



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ**  
**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**Η επίδραση της Τεχνητής Νοημοσύνης στον τομέα της  
Κυβερνοασφάλειας**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

του

**Γεώργιου Αριστείδη Τσατσαρώνη**

**Επιβλέπουσα:** Διαμαντοπούλου Βασιλική

**Μέλη εξεταστικής επιτροπής:** Καρύδα Μαρία, Σπαθούλας Γεώργιος

Σάμος, Σεπτέμβριος 2026

Η σελίδα αυτή είναι σκόπιμα λευκή.

## **Πρόλογος και ευχαριστίες**

Θα ήθελα να ευχαριστήσω πρώτα απ' όλα την επιβλέπουσα καθηγήτριά μου για την κατανόηση, την υποστήριξη και τις κατευθυντήριες οδηγίες που μου έδωσε. Με την πολύτιμη βοήθειά της κατάφερα παρά τις όποιες δυσκολίες να ολοκληρώσω το έργο αυτό.

Επίσης, θέλω να ευχαριστήσω την οικογένειά μου για όλη την υποστήριξη που μου προσέφερε τις δύσκολες στιγμές, τους φίλους μου, την ψυχολόγο μου και γενικά οποιονδήποτε με άκουσε όταν χρειαζόμουν βοήθεια. Χωρίς αυτούς, η εργασία αυτή δε θα πραγματοποιούταν ποτέ.

© 2026

του

Γεώργιου Αριστείδη Τσατσαρώνη

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή.

## Πίνακας περιεχομένων

|          |                                                                         |          |
|----------|-------------------------------------------------------------------------|----------|
| <b>1</b> | <b>Εισαγωγή .....</b>                                                   | <b>1</b> |
| 1.1      | Ο ρόλος της τεχνητής νοημοσύνης στην ασφάλεια στον κυβερνοχώρο.....     | 1        |
| 1.2      | Αντικείμενο διπλωματικής.....                                           | 2        |
| 1.3      | Πεδίο εφαρμογής της μελέτης .....                                       | 3        |
| 1.4      | Δομή της διπλωματικής .....                                             | 3        |
| <b>2</b> | <b>Μεθοδολογία εργασίας.....</b>                                        | <b>5</b> |
| 2.1      | Συλλογή Βιβλιογραφικού Υλικού.....                                      | 5        |
| 2.2      | Κριτήρια Επιλογής και Απόρριψης Πηγών .....                             | 6        |
| 2.3      | Επεξεργασία και Ανάλυση Δεδομένων.....                                  | 6        |
| 2.4      | Περιορισμοί της Μεθοδολογίας.....                                       | 7        |
| 2.5      | Μεθοδολογία σε μορφή πίνακα & διαγράμματος.....                         | 7        |
| <b>3</b> | <b>Επισκόπηση του περιεχομένου της Κυβερνοασφάλειας.....</b>            | <b>8</b> |
| 3.1      | Η έννοια της Κυβερνοασφάλειας.....                                      | 8        |
| 3.2      | Ένας νέος ορισμός για την Κυβερνοασφάλεια .....                         | 10       |
| 3.3      | Ιστορική αναδρομή .....                                                 | 10       |
| 3.4      | Βασικές έννοιες στην Κυβερνοασφάλεια .....                              | 11       |
| 3.4.1    | CIA.....                                                                | 12       |
| 3.4.2    | AAA.....                                                                | 13       |
| 3.4.3    | 3Ps .....                                                               | 14       |
| 3.5      | Στόχοι Κυβερνοασφάλειας .....                                           | 15       |
| 3.5.1    | Πρόληψη Απειλών (Threat Prevention) .....                               | 15       |
| 3.5.2    | Αναγνώριση απειλών και Σκλήρυνση Συστήματος (System Hardening).....     | 15       |
| 3.5.3    | Υιοθέτηση επιχειρησιακών και τεχνολογικών καινοτομιών .....             | 16       |
| 3.5.4    | Ετοιμότητα και Σχεδιασμός Έκτακτης Ανάγκης (Contingency Planning) ..... | 16       |
| 3.5.5    | Διαμοιρασμός Πληροφοριών Απειλών (Threat Intelligence Sharing).....     | 16       |
| 3.5.6    | Εξειδικευμένη εκπαίδευση προσωπικού .....                               | 16       |
| 3.5.7    | Ενίσχυση της Ανοχής σε Σφάλματα (Fault Tolerance).....                  | 16       |
| 3.5.8    | Μείωση ευπαθειών (Vulnerability Management) .....                       | 16       |
| 3.5.9    | Βελτίωση της Ευχρηστίας (Usability).....                                | 16       |
| 3.5.10   | Ισχυρή Αυθεντικοποίηση (Robust Authentication) .....                    | 16       |
| 3.5.11   | Αυτοματοποίηση Διαδικασιών Ασφαλείας (Security Automation).....         | 17       |
| 3.5.12   | Διασφάλιση Διαλειτουργικότητας (Interoperability).....                  | 17       |

|        |                                                                           |    |
|--------|---------------------------------------------------------------------------|----|
| 3.5.13 | Καταγραφή και Ανάλυση Δυσμενών Περιστατικών (Incident Logging & Analysis) | 17 |
| 3.5.14 | Εφαρμογή Πολυεπίπεδων Μέτρων Ασφαλείας (Defense-in-Depth)                 | 17 |
| 3.6    | Υποδομή κυβερνοασφάλειας και αρχιτεκτονική διαδικτύου (NIST)              | 18 |
| 3.6.1  | Πυρήνας πλαισίου                                                          | 18 |
| 3.6.2  | Επίπεδα υλοποίησης                                                        | 19 |
| 3.6.3  | Προφίλ                                                                    | 19 |
| 3.6.4  | Αναγνώριση                                                                | 19 |
| 3.6.5  | Προστασία                                                                 | 20 |
| 3.6.6  | Εντοπισμός                                                                | 21 |
| 3.6.7  | Ανταπόκριση                                                               | 21 |
| 3.6.8  | Ανάκτηση                                                                  | 22 |
| 3.7    | Απειλές και τρωτά σημεία για την ασφάλεια στον κυβερνοχώρο                | 23 |
| 3.7.1  | Κρίσιμες υποδομές                                                         | 23 |
| 3.7.2  | Ασφάλεια δικτύου                                                          | 25 |
| 3.7.3  | Ασφάλεια στο νέφος                                                        | 26 |
| 3.7.4  | Ασφάλεια εφαρμογών                                                        | 27 |
| 3.7.5  | Ασφάλεια του Διαδικτύου των πραγμάτων                                     | 29 |
| 3.8    | Στρατηγικές κυβερνοασφάλειας                                              | 29 |
| 3.8.1  | Αμυντικά μέτρα προστασίας                                                 | 30 |
| 3.8.2  | Επιθετικά μέτρα προστασίας                                                | 31 |
| 3.9    | Τάσεις στην κυβερνοασφάλεια                                               | 33 |
| 3.9.1  | Η ασφάλεια στον κυβερνοχώρο αλλάζει στις εταιρείες                        | 33 |
| 3.9.2  | Διακομιστές Ιστού (Web servers)                                           | 34 |
| 3.9.3  | Υπολογιστικό νέφος και οι υπηρεσίες του                                   | 34 |
| 3.9.4  | Προηγμένες επίμονες απειλές                                               | 34 |
| 3.9.5  | Δίκτυα κινητής τηλεφωνίας                                                 | 34 |
| 3.9.6  | Κρυπτογράφηση δεδομένων                                                   | 34 |
| 3.9.7  | Εφαρμογές για την ασφάλεια δεδομένων                                      | 35 |
| 4      | Τεχνητή νοημοσύνη                                                         | 36 |
| 4.1    | Ορισμός και βασικές έννοιες της τεχνητής νοημοσύνης                       | 36 |
| 4.1.1  | Μηχανική Μάθηση (Machine Learning ML)                                     | 38 |
| 4.1.2  | Βαθιά Μάθηση (Deep Learning DL)                                           | 41 |
| 4.1.3  | Μεγάλα γλωσσικά μοντέλα                                                   | 46 |
| 4.2    | Τεχνολογίες ΤΝ                                                            | 48 |
| 4.2.1  | Ρομποτική                                                                 | 48 |
| 4.2.2  | Εξόρυξη δεδομένων                                                         | 48 |

