωνία.

Οι πόροι και τα δεδομένα σε ένα σύστημα πρέπει να διαφυλαχτούν και προστατευτούν σύμφωνα με τρείς θεμελιώδεις ιδιότητες της ασφάλειας των πληροφοριών. Αυτές οι ιδιότητες περιλαμβάνουν την εμπιστευτικότητα (confidentiality), την ακεραιότητα (integrity) και την διαθεσιμότητα (availability). Αναλυτικότερα, η έννοια της εμπιστευτικότητας σχετίζεται με την προστασία των πληροφοριών από μη εξουσιοδοτημένη αποκάλυψη (αναγνώριση) τους. Η ακεραιότητα σχετίζεται με την προστασία από την μεταβολή των πληροφοριών, δηλαδή την τροποποίηση ή την διαγραφή των πληροφοριών χωρίς εξουσιοδότηση. Τέλος, η διαθεσιμότητα σχετίζεται με τη διασφάλιση ότι η εξουσιοδοτημένη πρόσβαση στις πληροφορίες (για ανάγνωση ή μεταβολή) θα είναι πάντα διαθέσιμη χωρίς να υπάρχουν εμπόδιο ή καθυστερήσεις [46].

Επίσης, για να επιτευχθεί η ασφάλεια σε ένα συστήματα πληροφορικής θα πρέπει να εφαρμοστούν επιτυχημένα τρεις συγκεκριμένοι μηχανισμοί, η αναγνώριση (identification), η αυθεντικοποίηση ή αλλιώς ο έλεγχος ταυτότητας (authentication) και η εξουσιοδότηση (authorization). Αναλυτικότερα, ο μηχανισμός της αναγνώρισης σχετίζεται με τη διαδικασία κατά την οποία μια οντότητα (π.χ. ένας πελάτης) παρουσιάζει την ταυτότητα του στο

σύστημα (π.χ. ένας εξυπηρετητής). Η αυθεντικοποίηση ή αλλιώς ο έλεγχος ταυτότητας σχετίζεται με τη διαδικασία κατά την οποία επιβεβαιώνεται η ταυτότητα που παρουσίαση η οντότητα στο σύστημα. Ο μηχανισμός εξουσιοδότησης σχετίζεται με την διαδικασία κατά την οποία λαμβάνεται από απόφαση σχετικά με το αν γίνεται αποδεκτό ή απορρίπτεται ένα αίτημα πρόσβασης μιας αυθεντικόποιημένης οντότητας με βάση τα δικαιώματα πρόσβασης που έχει ήδη εκχωρήσει και την πολιτική του συστήματος για τον έλεγχο πρόσβασης [46].

Σε αυτό το κεφάλαιο προσεγγίζονται οι σημαντικότερες προκλήσεις ασφαλείας και οι παραδοσιακές λύσεις ασφαλείας που μπορούν να εφαρμοστούν ή εφαρμόζονται για την αντιμετώπιση των προκλήσεων ασφαλείας .

4.1 Τύποι επιθέσεων ασφαλείας

Σε αυτή την ενότητα γίνεται εισαγωγή στους πιο κοινούς τύπους επιθέσεων που μπορεί να συμβούν σε δίκτυα 5G και γίνεται συχνή αναφορά σε αυτούς κατά την προσέγγιση και περιγραφή των προκλήσεων ασφαλείας. Οι επιθέσεις διακρίνονται σε παθητικές και ενεργές.

4.1.1 Παθητικές επιθέσεις

Οι παθητικές επιθέσεις είναι αυτές κατά τις οποίες ο επιτιθέμενος στοχεύει να λάβει πληροφορίες, αλλά δεν επιθυμεί να τροποποιήσει το περιεχόμενο των δεδομένων. Έτσι, είναι πολύ δύσκολο να εντοπιστεί, αφού δεν αλλοιώνει τα δεδομένα. Η απελευθέρωση μηνυμάτων, το sniffing, οι καταγραφείς πληκτρολόγησης (key loggers) είναι μερικές από τις πιο συνηθισμένες τεχνικές παθητικών επιθέσεων [47].

4.1.1.1 Eavesdropping (υποκλοπές)

Οι επιθέσεις υποκλοπής eaves-dropping (λαθροακρόασης, δηλαδή κρυφή ή λαθραία ακρόαση ιδιωτικών δεδομένων) αποτελούν μια από τις πιο κοινές παραβιάσεις της ιδιωτικής ζωής. Ουσιαστικά, ο εισβολέας παρεμβαίνει στα δεδομένα των νόμιμων χρηστών και μπορεί, απλώς, να μάθει το περιεχόμενο της επικοινωνίας [48]. Είναι ένας τύπος επίθεσης που γίνεται χωρίς την άδεια ή τη γνώση των νόμιμων χρηστών. Παραβιάζει τους κανόνες εμπιστευτικότητας και διακρίνεται σε τέσσερις τύπους[47]:

4.1.1.1.1 Απελευθέρωση μηνύματος (Release of message)

Όταν ένας χρήστης στέλνει ένα μήνυμα σε έναν άλλο χρήστη (π.χ. σε ένα φίλο ή ένα συνεργάτη) επιθυμεί μόνο αυτό το άτομο να μπορεί να διαβάσει αυτό το μήνυμα. Με τη

χρήση συγκεκριμένου μηχανισμού ασφαλείας, ο χρήστης, μπορεί να αποτρέψει την απελευθέρωση του περιεχομένου του μηνύματος. Για παράδειγμα, μπορεί να κωδικοποιήσει το μήνυμα χρησιμοποιώντας έναν αλγόριθμο. Ουσιαστικά, μέσω μιας παθητικής επίθεσης απελευθέρωσης μηνύματος, ο επιτιθέμενος παρακολουθεί τα μηνύματα [47].

4.1.1.1.2 Ανάλυση κυκλοφορίας (Traffic analysis)

Τα μηνύματα που μεταφέρονται μέσω του δικτύου είναι ευάλωτα ακόμη και αν είναι κρυπτογραφημένα [48]. Στον τύπο επιθέσεων με ανάλυση κυκλοφορίας, ο επιτιθέμενος δεν στοχεύει στα πραγματικά δεδομένα, αλλά παρακολουθεί την επικοινωνία του δικτύου για να εκτελέσει ανάλυση κίνησης, ώστε να προσδιορίσει την θέση των κόμβων κλειδιών (key nodes), τη δομή της δρομολόγησης και τα πρότυπα συμπεριφοράς. Για παράδειγμα, ο επιτιθέμενος μπορεί να πραγματοποιεί ανάλυση της κυκλοφορίας για να εντοπίσει ένα σταθμό βάσης. Μόλις εντοπίσει αυτό σταθμό βάσης, μπορεί να ξεκινήσει με ακρίβεια μια σειρά από ενεργές επιθέσεις [49].

4.1.1.1.3 Sniffing

Το sniffing αποτελεί μια μέθοδο μέσω της οποίας χρησιμοποιείται κάποιο πρόγραμμα ή μηχάνημα για την υποκλοπή της κίνησης των δεδομένων που μεταφέρονται από ένα δίκτυο. Ουσιαστικά, με αυτή τη μέθοδο ο επιτιθέμενος προσπαθεί να μάθει τι είδους μηνύματα ή δεδομένα μεταφέρονται από τον αποστολέα [47].

4.1.1.1.4 Keyloggers (καταγραφείς πληκτρολόγησης)

Το keylogger αποτελεί ένα πρόγραμμα που εκτελείται στο παρασκήνιο, καταγράφοντας όλες τις πληκτρολογήσεις. Μόλις καταγραφούν οι πληκτρολογήσεις, αποθηκεύονται κρυμμένες στον υπολογιστή για να ανακτηθούν αργότερα από τον επιτιθέμενο. Στη συνέχεια, ο επιτιθέμενος μελετά προσεκτικά τις πληκτρολογήσεις προσπαθώντας να βρει κωδικούς πρόσβασης ή άλλες πιθανές χρήσιμες πληροφορίες [47].

4.1.2 Ενεργές επιθέσεις

Ως ενεργές χαρακτηρίζονται οι επιθέσεις που κάνουν κάποια τροποποίηση στο αρχικό μήνυμα ή δημιουργούν κάποιο ψευδή μήνυμα [47]. Ένας εισβολέας μπορεί να επαναλάβει παλιές ροές δεδομένων, αλλάζοντας τα μηνύματα επικοινωνίας ή να αφαιρέσει κάποιο επιλεγμένο μέρος σημαντικών μηνυμάτων επικοινωνίας [48]. Αυτές οι επιθέσεις είναι πολύ περίπλοκες και δεν μπορούν να αποτραπούν εύκολα. Μπορούν να κατηγοριοποιηθούν σε

τρεις τύπους: διακοπή (interruption), κατασκευή (fabrication) και τροποποίηση (modification) [47].

4.1.2.1 Διακοπή (interruption)

Στις επιθέσεις διακοπής μια εξουσιοδοτημένη οντότητα προσποιείται ότι είναι μια άλλη οντότητα. Για παράδειγμα, υπάρχουν τρεις χρήστες Α, Β και Γ. Ο χρήστης Α μπορεί να εμφανίζεται ως χρήστης Γ και να στείλει ένα μήνυμα στον χρήστη Β. Ο χρήστης Β πιστεύει ότι το μήνυμα προήλθε από τον χρήστη Γ. Ουσιαστικά, ο χρήστης Α διακόπτει την επικοινωνία μεταξύ του χρήστη Β και Γ. Η διακοπή θέτει σε κίνδυνο τη διαθεσιμότητα των πόρων, καθώς χρησιμοποιούνται πόροι του συστήματος χωρίς να απαιτείται [47]. Από τις πιο κοινές επιθέσεις διακοπής αποτελεί η άρνηση υπηρεσίας, η κατανεμημένη άρνηση υπηρεσίας και οι εγχύσεις SQL.

4.1.2.1.1 Άρνηση υπηρεσίας (Denial of Service - DoS)

Οι επιθέσεις άρνησης υπηρεσίας (DoS) έχουν ως στόχο να καταστρέψουν ένα διακομιστή (server). Για να επιτευχθεί αυτό θέτουν επιπλέον κίνηση από το μέγιστο όριο επεξεργασίας του διακομιστή [48]. Καθώς, όταν ένα σύστημα λαμβάνει αιτήματα επικοινωνίας, είναι απασχολημένο προσπαθώντας να δημιουργήσει μια διαδρομή επικοινωνίας επιστροφής με τον εκκινητή (δηλαδή, αυτόν που ξεκινάει την επικοινωνία, ή αλλιώς αυτών που κάνει αίτημα επικοινωνίας με ένα σύστημα). Ο εκκινητής μπορεί να χρησιμοποιεί μια έγκυρη διεύθυνση IP ή μια μη έγκυρη. Έτσι, το σύστημα παραμένει σε κατάσταση αναμονής, καθώς δεν μπορεί να εξυπηρετήσει του νόμιμους χρήστες [47]. Με πιο απλά λόγια, μέσω των επιθέσεων DoS στέλνονται από τον παράνομο χρήστη υπερβολικά αιτήματα που έχουν ως αποτέλεσμα την υπερφόρτωση του δικτύου. Έτσι, το σύστημα σταματάει να απαντάει σε ορισμένα ή και σε όλα τα νόμιμα αιτήματα [48].

4.1.2.1.2 Κατανεμημένη Άρνηση Υπηρεσιών (Distributed Denial of Services - DDoS)

Οι επιθέσεις κατανεμημένης άρνησης υπηρεσίας (DDoS) πραγματοποιούνται σε δίκτυα υπολογιστών που συνδέονται στο διαδίκτυο. Αυτά τα δίκτυα αποτελούνται από υπολογιστές και άλλες συσκευές (όπως συσκευές IoT) που έχουν μολυνθεί από κακόβουλο λογισμικό και επιτρέπουν τον απομακρυσμένο έλεγχο από τον εισβολέα [50]. Αυτοί οι παραβιασμένοι υπολογιστές ή οι παραβιασμένες συσκευές στέλνουν πολλαπλά αιτήματα σε ένα στόχο με αποτέλεσμα την άρνηση της υπηρεσίες σε νόμιμους χρήστες [47]. Η διαφορά με τις απλές

DoS επιθέσεις είναι ότι τα παράνομα μηνύματα προέρχονται από πολλές και διαφορετικές πηγές. Έτσι, είναι δύσκολο να σταματήσει μια επίθεση μπλοκάροντας μια μόνο πηγή.

4.1.2.1.3 ΕγχύσειςSQL (SQL Injections)

Η έγχυση SQL είναι ένα τύπος επίθεσης κατά την οποία ο εισβολέας προσθέτει κώδικα «Δομημένης Γλώσσας Ερωτήματος» (“Structural Query Language” - SQL) σε ένα πλαίσιο (ιστού) εισαγωγής δεδομένων για να αποκτήσει πρόσβαση ή να κάνει αλλαγές στα δεδομένα. Η έγχυσηSQL επιτρέπει σε έναν εισβολέα να εκτελεί εντολές απευθείας στη βάση δεδομένων μιας εφαρμογής ιστού και να καταστρέψει τη λειτουργικότητα ή την εμπιστευτικότητα της [51]. Πρόκειται για ένα μηχανισμό επίθεσης στο διαδίκτυο στον οποίο μια κακόβουλη δήλωση SQL εισάγεται μέσω του τομέα δεδομένων εισόδου από τον πελάτη (client) στην εφαρμογή των δεδομένων [52]. Ουσιαστικά, είναι ένας τύπος επίθεσης, όπου ο επιτιθέμενος χρησιμοποιεί κακόβουλο κώδικα SQL για να χειριστή μια βάση δεδομένων (back-end) και να έχει πρόσβαση σε πληροφορίες που δεν προοριζόταν να εμφανιστούν. Αυτές οι πληροφορίες μπορεί να περιλαμβάνουν οποιονδήποτε αριθμό στοιχείων, όπως ευαίσθητα εταιρικά δεδομένα, λίστες χρηστών ή ιδιωτικά στοιχεία πελατών.

4.1.2.2 Κατασκευή (Fabrication)

Στις επιθέσεις κατασκευής οι κακόβουλοι χρήστες χρησιμοποιούν κάποια υπηρεσία πρόσβασης, για την οποία δεν πληρούν τις προϋποθέσεις. Οι επιθέσεις αυτές μπορούν να επιτευχθούν όταν απουσιάζουν οι κατάλληλοι μηχανισμοί ελέγχου ταυτότητας [47].

4.1.2.2.1 Επίθεση επανάληψης (replayattack)

Η επίθεση επανάληψης είναι μια μορφή ενεργής επίθεσης στο δίκτυο στην οποία μια έγκυρη μετάδοση δεδομένων επαναλαμβάνεται ή καθυστερεί κακόβουλα. Ένας εισβολέας συλλαμβάνει τα εξουσιοδοτημένα δεδομένα και τα στέλνει ξανά για προσωπική του χρήση. Για παράδειγμα, ο χρήστης Α θέλει να μεταφέρει κάποιο ποσό στον τραπεζικό λογαριασμό του χρήστη Γ. Και οι δύο χρήστες Α και Γ έχουν λογαριασμό στην τράπεζα Β. Ο χρήστης Α στέλνει ένα ηλεκτρονικό μήνυμα στην τράπεζα Β, ζητώντας μεταφορά χρημάτων. Ο χρήστης Γ μπορούσε να συλλάβει αυτό το μήνυμα και να στείλει ένα δεύτερο αντίγραφο στην τράπεζα Β, αλλά η τράπεζα Β δεν μπορούσε να έχει ιδέα ότι πρόκειται για μη εξουσιοδοτημένο μήνυμα. Έτσι, ο χρήστης Γ θα επωφεληθεί από τη μεταφορά χρημάτων δύο φορές. Μια επίθεση επανάληψης, μπορεί να αποτραπεί χρησιμοποιώντας ισχυρές ψηφιακές υπογραφές

που περιλαμβάνουν χρονικά γραμματόσημα και συμπεριλαμβάνουν μοναδικές πληροφορίες από την προηγούμενη συναλλαγή [47].

4.1.2.2.2 Μεταμφίεση (masquerading)

Η μεταμφίεση είναι ένας τύπος επίθεσης στην οποία ένα σύστημα αναλαμβάνει την ταυτότητα ενός άλλου. Είναι μια τεχνική που χρησιμοποιείται από τον εισβολέα για να προσποιείται τον εαυτό του ως εξουσιοδοτημένο άτομο, προκειμένου να αποκτήσεις πρόσβαση σε εμπιστευτικές πληροφορίες με παράνομο τρόπο [47]. Ουσιαστικά, αυτή η επίθεση αναφέρεται στη χρήση πλαστής ταυτότητας για την απόκτηση παράνομης πρόσβασης σε οποιονδήποτε προσωπικό υπολογιστή. Ένας εισβολέας ενεργεί ως μη εξουσιοδοτημένο σύστημα για να αποκτήσει πρόσβαση σε αυτό ή για να αυξήσει μεγαλύτερα δικαιώματα από αυτά που έχουν εγκριθεί. Η επίθεση με μεταμφίεση συνήθως περιλαμβάνει έναν άλλο τύπο ενεργής επίθεσης [48].

4.1.2.3 Τροποποίηση (modification)

Οι επιθέσεις τροποποίησης περιλαμβάνουν παραβίαση των στοιχείων. Τέτοιες επιθέσεις προκαλούν απώλεια ακεραιότητας [47][53], αλλά θα μπορούσαν επίσης να αντιπροσωπεύουν και απώλεια διαθεσιμότητας. Εάν ένας χρήστης έχει πρόσβαση σε ένα αρχείο με μη εξουσιοδοτημένο τρόπο και αλλάξει τα δεδομένα που περιέχει, θα επηρεάσει την ακεραιότητα των δεδομένων που περιέχονται στο αρχείο. Ωστόσο, εάν πρόκειται για ένα αρχείο διαμόρφωσης που διαχειρίζεται τον τρόπο συμπεριφοράς μιας συγκεκριμένης υπηρεσίας, ενδέχεται να επηρεαστεί η διαθεσιμότητα αυτής της υπηρεσίας αλλάζοντας τα περιεχόμενα του αρχείου [53].

4.1.2.3.1 Man in the middle

Η επίθεση man-in-the-middle αποτελεί μια ενεργή διαδικτυακή επίθεση κατά την οποία ο επιτιθέμενος παραβιάζει τη νόμιμη επικοινωνία ανάμεσα σε δυο φιλικά μέρη. Ο επιτιθέμενος επιχειρεί να υποκλέψει, να διαβάσει και να τροποποιήσει τις πληροφορίες που μεταδίδονται μεταξύ των δύο μερών, για παράδειγμα του χρήστη ενός δημόσιου δικτύου και οποιουδήποτε ιστότοπου που ζητήθηκε ή του χρήστη Α και του χρήστη Β. Ο εισβολέας χρησιμοποιεί τις πληροφορίες που συλλέγονται παράνομα για κλοπή ταυτότητας και άλλους τύπους απάτης [47].

4.2 Προκλήσεις ασφαλείας στις νέες τεχνολογίες

Τα δίκτυα 5G είναι πολύπλοκα και περιλαμβάνουν, όπως παρουσιάστηκε σε προηγούμενο κεφάλαιο, πλήθος διαφορετικών τεχνολογιών. Για το λόγο αυτό επιλέχθηκε η προσέγγιση των προκλήσεων ασφαλείας να ταξινομηθεί βάσει των νέων τεχνολογιών. Καθώς σε κάθε μια από αυτές τις τεχνολογίες εντοπίζονται διαφορετικές προκλήσεις ασφαλείας.

4.2.1 Προκλήσεις ασφαλείας στα Νέφη Κινητής (Mobile Clouds)

Λαμβάνοντας υπόψη ότι τα συστήματα υπολογιστικού νέφους περιλαμβάνουν διάφορους πόρους που μοιράζονται μεταξύ των χρηστών, είναι πιθανό ένας χρήστης να διαδίδει κακόβουλη κίνηση δεδομένων για να μειώσει την απόδοση ολόκληρου του συστήματος, να καταναλώσει περισσότερους πόρους ή να αποκτήσει κρυφά πρόσβαση σε άλλους χρήστες. Έτσι, καθώς, τα υπολογιστικά νέφη κινητής (Mobile Cloud Computing - MCC) μεταφέρουν τις έννοιες από τα υπολογιστικά νέφη στα συστήματα δικτύου 5G, δημιουργείται μια σειρά από ευπάθειες ασφαλείας, οι οποίες προκύπτουν κυρίως από την αρχιτεκτονική και την υποδομή [5].

Οι απειλές ασφαλείας για τα υπολογιστικά νέφη κινητής μπορούν να κατηγοριοποιηθούν σε απειλές διεπαφής (front-end) νέφους, υποστήριξης (back-end) νέφους και βάσης δικτύου. Στη συνέχεια, περιγράφονται οι απειλές ασφαλείας για νέφη κινητής με βάση αυτή την κατηγοριοποίηση.

Η διεπαφή (front-end) στα υπολογιστικά νέφη κινητής, είναι η πλατφόρμα πελάτη που αποτελείται από το κινητό τερματικό στο οποίο εκτελούνται εφαρμογές και διεπαφές που απαιτούνται για την πρόσβαση στις εγκαταστάσεις του νέφους. Σε αυτό το τμήμα, μια απειλή μπορεί να είναι στο φυσικό επίπεδο, όπου η πραγματική κινητή συσκευή και άλλα ενσωματωμένα στοιχεία υλικού είναι οι πρωταρχικοί στόχοι Επίσης, μπορεί να είναι μια απειλή που βασίζεται στις εφαρμογές, όπου κακόβουλο λογισμικό, λογισμικό υποκλοπής spyware κ.α. χρησιμοποιούνται από εχθρούς για τη διακοπή των εφαρμογών του χρήστη ή τη συλλογή ευαίσθητων πληροφοριών του χρήστη [54][55].

Η πλατφόρμα υποστήριξης (back-end) αποτελείται από διακομιστές νέφους, συστήματα αποθήκευσης δεδομένων, εικονικές μηχανές, επόπτες (hypervisors) και πρωτόκολλα που απαιτούνται για να προσφέρουν υπηρεσίες νέφους. Επισημαίνεται ότι επόπτης είναι ένα πρόγραμμα που χρησιμοποιείται για την εκτέλεση και τη διαχείριση ενός ή περισσότερων εικονικών μηχανών σε έναν υπολογιστή. Σε αυτήν την πλατφόρμα υποστήριξης, οι απειλές

ασφαλείας στοχεύουν κυρίως στους διακομιστές νέφους για κινητά [5]. Η τυπική επίθεση ασφαλείας σε αυτό το τοπίο περιλαμβάνει επίθεση HTTP και XML DoS. Η άρνηση εξυπηρέτησης HTTP και XML (HX-Dos) συνδυάζει μηνύματα HTTP και XML για την κατάκλιση της χωρητικότητας των υποδομών νέφους. Μια τέτοια επίθεση θα μπορούσε να ξεκινήσει σε υποδομές νέφους, λογισμικό ή πλατφόρμες. Άλλες απειλές σε αυτό το τοπίο είναι η αργή αυξανόμενη πολυμορφική στρατηγική επίθεσης DDoS (Slowly Increasing Polymorphic DDoS Attack Strategy - SIPDAS). Το SIPDAS είναι ένα είδος επίθεσης DoS που αποφεύγει τους αλγόριθμους ανίχνευσης προτύπων τροποποιώντας τη συμπεριφορά του δυναμικά [36].

Το Δίκτυο Ραδιοπρόσβασης Νέφους (Cloud Radio Access Network - C-RAN) είναι ένας άλλος βασικός τομέας ενδιαφέροντος για την ανάλυση των προκλήσεων ασφαλείας για τα νέφη κινητής των δικτύων 5G. Το C-RAN έχει τη δυνατότητα να καλύψει τις ανάγκες ανάπτυξης της βιομηχανίας για υψηλότερη κινητικότητα (mobility) σε συστήματα κινητής επικοινωνίας 5G. Ωστόσο, το C-RAN είναι επιρρεπές σε προκλήσεις ασφαλείας που σχετίζονται με τα εικονικά συστήματα και τη τεχνολογία υπολογιστικού νέφους. Για παράδειγμα, η κεντρική αρχιτεκτονική του C-RAN υφίσταται την απειλή ενός μοναδικού σημείου αποτυχίας (single-point of failure) [56], δηλαδή ο κίνδυνος δημιουργείται από κάποιο πιθανό ελάττωμα στο σχεδιασμό, την εφαρμογή ή τη διαμόρφωση του συστήματος, στο οποίο ένα σφάλμα ή μια δυσλειτουργία προκαλεί τη διακοπή λειτουργίας ενός ολόκληρου συστήματος [57]. Επίσης, απειλές όπως επιθέσεις εισβολής, όπου οι αντίπαλοι εισέρχονται στο εικονικό περιβάλλον για την παρακολούθηση, την τροποποίηση ή την εκτέλεση ρουτινών λογισμικού στην πλατφόρμα που δεν ανιχνεύονται [56].

Επιθέσεις εντοπίζονται, επίσης, και στη συνολική πλατφόρμα νέφους, οι οποίες καλούνται insider επιθέσεις. Ο όρος insiders σχετίζεται με την πρόσβαση του προσωπικού του παρόχου υπηρεσιών στους φυσικούς διακομιστές στους οποίους αποθηκεύονται τα δεδομένα χρήστη. Δηλαδή, όταν το προσωπικό, που θεωρείται έμπιστο, σκοπεύει να κάνει κατάχρηση ή κακή διαχείριση δεδομένων και πληροφοριών των χρηστών, αυτό, θα μπορούσε να αποτελέσει σημαντική απειλή για ολόκληρη την ιδέα του νέφους. Ένα κλασικό παράδειγμα αποτελεί η περίπτωση του Facebook Cambridge Analytica, όπου οι προσωπικές πληροφορίες σχεδόν 87 εκατομμυρίων χρηστών του Facebook χρησιμοποιήθηκαν για πολιτικά κίνητρα [36].

4.2.2 Προκλήσεις ασφαλείας σε SDN

Η τεχνολογία SDN συγκεντρώνει τις πλατφόρμες ελέγχου δικτύου και επιτρέπει τον προγραμματισμό στα δίκτυα επικοινωνίας [58]. Ουσιαστικά, το SDN διαχωρίζει τις

λειτουργίες ελέγχου δικτύου από το επίπεδο προώθησης δεδομένων σε μια κεντρική πλατφόρμα ελέγχου [8]. Με απλά λόγια, το δίκτυο προβάλλεται μέσω διαφορετικών εφαρμογών SDN με τις οποίες μπορεί να γίνει και η διαχείριση του δικτύου. Ωστόσο, αυτές οι λειτουργίες, δημιουργούν ευκαιρίες για διάσπαση και παραβίαση του δικτύου [8][36], κάτι που έχει ήδη αποδειχθεί σχετικά με την υλοποίηση του OpenFlow του SDN [36]. Για παράδειγμα, ο κεντρικός έλεγχος (π.χ. ελεγκτής OpenFlow) θα είναι μια ευνοϊκή επιλογή για επιθέσεις DoS ή DDoS και η έκθεση των Διεπαφών Προγραμματισμού Εφαρμογών (Application Programming Interfaces - API) σε ανεπιθύμητο λογισμικό μπορεί να καταργήσει ολόκληρο το δίκτυο [58]. Πιο συγκεκριμένα, μέσω τον διεπαφών προγραμματισμού, η συμπεριφορά του επιπέδου προώθησης μπορεί να παρακολουθείται και να ελέγχεται εξ αποστάσεως και η ανάπτυξη νέων λειτουργιών δικτύωσης μπορεί να απλοποιηθεί. Ωστόσο, η συγκέντρωση του επιπέδου ελέγχου οδηγεί επίσης σε προκλήσεις, όπως, η ανθεκτικότητα, η ασφάλεια και η επεκτασιμότητα [8].

Επίσης, ο ελεγκτής SDN τροποποιεί τους κανόνες ροής στη διαδρομή δεδομένων, επομένως, μπορεί εύκολα να αναγνωριστεί η κυκλοφορία του ελεγκτή. Αυτό καθιστά τον ελεγκτή μια ορατή οντότητα στο δίκτυο και έτσι, αποτελεί μια επιλογή για επιθέσεις DoS. Ο κεντρικός έλεγχος του δικτύου μπορεί, επίσης, να κάνει τον ελεγκτή ένα σημείο συμφόρησης για ολόκληρο το δίκτυο λόγω των επιθέσεων κορεσμού [5]. Ακόμη, εξαιτίας της επεκτασιμότητας του ελεγκτή, εντοπίζεται μια περίπτωση επίθεσης κατά την οποία πακέτα IP με τυχαία πεδία κεφαλίδας (header filed) αποστέλλονται συνεχώς στον ελεγκτή καθιστώντας το μη διαθέσιμο σε νόμιμα αιτήματα ρύθμισης ροής (flow setup) [36].

4.2.3 Προκλήσεις ασφαλείας στα κανάλια επικοινωνίας

Τα δίκτυα 5G αποτελούν ένα σύνθετο οικοσύστημα που θα περιλαμβάνει drone και ελέγχους εναέριας κυκλοφορίας, εικονική πραγματικότητα που βασίζεται σε νέφη, συνδεδεμένα οχήματα, έξυπνα εργοστάσια, ρομπότ σε νέφη, μεταφορές και υγεία. Έτσι, οι εφαρμογές χρειάζονται ασφαλή συστήματα επικοινωνίας, τα οποία να μπορούν να υποστηρίξουν πιο συχνό έλεγχο ταυτότητας (authentication) και ανταλλαγή πιο ευαίσθητων δεδομένων [5].

Επίσης, στα δίκτυα της νέας γενιάς θα εμπλακούν πολλοί νέοι παράγοντες, όπως οι φορείς εκμετάλλευσης δικτύων κινητής τηλεφωνίας (Mobile Network Operators - MNO) και οι φορείς εκμετάλλευσης νέφους. Σε ένα τέτοιο οικοσύστημα απαιτούνται αρκετά επίπεδα ελέγχου ταυτότητας (authentications) τόσο σε επίπεδο πρόσβασης δικτύου όσο και σε επίπεδο υπηρεσίας και απαιτείται συχνός έλεγχος ταυτότητας μεταξύ των φορέων[5].

Πριν από τα δίκτυα 5G, τα δίκτυα κινητής τηλεφωνίας είχαν ειδικά κανάλια επικοινωνίας με βάση σήραγγες (tunnels) GTP και IPsec. Οι διεπαφές επικοινωνίας, όπως οι X2, S1, S6, S7, οι οποίες χρησιμοποιούνται μόνο σε δίκτυα κινητής τηλεφωνίας, απαιτούν σημαντικό επίπεδο εμπειρογνωμοσύνης από κάποιον επιτιθέμενο, ώστε να αποκτήσει πρόσβαση σε αυτές τις διεπαφές. Ωστόσο, τα δίκτυα 5G που βασίζονται στην τεχνολογία SDN δεν θα έχουν τέτοιες αποκλειστικές διεπαφές αλλά μάλλον κοινές διεπαφές SDN, γεγονός που θα αυξήσει τις πιθανότητες εισβολής. Η επικοινωνία σε δίκτυα κινητής τηλεφωνίας 5G που βασίζεται σε SDN μπορεί να κατηγοριοποιηθεί σε τρία κανάλια επικοινωνίας, δηλαδή στο κανάλι δεδομένων, στο κανάλι ελέγχου και στο κανάλι μεταξύ ελεγκτών (όπως φαίνεται στο σχήμα (Εικόνα 2) που παρουσιάζεται η αρχιτεκτονική SDN). Στο τρέχον σύστημα SDN, αυτά τα κανάλια προστατεύονται χρησιμοποιώντας τις συνεδρίες TLS (Transport Layer Security)/SSL (Secure Sockets Layer), δηλαδή την ασφάλεια στο επίπεδο μεταφοράς και ασφάλεια στο επίπεδο υποδοχών. Ωστόσο, οι συνεδρίες TLS/SSL είναι ιδιαίτερα ευάλωτες σε επιθέσεις επιπέδων IP, σε επιθέσεις SDN Scanner και δεν διαθέτουν ισχυρούς μηχανισμούς ελέγχου ταυτότητας (authentication mechanisms) [5].

4.2.4 Προκλήσεις ασφαλείας σε NFV

Η τεχνολογία NFV διευκολύνει τη συνεργασία μεταξύ των παρόχων υποδομής, των φορέων πρόσβασης δικτύου και τον παρόχων υπηρεσιών. Επίσης, διευκολύνει την προσαρμοσμένη ανάπτυξη των πόρων του δικτύου. Πιο συγκεκριμένα, με την NFV, η πρόσβαση και η λειτουργικότητα του βασικού δικτύου αναπτύσσονται ως λογισμικό πάνω από την υποδομή υλικού που μπορεί να μοιραστεί μεταξύ διαφορετικών χειριστών δικτύου ή συγκεκριμένων εφαρμογών [8]. Ωστόσο, αν και η NFV αποτελεί μια πολύ σημαντική τεχνολογία για τα μελλοντικά δίκτυα κινητής τηλεφωνίας, έρχεται αντιμέτωπη με βασικές προκλήσεις ασφαλείας, όπως η εμπιστευτικότητα, η ακεραιότητα, η αυθεντικότητα και η μη απόρριψη [5].

Η ασφάλεια NFV βασίζεται στις δυνατότητες απομόνωσης του επιπέδου εικονικοποίησης για την αποφυγή παρεμβολών και διαρροών πληροφοριών μεταξύ του λογισμικού που εκτελείται σε διαφορετικές εικονικές μηχανές [8]. Μια από τις κύριες προκλήσεις στη χρήση της τεχνολογίας NFV σε δίκτυα κινητής τηλεφωνίας είναι ο δυναμικός χαρακτήρας των λειτουργιών εικονικού δικτύου (VNFs) που οδηγεί σε σφάλματα διαμόρφωσης (configuration errors) και συνεπώς σταματάει η ασφάλεια. Η δυναμικότητα σχετίζεται με την ικανότητα [59] των εικονικών μηχανών να δημιουργηθούν, να διαγραφούν και να μετακινηθούν εύκολα σε ένα δίκτυο. Αυτή η δυναμικότητα δημιουργεί περισσότερα προβλήματα ασφαλείας στα

συστήματα NFV, καθώς η παρακολούθηση μια κακόβουλης εικονικής μηχανής είναι πιο περίπλοκη. Σε κάθε περίπτωση, η κύρια πρόκληση των VNF είναι ότι ολόκληρο το δίκτυο μπορεί να μπει σε κίνδυνο αν το πρόγραμμα εποπτείας (hypervisor) δεχτεί επίθεση [8].

Ειδικότερα, προκλήσεις εντοπίζονται στον τρόπο λειτουργίας των εικονικών συστημάτων. Ένα εικονοποιημένο σύστημα δεν μπορεί να ασφαλιστεί, όπως ένα φυσικό σύστημα. Καθώς, πολλά εικονομοιημένα συστήματα μπορεί να εκτελούνται στο ίδιο στοιχείο του δικτύου, αλλά κάθε ένα από αυτά απαιτεί διαφορετικοί ασφάλεια [60]. Επομένως, δεν μπορούν να εφαρμοστούν οι ίδιες διαδικασίες ασφαλείας σε όλο το σύστημα. Στο [60] περιγράφεται ένα σενάριο, στο οποίο ένας διακομιστής που φιλοξενεί εικονικές μηχανές χωρίζεται σε πολλές ζώνες, καθεμία από τις οποίες έχει διαφορετικά επίπεδα ασφάλειας. Σε αυτήν την περίπτωση, μια ζώνη που πρέπει να συμμορφώνεται με συγκεκριμένο επίπεδο ασφάλειας δεν μπορεί να μετακινηθεί σε άλλο φυσικό διακομιστή, καθώς ο άλλος υπολογιστής ενδέχεται να μην προσφέρει την ίδια ασφάλεια.

Προκλήσεις εντοπίζονται επίσης στους επόπτες (hypervisors). Στην εικονικοποίηση του δικτύου, οι επόπτες χρησιμοποιούνται για τη χαρτογράφηση διαφόρων λειτουργιών δικτύου [61]. Ένας επόπτης μπορεί να δημιουργήσει και να εκτελέσει πολλαπλά λειτουργικά συστήματα επισκεπτών και ελέγχει τον απαραίτητο προγραμματισμό του επεξεργαστή (CPU) και την κατανομή μνήμης για αυτά τα συστήματα. Έτσι, ο επόπτης είναι η κύρια οντότητα σε ολόκληρο το εικονικοποιημένο σύστημα που βασίζεται στον επόπτη. Επομένως, εάν ένας επόπτης παραβιαστεί, τότε ολόκληρο το σύστημα μπορεί να τεθεί σε κίνδυνο [60]. Ο παραβιασμένος επόπτης μπορεί να στοχεύσει σε διάφορες επιθέσεις, όπως η εκμετάλλευση του λειτουργικού συστήματος του κεντρικού υπολογιστή για να καταστρέψει την απομόνωση ενός τεμαχισμένου δικτύου, η επίθεση DoS σε εικονικές μηχανές και επιθέσεις hopping σε εικονικές μηχανές [62].

4.2.5 Προκλήσεις σε παρόχους επικοινωνίας

Στα δίκτυα 5G υπάρχουν πολλοί πάροχοι επικοινωνίας που εμπλέκονται για την λειτουργία τους, όπως οι διαχειριστές των δικτύων κινητής τηλεφωνίας (Mobile Virtual Network Operators - MVNO), οι πάροχοι υπηρεσιών επικοινωνίας (Communication Service Providers - CSP) και οι πάροχοι υποδομής δικτύου (network infrastructure providers). Όλοι αυτοί οι πάροχοι υπηρεσιών έχουν διαφορετικές πολιτικές ασφάλειας και απορρήτου. Ο συγχρονισμός των αντίστοιχων πολιτικών απορρήτου μεταξύ αυτών των παραγόντων είναι μια πρόκληση στο δίκτυο 5G. Στις προηγούμενες γενιές, οι πάροχοι κινητής τηλεφωνίας είχαν άμεση πρόσβαση και έλεγχο όλων των στοιχείων του συστήματος. Ωστόσο, οι πάροχοι κινητής

τηλεφωνίας 5G χάνουν τον πλήρη έλεγχο των συστημάτων, καθώς βασίζονται σε νέους φορείς όπως τα CSP. Έτσι, οι φορείς εκμετάλλευσης 5G θα χάσουν την πλήρη διακυβέρνηση της ασφάλειας και της ιδιωτικότητας. Το απόρρητο των χρηστών και των δεδομένων αμφισβητείται σοβαρά σε κοινόχρηστα περιβάλλοντα, όπου η ίδια υποδομή μοιράζεται μεταξύ διαφόρων φορέων, για παράδειγμα VMNO και άλλων ανταγωνιστών. Επιπλέον, δεν υπάρχουν φυσικά όρια του δικτύου 5G, καθώς η αποθήκευση των δεδομένων και οι λειτουργίες NFV που χρησιμοποιούνται βασίζονται σε νέφος. Έτσι, οι χειριστές των δικτύων 5G δεν έχουν άμεσο έλεγχο της θέσης αποθήκευσης δεδομένων σε περιβάλλοντα νέφους. Ακόμη, καθώς διαφορετικές χώρες έχουν διαφορετικό επίπεδο μηχανισμών απορρήτου δεδομένων ανάλογα με το προτιμώμενο περιβάλλον τους, το απόρρητο αμφισβητείται εάν τα δεδομένα χρήστη αποθηκεύονται σε νέφος σε διαφορετική χώρα [5][36].

4.2.6 Προκλήσεις ασφαλείας σε massive MIMO

Το μαζικό MIMO αποτελεί μια από τις πιο σημαντικές τεχνολογίες που ενσωματώνονται στο φυσικό επίπεδο των δικτύων 5G. Σε ένα τεράστιο σύστημα MIMO, ένας σταθμός βάσης είναι συνήθως εξοπλισμένος με μεγάλο αριθμό στοιχείων κεραίας που ταυτόχρονα υποστηρίζουν μεγάλο αριθμό χρηστών. Οι ευπάθειες ασφαλείας ενός τεράστιου συστήματος MIMO χωρίζονται σε δύο κατηγορίες σε: 1) παθητικές επιθέσεις και 2) σε ενεργές επιθέσεις. Οι παθητικές επιθέσεις αναφέρονται σε υποκλοπές (eaves dropped) που γίνονται σε νόμιμες μεταδόσεις δεδομένων, δηλαδή ο εισβολέας προσπαθεί να υποκλέψει τα μεταδιδόμενα σήματα, χωρίς να μεταδίδει κανένα σήμα. Στις ενεργές επιθέσεις, ο εισβολέας μεταδίδει σήματα για να διακόψει ή να καταστρέψει μια μόνιμη μετάδοση. Οι ενεργές επιθέσεις μπορούν να χωριστούν περαιτέρω σε δύο κατηγορίες: επιθέσεις παρεμβολών (jamming attacks) και επιθέσεις πιλοτικής πλαστογράφησης (spoofing attacks). Ο στόχος μιας επίθεσης παρεμβολής είναι να διακόψει τη μετάδοση με την αποστολή μεγάλου όγκου δεδομένων προς το σταθμό βάσης ή τους χρήστες. Η πιλοτική πλαστογράφηση είναι μια έξυπνη μορφή ενεργής επίθεσης, όπου ο εισβολέας προσποιείται ότι είναι νόμιμος χρήστης [8][36].

4.3 Προκλήσεις ασφαλείας στο φυσικό επίπεδο

Πέρα από τις προκλήσεις ασφαλείας που εντοπίζονται στις διαφορετικές τεχνολογίες των δικτύων 5G, προκλήσεις προκύπτουν και σε επίπεδο αρχιτεκτονικής, όπως στο φυσικό επίπεδο σχετικά με τον έλεγχο ταυτότητας και το απόρρητο. Στις ακόλουθες υποενότητες παρουσιάζονται αναλυτικότερα.

4.3.1 Προκλήσεις ελέγχου ταυτότητας (authentication) στο φυσικό επίπεδο

Σε αυτή την ενότητα προσεγγίζονται οι προκλήσεις ελέγχου ταυτότητας στο φυσικό επίπεδο των δικτύων 5G. Ο έλεγχος ταυτότητας (authentication) είναι η διαδικασία κατά την οποία η ταυτότητα που απαιτείται επαληθεύεται [63].

Για να γίνει πιο κατανοητή η έννοια του ελέγχου ταυτότητας στο φυσικό επίπεδο γίνεται αναφορά στο συμβατικό μοντέλο της Alice και του Bob που χρησιμοποιείται ευρέως ως παράδειγμα στη βιβλιογραφία. Σε αυτό το παράδειγμα, η Alice είναι ο αποστολέας και ο Bob ο παραλήπτης και αποτελούν τις νόμιμες συσκευές, ενώ υπάρχει και ο αντίπαλος Eve, ο οποίος αποτελεί την παράνομη συσκευή. Η Alice και ο Bob επικοινωνούν μεταξύ τους με την παρουσία της Eve. Αυτός ο αντίπαλος (Eve) προσπαθεί να αποκτήσει νόμιμες πληροφορίες και να πλαστογραφήσει την Alice, ώστε να αποκτήσει παράνομα πλεονεκτήματα από τον Bob. Η Eve μπορεί επίσης να ανακτήσει τα μηνύματα. Ο κύριος στόχος του Bob είναι να αναγνωρίσει μοναδικά και ξεκάθαρα την Alice με τεχνικές ελέγχου ταυτότητας. Ουσιαστικά, η Alice υπερθέτει ετικέτες στα μηνύματά της για λόγους ελέγχου ταυτότητας. Ο Bob επικυρώνει την Alice μόνο όταν ανιχνεύει τις σωστές ετικέτες στο σήμα που έλαβε. Όταν ένα σήμα περιέχει μια ετικέτα ελέγχου ταυτότητας, λέμε ότι έχει επισημανθεί (tagged) [6][63][64]. Αν και οι κρυπτογραφικές τεχνικές βασισμένες σε ψηφιακά κλειδιά έχουν χρησιμοποιηθεί ευρέως τόσο για την ασφάλεια επικοινωνίας όσο και για τον έλεγχο ταυτότητας, ωστόσο σε πολλά αναδυόμενα σενάρια και σε διάφορες περιπτώσεις χρήσεις στα δίκτυα 5G ενδέχεται να μην μπορούν να καλύψουν την επιθυμητή απόδοση [64]. Πιο συγκεκριμένα, μια βασική αδυναμία των ψηφιακών διαπιστευτηρίων που βασίζονται στη συμβατική κρυπτογραφία είναι ότι, δεν μπορούν να εντοπιστούν εύκολα τα παραβιασμένα κλειδιά ασφαλείας, καθώς τα φυσικά χαρακτηριστικά των συσκευών επικοινωνίας και των χρηστών δεν λαμβάνονται υπόψη [65]. Με δεδομένη την ταχύτητα που αναπτύσσονται οι υπολογιστικές μηχανές χαμηλού κόστους (όπως για παράδειγμα οι συσκευές που χρησιμοποιούνται στο IoT), γίνεται όλο και πιο εύκολο να σπάσει (crack) το κλειδί ασφαλείας από την υποκλοπή σημάτων στα τυποποιημένα και στατικά πρωτόκολλα ασφαλείας. Επιπλέον, οι συμβατικές κρυπτογραφικές τεχνικές απαιτούν κατάλληλες διαδικασίες διαχείρισης κλειδιών για τη δημιουργία, διανομή, ανανέωση και ανάκληση ψηφιακών κλειδιών ασφαλείας, τα οποία μπορεί να οδηγήσουν σε υπερβολικές καθυστερήσεις στα δίκτυα μεγάλης κλίμακας, όπως τα δίκτυα 5G. Αυτός ο χρόνος καθυστέρησης μπορεί να επιφέρει σημαντικές επιπτώσεις στις επικοινωνίες, οι οποίες είναι ευαίσθητες στις καθυστερήσεις, όπως για παράδειγμα στον έλεγχο ενός δικτύου ηλεκτρικής

ενέργειας και στις επικοινωνίες των οχημάτων. Ακόμη, η υπολογιστική επιβάρυνση των κρυπτογραφικών μεθόδων που βασίζονται σε ψηφιακό κλειδί είναι ιδιαίτερα ανεπιθύμητη για συσκευές, οι οποίες έχουν περιορισμένη διάρκεια ζωής μπαταρίας και υπολογιστικής ικανότητας, όπως οι αισθητήρες IoT [64].