|          |                                                                       |           |
|----------|-----------------------------------------------------------------------|-----------|
| 4.2.3    | Υπολογιστική Όραση .....                                              | 48        |
| 4.2.4    | Εικονικοί πράκτορες .....                                             | 48        |
| 4.2.5    | Υλικό βελτιστοποιημένο για την TN.....                                | 48        |
| 4.2.6    | Διαχείριση αποφάσεων.....                                             | 48        |
| 4.2.7    | Πλατφόρμες βαθιάς μάθησης .....                                       | 49        |
| 4.2.8    | Βιομετρία.....                                                        | 49        |
| 4.2.9    | Ανάλυση κειμένου και NLP .....                                        | 49        |
| 4.2.10   | Αναπαράσταση γνώσης και συλλογιστική .....                            | 49        |
| <b>5</b> | <b>Χρήσεις της TN στην Κυβερνοασφάλεια .....</b>                      | <b>50</b> |
| 5.1      | Κακόβουλες χρήσεις της TN στην ασφάλεια στον κυβερνοχώρο .....        | 51        |
| 5.1.1    | Αντίπαλη TN στις απειλές κυβερνοασφάλειας.....                        | 51        |
| 5.2      | Ηθικές χρήσεις της TN στην κυβερνοασφάλεια .....                      | 58        |
| 5.2.1    | Αυτοματοποιημένη ανάλυση, πληροφόρηση και ανίχνευση για απειλές ..... | 58        |
| 5.2.2    | Αξιολόγηση ευπαθειών(Vulnerability assessment).....                   | 59        |
| 5.2.3    | TN για την αντίδραση σε συμβάντα και την αποκατάσταση .....           | 60        |
| 5.2.4    | Χρήση της Προγνωστικής ανάλυσης (Predictive Analytics) .....          | 61        |
| 5.2.5    | Εφαρμογές στον πραγματικό κόσμο και μελέτες περίπτωσης.....           | 61        |
| 5.3      | Προκλήσεις & Περιορισμοί .....                                        | 62        |
| 5.3.1    | Ευπάθεια στον Χειρισμό και την Κακόβουλη Χρήση .....                  | 62        |
| 5.3.2    | Περιορισμοί στην Ανίχνευση Απειλών .....                              | 63        |
| 5.3.3    | Δυνατότητα Κακόβουλης Χρήσης από Τρίτους .....                        | 63        |
| 5.3.4    | Περιορισμένη Κατανόηση και Χρήση από την Κοινωνία.....                | 64        |
| 5.3.5    | Υψηλό Κόστος Υλοποίησης.....                                          | 64        |
| <b>6</b> | <b>Συμπεράσματα και Προοπτικές για το μέλλον .....</b>                | <b>65</b> |
| 6.1      | Σύνοψη.....                                                           | 65        |
| 6.2      | Συμπεράσματα .....                                                    | 66        |
| 6.3      | Μελλοντικές κατευθύνσεις .....                                        | 67        |
|          | <b>Βιβλιογραφία.....</b>                                              | <b>68</b> |



## Λίστα Σχημάτων

|                                                                                                                                        |    |
|----------------------------------------------------------------------------------------------------------------------------------------|----|
| Σχήμα 1: Βήματα μεθοδολογίας εκπόνησης της εργασίας.....                                                                               | 7  |
| Σχήμα 2: Η τριάδα C.I.A. δείχνει τα βασικά στοιχεία της ασφάλειας: Εμπιστευτικότητα, Ακεραιότητα και Διαθεσιμότητα.....                | 13 |
| Σχήμα 3: Το μοντέλο AAA (Authentication, Authorization, Accounting) .....                                                              | 14 |
| Σχήμα 4: Το μοντέλο 3P (People, Practice, Plan) .....                                                                                  | 15 |
| Σχήμα 5: Πλαίσιο κυβερνοασφάλειας NIST .....                                                                                           | 18 |
| Σχήμα 6: Μηχανισμός Phishing .....                                                                                                     | 26 |
| Σχήμα 7: Γραφικός κύκλος ζωής της TN .....                                                                                             | 38 |
| Σχήμα 8: Συνελικτικό νευρωνικό δίκτυο.....                                                                                             | 42 |
| Σχήμα 9: Ένα επαναλαμβανόμενο νευρωνικό δίκτυο .....                                                                                   | 44 |
| Σχήμα 10: Η κυτταρική δομή μιας μονάδας μακράς βραχυπρόθεσμης μνήμης (LSTM) .....                                                      | 44 |
| Σχήμα 11: Η δομή των κυττάρων μιας ελεγχόμενης επαναλαμβανόμενης μονάδας (GRU) .....                                                   | 45 |
| Σχήμα 12: Νευρωνικό δίκτυο τροφοδότησης ενός κρυφού στρώματος με μία έξοδο .....                                                       | 46 |
| Σχήμα 13: Διάγραμμα επίθεσης αποφυγής (evasion attack) εναντίον εκπαιδευμένου μοντέλου μηχανικής μάθησης κατά τη φάση της δοκιμής..... | 56 |



## Λίστα Πινάκων

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| Πίνακας 1: Κριτήρια επιλογής και απόρριψης βιβλιογραφίας.....               | 6  |
| Πίνακας 2: Μεθοδολογία εκπόνησης εργασίας.....                              | 7  |
| Πίνακας 3: Πλαίσιο κυβερνοασφάλειας NIST .....                              | 23 |
| Πίνακας 4: Απειλές και τρωτά σημεία για την ασφάλεια στον κυβερνοχώρο ..... | 29 |
| Πίνακας 5: Μέτρα ασφαλείας 2 κατηγοριών .....                               | 33 |
| Πίνακας 6: Κακόβουλες χρήσεις της ΤΝ στην ασφάλεια στον κυβερνοχώρο .....   | 57 |



## Ακρωνύμια

| Ακρωνύμιο | Αγγλική Σημασία                                | Ελληνική Σημασία                               |
|-----------|------------------------------------------------|------------------------------------------------|
| TN        | Artificial Intelligence (AI)                   | Τεχνητή Νοημοσύνη                              |
| CIA       | Confidentiality, Integrity, Availability       | Εμπιστευτικότητα, Ακεραιότητα, Διαθεσιμότητα   |
| AAA       | Authentication, Authorization, Accounting      | Αυθεντικοποίηση, Εξουσιοδότηση, Καταγραφή      |
| MFA       | Multi-Factor Authentication                    | Έλεγχος Ταυτότητας Πολλαπλών Παραγόντων        |
| IoCs      | Indicators of Compromise                       | Ενδείξεις Παραβίασης                           |
| IoT       | Internet of Things                             | Διαδίκτυο των Πραγμάτων                        |
| NIST      | National Institute of Standards and Technology | Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας     |
| CSF       | Cybersecurity Framework                        | Πλαίσιο Κυβερνοασφάλειας                       |
| VPN       | Virtual Private Network                        | Εικονικό Ιδιωτικό Δίκτυο                       |
| DoS       | Denial of Service                              | Άρνηση Παροχής Υπηρεσίας                       |
| DDoS      | Distributed Denial of Service                  | Κατανεμημένη Άρνηση Παροχής Υπηρεσίας          |
| MITM      | Man-In-The-Middle                              | Επίθεση Μεσάζοντα                              |
| IDS       | Intrusion Detection System                     | Σύστημα Ανίχνευσης Εισβολών                    |
| SIEM      | Security Information and Event Management      | Διαχείριση Πληροφοριών και Συμβάντων Ασφαλείας |
| APT       | Advanced Persistent Threat                     | Προηγμένη Επίμονη Απειλή                       |
| 3P        | People, Plan, Practice                         | Άνθρωποι, Σχεδιασμός, Εκπαίδευση               |
| XSS       | Cross-Site Scripting                           | Διασταυρούμενη Ενσωμάτωση Κώδικα σε Ιστότοπο   |
| CUC       | Compromised User Credentials                   | Παραβιασμένα Διαπιστευτήρια Χρήστη             |
| SQL       | Structured Query Language                      | Δομημένη Γλώσσα Ερωτημάτων                     |
| CTI       | Cyber Threat Intelligence                      | Πληροφορίες για Κυβερνοαπειλές                 |
| AV        | Antivirus                                      | Αντιϊκό Λογισμικό                              |
| ML        | Machine Learning                               | Μηχανική Μάθηση                                |
| DL        | Deep Learning                                  | Βαθιά Μάθηση                                   |
| ANN       | Artificial Neural Network                      | Τεχνητό Νευρωνικό Δίκτυο                       |
| RNN       | Recurrent Neural Network                       | Αναδρομικό Νευρωνικό Δίκτυο                    |

|      |                                                         |                                                       |
|------|---------------------------------------------------------|-------------------------------------------------------|
| CNN  | Convolutional Neural Network                            | Συνελικτικό Νευρωνικό Δίκτυο                          |
| LSTM | Long Short-Term Memory                                  | Μνήμη Μακράς και Βραχείας Διάρκειας                   |
| GRU  | Gated Recurrent Unit                                    | Ελεγχόμενη Αναδρομική Μονάδα                          |
| NLP  | Natural Language Processing                             | Επεξεργασία Φυσικής Γλώσσας                           |
| FNN  | Feedforward Neural Network                              | Προωθητικό Νευρωνικό Δίκτυο                           |
| BERT | Bidirectional Encoder Representations from Transformers | Διπλής Κατεύθυνσης Αναπαραστάσεις από Μετασχηματιστές |
| GPT  | Generative Pre-trained Transformer                      | Παραγωγικός Προεκπαιδευμένος Μετασχηματιστής          |
| SOAR | Security Orchestration, Automation, and Response        | Συντονισμός, Αυτοματοποίηση και Απόκριση Ασφάλειας    |



## Περίληψη

Η Τεχνητή Νοημοσύνη (TN) μεταμορφώνει το τοπίο της Κυβερνοασφάλειας προσφέροντας προηγμένα εργαλεία για την ανίχνευση απειλών, την αντιμετώπιση περιστατικών και αυτοματοποιημένους μηχανισμούς άμυνας. Η παρούσα διπλωματική εργασία εξετάζει κριτικά τον διττό αντίκτυπο της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια, αναδεικνύοντας τα οφέλη της στην ενίσχυση των επιχειρήσεων ασφαλείας, καθώς και τους αναδυόμενους κινδύνους που συνδέονται με την κακόβουλη εκμετάλλευσή της. Μέσω μιας ολοκληρωμένης επισκόπησης της τρέχουσας βιβλιογραφίας και της ανάλυσης των σύγχρονων εφαρμογών TN, η μελέτη διερευνά βασικές τεχνολογίες όπως η μηχανική μάθηση, η βαθιά μάθηση και η επεξεργασία φυσικής γλώσσας, όπως εφαρμόζονται στην ανίχνευση εισβολών, την ανάλυση κακόβουλου λογισμικού και τα συστήματα αυτοματοποιημένης αντίδρασης. Εξετάζονται επίσης ηθικές προκλήσεις, συμπεριλαμβανομένων της προκατάληψης, της διαφάνειας και της πιθανής κατάχρησης. Η έρευνα παρέχει τελικά εφαρμόσιμες συστάσεις για την ανάπτυξη υπεύθυνων πρακτικών Κυβερνοασφάλειας με βάση την TN που εξισορροπούν την αποτελεσματικότητα, την ασφάλεια και τα ηθικά πρότυπα. Προσδιορίζονται οι μελλοντικές τάσεις και ερευνητικές κατευθύνσεις για να καθοδηγήσουν την εξέλιξη της TN στη διασφάλιση των ψηφιακών υποδομών.

**Λέξεις Κλειδιά:** Τεχνητή Νοημοσύνη, Κυβερνοασφάλεια, Μηχανική Μάθηση, Ηθική Τεχνητή Νοημοσύνη, Ανίχνευση Απειλών



## Abstract

Artificial Intelligence (AI) is transforming the Cybersecurity landscape by offering advanced tools for threat detection, incident response, and automated defense mechanisms. This thesis critically examines the dual impact of AI on cybersecurity, highlighting its benefits in enhancing security operations as well as the emerging risks associated with its malicious exploitation. Through a comprehensive review of current literature and analysis of state-of-the-art AI applications, the study explores key technologies such as machine learning, deep learning, and natural language processing as applied to intrusion detection, malware analysis, and automated response systems. Ethical challenges including bias, transparency, and potential misuse are also addressed. The research ultimately provides actionable recommendations for developing responsible AI-driven Cybersecurity practices that balance efficiency, security, and ethical standards. Future trends and research directions are identified to guide the evolution of AI in safeguarding digital infrastructures.

**Keywords:** Artificial Intelligence, Cybersecurity, Machine Learning, Ethical AI, Threat Detection

# 1

## *Εισαγωγή*

### *1.1 Ο ρόλος της τεχνητής νοημοσύνης στην ασφάλεια στον κυβερνοχώρο*

Ο όρος κυβερνοασφάλεια αναφέρεται σε ένα σύνολο τεχνολογιών, διαδικασιών και πρακτικών για την προστασία και την υπεράσπιση δικτύων, συσκευών, λογισμικού και δεδομένων από επιθέσεις, ζημιές ή μη εξουσιοδοτημένη πρόσβαση.

Η κυβερνοασφάλεια γίνεται όλο και πιο πολύπλοκη λόγω της εκθετικής αύξησης των διασυνδεδεμένων συσκευών, συστημάτων και δικτύων. Το γεγονός αυτό επιδεινώνεται από την πρόοδο της ψηφιακής οικονομίας και υποδομής, οδηγώντας σε σημαντική αύξηση των κυβερνοεπιθέσεων με σοβαρές συνέπειες.

Επιπλέον, υπάρχει συνεχής εξέλιξη των αντιπάλων που συνδέονται με τα κράτη και τους εγκληματίες, καθώς και την αυξανόμενη πολυπλοκότητα των κυβερνοεπιθέσεων, οι οποίες βρίσκουν νέους και επεμβατικούς τρόπους για να στοχεύουν ακόμη και τους πιο επιδέξιους στόχους. Η εξέλιξη αυτή οδηγεί σε αύξηση του αριθμού, της κλίμακας και του αντίκτυπου των κυβερνοεπιθέσεων και καθιστά αναγκαία την εφαρμογή της κυβερνοασφάλειας με βάση τις πληροφορίες για την παροχή δυναμικής άμυνας έναντι των εξελισσόμενων κυβερνοεπιθέσεων και τη διαχείριση μεγάλων δεδομένων.

Συμβουλευτικοί οργανισμοί ενθαρρύνουν επίσης τη χρήση πιο προληπτικών και προσαρμοστικών προσεγγίσεων, στρεφόμενοι προς αξιολογήσεις σε πραγματικό χρόνο, συνεχή παρακολούθηση και ανάλυση με βάση τα δεδομένα για τον εντοπισμό, την προστασία, την ανίχνευση, την αντιμετώπιση και την καταγραφή των κυβερνοεπιθέσεων για την πρόληψη μελλοντικών περιστατικών ασφαλείας.

Η τεχνητή νοημοσύνη (TN) είναι ένα ενδιαφέρον εργαλείο που μπορεί να παρέχει αναλύσεις και ευφυΐα για την προστασία από τις συνεχώς εξελισσόμενες κυβερνοεπιθέσεις αναλύοντας γρήγορα εκατομμύρια γεγονότων και παρακολουθώντας μια ευρεία ποικιλία απειλών στον κυβερνοχώρο για να προβλέψει και να δράσει πριν να ανακύψει πρόβλημα.

Για τον λόγο αυτό, η TN ενσωματώνεται όλο και περισσότερο στον ιστό της κυβερνοασφάλειας και χρησιμοποιείται σε διάφορες περιπτώσεις χρήσης για την αυτοματοποίηση εργασιών ασφαλείας ή την υποστήριξη του έργου των ανθρώπινων ομάδων ασφαλείας. Ο ανθηρός τομέας της ασφάλειας στον κυβερνοχώρο και ο αυξανόμενος ενθουσιασμός των ερευνητών τόσο από την TN όσο και από την ασφάλεια στον κυβερνοχώρο έχουν οδηγήσει σε πολυάριθμες μελέτες για την επίλυση προβλημάτων που σχετίζονται με τον εντοπισμό, την προστασία, την ανίχνευση, την αντιμετώπιση και την αποκατάσταση από επιθέσεις στον κυβερνοχώρο.

Ο αυξανόμενος αριθμός και η πολυπλοκότητα των επιθέσεων στον κυβερνοχώρο έχουν υπογραμμίσει την επείγουσα ανάγκη για καινοτόμες λύσεις για την ενίσχυση της ασφάλειας των ψηφιακών υποδομών. Μεταξύ αυτών των λύσεων, η TN αναδεικνύεται ως μια πολλά υποσχόμενη τεχνολογία με τη δυνατότητα να ενισχύσει σημαντικά την ασφάλεια στον κυβερνοχώρο και την αντιμετώπιση περιστατικών. Έχει την ικανότητα να βελτιώσει την ταχύτητα και την ακρίβεια της ανίχνευσης, της αντίδρασης και του μετριασμού των απειλών, μειώνοντας παράλληλα τον φόρτο εργασίας των επαγγελματιών ασφαλείας. [1]

## ***1.2 Αντικείμενο διπλωματικής***

Ο πρωταρχικός στόχος της εργασίας αυτής, είναι να εξετάσει κριτικά τον αντίκτυπο της TN στην ασφάλεια στον κυβερνοχώρο είτε με θετικό τρόπο, είτε κακόβουλο, αξιολογώντας τη συμβολή, τους περιορισμούς και τις μελλοντικές επιπτώσεις της. Συγκεκριμένα, η εργασία στοχεύει να αναλύσει το πώς οι τεχνολογίες TN ενισχύουν τις πρακτικές κυβερνοασφάλειας, από την προηγμένη ανίχνευση απειλών έως την αυτοματοποιημένη απόκριση σε συμβάντα και να αξιολογήσει τους περιορισμούς που προκύπτουν με την εξάρτηση από συστήματα TN.

Επιπλέον, η εργασία αυτή διερευνά τις ηθικές επιπτώσεις της TN στην ασφάλεια στον κυβερνοχώρο, εστιάζοντας σε ζητήματα όπως η μεροληψία, η διαφάνεια και η πιθανή χρήση εργαλείων που βασίζονται στην TN σε κακόβουλες επιθέσεις. Επίσης επιδιώκει να διερευνήσει το εξελισσόμενο τοπίο των εφαρμογών TN στον κυβερνοχώρο, εντοπίζοντας τομείς όπου η TN

μπορεί να γεφυρώσει τα υπάρχοντα κενά και προβλέποντας μελλοντικές τάσεις που μπορούν να διαμορφώσουν τη βιομηχανία. Διεξάγοντας μια συστηματική ανασκόπηση της σχετικής βιβλιογραφίας και αναλύοντας τα διαθέσιμα στοιχεία, η εργασία στοχεύει να παρέχει μια ολοκληρωμένη επισκόπηση τόσο των τεχνικών όσο και των ηθικών παραμέτρων που αφορούν την ΤΝ στην ασφάλεια στον κυβερνοχώρο.

Τελικά, φιλοδοξεί να προσφέρει συστάσεις για την ενίσχυση της αποτελεσματικότητας και της ακεραιότητας της ΤΝ στην ασφάλεια στον κυβερνοχώρο, καθοδηγώντας τους ενδιαφερόμενους φορείς και τους ερευνητές του κλάδου προς υπεύθυνες πρακτικές ανάπτυξης ΤΝ που εξισορροπούν την ασφάλεια, την αποτελεσματικότητα και τα ηθικά πρότυπα.