Για να ξεπεραστούν αυτές οι προκλήσεις, μια εναλλακτική προσέγγιση του ελέγχου ταυτότητας ενός χρήστη (πομπός) είναι η εκμετάλλευση των φυσικών επιπέδων των συνδέσεων επικοινωνίας. Τέτοια χαρακτηριστικά αναλογικού τομέα σχετίζονται με τη μοναδική ατέλεια των συσκευών επικοινωνίας και με το αντίστοιχο περιβάλλον, τα οποία είναι δύσκολο να πλαστογραφηθούν και να προβλεφθούν [64]. Αυτά τα χαρακτηριστικά φυσικού επιπέδου περιλαμβάνουν την απόκριση παλμού καναλιού (Channel Impulse Response- CIR), την απόκριση συχνότητας καναλιού (Channel Impulse Response - CFR) στην ορθογώνια διαίρεση πολλαπλών συχνοτήτων (Orthogonal Frequency Division Multiplexing - OFDM), τον δείκτη ισχύος λαμβανόμενου σήματος (Received Signal Strength Indicator - RSSI), την μετατόπιση συχνότητας φορέα (Carrier Frequency Offset - CFO) και το δακτυλικό αποτύπωμα ραδιοσυχνοτήτων (Radio Frequency Fingerprint - RFF) [6].

Αναλυτικότερα, στα δίκτυα κινητής τηλεφωνίας 5ης γενιάς χρησιμοποιείται η εξισορρόπηση πεδίου συχνότητας (Frequency Domain Equalization - FDE). Αποτελεί μια αποτελεσματική τεχνική για τον μετριασμό από την παρεμβολή μεταξύ συμβόλων (Inter Symbol Interference – ISI) [66], η οποία είναι μορφή παραμόρφωσης ενός σήματος στο οποίο ένα σύμβολο παρεμβαίνει με τα επόμενα σύμβολα. Σε ένα ψηφιακά διαμορφωμένο σήμα, ο ρυθμός συμβόλων (επίσης γνωστός ως ρυθμός baud ή ρυθμός διαμόρφωσης) είναι ο αριθμός διαμόρφωσης συμβόλων, κυματομορφής ή συμβάντων σηματοδότησης ανά μονάδα χρόνου. Κάθε σύμβολο μπορεί να αντιπροσωπεύει ή να μεταφέρει ένα ή περισσότερα σύμβολα. Με πιο απλά λόγια, ένα σύμβολο είναι μια κυματομορφή ή μια κατάσταση του καναλιού επικοινωνίας που παραμένει. Η τεχνική FDE, η οποία εφαρμόζεται στα συστήματα MIMO, συνήθως απαιτεί, εκ των προτέρων, τις πληροφορίες καναλιού, δηλαδή την απόκριση συχνότητας καναλιού (CFR). Η προσέγγιση της CFR μπορεί να γίνει αποκτώντας πρώτα την απόκτηση ώθησης καναλιού (CIR) και στη συνέχει να μεταφερθεί ξανά στον τομέα συχνότητας μέσω επεξεργασίας γρήγορου μετασχηματισμού Fourier (Fast Fourier Transport - FFT) [66].

Οι αποκρίσεις παλμού καναλιού (CIR) αποτελούν τις εξόδους ενός συστήματος κατά τη διάδοση ραδιοφωνικών σημάτων σε δίκτυα ευρείας ζώνης. Αυτές οι έξοδοι CIR περιλαμβάνουν καταγεγραμμένα δεδομένα με πληροφορίες σχετικά με τις διαδρομές

διάδοσης του σήματος [67]. Έτσι, η λειτουργία CIR δίνει τη δυνατότητα να εκτιμηθεί το περιβάλλον ενός καναλιού επικοινωνίας και να βελτιωθεί η απόδοση. Πιο συγκεκριμένα, το CIR είναι ένα γραμμικό χρονικό φίλτρο, το οποίο χαρακτηρίζει το κανάλι πολλαπλών διαδρομών του περιβάλλοντος λειτουργίας [68]. Σε ένα περίπλοκο σενάριο (όπως για παράδειγμα σε ένα βιομηχανικό περιβάλλον, οι συνθήκες διάδοσης σήματος και κατά συνέπεια τα αποτελέσματα των συλλεγόμενων CIR, μπορεί να ποικίλουν. Εξαιτίας της πολυπλοκότητας του περιβάλλοντος είναι δύσκολο να καταγραφεί σε ένα αναλυτικό μοντέλο [67]. Σε γενικές γραμμές, ένα CIR μπορεί να παρέχει ένα ολοκληρωμένο σύνολο πληροφοριών. Λύση σε αυτό τα πολύπλοκα δεδομένα CIR μπορεί να προσφέρουν οι λύσεις της μηχανικής μάθησης, ώστε να αναλυθούν αποτελεσματικότερα τα δεδομένα. Ο FFT είναι ένας αλγόριθμος που υπολογίζει τον διακριτό μετασχηματισμό Fourier (Discrete Fourier Transform - DFT) ή ο αντίστροφος (Inverse- IDFT). Η ανάλυση Fourier μετατρέπει ένα σήμα από τον αρχικό τομέα των συχνοτήτων και αντιστρόφως. Μια άλλη λύση αποτελεί η ορθογώνια πολυπλεξία διαίρεσης χρόνου (OFDM), η οποία αποτελεί μια συνηθισμένη λύση πολλαπλών μεταφορών για την αντιμετώπιση της αύξησης της καθυστέρησης του καναλιού και η οποία επίσης χρησιμοποιεί FFT. Η OFDM αποτελεί ένα τύπο ψηφιακής μετάδοσης και μια μέθοδο κωδικοποίησης σε πολλαπλές συχνότητες φορέα. Ένας φορέας είναι μια κυματομορφή που διαμορφώνεται με ένα σήμα φέρει πληροφορίες για το σκοπό της μεταφοράς πληροφοριών. Η OFDM χρησιμοποιείται ευρέως, μεταξύ άλλων, σε εφαρμογές κινητής επικοινωνίας [66]. Στα συμβατικά συστήματα OFDM μετρούν συνήθως τη μετατόπιση συχνότητας φορέα (CFO) [69].

Όπως έχει αναφερθεί και νωρίτερα, το 5G επιτρέπει την αποτελεσματική εφαρμογή των συστημάτων IoT. Στο πλαίσιο αυτό χρησιμοποιούνται διάφορες τεχνολογίες για την επικοινωνία μεταξύ των συσκευών, όπως τα ασύρματα τσιπ ραδιοσυχνοτήτων τύπου WiFi, Zigbee ή Bluetooth. Κάθε ασύρματο τσιπ παρέχει ένα byte πληροφοριών που σχετίζονται με την ισχύ του λαμβανόμενου σήματος (RSS), το λεγόμενο δείκτη ισχύος λαμβανόμενου σήματος (RSSI). Σε ένα δεδομένο χρονικό σημείο, σε ένα συγκεκριμένο ραδιοφωνικό κανάλι, το RSSI έχει μια δεδομένη στιγμή στο χρόνο. Σε γενικό πλαίσιο, ο RSSI είναι μια μονάδα μέτρησης ισχύος που υπάρχει σε ένα ραδιοφωνικό σήμα που λαμβάνεται [70].

Τέλος, η εξαγωγή δακτυλικών αποτυπωμάτων ραδιοσυχνοτήτων (Radio Frequency Fingerprint - RFF) είναι μια τεχνολογία που μπορεί να αναγνωρίσει τον μοναδικό πομπό ραδιοφώνου σε φυσικό επίπεδο, χρησιμοποιώντας μόνο εξωτερικές μετρήσεις χαρακτηριστικών για να ταιριάζει με τη βιβλιοθήκη χαρακτηριστικών. Το RFF αντικατοπτρίζει τις διαφορές μεταξύ των εξαρτημάτων υλικού των πομπών και περιέχει

πλούσια μη γραμμικά χαρακτηριστικά των εσωτερικών εξαρτημάτων μέσα στον πομπό. Η τεχνική RFF έχει εφαρμοστεί ευρέως για την ενίσχυση της ασφάλειας της επικοινωνίας ραδιοσυχνοτήτων [62].

4.3.2 Προκλήσεις απόρρητου (privacy) στο φυσικό επίπεδο

Από την πλευρά του χρήστη, οι κύριες ανησυχίες για το απόρρητο σχετίζονται κυρίως με τα δεδομένα, την τοποθεσία και την ταυτότητα [71]. Οι περισσότερες εφαρμογές για τα έξυπνα κινητά τηλέφωνα απαιτούν πρόσβαση σε προσωπικά στοιχεία του συνδρομητή πριν από την εγκατάσταση τους. Οι προγραμματιστές και οι εταιρίες των εφαρμογών ενημερώνουν το χρήστη για τους σκοπούς που πρόκειται να χρησιμοποιηθούν τα προσωπικά τους δεδομένα. Απειλές όπως, οι σημασιολογικές επιθέσεις πληροφοριών (semantic information attacks), οι επιθέσεις χρονισμού (timing attacks) και οι επιθέσεις ορίων (boundary attacks) στοχεύουν κυρίως το απόρρητο της τοποθεσίας των συνδρομητών [72]. Σε φυσικό επίπεδο, το απόρρητο της τοποθεσίας στα δίκτυα 5G, μπορεί να διαρρεύσει με αλγόριθμους επιλογής σημείου πρόσβασης (access point selection algorithms). Οι επιθέσεις στην διεθνή ταυτότητα κινητού συνδρομητή (International Mobile Subscriber Identity - IMSI) μπορούν να χρησιμοποιηθούν για να αποκαλύψουν την ταυτότητα ενός συνδρομητή, πιάνοντας το IMSI του εξοπλισμού χρήστη (User Equipment - UE) του συνδρομητή. Τέτοιες επιθέσεις μπορούν επίσης να προκληθούν με τη δημιουργία ενός ψεύτικου σταθμού βάσης που θεωρείται ως προτιμώμενος σταθμός βάσης από το UE και έτσι οι συνδρομητές θα ανταποκρίνονται με το IMSI τους [5].

4.4 Λύσεις ασφαλείας στις νέες τεχνολογίες

Αφού μελετήθηκαν οι προκλήσεις ασφαλείας που εντοπίζονται στις νέες τεχνολογίες, σε αυτή την ενότητα παρουσιάζονται οι λύσεις που μπορούν να εφαρμοστούν στην κάθε τεχνολογία για τον περιορισμό των προκλήσεων ασφαλείας. Οι λύσεις αυτές, αποτελούν κατά βάση παραδοσιακές προσεγγίσεις ασφαλείας. Ενώ στο επόμενο κεφάλαιο προσεγγίζονται οι λύσεις στις οποίες μπορούν να εφαρμοστούν τεχνικές μηχανικής μάθησης.

4.4.1 Λύσεις ασφαλείας σε massive MIMO

Για να επιτευχθούν πλήρως τα οφέλη της τεχνολογίας MIMO, το σύστημα πρέπει να ασφαλιστεί απέναντι στις σημαντικές προκλήσεις ασφαλείας που παρουσιάστηκαν νωρίτερα. Στο [73] παρουσιάζονται δυο διαφορετικές μεθόδους για την ανίχνευση ενεργή επίθεσης υποκλοπής.

Η πρώτη μέθοδος περιλαμβάνει την εκμετάλλευση της ελεγχόμενης τυχαιότητας (controlled randomness) με τη μετάδοση τυχαίων πιλότων (random pilots) για την ανίχνευση των ενεργών υποκλοπών. Δηλαδή, ο νόμιμος χρήστης μεταδίδει μια ακολουθία συμβόλων τυχαίας μετατόπισης φάσης (Phase Shift keying - PSK), η οποία αποτελεί το κλειδί για την ανίχνευση της υποκλοπής στον σταθμό βάσης [73]. Με πιο απλά λόγια, μέσω τις ελεγχόμενη τυχαιότητας υλοποιείται μια μέθοδος που χρησιμοποιεί ταυτόχρονα όλα τα διαφορετικά κλειδιά, κατά τη διάρκεια μιας συνεδρίας. Δηλαδή, ενώ στα κλασσικά συστήματα κρυπτογράφησης, η χρήση των κλειδιών γίνεται γραμμικά, δηλαδή, για κάθε συνεδρία χρησιμοποιείται ένα κλειδί (π.χ. το κλειδί 1 για την πρώτη περίοδο της συνεδρίας, το κλειδί 2 για τη δεύτερη και ούτω καθεξής). Αντίθετα, στην ελεγχόμενη τυχαιότητα, χρησιμοποιούνται όλα τα κλειδιά ταυτόχρονα στη συνολική διάρκεια της συνεδρίας [74]. Ωστόσο, το μειονέκτημα αυτής της μεθόδου είναι ότι συνεπάγεται τα γενικά έξοδα μετάδοσης επιπλέον τυχαίων ακολουθιών [36][[73]. Στη δεύτερη μέθοδο, που παρουσιάζεται στο [73], όταν λαμβάνεται το σήμα εκπαίδευσης από το νόμιμο χρήστη, ο σταθμός βάσης μπορεί να εφαρμόσει ένα διαμορφωτή δέσμης (beamformer), ο οποίος κατασκευάζεται με τέτοιο τρόπο ώστε το δείγμα που λαμβάνεται από τον νόμιμο χρήστη να ισούται με μια συμφωνημένη τιμή. Στην περίπτωση μιας ενεργής επίθεσης υποκλοπής, ο νόμιμος χρήστης θα παρατηρήσει μια πολύ μικρότερη τιμή. Η ανίχνευση ενεργής υποκλοπής μπορεί να αντιμετωπιστεί από συνεργατικούς σταθμούς βάσης. Σε τέτοια σενάρια, διαφορετικοί σταθμοί βάσης μπορούν να ανταλλάσσουν πληροφορίες και έτσι δίνουν την ευκαιρία να εκτιμηθούν από κοινού τα επίπεδα της νόμιμης πιλοτικής μόλυνσης (Pilot Contamination) που προκαλείται από τον χρήστη. Οι μέθοδοι μηχανικής μάθησης μπορούν επίσης να χρησιμοποιηθούν για τον εντοπισμό ενεργών επιθέσεων παραβίασης, όπως περιγράφονται αναλυτικά σε επόμενο κεφάλαιο.

Καθώς ο σταθμός βάσης massive MIMO μπορεί να εξυπηρετήσει ταυτόχρονα μεγάλο αριθμό χρηστών, είναι απαραίτητο να εξασφαλιστεί ένα μήνυμα από όλους τους χρήστες εκτός από τον προβλεπόμενο. Ο αλγόριθμος προκατασκευαστή (precoder), δηλαδή ο αλγόριθμος για τον διαμορφωτή δέσμης, που χρησιμοποιείται στο σταθμό βάσης πρέπει να σχεδιαστεί κατά τρόπο ώστε να επιτευχθεί αυτός ο στόχος. Αξίζει επίσης να εξεταστεί η πιθανότητα ένας υποκλοπέας να χρησιμοποιεί μαζικές συστοιχίες κεραιών για να παρακολουθεί τις πληροφορίες [36].

Στο [75] προτείνεται μια προσέγγιση ασφαλείας στο φυσικό επίπεδο που ονομάζεται περιστρεφόμενη αρχική φάση συμβόλου (Original Symbol Phase Rotated - OSPR). Η βασική ιδέα του σχήματος OSPR είναι η τυχαία περιστροφή της φάσης των αρχικών συμβόλων στο

σταθμό βάσης πριν μεταδοθούν, έτσι ώστε ο υποκλοπέας του μαζικού MIMO να μπερδευτεί από τα παρεμποδισμένα σήματα, τα οποία ενδέχεται να μην αντιπροσωπεύουν τα πραγματικά σύμβολα πληροφοριών. Ωστόσο, οι νόμιμοι χρήστες μπορούν να συμπεράνουν τις σωστές περιστροφές φάσης και να κάνουν τις κατάλληλες αντίστροφες λειτουργίες για να ανακτήσουν τα αρχικά σύμβολα.

4.4.2 Λύσεις ασφαλείας σε SDN

Η ασφάλεια του SDN είναι πολυδιάστατη. Οι λύσεις ασφαλείας μπορούν να διαχωριστούν σε λύσεις ασφαλείας για επίπεδο εφαρμογής, επίπεδο ελέγχου, επίπεδο δεδομένων και επίπεδο διεπαφής SDN.

Σε επίπεδο εφαρμογής, αναμφίβολα, οι κακόβουλες εφαρμογές δεν πρέπει να έχουν πρόσβαση ούτε στο δίκτυο, ούτε στο επίπεδο ελέγχου δικτύου SDN. Στη βιβλιογραφία υπάρχουν διάφορες προτάσεις για την επαλήθευση (verification) των εφαρμογών SDN πριν τους δοθεί η άδεια πρόσβασης από το επίπεδο ελέγχου για τη διαμόρφωση του δικτύου. Για παράδειγμα, στο [76] προτείνεται το PermOF το οποίο αποτελεί είναι ένα λεπτομερές σύστημα αδειών που θέτει όρια στις εφαρμογές για να λειτουργήσει εντός των καθορισμένων δικαιωμάτων του. Ο σχεδιασμός του PermOF παρέχει δικαιώματα ανάγνωσης, εγγραφής, ειδοποίησης και συστήματος σε διαφορετικές εφαρμογές για την επιβολή ελέγχου δικαιωμάτων. Έτσι, προστατεύει τις πλατφόρμες ελέγχου από κακόβουλες εφαρμογές. Ένα άλλο σύστημα αδειών για εφαρμογές SDN περιγράφεται στο [77] που διασφαλίζει ότι οι λειτουργίες του επιπέδου ελέγχου είναι διαθέσιμες μόνο σε αξιόπιστες εφαρμογές. Παρομοίως, στο [78] προτείνεται το FortNOX, το οποίο είναι μια επέκταση λογισμικού που παρέχει εξουσιοδότηση βάσει ρόλων μέσω ελέγχου ταυτότητας και επιβάλει περιορισμούς ασφαλείας για κάθε εφαρμογή OpenFlow.

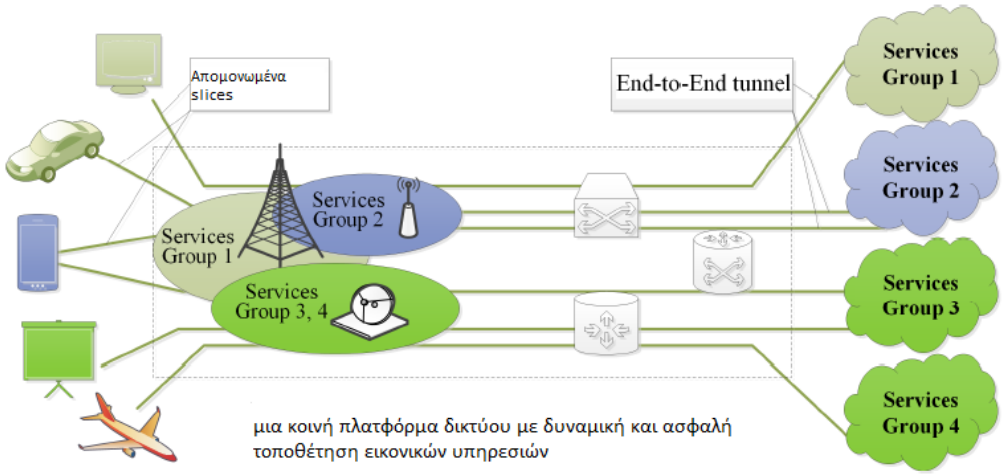
Σε επίπεδο ελέγχου (π.χ. ελεγκτής) υπάρχουν διάφορες προσεγγίσεις για την ενίσχυση της ασφάλειας, καθώς αυτό το επίπεδο έχει κεντρικό ρόλο. Στο [79] προτείνεται ένας ελεγκτής ενισχυμένης ασφάλειας, ο οποίος ονομάζεται SE-Floodlight και ο σκοπός του είναι ασφαλίσει το επίπεδο ελέγχου SDN. Για το σκοπό αυτό, ο ελεγκτής SE-Floodlight παρέχει μηχανισμούς για το διαχωρισμό των δικαιωμάτων, προσθέτοντας ένα API βορά (north-bound API), το οποίο σύμφωνα με την αρχιτεκτονική SDN που παρουσιάστηκε στο προηγούμενο κεφάλαιο αποτελεί μια εφαρμογή μεταξύ του επιπέδου εφαρμογών και του επιπέδου ελεγκτών. Ουσιαστικά, λειτουργεί ως μεσολαβητής μεταξύ της εφαρμογής και των επιπέδων δεδομένων επαληθεύοντας τους κανόνες ροής που δημιουργούνται από τις εφαρμογές. Ο ελεγκτής ROSEMARY [80] είναι ένα ισχυρό λειτουργικό σύστημα δικτύου για τον ελεγκτή

SDN, ο οποίος βασίζεται σε ένα σύστημα αδειών για εφαρμογές για τη διασφάλιση της λειτουργίας του ελεγκτή από λάθη ή κακόβουλες εφαρμογές.

Για το μετριασμό τους ζητήματος επεκτασιμότητας του ελεγκτή και τη βελτίωση της ανθεκτικότητας του απέναντι σε επιθέσεις DoS, στο [81] σχεδιάστηκε, αναπτύχθηκε και αξιολογήθηκε η εφαρμογή AVANT-GUARD, η οποία περιορίζει τα αιτήματα ροής, ξεχωρίζοντας τις αποτυχημένες περιόδους σύνδεσης TCP στο επίπεδο δεδομένων πριν από οποιαδήποτε ειδοποίηση στο επίπεδο ελέγχου. Για την επίτευξη αυτού του σκοπού χρησιμοποιεί ένα εργαλείο μετεγκατάστασης σύνδεσης (connection migration). Στη βιβλιογραφία προτείνονται διάφορες μέθοδοι για την ασφάλεια στο επίπεδο ελέγχου απέναντι σε επιθέσεις DoS, οι οποίες χρησιμοποιούν αλγορίθμους μηχανικής μάθησης και περιγράφονται σε επόμενο κεφάλαιο αναλυτικότερα.

4.4.3 Λύσεις ασφαλείας σε NFV

Η εικονικοποίηση μπορεί να αυξήσει σε μεγάλο βαθμό την ασφάλεια του χρήστη, των υπηρεσιών και του δικτύου. Ένας βασικός μηχανισμός είναι η χρήση τεμαχισμού (slicing) του δικτύου για τον διαχωρισμό της κίνησης διαφορετικών υπηρεσιών (όπως φαίνεται στην ακόλουθη εικόνα) ή τμημάτων δικτύου με βάση τις προτεραιότητες ασφαλείας [36].



Εικόνα 10: Ασφάλεια διόδων (tunnels) E2E για διαφορετικές υπηρεσίες [36]

Λόγω του κεντρικού ρόλου του επόπτη στα εικονοποιημένα συστήματα, η βασική επιλογή για την ενίσχυση της ασφάλειάς του είναι η ελάχιστη ή περιορισμένη έκθεση σε εικονικές μηχανές ή σε άλλα συστήματα. Για τους λόγους αυτούς, στο [82] πρότειναν το Xoar, μια τροποποιημένη έκδοση του Xen. Πρόκειται για μια πλατφόρμα που διασπά την εικονική μηχανή ελέγχου σε πολλαπλές εικονικές μηχανές με σκοπό να καταστήσει σαφή την έκθεση

σε κινδύνους, να διατηρήσει τα λιγότερα προνόμια πρόσβασης και έτσι να αυξήσει την ασφάλεια του συνολικού συστήματος [82]. Στο [83] παρουσιάζεται ο επόπτης OpenVirteX, ο οποίος λειτουργεί παρόμοια με την εφαρμογή OpenFlow SDN, ο επόπτης ενεργεί όπως ο ελεγκτής στο OpenFlow, όπου τα κομμάτια (slices) των χρηστών έχουν το δικό τους επίπεδο ελέγχου και δεδομένων. Το ενδιαφέρον γεγονός σε αυτήν τη λύση είναι η ικανότητα του ελεγκτή SDN που μπορεί να επιβλέπει τις δραστηριότητες του επόπτη και μπορεί εύκολα να παρακολουθεί τις ευπάθειες ασφαλείας [83].

Η δυναμική φύση των NFV και οι εικονικοποιημένοι πόροι μπορούν να χρησιμοποιηθούν ως πλεονέκτημα από την άποψη της ασφάλειας. Για παράδειγμα, η ευελιξία του NFV επιτρέπει την απομόνωση παραβιασμένων στοιχείων δικτύου ή ακόμη και ολόκληρων τμημάτων δικτύου μέσω του καθορισμού ζωνών ασφαλείας και της χρήσης διεύθυνσης κίνησης [84]. Στο [84], προτείνεται ένα πλαίσιο διαχείρισης και κρυπτογράφησης προσανατολισμένο στην ασφάλεια που διαχειρίζεται δυναμικά ολόκληρο τον κύκλο ζωής των λειτουργιών ασφαλείας, καθώς εξασφαλίζει υποδομές και πλατφόρμες που βασίζονται σε NFV. Ωστόσο, η ελαστικότητα του NFV σπάνια χρησιμοποιήθηκε για την αύξηση της ασφάλειας του δικτύου.

Τέλος, αναφορικά με τις προκλήσεις ασφαλείας στα κανάλια επικοινωνίας. Η ασφάλεια των NFV εξαρτάται σε μεγάλο βαθμό από την ασφάλεια των συστημάτων που χρησιμοποιούν τα MVNO. Ωστόσο, αυτό δεν είναι αρκετό από μόνο του. Για παράδειγμα, η διαλειτουργικότητα διαφόρων στοιχείων σε ένα σύστημα συνδυασμού και αντιστοίχισης θα απαιτήσει ένα πλαίσιο παρακολούθησης ασφάλειας και εμπιστοσύνης. Ένα πλαίσιο ασφάλειας και εμπιστοσύνης για VNF σε δίκτυα κινητής τηλεφωνίας που βασίζονται σε SDN προτείνεται στο [85]. Το προτεινόμενο πλαίσιο εφαρμόζει προσαρμοστικές τεχνολογίες αξιολόγησης και διαχείρισης εμπιστοσύνης και βιώσιμες αξιόπιστες τεχνολογίες υπολογιστών για τη διασφάλιση της εμπιστοσύνης της πλατφόρμας υπολογιστών και την επίτευξη ασφάλειας δικτύου SDN [85]. Ομοίως, ένα εννοιολογικό πλαίσιο διαχείρισης ασφάλειας που βασίζεται σε NFV παρουσιάζεται στο [84]. Το προτεινόμενο πλαίσιο προστατεύει τους πόρους ή άλλα NFV από απειλές ασφαλείας που προέρχονται από το Διαδίκτυο ή άλλους NFV επικυρώνοντας τα χαρακτηριστικά ασφαλείας τους. Ωστόσο, η περιορισμένη ανάπτυξη τέτοιων χειριστών σε πραγματικό χρόνο καθιστά δύσκολη την πλήρη κατανόηση του πλήρους φάσματος πιθανών απειλών ασφαλείας. Επομένως, υπάρχει λίγη έρευνα που επικεντρώνεται στις λύσεις ασφαλείας για MVNOs, εκτός από την αύξηση της ασφάλειας των NFV.

4.4.4 Λύσεις ασφαλείας σε νέφη κινητής

Οι επιθέσεις με τη μορφή HX-DoS, DDoS ή SIPDAS μπορούν γενικά να μετριαστούν ελέγχοντας συχνά την κατανάλωση υπολογιστικών πόρων καθώς και την ένταση των εισερχόμενων αιτημάτων. Αυτές οι λύσεις μπορούν να επιτευχθούν με μεθόδους μηχανικής μάθησης, οι οποίες αναλύονται περισσότερο στο επόμενο κεφάλαιο.

Η αντιμετώπιση των επιθέσεων insider σε δίκτυα 5G που βασίζονται σε νέφη αποτελούν τόσο κοινωνική πρόκληση όσο και τεχνική πρόκληση. Ενώ οι πάροχοι υπηρεσιών νέφους εργάζονται για την παροχή ασφαλών συνδέσεων και εφαρμογών (API) για τους χρήστες, υπάρχει η πιθανότητα κατάχρησης και άσκοπης χρήσης των δεδομένων στο νέφος από εξουσιοδοτημένο προσωπικό του παρόχου υπηρεσιών απαιτεί εξίσου ανησυχία. Αυτή η πρόκληση συγχέεται περαιτέρω με άλλα φυσικά περιστατικά όπως διαρροές και απώλειες δεδομένων. Από αυτήν την άποψη, η ανάγκη για πιο ισχυρές εγκαταστάσεις δημιουργίας αντιγράφων ασφαλείας και πολλαπλότητα αντιγράφων ασφαλείας σε διαφορετικές τοποθεσίες και πλατφόρμες αποτελεί μια στρατηγική που μπορεί πρακτικά να εφαρμοστεί, ώστε να μετριαστεί αυτή η πιθανή απειλή. Για άλλες επιθέσεις με insider, η εφαρμογή κατάλληλων ελέγχων και ελέγχων ρουτίνας σε συνδυασμό με ψηφιακή χρονική σήμανση και υπογραφές σε δεδομένα νέφους, θα μπορούσε να συμβάλει στον μετριασμό της πιθανότητας κατάχρησης και άσκοπης χρήσης δεδομένων νέφους από το εξουσιοδοτημένο προσωπικό [85].

4.5 Έλεγχος ταυτότητας εξοπλισμού χρήστη και δικτύου

Τα δίκτυα κινητής επικοινωνίας συνδέουν μεγάλο μέρος του παγκόσμιου πληθυσμού. Η ασφάλεια των κλήσεων, των δεδομένων, και των sms εξαρτάται από τις εγγυήσεις που παρέχονται από τα πρωτόκολλα πιστοποιημένης ανταλλαγής κλειδιών (Authentical Key Exchange). Για τα δίκτυα 5G, η ομάδα 3GPP έχει τυποποιήσει το πρωτόκολλο 5G AKA (Authentication and Key Agreement) για την επίτευξη αυτού του σκοπού. Το πρότυπο 5G αναπτύχθηκε σε δύο φάσεις.

Τον Ιούνιο του 2019, η 3GPP δημοσίευσε την τελική έκδοση v15.4.0[87] της έκδοσης 15, που αποτελεί την Τεχνική Προδιαγραφή μέσω της οποία καθορίζεται η αρχιτεκτονική και οι διαδικασίες ασφαλείας για τα δίκτυα 5G. Ο έλεγχος ταυτότητας στην έκδοση 15 βασίζεται σε νέες εκδόσεις των πρωτοκόλλων AKA (τα οποία, αποτελούν ουσιαστικά τα εξελιγμένα πρωτόκολλα που χρησιμοποιήθηκαν για το 4G (EPSAKA)) και ονομάζονται 5G AKA. Η έκδοση 15 περιλαμβάνει πρότυπα ασφαλείας για αυτόνομα και μη αυτόνομα βελτιωμένα

σενάρια κινητής ευρυζωνικότητας. Ενώ οι επόμενες εκδόσεις θα επικεντρώνονται σε πρότυπα ασφαλείας για μαζική επικοινωνίας μεταξύ μηχανημάτων και αξιόπιστες επικοινωνίες με χαμηλό χρόνο καθυστέρησης. Οι νέες δυνατότητες ασφαλείας στοχεύουν στην παροχή ασφάλειας E2E (επικοινωνία από άκρο σε άκρο), καθώς και την ενσωμάτωση πολλαπλών σημείων ελέγχου ταυτότητας και πρωτοκόλλων ασφαλείας για την υποστήριξη της ασφάλειας στο πλαίσιο της αρχιτεκτονικής που βασίζεται στην υπηρεσία (Service Based Architecture - SBA) [87].

Στην αρχιτεκτονική LTE, το EPS-AKA χρησιμοποιήθηκε για την αμοιβαία πιστοποίηση μεταξύ του UE και του δικτύου. Το AKA, που ξεκίνησε στο GSM, εξελίχθηκε με τις επόμενες γενιές και εξακολουθεί να θεωρείται ως ο πιο βιώσιμος μηχανισμός ελέγχου ταυτότητας και εξουσιοδότησης σε δίκτυα 5G. Το AKA βασίζεται σε συμμετρικά κλειδιά και εκτελείται σε SIM. Η μέθοδος Extensible Authentication Protocol (EAP)-AKA για δίκτυα 3G αναπτύχθηκε από το 3GPP για να υποστηρίξει το απόρρητο ταυτότητας και τον γρήγορο επαναληπτικό έλεγχο ταυτότητας. Το EPS-AKA παρείχε περισσότερη ασφάλεια, όπως η χρήση πολλαπλών κλειδιών σε διαφορετικά περιβάλλοντα, η ανανέωση των κλειδιών και αυτό χωρίς τη συμμετοχή του οικιακού δικτύου κάθε φορά. Το EPS-AKA έχει κάποιες προκλήσεις, όπως ο υπολογισμός και τα γενικά έξοδα επικοινωνίας και η καθυστέρηση, ωστόσο, δεν έχουν εμφανιστεί μέχρι στιγμής ορατές ευπάθειες και έτσι θα χρησιμοποιηθεί στο 5G [36].

4.6 Μέθοδοι ελέγχου ταυτότητας (authentication)

Συνοπτικά σε αυτή την ενότητα γίνεται αναφορά στις μεθόδους ελέγχου ταυτότητας. Υπενθυμίζεται ότι η αυθεντικοποίηση ή αλλιώς ο έλεγχος ταυτότητας σχετίζεται με τη διαδικασία κατά την οποία επιβεβαιώνεται η ταυτότητα που παρουσίαση η οντότητα στο σύστημα.

4.6.1 Έλεγχος ταυτότητας με κωδικό πρόσβασης (password authentication)

Ο κωδικός πρόσβασης (password) αποτελεί την ποιο συνηθισμένη μέθοδο και την πιο βασική μορφή ελέγχου ταυτότητας. Σε αυτή τη μέθοδο, αφού ένας χρήστης εισαγάγει το όνομα χρήστη του, πρέπει να πληκτρολογήσει έναν μυστικό κωδικό για να αποκτήσει πρόσβαση (π.χ. στο δίκτυο, στην εφαρμογή). Εάν κάθε χρήστης διατηρήσει τον κωδικό πρόσβασής του ιδιωτικό, θα αποτραπεί η μη εξουσιοδοτημένη πρόσβαση. Ωστόσο, ακόμη και οι μυστικοί κωδικοί πρόσβασης είναι ευάλωτοι σε εισβολές. Ένας κακόβουλος χρήστης μπορεί να

χρησιμοποιούν προγράμματα που δοκιμάζουν χιλιάδες πιθανούς κωδικούς πρόσβασης, αποκτώντας πρόσβαση όταν μαντέψουν το σωστό. Για το λόγο αυτό κάθε χρήστης θα πρέπει να χρησιμοποιεί περίπλοκους κωδικούς πρόσβασης που είναι δυσκολότερη στην παραβίασης τους [88]. Ένας κωδικός πρόσβασης μπορεί να είναι με τη μορφή προσωπικού κωδικού αναγνώρισης, γνωστός ως PIN (Personal Identification Numbers). Σε αυτή την περίπτωση συνήθως, οι χρήστες πρέπει να επιλέξουν μια σειρά τεσσάρων ψηφίων τα οποία πρέπει να απομνημονεύσουν. Στη συνέχεια, κάθε φορά που η κινητή συσκευή πρέπει να ξεκλειδώσει, το σύστημα ζητά, μέσω ενός πεδίου εισαγωγής, ο χρήστης πρέπει να συμπληρώσει αυτά τα ψηφία με τη σωστή σειρά για να εξουσιοδοτηθεί η πρόσβαση ολόκληρο το περιεχόμενο της συσκευής ή σε μια εφαρμογή [89].

Ο κωδικός πρόσβασης μπορεί επίσης να έχει τη μορφή βελτιωμένου κειμένου. Σε αντίθεση με το PIN, οι χρήστες πρέπει να επιλέξουν, τουλάχιστον, μια σειρά έξι χαρακτήρων που δεν περιορίζονται μόνο σε ψηφία. Στη συνέχεια, ο έλεγχος ταυτότητας εξαρτάται από την ίδια διαδικασία με το PIN.

Ακόμη, ο κωδικός πρόσβαση μπορεί να είναι γραφικός, δηλαδή ο χρήστης πρέπει να σχεδιάσει ένα συγκεκριμένο μοτίβο (π.χ. στην οθόνη αφής), όπως για παράδειγμα το μοτίβο ξεκλειδώματος που διαθέτουν οι συσκευές android [89].

4.6.2 Βιομετρικός έλεγχος ταυτότητας (biometrical authentication)

Τα βιομετρικά στοιχεία βασίζονται στα φυσικά χαρακτηριστικά ενός χρήστη. Τα πιο ευρέως διαθέσιμα βιομετρικά συστήματα χρησιμοποιούν δακτυλικά αποτυπώματα (finger print recognition), σάρωση αμφιβληστροειδούς ή ίριδας (του ματιού), γεωμετρία παλάμης και δακτύλων (hand/finger geometry), αναγνώριση αγγειακού μοτίβου (Vascular Pattern Recognition), δυναμική υπογραφή (dynamic signature), αναγνώριση φωνής (voice recognition) και ανίχνευση προσώπου. Δεδομένου ότι κανένας χρήστης δεν έχει τα ίδια ακριβή φυσικά χαρακτηριστικά, ο βιομετρικός έλεγχος ταυτότητας είναι εξαιρετικά ασφαλής [88].

Στην έλεγχο ταυτότητας με δακτυλικό αποτύπωμα, ο χρήστης πρέπει να ενημερώσει τη συσκευή για το μοτίβο του δακτυλικού αποτυπώματος, για ένα ή περισσότερα δάχτυλα. Για το σκοπό αυτό, πρέπει να βάλει κάθε ένα από τα δάχτυλα που θέλει να χρησιμοποιήσει για να ξεκλειδώσει τη συσκευή του στον αισθητήρα, πολλές φορές. Έτσι, κάθε φορά που απαιτείται έλεγχος ταυτότητας, ο χρήστης τοποθετεί οποιοδήποτε δάχτυλο που έχει προηγουμένως καταγραφεί στον αισθητήρα. Η τεχνολογία αυτή χρησιμοποιεί μοναδικά μοτίβα δακτυλικών

αποτυπωμάτων που υπάρχουν στα δάχτυλα κάθε ανθρώπου για την πιστοποίηση χρηστών. Οι κορυφογραμμές που συνθέτουν το μοτίβο είναι παραδοσιακά ταξινομημένες σε βρόχους, καμάρες και μοτίβα στροβιλισμού.

Εκτός από τη διαδικασία ελέγχου ταυτότητας δακτυλικών αποτυπωμάτων, ένας χρήστης μπορεί να ενημερώσει τη συσκευή του για τα χαρακτηριστικά του προσώπου του. Στη συνέχεια, όταν το πρόσωπο αναγνωρίζεται από το σύστημα, ο χρήστης έχει πρόσβαση στη συσκευή ή την εφαρμογή.

Οι τεχνικές αναγνώρισης φωνής εστιάζουν επίσης σε βιομετρικά χαρακτηριστικά, δηλαδή στα χαρακτηριστικά που παράγονται από την ομιλία. Αυτά τα χαρακτηριστικά εξαρτώνται από τη διάσταση της φωνητικής οδού, του στόματος και των ρινικών κοιλοτήτων, αλλά επίσης βασίζονται στο βήμα φωνής, το στυλ ομιλίας και τη γλώσσα. Υπάρχουν δύο κορυφαίες μέθοδοι για την επεξεργασία της αναγνώρισης φωνής, στη μια ο χρήστης πρέπει να πει μια συγκεκριμένη λέξη κλειδί και στην άλλη η αναγνώριση του χρήστη γίνεται ανεξάρτητα από το τι θα πει [89].

4.6.3 Έλεγχος ταυτότητας δύο παραγόντων (two-factor authentication - 2FA)

Ο έλεγχος ταυτότητας δύο παραγόντων βασίζεται στους κωδικούς πρόσβασης για τη δημιουργία μιας πολύ πιο ισχυρής λύσης ασφαλείας. Απαιτεί τόσο έναν κωδικό πρόσβασης όσο και την κατοχή ενός συγκεκριμένου φυσικού αντικειμένου για την απόκτηση πρόσβασης [88].

4.6.4 Έλεγχος ταυτότητας με token (τεκμήριο)

Τα συστήματα token χρησιμοποιούν μια ειδικά σχεδιασμένη φυσική συσκευή για τον έλεγχο δύο παραγόντων. Αυτό μπορεί να είναι ένα dongle που έχει εισαχθεί στη θύρα USB του υπολογιστή ή μια έξυπνη κάρτα που περιέχει αναγνώριση ραδιοσυχνοτήτων ή τσιπ επικοινωνίας [88][89].

4.6.5 Έλεγχος ταυτότητας μοναδικής σύνδεσης (single sign-on (SSO))

Η μοναδική σύνδεση (SSO) επιτρέπει στον χρήστη να εισάγει τα διαπιστευτήριά του μόνο μία φορά για να αποκτήσει πρόσβαση σε πολλές εφαρμογές. Σκεφτείτε έναν υπάλληλο που χρειάζεται πρόσβαση σε email και αποθήκευση νέφους σε ξεχωριστούς ιστότοπους. Εάν οι δύο ιστότοποι συνδέονται με SSO, ο χρήστης θα έχει αυτόματα πρόσβαση στον ιστότοπο αποθήκευσης νέφους αφού συνδεθεί στον πελάτη email. Το SSO εξοικονομεί χρόνο και κάνει τους χρήστες ευχαριστημένους αποφεύγοντας την επανειλημμένη εισαγωγή κωδικών.

Ωστόσο, μπορεί επίσης να εισάγει κινδύνους ασφαλείας, καθώς ένας μη εξουσιοδοτημένος χρήστης που αποκτά πρόσβαση σε ένα σύστημα μπορεί να διεισδύσει και σε άλλα [88].

5 *Μηχανική μάθηση και ασφάλεια δικτύων 5G*

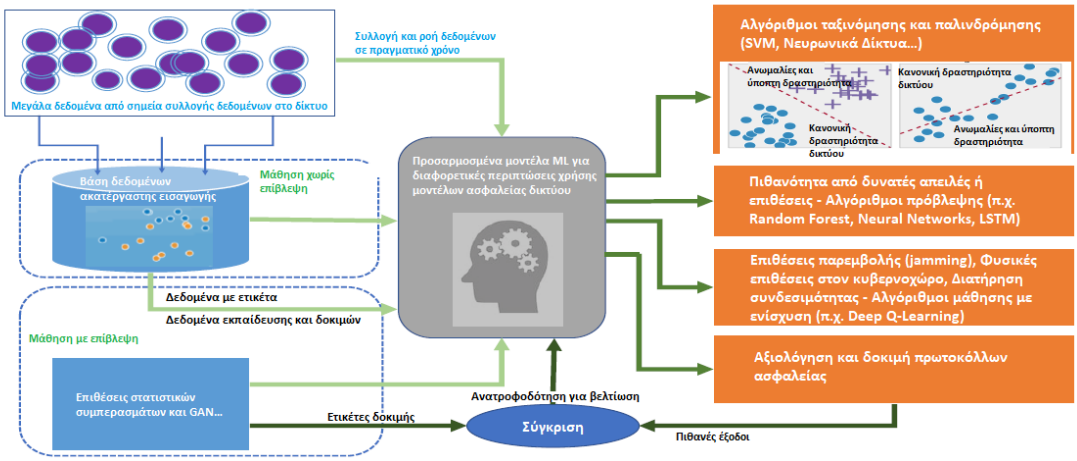
Η ιδέα για τη χρήση της τεχνητής νοημοσύνης και της μηχανικής μάθησης δεν είναι καινούρια, όμως, η εξέλιξη των αλγορίθμων βαθιάς μάθησης έχουν αλλάξει το σκοπό χρήσης τους. Ειδικότερα, πριν αναπτυχθούν οι αλγόριθμοι βαθιάς μάθησης, οι περισσότερες από τις μεθόδους μηχανικής μάθησης ήταν επικεντρωμένες στη μοντελοποίηση των μοτίβων επίθεσης που δεν έχουν δυναμικά χαρακτηριστικά. Όμως, με τη βαθιά μάθηση, τα συστήματα δικτύου μπορούν να γίνουν πιο ανθεκτικά σε νέες εξελιγμένες απειλές και επιθέσεις με δυναμικά χαρακτηριστικά. Πιο συγκεκριμένα, οι επιτιθέμενοι χρησιμοποιούν εξελιγμένες τεχνικές επίθεσης, όπως η απόκρυψη, ο πολυμορφισμός ή η πλαστοπροσωπία για να αποφύγουν τον εντοπισμό τους [7], αυτά θεωρούνται ως δυναμικά χαρακτηριστικά.

Από τη φύση τους, οι αλγόριθμοι μηχανικής μάθησης απαιτούν όσο το δυνατών περισσότερα δεδομένα για την εκπαίδευση του μοντέλου τους και την αποτελεσματική τους λειτουργία. Στην εποχή του 5G, γίνεται εύκολα η παραγωγή, η αποθήκευση και η διαχείριση, καθώς υπάρχει υψηλή υπολογιστική ισχύ, εκθετική ανάπτυξη δεδομένων και πάρα πολλές πηγές δεδομένων. Ένα δίκτυο 5G μπορεί να διατηρηθεί, να προσπελαστεί και να αναλυθεί για πιθανές απειλές, επιθέσεις και ευπάθειες χρησιμοποιώντας τη μηχανική μάθηση με χαμηλό υπολογιστικό κόστος και προσιτή υποδομή.

Στις ακόλουθες ενότητες προσεγγίζεται η ασφάλεια στις νέες τεχνολογίες και έλεγχος ταυτότητας στα δίκτυα 5G με τη βοήθεια των τεχνικών της μηχανικής μάθησης.

5.1 Μηχανική μάθηση και ασφάλεια στις νέες τεχνολογίες

Στο ακόλουθο σχήμα συνοψίζονται οι διάφορες εφαρμογές μηχανικής μάθησης στην ασφάλεια του δικτύου.



Εικόνα 11: Διαφορετικά σενάρια εφαρμογών και περιπτώσεις χρήσης της μηχανικής μάθησης για την ασφάλεια δικτύου (Επεξεργασία από το [7])

Αναλυτικότερα, τα μοντέλα μηχανικής μάθησης μπορούν να χρησιμοποιηθούν για τον εντοπισμό ύποπτων δραστηριοτήτων σε πραγματικό χρόνο, αναλύοντας μοτίβα και παραμέτρους από τη δραστηριότητα του δικτύου. Οι αλγόριθμοι ταξινόμησης μπορούν να χρησιμοποιηθούν για την ανίχνευση ανωμαλιών παρακολουθώντας τις παραμέτρους του δικτύου, όπως αρχεία καταγραφής απόδοσης και σφάλματος δικτύου. Οι αλγόριθμοι ομαδοποίησης μπορούν να χρησιμοποιηθούν για την κατηγοριοποίηση διαφόρων ειδών απειλών και κενών στην ασφάλεια του δικτύου. Οι επιθέσεις στατιστικών συμπερασμάτων και τα γενετικά δίκτυα αντιπαραθέσεων (Generative Adversarial Network - GAN) μπορούν να δημιουργήσουν πλαστά σύνολα δεδομένων για να αναπτύξουν και να αξιολογήσουν νέα μέτρα ασφαλείας, καθώς και να δοκιμάσουν και να εφαρμόσουν εξελιγμένα πρωτόκολλα και αλγόριθμους ασφαλείας.

Στις ακόλουθες ενότητες παρουσιάζονται πιο εξειδικευμένα οι εφαρμογές της μηχανικής μάθησης για τα ζητήματα ασφαλείας των δικτύων 5G, τα οποία παρουσιαστήκαν στο προηγούμενο κεφάλαιο.

5.1.1 Μηχανική μάθηση και ασφάλεια σε massive MIMO

Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται συνήθως για την ανακάλυψη ενός μοτίβου σε υπάρχοντα δεδομένα, την πρόβλεψη τιμών ή την εξαγωγή χαρακτηριστικών, τα οποία είναι πολύ χρήσιμα εργαλεία για τον εντοπισμό ενεργών επιθέσεων από τους αντιπάλους [8]. Εν συντομία, η μηχανική μάθηση θα μπορούσε να βοηθήσει το μαζικό MIMO για την εκτίμηση πιλοτικών επιθέσεων και των εχθρικών καναλιών, όπως αναφέρθηκε και νωρίτερα.

Αναλυτικότερα, στο [69], προτείνεται ένα πιλοτικό σχέδιο κατανομής βασισμένο στη βαθιά μάθηση (Deep Learning based Pilot Allocation Scheme - DL-PAS) για τα συστήματα μαζικού MIMO. Το προτεινόμενο DL-PAS βελτιώνει την απόδοση σε κυψελοειδή δίκτυα με σοβαρή μόλυνση πιλότου, μαθαίνοντας τη σχέση μεταξύ της ανάθεσης πιλότου και του μοτίβου τοποθεσίας των χρηστών. Χρησιμοποιείται μια μέθοδος μάθησης με επίβλεψη, όπου οι εισαγωγές χαρακτηριστικών είναι οι τοποθεσίες των χρηστών σε όλα τα κελιά και οι ετικέτες εξόδου είναι οι πιλοτικές εργασίες. Συγκεκριμένα, οι προκατασκευασμένες βέλτιστες πιλοτικές εκχωρήσεις με συγκεκριμένες τοποθεσίες των χρηστών παρέχονται μέσω μιας εξαντλητικής μεθόδου αναζήτησης (exhaustive search ή Brute-force search) ως δεδομένα εκπαίδευσης. Στη συνέχεια, το προτεινόμενο DL-PAS παρέχει μια σχεδόν βέλτιστη πιλοτική εκχώρηση από την παραγόμενη συνάρτηση συμπερασμάτων αναλύοντας τα δεδομένα εκπαίδευσης.

Στο [65], παρουσιάζεται η ανίχνευση πιλοτικής επίθεσης για μη ορθογώνια πολλαπλή πρόσβαση (non-orthogonal multiple access– NOMA) σε mmWave μαζικό MIMO 5G. Η NOMA έχει θεωρηθεί ως βασική τεχνολογία στην επικοινωνία 5G. Η ανίχνευση επιθέσεων μόλυνσης πιλότου αντιμετωπίζει νέες προκλήσεις, λόγω των νέων χαρακτηριστικών του NOMA, όπως τα σήματα υπέρθεσης (superposed signals) με πολλούς χρήστες. Το SVM (αλγόριθμος μηχανικής μάθησης) χρησιμοποιήθηκε για την ανίχνευση επιθέσεων πιλοτικής μόλυνσης.

Μια λύση για την ανίχνευση ενεργής επίθεσης μέσω εχθρικών καναλιών είναι γνωστή ως δακτυλικό αποτύπωμα συσκευής (device finger printing). Οι εξαρτώμενες από τη συσκευή ραδιομετρικές όπως οι διαφορές συχνότητας και η αλλαγή φάσης μπορούν να χρησιμοποιηθούν ως μοναδικά δακτυλικά αποτυπώματα [8]. Στο [90] οι συγγραφείς πρότειναν μια μη παραμετρική μέθοδος Bayesian για τον εντοπισμό του αριθμού των συσκευών καθώς και την ταξινόμηση πολλών συσκευών με μάθηση χωρίς επίβλεψη. Οι

συγγραφείς απέδειξαν την αποτελεσματικότητα της μεθόδου κατά των επιθέσεων Sybil και Masquerade χρησιμοποιώντας τόσο προσομοίωση όσο και πειραματικές μετρήσεις.

Μια πρόκληση για τη χρήση αλγορίθμων μηχανικής μάθησης για τη διασφάλιση ενός μαζικού MIMO είναι η υψηλή επιβάρυνση λόγω του μεγάλου όγκου δεδομένων εκπαίδευσης που απαιτούνται από τους αλγόριθμους μηχανικής μάθησης. Αυτό καθίσταται κρίσιμο για μεγάλο αριθμό ροών κεραίας που παράγονται σε ένα τεράστιο σύστημα MIMO. Για παράδειγμα, ένας σταθμός βάσης 64 κεραιών θα απαιτεί ξεχωριστά δεδομένα εκπαίδευσης για κάθε κεραία, δηλαδή 64 φορές περισσότερα δεδομένα και επεξεργασία που απαιτείται από ένα τυπικό σύστημα ανίχνευσης εισβολής. Επομένως, είναι πιο ευάλωτοι σε επιθέσεις μπλοκαρίσματος όταν το σύστημα ήδη υποφέρει από υψηλά γενικά δεδομένα [91].

Στην περίπτωση της μάθησης με επίβλεψη, ένα μαζικό MIMO πρέπει να εκπαιδευτεί με απουσία υποκλοπών. Εάν η απουσία ενός εισβολέα δεν είναι σίγουρη, πρέπει να υιοθετηθεί μια προσέγγιση μάθησης χωρίς επίβλεψη. Επίσης, δεν είναι δυνατό να εντοπιστεί μια νέα επίθεση όταν δεν υπάρχουν σχετικά δεδομένα εκπαίδευσης, με μεθόδους μάθησης με επίβλεψη. Ωστόσο, οι αλγόριθμοι της μάθησης χωρίς επίβλεψη είναι λιγότερο ακριβείς και αξιόπιστοι από αυτούς της μάθησης με επίβλεψη, επειδή τα δεδομένα εισαγωγής δεν είναι γνωστά και επισημαίνονται εκ των προτέρων. Ο χρήστης πρέπει να ερμηνεύει και να επισημαίνει το συγκεντρωτικό αποτέλεσμα της μάθησης χωρίς επίβλεψη. Λόγω της σημασίας των υπηρεσιών ασφαλείας, μια αναξιόπιστη μέθοδος δεν είναι η ιδανική λύση. Επιπλέον, οι αλγόριθμοι μάθησης χωρίς επίβλεψη είναι πιο περίπλοκοι σε σύγκριση με τους αλγόριθμους μάθησης με επίβλεψη [8].

5.1.2 Μηχανική μάθηση και ασφάλεια σε SDN

Η ενσωμάτωση της τεχνολογίας SDN στα δίκτυα 5G, όπως αναλύσαμε νωρίτερα, συνοδεύονται με πολλές προκλήσεις ασφαλείας. Στη βιβλιογραφία εντοπίζεται μεγάλος αριθμός ερευνών που προσεγγίζουν εφαρμογές της μηχανικής μάθησης για την αντιμετώπιση των ζητημάτων ασφαλείας για την τεχνολογία SDN. Για παράδειγμα, τα δέντρα αποφάσεων (decision trees), η κατηγοριοποίηση Bayesian (και συγκεκριμένα naive Bayes) και το SVM αποτελούν αλγόριθμους, οι οποίοι όταν εκπαιδεύονται με ιστορικά δεδομένα, έχουν τη δυνατότητα να προβλέπουν και να ταξινομούν με ακρίβεια την κυκλοφορία δικτύου σε πραγματικό χρόνο, όπως DDoS και μη κακόβουλη κυκλοφορία [92].

Στο [93] έδειξαν πως το SDN μπορεί να επωφεληθεί από τεχνικές μάθησης με ενίσχυση. Συγκεκριμένα, υλοποιήθηκε η συλλογή μετρήσεων δικτύου και η ομαδοποίησή τους σε προφίλ, καθεμία από τις οποίες έχει ένα σύνολο ενεργειών που χειρίζονται προβλήματα με τη

χρήση της μάθησης με ενίσχυση, την NFV και έναν ελεγκτή SDN. Ουσιαστικά, το σύστημα που αναπτύχθηκε βασίζεται σε έναν πράκτορα που βρίσκεται στην αρχιτεκτονική NFV και συλλέγει αποδείξεις ανωμαλιών από μετρήσεις δικτύου, δημιουργώντας τα προφίλ δικτύου. Τα προφίλ δικτύου προσπαθούν να ιεραρχήσουν την αντιμετώπιση διαφορετικών απειλών, διατηρώντας τον πράκτορα επικεντρωμένο στην εξάλειψη πρώτα των πιο σημαντικών ανωμαλιών. Παράλληλα με αυτή την ενέργεια Μια τέτοια ιεραρχία βοηθά επίσης να αποφευχθεί ότι μια ενέργεια που αποσκοπεί στην εξάλειψη μιας απειλής καταλήγει να ακυρώσει τα αποτελέσματα των ενεργειών άλλων προφίλ.

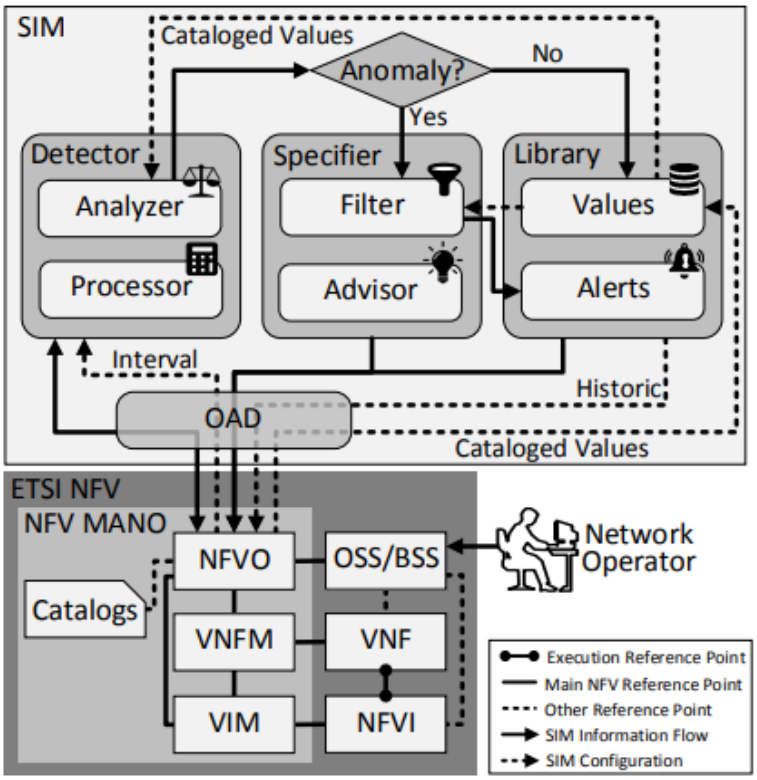
Στο [94] χρησιμοποιήθηκαν αυτό-οργανωμένοι χάρτες (self-organizing maps - SOM) για την ανίχνευση των επιθέσεων DoS και DDoS με βάση τα χαρακτηριστικά ροής κίνησης στο επίπεδο ελέγχου. Επισημαίνεται ότι, οι SOM αποτελούν ένα τύπο νευρωνικού δικτύου που εκπαιδεύεται μέσω μάθησης χωρίς επίβλεψη. Η μέθοδος που παρουσιάστηκε στο [94] χαρακτηρίζεται από τους συγγραφείς ως ελαφριά (light weight) καθώς κατά την εξαγωγή των πληροφοριών η επιβάρυνση του επιπέδου ελέγχου είναι χαμηλότερη σε σχέση με τις παραδοσιακές προσεγγίσεις (όπως, αυτές που παρουσιάστηκαν στο προηγούμενο κεφάλαιο). Άλλα σημαντικά πλεονεκτήματα αυτής της μεθόδου αποτελούν το υψηλό ποσοστό ανίχνευσης επιθέσεων και το πολύ χαμηλό ποσοστό ψευδών συναγερμών που λαμβάνονται από την ανάλυση της ροής με τη χρήση του SOM.

5.1.3 Μηχανική μάθηση και ασφάλεια σε NFV

Τα στοιχεία NFV που σχετίζονται με την μηχανική μάθηση στην αρχιτεκτονική ασφαλείας περιλαμβάνουν τον ελεγκτή ασφαλείας NFV, ο οποίος ενοποιεί πολιτικές ασφαλείας σε ολόκληρο το σύστημα και υπηρεσίες αναλυτικής ασφάλειας, οι οποίες λαμβάνουν απομακρυσμένη παρακολούθηση σε συστήματα NFV και εφαρμόζουν τη μηχανική μάθηση για τον εντοπισμό αναδυόμενων απειλών. Υπάρχουν διάφορες εφαρμογές μηχανικής μάθησης για την ασφάλεια στο NFV, όπως η ανίχνευση ανωμαλιών εντός της κυκλοφορίας ελέγχου και της αλυσίδας υπηρεσιών, η ανάλυση συμπεριφοράς αντιπαράθεσης σε εικονικά honeypots καθώς και ανίχνευση εισβολής βάσει εικονικής μηχανής και κεντρικού υπολογιστή.

Σχετικά, την ανίχνευση ανωμαλιών εντός της κυκλοφορίας ελέγχου και της αλυσίδας υπηρεσιών. Στο [95] προτείνεται ένα πλαίσιο λειτουργιών που επιτρέπει στους ενορχηστρωτές NFV να αναλύουν στοιχεία NFV και να εκτελούν προτεινόμενες ενέργειες με στόχο τη διατήρηση της ακεραιότητας των υπηρεσιών στο δίκτυο. Ο ενορχηστρωτής (orchestrator) NFV (NFVO) είναι ένα βασικό στοιχείο του αρχιτεκτονικού πλαισίου NFV

MANO (διαχείριση λειτουργιών εικονικοποίησης δικτύου και ενορχήστρωση δικτύου), το οποίο βοηθά στην τυποποίηση των λειτουργιών της εικονικής δικτύωσης για την αύξηση της διαλειτουργικότητας των στοιχείων δικτύωσης που καθορίζονται από λογισμικό (SDN). Το NFVO εκτελεί ενορχήστρωση πόρων και ενορχήστρωση υπηρεσιών δικτύου, καθώς και άλλες λειτουργίες. Με πιο απλά λόγια, το NFVO είναι ένα κεντρικό συστατικό μιας λύσης που βασίζεται σε NFV. Συγκεντρώνει διαφορετικές λειτουργίες για να δημιουργήσει μια ενιαία υπηρεσία ενορχήστρωσης που περιλαμβάνει ολόκληρο το πλαίσιο και έχει καλά οργανωμένη χρήση πόρων [96]. Το πλαίσιο που προτείνεται στο [95] βασίζεται στην προσθήκη μιας νέας μονάδας που ονομάζεται SIM (Service Function Chains Integrity Module) στην αρχιτεκτονική NFV MANO, όπως απεικονίζεται στο παρακάτω σχήμα. Το SIM επικοινωνεί απευθείας με το NFVO, χρησιμοποιώντας τυπικά API για να ζητήσει πληροφορίες σχετικά με τα στοιχεία NFV λειτουργίας και για την προώθηση των αποτελεσμάτων της ανίχνευσης ανωμαλιών.



Εικόνα 12: (Service Function Chains Integrity Module) αρχιτεκτονική και εσωτερικά στοιχεία [95]

Η επιλογή για μια συγκεκριμένη τεχνική ανίχνευσης ανωμαλιών εξαρτάται από τα σενάρια δικτύου και τις πληροφορίες που παρακολουθούνται. Τεχνικές που απαιτούν μάθηση με

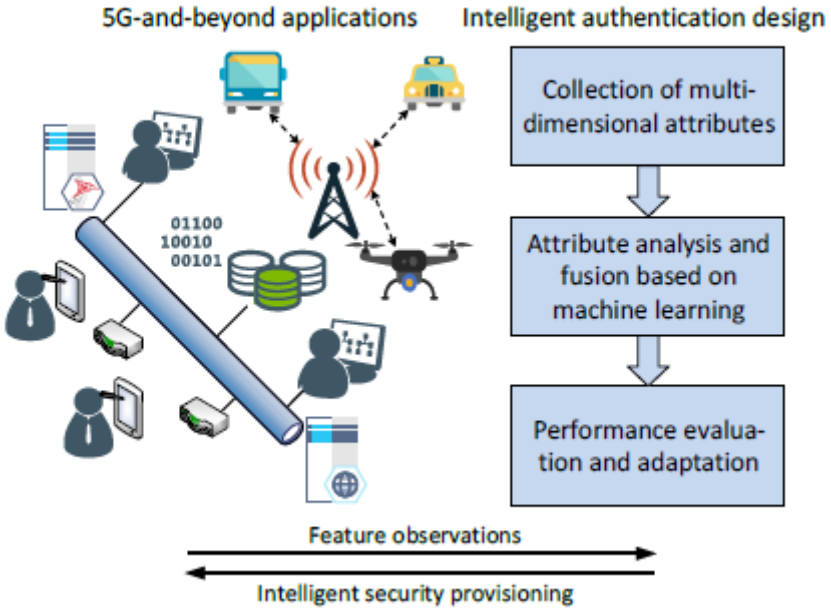
επίβλεψη ή στατιστική μοντελοποίηση σχετικά με τη λειτουργία του δικτύου ενδέχεται να μην είναι κατάλληλες για σενάρια NFV λόγω της δυναμικής τους συμπεριφοράς. Ωστόσο, στο πλαίσιο της ανάλυσης των πληροφοριών, οι τεχνικές μηχανικής μάθησης μπορούν να εφαρμοστούν για να μάθουν με την πάροδο του χρόνου ποια είναι η καλύτερη δράση που πρέπει να εκτελεστεί με βάση το ιστορικό ανιχνευμένων ανωμαλιών. Συμπληρωματικά, μπορούν να ληφθούν λεπτομερή συμπεράσματα, όπως η αποκάλυψη της προέλευσης μιας επίθεσης DoS αναλύοντας τα στοιχεία που λείπουν και την τελευταία πρόσβαση σε αυτά τα στοιχεία. Επίσης, τεχνικές μηχανικής μάθησης μπορούν να εφαρμοστούν στον προσδιοριστή (specificer) του προτεινόμενου πλαισίου, ο οποίος προσδιορίζει τα ανώμαλα στοιχεία και επιλέγει την καταλληλότερη ενέργεια που πρέπει να ληφθεί [95].

Σχετικά με την ανάλυση συμπεριφοράς αντιπαράθεσης σε εικονικά honeypots, στο [97] παρουσιάζεται ένα πλαίσιο, το οποίο ονομάζεται Mouseword και συνδυάζει τις τεχνολογίες NFV και SDN για να δημιουργήσει ένα περιβάλλον ικανό να συνδυάζει και να μεταδίδει πραγματική και συνθετική κίνηση, καθώς και να συλλέγει και να επισημαίνει αυτήν την κίνηση προκειμένου να χρησιμοποιηθούν για την εκπαίδευση και την επικύρωση αλγορίθμων μηχανικής μάθησης που θα εφαρμοστούν στην ανίχνευση απειλών για την ασφάλεια στον κυβερνοχώρο. Ο ενορχηστρωτής δικτύου ανοιχτού κώδικα OSM χρησιμοποιείται για τον έλεγχο και τη διαχείριση του πλαισίου και για την ανάπτυξη σεναρίων εκπαίδευσης και επικύρωσης.

Όπως αναφέρθηκε και νωρίτερα, οι τεχνικές μηχανικής μάθησης δεν προτείνονται ευρέως στο πλαίσιο της τεχνολογίας NFV και σε σενάρια κυκλοφορίας δικτύου, καθώς η έλλειψη επισημασμένων δεδομένων καθιστά αδύνατη την χρήση αλγορίθμων μηχανικής μάθησης με επίβλεψη, όπου οι επισημασμένες ροές κυκλοφορίας είναι απαραίτητες για τις διαδικασίες εκπαίδευσης και επικύρωσης. Επιπλέον, οι μη επιτηρούμενοι αλγόριθμοι χρειάζονται επισημασμένα δεδομένα για εγκάρσια επικύρωση με βάση την περιοχή κάτω από το χαρακτηριστικό λειτουργίας του δέκτη. Το Mouseworld είναι το πρώτο πλαίσιο που συνδυάζει τις τεχνολογίες NFV και SDN για την αυτόματη εκπαίδευση και επικύρωση αλγορίθμων που σχετίζονται με την ασφάλεια χρησιμοποιώντας ετικέτες δεδομένων [97]. Βέβαια, το συγκεκριμένο πλαίσιο δεν αναπτύχθηκε πρακτικά και δεν δίνονται περισσότερες λεπτομέρειες σχετικά με την επιλογή συγκεκριμένων αλγορίθμων μηχανικής μάθησης.

5.2 Μηχανική μάθηση και έλεγχος ταυτότητας (authentication)

Στο παρακάτω σχήμα, παρουσιάζουμε το σχεδιασμό προσεγγίσεων έξυπνου ελέγχου ταυτότητας με μηχανική μάθηση χρησιμοποιώντας πολυδιάστατα χαρακτηριστικά και βελτιστοποιώντας την διαδικασία ελέγχου ταυτότητας.



Εικόνα 13: Διάγραμμα πλαισίου ευφυούς σχεδιασμού ελέγχου ταυτότητας [6]

Στην πρώτη φάση γίνεται η συλλογή των χρονικά μεταβαλλόμενων πολυδιάστατων χαρακτηριστικών για έλεγχο ταυτότητας, ο οποίος μπορεί να εκτιμηθεί ότι δεν έχει θορύβους και σφάλματα μέτρησης [6]. Στα παραδείγματα περιλαμβάνονται τα χαρακτηριστικά φυσικού επιπέδου, η επιλογή δικτύου σε ετερογενή ασύρματα δίκτυα και τα μοτίβα κινητικότητας. Σε ένα συγκεκριμένο σενάριο ασύρματης επικοινωνίας 5G, αυτά τα χαρακτηριστικά που παρέχουν περισσότερες πληροφορίες για έλεγχο ταυτότητας μπορούν να επιλεγούν πρώτα. Αναλυτικά, τα ανεξάρτητα χαρακτηριστικά που έχουν ευρύτερο φάσμα διανομής και μεγαλύτερη ακρίβεια εκτίμησης θα μπορούσαν να προσφέρουν περισσότερες πληροφορίες για τη διάκριση διαφορετικών πομπών. Χρησιμοποιώντας πολυδιάστατα χαρακτηριστικά καθώς και κοινή χρήση πληροφοριών μεταξύ διαφορετικών επιπέδων και δικτύων, βελτιώνεται η αξιοπιστία του ελέγχου ταυτότητας. Ο σχεδιασμός του έξυπνου ελέγχου ταυτότητας βασίζεται μόνο στα δεδομένα εκτίμησης των χαρακτηριστικών χωρίς να

απαιτείται ακριβής δομή των χρονικών μεταβλητών χαρακτηριστικών, όπως για παράδειγμα του μοντέλου καναλιού, με αποτέλεσμα την επικύρωση συσκευής χωρίς μοντέλο.

Στη δεύτερη φάση τα πολυδιάστατα χαρακτηριστικά μπορούν να συγχωνευτούν για έλεγχο ταυτότητας βάσει τεχνικών μηχανικής μάθησης [6]. Ένα παράδειγμα δίνεται στο [64], το οποίο είναι ένα πρόγραμμα ελέγχου ταυτότητας φυσικού επιπέδου βασισμένο σε μηχανική μάθηση πυρήνα. *(Περισσότερα παραδείγματα μηχανικής μάθησης για έξυπνο έλεγχο ταυτότητας παρουσιάζονται στην επόμενη ενότητα).* Λαμβάνοντας υπόψη τις χρονικές μεταβολές των συνθηκών του δικτύου, όπως οι περιορισμοί πόρων και οι αβεβαιότητες, τα χαρακτηριστικά μπορούν να επιλεγούν ευκαιριακά για την ταυτόχρονη αντιμετώπιση τόσο των γενικών επικοινωνιών όσο και της διαχείρισης ασφάλειας. Επιπλέον, η ανάπτυξη ενός κατάλληλου αλγορίθμου μηχανικής μάθησης και η μείωση της διάστασης του συστήματος ελέγχου ταυτότητας ωφελούν επίσης την απόδοση της επικοινωνίας, έτσι θα επιτευχθεί οικονομικός έλεγχος ταυτότητας.

Στην τρίτη φάση γίνεται ο έλεγχος ταυτότητας των χρηστών για να ελεγχθεί αν είναι νόμιμος ή παράνομος, δηλαδή στο παράδειγμα που παρουσιάστηκε νωρίτερα, γίνεται ο έλεγχος ταυτότητας της Alice και του Spoofer, ο οποίος μπορεί να πραγματοποιηθεί με βάση τη νέα συλλογή των πολυδιάστατων χαρακτηριστικών. Για να επιτευχθεί αυτό, το μοντέλο παλινδρόμησης (regression) ή ταξινόμησης (classification) θα πρέπει να βασίζεται στα δεδομένα εκπαίδευσης που συλλέγονται από τηνAlice και τον Spoofer. Στη συνέχεια, μπορεί να γίνει αξιολόγηση της απόδοσης του ελέγχου ταυτότητας και η διαδικασία του ελέγχου ταυτότητας να προσαρμοστεί στο περίπλοκο χρονικά μεταβαλλόμενο περιβάλλον με την εξερεύνηση της μηχανικής εκμάθησης για την παρακολούθηση των παραλλαγών των πολυδιάστατων χαρακτηριστικών [6].

5.2.1 Βιομετρικός έλεγχος ταυτότητας με μηχανική μάθηση

Οι παραδοσιακές μορφές ελέγχου ταυτότητας (authentication), όπως αυτές που αναφέρθηκαν στο προηγούμενο κεφαλαίο, δεν έχουν τη δυνατότητα να ανταποκριθούν στις ανάγκες των χρηστών στη σύγχρονη εποχή του διαδικτύου. Συχνά, οι χρήστες χρησιμοποιούν εύκολους κωδικούς, ώστε, πρακτικά, να μπορούν να τους θυμούνται εύκολα, καθώς πρέπει να θυμούνται μεγάλο αριθμό κωδικών. Αυτό έχει ως αποτέλεσμα, οι κωδικοί αυτοί να είναι ευάλωτοι σε επιθέσεις από άτομα χωρίς εξουσιοδότηση. Για την αντιμετώπιση αυτού του ζητήματος, τις τελευταίες δεκαετίες έχουν προταθεί διάφορες μέθοδοι ελέγχου ταυτότητας, οι οποίες βασίζονται στα βιομετρικά χαρακτηριστικά των χρηστών, τα οποία είναι φυσικά χαρακτηριστικά του ανθρώπινου σώματος, όπως, τα δακτυλικά αποτυπώματα, η ίριδα του

ματιού, το πρόσωπο, η παλάμη, όπως περιγράφηκαν στο προηγούμενο κεφάλαιο. Ωστόσο, υπάρχουν και άλλες βιομετρικές μέθοδοι ελέγχου ταυτότητας, οι οποίες δεν είναι μοναδικές, αλλά με τη βοήθεια της μηχανικής μάθησης μπορούν να χρησιμοποιηθούν ως μέθοδοι ελέγχου ταυτότητας, όπως το σήμα της καρδιάς, ο τρόπος γραφής ή η υπογραφή [117].

Ο τρόπος γραφής ή ο τρόπος πληκτρολόγησης διαφέρει από άνθρωπο σε άνθρωπο. Όμως, αυτό το βιομετρικό χαρακτηριστικό δεν είναι μοναδικό και επομένως δεν αποτελεί από μόνο του μια μέθοδο ελέγχου ταυτότητας. Ωστόσο, μέσω συνεχής παρακολούθησης του τρόπου γραφής ή πληκτρολόγησης ενός ατόμου μπορεί να χρησιμοποιηθεί για τον έλεγχο της ταυτότητας του χρήστη [117]. Ουσιαστικά, πρόκειται για μια συμπεριφορική μέτρηση μέσω της οποίας αναγνωρίζεται ένας χρήστης βάσει των χαρακτηριστικών της πληκτρολόγηση του [118]. Δηλαδή, όταν ένας χρήστης πληκτρολογεί αναγνωρίζονται συγκεκριμένα μοτίβα, όπως ο χρόνος ανάμεσα στις διαδοχικές πιέσεις των πλήκτρων, η διάρκειας πίεσης του κάθε πλήκτρου, η πίεση που εφαρμόζεται σε κάθε πλήκτρο και η τοποθέτηση των δακτύλων μπορούν να μετρηθούν, ώστε να δημιουργηθεί ένα μοναδικό προφίλ (μοναδική υπογραφή) πληκτρολόγησης του χρήστη [117][119].

Στο [119] παρουσιάζονται οι μέθοδοι ανάλυσης δεδομένων που μπορούν να χρησιμοποιηθούν στη δυναμική πληκτρολόγηση. Αυτές οι μέθοδοι ανάλυσης δεδομένων περιλαμβάνουν την αναγνώριση προτύπων (pattern recognition), τις μεθόδους μέθοδοι εντοπισμού εξωκείμενων τιμών (Outlier Detection και Novelty Detection) και One-Class SVM, τη συνάρτηση Gaussian Mixture Model (GMM), τη μέθοδο ανάλυσης των κύριων συστατικών. Πιο συγκεκριμένα, η αναγνώριση προτύπων αποτελεί ένα τομέα της μηχανικής μάθηση που βασίζεται στην επαναληψιμότητα των δεδομένων. Συχνά, θεωρείται συνώνυμο της μηχανικής μάθησης. Στις περισσότερες περιπτώσεις, ένα σύστημα αναγνώρισης προτύπων εκπαιδεύεται από δεδομένα με ετικέτες, δηλαδή πρόκειται για μάθηση με επίβλεψη, όμως όταν αυτά τα δεδομένα δεν είναι διαθέσιμα μπορεί να χρησιμοποιηθεί μάθηση χωρίς επίβλεψη, ώστε να αναγνωριστούν τα πρότυπα. Οι μέθοδοι εντοπισμού εξωκείμενων τιμών χρησιμοποιούνται για τη λήψη αποφάσεων σχετικά με μια νέα παρατήρηση. Στο [119] χρησιμοποιήθηκε η μέθοδος novelty detection για τον εντοπισμό των ανωμαλιών σε νέες παρατηρήσεις για δεδομένα τα οποία δεν έχουν μολυνθεί από εξωκείμενες τιμές, δηλαδή, οι παρατηρήσεις που ξεπερνούν τα όρια των ακραίων τιμών (π.χ. min και max). Για την εφαρμογή της μεθόδου novelty detection η συνάρτηση One-Class SVM αποτελεί μια διαδομένη τεχνική. Για τον εντοπισμό των εξωκείμενων τιμών χρησιμοποιήθηκε το μοντέλο Gaussian Mixture, το οποίο αποτελεί μια παραμετρική συνάρτηση πυκνότητας έχει χρησιμοποιηθεί για την αναπαράσταση χαρακτηριστικών σε

βιομετρικά συστήματα ειδικά σε αναγνώριση φωνής ή ήχων [120]. Τέλος, στο σύστημα που παρουσιάζεται στο [119] χρησιμοποιήθηκε η μέθοδος PCA (Principal Component Analysis) για τη μείωση των διαστάσεων των δεδομένων από 3 σε 2. Ενδεικτικά, αναφέρεται ότι η μέθοδος PCA βασίζεται στην στατιστική και εφαρμόζει ορθογωνικό μετασχηματισμό για τη μετατροπή ενός συνόλου συσχετισμένων παρατηρήσεων σε ένα σύνολο γραμμικών ασυσχέτιστων τιμών (principal components).

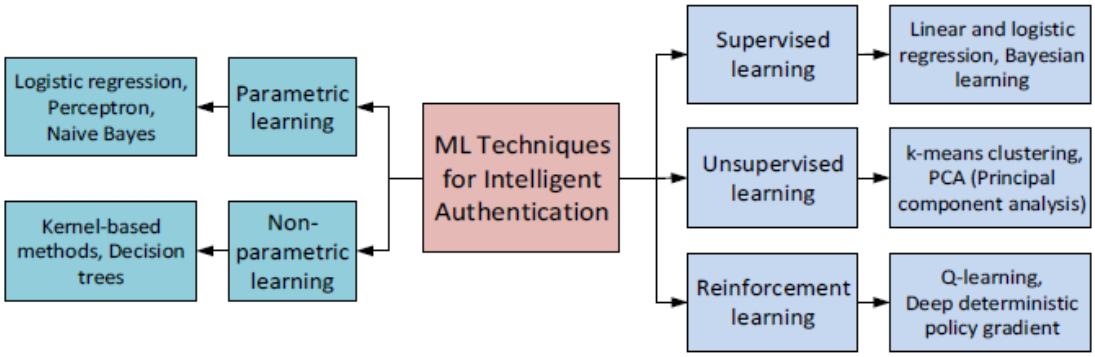
Άλλες τεχνικές της μηχανικής που έχουν χρησιμοποιηθεί για τον έλεγχο ταυτότητας με πληκτρολόγηση αποτελούν τα νευρωνικά δίκτυα. Πιο συγκεκριμένα στο [121] εφαρμόστηκε ο αλγόριθμος Perceptron, ο οποίες παρέχει γραμμικές συναρτήσεις απόφασης και χρησιμοποιήθηκε για την ταξινόμηση των χρηστών. Μέσω της εφαρμογής αυτού του αλγορίθμου επιτεύχθηκε σφάλμα 2%.

Στο [122] χρησιμοποιήθηκε το μοντέλο Bacκ Propagation. Στη μηχανική μάθηση, το back propagation είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος για την εκπαίδευση νευρωνικών δικτύων. Κατά την τοποθέτηση ενός νευρωνικού δικτύου, το back propagation υπολογίζει τη διαβάθμιση της λειτουργίας απώλειας σε σχέση με τα βάρη του δικτύου για ένα μόνο παράδειγμα εισόδου-εξόδου [123]. Η εφαρμογή αυτού του αλγορίθμου για την μελέτη της πληκτρολόγησης του χρήστη στο [122] έδειξε σημαντική δυναμική πρόβλεψης. Στο [124] εφαρμόστηκε επίσης ο αλγόριθμος back propagation σε νευρωνικό δίκτυο για την αναγνώριση ενός έγκυρου χρήστη ή όχι σύμφωνα με τον τρόπο πληκτρολόγηση του κωδικού πρόσβασης (password).

Παρόμοια, τεχνικές μηχανικής μάθησης μπορούν να εφαρμοστούν και για άλλες μεθόδους βιομετρικού ελέγχου ταυτότητας, όπως για παράδειγμα, μετρήσεις από τον τρόπο χρήσης του ποντικιού, όπως η αδράνεια, η κατανομή των θέσεων του δείκτη στην οθόνη, οι αποστάσεις και οι κατευθύνσεις της κίνησης, ο χρόνος του διπλού κλικ, και άλλα. Επίσης, παρόμοιες τεχνικές μπορούν να χρησιμοποιηθούν για την παρατήρηση της συμπεριφοράς του χρήστη κατά την χρήση ενός συγκεκριμένο λογισμικού, όπως για παράδειγμα, κατά τη σύνταξη ενός μηνύματος ηλεκτρονικού ταχυδρομείου, η αλληλεπίδραση κατά την περιήγηση στο διαδίκτυο μέσω ενός σελιδομετρητή (browser) και άλλα [119].

5.2.2 Παραδείγματα μηχανικής μάθησης και ελέγχου ταυτότητας

Οι αλγόριθμοι μηχανικής μάθησης και τα παραδείγματα εφαρμογής τους για τον έλεγχο ταυτότητας μπορούν να κατηγοριοποιηθούν σε παραμετρική και μη παραμετρική μάθηση, μάθηση με ενίσχυση, μάθησης με και χωρίς επίβλεψη, όπως παρουσιάζεται στο ακόλουθο σχήμα.



Εικόνα 14: Κατηγορίες για έξυπνο έλεγχο ταυτότητας με μηχανική μάθηση [6]

Επισημαίνεται ότι, οι παραμετρικές και οι μη παραμετρικές μέθοδοι μάθησης υποδηλώνουν εάν υπάρχουν συγκεκριμένες μορφές των λειτουργιών εκπαίδευσης, ενώ η μάθηση με και χωρίς επίβλεψη υποδεικνύει εάν υπάρχουν επισημασμένα δείγματα στη βάση δεδομένων.

5.2.2.1 Παραμετρικές μέθοδοι μάθησης

Οι παραμετρικές μέθοδοι μάθησης απαιτούν συγκεκριμένη μορφή εκπαιδευτικών λειτουργιών. Παραδείγματα τέτοιων μεθόδων αποτελούν η λογιστική παλινδρόμηση (logistic regression), η γραμμική διακριτή ανάλυση (linear discriminant analysis), η perceptron και οι αλγόριθμοι Naive Bayes [98]. Όταν οι λειτουργίες εκπαίδευσης, που σχετίζονται με τα δείγματα εκπαίδευσης (δηλαδή, τη συλλογή πολυδιάστατων χαρακτηριστικών), επιλέγονται κατάλληλα, οι παραμετρικές μέθοδοι μάθησης μπορούν να είναι πιο ακριβείς, απλούστερες και απαιτούν λιγότερα δεδομένα εκπαίδευσης από τις μη παραμετρικές μεθόδους μάθησης [6]. Για παράδειγμα, στο [99] προτείνεται ένα σύστημα ελέγχου ταυτότητας που βασίζεται σε λογιστική παλινδρόμηση, χρησιμοποιώντας το RSSI των πομπών για τη βελτίωση του ελέγχου ταυτότητας, υποθέτοντας ότι όλοι οι ραδιοκόμβοι είναι στατικοί.

Στα έξυπνα σχήματα ελέγχου ταυτότητας, οι παραμετρικές μέθοδοι μάθησης θα μπορούσαν να μοντελοποιήσουν τα χαρακτηριστικά ανεξάρτητα με βάση τη συγκεκριμένη μορφή λειτουργιών εκπαίδευσης, έτσι ώστε να μπορούν να παρακαμφθούν οι αβεβαιότητες που προκαλούνται από το πολύπλοκο χρονικά μεταβαλλόμενο περιβάλλον, καθώς και τα γενικά έξοδα και την πολυπλοκότητα της εκπαίδευσης μπορούν να προσαρμοστούν. Ωστόσο, αυτό το είδος μεθόδων μάθησης μπορεί να είναι περιορισμένο σε σενάρια ασύρματης επικοινωνίας 5G για τα οποία οι στατιστικές ιδιότητες και η προηγούμενη γνώση των χαρακτηριστικών δεν είναι διαθέσιμες [6].

5.2.2.2 Μη παραμετρικές μέθοδοι μάθησης

Σε αντίθεση με τις παραμετρικές μεθόδους μάθησης, οι μη παραμετρικές μέθοδοι μάθησης δεν καθορίζονται εκ των προτέρων, αλλά καθορίζονται από τα διαθέσιμα δεδομένα. Μερικά παραδείγματα περιλαμβάνουν του αλγορίθμους kernel estimator, k-πλησιέστερους γείτονες (k-nearest neighbors) και δέντρα αποφάσεων (decision trees) [6]. Ένα σχέδιο βασισμένο σε kernel για έξυπνη διαδικασία ελέγχου ταυτότητας φυσικού επιπέδου προτείνεται στο [64] για τη σύντηξη πολλαπλών χαρακτηριστικών φυσικού επιπέδου χωρίς να απαιτείται η γνώση στατικών ιδιοτήτων. Η διάσταση ενός συστήματος ελέγχου ταυτότητας πολλαπλών χαρακτηριστικών μειώνεται από τη λειτουργία πυρήνα και η προκύπτουσα διαδικασία ελέγχου ταυτότητας μπορεί να μοντελοποιηθεί ως γραμμικό σύστημα, μειώνοντας έτσι την πολυπλοκότητα υπολογισμού της διαδικασίας ελέγχου ταυτότητας, παρόλο που χρησιμοποιείται μεγάλος αριθμός χαρακτηριστικών. Το πιο σημαντικό, ο προτεινόμενος αλγόριθμος μάθησης kernel παρακολουθεί τα χαρακτηριστικά που ποικίλλουν από το χρόνο για να βελτιώσει την ασφάλεια με βάση τη συνεχή επικύρωση της συσκευής [64].

Οι μη παραμετρικές μέθοδοι μάθησης εκπαιδεύονται δυναμικά από διαφορετικά χρονικά περιβάλλοντα χωρίς να απαιτούνται υποθέσεις σχετικά με τα εκπαιδευτικά μοντέλα. Το πλεονέκτημα αυτών των μεθόδων είναι ότι παρέχουν μεγαλύτερη ευελιξία για τον έξυπνο έλεγχο ταυτότητας, ειδικά σε αυτά τα σενάρια ελέγχου ταυτότητας σε πραγματικό χρόνο, όπου οι υπολογιστικοί πόροι και ο διαθέσιμος χρόνος για τη λήψη των στατιστικών ιδιοτήτων των χαρακτηριστικών και των λειτουργιών εκπαίδευσης είναι περιορισμένοι. Ωστόσο, σε σύγκριση με τις παραμετρικές μεθόδους μάθησης, απαιτούν περισσότερα δεδομένα εκπαίδευσης (δηλαδή, συλλογή πολυδιάστατων πληροφοριών ή / και τις αντίστοιχες ετικέτες τους) [6].

5.2.2.3 Αλγόριθμοι μάθησης με επίβλεψη

Στο [98] προτείνονται αλγόριθμοι μηχανικής μάθησης με επίβλεψη για την ανίχνευση επιθέσεων έγχυσης ψευδών δεδομένων (false data injection attacks). Οι αλγόριθμοι που χρησιμοποιήθηκαν ήταν ο perceptron (ο οποίος στα ελληνικά καλείται «αντίληπτρο») και πρόκειται για ένα γραμμικό δυαδικό ταξινομητή, ο αλγόριθμος κ-κοντινότερων γειτόνων (k-nearest neighbors) που πρόκειται για ένα αλγόριθμο κατηγοριοποίησης και οι διανυσματικές υποστηρικτικές μηχανές (SVM). Οι αλγόριθμοι αυτοί χρησιμοποιούνται για να ταξινομήσουν τις μετρήσεις ως ασφαλείς ή ως επιθέσεις [98].

5.2.2.4 Αλγόριθμοι μάθησης χωρίς επίβλεψη

Για την προσέγγιση και κατανόηση των μεθόδων μάθησης χωρίς επίβλεψη στον έλεγχο ταυτότητας φυσικού επιπέδου χρησιμοποιείται ξανά το παράδειγμα του Bob και της Alice που αναφέρθηκε και νωρίτερα. Ειδικότερα, όταν ο Bob, ο οποίος είναι ο εκπαιδευόμενος, λαμβάνει αποκλειστικά δεδομένα εκπαίδευσης χωρίς ετικέτα, χρησιμοποιώντας τον αλγόριθμο ομαδοποίησης K-means [6].

Η ομαδοποίηση είναι ένα κοινό πρόβλημα σε δίκτυα 5G, ειδικά σε ετερογενή σενάρια που σχετίζονται με διαφορετικά μεγέθη κυψελών (cells) και σε επικοινωνίες D2D. Για παράδειγμα, τα small cells πρέπει να συγκεντρωθούν προσεκτικά για να αποφευχθούν παρεμβολές χρησιμοποιώντας συντονισμένη μετάδοση πολλαπλών σημείων (Coordinated Multi-Point - CoMP) [100]. Στο [101] για τον διαχωρισμό των σημείων πρόσβασης πλέγματος (Mess Access Points - MAPs) σε διάφορες ομάδες. Στο προτεινόμενο σχήμα, η λειτουργία ξεκινάει από ένα αρχικό σημείο πρόσβασης πύλης (Gateway Access Point–GAP), το οποίο μπορεί να προσδιοριστεί με μια τυχαία επιλογή από τα σημεία πρόσβασης πλέγματος (MAPs) ή μπορεί να προσδιοριστεί (πιο έξυπνα) με ένα κριτήριο αρχικοποίησης. Στη συνέχεια, κάθε MAP αντιστοιχίζεται στο πλησιέστερο GAP του. Εάν βρίσκεται κοντά σε πολλά κατάλληλα GAP, τότε επιλέγεται το GAP που διαθέτει άμεσα ένα εικονικό κανάλι. Τέλος, χρησιμοποιώντας έναν απλό αλγόριθμο K-means, οι MAP χωρίζονται σε ομάδες k που σχετίζονται με τα πλησιέστερα GAP[101].