### ***1.3 Πεδίο εφαρμογής της μελέτης***

Η εργασία εστιάζει στον ρόλο της ΤΝ στις σύγχρονες πρακτικές κυβερνοασφάλειας, διερευνώντας τόσο τις δυνατότητές της όσο και τους περιορισμούς της στη διαφύλαξη ψηφιακών περιβαλλόντων. Το εύρος αυτής της εργασίας περιλαμβάνει την αναφορά σε βασικές τεχνολογίες ΤΝ, συμπεριλαμβανομένης της μηχανικής εκμάθησης, της βαθιάς μάθησης και της επεξεργασίας φυσικής γλώσσας, και τις συγκεκριμένες εφαρμογές τους στην ασφάλεια του κυβερνοχώρου, όπως η ανίχνευση εισβολών, η ανάλυση κακόβουλου λογισμικού και τα αυτοματοποιημένα συστήματα απόκρισης. Επιπλέον, η εργασία αυτή εξετάζει τις προκλήσεις που σχετίζονται με την υιοθέτηση της ΤΝ, ιδιαίτερα τεχνικούς περιορισμούς όπως η επεκτασιμότητα και η ακρίβεια, καθώς και ηθικές ανησυχίες όπως η μεροληψία και η διαφάνεια.

### ***1.4 Δομή της διπλωματικής***

Αυτή η εργασία διαρθρώνεται σε έξι συνολικά κεφάλαια, καθένα από τα οποία εστιάζει σε διαφορετικές πτυχές της ΤΝ στην ασφάλεια στον κυβερνοχώρο.

Στο Κεφάλαιο 1 πραγματοποιείται μια εισαγωγή στο θέμα, περιγράφοντας τον ρόλο της ΤΝ, τους ερευνητικούς στόχους και το πεδίο εφαρμογής.

Στο Κεφάλαιο 2 εξηγείται ο τρόπος περισυλλογής της βιβλιογραφίας όλης της εργασίας καθώς και τα κριτήρια διατήρησης/απόρριψης αυτής.

Στο Κεφάλαιο 3 γίνεται, στη συνέχεια, μια περιεκτική βιβλιογραφική ανασκόπηση, εξετάζοντας την τρέχουσα έρευνα, τα θεωρητικά πλαίσια και τις πρακτικές εφαρμογές της ΤΝ στην ασφάλεια στον κυβερνοχώρο.

Το Κεφάλαιο 4 εμβαθύνει στις βασικές αρχές της ΤΝ, συζητώντας τις βασικές τεχνολογίες όπως η μηχανική μάθηση και η βαθιά μάθηση.

Το Κεφάλαιο 5 διερευνά τόσο κακόβουλες όσο και ηθικές εφαρμογές της ΤΝ στην ασφάλεια στον κυβερνοχώρο, συμπεριλαμβανομένων των πιθανών κινδύνων και των βέλτιστων

πρακτικών, καθώς και τις προκλήσεις και τους περιορισμούς της ΤΝ, καλύπτοντας τεχνικούς, ηθικούς και λειτουργικούς περιορισμούς.

Τέλος, το Κεφάλαιο 6 ολοκληρώνεται με μια σύνοψη βασικών ευρημάτων, συστάσεις για τη βελτίωση της ΤΝ στην ασφάλεια στον κυβερνοχώρο και πληροφορίες για τις μελλοντικές τάσεις.

# 2

## *Μεθοδολογία εργασίας*

Η παρούσα διπλωματική εργασία βασίζεται σε μία ποιοτική μεθοδολογική προσέγγιση, με στόχο την κριτική επισκόπηση της βιβλιογραφίας και την ανάλυση των σύγχρονων τάσεων στη χρήση της ΤΝ στον τομέα της ασφάλειας στον κυβερνοχώρο. Η μεθοδολογία που ακολουθήθηκε περιλαμβάνει τα ακόλουθα βασικά στάδια:

### *2.1 Συλλογή Βιβλιογραφικού Υλικού*

Η συλλογή των πηγών πραγματοποιήθηκε μέσω:

- Ακαδημαϊκών βάσεων δεδομένων (όπως IEEE Xplore, ScienceDirect, SpringerLink)
- Κύριας μηχανής αναζήτησης το Google Scholar
- Επιστημονικών περιοδικών και συνεδρίων (ACM, Elsevier, IEEE S&P)
- Επίσημων εκθέσεων οργανισμών ασφάλειας στον κυβερνοχώρο (π.χ. ENISA, NIST, ISO)
- Τεχνικών white papers από κορυφαίες εταιρείες στον χώρο της κυβερνοασφάλειας (π.χ. IBM, Palo Alto Networks)

Χρησιμοποιήθηκαν λέξεις-κλειδιά όπως: «*cybersecurity*», «*artificial intelligence*», «*machine learning in cybersecurity*», «*adversarial AI*», «*bias in AI*», «*ethical AI*», «*deep learning*», «*malicious uses of AI*», «*automated threat detection*», και ακρωνύμια όπως: «*AI*», «*ML*», «*DL*», «*NLP*», «*ANN*», «*NIST*».

## 2.2 Κριτήρια Επιλογής και Απόρριψης Πηγών

Για να διασφαλιστεί η ποιότητα και η σχετικότητα του υλικού:

- Διατηρήθηκαν μόνο πηγές από το 2015 και έπειτα, ώστε να καλύπτονται οι πρόσφατες τεχνολογικές εξελίξεις
- Έγινε έλεγχος της εγκυρότητας των πηγών (peer-reviewed άρθρα ή επίσημες δημοσιεύσεις)
- Απορρίφθηκαν άρθρα, με βάση τα παρακάτω κριτήρια:

| Φάση Ελέγχου                          | Κριτήρια Επιλογής                                                                       | Κριτήρια Απόρριψης                                               |
|---------------------------------------|-----------------------------------------------------------------------------------------|------------------------------------------------------------------|
| Πριν την ανάγνωση                     | Επιστημονικά άρθρα, βιβλιοκεφάλαια, δημοσιεύσεις σε περιοδικά ή συνέδρια                | Άρθρα χωρίς περίληψη και εισαγωγή                                |
| Κατά την ανάγνωση περίληψης/εισαγωγής | Συγκεκριμένη αναφορά στην ΤΝ και την κυβερνοασφάλεια                                    | Άσχετο περιεχόμενο παρά τον τίτλο ή αναφορές σε άλλα πεδία       |
| Κατά την ανάγνωση πλήρους κειμένου    | Peer-reviewed, με επιστημονική τεκμηρίωση και παραπομπές                                | Ειδησεογραφικά άρθρα, blog posts, άρθρα χωρίς βιβλιογραφία       |
|                                       | Συγγραφείς που ανήκουν σε αναγνωρισμένους φορείς (πανεπιστήμια, οργανισμούς, εταιρείες) | Συγγραφείς χωρίς αναγνωρισμένη σύνδεση με τον κλάδο              |
|                                       | Εξειδικευμένη εστίαση σε ζητήματα ΤΝ ή κυβερνοασφάλειας                                 | Γενικά άρθρα περί «ασφάλειας» χωρίς τεχνική ή θεματική εμβάθυνση |

Πίνακας 1: Κριτήρια επιλογής και απόρριψης βιβλιογραφίας

## 2.3 Επεξεργασία και Ανάλυση Δεδομένων

Η ανάλυση της βιβλιογραφίας έγινε με βάση θεματικές κατηγορίες, όπως:

- Εφαρμογές της ΤΝ στην ανίχνευση και απόκριση απειλών
- Ηθικά ζητήματα της ΤΝ (π.χ. μεροληψία, αδιαφάνεια)
- Περιορισμοί και προκλήσεις χρήσης της ΤΝ στην ασφάλεια

Χρησιμοποιήθηκε η μέθοδος της συγκριτικής ανάλυσης για να εντοπιστούν αντιφάσεις, ομοιότητες και κενά στη διεθνή βιβλιογραφία.

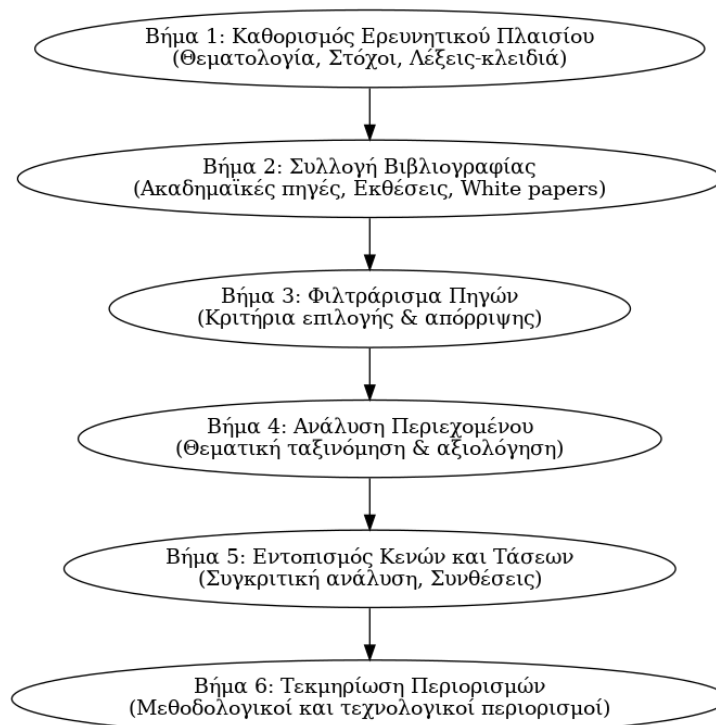
## 2.4 Περιορισμοί της Μεθοδολογίας

Παρόλο που η προσέγγιση αυτή προσφέρει βάθος στη θεωρητική κατανόηση, δεν περιλαμβάνει πρωτογενή ερευνητικά δεδομένα. Επιπλέον, η ταχεία εξέλιξη της TN ενδέχεται να καταστήσει ορισμένα συμπεράσματα παρωχημένα σε μικρό χρονικό διάστημα.

## 2.5 Μεθοδολογία σε μορφή πίνακα & διαγράμματος

| Βήμα                  | Περιγραφή                                                                               |
|-----------------------|-----------------------------------------------------------------------------------------|
| Συλλογή Βιβλιογραφίας | Αναζήτηση σε βάσεις δεδομένων (IEEE, ACM, ENISA, NIST, κ.ά.) με σχετικές λέξεις-κλειδιά |
| Καθορισμός Κριτηρίων  | Επιλογή πηγών από το 2015+, peer-reviewed ή επίσημες εκθέσεις, αποφυγή ανεπαρκών άρθρων |
| Ανάλυση Περιεχομένου  | Ομαδοποίηση σε θεματικές ενότητες βάσει της χρήσης της TN στην κυβερνοασφάλεια          |
| Συγκριτική Ανάλυση    | Εντοπισμός ομοιοτήτων, διαφορών, καινοτομιών και κενών στη βιβλιογραφία                 |
| Καταγραφή Περιορισμών | Τεχνικά και μεθοδολογικά όρια, έλλειψη πρωτογενούς έρευνας, ταχεία απαρχαίωση δεδομένων |

**Πίνακας 2: Μεθοδολογία εκπόνησης εργασίας**



**Σχήμα 1: Βήματα μεθοδολογίας εκπόνησης της εργασίας**

# 3

## *Επισκόπηση του περιεχομένου της Κυβερνοασφάλειας*

### *3.1 Η έννοια της Κυβερνοασφάλειας*

Η ασφάλεια στον κυβερνοχώρο είναι ένας ευρέως χρησιμοποιούμενος όρος, του οποίου οι ορισμοί είναι εξαιρετικά μεταβλητοί, συχνά υποκειμενικοί και ενίοτε μη κατατοπιστικοί. Η απουσία ενός συνοπτικού και ευρέως αποδεκτού ορισμού που να αποτυπώνει την πολυδιάστατη φύση της, εμποδίζει την τεχνολογική και επιστημονική πρόοδο του τομέα αυτού. Παράλληλα, ενισχύει μια στενά τεχνική θεώρηση, διαχωρίζοντας επιστημονικούς κλάδους που θα έπρεπε να συνεργάζονται για την επίλυση των σύνθετων προκλήσεων της. [2]

Καθώς το μεγαλύτερο μέρος των πληροφοριών παγκοσμίως είναι πλέον αποθηκευμένο σε ψηφιακά συστήματα και βάσεις δεδομένων, η απειλή παραβίασής τους αποτελεί έναν από τους πιο κρίσιμους κινδύνους σε διεθνές επίπεδο. Από το γεγονός αυτό αναδύεται μια επιτακτική ανάγκη για εμπειρογνώμονες στον τομέα. Ωστόσο, το ακριβές αντικείμενο της κυβερνοασφάλειας, καθώς και οι δεξιότητες που απαιτούνται από την αγορά εργασίας, παραμένουν συχνά ασαφή.

Υπάρχουν πολλοί ορισμοί για την ασφάλεια στον κυβερνοχώρο, τους κινδύνους που ενέχει και τη σημασία διαφόρων παραγόντων. Η κυβερνοασφάλεια είναι η πρακτική της υπεράσπισης υπολογιστών, διακομιστών, κινητών συσκευών, ηλεκτρονικών συστημάτων, δικτύων και δεδομένων από κακόβουλες επιθέσεις. Είναι επίσης γνωστή ως ασφάλεια τεχνολογίας πληροφοριών ή ασφάλεια ηλεκτρονικών πληροφοριών. Ο όρος εφαρμόζεται σε ποικίλα πλαίσια, από τις επιχειρήσεις έως την κινητή πληροφορική και μπορεί να χωριστεί σε μερικές κοινές κατηγορίες.

Το πρώτο βήμα για την κατανόηση του ορισμού της κυβερνοασφάλειας είναι να εξετάσουμε τι προσπαθεί να κάνει: δηλαδή να υπερασπιστεί τα δεδομένα από κακόβουλες επιθέσεις. Το επόμενο κομμάτι του ορισμού είναι το τι πρέπει να υπερασπιστεί. Με τον πολλαπλασιασμό των δεδομένων, των συσκευών ελέγχου και την επέκταση του Διαδικτύου των πραγμάτων (IoT), οι τομείς που πρέπει να προστατευθούν περιλαμβάνουν όλες τις συνδεδεμένες συσκευές, καθώς και όλα τα ανεξάρτητα συστήματα ή δεδομένα. Το IoT πρόκειται για την έννοια της σύνδεσης οποιασδήποτε συσκευής στο Διαδίκτυο (ή/και μεταξύ αυτών). Αυτό περιλαμβάνει μεγάλο εύρος συσκευών, από κινητές συσκευές και έξυπνες οικιακές συσκευές (ψυγεία, φωτισμό, ηχεία) μέχρι οχήματα και συστήματα ελέγχου [3], [4]. Συνεπώς, στον σύγχρονο κόσμο της κυβερνοασφάλειας, κάθε συνδεδεμένη ή αυτόνομη ηλεκτρονική συσκευή αποτελεί εν δυνάμει στόχο και χρήζει ανάλογης θωράκισης.

Επιπλέον, σημειώνονται έξι γενικοί τομείς που περιλαμβάνονται στην κυβερνοασφάλεια:

Η ασφάλεια δικτύου είναι η πρακτική της εξασφάλισης ενός δικτύου υπολογιστών από εισβολείς, είτε πρόκειται για στοχευμένους επιτιθέμενους είτε για καιροσκοπικό κακόβουλο λογισμικό.

Η ασφάλεια εφαρμογών επικεντρώνεται στη διατήρηση του λογισμικού και των συσκευών απαλλαγμένων από απειλές. Καθώς μια παραβιασμένη εφαρμογή μπορεί να επιτρέψει την πρόσβαση σε ευαίσθητα δεδομένα, η ασφάλεια πρέπει να λαμβάνεται υπόψη ήδη από το στάδιο του σχεδιασμού της.

Η ασφάλεια πληροφοριών προστατεύει την ακεραιότητα και την ιδιωτικότητα των δεδομένων, τόσο κατά την αποθήκευση όσο και κατά τη μεταφορά τους.

Η επιχειρησιακή ασφάλεια περιλαμβάνει τις διαδικασίες και τις αποφάσεις που αφορούν τον χειρισμό και την προστασία των δεδομένων. Περιλαμβάνει τα δικαιώματα πρόσβασης των χρηστών στο δίκτυο, καθώς και τα πρωτόκολλα για τον διαμοιρασμό και την αποθήκευση των πληροφοριών.

Η ανάκαμψη από καταστροφές και η επιχειρησιακή συνέχεια καθορίζουν τον τρόπο με τον οποίο ένας οργανισμός ανταποκρίνεται σε ένα περιστατικό ασφάλειας στον κυβερνοχώρο ή σε οποιοδήποτε άλλο συμβάν προκαλεί απώλεια λειτουργιών ή δεδομένων. Οι πολιτικές αποκατάστασης από καταστροφές υπαγορεύουν τον τρόπο με τον οποίο ο οργανισμός αποκαθιστά τις λειτουργίες και τις πληροφορίες του, ώστε να επανέλθει στην ίδια λειτουργική ικανότητα όπως πριν από το συμβάν. Η επιχειρησιακή συνέχεια είναι το σχέδιο στο οποίο καταφεύγει ο οργανισμός ενώ προσπαθεί να λειτουργήσει χωρίς συγκεκριμένους πόρους.

Η εκπαίδευση των τελικών χρηστών αντιμετωπίζει τον πιο απρόβλεπτο παράγοντα ασφάλειας στον κυβερνοχώρο: τους ανθρώπους. Οποιοσδήποτε μπορεί από ανθρώπινο λάθος να θέσει σε κίνδυνο ένα κατά τα άλλα ασφαλές σύστημα μη ακολουθώντας τις ορθές πρακτικές ασφαλείας. Η ενημέρωση των χρηστών για βασικές πρακτικές, όπως η αποφυγή ύποπτων συνημμένων email ή η μη χρήση άγνωστων μονάδων USB, είναι ζωτικής σημασίας για την ασφάλεια κάθε οργανισμού.