5.2.2.5 Αλγόριθμοι μάθησης με ενίσχυση

Η μάθηση με ενίσχυση δεν απαιτεί ακριβείς εισόδους και εξόδους και ακριβείς ενημερώσεις παραμέτρων. Στο [99] διερευνήθηκε ο έλεγχος ταυτότητας φυσικού επιπέδου που εκμεταλλεύεται πληροφορίες ραδιοφωνικού καναλιού (όπως οι ενδείξεις της έντασης του σήματος που λαμβάνονται) για να ανιχνεύσει επιθέσεις πλαστογραφίας σε ασύρματα δίκτυα. Το πρόγραμμα ελέγχου ταυτότητας που προτείνεται βασίζεται στον αλγόριθμο Q-learning, όπου χρησιμοποιεί τους δείκτες ισχύος των λαμβανόμενων σημάτων RSSI για την βελτίωση της ακρίβειας του ελέγχου ταυτότητας και την ανίχνευση της πλαστογράφησης. Τόσο η προσομοίωση όσο και τα πειραματικά αποτελέσματα έχουν επικυρώσει την αποτελεσματικότητα των προτεινόμενων στρατηγικών. Ωστόσο, όπως αναφέρεται στο [6], αυτό το σχήμα είναι στατικό ως προς τον έλεγχο ταυτότητας και μπορεί να λειτουργήσει μόνο όταν οι διαθέσιμη πόροι και ο χρόνος είναι περιορισμένοι.

Η βελτίωση του ελέγχου ταυτότητας μπορεί να επιτευχθεί με τη χρήση πολυδιάστατων χαρακτηριστικών σε υψηλότερα επίπεδα από το φυσικό, όπως για παράδειγμα στο επίπεδο

εφαρμογής, καθώς και με τη διερεύνηση ακριβέστερων πληροφοριών για τον χρήστη που αποστέλλει μηνύματα (δηλαδή, η Alice). Όμως, σε αυτή την περίπτωση, είναι πολύ πιθανό το απόρρητο της να μην είναι εγγυημένο. Με πιο απλά λόγια, ο Bob μπορεί να συλλέξει τις πληροφορίες των χαρακτηριστικών της Alice, όπως συμπεριφορά χρήστη και λειτουργίες που σχετίζονται με την τοποθεσία, για την ανάλυση των συνηθειών της, των τοποθεσιών της και άλλων ευαίσθητων πληροφοριών κατά τη διαδικασία ελέγχου ταυτότητας. Αυτό, όμως, οδηγεί σε διαρροή απορρήτου [102].

Επομένως, η ανάπτυξη ενός συστήματος ελέγχου ταυτότητας για τη διατήρηση του απορρήτου που βασίζεται σε μεθόδους κάλυψης (masking methods) [103] είναι χρήσιμη για την προστασία των ιδιωτικών πληροφοριών της Alice κατά τη διαδικασία ελέγχου ταυτότητας με την προσθήκη μοτίβων συσκότισης (obfuscation patterns) στις μετρήσεις χαρακτηριστικών [6]. Ωστόσο, υπάρχει μια ανταλλαγή μεταξύ του απορρήτου και του βοηθητικού προγράμματος δεδομένων (data utility) [103], που έχει ως αποτέλεσμα την αντιστάθμιση ανάμεσα στη διατήρηση του απορρήτου και την απόδοση στον έλεγχο ταυτότητας. Πιο συγκεκριμένα, με την προσθήκη μοτίβων συσκότισης στις μετρήσεις των χαρακτηριστικών, οι πληροφορίες θα κυκλοφορήσουν για να διατηρηθεί το απόρρητο της Alice, ενώ η απόδοση του ελέγχου ταυτότητας θα μειωθεί, καθώς ο οποίος βασίζεται στη μειωμένη ακρίβεια των χαρακτηριστικών [6].

Στο [6] προτείνεται μια καλύτερη αντιστάθμιση για τη διατήρηση του απορρήτου και την απόδοση του ελέγχου ταυτότητας. Πρόκειται, για μια τεχνική προστασίας απορρήτου που βασίζεται στην κάλυψη (mask) με χωρική και χρονική συνάθροιση μετρήσεων χαρακτηριστικών για έξυπνο έλεγχο ταυτότητας. Αυτή η τεχνική δημιουργεί και συγκεντρώνει μοτίβα συσκότισης για κάθε χαρακτηριστικό της Alice σε συγκεκριμένο χρονικό διάστημα t ως $\mathbf{H}[t]=H[t]+\gamma[t]$, όπου $\mathbf{H}=(H_1, H_2, \ldots, H_N)^T$ και $\gamma=(\gamma_1, \gamma_2, \ldots, \gamma_N)^T$, αντιπροσωπεύουν τις εκτιμήσεις των χαρακτηριστικών N που χρησιμοποιήθηκαν και δημιούργησαν μοτίβα συσκότισης, αντίστοιχα. Το πιο σημαντικό είναι ότι τα μοτίβα συσκότισης δημιουργούνται για να διασφαλιστεί ότι η σύντηξη του H [t] για έλεγχο ταυτότητας με μηχανική μάθηση θα πρέπει να είναι κοντά στη σύντηξη των αρχικών εκτιμήσεων χαρακτηριστικών H [t]. Για να επιτευχθεί αυτό, απαιτείται ανταλλαγή πληροφοριών μεταξύ διαφορετικών επιπέδων αρχιτεκτονικής δικτύου. Όταν ένα χαρακτηριστικό δεν είναι ευαίσθητο στην προστασία της ιδιωτικής ζωής για την Αλίκη, το μοτίβο συσκότισης που βρίσκεται πάνω σε αυτό το χαρακτηριστικό πρέπει να οριστεί στο μηδέν, έτσι ώστε να διασφαλιστεί η χωρητικότητα του συστήματος. Συμπερασματικά, μέσω της τεχνικής προστασίας της ιδιωτικής ζωής με βάση τη μάσκα με τη χωρική και τη χρονική

συνάθροιση, κάθε ευαίσθητο χαρακτηριστικό της Alice έχει όλες τις ιδιότητες των τυχαίων σειρών αριθμών, επομένως δεν αποκαλύπτει καμία πληροφορία στον Bob.

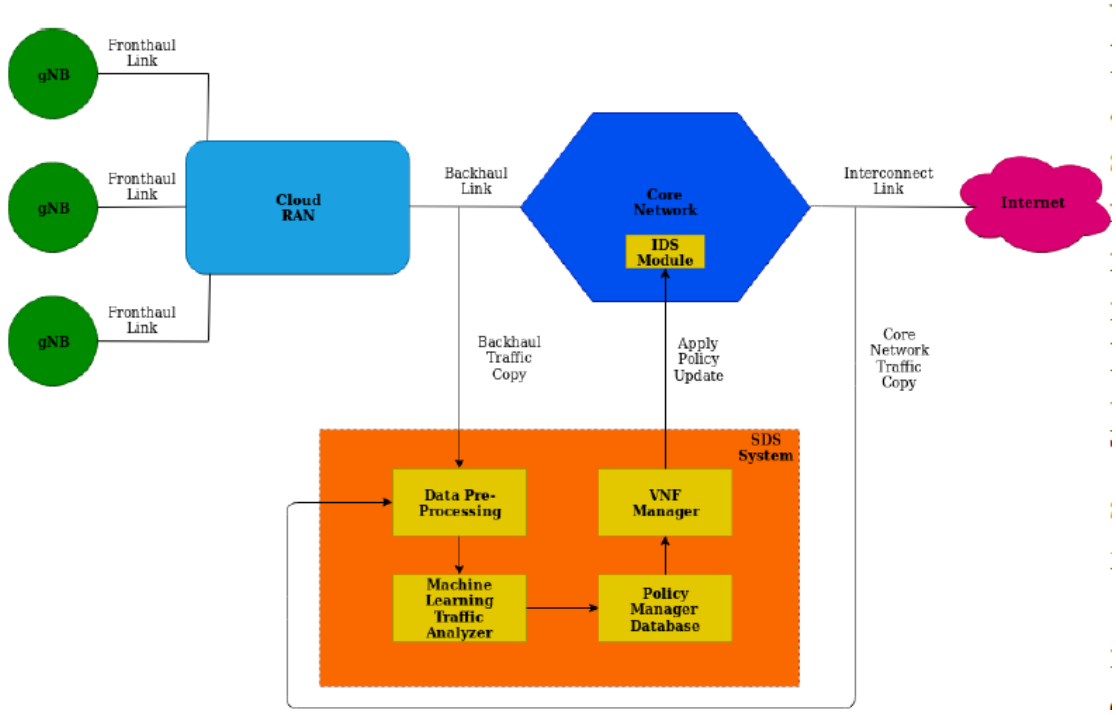
6 *Μελέτες περιπτώσεων*

6.1 Μηχανική μάθηση για ανίχνευση ανωμαλιών σε δίκτυο 5G

Στο [10] προτείνεται ένα πλαίσιο ασφάλειας που καθορίζεται από λογισμικό (Software Defined Security – SDS) ως μέσο για την παροχή ενός αυτοματοποιημένου, ευέλικτου και κλιμακούμενου συστήματος άμυνας δικτύου. Στο SDS χρησιμοποιήθηκαν οι τρέχουσες εξελίξεις της μηχανικής μάθησης για να σχεδιαστεί ένα συνελικτικό νευρωνικό δίκτυο (Convolutional Neural Network - CNN) χρησιμοποιώντας την νευρωνική αρχιτεκτονική αναζήτησης (Neural Architecture Search – NAS) για την ανίχνευση της ανώμαλης κίνησης στο δίκτυο.

6.1.1 Αρχιτεκτονική SDS για δίκτυα 5G

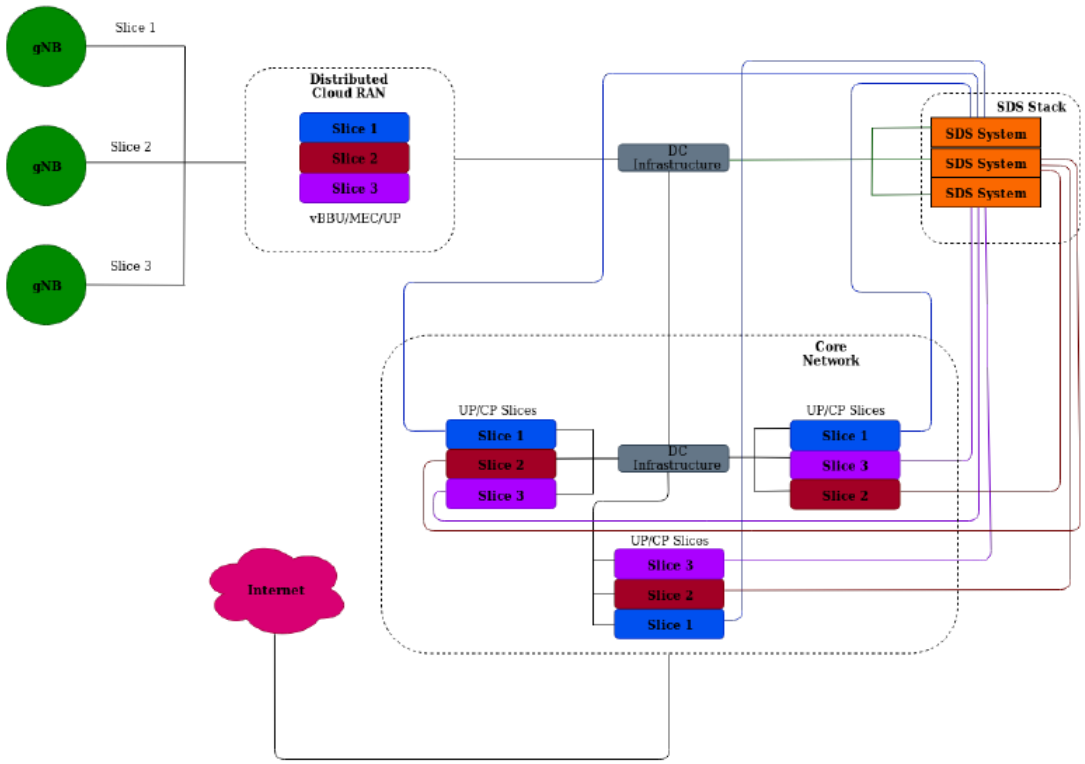
Τα δίκτυα 5G και τα κύρια στοιχεία τους, όπως το νέφος δικτύου Ραδιοπρόσβασης (C-RAN) και το κεντρικό δίκτυο εικονικοποιούνται, επομένως, ορίζονται πλήρως μέσω λογισμικού. Μια παρόμοια προσέγγιση μπορεί να ακολουθηθεί για την εφαρμογή ενός αυτοματοποιημένου συστήματος ασφαλείας μέσω SDS. Το σχήμα παρακάτω δείχνει μια πιθανή εφαρμογή ενός συστήματος SDS σε ένα δίκτυο 5G.



Εικόνα 15: Δίκτυο 5G με ασφάλεια που καθορίζεται από λογισμικό (SDS) [10]

Σύμφωνα με το παραπάνω σχήμα ένα αντίγραφο με επαρκή ποσότητα κίνησης, η οποία προέρχεται από το σύνδεσμο (link) backhaul και από το σύνδεσμο του βασικού δικτύου (core network) μπορεί να αναλυθεί και να παρέχει ανίχνευση ανωμαλιών σε ένα δίκτυο από άκρο σε άκρο. Λαμβάνεται ένα αντίγραφο δεδομένων για ανάλυση και για τη δημιουργία προφίλ καθορισμού καλοήθους (benign) και ανώμαλης (anomalous) κίνησης για το μοντέλο, επίσης με την αντιγραφή δεδομένων δεν θα υπάρξουν επιπτώσεις στην απόδοση του δικτύου ενώ το μοντέλο αναλύει τα δεδομένα. Στη συνέχεια, τα δεδομένα προ-επεξεργάζονται για να είναι σε μορφή κατάλληλη για το μοντέλο μηχανικής μάθησης και αναλύονται για ανωμαλίες, οι οποίες αποθηκεύονται στη βάση δεδομένων του διαχειριστή πολιτικής (policy manager data base) με τις αντίστοιχες δυνατότητες επισκεψιμότητας. Έπειτα, αυτές οι πολιτικές αποστέλλονται σε έναν διαχειριστή (manager) VNF, ο οποίος στη συνέχεια μπορεί να ενημερώσει την κατάλληλη μονάδα IDS (Intrusion Detection System) στο κεντρικό δίκτυο. Με βάση το χρόνο που απαιτείται για την επεξεργασία των δεδομένων από το μοντέλο, μπορούν να καθοριστούν χρονοδιαγράμματα για την εκτέλεση του μοντέλου, ώστε να διασφαλιστεί ότι οι πολιτικές στην ενότητα IDS είναι ενημερωμένες και για την περαιτέρω βελτίωση της μάθησης του μοντέλου μηχανικής μάθησης. Τα βασικά οφέλη είναι η δυνατότητα αυτοματοποίησης της ανίχνευσης, των ενημερώσεων βάσεων δεδομένων και της κατάλληλης δράσης τυχόν κακόβουλων ροών.

Στο παρακάτω σχήμα απεικονίζεται πως το προτεινόμενο σύστημα SDS μπορεί να αναπτυχθεί, επίσης, σε συγκεκριμένα τεμάχια δικτύου για την παρακολούθηση των ροών κυκλοφορίας και τη δημιουργία προφίλ καλοήθης και ανώμαλης κίνησης με βάση τις απαιτούμενες προδιαγραφές για το συγκεκριμένο τμήμα.



Εικόνα 16: Εφαρμογή ασφάλειας που καθορίζεται από λογισμικό (SDS) σε τεμαχισμό δικτύου [10]

Όπως φαίνεται και στο παραπάνω σχήμα, η διάταξη του διαγράμματος εστιάζεται στο διαχωρισμό του επιπέδου ελέγχου (Control Plane - CP) και του επιπέδου χρήστη (User Plane - UP), με το επίπεδο χρήστη (UP) να βρίσκεται είτε στο βασικό δίκτυο (core network) είτε στο C-RAN. Το επίπεδο ελέγχου (CP) βρίσκεται στο βασικό δίκτυο για να συγκεντρώσει τον έλεγχο του δικτύου. Τα στοιχεία C-RAN διανέμονται, συμπεριλαμβανομένων των εφαρμογών vBBU (virtualized Base Band Units), MEC (Mobile Edge Computing) και UP. Οι έγχρωμες γραμμές υποδεικνύουν τις λογικές συνδέσεις μεταξύ του συστήματος SDS και των διαφόρων στοιχείων του δικτύου, τα δεδομένα των τεμαχίων είναι προσβάσιμα από το πρώτο κέντρο δεδομένων (Data Center - DC) για την παρακολούθηση της επιστροφής κίνησης από το C-RAN, καθώς και από την κατανομή των τεμαχίων του δικτύου στο βασικό δίκτυο.

6.1.2 Ανίχνευση ανωμαλιών από Συνελικτικό Νευρωνικό Δίκτυο (CNN)

Όπως μελετήθηκε στο κεφάλαιο της μηχανικής μάθησης, η βαθιά μάθηση (deep learning) είναι ένας τομέας της μηχανικής μάθησης που περιλαμβάνει το σχεδιασμό πολύ επίπεδων νευρωνικών δικτύων, τα οποία, κατά κύριο λόγο, βασίζονται σε μαθηματικές δομές τύπου νευρώνων που χρησιμοποιούν πολλές μεταβλητές για την επίλυση μιας σύνθετης εξίσωσης. Για την ανάπτυξη ενός νευρωνικού δικτύου για την ταξινόμηση κειμένου ή εικόνων απαιτείται σημαντική ποσότητα αρχιτεκτονικής μηχανικής για την απόκτηση ενός δικτύου που ταιριάζει καλύτερα στο παρεχόμενο σύνολο δεδομένων και έχει επαρκές επίπεδο ακρίβειας [104].

Η ανίχνευση εισβολής δικτύου σχετίζεται με το ζήτημα της παρακολούθησης και της διαφοροποίησης των κανονικών ροών δικτύου από μη φυσιολογικές ροές που μπορούν θέτουν σε κίνδυνο την ασφάλεια ενός συστήματος. Τόσο οι κυβερνήσεις όσο και οι οργανισμοί επενδύουν σε μεγάλο βαθμό για να βρουν μια αξιόπιστη λύση για την προστασία των περιουσιακών στοιχείων και των πόρων τους από κακόβουλη πρόσβαση, γεγονός που έφερε τα συστήματα ανίχνευσης εισβολών (Intrusion Detection Systems - IDS) στο προσκήνιο του τοπίου της ασφάλειας στον κυβερνοχώρο [105]. Όπως προτείνεται από τον Denning (1987) [106], την ιδέα για την ανάπτυξη συστημάτων ανίχνευσης εισβολής (IDS) που χρησιμοποιούν τεχνικές μηχανικής μάθησης, αποτελεί ο εντοπισμός μη φυσιολογικών μοτίβων χρήσης και ανώμαλης κίνησης που μπορεί να σηματοδοτήσουν απόπειρα εισβολής στο δίκτυο. Αυτή η ιδέα οδήγησε στη δημιουργία ενός νέου τύπου IDS που βασίζεται σε μαθησιακούς αλγόριθμους, αντί να ενημερώνει με μη αυτόματα τις υπογραφές από προηγούμενους εισβολείς που είχαν εντοπιστεί. Κατά τις τελευταίες τρεις δεκαετίες, εφαρμόστηκαν διάφορες τεχνικές μηχανικής μάθησης σε μια συμβατική προσέγγιση για την ανάπτυξη μοντέλων ανίχνευσης ανωμαλιών δικτύου. Αυτές οι προσεγγίσεις χρησιμοποίησαν αλγορίθμους μηχανικής μάθησης με επίβλεψη, χωρίς επίβλεψη και με ημιεπίβλεψη, ώστε να προτείνουν μια λύση για ανίχνευση ανωμαλιών [105].

Επομένως, η ανίχνευση ανωμαλιών δεν είναι ένας νέος τομέας μελέτης σε εφαρμογές μηχανικής μάθησης και η τρέχουσα έρευνα έχει διερευνήσει μια ποικιλία εφαρμογών που βασίζονται στη μηχανική μάθηση. Ωστόσο, προκύπτουν ορισμένα κοινά ζητήματα, όπως χαμηλά επίπεδα ακρίβειας λόγω μη βέλτιστου σχεδιασμού μοντέλου, μη ρεαλιστικά υψηλά επίπεδα ακρίβειας λόγω έλλειψης γενίκευσης μοντέλου και υπερβολικής προσαρμογής, καθώς και η χρήση παρωχημένων και απλοϊκών συνόλων δεδομένων. Όπως φαίνεται στο [107] η ακρίβεια άνω του 99% επιτυγχάνεται χρησιμοποιώντας ένα νευρωνικό δίκτυο πολλαπλών επιπέδων, ωστόσο το σύνολο δεδομένων που χρησιμοποιείται είναι το σύνολο

δεδομένων KDD99, ένα σύνολο δεδομένων που είναι 20 ετών και δεν αντιπροσωπεύει τα τρέχοντα δυναμικά περιβάλλοντα δικτύου [10].

Η ίδια η ανίχνευση ανωμαλιών μπορεί να μοντελοποιηθεί ευκολότερα ως πρόβλημα ταξινόμησης στην μάθηση με επίβλεψη [105].Υπενθυμίζεται ότι η μάθηση με επίβλεψη σημαίνει ότι τα δεδομένα με ετικέτα χρησιμοποιούνται για την εκπαίδευση του μοντέλου ανίχνευσης ανωμαλιών. Ο στόχος αυτού του τύπου εκπαίδευσης είναι να ταξινομηθούν τα δεδομένα δοκιμής ως ανώμαλα ή κανονικά με βάση ένα συγκεκριμένο σύνολο χαρακτηριστικών.

Στη συγκεκριμένη μελέτη περίπτωσης, το πρόβλημα ανίχνευσης ανωμαλιών προσεγγίζεται (όπως παρουσιάζεται αναλυτικότερα στη συνέχεια) από την σκοπιά της μάθησης με επίβλεψη και χρησιμοποιεί μια αρχιτεκτονική CNN σχεδιασμένη χρησιμοποιώντας την αρχιτεκτονική NAS για να βελτιστοποιήσει τα υψηλότερα δυνατά επίπεδα ακρίβειας.

Ο αποτελεσματικός σχεδιασμός ενός τέτοιου μοντέλου απαιτεί σημαντικό βαθμό αρχιτεκτονικής μηχανικής [104]. Στο [108] δείχνει ότι ο σχεδιασμός βασικών CNN όπου προστίθενται επιπλέον στρώματα για δοκιμές δεν βελτιώνει την ακρίβεια, δίνοντας υποβέλτιστα αποτελέσματα σε ποσοστό ανίχνευσης κάτω από 80%. Στο [109] αποδεικνύουν την αποτελεσματικότητα της δειγματοληψίας πάνω-κάτω (up and down sampling) στα δεδομένα για την εξισορρόπηση όγκων ανωμαλιών και καλοηθών δεδομένων, επιτυγχάνοντας ένα ποσοστό ανίχνευσης 99,99% χρησιμοποιώντας αλγόριθμο τυχαίου δάσους (random forest) και 99,30% χρησιμοποιώντας βαθιά νευρωνικά δίκτυα τριών στρωμάτων, αυτά τα πολύ υψηλά αποτελέσματα είναι απίθανο να αντιπροσωπεύουν επίπεδα ανίχνευσης πραγματικού κόσμου και δίνουν την εντύπωση ενός υπερσύγχρονου μοντέλου και έλλειψης γενίκευσης [10].

Η αποτελεσματική ταξινόμηση τόσο της καλοήθους όσο και της ανώμαλης κυκλοφορίας είναι επίσης ένα ζήτημα, στις περισσότερες περιπτώσεις, τα μοντέλα μπορούν να προσδιορίσουν την επισήμανση καλοήθης κυκλοφορίας με πολύ υψηλή ακρίβεια (99-100%), ωστόσο ο καθορισμός της ανώμαλης κυκλοφορίας μπορεί να είναι πιο δύσκολος, όπως φαίνεται στο [110], όπου ο αλγόριθμος τυχαίου δάσους (random forest) εφαρμόζεται στο σύνολο δεδομένων UNSW-NB15, η καλοήθης κυκλοφορία ταξινομήθηκε με ακρίβεια 99%, ωστόσο η ανώμαλη κίνηση ταξινομήθηκε στο 82%, αυτό σημαίνει ότι το 18% της ανώμαλης κυκλοφορίας ουσιαστικά δεν ανιχνεύτηκε.

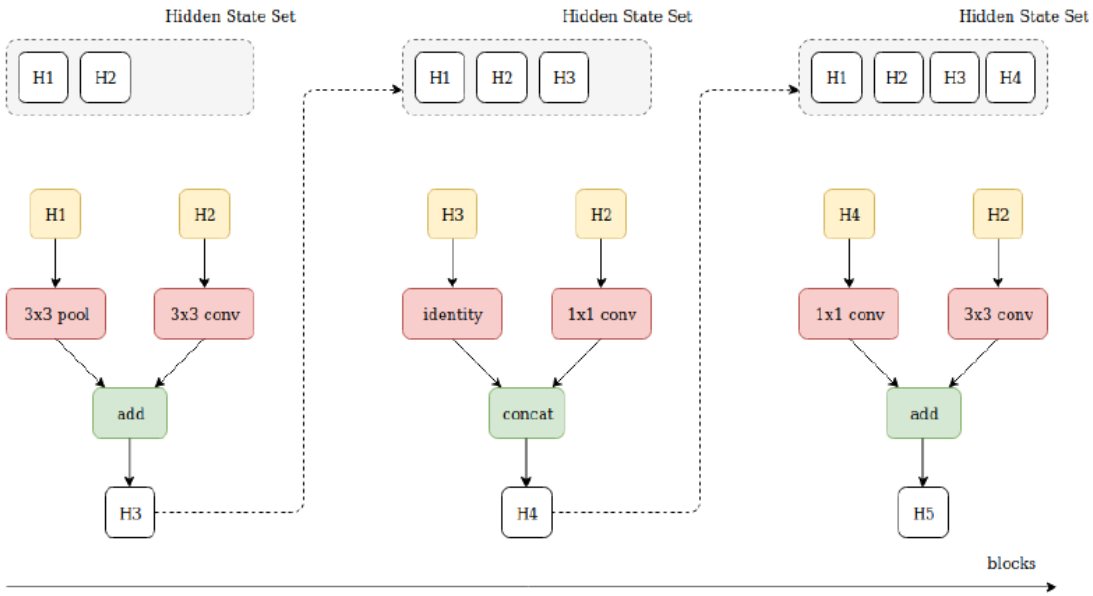
Η προσέγγιση αυτής της μελέτης περίπτωσης διαρθρώνει και αντιμετωπίζει ορισμένα από αυτά τα κοινά ζητήματα. Αυτό γίνεται με δύο βασικούς τρόπους επιλέγοντας το πιο

ενημερωμένο σύνολο δεδομένων IDS, το CICIDS2018 που προσομοιώνει ένα πραγματικό περιβάλλον και εξηγείται λεπτομερώς παρακάτω. Και δεύτερον, χρησιμοποιώντας ένα μοντέλο CNN που βασίζεται στο NAS, το οποίο έχει επιτύχει μερικά από τα υψηλότερα επίπεδα ακρίβειας στο σύνολο δεδομένων ImageNet και χρησιμοποιεί έναν ελεγκτή για την αυτόνομη βελτιστοποίηση παραμέτρων για το μοντέλο. Με αυτήν την προσέγγιση, το βέλτιστο μοντέλο μπορεί να δημιουργηθεί για ένα συγκεκριμένο σύνολο δεδομένων[10].

6.1.3 Υλοποίηση AutoML & NAS

Στη συγκεκριμένη έρευνα, χρησιμοποιήθηκαν οι πλατφόρμες AutoML Vision και το Vision Edge της Google για το σχεδιασμό ενός μοντέλου, την εκπαίδευση, την επικύρωση και την δοκιμή. Η βασική αρχιτεκτονική αυτών των πλατφορμών είναι το NASNet (Neural Architecture Search Network) και το MNasNet (Mobile Neural Architecture Search Network).

Η αναζήτηση νευρωνικής αρχιτεκτονικής μπορεί να οριστεί ως μια μέθοδος με βαθμίδες για την εύρεση βελτιστοποιημένων αρχιτεκτονικών. Η δομή και η συνδεσιμότητα ενός νευωνικού δικτύου μπορούν να καθοριστούν από μια συμβολοσειρά μεταβλητού μήκους. Έτσι, είναι δυνατό να χρησιμοποιηθεί ένα Επαναλαμβανόμενο Νευρωνικό Δίκτυο (Recurrent Neural Network - RNN) για τη δημιουργία αυτής της συμβολοσειράς, όπως φαίνεται στο ακόλουθο σχήμα.



Εικόνα 17: Αναζήτηση δημιουργίας SpaceBlock [10]

Το δίκτυο που καθορίζεται από τη συμβολοσειρά είναι γνωστό ως θυγατρικό δίκτυο και η εκπαίδευση του πραγματικού συνόλου δεδομένων με το θυγατρικό δίκτυο θα έχει ως αποτέλεσμα προοδευτική αύξηση της ακρίβειας στο σύνολο των δεδομένων δοκιμής. Αυτή η ακρίβεια μπορεί να χρησιμοποιηθεί ως σήμα ανταμοιβής για τον υπολογισμό της διαβάθμισης πολιτικής για την ενημέρωση του ελεγκτή. Επομένως, στην επόμενη επανάληψη, ο ελεγκτής θα δώσει μεγαλύτερη πιθανότητα σε αρχιτεκτονικές που λαμβάνουν μεγαλύτερη ακρίβεια [111]. Με απλά λόγια, αυτό σημαίνει ότι, ο ελεγκτής μπορεί να μάθει να βελτιώνει την αναζήτησή του με την πάροδο του χρόνου και να βελτιστοποιεί την τοποθέτηση στρωμάτων και μπλοκ του νευρικού δικτύου.

Σχετικά με την εφαρμογή, η νευρωνική αρχιτεκτονική αναζήτηση (NAS) χρησιμοποιεί τον ελεγκτή για τη δημιουργία ενός συνόλου αρχιτεκτονικών παραμέτρων του δικτύου. Στην περίπτωση ενός CNN μπορεί να προβλέψει το ύψος του φίλτρου, το πλάτος του φίλτρου, το ύψος του βήματος, το πλάτος του βήματος και έναν αριθμό φίλτρων ανά στρώση [111]. Αυτή η διαδικασία επαναλαμβάνεται συνεχώς μέχρι ο αριθμός των επιπέδων να υπερβεί μια συγκεκριμένη τιμή. Αξίζει να σημειωθεί ότι, αυτή η τεχνική εφαρμόζεται σε ένα δείγμα του συνόλου δεδομένων, καθώς η εφαρμογή της NAS σε ένα πολύ μεγάλο σύνολο δεδομένων απαιτεί μεγάλο αριθμό υπολογιστικών πόρων [104].

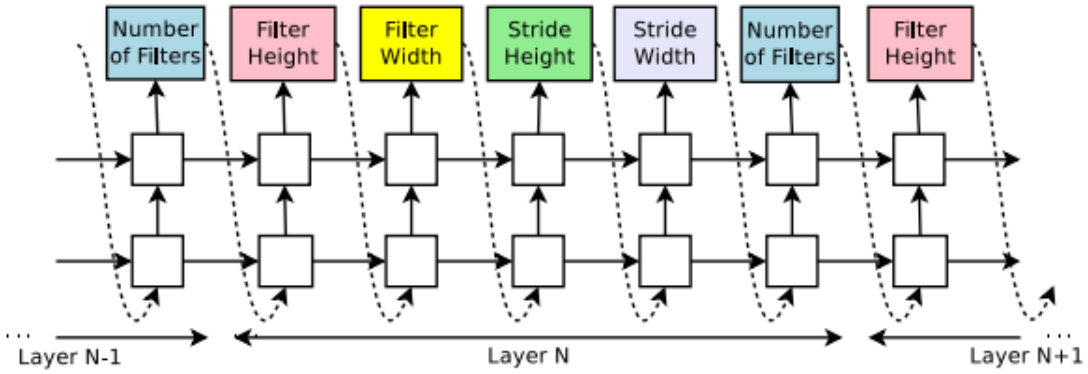
Ο χώρος αναζήτησης στο NAS ορίζεται έτσι ώστε η πολυπλοκότητα της αρχιτεκτονικής να είναι ανεξάρτητη από το βάθος του δικτύου και το μέγεθος των εικόνων εισαγωγής. Αυτό το επιτυγχάνει διασπώντας όλα τα CNN στο χώρο αναζήτησης σε κελιά με ίδια δομή αλλά διαφορετικά βάρη όπως φαίνεται στο προηγούμενο σχήμα (Εικόνα 17) [104]. Επομένως, η αναζήτηση της βέλτιστης αρχιτεκτονικής μπορεί να μειωθεί σε αναζήτηση της καλύτερης αρχιτεκτονικής κελιού. Ψάχνοντας για κάθε συγκεκριμένη αρχιτεκτονική κελιού, η ταχύτητα αυξάνεται πολύ και το κελί είναι πιο πιθανό να έχει καλύτερη γενίκευση. Με βάση αυτήν την ατομική προσέγγιση εκπαίδευσης κυψελών, τα δίκτυα μπορούν να βελτιστοποιηθούν για ταχύτητα ή ακρίβεια ανάλογα με το μέγεθος του χώρου αναζήτησης. Αυτό επιτρέπει στο νευρωνικό δίκτυο να επιτύχει ένα πολύ υψηλό επίπεδο ακρίβειας στο σύνολο δεδομένων επικύρωσης ImageNet στο 82,7% κορυφαία ακρίβεια [112]. Το ImageNet είναι ένα σύνολο δεδομένων εικόνας που οργανώνεται σύμφωνα με την ιεραρχία του WordNet. Κάθε ουσιαστική έννοια στο WordNet, που πιθανώς περιγράφεται από πολλές λέξεις ή φράσεις λέξεων, ονομάζεται σύνολο. Υπάρχουν περισσότερα από 100.000 σύνολα στο WordNet, τα περισσότερα από αυτά είναι ουσιαστικά πάνω από 80.000. Οι εικόνες κάθε έννοιας ελέγχονται από την ποιότητα και επισημαίνονται από τον άνθρωπο. Το ImageNet είναι η μεγαλύτερη βάση δεδομένων για εικόνες με ετικέτες που περιέχουν πάνω από 14

εκατομμύρια εικόνες και χρησιμοποιείται ευρέως στην παροχή αναφοράς για τον προσδιορισμό της απόδοσης διαφορετικών μοντέλων CNN .

Το MnasNet επεκτείνει την έννοια του χώρου αναζήτησης NAS εφαρμόζοντας τον παραγοντοποιημένο ιεραρχικό χώρο αναζήτησης (factorized hierarchical search space). Ο παραγοντοποιημένος ιεραρχικός χώρος αναζήτησης βοηθάει την πρόσθετη ποικιλία επιπέδων σε όλο το δίκτυο και εξισορροπεί το μέγεθος του συνολικού χώρου αναζήτησης. Αυτή η προσέγγιση φέρνει περισσότερη ευελιξία στο NAS καθώς τα μοντέλα μπορούν να σχεδιαστούν για να εξισορροπήσουν την ταχύτητα και την ακρίβεια. Μέχρι στιγμής αυτή η προσέγγιση έχει το μεγαλύτερο πλεονέκτημα της ταχύτητας. Στο σύνολο δεδομένων ImageNet, η αρχιτεκτονική του MNasNet πέτυχε ποσοστό 75,2% μέγιστη ακρίβεια, σε σύγκριση με τις παραδοσιακές αρχιτεκτονικές νευρωνικών δικτύων κινητής τηλεφωνίας είναι 1,8 φορές ταχύτερη και 0,5% υψηλότερη ακρίβεια από το MobileNet V2. Σε σύγκριση με τα αποτελέσματα του NASNet ήταν 7,5% χαμηλότερη ακρίβεια, ωστόσο 2,3 φορές ταχύτερη στην επεξεργασία εικόνων μέσα στην αρχιτεκτονική [113].

6.1.3.1 Γενικές περιγραφές υποδείγματος με έναν ελεγκτή RNN

Όπως αναφέρθηκε νωρίτερα, στη νευρωνική αρχιτεκτονική αναζήτησης (NAS), χρησιμοποιείται ένας ελεγκτή για τη δημιουργία αρχιτεκτονικών υπερπαραμέτρων νευρωνικών δικτύων. Για να είναι ευέλικτος, ο ελεγκτής υλοποιείται ως επαναλαμβανόμενο νευρωνικό δίκτυο. Υποθέτοντας ότι θέλουμε να γίνει πρόβλεψη τροφοδότηση νευρικού δικτύου προς τα εμπρός με μόνο συνελικτικά, μπορεί να χρησιμοποιηθεί ο ελεγκτής για τη δημιουργία των υπερπαραμέτρων του ως ακολουθία συμβόλων:



Εικόνα 18: Ελεγκτής RNN[111]

Το παραπάνω σχήμα της εικόνας δείχνει πως ο ελεγκτής RNN δειγματίζει ένα απλό συνελικτικό δίκτυο. Όπου, προβλέπει το ύψος του φίλτρου, το πλάτος του φίλτρου, το ύψος

του βήματος, το πλάτος του βήματος και τον αριθμό των φίλτρων για ένα στρώμα και επαναλήψεις. Κάθε πρόβλεψη πραγματοποιείται από έναν ταξινομητή (classifier) softmax και στη συνέχεια τροφοδοτείται στο επόμενο βήμα ως είσοδος.

Σε αυτή τη μελέτη, η διαδικασία δημιουργίας μιας αρχιτεκτονικής σταματά εάν ο αριθμός των επιπέδων υπερβαίνει μια συγκεκριμένη τιμή. Αυτή η τιμή ακολουθεί ένα χρονοδιάγραμμα, και αυξάνεται καθώς προχωρά η εκπαίδευση. Μόλις ο ελεγκτής RNN ολοκληρώσει τη δημιουργία μιας αρχιτεκτονικής, δημιουργείται και εκπαιδεύεται ένα νευρωνικό δίκτυο με αυτήν την αρχιτεκτονική. Κατά τη συνέλιξη, καταγράφεται η ακρίβεια του δικτύου σε ένα παλιό σύνολο επικύρωσης. Οι παράμετροι του ελεγκτή RNN, θ_c, στη συνέχεια βελτιστοποιούνται προκειμένου να μεγιστοποιηθεί η αναμενόμενη ακρίβεια επικύρωσης των προτεινόμενων αρχιτεκτονικών. Στην επόμενη ενότητα, περιγράφεται μια μέθοδος διαβάθμισης πολιτικής που χρησιμοποιήθηκε από στο [111] για ενημέρωση των παραμέτρων θ_c έτσι ώστε ο ελεγκτής RNN να παράγει καλύτερες αρχιτεκτονικές με την πάροδο του χρόνου.

6.1.3.2 Εκπαίδευση με ενίσχυση

Η λίστα των διακριτικών που προβλέπει ο ελεγκτής θεωρείται ως μια λίστα ενεργειών a_1:T για το σχεδιασμό μιας αρχιτεκτονικής για ένα θυγατρικό δίκτυο. Κατά τη σύγκλιση, αυτό το θυγατρικό δίκτυο θα επιτύχει μια ακρίβεια R σε ένα παλιό σύνολο δεδομένων. Μπορεί να χρησιμοποιηθεί αυτή η ακρίβεια R ως σήμα ανταμοιβής και να χρησιμοποιηθεί μάθηση με ενίσχυση για την εκπαίδευση του ελεγκτή. Πιο συγκεκριμένα, για να βρεθεί η βέλτιστη αρχιτεκτονική, ζητείται από τον ελεγκτή να μεγιστοποιήσει την αναμενόμενη ανταμοιβή του, που αντιπροσωπεύεται από το $J(\theta_c)$:

$$J(\theta_c) = E_{P(a_{1:T};\theta_c)}[R]$$

Δεδομένου ότι το σήμα ανταμοιβής R δεν είναι διαφοροποιήσιμο, πρέπει να χρησιμοποιηθεί μια μέθοδος διαβάθμισης πολιτικής για την επανάληψη ενημέρωσης θ_c. Στο [111], χρησιμοποιήθηκε ο κανόνας ενίσχυσης (Reinforce)από τον Williams (1992):

$$\nabla_{\theta_c} J(\theta_c) = \sum_{t=1}^{T} E_{P(a_{1:T};\theta_c)}\left[\nabla_{\theta_c} log\, P\left(a_t | a_{(t-1):1}; \theta_c\right) R\right]$$

Μια εμπειρική προσέγγιση της παραπάνω συνάρτησης είναι:

$$\frac{1}{m}\sum_{k=1}^{m}\sum_{t=1}^{T} \nabla_{\theta_c} log\, P\left(a_t | a_{(t-1):1}; \theta_c\right) R_k$$

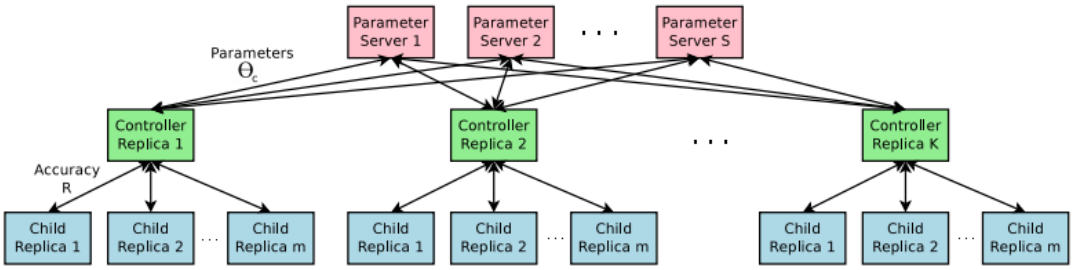
Όπου m είναι ο αριθμός των διαφορετικών αρχιτεκτονικών που ο ελεγκτής δειγματίζει σε μία παρτίδα και T είναι ο αριθμός των υπερπαραμέτρων που πρέπει να προβλέψει ο ελεγκτής για να σχεδιάσει μια αρχιτεκτονική νευρωνικού δικτύου.

Η ακρίβεια επικύρωσης που επιτυγχάνει η αρχιτεκτονική του νευρωνικού δικτύου k-th μετά την εκπαίδευση σε ένα σύνολο δεδομένων εκπαίδευσης είναι R_k. Η παραπάνω ενημέρωση είναι μια εκτίμηση για την κλίση της συνάρτησης, αλλά έχει πολύ υψηλή διακύμανση. Για να μειωθεί αυτή η διακύμανση της εκτίμησης χρησιμοποιήθηκε μια συνάρτηση βάσης

$$\frac{1}{m}\sum_{k=1}^{m}\sum_{t=1}^{T}\nabla_{\theta_c} log\, P\left(a_t|a_{(t-1):1};\theta_c\right)(R_k - b)$$

Σε αυτό το έργο, το βασικό b είναι ένας εκθετικός κινούμενος μέσος όρος για την ακρίβεια των προηγούμενων αρχιτεκτονικών.

Στην αναζήτηση νευρικής αρχιτεκτονικής, κάθε ενημέρωση διαβάθμισης στις παραμέτρους ελεγκτή θ_c αντιστοιχεί στην εκπαίδευση ενός θυγατρικού δικτύου σε συνέλιξη. Καθώς η εκπαίδευση ενός θυγατρικού δικτύου (child network) (δηλαδή, ενός δικτύου που περιλαμβάνει όλα τα χαρακτηριστικά και τις λειτουργίες του γονικού δικτύου και μπορούν να γίνουν προσαρμογές σε αυτό χωρίς να επηρεαστεί το γονικό δίκτυο) μπορεί να διαρκέσει ώρες, χρησιμοποιείται κατανεμημένη εκπαίδευση και ασύγχρονες ενημερώσεις παραμέτρων προκειμένου να επιτευχθεί η διαδικασία εκπαίδευσης του ελεγκτή. Πιο συγκεκριμένα, χρησιμοποιείται ένα σχήμα παραμέτρων-διακομιστή όπου υπάρχει ένας διακομιστής παραμέτρων S θραυσμάτων, που αποθηκεύουν τις κοινόχρηστες παραμέτρους για τα αντίγραφα του ελεγκτή K. Κάθε αντίγραφο του ελεγκτή δειγματίζει διαφορετικές θυγατρικές αρχιτεκτονικές που εκπαιδεύονται παράλληλα. Στη συνέχεια, ο ελεγκτής συλλέγει διαβαθμίσεις σύμφωνα με τα αποτελέσματα αυτής της μικρής παρτίδας αρχιτεκτονικών m σε συνέλιξη και τις στέλνει στον διακομιστή παραμέτρων για να ενημερώσει τα μέτρα σε όλα τα αντίγραφα (replica) του ελεγκτή. Κατά την εφαρμογή, η συνέλιξη κάθε θυγατρικού δικτύου επιτυγχάνεται όταν η εκπαίδευσή του υπερβαίνει έναν ορισμένο αριθμό εποχών. Αυτό το σχήμα που περιγράφηκε παραπάνω συνοψίζεται στο ακόλουθο σχήμα.



Εικόνα 19: Κατανεμημένη εκπαίδευση για την NAS[111]

Συνοπτικά λοιπόν, χρησιμοποιείται ένα σύνολο S παραμέτρων διακομιστών (parameter servers) για την αποθήκευση και την αποστολή παραμέτρων σε αντίγραφα ελεγκτή K (controller replica). Στη συνέχεια, κάθε αντίγραφο του ελεγκτή δειγματίζει αρχιτεκτονικές m και εκτελεί παράλληλα τα πολλαπλά θυγατρικά μοντέλα (child replica). Η ακρίβεια κάθε θυγατρικού μοντέλου καταγράφεται για τον υπολογισμό των διαβαθμίσεων σε σχέση με τα θ_c, τα οποία στη συνέχεια αποστέλλονται πίσω στους διακομιστές παραμέτρων (parameter servers) [111].

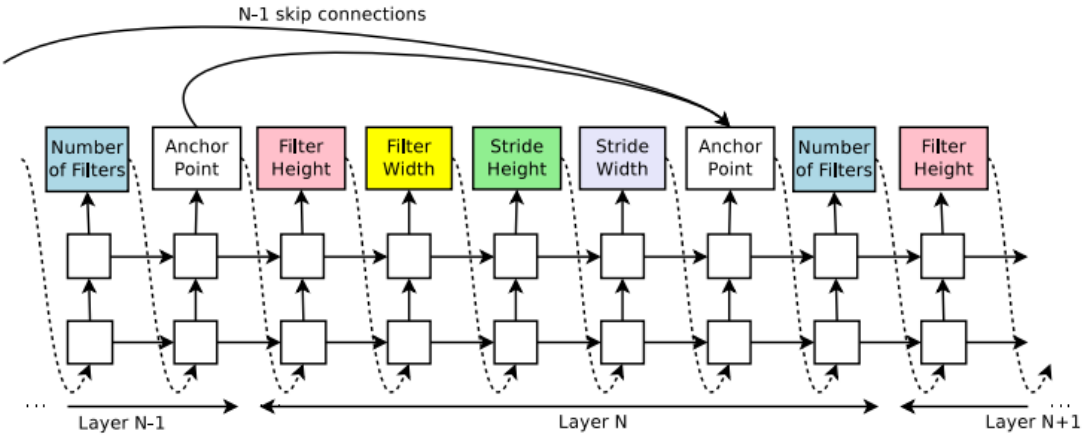
6.1.3.3 Συνδέσεις παράλειψης

Στο πρώτο σχήμα που παρουσιάστηκε για τον ελεγκτή δεν περιλαμβάνονται συνδέσεις παράλειψης (skip connections) ή επίπεδα διακλάδωσης (branching layers), τα οποία χρησιμοποιούνται σε σύγχρονες αρχιτεκτονικές όπως το GoogleNet και το ResidualNet. Σε αυτήν την ενότητα παρουσιάζεται μια μέθοδος που επιτρέπει στον ελεγκτή RNN να προτείνει συνδέσεις παράλειψης ή επίπεδα διακλάδωσης, διευρύνοντας έτσι τον χώρο αναζήτησης. Για να επιτραπεί αυτή η ενέργεια στον ελεγκτή, δηλαδή να προβλέψει τέτοιες συνδέσεις, χρησιμοποιήθηκε ένα σύνολο επιλογής τύπου προσοχής (set-selection type attention), το οποίο βασίζεται στο μηχανισμό προσοχής (attention mechanism).

Πιο συγκεκριμένα, στο επίπεδο N, προστίθεται ένα σημείο άγκυρα (anchor point) που έχει N-1 σιγμοειδή (συνάρτηση S) που βασίζεται στο περιεχόμενο για να υποδείξει τα προηγούμενα επίπεδα που πρέπει να συνδεθούν. Κάθε σιγμοειδή είναι συνάρτηση του τρέχοντος κρυφού σταδίου του ελεγκτή και των προηγούμενων κρυφών σταθμών των προηγούμενων σημείων άγκυρας N-1: P(Layer j is an input to layer i) = sigmoid (u^T tanh (W_{prev}*h_j+W_{curr}*h_i))

Όπου, το h_j αντιπροσωπεύει την κρυφή κατάσταση του ελεγκτή στο σημείο άγκυρας για το επίπεδο j-th, όπου το j κυμαίνεται από 0 έως N-1. Στη συνέχεια, λαμβάνονται δείγματα από αυτά τα σιγμοειδή για να αποφασιστεί ποια προηγούμενα στρώματα θα χρησιμοποιηθούν ως είσοδοι στο τρέχον επίπεδο. Οι πίνακες W_{prev}, W_{curr} και v είναι εκπαιδευτικές παράμετροι.

Καθώς οι συνδέσεις αυτές ορίζονται επίσης από κατανομές πιθανότητας, η μέθοδος ενίσχυσης εξακολουθεί να ισχύει χωρίς σημαντικές τροποποιήσεις. Το ακόλουθο δείχνει πώς ο ελεγκτής RNN χρησιμοποιεί τις συνδέσεις παράλειψης (skip connections) για να αποφασίσει ποια επίπεδα θέλει ως εισόδους στο τρέχον επίπεδο.



Εικόνα 20: Δημιουργία συνδέσεων παράλειψης από τον ελεγκτή RNN[111]

Στο πλαίσιο αυτό, εάν ένα επίπεδο έχει πολλά επίπεδα εισόδου, τότε όλα τα επίπεδα εισόδου συνδυάζονται στη διάσταση βάθους. Η παράλειψη συνδέσεων μπορεί να προκαλέσει αποτυχίες σύνταξης, όπου ένα επίπεδο δεν είναι συμβατό με ένα άλλο επίπεδο ή ένα επίπεδο ενδέχεται να μην έχει καμία είσοδο ή έξοδο. Για να παρακαμφθούν αυτά τα ζητήματα, χρησιμοποιούνται τρεις απλές τεχνικές. Πρώτον, εάν ένα επίπεδο δεν είναι συνδεδεμένο σε κανένα επίπεδο εισόδου, τότε η εικόνα χρησιμοποιείται ως το επίπεδο εισόδου. Δεύτερον, στο τελικό επίπεδο λαμβάνονται όλες οι έξοδοι επιπέδου που δεν έχουν συνδεθεί και συνδυάζονται προτού σταλεί στον ταξινομητή. Τέλος, εάν τα στρώματα εισόδου που θα συνδυαστούν έχουν διαφορετικά μεγέθη, τα μικρά στρώματα συμπληρώνονται με μηδενικά, έτσι ώστε τα συνδυασμένα επίπεδα να έχουν τα ίδια μεγέθη [111].

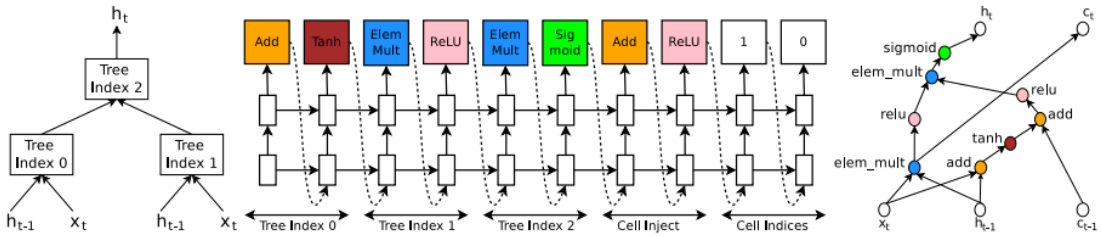
6.1.3.4 Γενική αρχιτεκτονική επαναλαμβανόμενων κελιών

Σε αυτήν την ενότητα, τροποποιείται η παραπάνω μέθοδο για τη δημιουργία επαναλαμβανόμενων κελιών. Σε κάθε βήμα t, ο ελεγκτής πρέπει να βρει μια λειτουργική φόρμα για h_t που παίρνει x_t και h_t-1 ως εισόδους. Ο απλούστερος τρόπος είναι αν ισχύει ότι $h_t = \tanh(W_1*x_t+W_2*h_t\text{-}1)$, που είναι ο σχηματισμός ενός βασικού επαναλαμβανόμενου κελιού. Ένα πιο περίπλοκο σκεύασμα είναι το ευρέως χρησιμοποιούμενο LSTM.

Οι υπολογισμοί για βασικά κελιά RNN και LSTM μπορούν να γενικευθούν ως ένα δέντρο βημάτων που λαμβάνουν x_t και h_t-1 ως εισόδους και παράγουν h_t ως τελική έξοδο. Ο

ελεγκτής RNN πρέπει να επισημαίνει κάθε κόμβο στο δέντρο με μια μέθοδο συνδυασμού (προσθήκη, στοιχειώδης πολλαπλασιασμός κ.α.) και μια συνάρτηση λειτουργίας (activation function) (tanh, σιγμοειδής κ.λπ.) για να συγχωνεύσει δύο εισόδους και να παράγει μία έξοδο. Στη συνέχεια, δύο έξοδοι τροφοδοτούνται ως είσοδοι στον επόμενο κόμβο στο δέντρο. Για να επιτραπεί στον ελεγκτή RNN να επιλέξει αυτές τις μεθόδους και τις συναρτήσεις, δημιουργήθηκε ένα ευρετήριο κόμβων στο δέντρο (index tree) με μια σειρά, έτσι ώστε ο ελεγκτής RNN να μπορεί να επισκέπτεται κάθε κόμβο έναν προς έναν και να επισημαίνει τις απαραίτητες υπερπαραμέτρους.

Σύμφωνα με την κατασκευή του LSTM, χρειάζονται επίσης μεταβλητές κυψελών c_t-1 και c_t για την αναπαράσταση των καταστάσεων μνήμης. Για να ενσωματωθούν αυτές τις μεταβλητές, χρειάζεται ο ελεγκτής RNN για να προβλέψει ποιοι κόμβοι στο δέντρο θα συνδέσουν αυτές τις δύο μεταβλητές. Αυτές οι προβλέψεις μπορούν να γίνουν στα δύο τελευταία μπλοκ του ελεγκτή RNN. Για να γίνει πιο ξεκάθαρη αυτή η διαδικασία παρουσιάζονται παρακάτω ένα σχήμα, για μια δομή δέντρου που έχει δύο κόμβους φύλλων και έναν εσωτερικό κόμβο.



Εικόνα 21: Παράδειγμα επαναλαμβανόμενου κελιού [111]

Οι κόμβοι φύλλων χαρακτηρίζονται με 0 και 1, και ο εσωτερικός κόμβος με 2. Ο ελεγκτής RNN πρέπει πρώτα να προβλέψει 3 μπλοκ, κάθε μπλοκ που καθορίζει μια μέθοδο συνδυασμού και μια λειτουργία ενεργοποίησης για κάθε δείκτη δέντρου. Μετά από αυτό πρέπει να προβλέψει τα τελευταία 2 μπλοκ που καθορίζουν τον τρόπο σύνδεσης ct και ct − 1 σε προσωρινές μεταβλητές μέσα στο δέντρο.

Ειδικότερα, σύμφωνα με τις προβλέψεις του ελεγκτή RNN σε αυτό το παράδειγμα, θα προκύψουν τα ακόλουθα βήματα υπολογισμού:

- Ο ελεγκτής προβλέπει τις Add και Tanh για το ευρετήριο δέντρων (index tree) 0, αυτό σημαίνει ότι πρέπει να υπολογιστεί:

$$a_0 = tanh(W_1 * x_t + W_2 * h_{t-1})$$

- Ο ελεγκτής προβλέπει τις ElemMult και ReLU για το ευρετήριο δέντρων (index tree)1, αυτό σημαίνει ότι πρέπει να υπολογιστεί:

$$a_1 = ReLU\big((W_3 * x_t) \odot (W_4 * h_{t-1})\big)$$

- Ο ελεγκτής προβλέπει 0 για το δεύτερο στοιχείο του "CellIndex", Add και ReLU για στοιχεία στο "CellInject", που σημαίνει ότι πρέπει να υπολογιστεί ένα νέο

$$a_0^{new} = ReLU(a_0 + c_{t-1})$$

Σημειώνεται ότι δεν περιλαμβάνονται παράμετροι εκπαίδευσης για τους εσωτερικούς κόμβους (internal nodes) του δέντρου.

- Ο ελεγκτής προβλέπει ElemMult και Sigmoid για το index tree 2, αυτό σημαίνει ότι πρέπει να υπολογιστεί

$$a_2 = sigmoid(a_0^{new} \odot a_1)$$

Δεδομένου ότι ο μέγιστος δείκτης στο δέντρο είναι 2, το h_t ορίζεται σε a_2

- Ο ελεγκτής RNN προβλέπει 1 για το πρώτο στοιχείο του "CellIndex", αυτό σημαίνει ότι πρέπει να ρυθμιστεί το c_t στην έξοδο του δέντρου στο index 1 πριν από την ενεργοποίηση, δηλαδή,

$$c_t = (W_3 * x_t) \odot (W_4 * h_{t-1})$$

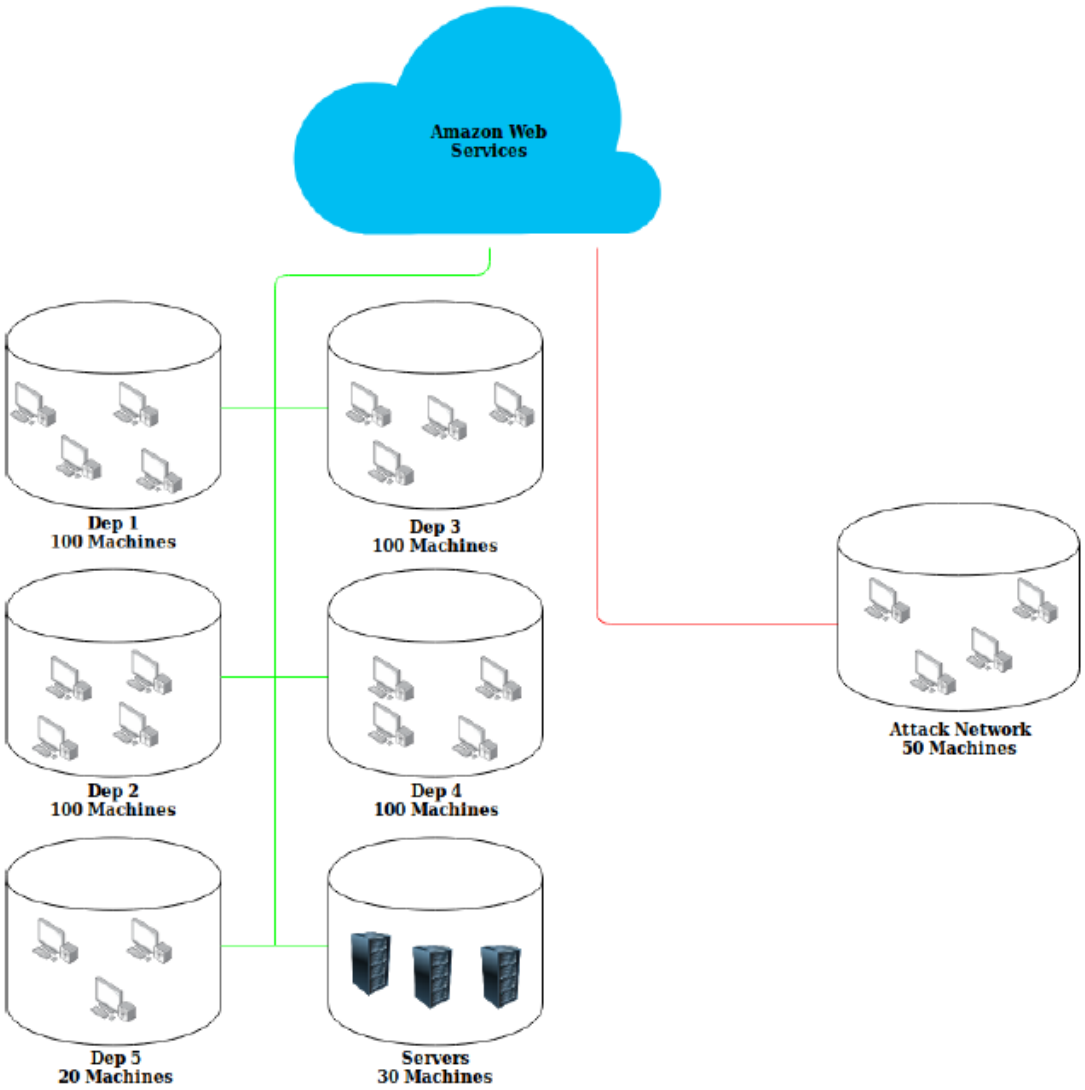
Στη συνέχεια παρουσιάζεται η πειραματικά εφαρμογή των [10] που βασίστηκε στη θεωρία που παρουσιάστηκε σε αυτή την ενότητα και βασίζεται στο [111]. Τα αποτελέσματα από το [111] δεν παρουσιάζονται καθώς δεν αναφέρονται σε δίκτυα 5G.

6.1.4 Σύνολα δεδομένων για την ανίχνευση ανωμαλιών

Η ανίχνευση ανωμαλιών είναι ένας από τους πιο υποσχόμενους τομείς έρευνας για την ανίχνευση νέων επιθέσεων. Ωστόσο, η υιοθέτησή του σε πραγματικές εφαρμογές εμποδίζεται λόγω της πολυπλοκότητας του συστήματος που απαιτεί μεγάλο αριθμό δοκιμών, συντονισμού και αξιολόγησης. Επομένως, για ερευνητικούς σκοπούς, ένα προσομοιωμένο σύστημα μπορεί να σχεδιαστεί με ένα ολοκληρωμένο σύνολο παρεμβολών και ανώμαλης συμπεριφοράς μαζί με κανονική κίνηση για ανάλυση της ανίχνευσης ανωμαλιών. Καθώς οι συμπεριφορές δικτύου και το κακόβουλο λογισμικό αλλάζουν, καθίσταται απαραίτητο να υπάρχει ένα περιβάλλον που προσομοιώνει με μεγαλύτερη ακρίβεια ένα σενάριο πραγματικού κόσμου. Τα δεδομένα που μπορούν στη συνέχεια να ληφθούν από το σύστημα

είναι δυναμικά και παρέχουν πιο ουσιαστική και ρεαλιστική εικόνα για καλοήθη και ανώμαλη συμπεριφορά κίνησης δικτύου. Δυστυχώς, τα παραδοσιακά σύνολα δεδομένων IDS δεν σχεδιάστηκαν με αυτόν τον τρόπο, για παράδειγμα το σύνολο δεδομένων KDDCUP99 ή το σύνολο δεδομένων ADFA-IDS δημιουργήθηκαν σε περιβάλλον δοκιμών που αποτελούταν μόνο από συνδέσμους LAN και ένα επιθετικό και ένα αμυντικό σύστημα, αυτή η προσέγγιση αντιπροσωπεύει ένα στατικό περιβάλλον και παρέχει μη βέλτιστα και λιγότερο ρεαλιστικά αποτελέσματα [10].

Το σύνολο δεδομένων IDS-2018 από το Canadian Institute of Cyber security είναι ένα σύνολο δεδομένων που προέρχεται από ένα προσομοιωμένο περιβάλλον που επιχειρεί να αντιμετωπίσει αυτές τις ελλείψεις. Ο κύριος στόχος αυτού του συνόλου δεδομένων είναι η χρήση μιας συστηματικής προσέγγισης για τη δημιουργία ενός ποικίλου και περιεκτικού συνόλου δεδομένων αναφοράς για τον εντοπισμό εισβολής που βασίζεται στη δημιουργία προφίλ καλοήθη και ανώμαλης επισκεψιμότητας. Το ίδιο το περιβάλλον αποτελείται από 50 μηχανήματα επίθεσης σε μια οργάνωση θυμάτων με 5 τμήματα που περιλαμβάνουν 420 μηχανήματα και 30 διακομιστές. Το σύνολο δεδομένων λαμβάνει πακέτα συλλογής κίνησης δικτύου και αρχείων καταγραφής κάθε μηχανήματος, καθώς και την εξαγωγή 80 δυνατοτήτων δικτύου που οργανώνονται ως ροές. Το παρακάτω σχήμα δείχνει τη συνολική τοπολογία δικτύου που είναι ένα κοινό δίκτυο LAN σε μια πλατφόρμα cloud AWS (AmazonWeb Services). Έξι υποδίκτυα εγκαθίστανται με την ένδειξη Dep1 έως Dep5 και Servers. Τα μηχανήματα Dep1 έως Dep4 διαθέτουν λειτουργικά συστήματα Windows 8/10, το Dep5 διαθέτει όλα τα μηχανήματα Linux που εκτελούν Ubuntu, οι διακομιστές έχουν διαφορετικούς διακομιστές MS Windows, όπως διακομιστές εφαρμογών, ενεργό κατάλογο και email. Το δίκτυο επιτιθέμενων διαθέτει μηχανήματα Windows 8/10 και μηχανήματα Ubuntu.



Εικόνα 22: Τοπολογία δικτύου CICIDS2018 [10]

6.1.5 Προφίλ και δυνατότητες δικτύου

Τα πρωτόκολλα που προσομοιώνονται στο περιβάλλον είναι: HTTPS (Hyper-Text Transfer Protocol Secure - Πρωτόκολλο Μεταφοράς Ασφαλής Υπερκειμένου), HTTP (Hyper Text Transfer Protocol - Πρωτόκολλο Μεταφοράς Υπερκειμένου), SMTP (Simple Mail Transfer Protocol - Πρωτόκολλο Μεταφοράς Ηλεκτρονικού Ταχυδρομείου), POP3 (Post Office Protocol 3 - Πρωτόκολλο ταχυδρομείου), IMAP (Internet Message Access Protocol - Πρωτόκολλο πρόσβασης σε μηνύματα Διαδικτύου), SSH (Secure Shell), FTP (File Transfer Protocol – Πρωτόκολλο Μεταφοράς Αρχείων). Οι τύποι κυκλοφορίας χωρίζονται σε δύο προφίλ, σε προφίλ Β (καλοήθης κυκλοφορία) και σε προφίλ Μ (κακόβουλη κίνηση).

Το προφίλ Β (καλοήθης κυκλοφορία), περιγράφει τους κανονικούς τύπους κυκλοφορίας που προσομοιώνονται μέσω ενός αριθμού αλγορίθμων μηχανικής μάθησης με διαφορετικά πρωτόκολλα δικτύου. Αναλυτικότερα, το προφίλ αυτό εξομοιώνει τη συμπεριφορά των χρηστών χρησιμοποιώντας διάφορες τεχνικές στατιστικής ανάλυσης μηχανικής εκμάθησης όπως το K-Means, το Random Forest, το SVM και το J48.Οι δυνατότητες δικτύου που συλλέγονται περιλαμβάνουν το μέγεθος του πρωτοκόλλου πακέτου, τον αριθμό των πακέτων ανά ροή, διάφορα μοτίβα στο ωφέλιμο φορτίο, το μέγεθος του ωφέλιμου φορτίου και την κατανομή χρόνου αιτήματος ενός πρωτοκόλλου.

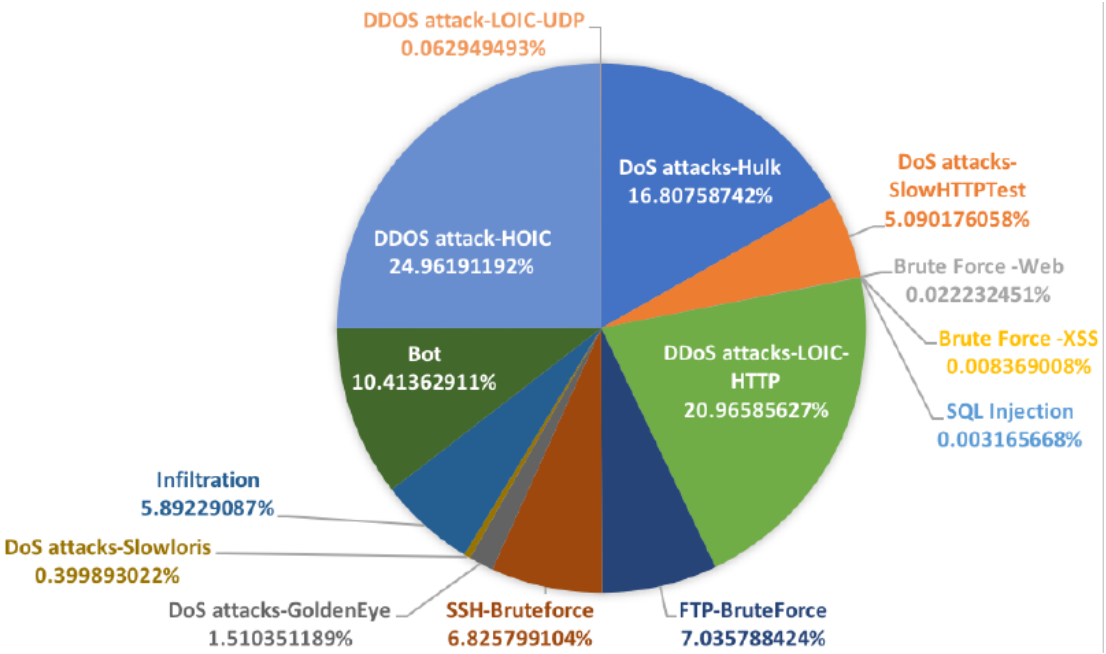
Οι επιθέσεις που χρησιμοποιούνται στο προφίλ Μ είναι κοινές επιθέσεις που χρησιμοποιούνται από κακόβουλους ηθοποιούς καθώς και δοκιμαστές διείσδυσης. Καλύπτουν ένα ευρύ φάσμα σεναρίων από επιθέσεις βάσει δικτύου, διάφορες μορφές HTTP DoS και DDoS, επιθέσεις με βάση το διαδίκτυο και εκτεταμένες ευπάθειες. Καλύπτουν επίσης πτυχές του OWASP top 10 2019, συμπεριλαμβανομένων επιθέσεων με εγχύσεις SQL, κατεστραμμένου ελέγχου ταυτότητας λόγω κακής διαχείρισης κωδικού πρόσβασης, επιτρέποντας ευκολότερες επιθέσεις brute force και εσφαλμένες διαμορφώσεις ασφαλείας που επιτρέπουν ευπάθειες [114].

Το προφίλ Μ περιγράφει το σενάριο επίθεσης για ανώμαλη κίνηση, όπου προσομοιώνονται πέντε διαφορετικά σενάρια επίθεσης [115]: 1) την εσωτερική διείσδυση δικτύου, η οποία εκμεταλλεύεται την ευπάθεια της εφαρμογής στέλνοντας κακόβουλα αρχεία μέσω email. Το πλαίσιο Metasploit χρησιμοποιείται για εκμετάλλευση επιτρέποντας την εκτέλεση ενός backdoor στον υπολογιστή του θύματος, 2) HTTP DoS – Slowloris, LOIC και HOIC που προκαλούν άρνηση υπηρεσίας, αυτά τα εργαλεία είναι σε θέση να κάνουν τους διακομιστές ιστού απρόσιτους. Το Slowloris μπορεί να το κάνει αυτό με ένα μόνο μηχάνημα και είναι πιο αποτελεσματικό σε σχέση με τους διακομιστές Apache. Οι διακομιστές Apache είναι οι

δεύτεροι πιο συνηθισμένοι διακομιστές ιστού στο διαδίκτυο που αντιπροσωπεύουν το 26,73% των διακομιστών ιστού [10]. 3) Επιθέσεις εφαρμογών ιστού, δηλαδή επιθέσεις που βασίζονται σε εφαρμογές Web που δοκιμάστηκαν χρησιμοποιώντας την εφαρμογή Damn Vulnerable Web (DVWA) για έγχυση SQL, έγχυση εντολών και μεταφόρτωση αρχείων χωρίς εμπόδια. 4) Επίθεση Brute, στην οποία χρησιμοποιείται ένα λεξικό επίθεσης brute force που περιέχει 90 εκατομμύρια λέξεις εναντίον κύριων διακομιστών για να γίνει προσπάθεια απόκτησης πληροφοριών λογαριασμού SSH και MySQL. 5) Οι επιθέσεις τελευταίων ενημερώσεων, οι οποίες είναι γνωστές ως τρωτά σημεία που μπορούν να επηρεάσουν χιλιάδες συσκευές υπό ορισμένες συνθήκες, όπου εκτελούν παλαιότερες, ξεπερασμένες εκδόσεις λογισμικού. Το Heartleech χρησιμοποιήθηκε σε αυτό το περιβάλλον για τη σάρωση συστημάτων ευάλωτων στο σφάλμα Heartbleed, μόλις βρεθούν συστήματα, τότε μπορούν να αξιοποιηθούν και να εξαχθούν τα δεδομένα [10].

Για τον ορισμό των δυνατοτήτων από αυτά τα προφίλ, οι αρχικές συλλογές ακατέργαστων πακέτων μετατρέπονται σε ροές δικτύου για ευκολότερη ανάλυση. Η χρήση αμφίδρομων ροών CIC Flow Meter δημιουργούνται όπου το πρώτο πακέτο καθορίζει τις κατευθύνσεις προς τα εμπρός (πηγή προς προορισμό) και προς τα πίσω (προορισμός προς πηγή). Επομένως, από τα 83 στατιστικά χαρακτηριστικά που συγκεντρώθηκαν από τις ροές, όπως η διάρκεια, ο αριθμός των πακέτων, ο αριθμός των byte, το μήκος των πακέτων, αυτά υπολογίζονται ξεχωριστά τόσο για την εμπρόσθια όσο και για την αντίστροφη κατεύθυνση. Για ροές TCP τερματίζονται κατά τη διάρκεια της σύνδεσης μόλις ληφθεί ένα πακέτο FIN και οι ροές UDP τερματίζονται με χρονικό όριο ροής.

Οι ερευνητές στο [10] ανέλυσαν όλες τις επισημασμένες ροές δικτύου σε δύο ροές για ανάλυση, καλοήθης και ανώμαλες ροές. Η καλοήθης ροή αποτελείται από όλη την κυκλοφορία που περιγράφεται στο προφίλ Β και η ανώμαλη ροή είναι όλη η κυκλοφορία που περιγράφεται στο προφίλ Μ. Διαφορετικές επιθέσεις συμβαίνουν σε διαφορετικές ημέρες από ένα σύνολο 10 ημερών, δηλαδή 240 ωρών, αυτές οι επιθέσεις διασκορπίζονται τυχαία σε καλοήθη κυκλοφορία. Συνολικά υπάρχουν 2748235 ανώμαλες ροές και 6584535 καλοήθεις ροές δίνοντας συνολικά 9332770 ροές στο σύνολο δεδομένων. Πρόκειται για διαχωρισμό της καλοήθους κυκλοφορίας 70,55% και της ανωμαλίας 29,45%. Το παρακάτω διάγραμμα δείχνει την ανάλυση του όγκου κυκλοφορίας στο σύνολο δεδομένων κακόβουλης κυκλοφορίας.



Εικόνα23: Όγκοι κακόβουλης κυκλοφορίας [10]

6.1.6 Χρόνος διαδρομής άφιξης και επιλογή δυνατοτήτων

Ο χρόνος διαδρομής άφιξης (Inter-Arrival Time - IAT) μπορεί να οριστεί ως ο μέσος όρος πλαισίων, πακέτων ή ροών που φτάνουν σε έναν κεντρικό υπολογιστή για μια συγκεκριμένη χρονική περίοδο [116]. Εξετάζοντας αυτό το χαρακτηριστικό και άλλες στατιστικές μορφές του IAT, όπως η μέση, ελάχιστη, μέγιστη και τυπική απόκλιση του IAT μιας ροής δικτύου, η καλοήθης κυκλοφορία μπορεί να μοντελοποιηθεί για να συμμορφώνεται με την κατανομή Weibull. Με μοντελοποίηση της καλοήθης κυκλοφορίας στη διανομή Weibull, η ανώμαλη κυκλοφορία μπορεί να εντοπιστεί καθώς θα προκαλέσει ανωμαλίες και αποκλίσεις στη διανομή. Αυτή η συσχέτιση είναι αναγνωρίσιμη σε πακέτα, ροές και περιόδους σύνδεσης για πρωτόκολλα μεταφοράς TCP (Transmission Control Protocol) και UDP (User Datagram Protocol) στην κυκλοφορία στο Διαδίκτυο. Επομένως, αυτά τα χαρακτηριστικά ροής δικτύου IAT μπορεί να είναι ενδεικτικά της διαφοράς στις καλοήθεις και ανώμαλες ροές [10].

Λαμβάνοντας υπόψη τις προηγούμενες μελέτες σχετικά με τη συμπεριφορά ροής κυκλοφορίας κατά την άφιξη, αυτές οι έννοιες μπορούν να επεκταθούν στα τρέχοντα μοντέλα μηχανικής μάθησης για να παρέχουν σαφώς καθορισμένα επισημασμένα δεδομένα σχετικά με την ταξινόμηση μεταξύ ανώμαλων και καλοηθών ροών κυκλοφορίας. η επιλογή χαρακτηριστικών συνεπώς περιελάμβανε μια διαδικασία επιλογής δύο μερών. Το πρώτο

μέρος είναι η επιλογή τυπικών χαρακτηριστικών που παρέχουν βασικές πληροφορίες σχετικά με τη ροή. Το δεύτερο μέρος περιλαμβάνει την επιλογή ενός περιορισμένου αριθμού χαρακτηριστικών που καταδεικνύουν σαφείς διαφορές στις τιμές μεταξύ μιας καλοήθους και ανώμαλης ροής. Όπως έχει αποδειχθεί στο παρελθόν, τα δεδομένα ροής IAT και πιο συγκεκριμένα στατιστικές παραλλαγές των δεδομένων ροής IAT μπορούν να χρησιμοποιηθούν για την περαιτέρω ανάλυση αυτών των συσχετίσεων. Αυτή η απόφαση για τον περιορισμό της επιλογής χαρακτηριστικών είναι να παρέχει στο μοντέλο μηχανικής μάθησης καθαρά δεδομένα και να απομακρύνει τον υπερβολικό θόρυβο στα δεδομένα που δεν έχουν νόημα για τη συσχέτιση της σχέσης μεταξύ ανώμαλων και καλοηθών ροών. Με αυτόν τον τρόπο μπορεί να σχεδιαστεί ένα πιο αποτελεσματικό μοντέλο, με μεγαλύτερη ακρίβεια και μεγαλύτερη ταχύτητα. Επομένως, έχουν επιλεγεί 20 δυνατότητες μαζί με μια επιπλέον στήλη ετικέτας για την ταξινόμηση κάθε τύπου ροής, οι οποίες είναι:

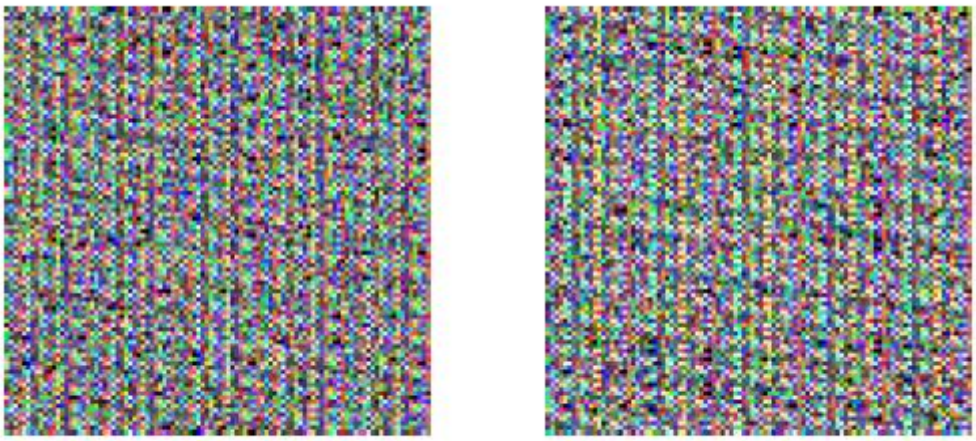
1) Βασικές δυνατότητες ροής (Basic Flow Features): Θύρα προορισμού (destination port), πρωτόκολλο (protocol), διάρκεια ροής (flow duration), συνολικά πακέτα προώθησης (total forward packets), συνολικά πακέτα προς τα πίσω (Total Backward Packets), ροές Pkts/s (δηλαδή, ροές πακέτων ανά δευτερόλεπτο).

2) Στατιστικά μεταδεδομένα IAT (IAT Statistical Metadata): Μέση IAT ροής, τυπική απόκλιση IAT ροής, μέγιστη ροή IAT, ελάχιστη ροή IAT, σύνολο IAT ροής, μέση πρόοδος IAT προς τα εμπρός, τυπική απόκλιση IAT προς τα εμπρός, μέγιστη IAT προς τα εμπρός, ελάχιστη προς τα εμπρός IAT, μέση προς τα πίσω IAT, τυπική απόκλιση IAT προς τα πίσω, μέγιστη IAT προς τα πίσω και ελάχιστη IAT προς τα πίσω.

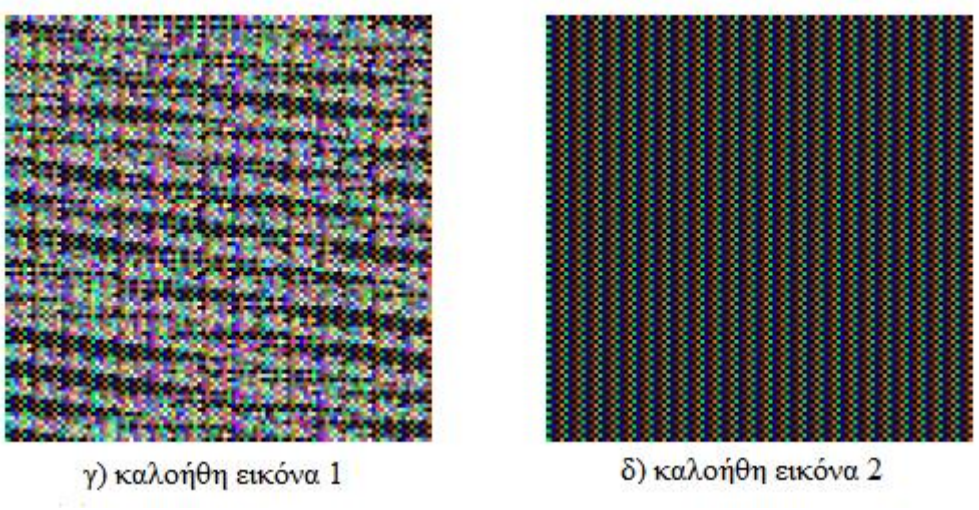
6.1.7 Προ-επεξεργασία συνόλου δεδομένων

Η προεπεξεργασία των συνόλων δεδομένων περιλαμβάνει τη μετατροπή των δεδομένων εισαγωγής στη σωστή μορφή κατάλληλη για το CNN, η οποία στην περίπτωση αυτή είναι μια εικόνα 100x100x3. Οι καθορισμένες 20 δυνατότητες από το σύνολο δεδομένων εξάγονται σε μορφή αρχείου CSV. Οι εισόδους CSV αναδιαμορφώνονται σε εικόνες RGB μεγέθους 100x100x3, τυχόν επιπλέον δεδομένα που έχουν απομείνει κάτω από αυτό το μέγεθος απορρίπτονται καθώς όλες οι εικόνες για το CNN πρέπει να έχουν το ίδιο μέγεθος εισόδου. Αυτό το μέγεθος εικόνας επιλέχθηκε λόγω της παροχής ενός μεγάλου όγκου δειγμάτων εικόνων για την ποσότητα των διαθέσιμων δεδομένων (πάνω από 1000 δείγματα εικόνων). Σε γενικές γραμμές, η αντιστάθμιση μεταξύ της χρήσης υψηλότερης σε σύγκριση με μια εικόνα χαμηλότερης ανάλυσης είναι ότι μια εικόνα υψηλότερης ανάλυσης θα περιέχει πολλές λεπτομέρειες κατά την επεξεργασία από το νευρωνικό δίκτυο, ωστόσο αυτό θα διαρκέσει

περισσότερο τόσο για τις φάσεις εκπαίδευσης όσο και για τις δοκιμές. Μια εικόνα χαμηλότερης ανάλυσης θα παρέχει λιγότερες λεπτομέρειες, αλλά περισσότερες αναπαραστάσεις χαρακτηριστικών και το νευρωνικό δίκτυο θα είναι σε θέση να εκπαιδεύσουν και να δοκιμάσουν τα δεδομένα με ταχύτερο ρυθμό. Σε αυτή την περίπτωση, τα δείγματα είναι autoML και αυξάνουν όλες τις εικόνες σε μέγεθος εικόνας εισόδου 224x224x3, επομένως υπάρχουν μόνο δύο εκτιμήσεις, πρώτον ο όγκος των εικόνων είναι πάνω από 1000 και ότι καταγράφονται επαρκείς λεπτομέρειες χαρακτηριστικών. Δύο παραδείγματα φαίνονται παρακάτω σχήματα σχετικά με το πώς φαίνεται μια ανώμαλη εικόνα σε σύγκριση με μια καλοήθη εικόνα. Σε γραφικό επίπεδο, οι ανώμαλες εικόνες είναι τυχαίες και θορυβώδεις, ενώ οι καλοήθεις εικόνες είναι πιο κανονικές και περιέχουν ορισμένα αναγνωρίσιμα μοτίβα.



Εικόνα 23: Παραδείγματα ανώμαλης εικόνας [10]



Εικόνα 24: Παραδείγματα καλοήθεις εικόνας [10]

6.1.8 Αποτελέσματα ανίχνευσης ανωμαλίας

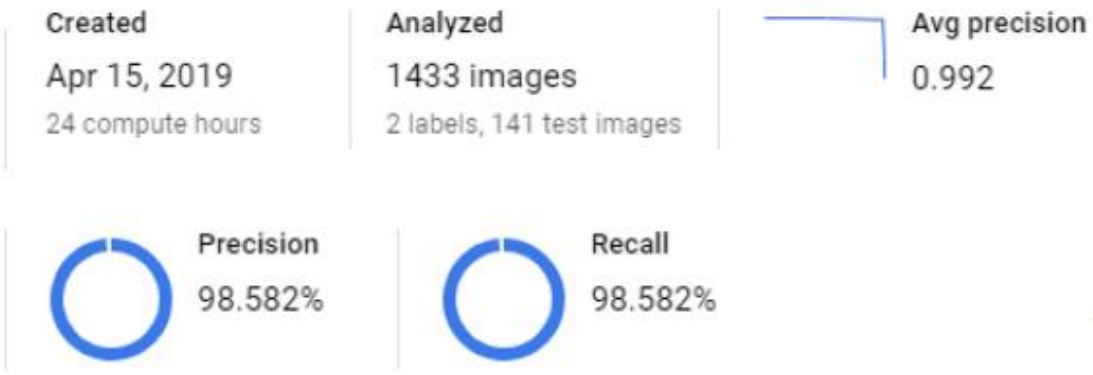
Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα του μοντέλου AutoML Visionπου εφαρμόστηκαν στα δεδομένα επεξεργασμένης εικόνας. Για τη μεγιστοποίηση της ακρίβειας του μοντέλου εκτελέστηκε προσομοίωση NASnet 24 ωρών και προσομοίωση MNasNet 3 ωρών για να συγκριθούν τα αποτελέσματα που επιτεύχθηκαν.

Αφού έχει γίνει προεπεξεργασία του συνόλου δεδομένων σε εικόνες, αυτές οι εικόνες μπορούν πλέον να μεταφορτωθούν στο Google Auto MLVision. Οι εικόνες οργανώνονται με δομή φακέλων για καλοήθεις και ανώμαλες, αυτές οι εικόνες μεταφορτώνονται στη συνέχεια στο Google Cloud Bucket storage μαζί με ένα αρχείο CSV για τη χαρτογράφηση διαδρομών αρχείων εξόδου στη σωστή ετικέτα. Το AutoML Vision αναδιαμορφώνει τα μεγέθη εικόνας εισόδου στα αναμενόμενα μεγέθη εισόδου των μοντέλων 224x224x3.

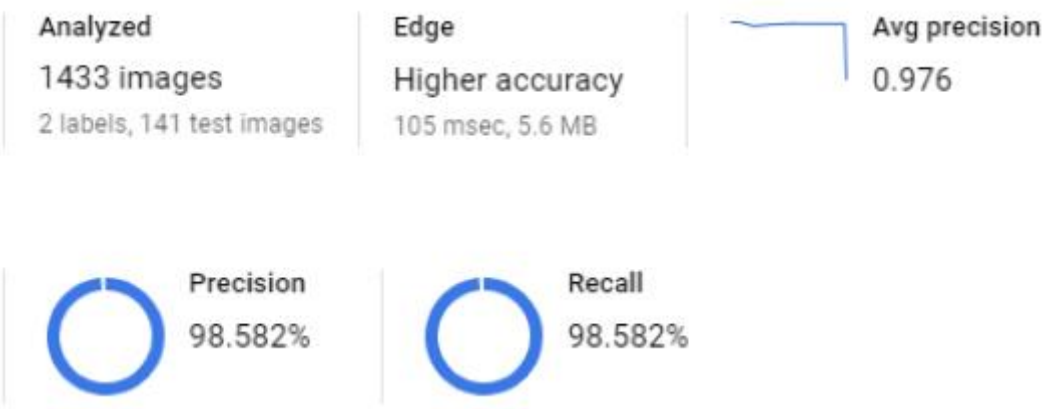
Στην παράρτημα δίνεται η εικόνα με την διάταξη του μοντέλου CNN με σχεδιασμό NAS.

Με χρόνο εκπαίδευσης 3 ωρών και χρήση 1433 εικόνων 100x100x3 χωρισμένες σε 925 καλοήθεις εικόνες και 508 ανώμαλες εικόνες. Οι 141 δοκιμαστικές εικόνες πέτυχαν μέση ακρίβεια 97,6%, μέγιστη ακρίβεια 98,582% και μέγιστη ανάκληση 98,582% για ολόκληρο το μοντέλο. Καθώς το μοντέλο έχει αναπτυχθεί στην πλατφόρμα MNasNet Edge, έχει πιο καλή ταχύτητα σε σύγκριση με τα παραδοσιακά μοντέλα NAS και δεν απαιτεί τεράστια ποσότητα υπολογιστικής ισχύος και πολλές ημέρες για εκπαίδευση. Η εκπαίδευση του μοντέλου για μεγαλύτερη ακρίβεια αποδίδει χρόνο επεξεργασίας 105ms ανά εικόνα για ένα κινητό τηλέφωνο Pixel 1.

Η εκτέλεση του μοντέλου στη δοκιμή 24 ωρών απέδωσε παρόμοια αποτελέσματα, εκτός από τη μέση ακρίβεια που ήταν ελαφρώς υψηλότερη στο 99,2%, αυτό οφείλεται στο ότι η περιοχή κάτω από την καμπύλη ακρίβειας ανάκλησης ήταν μεγαλύτερη όπως φαίνεται στα παρακάτω σχήματα.