Οι συγγραφείς στο [2], πραγματοποίησαν μια ολοκληρωμένη βιβλιογραφική ανασκόπηση και προτείνουν τα εξής ορισμό για την κυβερνοασφάλεια. «Η κυβερνοασφάλεια είναι η οργάνωση και η συλλογή πόρων, διαδικασιών, και δομών που χρησιμοποιούνται για την προστασία του

κυβερνοχώρου και των συστημάτων που βασίζονται σε αυτόν από περιστατικά που παραβιάζουν τα προκαθορισμένα δικαιώματα ιδιοκτησίας». [3]

### **3.2 Ένας νέος ορισμός για την Κυβερνοασφάλεια**

Ένας σύγχρονος ορισμός για την ασφάλεια στον κυβερνοχώρο, ο οποίος ενσωματώνει βασικές έννοιες από τη διεθνή βιβλιογραφία και προκύπτει μέσα από μια διεπιστημονική θεώρηση, διατυπώνεται ως εξής: «Η κυβερνοασφάλεια είναι η οργάνωση και η συλλογή πόρων, διαδικασιών και δομών που χρησιμοποιούνται για την προστασία του κυβερνοχώρου, καθώς και των συστημάτων που υποστηρίζονται από αυτόν, από περιστατικά τα οποία προσκρούουν στα νομικά και πραγματικά δικαιώματα ιδιοκτησίας (de jure και de facto)».

Ο παραπάνω ορισμός αναλύεται στα ακόλουθα βασικά μέρη:

- «...η οργάνωση και η συλλογή πόρων, διαδικασιών και δομών»: Η διάσταση αυτή αναδεικνύει την εγγενή πολυπλοκότητα της κυβερνοασφάλειας, η οποία περιλαμβάνει αλληλεπιδράσεις τόσο μεταξύ των ανθρώπων όσο και μεταξύ συστημάτων, αλλά και τη διάδραση ανθρώπου-μηχανής. Αποφεύγοντας τον αυστηρό προσδιορισμό συγκεκριμένων τεχνολογιών ή μεθόδων, ο ορισμός παραμένει μη κανονιστικός (non-prescriptive) και ευέλικτος, αναγνωρίζοντας τη δυναμική φύση του πεδίου.
- «...που χρησιμοποιούνται για την προστασία του κυβερνοχώρου και των συστημάτων που υποστηρίζονται από αυτόν...»: Το σκέλος αυτό διευρύνει την έννοια της προστασίας πέρα από τα στενά όρια των παραδοσιακών υπολογιστικών υποδομών. Περιλαμβάνει τόσο τα κλασικά ψηφιακά περιβάλλοντα όσο και συστήματα που εκ πρώτης όψεως δεν θεωρούνται μέρος του κυβερνοχώρου, όπως τα βιομηχανικά συστήματα ελέγχου και τα κυβερνο-φυσικά συστήματα (Cyber-Physical Systems).
- «...από περιστατικά»: Η πτυχή αυτή αναγνωρίζει ότι η προστασία οφείλει να καλύπτει ολόκληρο το φάσμα των απειλών: από κακόβουλες και σκόπιμες επιθέσεις, μέχρι τυχαία σφάλματα, απρόβλεπτα συμβάντα και φυσικές καταστροφές. [2]

### **3.3 Ιστορική αναδρομή**

Η ιστορία της ασφάλειας στον κυβερνοχώρο ξεκίνησε τη δεκαετία του 1960, όταν οι υπολογιστές άρχισαν να χρησιμοποιούνται για την αποθήκευση και την ανταλλαγή ευαίσθητων πληροφοριών. Εκείνη την εποχή, οι υπολογιστές ήταν σπάνιοι, μεγάλοι και λειτουργούσαν κυρίως σε κυβερνήσεις ή μεγάλα ιδρύματα. Καθώς οι άνθρωποι άρχισαν να μοιράζονται πρόσβαση σε υπολογιστές και να συνδέουν συστήματα εξ αποστάσεως, οι ειδικοί συνειδητοποίησαν ότι οι πληροφορίες θα μπορούσαν να κλαπούν ή να διαρρεύσουν κατά λάθος. Οι πρώτες μελέτες επεσήμαναν ότι δεν υπήρχε τέλειος τρόπος για να γίνουν τα συστήματα υπολογιστών ασφαλή, ειδικά όταν μοιράζονταν ή χρησιμοποιούνταν σε ανοιχτά περιβάλλοντα. Αυτές οι ανησυχίες οδήγησαν σε προσπάθειες για τη δημιουργία κανόνων, προστατευτικών μέτρων και τεχνικών εγγυήσεων για τη διατήρηση της εμπιστευτικότητας πληροφοριών.

Στις δεκαετίες του 1970 και του 1980, οι ανησυχίες για την ασφάλεια των υπολογιστών αυξήθηκαν καθώς εμφανίστηκαν νέοι κίνδυνοι. Οι άνθρωποι έμαθαν πώς να παραβιάζουν συστήματα από απομακρυσμένες τοποθεσίες και περιστατικά έδειξαν ότι τα δίκτυα υπολογιστών μπορούσαν να χειραγωγηθούν ή να ελεγχθούν από εξωγενείς παράγοντες. Οι ψευδείς συναγερμοί, οι κλεμμένες πληροφορίες και η μη εξουσιοδοτημένη πρόσβαση κατέστησαν σαφές ότι ακόμη και τα στρατιωτικά συστήματα δεν ήταν ασφαλή. Οι κυβερνητικοί ηγέτες προσπάθησαν να αντιδράσουν θέτοντας κανόνες για την προστασία των συστημάτων υπολογιστών, αλλά προέκυψαν συζητήσεις σχετικά με το ποιος θα έπρεπε να είναι υπεύθυνος. Τελικά, ψηφίστηκαν νέοι νόμοι για να κατανεμηθούν οι ευθύνες μεταξύ των διαφόρων τμημάτων της κυβέρνησης και να διασφαλιστεί ότι τόσο τα πολιτικά όσο και τα στρατιωτικά συστήματα λάμβαναν την κατάλληλη προστασία.

Τη δεκαετία του 1990, το επίκεντρο μετατοπίστηκε στον τρόπο με τον οποίο τα συστήματα πληροφοριών και επικοινωνιών θα μπορούσαν να χρησιμοποιηθούν σε γεωπολιτικές συγκρούσεις. Οι στρατιωτικοί αναλυτές άρχισαν να διερευνούν τη δυνατότητα εξουδετέρωσης εχθρικών υποδομών μέσω κυβερνοεπιθέσεων αντί της χρήσης συμβατικών όπλων. Ταυτόχρονα, άλλες χώρες έγιναν καχύποπτες απέναντι στην τεχνολογία ξένης προέλευσης και φοβήθηκαν ότι θα μπορούσε να χρησιμοποιηθεί εναντίον τους μέσω ευπαθειών και κακόβουλου κώδικα. Οι εξελίξεις αυτές οδήγησαν σε μια αυξανόμενη συνειδητοποίηση ότι οι σύγχρονες συγκρούσεις μπορούν να περιλαμβάνουν νέα τεχνολογικά εργαλεία πέραν των συμβατικών. Ορισμένοι άρχισαν να προετοιμάζονται για αυτό το νέο είδος σύγκρουσης, όπου οι κυβερνοεπιχειρήσεις θα μπορούσαν να επιφέρουν καίρια πλήγματα εξ αποστάσεως, χωρίς την ανάγκη άμεσης στρατιωτικής αντιπαράθεσης.

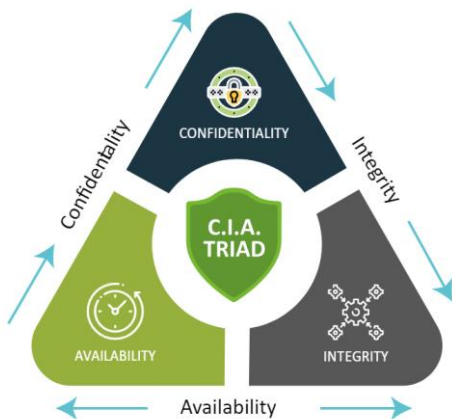
Στα τέλη της δεκαετίας του 1990, κατέστη πλέον σαφές το πόσο εκτεθειμένα ήταν συστήματα κρίσιμων εθνικών υποδομών σε διαδικτυακές απειλές. Ειδικές επιχειρησιακές ασκήσεις έδειξαν ότι οι επιθέσεις στα δίκτυα υπολογιστών θα μπορούσαν να επηρεάσουν τομείς ζωτικής σημασίας, όπως τα δίκτυα ηλεκτροδότησης, ύδρευσης, οι τραπεζικές συναλλαγές και οι τηλεπικοινωνίες. Η ολοένα αυξανόμενη συχνότητα πραγματικών περιστατικών παραβίασης απέδειξε ότι αυτές οι απειλές δεν ήταν απλώς θεωρητικές. Ως αποτέλεσμα, οι κυβερνήσεις προχώρησαν στη συγκρότηση εξειδικευμένων υπηρεσιών και τη δημιουργία ολοκληρωμένων εθνικών στρατηγικών κυβερνοάμυνας, εγκαινιάζοντας έτσι μια περίοδο σοβαρών και οργανωμένων πρωτοβουλιών για τη συστηματική θωράκιση του κυβερνοχώρου. [5]

### ***3.4 Βασικές έννοιες στην Κυβερνοασφάλεια***

Στον τομέα της κυβερνοασφάλειας, υπάρχουν ορισμένες θεμελιώδεις έννοιες και βασικές αρχές, που διασφαλίζουν την προστασία των ψηφιακών περιουσιακών στοιχείων. Αυτές οι βασικές αρχές είναι ζωτικής σημασίας για την κατανόηση του τρόπου με τον οποίο σχεδιάζονται και εφαρμόζονται τα συστήματα κυβερνοασφάλειας. Δύο από τα πιο σημαντικά πλαίσια που χρησιμοποιούνται είναι η τριάδα εμπιστευτικότητας, ακεραιότητας και διαθεσιμότητας (CIA) και το μοντέλο αυθεντικοποίησης, εξουσιοδότησης και καταγραφής (AAA). Αυτά τα πλαίσια παρέχουν μια ολοκληρωμένη προσέγγιση για τη διαφύλαξη των δεδομένων, των συστημάτων και

των χρηστών. Επίσης, η κυβερνοασφάλεια περιστρέφεται γύρω από την ανθρώπινη και επιχειρησιακή πλευρά της τεχνολογίας. Λαμβάνει υπόψη της τα 3P, δηλαδή τους ανθρώπους, το σχεδιασμό και την εκπαίδευση. [6]

### 3.4.1 CIA



Η τριάδα CIA (Εμπιστευτικότητα, Ακεραιότητα, Διαθεσιμότητα) είναι το ενοποιητικό χαρακτηριστικό για την ασφάλεια στον κυβερνοχώρο. Χρησιμοποιείται για την αξιολόγηση της ασφάλειας ενός οργανισμού, χρησιμοποιώντας τους τρεις βασικούς πυλώνες που σχετίζονται με την ασφάλεια. Την εμπιστευτικότητα, την ακεραιότητα και τη διαθεσιμότητα. Αυτά τα τρία χαρακτηριστικά έχουν συγκεκριμένες απαιτήσεις και λειτουργίες. [7]

#### 3.4.1.1 Εμπιστευτικότητα (Confidentiality - C)

Το γράμμα C αντιπροσωπεύει την εμπιστευτικότητα. Η ασφάλεια στον κυβερνοχώρο απαιτεί προστασία της ιδιωτικότητας των δεδομένων και των πληροφοριών. Η πρόσβαση σε ευαίσθητα στοιχεία, όπως ονόματα χρηστών, διαπιστευτήρια αυθεντικοποίησης (κωδικοί πρόσβασης), ιατρικά αρχεία ή επιχειρησιακά δεδομένα, πρέπει να επιτρέπεται αυστηρά και μόνο σε εξουσιοδοτημένους χρήστες, διεργασίες και συσκευές. Η αρχή αυτή στοχεύει στον αποκλεισμό της μη εξουσιοδοτημένης ανάγνωσης ή αποκάλυψης πληροφοριών, αποτρέποντας περιστατικά διαρροής που θα μπορούσαν να θέσουν σε κίνδυνο την ασφάλεια του εκάστοτε οργανισμού.

#### 3.4.1.2 Ακεραιότητα (Integrity - I)

Το γράμμα I στο μοντέλο CIA αντιστοιχεί στην ακεραιότητα. Η ασφάλεια στον κυβερνοχώρο απαιτεί να αισθανόμαστε ασφαλείς ότι τα δεδομένα που μεταδίδονται, υποβάλλονται σε επεξεργασία και αποθηκεύονται δεν έχουν αλλάξει από την αρχική τους μορφή είτε τυχαία είτε κακόβουλα. Για παράδειγμα, ακόμη και η αλλοίωση ενός μοναδικού bit κατά τη μετάδοση μπορεί να μεταβάλει πλήρως το νόημα ενός μηνύματος ή να καταστήσει το αρχείο μη αναγνώσιμο και κατεστραμμένο.

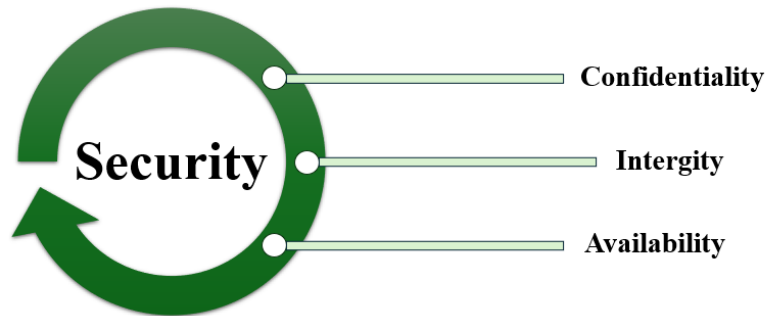
#### 3.4.1.3 Διαθεσιμότητα (Availability - A)

Το τελευταίο γράμμα A εκπροσωπεί τη διαθεσιμότητα. Η διαθεσιμότητα εγγυάται ότι οι πληροφοριακοί πόροι, τα συστήματα, το υλικό και οι εφαρμογές παραμένουν προσβάσιμα και λειτουργικά όποτε τα χρειάζονται οι εξουσιοδοτημένοι χρήστες για την εκτέλεση των καθηκόντων τους. Παράλληλα, περιλαμβάνει μηχανισμούς εξισορρόπησης φορτίου (load balancing), εφεδρείας και ανοχής σφαλμάτων (fault tolerance), ώστε τα συστήματα να συνεχίζουν να λειτουργούν απρόσκοπτα ή να ανακάμπτουν άμεσα σε περιπτώσεις κυβερνοεπιθέσεων, αστοχιών υλικού ή φυσικών καταστροφών. [7]

**Σχήμα 2: Η τριάδα C.I.A. δείχνει τα βασικά στοιχεία της ασφάλειας: Εμπιστευτικότητα, Ακεραιότητα και Διαθεσιμότητα**

### 3.4.2 AAA

Οι στόχοι της ασφάλειας στον κυβερνοχώρο υλοποιούνται με τη χρήση του μοντέλου AAA ή τριπλού Α. [7]



#### 3.4.2.1 Αυθεντικοποίηση (Authentication)

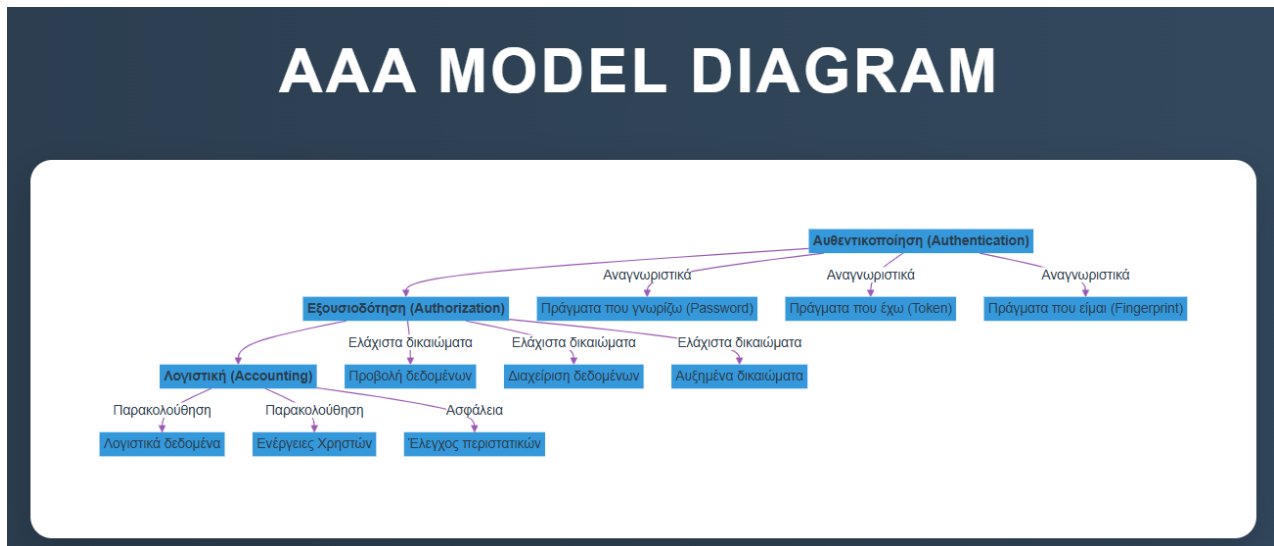
Το πρώτο Α αναφέρεται στην αυθεντικοποίηση, η οποία αποτελεί τη διαδικασία επαλήθευσης της ταυτότητας ενός χρήστη. Όταν κάποιος ισχυρίζεται ότι είναι κάποιος, αυτό ονομάζεται ταυτοποίηση, αλλά όταν το αποδεικνύει, αυτό είναι πιστοποίηση. Ενώ η ταυτοποίηση (Identification) αφορά απλώς τη δήλωση μιας ταυτότητας (π.χ. εισαγωγή ονόματος χρήστη), η αυθεντικοποίηση απαιτεί την έμπρακτη απόδειξή της. Η διαδικασία αυτή βασίζεται συνήθως σε έναν ή περισσότερους παράγοντες, όπως ένας κωδικός πρόσβασης, μια έξυπνη κάρτα ή βιομετρικά χαρακτηριστικά (π.χ. δαχτυλικό αποτύπωμα). Ο συνδυασμός δύο ή περισσότερων παραγόντων διαφορετικής κατηγορίας αποτελεί τον έλεγχο ταυτότητας πολλαπλών παραγόντων (Multi-Factor Authentication - MFA), ο οποίος μειώνει δραστικά την πιθανότητα μη εξουσιοδοτημένης πρόσβασης και πλαστοπροσωπίας.