Εικόνα 25: Αποτελέσματα δοκιμής 24 ωρών [10]



Εικόνα 26: Αποτελέσματα δοκιμής 3 ωρών [10]

Αυτό σημαίνει ότι το μοντέλο βελτιστοποιείται στο μέγιστο των δυνατοτήτων του με τα παρεχόμενα δεδομένα. Η εκπαίδευση με πρόσθετα σύνολα δεδομένων και περισσότερα δεδομένα θα κάνει μόνο το μοντέλο να έχει υψηλότερο επίπεδο απόδοσης. Όσον αφορά την εφαρμογή στον πραγματικό κόσμο, θα υπάρχει πάντα μια διαφορά μεταξύ ακρίβειας και ταχύτητας. Ωστόσο, στην περίπτωση αυτή είναι επιθυμητή η μικρή απώλεια στην ακρίβεια για πολύ μεγαλύτερο κέρδος στην απόδοση. Αυτό σημαίνει ότι η χρήση πόρων μπορεί να ελαχιστοποιηθεί, οι απειλές μπορούν να εντοπιστούν νωρίτερα και η επακόλουθη πρόσθετη εκπαίδευση και βελτίωση του μοντέλου μπορεί να ολοκληρωθεί με ταχύτερο ρυθμό.

Οι πιο συνηθισμένες μετρήσεις που χρησιμοποιούνται για τον προσδιορισμό της συνάφειας των αποτελεσμάτων, όπως ακρίβεια, ανάκληση, βαθμολογία f1 και καμπύλη ανάκλησης ακριβείας. Η ακρίβεια και η ανάκληση καθορίζονται από τα ακόλουθα στατιστικά:

True Positive (TP – Πραγματικά θετικό): Μια ανώμαλη εικόνα ταξινομείται από το μοντέλο ως ανωμαλία όταν το αποτέλεσμα είναι True Positive.

False Positive (FP – Ψευδώς θετικό): Μια ανώμαλη εικόνα ταξινομείται από το μοντέλο ως καλοήθης όταν το αποτέλεσμα είναι False Positive.

True Negative (TN – Πραγματικά αρνητικό): Μια καλοήθης εικόνα ταξινομείται από το μοντέλο ως καλοήθης όταν το αποτέλεσμα είναι True Negative.

False Negative (FN – Ψευδώς αρνητικό): Μια καλοήθης εικόνα ταξινομείται από το μοντέλο ως ανωμαλία όταν το αποτέλεσμα είναι False Negative.

Η ακρίβεια μπορεί να οριστεί ως το ποσοστό των θετικών προβλέψεων που είναι σωστές και η ανάκληση μπορεί να οριστεί ως το ποσοστό των θετικών περιπτώσεων που εντόπισε ο ταξινομητής. Μαθηματικά αυτό μπορεί να υπολογιστεί ως:

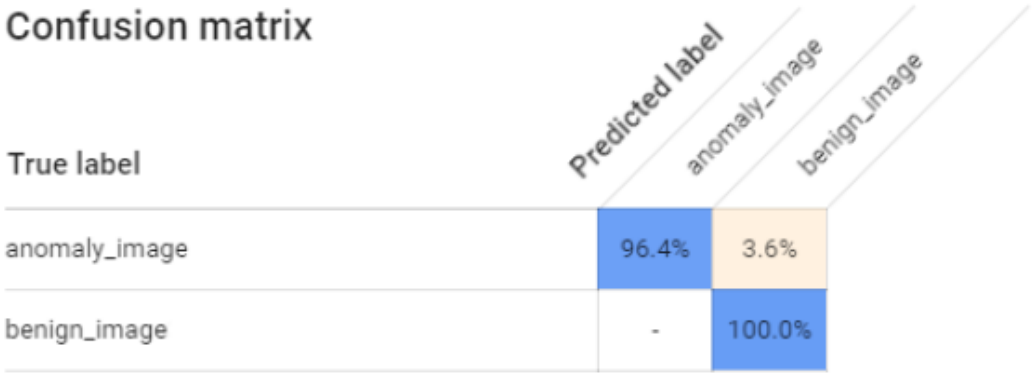
$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \quad \mathrm{Re}call = \frac{TP}{TP + FN}$$

Και από αυτά τα στατιστικά στοιχεία μπορεί να υπολογιστεί και η βαθμολογία F1 που παρέχει τον αρμονικό μέσο ακρίβειας και ανάκλησης:

$$F1score = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Ο πίνακας σύγχυσης στο παρακάτω σχήμα δείχνει την ποσοστιαία πρόβλεψη του πότε ο ταξινομητής επέλεξε τη σωστή απάντηση, στην περίπτωση αυτή το 96,4% των ανώμαλων εικόνων αναγνωρίζονται ως ανώμαλες εικόνες και το 3,6% των ανώμαλων εικόνων

αναγνωρίστηκαν εσφαλμένα ως καλοήθεις εικόνες. Για τις καλοήθεις εικόνες, το 0% των καλοηθών εικόνων είχε προβλεφθεί εσφαλμένα ως εικόνες ανωμαλίας και το 100% των καλοηθών εικόνων ταχτοποιήθηκαν σωστά.



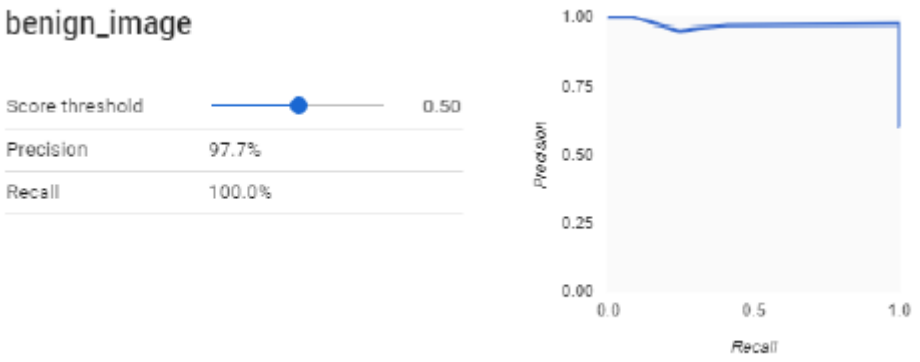
Εικόνα 27: πίνακας αναγνώρισης καλοηθών και ανώμαλων εικόνων [10]

Τα βασικά στατιστικά στοιχεία απόδοσης παρουσιάζονται στον παρακάτω πίνακα για ευκολότερη ορατότητα και σύγκριση. Η καλοήθης κυκλοφορία έχει ανάκληση ακρίβειας 100%, αλλά έχει ακρίβεια 97,7% λόγω της εσφαλμένης ταξινόμησης ενός μικρού ποσοστού εικόνων ανωμαλιών ως καλοήθων. Η κίνηση της ανωμαλίας έχει ακρίβεια 100%, επειδή όλες οι ανωμαλίες που εντοπίστηκαν ταξινομήθηκαν σωστά, αλλά έχει ανάκληση 96,4% λόγω της έλλειψης ορισμένων ανωμαλιών και της εσφαλμένης ταξινόμησής τους ως καλοήθους.

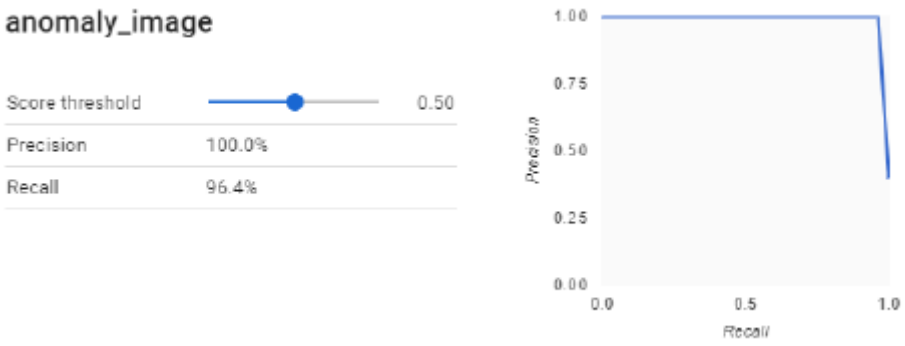
Πίνακας 1: Στατιστικά απόδοσης[10]

Αποτελέσματα 24 και 3 ωρών			
Τύπος κυκλοφορίας	**Ακρίβεια**	**Ανάκληση**	**Βαθμολογία F1**
Καλοήθης	0.977	1	0.988
Ανώμαλη	1	0.964	0.982
Μέση τιμή	0.9885	0.982	0.985

Η ακρίβεια έναντι της ανάκλησης είναι μια αντιστάθμιση και αυτό φαίνεται στα παρακάτω, όπου παρουσιάζονται οι καμπύλες ανάκλησης ακρίβεια. Το όριο βαθμολογίας ορίζεται στο 0,5 για να ισορροπήσει ομοιόμορφα αυτές τις μετρήσεις. Για να απλοποιηθούν αυτές οι μετρήσεις σε έναν αριθμό, μπορεί να χρησιμοποιηθεί η βαθμολογία F1, η οποία ζυγίζει ομοιόμορφα τόσο την ακρίβεια όσο και την ανάκληση, οι τιμές βαθμολογίας F1 εμφανίζονται παραπάνω. Η βαθμολογία F1 είναι μια σημαντική μέτρηση για αυτό το μοντέλο, καθώς τόσο η ανάκληση όσο και η ακρίβεια πρέπει να ληφθούν υπόψη κατά την ανίχνευση της ανώμαλης κυκλοφορίας. Σε μια πραγματική εφαρμογή, ένα IDS πρέπει να ελαχιστοποιήσει το ποσό των καλοήθων ροών κυκλοφορίας που αναγνωρίζονται ως ανωμαλίες όσο το δυνατόν περισσότερο, ενώ προσπαθούν ακόμη να μεγιστοποιήσουν τα ποσοστά ανίχνευσης της πραγματικής ανώμαλης κίνησης.



Εικόνα 28: Καλοήθη δεδομένα [10]



Εικόνα 29: Ανώμαλα δεδομένα [10]

6.1.9 Συμπεράσματα περίπτωσης

Υπάρχουν πολλοί τομείς που πρέπει να ληφθούν υπόψη κατά την προσπάθεια ασφάλειας στο 5G, καθώς το δίκτυο είναι τόσο διαφορετικό από τα δίκτυα της προηγούμενης γενιάς, είναι πιο δύσκολο να εφαρμοστεί αποτελεσματικά η ασφάλεια. Στην συγκεκριμένη περίπτωση εξετάστηκε η παρακολούθηση της κυκλοφορίας ενός δικτύου, ωστόσο μια περεταίρω έρευνα θα περιλάμβανε την παρακολούθηση της κυκλοφορίας του δικτύου από άκρο σε άκρο. Βέβαια, και στις δύο περιπτώσεις αυτό δεν αποτελεί μια οριστική λύση ασφαλείας, αλλά μόνο ένα μέρος της συνολικής αρχιτεκτονικής ασφαλείας που απαιτείται για την ασφάλεια ενός δικτύου.

Το μοντέλο που παρουσιάστηκε σχεδιάστηκε με τα επιλεγμένα βασικά και IAT χαρακτηριστικά, όπου με τη χρήση της αυτόματης μηχανικής μάθησης (autoML) ταξινομήθηκαν σωστά όλες οι καλοήθεις ροές κυκλοφορίας, κάτι που είναι πολύ καλό αποτελέσματα, ωστόσο για τις ανώμαλες ροές κυκλοφορίας ταξινομήθηκε σωστά το 96,4%, που σημαίνει ότι υπάρχει περιθώριο για βελτίωση.

Συγκρίνοντας τις δύο αρχιτεκτονικές που εφαρμόστηκαν, MNasNet (3 ώρες λειτουργίας) και NASNet 24 ώρες λειτουργίας), υπήρχε ελάχιστη διαφορά μεταξύ τους, το οποίο μπορεί να οφείλεται στο μέγεθος του συνόλου δεδομένων. Για περαιτέρω επικύρωση μοντέλου, οι δοκιμές μπορούν να διεξαχθούν με ένα μεγαλύτερο σύνολο δεδομένων, οι δοκιμές μπορούν επίσης να διεξαχθούν με διαφορετικά σύνολα δεδομένων για να διασφαλιστεί ένας λογικός βαθμός γενίκευσης στο μοντέλο και για να ελεγχθούν τυχόν προβλήματα τοποθέτησης.

Επίσης, αυτό το μοντέλο θα μπορούσε να επεκταθεί περαιτέρω με εφαρμογή τεχνικών μάθησης χωρίς επίβλεψη για τη δημιουργία ενός μοντέλου με ημιεπίβλεψη, καθώς στην πραγματικότητα η πλειονότητα της κυκλοφορίας δικτύου είναι δεδομένα χωρίς ετικέτα και η

προεπεξεργασία δεδομένων χωρίς ετικέτα σε καθαρά και οργανωμένα δεδομένα με ετικέτα είναι μια χρονοβόρα διαδικασία [10].

6.2 Συνεχής έλεγχος ταυτότητας με δυναμική πληκτρολόγηση σε διαδικτυακές εφαρμογές

Σε αυτή την περίπτωση μελετάται ο συνεχής έλεγχος ταυτότητας ή αλλιώς η συνεχής αυθεντικοποίηση σε χρήστες διαδικτυακών εφαρμογών, όπως αυτό μελετήθηκε και σχεδιάστηκε στο [119]. Σε αυτή τη μελέτη-έρευνα σχεδιάστηκε ένα σύστημα συνεχούς ελέγχου ταυτότητας βάσει της συμπεριφοράς του χρήστη και συγκεκριμένα βάσει του τρόπου πληκτρολόγησης του χρήστη. Με την έννοια συνεχής νοείται ότι ο έλεγχος ταυτότητας πραγματοποιείται συνεχώς, μέσω της παρακολούθησης της συμπεριφοράς του χρήστη καθ' όλη τη διάρκεια της σύνδεσης του χρήστη του συστήματος, δηλαδή, ο έλεγχος ταυτότητας δεν γίνεται μόνο κατά την είσοδο (log-in) του χρήστη στο σύστημα. Ουσιαστικά, το σύστημα συνεχούς ελέγχου ταυτότητας που σχεδιάστηκε εντοπίζει κάποια πιθανή απόκλιση από την τυπική συμπεριφορά, και έτσι εντοπίζει αν είναι αυθεντικός ή όχι ένας χρήστης.

6.2.1 Μεθοδολογία

Σε αυτή την ενότητα παρουσιάζεται η μεθοδολογία που ακολουθήθηκε στο [119] για το σχεδιασμό του συστήματος συνεχούς ελέγχου ταυτότητας.

6.2.1.1 1ο βήμα: χρήση δυναμικής πληκτρολόγησης χρηστών

Στο πρώτο βήμα έγινε η επιλογή του βιομετρικού στοιχείου συμπεριφοράς του χρήστη που χρησιμοποιήθηκε για τον έλεγχο ταυτότητας του χρήστη. Το στοιχείο αυτό ανήκει στη δυναμική πληκτρολόγηση, (σημειώνεται ότι, η δυναμική πληκτρολόγηση παρουσιάστηκε στο προηγούμενο κεφάλαιο), και συγκεκριμένα εφαρμόστηκε ο συνεχής έλεγχος ταυτότητα με του χρόνους των keystroke digraphs του κάθε χρήστη. Ο όρος keystroke digraphs αναφέρεται στην διαδοχική πληκτρολόγηση δύο πλήκτρων, όπως για παράδειγμα το "ka" και το "at". Στα ελληνικά ο όρος αυτός μπορεί να μεταφραστεί ως δίγραφοι πληκτρολογίου.

Σε ευρύτερο πλαίσιο, η παρακολούθηση των keystroke digraphs μπορεί να γίνει συλλογή αρκετών χρονικών πληροφοριών, όπως η χρονική διαφορά μεταξύ της πίεσης του πρώτου και του δεύτερου πλήκτρου, ή οι χρόνοι από την ελευθέρωση του πρώτου και του δεύτερου πλήκτρου.

Στη συγκεκριμένη περίπτωση, στο [119], από κάθε keystroke digraph έγινε εξαγωγή τριών χρονικών πληροφοριών: 1) χρόνος πίεσης πρώτου πλήκτρου, 2) χρόνος πίεσης δεύτερου πλήκτρου, και 3) χρόνος μεταξύ ελευθέρωσης πρώτου πλήκτρου και πίεσης δεύτερου πλήκτρου.

Η αναπαράσταση αυτών των χρονικών πληροφοριών έγινε σε μορφή διανύσματος Nx3, όπου το Nσυμβολίζει το πλήθος των δειγμάτων για κάθε digraph. Για κάθε λέξει με nγράμματα υπάρχουνn-1 diagraphs, για παράδειγμα η λέξει “world” έχει 5 γράμματα και 4 diagraphs (wo, or, rl, ld). Άρα, για τη λέξη“world” θα σχηματιστούν 4 διανύσματα με διάσταση 1x3.

Τα στοιχεία που καταγράφονται και αποθηκεύονται στο σύστημα από την πληκτρολόγηση των χρηστών είναι: 1) το όνομα του πλήκτρου (π.χ. για το πλήκτρο ‘a’ το όνομα είναι KeyA), 2) το γεγονός (event), το οποίο μπορεί να είναι η πίεση του πλήκτρου (KeyDown) ή η ελευθέρωση του πλήκτρου (KeyUp), 3) η χρονοσφραγίδα (timestamp), δηλαδή το πλήθος των δευτερολέπτων σε μια συγκεκριμένη μέρα. Στη συνέχεια δίνεται ένα σχετικό παράδειγμα πληκτρολόγησης για τη λέξει “you”.

{event: ‘KeyDown’, key: ‘KeyY’, timestamp: 135345554}
{event: ‘KeyUp’, key: ‘KeyY’, timestamp: 135345754}
{event: ‘KeyDown’, key: ‘KeyO’, timestamp: 1353455800}
{event: ‘KeyDown’, key: ‘KeyU’, timestamp: 1353455830}
{event: ‘KeyUp, key: ‘KeyO’, timestamp: 1353455900}
{event: ‘KeyUp’, key: ‘KeyU’, timestamp: 1353455950}

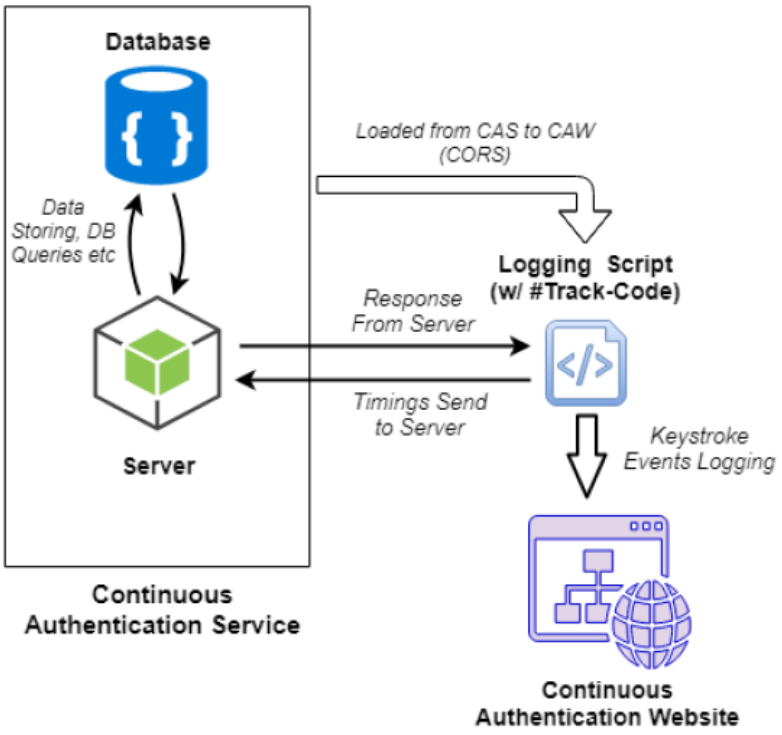
Εικόνα 30: Καταγραφή γεγονότων (events) πληκτρολόγησης [119]

6.2.1.2 2ο βήμα: σχεδίαση υπηρεσίας συνεχούς ελέγχου ταυτότητας

Για την συλλογή των μετρήσεων του χρήστη, δηλαδή στην συγκεκριμένη περίπτωση τα keystroke digraphs, σχεδιάστηκε μια νέα web υπηρεσία, στην οποία μπορεί να συνδεθεί ο διαχειριστής-ιδιοκτήτης (admin) ενός ιστότοπου (στον οποίο ενδεχομένως υπάρχουν ευαίσθητα δεδομένα), και έτσι μπορεί να εφαρμοστεί ο συνεχείς έλεγχος ταυτότητας δυναμικής πληκτρολόγησης. Μέσω αυτής της υπηρεσίας γίνεται συνεχής συλλογή δεδομένων πληκτρολόγησης σε πραγματικό χρόνο (real time) και εφαρμόζεται συνεχής έλεγχος ταυτότητας. Με αυτό τον τρόπο, όταν η πληκτρολόγηση ενός χρήστη αποκλίνει από την τυπική του πληκτρολόγηση, το σύστημα ελέγχου ταυτότητας αντιλαμβάνεται αυτή την ανωμαλία και λαμβάνει αποφάσεις για να προβεί στις κατάλληλες ενέργειες.

Στο [119] ο συγγραφέας ονομάζει την υπηρεσία ελέγχου ταυτότητας ως CAS (Continuous Authentication Service) και τον ιστότοπο στον οποίο γίνεται η εφαρμογή της υπηρεσίας ως

CAW (Continuous Authentication Website). Επίσης, ο ιδιοκτήτης του CAW καλείται admin και ο χρήστης που περιηγείται και κάνει ενέργειες στον CAW ονομάζεται subject, και σε αυτό γίνεται η εφαρμογή του συνεχούς ελέγχου ταυτότητας. Στην ακόλουθη εικόνα δίνει η γενική δομή και η αλληλεπίδραση μεταξύ CAS και CAW.



Εικόνα 31: Δομή υπηρεσίας συνεχούς ελέγχου ταυτότητας (CAS) [119]

Όπως φαίνεται στην παραπάνω εικόνα, η επικοινωνία ανάμεσα στο CAS και το CAW γίνεται με τη βοήθεια ενός logging script, η οποία φορτώνεται από το CAS στο CAW με ένα αίτημα CORS [119]. Το CORS είναι η συντομογραφία του Cross-Origin Resource Sharing, δηλαδή του διαμοιρασμού πόρων μεταξύ προέλευσης. Πιο συγκεκριμένα, το CORS αποτελεί ένα μηχανισμό HTTP κεφαλής (HTTP-header), ο οποίος επιτρέπει σε ένα διακομιστή (server) να υποδεικνύει οποιαδήποτε άλλη προέλευση (τομέα (domain), σχήμα (scheme) ή θύρα (port) από τη δική του από την οποία ένα πρόγραμμα περιήγησης πρέπει να επιτρέπει τη φόρτωση πόρων. Το CORS βασίζεται επίσης σε έναν μηχανισμό με τον οποίο τα προγράμματα περιήγησης υποβάλλουν ένα αίτημα “preflight” στον διακομιστή που φιλοξενεί τον πόρο πολλαπλής προέλευσης, προκειμένου να ελέγξει εάν ο διακομιστής θα επιτρέψει το πραγματικό αίτημα. Σε αυτό το preflight, το πρόγραμμα περιήγησης στέλνει κεφαλίδες που υποδεικνύουν τη μέθοδο HTTP και κεφαλίδες που θα χρησιμοποιηθούν στο πραγματικό αίτημα [125].

Επομένως, το logging script τοποθετείται στο CAW και αναλαμβάνει την καταγραφή των μετρήσεων πληκτρολόγησης του χρήστη, οι οποίες αποστέλλονται περιοδικά στο CAS, στο οποίο γίνεται αποθήκευση αυτών των μετρήσεων σε μια βάση δεδομένων. Στη συνέχεια, από το CAS επιστρέφονται κατάλληλες απαντήσεις σχετικά με τα δεδομένα πληκτρολόγησης του χρήστη, δηλαδή αν ο χρήστης είναι ή όχι ύποπτος.

Όπως φαίνεται και στην παραπάνω σχεδιάγραμμα της εικόνας (Εικόνα 31), το logging script περιλαμβάνει την ιδιότητα track code, η οποία ουσιαστικά αποτελεί ένα μοναδικό κώδικα αναγνώρισης για τον κάθε CAW. Έτσι, ο admin δεν χρειάζεται να κάνει κάποια επιπλέον ενέργεια για να αποκτήσει την συγκεκριμένη υπηρεσία ελέγχου ταυτότητας. Ενώ για τον κάθε χρήστη ενός συγκεκριμένου CAW χρησιμοποιείται το username, το οποίο πρέπει επίσης να είναι μοναδικό, το οποίο επίσης αποθηκεύεται στη βάση δεδομένων του CAS.

Αξίζει να σημειωθεί ότι αν ένας χρήστης με κακόβουλη πρόθεση μάθει το track code ενός ιστότοπου, το οποίο είναι αρκετά εύκολο, καθώς το script εμφανίζεται στον πρόγραμμα περιήγησης πελάτη (client browser). Έτσι, μπορεί να εφαρμόσει σε ένα δικό του ιστότοπο και να στείλει μη αληθή δεδομένα στο CAS, γεγονός που θα οδηγήσει στην καταστροφή των υπαρχόντων δεδομένων. Έτσι, ο ερευνητής στο [119] έλαβε συγκεκριμένα μέτρα για την αποφυγή μιας τέτοιας κακόβουλης επίθεσης. Συγκεκριμένα, εκτός από το track code συμπεριλαμβάνεται και το όνομα τομέα (domain name) του ιστότοπου.

Σχετικά με την οργάνωση της βάσης δεδομένων στην οποία θα αποθηκεύονται τα δεδομένα από την πληκτρολόγηση των χρηστών χρησιμοποιήθηκε η MongoDB. Επισημαίνεται ότι η MongoDB αποτελεί μια noSQL βάση δεδομένων η οποία χρησιμοποιεί τύπο εγγράφου JSON, δηλαδή, η βάση δεδομένων οργανώνεται σε συλλογές, κάθε συλλογή (collection) περιέχει JSON έγγραφα, και κάθε έγγραφο (document) περιλαμβάνει ζευγάρια από πεδία (field) και τιμές (value). Στην ακόλουθη εικόνα δίνεται σχετικό παράδειγμα.

Εικόνα 32: Παράδειγμα οργάνωσης MongoDB [119]

Η βάση δεδομένων του συστήματος ελέγχου ταυτότητας με δυναμική πληκτρολόγηση θα πρέπει να διαθέτει: 1) λογαριασμούς admin (δηλαδή, για τους ιδιοκτήτες των ιστότοπων που επιθυμούν να συμπεριλάβουν την υπηρεσία, 2) δυνατότητα δημιουργίας πολλαπλών project των admin, 3) αποθήκευση δεδομένων πληκτρολόγησης του κάθε subject (χρήστη)και αντιστοίχηση του κάθε χρήστη με συγκεκριμένο Project, 4) αποθήκευση δεδομένων εκπαίδευσης για τη δημιουργία του προφίλ πληκτρολόγησης των χρηστών. Επομένως, δημιουργήθηκαν δύο συλλογές (collections): 1) Admin Collection και 2) Keystrokes Collection.

Στην ακόλουθη εικόνα παρατίθεται σε μορφή JSON εγγράφου το Admin Collection, όπου για κάθε χρήστη αποθηκεύονται ορισμένα προσωπικά στοιχεία, δηλαδή, username, firstname, lastname, password, date, καθώς και το projectγια το οποίο αποθηκεύονται συγκεκριμένα στοιχεία, δηλαδή, siteurl (δηλαδή το όνομα τομέα), track_code, οι υπόλοιπες παράμετροι σχετίζονται με το μοντέλο εκπαίδευσης και περιγράφονται πιο κάτω.

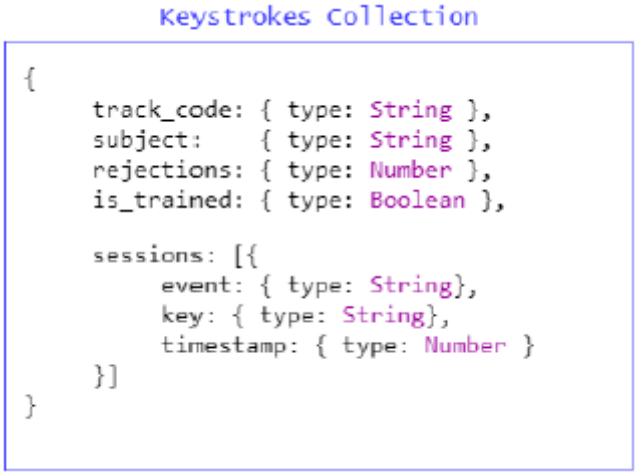
Admin Collection

```
{
     username:  { type: String },
     firstname: { type: String },
     lastname:  { type: String },
     password:  { type: String },
     date:      { type: Date },

     projects: [{
          date: { type: Date },
          siteurl: { type: String },
          track_code: { type: String },
          last_train_date: { type: Date },
          enable_keyguard_auth_flag: { type: Boolean },
          testing_threshold: { type: Number },
          keystroke_code_collect_period: { type: Number },
          training_algorithm: { type: String },
          timing_limits: {
              key_hold: {
                          min: { type: Number },
                          max: { type: Number }
                        },
              up_down:  {
                          min: { type: Number },
                          max: { type: Number }
                        }
          },
          training_parameters: {
              GMM: {
                       delta: { type: Number },
                       n_components: { type: Number }
                   },
              ONE_CLASS_SVM: {
                                  gamma: { type: Number }
                              }
          }
     }]

}
```

Εικόνα 33: Admin Collection[119]

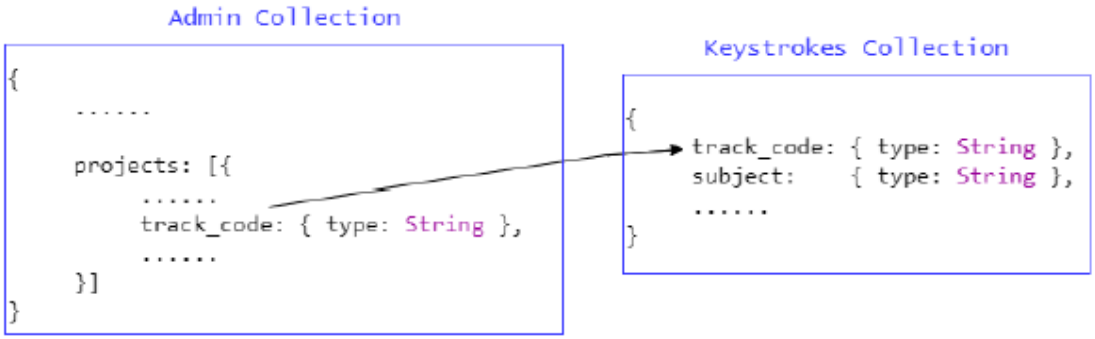
Στην ακόλουθη εικόνα παρουσιάζεται σε τύπο εγγράφου JSON η μορφή μιας Keystrokes Collection, όπου για κάθε χρήστη (subject) του ιστότοπου (CAW) αναπτύσσεται και μια

διαφορετική Keystrokes Collection και στην αποθηκεύονται το track_code, το subject (δηλαδή, το username του χρήστη), sessions (δηλαδή, οι συνεδρίες της πληκτρολόγησης, οι υπόλοιπες παράμετροι σχετίζονται με το μοντέλο εκπαίδευσης και περιγράφονται πιο κάτω.



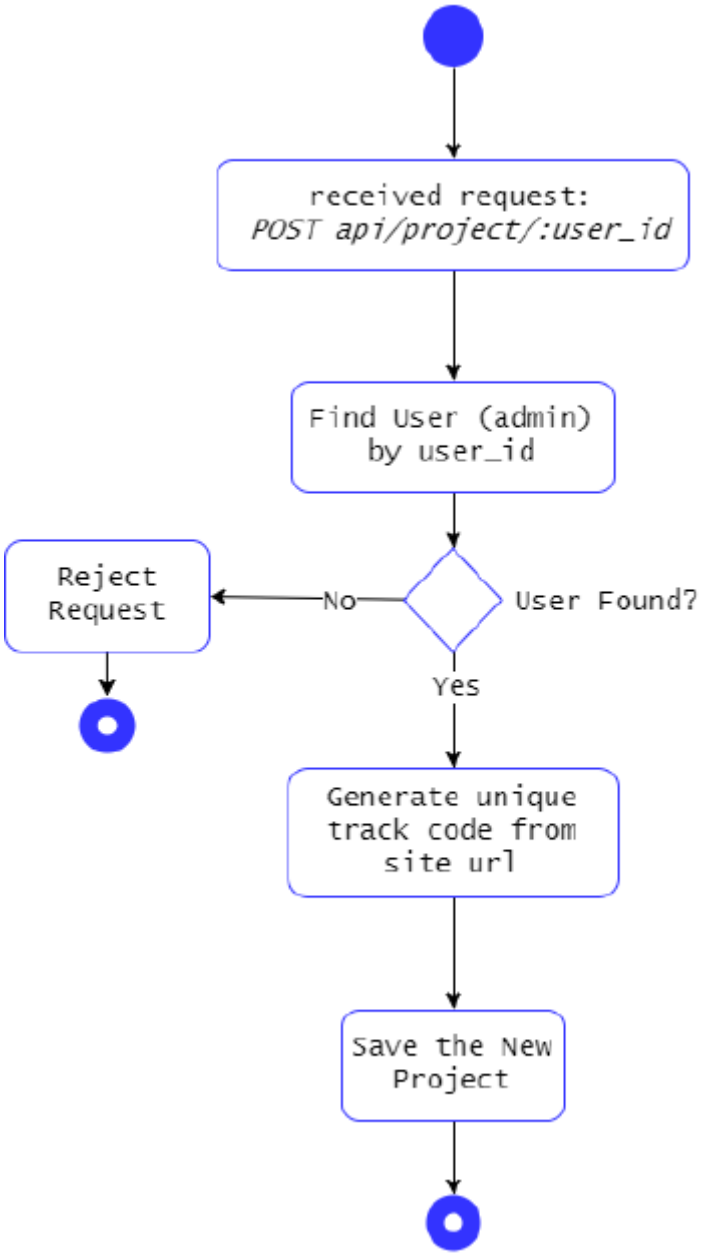
Εικόνα 34: Keystrokes Collection[119]

Στην παρακάτω εικόνα παρουσιάζεται ο τρόπος με τον οποίο συνδέονται οι δυο συλλογές μέσω του track_code.



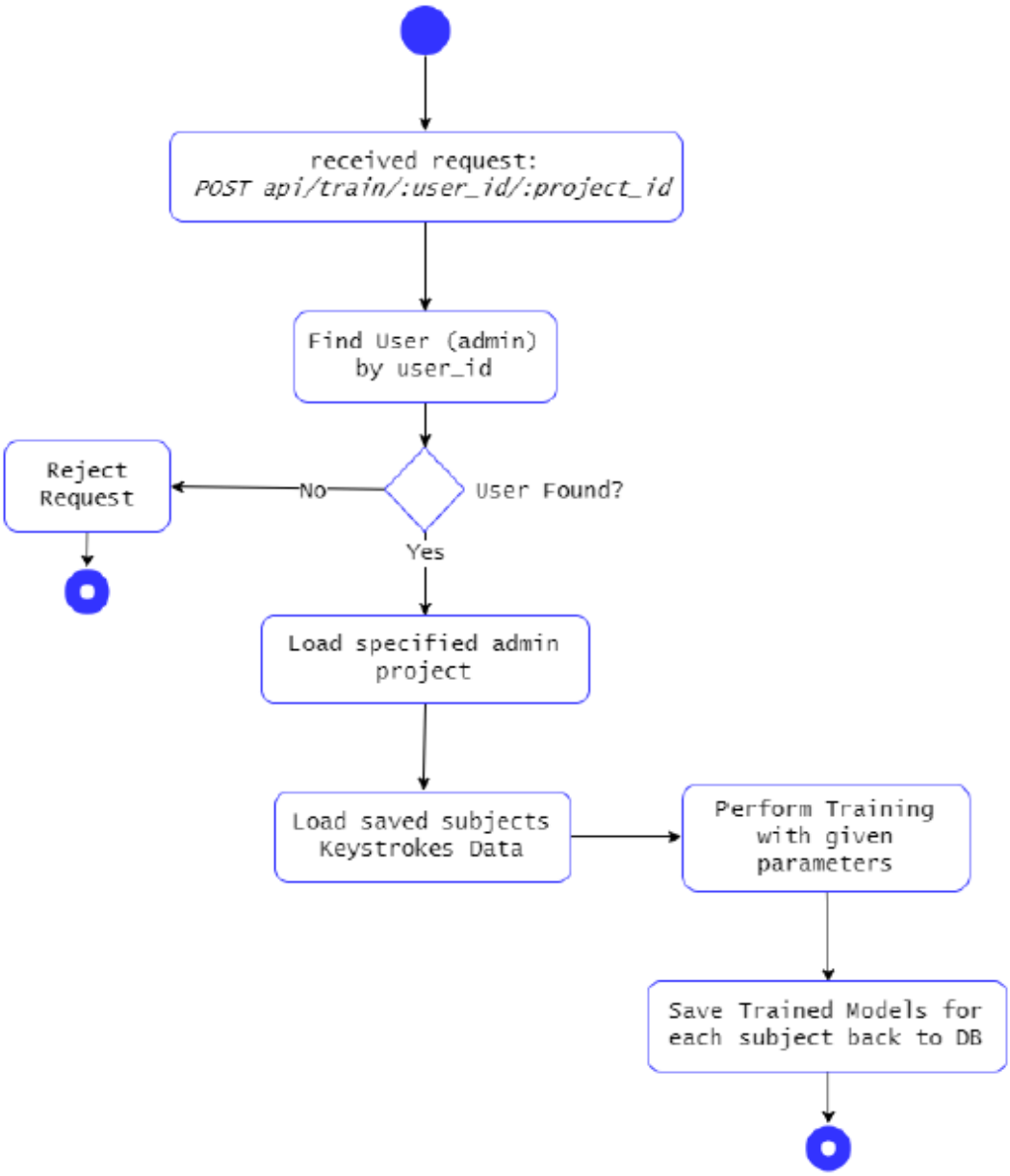
Εικόνα 35: Σύνδεση δύο συλλογών (Adminκαι Keystrokes)[119]

Για τις ανάγκες ανάπτυξης του back-end server, χρησιμοποιήθηκε η γλώσσα προγραμματισμού Node.JS της έκδοσης 8.9.0. Μέσω αυτή της τεχνολογίας αναπτύχθηκαν δύο εφαρμογές (API's): admin API και collect API. Στο διάγραμμα της ακόλουθης εικόνας αποτυπώνεται ένα ενδεικτικό διάγραμμα ροής για τη λειτουργία της δημιουργίας ενός project.



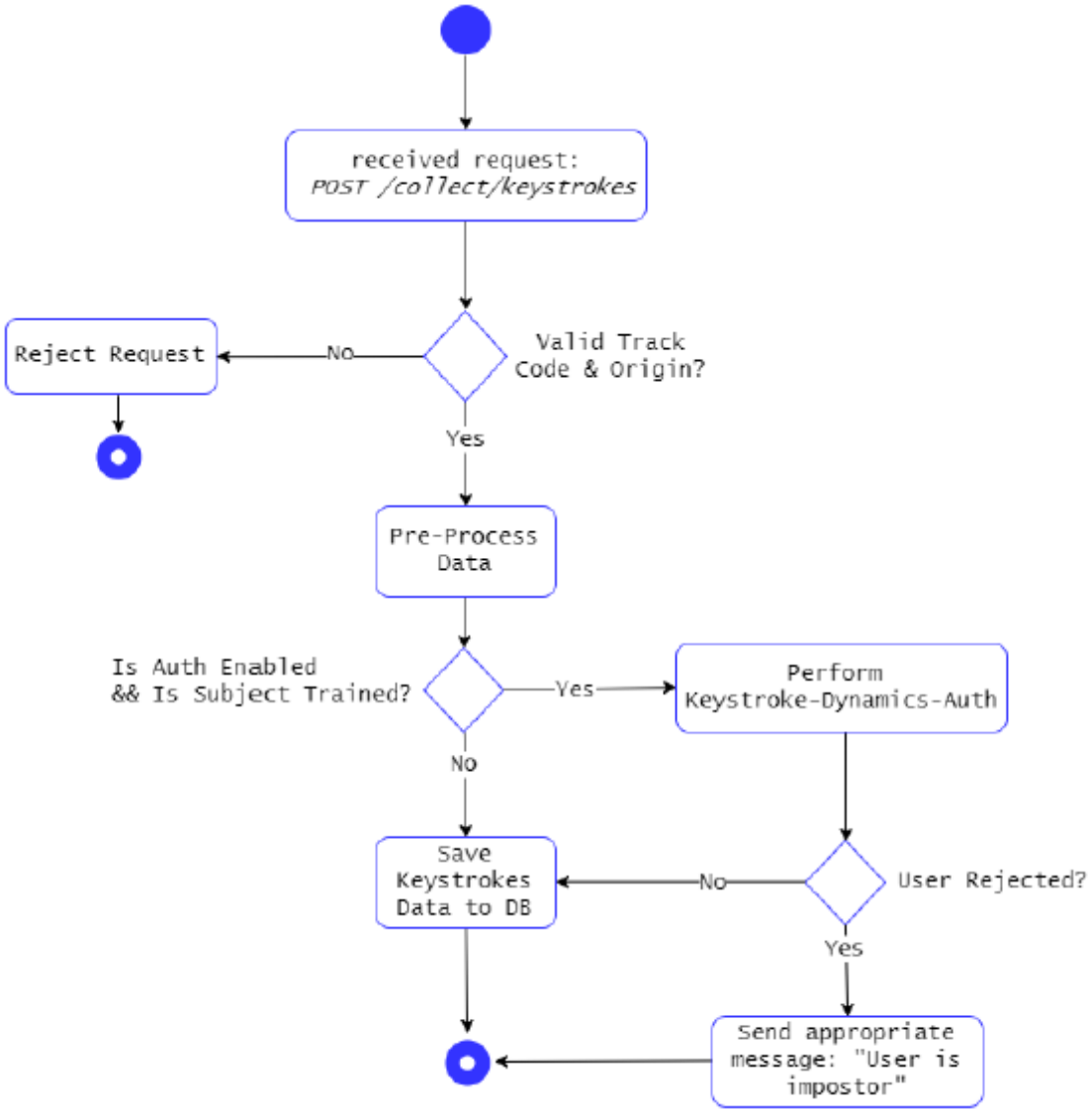
Εικόνα 36: Διάγραμμα ροής για τη δημιουργία project στο admin API [119]

Στο διάγραμμα της ακόλουθης εικόνας παρουσιάζεται το διάγραμμα ροής για την εκπαίδευση των δεδομένων που συλλέγονται από τους χρήστες.



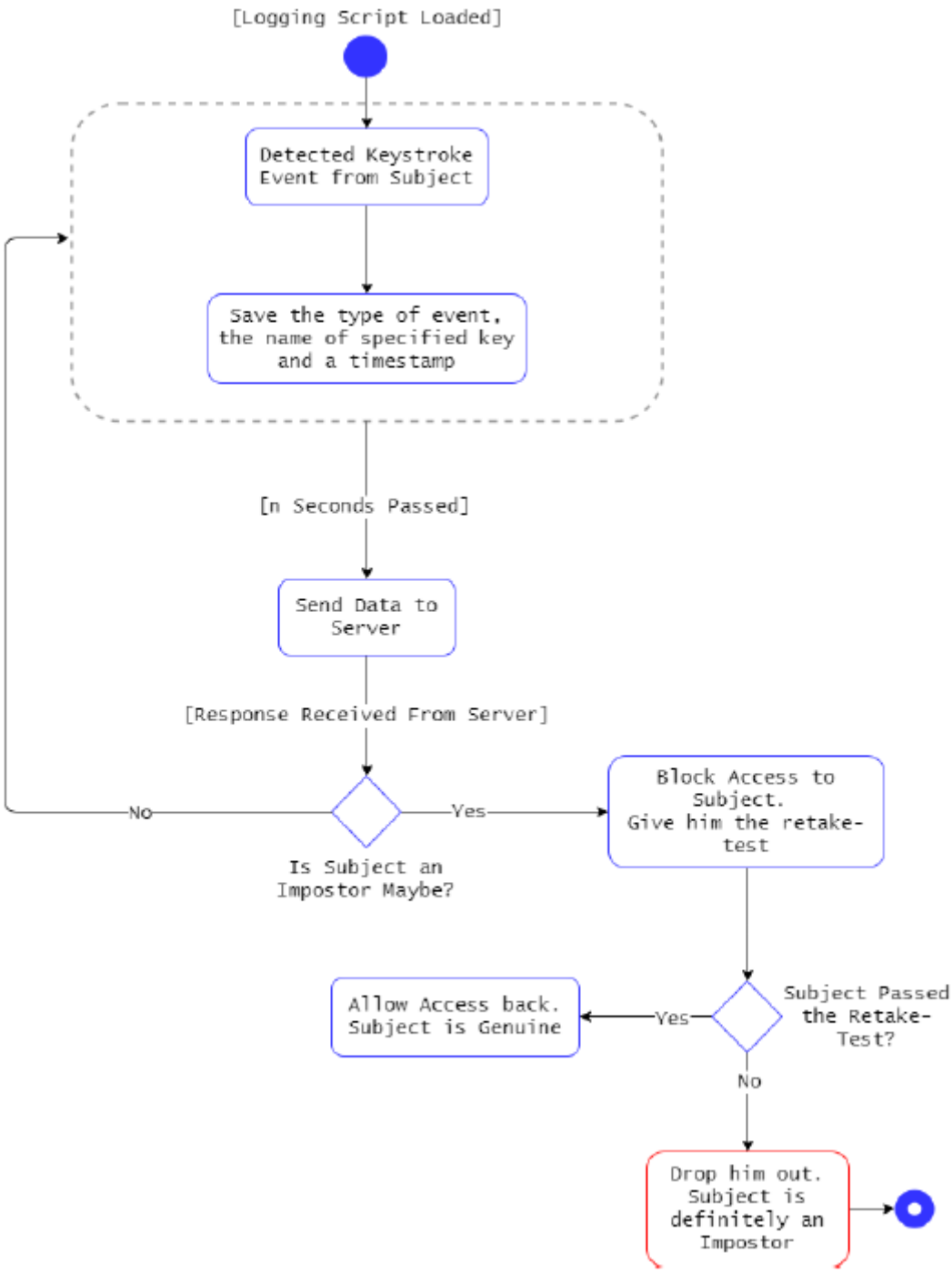
Εικόνα 37: Διάγραμμα ροής εκπαίδευσης δεδομένων των subjectsστο admin API [119]

Το collect API επικοινωνεί με το logging script, όπως εξηγήθηκε νωρίτερα. Στον διάγραμμα ροής της ακόλουθης εικόνας παρουσιάζεται η λειτουργία του.



Εικόνα 38: Διάγραμμα ροής για τη λειτουργία του CollectAPI [119]

Στην ακόλουθη εικόνα παρουσιάζεται το διάγραμμα ροής του logging script, η αποστολή του οποίου εξηγήθηκε νωρίτερα.



Εικόνα 39: Διάγραμμα ροής για τη λειτουργία του logging script [119]

6.2.1.3 3ο βήμα: εξαγωγή δεδομένων και προεπεξεργασία

Όπως ήδη έχει γίνει αντιληπτό από τη μέχρι τώρα περιγραφή, τα δεδομένα που λαμβάνονται από τον διακομιστή μέσω του logging script αποτελούν απλές καταγραφές γεγονότων (events), δηλαδή δεν είναι σε διανυσματική μορφή, η οποία να μπορεί να αναλυθεί και γίνει η

αναγνώριση προτύπων. Επομένως, πριν από την εκπαίδευση των δεδομένων, θα πρέπει πρώτα να γίνει μετάφραση των γεγονότων. Ουσιαστικά, τα δεδομένα που λαμβάνονται από το logging script έχουν την εξής μορφή:

```
[ { event: { type: String }, key: { type: String }, timestamp: { type: Number } } ]
```

Έτσι, χρησιμοποιήθηκε ο ακόλουθος αλγόριθμος για την μετάφραση τον δεδομένων, δηλαδή, ο αλγόριθμος εξαγωγής για τα χρονικά χαρακτηριστικά των digraphs και τα γεγονότα πληκτρολόγησης (keystroke events).

Algorithm 1: Extraction of Digraph Timings

Input : Array of Keystroke Events *arr*
Output: Digraph Timings

while *length of arr* ≥ 2 **do**
  $key\_1\_down\_event \longleftarrow arr[0]$;
  Find the respective $key\_1\_up\_event$;
  Find the following $key\_2\_down\_event$ *after* $key\_1\_down\_event$;
  Find the respective $key\_2\_up\_event$;
  $digraph \longleftarrow key\_1\_down\_event.key + key\_2\_down\_event.key$;
  $key\_1\_hold_t \longleftarrow key\_1\_up\_event.timestamp - key\_1\_down\_event.timestamp$;
  $key\_2\_hold_t \longleftarrow key\_2\_up\_event.timestamp - key\_2\_down\_event.timestamp$;
  $digraph\_up\_down_t \longleftarrow key\_2\_down\_event.timestamp - key\_1\_up\_event.timestamp$;
  $V \longleftarrow [key\_1\_hold_t, key\_2\_hold_t, digraph\_up\_down_t]$;
  Store vector V of Digraph Appropriately;
  Pop $key\_1\_down\_event$ *and* $key\_1\_up\_event$ *from arr;*
end

Εικόνα 40: Αλγόριθμος εξαγωγής για τα χρονικά χαρακτηριστικά των digraphs και τα γεγονότα πληκτρολόγησης (keystroke events) [119]

6.2.1.4 4ο βήμα: εκπαίδευση και πρόβλεψη

Η υπηρεσία συνεχούς ελέγχου ταυτότητας διαθέτει δύο διαφορετικούς αλγορίθμους για την εκπαίδευση και την πρόβλεψη από τα δεδομένα πληκτρολόγησης.

Ο πρώτος είναι ο αλγόριθμος One Class SVM, ο οποίος αναλαμβάνει και την εκπαίδευση και την πρόβλεψη, και ο άλλος είναι ο Gaussian Mixture Models που αναλαμβάνει την εκπαίδευση, ενώ η πρόβλεψη γίνεται με την εφαρμογή κατάλληλα προσαρμοσμένου

αλγόριθμου. Για τους οποίους έχουν δοθεί περισσότερες πληροφορίες στο προηγούμενο κεφάλαιο.

Ειδικότερα, η τεχνική της εκπαίδευσης και πρόβλεψης με τη χρήση του αλγορίθμου One-Class SVM χρησιμοποιήθηκε για την εξάλειψη των εξωκείμενων τιμών (outliers) από ένα σύνολο δεδομένων. Υπενθυμίζεται ότι οι εξωκείμενες τιμές αποτελούν παρατηρήσεις, οι οποίες ξεπερνούν τα όρια των ακραίων τιμών (π.χ. min και max). Η βασική διαδικασία αυτής της τεχνικής περιλαμβάνει τέσσερα βήματα: 1) Ο διακομιστής λαμβάνει νέα δεδομένα πληκτρολόγησης για έλεγχο ταυτότητας, 2) ο αλγόριθμος One-Class SVM αναλαμβάνει την εκπαίδευση των αποθηκευμένων χαρακτηριστικών χρόνου των digraphs, και ορίζει το επίπεδο διαχωρισμού (δηλαδή, τα όρια), 3) ελέγχει μέσω δοκιμών αν τα νέα δεδομένα πληκτρολόγησης είναι εντός των ορίων. Στην περίπτωση που είναι εντός ορίων χαρακτηρίζονται ως inliers, ενώ σε διαφορετική περίπτωση ως outliers.4) Στο τέλος, γίνεται εξαγωγή της μετρικής Similarity Score, η οποία στη συγκεκριμένη περίπτωση περιλαμβάνει το ποσοστό των σημείων των digraphs του subject που θεωρήθηκαν ως inliers.

Ο αλγόριθμος One-Class SVM περιλαμβάνει την ακτινική συνάρτηση βάσης του πυρήνα (radial basis function kernel – RBF Kernel), η οποία αποτελεί μια δημοφιλή συνάρτηση πυρήνα που χρησιμοποιείται σε διάφορους αλγορίθμους μηχανικής μάθησης, και συνήθως για στους SVM [126]. Σημειώνεται ότι, οι αλγόριθμοι SVM, δηλαδή οι αλγόριθμοι Μηχανών Διανυσμάτων Υποστήριξης (Support Vector Machines), αποτελούν ένα τύπο μάθησης με επίβλεψη που χρησιμοποιείται για την ταξινόμηση, την παλινδρόμηση και την ανίχνευση ακραίων τιμών. Κατά την εκπαίδευση ενός αλγορίθμου SVM με RBF Kernel, πρέπει να ληφθεί σοβαρά υπόψη η παράμετρος gamma [127].

Στην συγκεκριμένη περίπτωση, η παράμετρος gamma ορίζεται μέσα από τον πίνακα ελέγχου (dashboard) που διαθέτει η επαφή της υπηρεσίας που αναπτύχθηκε στο [119]. Η παράμετρος gamma καθορίζει πόσο μακριά φτάνει η επίδραση ενός παραδείγματος εκπαίδευσης, με χαμηλές τιμές που σημαίνει «μακριά» και υψηλές τιμές που σημαίνει «κοντά». Σε ευρύτερο πλαίσιο, η συμπεριφορά του μοντέλου είναι πολύ ευαίσθητη στην gamma [127]. Ουσιαστικά, όταν η παράμετρος gammaέχει μεγάλη τιμή, τότε δημιουργούνται πολλά μικρά επίπεδα για τα δεδομένα εκπαίδευσης. Ωστόσο, όταν συμβαίνει αυτό υπάρχει ο κίνδυνος να γίνει πολύ αυστηρός ο έλεγχος, με αποτέλεσμα οι νέες τιμές των δεδομένων πληκτρολόγησης να είναι δύσκολο να καταχωρηθούν ως εσωκείμενες (inliers). Το αντίθετο συμβαίνει όταν togammaείναι μικρό. Ουσιαστικά, αν ο admin ορίσει μεγάλη τιμή για την παράμετρο gamma

αυτό σημαίνει ότι η εκπαίδευση των δεδομένων θα είναι πιο αυστηρή και κατ' επέκταση πιο αυστηρό έλεγχο ταυτότητας [119].

Στην παρακάτω εικόνα παρουσιάζεται ο αλγόριθμος εκπαίδευσης και ελέγχου ταυτότητας με One-Class SVM.

Algorithm 2: Train & Test with One-Class SVM

Input : The Array of new Keystroke Events to be tested *events_test*
Input : The saved digraph timings for this subject *digraphs_train*
Output: *Similarity Score*

```
digraphs_test ⟵ extract(events_test) ;      // extract timings for
 all digraphs in events_test
gamma ⟵ load_gamma_from_db();
inliers_count ⟵ 0;
foreach digraph in digraph_test do
    Fit OneClass SVM to the particular points of digraphs_train;
    Predict test points;
    inliers_count ⟵ inliers_count + [test points that are considered
     as inliers];
end
similarity_score ⟵ inliers_count / total_count ;
return similarity_score;
```

Εικόνα 41: Αλγόριθμος εκπαίδευσης και ελέγχου ταυτότητας με One-Class *SVM [119].*

Στην δεύτερη τεχνική που αναπτύχθηκε και δοκιμάστηκε, χρησιμοποιήθηκε η τεχνική εκπαίδευσης με τη χρήση του Gaussian Mixture Models και ελέγχου ταυτότητας με προσαρμοσμένο μοντέλο. Η κεντρική ιδέα σε αυτή την προσέγγιση είναι ότι τα δεδομένα πληκτρολόγησης του χρήστη βάσει στατιστικής θα είναι εντός του μέσου όρου τους κατά n τυπικές αποκλίσεις. Φυσικά, στην περίπτωση που ισχύει κανονική κατανομή. Ουσιαστικά, ο Gaussian Mixture Models χρησιμοποιείται με σκοπό να βρεθούν οι κανονικές κατανομές των χρονικών δεδομένων ενός digraph,καθώς και το πλήθος των συστατικών (components).

Για να γίνει αυτό θα πρέπει τα inliers να σταθμιστούν με ένα βάρος. Ως inliers θεωρούνται τα σημεία στο μέσο όρο μιας κατανομής ενός component στα όρια μιας περιοχής ανοχής, η οποία ορίζεται ως δ, όπου το δ αποτελεί ένα θετικό δεκαδικό αριθμό η μέτρηση του οποίου γίνεται σε τυπικές αποκλίσεις, και εκφράζει την απόσταση των τυπικών αποκλίσεων από μια κατανομή που απαιτείται για να θεωρηθεί ένα σημείο inlier.

Για την εύρεση της απόσταση ενός τυχαίου σημείου από μια τυχαία κατανομή χρησιμοποιήθηκε η απόσταση Mahalanobis:

$$MHL_M(\overline{x}) = \sqrt{(\overline{x} - \overline{\mu})^T S^{-1} (\overline{x} - \overline{\mu})}$$

Όπου, το x συμβολίζει το διάνυσμα των σημείων για ένα digraph, το μ συμβολίζει το μέσο όρο της κατανομής και το S τον πίνακα συμμεταβλητότητας του συστατικά (component).

Με πιο απλά λόγια, το διάνυσμα x αποτελεί το inlier που έχει σταθμιστεί με ένα βάρος, σε μια κατανομή M κάποιου συστατικού (component), στην περίπτωση που ισχύει ότι:

$$MHL_M(\overline{x}) \leq \delta$$

Στην περίπτωση που ισχύει η παραπάνω σχέση, αν ένα σημείο οριστεί ως inlier σε μια μη στατιστικά σημαντική κατανομή, τότε, δεν θα αποδοθεί από το σημείο σημαντική αύξηση του Similarity Score. Επομένως, στην τεχνική Gaussian Mixture το Similarity Score δίνει ως αποτέλεσμα ένα συγκεκριμένο δείκτη ομοιότητας, ο οποίος προκύπτει από το πλήθος των σημείων, τα οποία είναι κοντά στις περιοχές ανοχής των κατανομών, και το οποίο έχει πολλαπλασιαστεί με το βάρος της κάθε κατανομής. Αντίθετα, στην προηγούμενη περίπτωση, δηλαδή στην τεχνική με τη χρήση του one-class SVM, το Similarity Score προκύπτει από το ποσοστό των inliers.

Αξίζει να σημειωθεί ότι, και σε αυτή την περίπτωση, ο χρήστης (admin) μπορεί να ορίσει το πλήθος των συστατικών (components) εκπαίδευσης και την παράμετρο δ για τον ορισμό της περιοχής ανοχής. Όπως γίνεται αντιληπτό, όσο πιο μεγάλη τιμή δοθεί στην παράμετρο δ, το μεγαλύτερη θα είναι η ανοχή του συστήματος ως προς τον προσδιορισμό των inliers.

Στην ακόλουθη εικόνα δίνεται ο αλγόριθμος Gaussian Mixture Model.

Algorithm 3: Train & Test with Gaussian Mixture Models

Input : The Array of new Keystroke Events to be tested *events_test*
Input : The saved digraph timings for this subject *digraphs_train*
Output: *Similarity Score*

```
1  digraphs_test ⟵ extract(events_test) ;      // extract timings for
     all digraphs in events_test
2  δ ⟵ load_delta_from_db();
3  n_components ⟵ load_n_components_from_db();
4  score ⟵ [[0...0]] ;      // NxM matrix where N the count of unique
     digraphs and M the number of GMM components
5  foreach digraph in digraph_test do
6      Fit GMM Model with n_components;
7      score_comp ⟵ [0...0] ;    // 1xM where M number of components
8      for m:1 to n_components do
9          Load weight, means and covariance matrix of this component as
             w, μ, S;
10         pass ⟵ 0;
11         foreach test_point in test_points do
12             if MLH(test_point, μ, S) ≤ δ then
13                 pass ⟵ pass + 1 ;
14         end
15         score_comp[m] ⟵ w * pass ;          // Scale with weight of
             component
16     end
17     Append score_comp to score vertically;
18 end
19 similarity_score ⟵ sum(score) / total_count
20 return similarity_score;
```

Εικόνα 42: Αλγόριθμος εκπαίδευσης και πρόβλεψης Gaussian Mixture Model [119]

6.2.1.5 5ο βήμα: συλλογή δεδομένων

Για να εξεταστεί και να αξιολογηθεί το σύστημα συνεχούς ελέγχου ταυτότητας που αναπτύχθηκε στο [119] συλλέχθηκαν δεδομένα πληκτρολόγησης από ένα ικανοποιητικό πλήθος ατόμων. Η συλλογή των δεδομένων πραγματοποιήθηκε με τον τρόπο που εξηγήθηκε στις προηγούμενες υποενότητες. Ως πηγές δεδομένων χρησιμοποιήθηκαν δύο ιστότοποι, ο SafeShop, οποίος κατασκευάστηκε από τον ερευνητή για της ανάγκες της συγκεκριμένη μελέτης περίπτωσης και ένα υπάρχον διαδικτυακός τόπος των μεταπτυχιακών φοιτητών του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης (ΑΠΘ), ο RAT.

6.2.2 *Συλλογή δεδομένων και σύνολά δεδομένων*

Τα δεδομένα συλλέχθηκαν από δύο πηγές, τους ιστότοπους Safeshop και RAT. Από το συνδυασμό αυτών των δύο συνόλων δεδομένων (datasets) έγινε συλλογή περίπου 59.000 δεδομένων πληκτρολόγησης από 37 χρήστες συνολικά. Η διαδικασία συλλογής των δεδομένων διήρκησε συνολικά 3 μήνες. Σημειώνεται ότι τα δεδομένα από τον πειραματικό ιστότοπο SafeShop ήταν πολύ περισσότερα από τον ιστότοπο RAT, καθώς στον δεύτερο οι χρήστες κλήθηκαν να πληκτρολογήσουν μόνο ορισμένες λέξεις κλειδιά. Στον πίνακα της παρακάτω εικόνας δίνονται συνοπτικά τα δεδομένα πληκτρολόγησης από τους δυο ιστότοπους.

Dataset	Subjects	Total Keystroke Events	Keystroke Events / Subject
Safeshop	9	37000	4101.1
RAT	28	22000	766.6
Total	37	59000	1594.6

Εικόνα 43: Συνοπτικά σύνολο δεδομένων πληκτρολόγησης [119]

6.2.3 *Αποτελέσματα*

Για την σύγκριση των τελικών αποτελεσμάτων εφαρμόστηκαν και οι δυο αλγόριθμοι που παρουσιάστηκαν νωρίτερα. Τα αποτελέσματα σχετικά με την αποτελεσματικότητα των δύο μεθόδων παρουσιάζονται στον πίνακα της παρακάτω εικόνας και κάθε σύνολο δεδομένων και την συγχώνευση των δυο συνόλων δεδομένων.

		3 Διαστάσεις		2 Διαστάσεις	
Dataset	**Αλγόριθμος**	**FAR %**	**FRR %**	**FAR %**	**FRR %**
Safeshop	One-C SVM	0.61	0.75	2.11	2.01
	GMM	0.62	0.80	1.35	1.71
RAT	One-C SVM	6.57	7.01	8.11	9.97
	GMM	7.03	7.21	8.01	9.16
S+R	One-C SVM	4.01	4.23	5.86	5.72
	GMM	4.61	4.70	5.11	5.71

Εικόνα 44: Σύγκριση τελικών αποτελεσμάτωνOne-class SVM και GSM

Σε αυτό το σημείο επισημαίνεται ότι, το FAR, σημαίνει False Accept Rate, δηλαδή η πιθανότητα της λανθασμένης αντιστοίχισης της εισόδου ενός μοτίβου με ένα από τα πρότυπα της βάσης δεδομένων από το σύστημα. Με πιο απλά λόγια, σε ένα σύστημα που γίνεται χρήση δεικτών ομοιότητας (similarity scores), όπως στη συγκεκριμένη περίπτωση, αν ένας χρήστης δεν είναι αυθεντικός, αλλά το similarity score είναι πιο υψηλό από το ένα συγκεκριμένο κατώτερο όριο, τότε ο χρήστης αναγνωρίζεται λανθασμένα από το σύστημα ως αυθεντικός. Αντίθετα, το FRR, σημαίνει False Reject Rate, δηλαδή η πιθανότητα αποτυχίας της λανθασμένης αντιστοίχισης της εισόδου ενός μοτίβου με ένα από τα πρότυπα της βάσης δεδομένων από το σύστημα. Με πιο απλά λόγια, το FRR μετράει το ποσοστό των έγκυρων εισόδων, τις οποίες δεν αποδέχεται το σύστημα.

6.2.4 Συμπεράσματα

Η χρήση του αλγορίθμου One-Class SVM και του GMM είναι αποδοτική για τον εντοπισμό ανωμαλιών σε νέα δεδομένα πληκτρολόγησης. Συγκριτικά, τα αποτελέσματα απόδοσης είναι σχετικά όμοια για τους δύο αλγορίθμους, καλύτερα αποτελέσματα δίνει ο One-Class SVM στην περίπτωση των 3 διαστάσεων και ο GMM στην περίπτωση των 2 διαστάσεων. Βέβαια, με ελάχιστες διαφορές. Ωστόσο, ο αλγόριθμος GMM απαιτεί για κάθε πλήθος συστατικών (components) απαιτεί διαφορετικές ρυθμίσεις παραμέτρων. Ωστόσο, σε κάθε περίπτωση για να αποδώσουν καλύτερα οι αλγόριθμοι θα πρέπει να γίνει συλλογή μεγάλης ποσότητας δεδομένων. Αυτό το συμπέρασμα προκύπτει από το γεγονός ότι στην περίπτωση των συνόλων δεδομένων από το RAT η αποτελεσματικότητα των δεδομένων ήταν χαμηλότερη σε σχέση με τα δεδομένα από το SafeShop.

6.3 Έλεγχος ταυτότητας με δεδομένα ηλεκτροκαρδιογραφήματος

Στο [117] έγινε εφαρμογή πειραματικών διαδικασιών ελέγχου ταυτότητας με την χρήση της μηχανικής μάθησης και τη χρήση δεδομένων από βιομετρικά χαρακτηριστικά, και συγκεκριμένα από δεδομένα ηλεκτροκαρδιογραφήματος.

6.3.1 Μεθοδολογία

Για την υλοποίηση των πειραματικών διαδικασιών εφαρμόστηκαν τρεις διαφορετικές προσεγγίσεις για τον έλεγχο ταυτότητας των χρηστών σύμφωνα με τα δεδομένα

ηλεκτροκαρδιογραφήματος. Κάθε διαφορετική προσέγγιση περιλαμβάνει τρία διαφορετικά στάδια στη διαδικασία της, την προεπεξεργασία (preprocess), την εξαγωγή χαρακτηριστικών (feature extraction) και την ταξινόμηση (classification). Στο στάδιο της προεπεξεργασίας γίνεται η επεξεργασία του σήματος που λαμβάνεται από τον χρήστη, ώστε να μπορεί να γίνει η επεξεργασία του από τα επόμενα βήματα. Στο στάδιο της εξαγωγής χαρακτηριστικών εφαρμόζεται ο μετασχηματισμός μέσω του οποίου εξάγονται τα απαραίτητα χαρακτηριστικά από το σήμα. Τέλος, στο στάδιο της ταξινόμησης, στα εξαγόμενα χαρακτηριστικά του προηγούμενου σταδίου γίνεται εφαρμογή κάποιου αλγορίθμου μηχανικής μάθησης ή εισαγωγή των δεδομένων σε κάποιο νευρωνικό δίκτυο. Στη συνέχεια παρουσιάζονται αυτές οι προσεγγίσεις. Στη συγκεκριμένη εργασία παρουσιάζονται οι δύο πρώτες προσεγγίσεις, καθώς έχουν μεγαλύτερο ενδιαφέρον σε σχέση με το ενδιαφέρον της παρούσας εργασίας.

6.3.1.1 Μηχανική μάθηση και εξαγωγή χαρακτηριστικών χωρίς σταθερή βάση σύγκρισης

Σε αυτή τη μέθοδο η προεπεξεργασία και η εξαγωγή των χαρακτηριστικών των σημάτων γίνεται στο πεδίο των συχνοτήτων, η ταξινόμηση των χαρακτηριστικών γίνεται μέσω διάφορων αλγορίθμων μηχανικής μάθησης.

Προεπεξεργασία

Κατά το στάδιο της προεπεξεργασίας ορίζεται αρχικά ένα σύνολο καρδιογραφημάτων από 50 άτομα.

$$P_i, για\ i - 1,2,\dots,50$$

Για κάθε άτομο έχουν ληφθεί 4 διαφορετικές δειγματοληψίες.

$$S_{i,j}, για\ i = 1,2,\dots,50\ και\ j = 1,2,\dots,4$$

Όπου, το iαντιστοιχεί στο άτομο της δειγματοληψίας j.

Στη συνέχεια, το σήμα κάθε δειγματοληψίας φιλτράρεται στη ζώνη συχνοτήτων 0.5-40Hz, ώστε να γίνει απόρριψη των πηγών θορύβου (π.χ. Baseline wander) και θόρυβος που μπορεί να υπάρχει από τις γραμμές του ηλεκτρικού ρεύματος που βρίσκεται συνήθως στα 50Hz ή 60Hz (ανάλογα με την περιοχή), μέσω ενός ζωνοπερατού φίλτρου Butterworth, το οποίο αποτελεί ένα φίλτρο επεξεργασίας σήματος. (Για το εν λόγω φίλτρο δεν γίνεται περεταίρω ανάλυση, καθώς δεν αποτελεί αντικείμενο της παρούσας εργασίας διπλωματικής εργασίας).

$$F_{i,j} = B\{S_{i,j}\}, για\ i = 1,2,\dots,50\ \ και\ \ j = 1,2,\dots,4$$

Έπειτα, σε κάθε δειγματοληπτικό σήμα $S_{i,j}$γίνεται διαχωρισμός σε επιμέρους τμήματα. Κάθε ένα από αυτά τα τμήματα έχει διάρκεια 5 δευτερόλεπτα. Έτσι, κάθε τμήμα θα έχει αρκετούς παλμούς.

Εξαγωγή χαρακτηριστικών

Στο στάδιο της εξαγωγής χαρακτηριστικών εφαρμόζονται τρεις διαφορετικοί μετασχηματισμοί σε κάθε επιμέρους τμήμα που παράχθηκε κατά το στάδιο της προεξεργασίας. Στη συνέχεια, γίνεται η εξαγωγή χαρακτηριστικών, τα οποία θα χρησιμοποιηθούν στο επόμενο στάδιο της ταξινόμησης.

Ο μετασχηματισμός των επιμέρους τμημάτων γίνεται μέσω του διακριτού μετασχηματισμού Fourier, του διακριτού μετασχηματισμού συνημίτονου (Cosine), και του μετασχηματισμού Wavelet αντίστοιχα:

$$y_{1,j} = F\{F_{i,j}\}$$
$$y_{2,j} = C\{F_{i,j}\}$$
$$y_{3,j} = W\{F_{i,j}\}$$

Περισσότερα για τους παραπάνω μετασχηματισμούς δίνονται στο [117], οι οποίοι δεν αναλύονται περισσότερο, καθώς δεν αποτελούν αντικείμενο της παρούσας εργασίας.

Με την ολοκλήρωση του μετασχηματισμού λαμβάνονται οι συντελεστές που θα χρησιμοποιηθούν για τον έλεγχο ταυτότητας. Από κάθε μετασχηματισμό γίνεται λήψη των 20 μεγαλύτερων συντελεστών.

$$coef_{1,i} = sort\{y_1[1:20]\}$$
$$coef_{2,i} = sort\{y_2[1:20]\}$$
$$coef_{3,i} = sort\{y_{3,level3}[1:20]\}$$

Ταξινόμηση

Στο στάδιο της ταξινόμησης γίνεται η δημιουργία των αρχείων εκπαίδευσης βάσει των παραπάνω συντελεστών. Για κάθε χρήστηP_i, δημιουργούνται δυο σύνολα δεδομένων, το $train_i$ και το $test_i$. Το 60% των συντελεστών των ατόμων P_i εισάγονται στο αρχείο $train_i$ και λαμβάνουν την ετικέτα 0, παρόμοια το 40% των συντελεστών των ατόμων P_i εισάγονται στο αρχείο $test_i$. Επίσης, στο αρχείο $train_i$ αποθηκεύεται το 60%των συντελεστών των ατόμων P_k, όπου $k \neq i$, ώστε να υπάρχουν ισάριθμοι συντελεστές του ατόμου i και του ατόμου k και λαμβάνουν επισήμανση με ετικέτα 1. Παρόμοια, στο $test_i$ γίνεται εισαγωγή του 40% των

συντελεστών των ατόμων P_k, όπου k≠i, ώστε να υπάρχουν ισάριθμοι συντελεστές του ατόμου i και του ατόμουk.

Η εκπαίδευση των παραπάνω δεδομένων στα αντίστοιχα αρχεία γίνεται με την εφαρμογή των αλγορίθμων K-Means, MLP, RBF Network, Random Forest στο περιβάλλον Weka.

6.3.1.2 Βαθιά μάθηση και εξαγωγή χαρακτηριστικών χωρίς σταθερή βάση σύγκρισης

Σε αυτή τη μέθοδο τα στάδιο της προεπεξεργασίας και το στάδιο της εξαγωγής χαρακτηριστικών ακολουθεί ακριβώς την ίδια διαδικασία με την προηγούμενη μέθοδο. Η διαφορά εντοπίζεται στο στάδιο της ταξινόμησης.

Ταξινόμηση

Σε αυτή τη μέθοδο εφαρμόζεται η τεχνική της βαθιάς μάθησης και ειδικότερα το νευρωνικό δίκτυο. Πιο συγκεκριμένα, για την εκπαίδευση του νευρωνικού δικτύου απαιτείται η δημιουργία και η συνένωση των τριών αρχείων που δημιουργήθηκαν στο προηγούμενο στάδιο της εξαγωγής των χαρακτηριστικών. Ουσιαστικά, σε κάθε σειρά του αρχείου θα περιέχονται οι συντελεστές (coef) και από τους τρεις μετασχηματισμούς που πραγματοποιήθηκαν σε κάθε τμήμα του παλμικού σήματος. Σκοπός στη δημιουργία αυτού του αρχείου είναι για την εκμετάλλευση του μεγάλου βαθμού ελευθερίας που χαρακτηρίζει ένα νευρωνικό δίκτυο.

Το δίκτυο κατασκευάστηκε με 60 νευρώνες εισόδου με τρία κρυφά επίπεδα με 100, 200 και 100 νευρώνες στο κάθε ένα. Επίσης, κατασκευάστηκαν 2 νευρώνες εξόδου, οι οποίοι δίνουν 0 αν ο χρήστης είναι αυθεντικός (δηλαδή πραγματοποιήθηκε με επιτυχία ο έλεγχος ταυτότητας) και δίνουν 1 αν ο χρήστη δεν είναι αυθεντικός (αποτυχημένος έλεγχος ταυτότητας).

6.3.2 Συλλογή δεδομένων - προεπεξεργασία

Στο [117]έγινε λήψη της Diagnostic ECG Database από την Physionet (www.physionet.org). Αυτή η βάση δεδομένων χρησιμοποιείται κυρίως για να ανιχνεύονται καρδιακές παθήσεις και είναι η μεγαλύτερη βάση δεδομένων για ηλεκτροκαρδιογραφήματα που διατίθεται δωρεάν. Αυτή η βάση δεδομένων περιέχει 549 δειγματοληψίες από 290 άτομα, κάθε άτομα αντιπροσωπεύεται από 1 έως 5 δειγματοληψίες. Για το λόγο αυτό και παράλληλα για να υπάρξει ομοιομορφία και πλήθος εγγραφών στα δεδομένα του συνόλου, επιλέχθηκαν μόνο τα άτομα για τα οποία είχαν ληφθεί 4 δειγματοληψίες. Στη συνέχεια, τα δειγματοληπτικά σήματα με τη βοήθεια του toolbox της Physiobank της Physionet και συγκεκριμένα με το wfdb2mat (www.physionet.org/physiotools/wag/wfdb2m-1.htm) πραγματοποιήθηκε

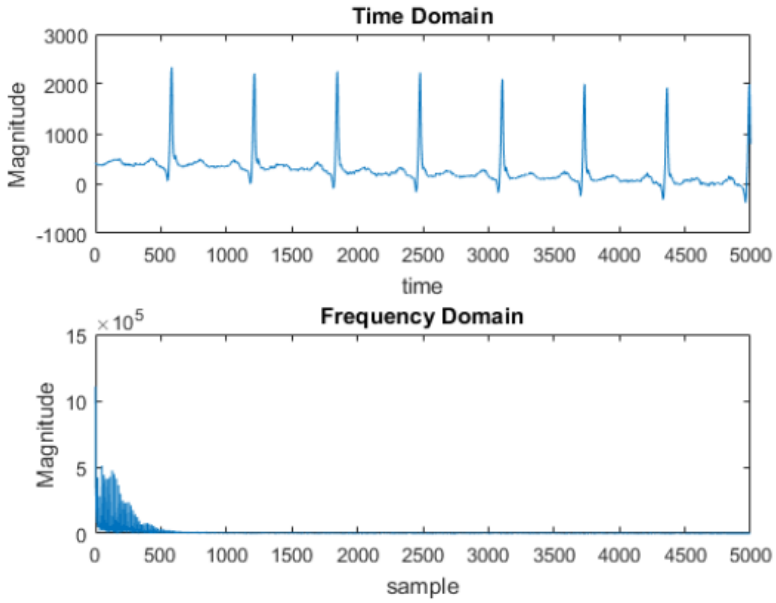
μετατροπή των σημάτων. Πιο συγκεκριμένα, η αρχική μορφή των αρχείων με τα ψηφιακά σήματα του δείγματος ήταν δυαδική (.dat) και μετατράπηκαν σε κατάλληλη μορφή, ώστε να μπορεί να γίνει η επεξεργασία τους από το λογισμικό Matlab (.mat). Συνοψίζοντας, το σύνολο δεδομένων περιλαμβάνει πληροφορίες για 50 άτομα, με 4 δειγματοληψίες για το κάθε άτομο και 115 δευτερόλεπτα για την κάθε μια.

Σε αυτά τα σήματα γίνεται η επεξεργασία με την εφαρμογή του φίλτρου Butterworth και διαχωρίζονται επιμέρους τμήματα των 5 δευτερολέπτων, έτσι ώστε να είναι έτοιμα για την εφαρμογή των μετασχηματισμών.

6.3.3 Εξαγωγή χαρακτηριστικών - Μετασχηματισμοί

6.3.3.1 Διακριτός μετασχηματισμός Fourier

Σε κάθε ένα από τα επιμέρους τμήματα των σημάτων εφαρμόζεται ο διακριτός μετασχηματισμός Fourier από το οποίο λαμβάνεται το καρδιακό σήμα στο πεδίο του χρόνου και στο πεδίο της συχνότητας, όπως φαίνεται στα διαγράμματα της παρακάτω εικόνας.



Εικόνα 45: Καρδιακός παλμός στο πεδίο του χρόνου και στο πεδίο των συχνοτήτων [117]

Όπως παρατηρείται στα παραπάνω διαγράμματα, στο πεδίο των συχνοτήτων, όλη η πληροφορίας είναι συγκεντρωμένοι στις χαμηλές συχνότητες. Στη συνέχεια, μετά το μετασχηματισμό, γίνεται εισαγωγή των 20 μεγαλύτερων συντελεστών κάθε τμήματος σε

αρχεία τύπου Weka, για την εκπαίδευση των αλγορίθμων μηχανικής μάθησης στο επόμενο στάδιο.

6.3.3.2 Διακριτός μετασχηματισμός συνημίτονου (cosine)

Ακολουθείται ακριβώς η ίδια διαδικασία με το μετασχηματισμό Fourier.

6.3.3.3 Αποσύνθεση – μετασχηματισμός Wavelet

Ακολουθείται ακριβώς η ίδια διαδικασία με το μετασχηματισμό Fourier.

6.3.4 Ταξινόμηση

6.3.4.1 Μηχανική μάθηση

Από τους παραπάνω μετασχηματισμούς δημιουργήθηκαν ορισμένα αρχεία τύπου Weka. Η ταξινόμηση των δεδομένων σε αυτή τη μέθοδο γίνεται σε δύο κλάσεις, δηλαδή ένα άτομα κατά τον έλεγχο ταυτότητας θα επαληθεύεται ή θα απορρίπτεται. Για να χαρακτηρίζεται από συνέπεια η διαδικασία εκπαίδευσης τα δεδομένα των δυο κλάσεων θα πρέπει να είναι ίσα προς τα μεγέθη. Η εκπαίδευση στο Weka γίνεται με τους αλγορίθμους Multi-Layer-Perceptron, K-means, RBF Network και Random Forest.

Τα αποτελέσματα του κάθε μετασχηματισμού και του κάθε αλγορίθμου δίνονται στον ακόλουθο πίνακα της εικόνας.

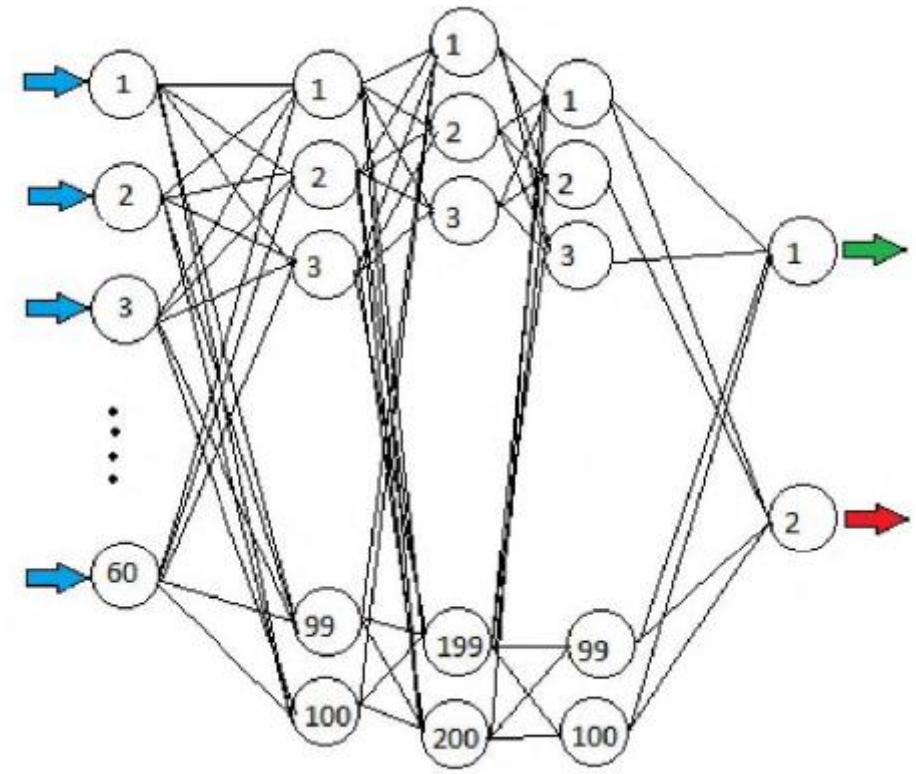
Transformation	Classification	Correct %	TP rate	FP rate	ROC Area
Cosine	KNN	81.616	0.832	0.199	0.817
Cosine	MLP	81.409	0.86	0.233	0.871
Cosine	RBFN	80.233	0.912	0.301	0.855
Cosine	RandForest	83.993	0.881	0.199	0.903
Fourier	KNN	82.53	0. 867	0.214	0.827
Fourier	MLP	82.601	0.87	0.215	0.886
Fourier	RBFN	81.076	0.911	0.284	0.871
Fourier	RandForest	85.204	0.89	0.183	0.914
Wavelet	KNN	86.974	0. 908	0.166	0.871
Wavelet	MLP	85.753	0. 887	0.17	0.918
Wavelet	RBFN	85.873	0. 919	0.198	0.909
Wavelet	RandForest	88.447	0. 917	0.146	0.952

Εικόνα 46: Συγκριτικά αποτελέσματα μετασχηματισμών και αλγορίθμων [117]

Παρατηρώντας τον παραπάνω πίνακα διαπιστώνεται ότι τα καλύτερα αποτελέσματα προέρχονται από το μετασχηματισμό Wavelet συνδυαστικά με τον αλγόριθμο Random Forest, όπου το ποσοστό επιτυχίας των αποτελεσμάτων είναι 88.447%

6.3.4.2 Βαθιά μάθηση

Σε αυτή τη μέθοδο δημιουργήθηκε ένα μοντέλο εκπαίδευσης με βαθιά μάθηση σε ένα νευρωνικό δίκτυο. Η υλοποίηση της μεθόδου έγινε με τη βοήθεια της βιβλιοθήκης Python Tensor flow της Google. Όπως αναφέρθηκε και νωρίτερα, τα αρχεία που δημιουργήθηκαν από τους προηγούμενους μετασχηματισμούς ενώθηκαν σε ένα ενιαίο αρχείο, καθώς η βαθιά μάθηση και τα νευρωνικά δίκτυα δίνουν τη δυνατότητα μεγάλου αριθμού εισόδων. Υπενθυμίζεται ότι, το δίκτυο αποτελείται από 60 νευρώνες εισόδου για τους συντελεστές των τριών μετασχηματισμών, 3 κρυφά επίπεδα με 100, 200 και 100 κρυφούς νευρώνες, καθώς και 2 νευρώνες εξόδου για τις δύο κλάσεις. Στο ακόλουθο σχήμα απεικονίζεται η δομή του νευρωνικού δικτύου.



Εικόνα 47: Δομή νευρωνικού δικτύου βαθιάς μάθησης [117]

Η ακρίβεια του νευρωνικού δικτύου λαμβάνοντας ως δεδομένα εισόδου τα ίδια δεδομένα και trainκαι testμε την προηγούμενη μέθοδο μηχανικής μάθησης είναι 0.8016%.

6.3.5 Συμπεράσματα

Τα καλύτερα αποτελέσματα σχετικά με την ακρίβεια του ελέγχου ταυτότητας προέκυψε από τη μέθοδο κατά την οποία χρησιμοποιήθηκε ο μετασχηματισμός Wavelet και ο αλγόριθμος μηχανικής μάθησης Random Forest, όπου η ακρίβεια της μεθόδου αυτής ήταν 88.447%.

Σύμφωνα με του συγγραφέα στο [117] τα αποτελέσματα αυτά ήταν πιο χαμηλά συγκριτικά με την αντίστοιχη βιβλιογραφία με παρόμοιες μεθόδους. Ωστόσο, σε αυτές τις περιπτώσεις είχαν χρησιμοποιηθεί μέθοδοι με πολύ πιο μικρά σύνολα δεδομένα (π.χ. για 10 άτομα). Η αρχική υπόθεση στο [117] ήταν ότι η μέθοδος του νευρωνικού δικτύου με βαθιά μάθηση ως προς την ακρίβεια του ελέγχου ταυτότητας. Ωστόσο, η ακρίβεια των αποτελεσμάτων ήταν 0,8016%, ποσοστό πολύ πιο μικρό σε σχέση με τη μέθοδο της μηχανικής μάθησης. Αυτό οφείλεται στο γεγονός ότι ο αριθμός των δεδομένων, καθώς η βαθιά μάθηση απαιτεί μεγάλο αριθμό δεδομένων για να μπορέσει να εκπαιδευτεί σωστά.

7 *Σύγκριση λύσεων ασφαλείας*

Στα προηγούμενα κεφάλαια μελετήθηκαν οι προκλήσεις ασφάλειας που εντοπίζονται στη δίκτυα 5G για τις βασικές νέες τεχνολογίες, καθώς και για τον έλεγχο ταυτότητας. Επίσης, αναζητήθηκαν λύσεις για αυτά τα ζητήματα ασφαλείας, οι οποίες διακρίνονται σε τυπικές και σε αυτές που βασίζονται στην μηχανική μάθηση. Σε αυτό το κεφάλαιο γίνεται σύγκριση μεταξύ αυτών των μεθόδων, ώστε να εξαχθούν συμπεράσματα σχετικά με τα πλεονεκτήματα και τα μειονεκτήματα τους.

7.1 Σύγκριση τυπικών μεθόδων και μεθόδων μηχανικής μάθησης για τις νέες τεχνολογίες

Σε αυτή την ενότητα γίνεται σύγκριση των λύσεων ασφαλείας για τις βασικές νέες τεχνολογίες των δικτύων 5G, δηλαδή SDN, NFV και massive MIMO. Σε ευρύτερο πλαίσιο, ειδικά για τις τεχνολογίες SDN και massive MIMO, οι τυπικές λύσεις δεν μπορούν να δώσουν επαρκείς λύσεις για τα συστήματα αυτών των τεχνολογιών, ενώ η μηχανική μάθηση έχει πολλά πλεονεκτήματα και προσφέρει νέες δυνατότητας. Στην περίπτωση των συστημάτων της τεχνολογίας NFV, η μηχανική μάθηση δεν είναι κάτι που προτείνεται ευρέως για την ασφάλεια των συστημάτων NFV σε πραγματικό χρόνο, ωστόσο μπορεί να συμβάλει σε διάφορους τομείς για τα συστήματα αυτά. Στη συνέχεια, περιγράφονται αναλυτικότερα η σύγκριση των τυπικών μεθόδων και των μεθόδων μηχανικής μάθησης για την κάθε τεχνολογία ξεχωριστά.

Από τη μελέτη της βιβλιογραφίας, οι τυπικές λύσεις που προτείνοντας για τα συστήματα SDN, επιβαρύνουν το επίπεδο ελέγχου, καθώς σε κάθε λύση που προτείνεται (π.χ. συστήματα αδειών ή ελέγχου ταυτότητας για εφαρμογές), η άδεια πρόσβασης δίνεται από το επίπεδο ελέγχου. Βέβαια, προτείνονται κάποιες λύσεις για την αντιμετώπιση αυτού του ζητήματος, δηλαδή την επιβάρυνση του επιπέδου ελέγχου, όπως για παράδειγμα η εφαρμογή AVANT-GUARD, η οποία περιορίζει τα αιτήματα ροής, ξεχωρίζοντας τις αποτυχημένες περιόδους σύνδεσης TCP στο επίπεδο δεδομένων, πριν από οποιαδήποτε ειδοποίηση του επιπέδου ελέγχου. Ωστόσο, αυτή η μέθοδος παραμένει στατική και είναι σχεδιασμένη μόνο για τον έλεγχο ενός συγκεκριμένου χαρακτηριστικού, όπως και οι υπόλοιπες τυπικές τεχνικές.

Αντίθετα, οι τεχνικές μηχανικής μάθησης που προτείνονται στη βιβλιογραφία για την ασφάλεια στην τεχνολογία SDN δίνουν τη δυνατότητα ακριβούς ταξινόμησης της κυκλοφορίας του δικτύου, όπως επιθέσεις DoS και κακόβουλη κυκλοφορία. Επίσης, οι μέθοδοι μηχανικής μάθησης που προτείνονται, όπως οι αυτόοργανωμένοι χάρτες, περιορίζουν την επιβάρυνση του επιπέδου ελέγχου σε σχέση με τις τυπικές τεχνικές.

Οι περισσότερες τυπικές τεχνικές για την ασφάλεια των συστημάτων NFV εκμεταλλεύονται τις εικονοποιημένες λειτουργίες που προσφέρει αυτή η τεχνολογία, διασπώντας το σύστημα σε μικρότερα, ώστε να εντοπίζονται ευκολότερα και να παρακολουθούνται οι ευπάθειες από το σύστημα. Οι τεχνικές μηχανικής μάθησης δεν προτείνονται ευρέως στο πλαίσιο της τεχνολογίας NFV και σε σενάρια κυκλοφορίας δικτύου, καθώς η έλλειψη επισημασμένων δεδομένων καθιστά αδύνατη την χρήση αλγορίθμων μηχανικής μάθησης με επίβλεψη, όπου οι επισημασμένες ροές κυκλοφορίας είναι απαραίτητες για τις διαδικασίες εκπαίδευσης και επικύρωσης. Επιπλέον, οι μη επιτηρούμενοι αλγόριθμοι χρειάζονται επισημασμένα δεδομένα. Ωστόσο, στο πλαίσιο της ανάλυσης των πληροφοριών, οι τεχνικές μηχανικής μάθησης μπορούν να εφαρμοστούν για να μάθουν με την πάροδο του χρόνου ποια είναι η καλύτερη δράση που πρέπει να εκτελεστεί με βάση το ιστορικό ανιχνευμένων ανωμαλιών.

Κατά την μελέτη των λύσεων για τις πιλοτικές επιθέσεις στα συστήματα massive MIMO προτείνονται τεχνικές τυπικές και μηχανικής μάθησης. Στην πρώτη περίπτωση, η ανίχνευση γίνεται κατά κύριο λόγο στο σταθμό βάσης με ελεγχόμενη τυχαιότητα (controlled randomness)με τη μετάδοση πιλότων. Αυτή η μέθοδος είναι καλύτερη από τα κλασικά συστήματα κρυπτογράφησης που χρησιμοποιούν ένα κλειδί, ενώ αυτή χρησιμοποιεί πολλά κλειδιά ταυτόχρονα. Ωστόσο απαιτούνται περισσότεροι πόροι για την μετάδοση των επιπλέον τυχαίων ακολουθιών. Από την άλλη, οι προτεινόμενες μέθοδοι μηχανικής μάθησης (βαθιά μάθηση με επίβλεψη, αλγόριθμος SVM, μέθοδος Bayesian) για την ανίχνευση των πιλοτικών

επιθέσεων πλαστοπροσωπίας, για τις οποίες έχει αποδειχθεί η αποτελεσματικότητα τους σε πειραματικό επίπεδο. Μια πρόκληση για τη χρήση αλγορίθμων μηχανικής μάθησης για τη διασφάλιση ενός μαζικού MIMO είναι η υψηλή επιβάρυνση λόγω του μεγάλου όγκου δεδομένων εκπαίδευσης. Στην περίπτωση της μάθησης με επίβλεψη, ένα μαζικό MIMO πρέπει να εκπαιδευτεί με απουσία υποκλοπών. Εάν η απουσία ενός εισβολέα δεν είναι σίγουρη, πρέπει να υιοθετηθεί μια προσέγγιση μάθησης χωρίς επίβλεψη. Ωστόσο, οι αλγόριθμοι της μάθησης χωρίς επίβλεψη είναι λιγότερο ακριβείς και αξιόπιστοι και πιο πολύπλοκοι από αυτούς της μάθησης με επίβλεψη

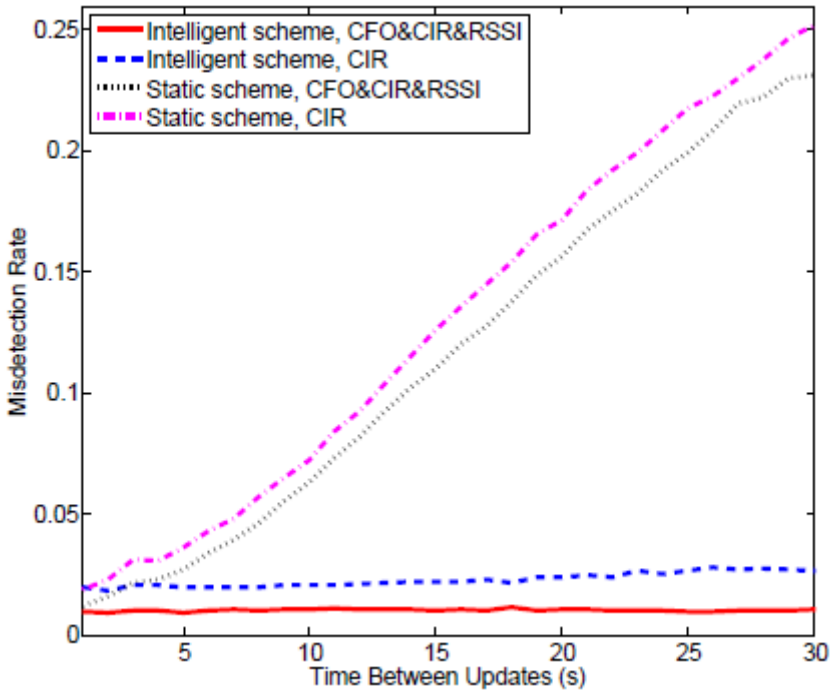
7.2 Σύγκριση παραδοσιακού και έξυπνου ελέγχου ταυτότητας

Σε αντίθεση με τις παραδοσιακές τεχνικές ελέγχου ταυτότητας, οι τεχνικές έξυπνου ελέγχου ταυτότητας με τη βοήθεια της μηχανικής μάθησης χρησιμοποιούν και τα ποικίλα χαρακτηριστικά, αλλά και τις προσαρμοστικές διαδικασίες ελέγχου ταυτότητας και εκπαιδεύονται από τα δεδομένα χωρίς να απαιτείται ένα ακριβές μοντέλο χαρακτηριστικών. Το πιο σημαντικό είναι ότι οι έξυπνες τεχνικές ελέγχου ταυτότητας με τη βοήθεια της μηχανικής μάθησης επιτυγχάνουν οικονομική, πιο αξιόπιστη, χωρίς μοντέλα, συνεχή και επικύρωση συσκευής με επίγνωση της κατάστασης στο περίπλοκο χρονικά διαφορετικό περιβάλλον (σε πραγματικό χρόνο) [64]. Το μειονέκτημα είναι ότι, συνήθως απαιτούν την κατάλληλη επιλογή αλγορίθμων μηχανικής μάθησης και ο σχεδιασμός των διαδικασιών ελέγχου ταυτότητας είναι περίπλοκος, ειδικά για τις συσκευές που είναι ευαίσθητες σε καθυστέρηση και τις συσκευές με περιορισμένους πότους [6]. Στον ακόλουθο πίνακα παρουσιάζεται συνοπτικά η σύγκριση μεταξύ των παραδοσιακών τεχνικών και των έξυπνων τεχνικών με μηχανική μάθηση για τον έλεγχο ταυτότητας στα δίκτυα 5G.

Πίνακας 2: Σύγκριση μεταξύ του παραδοσιακού ελέγχου ταυτότητας φυσικού επιπέδου και των έξυπνων τεχνικών ελέγχου ταυτότητας με μηχανική μάθηση σε δίκτυα 5G [6]

Τεχνικές ελέγχου ταυτότητας	**Τυπικά χαρακτηριστικά**	**Μηχανισμοί ασφαλείας**	**Πρόσθετες απαιτήσεις**	**Περιορισμοί εφαρμογής**
Παραδοσιακές τεχνικές	Βασίζονται σε μοντέλο, στατικές	Δοκιμές υπόθεσης	Στατιστικές ιδιότητες και ακριβή μοντέλα για διάφορα χαρακτηριστικά	Περιορίζονται σε πολύπλοκα δυναμικά περιβάλλοντα με ανεπαρκή στατιστικά στοιχεία χαρακτηριστικών
Έξυπνες τεχνικές με μηχανική μάθηση	Οικονομικά αποδοτικό, αξιόπιστο, χωρίς μοντέλα, συνεχές, με γνώμονα την κατάσταση	Αυτόματη προσαρμογή, μάθηση από δεδομένα χωρίς ένα ακριβές μοντέλο χαρακτηριστικών	Κατάλληλη επιλογή αλγορίθμου μηχανικής μάθησης και περίπλοκος σχεδιασμός διαδικασίας ελέγχου ταυτότητας	Περιορίζονται στη μακρά συνεχή διαδικασία ελέγχου ταυτότητας μεταξύ συσκευών χαμηλής ισχύος

Στο [6] για να αποδείξουν την προσέγγιση του έξυπνου ελέγχου ταυτότητας, σύγκριναν τα αποτελέσματα προσομοίωσης από το πρόγραμμα έξυπνου ελέγχου ταυτότητας με μάθηση kernel και του σχήματος στατικού ελέγχου ταυτότητας που βασίζεται και στα πολλαπλά χαρακτηριστικά (δηλαδή, CFO, CIR και RSSI) και σε ένα χαρακτηριστικό (δηλαδή, CIR).Οι εκτιμήσεις των χαρακτηριστικών δημιουργήθηκαν με βάση τα υπάρχοντα έργα [64][65] και οι εκτιμήσεις χαρακτηριστικών του Spoofer δεν απαιτούνται για την εκπαίδευση αυτού του έξυπνου σχήματος ελέγχου ταυτότητας, καθώς παρακολουθεί τα χαρακτηριστικά της Alice και επικυρώνει τις εκτιμήσεις που είναι διαφορετικές κατά την ερμηνεία από την Aliceκαι τον Spoofer.Οι εκτιμήσεις των χαρακτηριστικών του Spoofer για τον υπολογισμό του ποσοστού εσφαλμένης ανίχνευσης δημιουργήθηκαν τυχαία εντός των ορίων εκτίμησης. Στο ακόλουθο σχήμα δίνονται τα αποτελέσματα αυτών των μετρήσεων.



Εικόνα 48: Σύγκριση απόδοσης ελέγχου ταυτότητας μεταξύ του ευφυούς σχήματος ελέγχου ταυτότητας και του συστήματος στατικού ελέγχου ταυτότητας[6].

Στο παραπάνω σχήμα παρατηρείται ότι και με τον έξυπνο έλεγχο, αλλά και με τον στατικό, επιτυγχάνεται χαμηλότερος αριθμός εσφαλμένης ανίχνευσης όταν γίνεται διερεύνηση πολλαπλών χαρακτηριστικών. Αυτό οφείλεται στο γεγονός ότι η ακρίβεια ελέγχου ταυτότητας μπορεί να αυξηθεί σε νόμιμες συσκευές σε σύνθετα χρονικά μεταβαλλόμενα περιβάλλοντα μέσω της διερεύνησης πολλαπλών χαρακτηριστικών, καθώς είναι πιο δύσκολο για τον Spoofer να προβλέψει και να μιμηθεί πολλά χαρακτηριστικά της Alice. Παρατηρείται επίσης ότι, κατά την αύξηση του χρόνου μεταξύ των ενημερώσεων, ο ρυθμός εσφαλμένης ανίχνευσης του ευφυούς σχήματος ελέγχου ταυτότητας παραμένει σταθερός και ισχυρός, ενώ αυτός του σχήματος στατικού ελέγχου ταυτότητας αυξάνεται σημαντικά. Αυτό αποδεικνύει ότι χωρίς προσαρμογή που να είναι ενήμερη για την κατάσταση, η απόδοση του σχήματος στατικού ελέγχου ταυτότητας θα μειωθεί σημαντικά σε πολύπλοκα περιβάλλοντα που ποικίλλουν χρονικά, περιορίζοντας έτσι την εφαρμογή του σε ασύρματα δίκτυα 5G.

Προκειμένου να επιτευχθεί συνεχής έλεγχος ταυτότητας, το σχήμα στατικού ελέγχου ταυτότητας πρέπει να επαναλάβει τη διαδικασία ελέγχου ταυτότητας συλλέγοντας εκτιμήσεις χαρακτηριστικών, λαμβάνοντας τις στατιστικές ιδιότητες των χαρακτηριστικών, αντλώντας νέο όριο δοκιμής και, στη συνέχεια, να επιβεβαιώσει ξανά τις συσκευές. Με την επανάληψη

αυτής της διαδικασίας, το σχήμα στατικού ελέγχου ταυτότητας απαιτεί περισσότερους πόρους υπολογισμού και μεγαλύτερο χρονικό διάστημα [6].

8 *Επίλογος*

Σε αυτό το κεφάλαιο συνοψίζονται όσα μελετήθηκαν στο πλαίσιο της παρούσας διπλωματικής και παρατίθενται τα συμπεράσματα που προέκυψαν από την ολοκλήρωση της μελέτης. Επίσης, στο τέλος του κεφαλαίου προτείνονται ορισμένες ιδέες για μελλοντική έρευνα και επέκταση της διπλωματικής εργασίας.

8.1 Σύνοψη και συμπεράσματα

Στο πλαίσιο αυτής της εργασίας μελετήθηκαν τα δίκτυα 5G ως προς τα βασικά χαρακτηριστικά τους, τις βασικές τεχνολογίες που συμπεριλαμβάνονται σε αυτά και τις σχετικές αρχιτεκτονικές. Ως μια νέα τεχνολογία κινητής τηλεφωνίας, το 5G έχει πολύ διαφορετική δομή δικτύου, καθώς και διαφορετικές δυνατότητες και απαιτήσεις δικτύου, σε σχέση με τις προηγούμενες γενιές δικτύων κινητής τηλεφωνίας (3G, 4G, LTE) και ενσωματώνεται ένας μεγάλος αριθμός νέων τεχνολογιών. Βέβαια, εκτός από την εισαγωγή αυτών των νέων τεχνολογιών το 5G μπορεί να συνδέσει διάφορα χαρακτηριστικά και λειτουργίες της κοινωνίας στο διαδίκτυο, όπως οχήματα, οικιακές συσκευές, υγειονομική περίθαλψη, βιομηχανία, επιχειρήσεις και πολλά άλλα. Καθώς, τα δίκτυα 5G μπορούν να προσφέρουν προσβασιμότητα από παντού (σχεδόν) με υψηλές ταχύτητες και πολύ χαμηλό χρόνο καθυστέρησης. Ωστόσο, αυτές οι εξελίξεις εισάγουν νέες απειλές και τρωτά σημεία ασφαλείας που εντοπίζονται στα δίκτυα 5G.

Για το λόγο αυτό, κρίθηκε σημαντικό να μελετηθούν οι προκλήσεις ασφαλείας που εντοπίζονται στα δίκτυα 5G και να εντοπιστούν οι λύσεις ασφαλείας που μπορούν να συμβάλουν στην αντιμετώπιση αυτών των προκλήσεων, που αποτελεί και τη σημαντικότερη συνεισφορά της παρούσας εργασίας. Η προσέγγιση των προκλήσεων ασφαλείας

ταξινομήθηκε βάσει των νέων τεχνολογιών και σε επίπεδο ελέγχου ταυτότητας. Για την αντιμετώπιση αυτών των ζητημάτων εντοπίστηκαν και παρουσιάστηκαν λύσεις που βασίζονται σε τυπικές τεχνικές και λύσεις που βασίζονται σε τεχνικές της μηχανικής μάθησης.

Συμπερασματικά από την σύγκριση των τυπικών τεχνικών ασφαλείας με τις λύσεις ασφαλείας που βασίζονται στη μηχανική μάθηση, διαπιστώθηκε ότι, οι τυπικές τεχνικές που προτείνονται για τις νέες τεχνολογίες 5G βασίζονται σε συγκεκριμένα μοντέλα, είναι στατικές, οι μηχανισμοί ασφαλείας περιλαμβάνουν δοκιμές υποθέσεων ή ακολουθίες κλειδιών και απαιτούν ακριβή μοντέλα για τον έλεγχο συγκεκριμένων χαρακτηριστικών. Επίσης, επιβαρύνουν τα συστήματα ελέγχου και για την αποτελεσματικότητα τους απαιτούν περισσότερους υπολογιστικούς πόρους.

Σε αυτά τα ζητήματα, λύσεις μπορούν να δοθούν από τις δυνατότητες και τα πλεονεκτήματα της μηχανικής μάθησης. Τα μοντέλα μηχανικής μάθησης μπορούν να χρησιμοποιηθούν για τον εντοπισμό ύποπτων δραστηριοτήτων σε πραγματικό χρόνο, αναλύοντας μοτίβα και παραμέτρους από τη δραστηριότητα του δικτύου. Οι αλγόριθμοι ταξινόμησης μπορούν να χρησιμοποιηθούν για την ανίχνευση ανωμαλιών παρακολουθώντας τις παραμέτρους του δικτύου, όπως αρχεία καταγραφής απόδοσης και σφάλματος δικτύου. Οι αλγόριθμοι ομαδοποίησης μπορούν να χρησιμοποιηθούν για την κατηγοριοποίηση διαφόρων ειδών απειλών και κενών στην ασφάλεια του δικτύου. Αυτά αποτελούν μόνο μερικά παραδείγματα των εφαρμογών μηχανικής μάθησης που μπορούν να συμβάλουν στην ασφάλεια των δικτύων 5G. Σε σχέση με τις τυπικές μεθόδους, κατά κύριο λόγο, η μηχανική μάθηση μειώνει την επιβάρυνση των συστημάτων ελέγχου και απαιτεί λιγότερους υπολογιστικούς πόρους. Βέβαια, σε ορισμένες περιπτώσεις, η μηχανική μάθηση μπορεί επιβαρύνει το σύστημα κατά την εκπαίδευση εξαιτίας του μεγάλου όγκου δεδομένων. Ωστόσο, σε κάθε περίπτωση, οι μέθοδοι μηχανικής μάθησης μπορούν να προσφέρουν ασφάλεια σε πραγματικό χρόνο ανιχνεύοντας την κυκλοφορία των δεδομένων, δεν απαιτούν συγκεκριμένο μοντέλο, αλλά εκπαιδεύονται από τα ίδια τα δεδομένα και η αποτελεσματικότητα τους έγκειται στο υψηλό ποσοστό ανίχνευσης απειλών, όπως παρουσιάστηκε και στη μελέτη περίπτωσης που περιγράφηκε αναλυτικά.

Για τον έλεγχο ταυτότητας στα δίκτυα 5G μελετήθηκαν αρχικά οι προκλήσεις για τις τυπικές τεχνικές ελέγχου ταυτότητας και τα πλεονεκτήματα του ελέγχου ταυτότητας με μηχανική μάθηση. Επίσης, παρουσιάστηκαν τα παραδείγματα μηχανικής μάθησης για έλεγχο ταυτότητας, τα οποία ταξινομήθηκαν σε παραμετρικές και μη παραμετρικές μεθόδους

μάθησης, καθώς και τεχνικές μάθησης με και χωρίς επίβλεψη και με ενίσχυση. Συμπερασματικά, σε αντίθεση με τις παραδοσιακές τεχνικές ελέγχου ταυτότητας, οι τεχνικές ελέγχου ταυτότητας με τη βοήθεια της μηχανικής μάθησης χρησιμοποιούν τα ποικίλα χαρακτηριστικά, αλλά και τις προσαρμοστικές διαδικασίες ελέγχου ταυτότητας και εκπαιδεύονται από τα δεδομένα χωρίς να απαιτείται ένα ακριβές μοντέλο χαρακτηριστικών. Το πιο σημαντικό είναι ότι οι αυτές οι τεχνικές ελέγχου ταυτότητας με τη βοήθεια της μηχανικής μάθησης επιτυγχάνουν οικονομική και πιο αξιόπιστη ασφάλεια, με συνεχή έλεγχο και επικύρωση των συσκευών με επίγνωση της κατάστασης στο περίπλοκο χρονικά διαφορετικό περιβάλλον. Το μειονέκτημα είναι ότι, συνήθως απαιτούν την κατάλληλη επιλογή αλγορίθμων μηχανικής μάθησης και περίπλοκο σχεδιασμό διαδικασιών ελέγχου ταυτότητας, ειδικά για συσκευές ευαίσθητες σε καθυστέρηση και συσκευές με περιορισμένους πόρους.

8.2 Μελλοντικές επεκτάσεις

Σε αυτή τη διπλωματική εργασία μελετήθηκαν ορισμένες λύσεις ασφαλείας που βασίζονται στη μηχανική μάθηση βάσει προηγούμενων εμπειρικών ερευνών. Σε μια μελλοντική μελέτη θα μπορούσε να πραγματοποιηθεί ο σχεδιασμός και η υλοποίηση ενός σεναρίου για τον έλεγχο ταυτότητας με τεχνικές της μηχανικής μάθησης.

9 *Βιβλιογραφία*

[1] E. Ibarrola, M. Davis, C. Voisin, C. Close, L. Cristobo. A machine learning management model for QoE enhancement in next-generation wireless ecosystems. 10th ITU Academic Conference ITU Kaleidoscope: Machine learning for a 5g future, 26-28 Nov. 2018, Santa Fe,, Argentina.

[2] D. Forte, A. de Donno. Mobile Network Investigations. In E. Casey (Ed.) Handbook of Digital Forensics and Investigation, pp. 437-516, Academic Press, 2010.

[3] Thales Group. "Introducing 5G technology and networks (speed, use cases and rollout)", available at: https://www.thalesgroup.com/en/markets/digital-identity-and-security/mobile/inspired/5G, Feb. 2020.

[4] M.Z. Noohani, K.U. Magsi. A Review Of 5G Technology: Architecture, Security and wide Applications. International Research Journal of Engineering and Technology (IRJET), 7(5), May 2020.

[5] Ahmad, T. Kumar, L. Madhusanka, J. Okwuibe, M. Ylianttila, A. Gurtov. 5G Security: Analysis of Threats and Solutions. IEEE Conference on Standards for Communications and Networking (CSCN), pp. 193-199, Sep. 18 – 21, 2017, Helsinki, Finland.

[6] H. Fang, X. Wang, S. Tomasin. Machine Learning for Intelligent Authentication in 5G-and-Beyond Wireless Networks. IEEE Wireless Communications, 26(5), pp. 55-61, Oct. 2019.

[7] N. Haider, Z. Baig, M. Imran. Artificial Intelligence and Machine Learning in 5G Network Security: Opportunities, advantages, and future research trends. 2020. Available at: https://arxiv.org/pdf/2007.04490

[8] J. Suomalainen, A. Juhola, S. Shahabuddin, A. Mämmelä, I. Ahmad. Machine Learning Threatens 5G Security. IEEE Access, 6, pp. Oct. 2020.

[9] D. Basin, J. Dreier, L. Hirschi, S. Radomirović, R. Sasse, V. Stettler. A Formal Analysis of 5G Authentication. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications, pp. 1383–1396, Toronto Canada, October, 2018.

[10] J. Lam, R. Abbas. Machine Learning based Anomaly Detection for 5G Networks. 2020. arXiv:2003.03474

[11] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, R. Jain. Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things. IEEE Internet of Things Journal, 6(4), pp. 6822-6834, Aug. 2019.

[12] Smola, S.V.N. Vishwanathan. Introduction to Machine Learning. University Press, Cambridge, 2008.

[13] S. Divya, S. Jyothi. Machine Learning Algorithms in Big data Analytics. International Journal of Computer Sciences and Engineering, 6(1), pp. 63-70, Jan. 2018.

[14] H.D. Wehle. Machine Learning, Deep Learning, and AI: What's the Difference? DataScientistInnovationDay, 2017.

[15] Α. Γεωργούλη. Μηχανική Μάθηση. Στο Α. Γεωργούλη (Επιμ.) Τεχνητή νοημοσύνη, Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αθήνα, 2015.

[16] N.C. Luong, D.T. Hoang, S. Gong, D. Niyato, P. Wang, Y.C. Liang, D.I. Kim. Applications of deep reinforcement learning in communications and networking: A survey. IEEE Communications Surveys & Tutorials, 21(4), pp. 3133-3174, May 2018.

[17] Κ. Διαμαντάρας. Μηχανική Μάθηση – Μάθημα 1: Βασικές έννοιες. (Σημειώσεις μαθήματος). ΤΕΙ Θεσσαλονίκης - Τμήμα Πληροφορικής, 2011.

[18] F.Y. Osisanwo, J.E.T. Akinsola, O. Awodele, J.O. Hinmikaiye, O. Olakanmi, J. Akinjobi. Supervised Machine Learning Algorithms: Classification and Comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3), pp. 128-138, June 2017.

[19] M.A. Alsheikh, S. Lin, D. Niyato, H.P. Tan. Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Application. IEEE Communications Surveys and Tutorials, 16(4), pp. 1996-2018, 2014.

[20] S.J. Russel, P. Norvig. Artificial Intelligence: A Modern Approach, Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[21] S. Russel, P. Norvig and E. Davis. Artificial Intelligence: A modern Approach, 2nd ed. NewJersey: PrenticeHall, 2003, pp. 825

[22] Β. Σταματόπουλος. Μέθοδοι Μηχανικής Μάθησης για την Πρόβλεψη Αυτοματοποίησης Επαγγελμάτων. (Αδημοσίευτη Πτυχιακή Εργασία). Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών/Τμήμα Πληροφορικής και Τηλεπικοινωνιών. Αθήνα, 2018.

[23] Ι. Φενέρη. Η ενισχυτική μάθηση σε συστήματα πολλαπλών πρακτόρων. (Αδημοσίευτη διπλωματική εργασία). Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης/Τμήμα Πληροφορικής. Θεσσαλονίκη, 2006.

[24] C. Zhang, P. Patras, H. Haddadi. Deep Learning in Mobile and Wireless Networking: A Survey. IEEE Communications Surveys & Tutorials, 21(3), pp. 2224-2287, Mar. 2019.

[25] S. Albawi, T. A. Mohammed, S. Al-Zawi. Understanding of a Convolutional Neural Network. International Conference on Engineering and Technology (ICET), 21-23 Aug. 2017, Antalya, Turkey, doi: 10.1109/ICEngTechnol.2017.8308186

[26] S.J. Pan, Q. Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), pp. 1345-1359, Oct. 2009.

[27] N. Panwar, S. Sharma, A.K. Singh. A survey on 5G: The next generation of mobile communication. Physical Communication, 18(2), pp. 64-84, March 2016.

[28] NGMN Alliance. Next generation mobile networks, 5G white paper, 2015.

[29] D. Li. 5G and intelligence medicine—how the next generation of wireless technology will reconstruct healthcare? Precision Clinical Medicine, 2(4), pp. 205-208, Dec. 2019.

[30] M. Series. IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU 2083-0, Sep. 2015.

[31] Akyildiz, S. Nie, S. Lin, M. Chandrasekaran. 5G roadmap: 10 key enabling technologies. Computer Networks, 106, pp. 17-48, Sep. 2016.

[32] M. Umar, J. Ferzund, T. Sajjad, S.M. Owais. Dealing Issues of Mobile Cloud Computing using 5G Technology. IJCSNS International Journal of Computer Science and Network Security, 17(8), pp. 246-250, Aug. 2017.

[33] B. Varghese, R. Buyya. Next generation cloud computing: New trends and research directions. Future Generation Computer Systems, 79(3), 2018, pp. 849-861.

[34] Y. Riahi, S. Riahi. Big Data and Big Data Analytics: Concepts, Types and Technologies. International Journal of Research and Engineering, 5(9), 2018, pp. 524-528.

[35] J. Okwuibe, M. Liyanage, I. Ahmad, M. Ylianttila. Cloud and MEC Security. In M. Liyanage et al. (Eds.) A Comprehensive Guide to 5G Security, pp. 373-397, 2018, Wiley Telecom

[36] Ahmad, S. Shahabuddin, T. Kumar, J. Okwuibe, A. Gurtov, M. Ylianttila. Security for 5G and Beyond. IEEE Communications Surveys & Tutorials, 21(4), pp. 3682-3722, May 2019.

[37] Y. Niu, Y. Li, D. Jin, L. Su, A. Vasilakos. A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges. WirelessNetworks, 21(8), pp. 2657-2676, 2015.

[38] EETT, https://www.eett.gr

[39] ThinkTech, "Τι είναι οι σταθμοί βάσης" Διαθέσιμο στο https://thinktech.gr/%CF%84%CE%B9-%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9-%CE%BF%CE%B9-%CF%83%CF%84%CE%B1%CE%B8%CE%BC%CE%BF%CE%AF-%CE%B2%CE%AC%CF%83%CE%B7%CF%82-%CE%BA%CE%B9%CE%BD%CE%B7%CF%84%CE%AE%CF%82/, 2016.

[40] F. Usmani. "Why small cells will power 5G", Available at: https://www.evaluationengineering.com/applications/5g-test/article/21044438/why-small-cells-will-power-5g, 2019.

[41] S. Khan. The Backbone of 5G Networks: A Guide to Small Cell Technology. Telit, March 12, 2020. Available at https://www.telit.com/blog/5g-networks-guide-to-small-cell-technology/

[42] Y. Yang, L. Wu, G. Yin, L. Li, H. Zhao. A Survey on Security and Privacy Issues in Internet-of-Thind. Internet of Things Journal, 4(5), 2017.

[43] R. Mehta, J. Sahni, K. Khanna. Internet of things: Vision, applications and research challenges. Procedia Computer Science, 132, pp. 1263-1269, 2018.

[44] U. Kar, D. Sanyal. An overview of device-to-device communication in cellular networks. ICT Express, 4(4), pp. 203-208, Dec. 2018.

[45] Y. Yu. Mobile edge computing towards 5G: Vision, recent progress, and open challenges. ChinaCommunications, 13(2), pp. 89–99, 2016.

[46] Ι. Μαυρίδης. Βασικές έννοιες και ζητήματα ασφάλειας. Στο Ι. Μαυρίδης (Επιμ.). Ασφάλεια πληροφοριών στο διαδίκτυο, 2015, Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών.

[47] K. Ahmad, S. Verma, N. Kumar, J. Shekhar. Classification of Internet Security Attacks. Proceedings of the 5th National Conference Computing For Nation Development, 10-11 Mar. 2011, Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi.

[48] F. Shahzad, M. Pasha, A. Ahmad. A Survey of Active Attacks on Wireless Sensor Networks and their Countermeasures. International Journal of Computer Science and Information Security, 14(12), pp.54-65, Dec. 2016.

[49] J. Teng, W. Gu, D. Xuan. Defending Against Physical Attacks in Wireless Sensor Networks. In Handbook on Securing Cyber-Physical Critical Infrastructure, pp. 251-279, Elsevier Inc., 2012.

[50] Cloudflare. "What is a DDoS Attack?". Available at: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/

[51] Tajpour, S. Ibrahim, M. Marsom. SQL Injection Detection and Prevention Techniques. International Journal of Advancements in Computing Technology, 3(7), pp. 82-91, 2011.

[52] O. Bizimana, T. Belkhouja. SQL injections and mitigationsScanning and Exploitation using SQLmap. CS 539: Applied Security Concepts, University of Idaho, 2017.

[53] J. Andress. The Basics of Information Security: Understanding the Fundamentals of Infosec in Theory and Practice. 2nd Ed., 2014, Elsevier Inc.

[54] M. L. Polla, F. Martinelli, D. Sgandurra. A Survey on Security for Mobile Devices. IEEE Communications Surveys Tutorials, 15(1), pp. 446-471, Jan. 2013.

[55] S. S. Vikas, K. Pawan, A. K. Gurudatt, G. Shyam.Mobile cloud computing: Security threats. In 2014 International Conference on Electronics and Communication Systems (ICECS), Feb 2014, pp. 1-4.

[56] V. Sucasas, G. Mantas, J. Rodriguez. Security Challenges for Cloud Radio Access Networks. In Backhauling/Fronthauling for Future Wireless Systems, pp. 195-211, 2016, doi: 10.1002/9781119170402.ch9

[57] M. Rouse. Single Point of Failure (SPOF). Available at: https://searchdatacenter.techtarget.com/definition/Single-point-of-failure-SPOF, Feb. 2012.

[58] I. Ahmad, S. Namal, M. Ylianttila, A. Gurtov. Security in Software Defined Networks: A Survey. IEEE Communications Surveys Tutorials, 17(4), pp. 2317-2346, Aug. 2015.

[59] W. Yang, C. Fung. A survey on security in network functions virtualization. In 2016 IEEE NetSoft Conference and Workshops (NetSoft), Jun. 2016, pp. 15-19.

[60] S. J. Vaughan-Nichols. Virtualization Sparks Security Concerns. Computer, 41(4), pp. 13-15, Aug. 2008.

[61] M. Casado, T. Koponen, R. Ramanathan, S. Shenker. Virtualizing the Network Forwarding Plane. ACM PRESTO 2010, 30 Nov. 2010, Philadelphia, USA.

[62] Z. Huang, X. Wamg, G. Huang. Radio Frequency Fingerprint Extraction Based on Multidimension Permutation Entropy. International Journal of Antennas and Propagation, 2017, https://doi.org/10.1155/2017/1538728

[63] P.L. Yu, J. S. Baras, B.M. Salder. Multicarrier Authentication at the Physical Layer. In 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks, 23-26 June, Newport Beach, CA, USA, 2008.

[64] H. Fang, X. Wang, L. Hanzo. Learning-aided Physical Layer Authentication as an Intelligent Process. IEEE Transactions on Communications, 67(3), pp. 2260-2273, March 2019b.

[65] X. Wang, P. Hao, L. Hanzo. Physical-layer authentication for wireless security enhancement: current challenges and future developments. IEEE Communications Magazine, 54(6), pp. 152-158, June 2016.

[66] Y. Yang, Z. Shi, Y. H. Chew, T. T. Tjhung. Channel Frequency Response Estimation for MIMO Systems with Frequency-Domain Equalization. EURASIP Journal on Advances in Signal Processing, 2011, doi:10.1155/2011/501703

[67] S. Kram, M. Stahkle, T. Feigl, J. Seitz, J. Thielecke. UWB Channel Impulse Responses for Positioning in Complex Environments: A Detailed Feature Analysis. Sensors, 19(24), 2019.

[68] Y.J. Lin, P.H. Tsend, Y.C. Chan, F.S. Wu. Super-Resolution-Aided Positioning Fingerprinting Based on Channel Impulse Response Measurement. IEEE Wireless Communications and Networking Conference (WCNC), 19-22 Mar. 2017, San Francisco, CA.

[69] J.W. Kim, J.W. Moon, S. Bahng, Y. Bang, Y. Park. A research on carrier frequency offset estimation for 5G telecommunication. International Conference on Information and Communication Technology Convergence (ICTC), 22-24 Oct. 2014, Busan, South Korea.

[70] A. Buchman, C. Lung. On the relationship between received signal strength and received signal strength index of IEEE 802.11 compatible radio transceivers. Carpathian Journal of Electronic and Computer Engineering, 6(2), pp. 15-20, 2013.

[71] T. Kumar, M. Liyanage, A. Braeken, I. Ahmad, M. Ylianttila. From Gadget to Gadget-Free Hyperconnected World: Conceptual Analysis of User Privacy Challenges. In 2017 European Conference on Networks and Communications (EuCNC), June 2017, pp. 1-6.

[72] R. Yu, Z. Bai, L. Yang, P. Wang, O. A. Move, Y. Liu. A Location Cloaking Algorithm Based on Combinatorial Optimization for Location-Based Services in 5G Networks. IEEE Access, 4, pp. 6515-6527, Oct. 2016.

[73] D. Kapetanovic, G. Zheng, F. Rusek. Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks. IEEE Communications Magazine, 53(6), pp. 21-27, June 2015.

[74] Α. Βογιατζής. Αρχιτεκτονική Ασφαλείας Δικτύων Ενσωματωμένων Συστημάτων. (Αδημοσίευτη Διδακτορική Διατριβή). Πανεπιστήμιο Πατρών. Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών, Πάτρα, 2008.

[75] B. Chen and C. Zhu and W. Li and J. Wei and V. C. M. Leung L. T. Yang. Original Symbol Phase Rotated Secure Transmission Against Powerful Massive MIMO Eavesdropper. IEEE Access, 4, pp. 3016-3025, June 2016.

[76] X. Wen, Y. Chen, C. Hu, C. Shi, Y. Wang. Towards a Secure Controller Platform for Openflow Applications. Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, pp. 171-172, 2013. New York, USA.

[77] S. Scott-Hayward, C. Kane, S. Sezer. Operation Checkpoint: SDN Application Control. IEEE 22nd International Conference on Network Protocols, 21-24 Oct. 2014, Raleigh, NC, USA.

[78] P. Porras, S. Shin, V. Yegneswaran, M. Fong, M. Tyson, G. Gu. A Security Enforcement Kernel for OpenFlow Networks. Proceedings of the First Workshop on

Hot Topics in Software Defined Networks, pp. 121-126, 13 Aug. 2012, Helsinki, Finland.

[79] Porras, P. Toward a More Secure SDN Control Layer — SRI International's View. Sdxcentral, Available at: https://www.sdxcentral.com/articles/contributed/toward-secure-sdn-control-layer/2013/10/, 2013.

[80] S. Shin, Y. Song, T. Lee, S. Lee, J. Chung, P. Porras, V. Yegneswaran, J. Noh, B.B. Kang. Rosemary: A Robust, Secure, and High-Performance Network Operating System. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 78-89, Nov. 2014.

[81] S. Shin, V. Yegneswaran, P. Porras, G. Gu. AVANT-GUARD: Scalable and Vigilant Switch Flow Management in Software-defined Networks. Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, pp. 413-424, Nov. 2013.

[82] P. Colp, M. Nanavati, J. Zhu, W. Aiello, G. Coker, T. Deegan, P. Loscocco, A. Warfield. Breaking Up is Hard to Do: Security and Functionality in a Commodity Hypervisor. Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, pp. 189-202, Oct. 2011.

[83] A. Al-Shabibi, M. D. Leenheer, M. Gerola, A. Koshibe, W. Snow, G. Parulkar. OpenVirteX: A Network Hypervisor. Open Networking Summit, Mar. 2014, Santa Clara, CA, USA.

[84] M. Pattaranantakul, R. He, A. Meddahi, Z. Zhang. SecMANO: Towards Network Functions Virtualization (NFV) Based Security MANagement and Orchestration. 2016 IEEE Trustcom/BigDataSE/ISPA, 23-26 Aug. 2016, Tianjin, China.

[85] Z. Yan, P. Zhang, A. V. Vasilakos. A security and trust framework for virtualized networks and software-defined networking. Security and communication networks, 9(16), pp. 3059-3069, Mar. 2016.

[86] Cloud Security Alliance. “Top Threats to Cloud Computing V1.0”. Available at: https://ioactive.com/wp-content/uploads/2018/05/csathreats.v1.0-1.pdf , Mar. 2010.

[87] ETSI TS 133 501 V15.4.0 (2019-05). “Security architecture and procedures for 5G System (3GPP TS 33.501 version 15.4.0 Release 15) ”. Available at:

https://www.etsi.org/, May 2019.

[88] SolarWinds MSP. “Common Network Authentication Methods”, Available at: https://www.solarwindsmsp.com/blog/network-authentication-methods, Apr. 2019.

[89] W. Meng, X. Luo, S. Furnell, J. Zhou. Protecting Mobile Networks and DevicesQ Challenges and Solutions. CRC Press, 2017.

[90] N. T. Nguyen, G. Zheng, Z. Han, R. Zheng. Device fingerprinting to enhance wireless security using nonparametric Bayesian method. Proceedings IEEE INFOCOM, pp. 1404-1412, 10-15 April 2011, Shanghai, China.

[91] Y. Wu, A. Khisti, C. Xiao, G. Caire, K. K. Wong, X. Gao. A survey of physical layer security techniques for 5G wireless networks and challenges ahead. IEEE Journal on Selected Areas in Communications, 36(4), pp. 679-695, April 2018.

[92] S. Gangadhar, P.G Sterbenz. Machine Learning Aided Traffic Tolerance to Improve Resilience for Software Defined Networks, 9th International Workshop on Resilient Networks Design and Modeling (RNDM), 4-6 Sept. 2017, Alghero.

[93] Lauren et al. 2018

[94] R. Braga, E. Mota, A. Passito. Lightweight DDoS Flooding Attack Detection Using NOX/OpenFlow. IEEE Local Computer Network Conference, pp. 408-415, 10-14 Oct. 2010, Denver, CO, USA.

[95] L. Bondan, T. Wauters, B. Volckaert, F. De Turck, L. Z. Granville. Anomaly detection framework for SFC integrity in NFV environments. IEEE Conference on Network Softwarization (NetSoft), 3-7 July 2017, Bologna, Italy.

[96] SDxCentral Studios. “What Is an NFV Orchestrator (NFVO)? Definition”. Available at: https://www.sdxcentral.com/networking/nfv/definitions/nfv-orchestrator-nfvo-definition/ , Mar. 2016.

[97] A. Pastor, A. Mozo, D. R. Lopez, J. Folgueira, A. Kapodistria. The Mouseworld, a security traffic analysis lab based on NFV/SD. ARES 2018: Proceedings of the 13th International Conference on Availability, Reliability and Security, 57, pp. 1-6, Aug. 2018.

[98] M. Ozay, I. Esnaola, F. T. YarmanVural, S. R. Kulkarni, H. V. Poor. Machine Learning Methods for Attack Detection in the Smart Grid. IEEE Trans. Neural Networks and Learning Systems, 27(8), pp. 1773-1786, 2015.

[99] L. Xiao, Y. Li, G. Han, G. Liu, W. Zhuang. PHY-Layer Spoofing Detection with Reinforcement Learning in Wireless Networks. IEEE Transactions on Vehicular Technology, 65(12), pp. 10037-10047, Feb. 2016.

[100] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, L. Hanzo. Machine Learning Paradigms for Next-Generation Wireless Networks. IEEE Wireless Communications Magazine, 24(2), pp. 98-105, 2018.

[101] M. Xia, Y. Owada, M. Inoue, H. Harai. Optical and wireless hybrid access networks: Design and optimization. IEEE/OSA Journal of Optical Communications and Networking, 4(10), pp. 749-759, Oct. 2012.

[102] A. Iacovazzi, A. Baiocchi. Internet Traffic Privacy Enhancement with Masking: Optimization and Tradeoffs. IEEE Transactions on Parallel and Distributed Systems, 25(2), pp. 353-362, Feb. 2014.

[103] F. Knirsch, G. Eibl, D. Engel. Error-Resilient Masking Approaches for Privacy Preserving Data Aggregation. IEEE Transactions on Smart Grid, 9(4), pp. 3351-3361, July 2018.

[104] B. Zoph, V. Vasudevan, J. Shlens, Q.V. Le. Learning Transferable Architectures for Scalable Image Recognition. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[105] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Hun, M.M. Iqbal, K. Han. Enhanced Network Anomaly Detection Based on Deep Neural Networks. IEEE Access, 6, pp. 48231-48246, Aug. 2018.

[106] D.E. Denning. An Intrusion-Detection Model. IEEE Transactions on Software Engineering, SE-13(2), Feb. 1987.

[107] A. Dawoud, A. Shahristani, C. Raun. Deep Learning for Network Anomalies Detection. International Conference on Machine Learning and Data Engineering (iCMLDE), 3-7 Dec. 2018, Sydney, Australia.

[108] D. Kwon, K. Natarajan, S. C. Suh, H. Kim. An Empirical Study on Network Anomaly Detection Using Convolutional Neural Networks. IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 2-6 July 2018, Vienna, Austria.

[109] R. Abdulhammed, M. Faezipour, A. Abuzneid, A. AbuMallouh, IEEE Sensors Letters, 3(1), Jan. 2019.

[110] I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy, H. Ming. AD-IoT: Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning.

IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 7-9 Jan. 2019, Las Vegas, USA.

[111] B. Zoph, Q.V. Le. Neural Architecture Search with Reinforcement Learning. Under review as a conference paper at ICLR, 2017.

[112] M. Caron, P. Bojanowski, J. Mairal, A. Joulin. Unsupervised Pre-Training of Image Features on Non-Curated Data. IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2019.

[113] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, Q.V. Le. MnasNet: Platform-Aware Neural Architecture Search for Mobile.

[114] Sucuri Guides. "OWASP Top 10 Security Risks & Vulnerabilities", Available at: https://sucuri.net/guides/owasp-top-10-security-vulnerabilities-2020/, Feb. 2020.

[115] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018. Available at: https://www.unb.ca/cic/datasets/ids-2018.html

[116] Simul8. "What's The Difference Between Arrival Rates and Inter Arrival Times?". Available at: https://support.simul8.com/whats-the-difference-between-arrival-rates-and-inter-arrival-times/

[117] Η. Χ. Χαματίδης. Χρήση μεθόδων μηχανικής μάθησης για την ανάπτυξη συστήματος αυθεντικοποίησης μέσω δεδομένων ηλεκτροκαρδιογραφήματος. (Αδημοσίευτη Πτυχιακή Εργασία). Πανεπιστήμιο Θεσσαλίας/ Σχολή Θετικών Επιστημών, Λαμία, Μάιος 2018.

[118] G. Gupta, R. Joyce. Identity authentication based on keystroke latencies. CommunicationsoftheACM, 33(2), pp. 168-176, 1990.

[119] Α. Κακούρης. Συνεχής αυθεντικοποίηση χρήστη στη διαδικτυακές εφαρμογές βάσει συμπεριφοράς. (Αδημοσίευτη Διπλωματική Εργασία). Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη, Μάρτιος 2018.

[120] D. Reynolds. Gaussian mixture models. Encyclopedia of Biometrics, 2009.

[121] S. Bleha, M. Obaidat. Performance of the Perceptron algorithm for the classification of computer users. Proceedings of the ACM/SIGAPP Symposium of Applied Computing: Technological Challenges of the 1990s, pp. 863-866, 1992.

[122] M. Brown. User identification via keystroke characteristics of typed names using neural networks. Man-Machine Studies, 39, pp. 999-1014, 1993.

[123] Wikipedia, Feedforward Neural Network, Available at: https://en.wikipedia.org/wiki/Feedforward_neural_network

[124] D.T. Lin. Computer-access authentication with neural network based keystroke identity verification. Proceedings of the International Conference of Neural Networks, pp. 174-178, 1997.

[125] MDN Web Docs, Cross-Origin Resource Sharing (CORS), Available at: https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS

[126] Wikipedia, Radial basis function kernel, Available at: https://en.wikipedia.org/wiki/Radial_basis_function_kernel#cite_note-primer-2

[127] Scitik learn, Support Vector Machines, Available at: https://scikit-learn.org/stable/modules/svm.html

Παράρτημα: Μοντέλο CNN με αρχιτεκτονική NAS