#### 3.4.2.2 Εξουσιοδότηση (Authorization)

Το δεύτερο Α στο μοντέλο AAA αντιστοιχεί στην εξουσιοδότηση. Αυτή αφορά την παροχή κατάλληλου επιπέδου πρόσβασης για κάθε χρήστη με βάση τα διαπιστευτήριά του. Η έννοια αυτή συνδέεται άμεσα με την αρχή των ελάχιστων προνομίων (Principle of Least Privilege), σύμφωνα με την οποία σε κάθε χρήστη, συσκευή ή λογισμικό παραχωρούνται μόνο τα δικαιώματα που απαιτούνται για τη διεκπεραίωση του εκάστοτε έργου. Οποιαδήποτε εξουσιοδότηση πέραν των απαραίτητων αυξάνει τον κίνδυνο τυχαίας ή κακόβουλης παραβίασης της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας του συστήματος.

#### 3.4.2.3 Καταγραφή (Accounting)

Το τελευταίο Α του μοντέλου είναι η καταγραφή, δηλαδή η παρακολούθηση του τι κάνουν οι χρήστες ενώ είναι συνδεδεμένοι σε ένα σύστημα. Η παρακολούθηση, καταγραφή και αρχειοθέτηση της δραστηριότητας των χρηστών κατά τη διάρκεια της σύνδεσής τους. Η τήρηση αρχείων καταγραφής (logs) αποτελεί διεργασία υψηλής σημασίας. Από πλευράς ψηφιακής εγκληματολογίας (digital forensics), η ανίχνευση των γεγονότων που οδήγησαν σε ένα περιστατικό κυβερνοασφάλειας μπορεί να αποδειχθεί πολύ πολύτιμη σε μια έρευνα.



### Σχήμα 3: Το μοντέλο AAA (Authentication, Authorization, Accounting)

### 3.4.3 3Ps

#### 3.4.3.1 Οι άνθρωποι (People)

Η βασική ιδέα σχετικά με την εμπλοκή των ανθρώπων στην κυβερνοασφάλεια είναι η προετοιμασία του ανθρώπινου δυναμικού κατά των απειλών στον κυβερνοχώρο και η προστασία των ευαίσθητων δεδομένων. Δεν είναι ασυνήθιστο για τους οργανισμούς να πέφτουν θύματα ηλεκτρονικού ψαρέματος (phishing). Μέσα από προγράμματα κατάρτισης, σεμινάρια, ενημερωτικά δελτία και οδηγίες βέλτιστων πρακτικών (best practices), δημιουργείται μια εταιρική κουλτούρα ασφάλειας που λειτουργεί ως πρώτη γραμμή άμυνας σε απόπειρες παραβίασης.

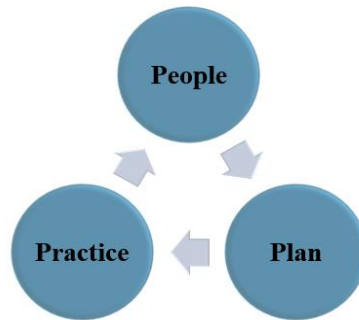
#### 3.4.3.2 Σχεδιασμός (Plan)

Προκειμένου να αντιμετωπιστεί ένα περιστατικό στον κυβερνοχώρο, θα πρέπει να υπάρχει ένα ολοκληρωμένο σχέδιο δράσης που να δίνει σαφείς απαντήσεις σε τεχνικά και επιχειρησιακά ερωτήματα. Η χρήση ενός σχεδίου θα μπορούσε να βοηθήσει στην ανάκτηση δεδομένων και την αποκατάσταση συγκεκριμένων κρίσιμων λειτουργιών. Μπορεί να καθοδηγήσει, να καθιερώσει πρωτόκολλα και να εξασφαλίσει συντονισμένη αντίδραση σε περιόδους κρίσης.

#### 3.4.3.3 Πρακτική εξάσκηση (Practice)

Η ύπαρξη ενός σχεδίου δεν επαρκεί εάν αυτό δεν έχει δοκιμαστεί στην πράξη. Η πρακτική εξάσκηση αφορά τη συστηματική υλοποίηση επιχειρησιακών ασκήσεων και προσομοιώσεων

ασφαλείας, ώστε να εντοπιστούν τυχόν αδυναμίες ή κενά στις διαδικασίες. Οι ασκήσεις αυτές περιλαμβάνουν υποθετικά σενάρια επιθέσεων (tabletop exercises), αξιολόγηση ενδείξεων παραβίασης, διαδικασίες λήψης αποφάσεων υπό πίεση και την πρακτική εφαρμογή πρωτοκόλλων αντιμετώπισης για συναφή περιστατικά.



**Σχήμα 4: Το μοντέλο 3P (People, Plan, Practice)**

### **3.5 Στόχοι Κυβερνοασφάλειας**

Κεντρικός στόχος της κυβερνοασφάλειας είναι η διατήρηση ενός ασφαλούς και ανθεκτικού ψηφιακού περιβάλλοντος, το οποίο θα μπορεί να εγγυηθεί την εύρυθμη λειτουργία κρίσιμων υποδομών. Για την ανάκαμψη από περιστατικά και επιθέσεις στον κυβερνοχώρο, θα πρέπει να υπάρχει η κατάλληλη αντίδραση, επίλυση και ανάκαμψη. Για το λόγο αυτό είναι ανάγκη να υπάρχει το κατάλληλο θεσμικό και νομικό πλαίσιο.

Ακολουθούν ορισμένοι στόχοι που οδηγούν στην πρόληψη από απειλές στον κυβερνοχώρο και την προστασία από κυβερνοεπιθέσεις. [6]

#### **3.5.1 Πρόληψη Απειλών (Threat Prevention)**

Περιλαμβάνει τον σχεδιασμό, την υλοποίηση και την επιβολή αυστηρών πρωτοκόλλων ελέγχου δικτύου. Απαιτείται η συνεχής παρακολούθηση ενδείξεων παραβίασης (indicators of compromise - IoCs), η θέσπιση διαδικασιών αναφοράς περιστατικών, η υιοθέτηση βέλτιστων πρακτικών και η χρήση σύγχρονων εργαλείων εντοπισμού κακόβουλου λογισμικού.

#### **3.5.2 Αναγνώριση απειλών και Σκλήρυνση Συστήματος (System Hardening)**

Ένας από τους πρωταρχικούς στόχους της ασφάλειας στον κυβερνοχώρο είναι η αναγνώριση απειλών προκειμένου να θωρακιστεί το σύστημα. Η διαδικασία αυτή διασφαλίζει την αξιολόγηση του κινδύνου και τη λήψη μέτρων ασφαλείας. Σκοπός της σκλήρυνσης του συστήματος είναι να μετριάσει ορισμένους κινδύνους που σχετίζονται με την ασφάλεια. Ορισμένες φορές χρησιμοποιείται μια προηγμένη προσέγγιση σκλήρυνσης του συστήματος, η οποία περιλαμβάνει την αναδιαμόρφωση (formatting) σκληρών δίσκων και την εγκατάσταση μόνο εγκεκριμένων προγραμμάτων στο σύστημα (whitelisting).

### **3.5.3 Υιοθέτηση επιχειρησιακών και τεχνολογικών καινοτομιών**

Η εισαγωγή δυναμικών προσεγγίσεων προς τη διαχείριση των κινδύνων στον κυβερνοχώρο προστατεύει τις υποδομές του κυβερνοχώρου από συγκεκριμένες κυβερνοεπιθέσεις.

### **3.5.4 Ετοιμότητα και Σχεδιασμός Έκτακτης Ανάγκης (Contingency Planning)**

Η ιδέα του σχεδιασμού έκτακτης ανάγκης είναι ουσιαστικά η ετοιμότητα για κυβερνοεπιθέσεις. Μπορεί να περιλαμβάνει πολιτικές, βέλτιστες πρακτικές, διαδικασίες και σχέδια ανάκαμψης.

### **3.5.5 Διαμοιρασμός Πληροφοριών Απειλών (Threat Intelligence Sharing)**

Ο έγκαιρος και αξιόπιστος διαμοιρασμός πληροφοριών σχετικά με τρωτά σημεία, τρέχουσες επιθέσεις και ειδοποιήσεις ασφαλείας (security advisories), τόσο εντός του οργανισμού όσο και σε συνεργασία με αρμόδιους φορείς, ενισχύει την άμυνα εναντίον επιθέσεων.

### **3.5.6 Εξειδικευμένη εκπαίδευση προσωπικού**

Το εργατικό δυναμικό πρέπει να είναι εφοδιασμένο με εξειδικευμένη εκπαίδευση ασφαλείας. Οι πληροφορίες και οι υπηρεσίες πρέπει να παρέχονται επίσης στους κοινούς εταίρους ώστε το εκάστοτε προσωπικό να είναι προετοιμασμένο κατά τη διάρκεια περιστατικών στον κυβερνοχώρο.

### **3.5.7 Ενίσχυση της Ανοχής σε Σφάλματα (Fault Tolerance)**

Η ανθεκτικότητα των συστημάτων αξιολογείται μέσω τακτικών ελέγχων τρωτότητας (vulnerability assessments) και προσομοιώσεων εισβολής, ώστε οι υποδομές να παραμένουν λειτουργικές ακόμη και κατά τη διάρκεια κυβερνοεπιθέσεων.

### **3.5.8 Μείωση ευπαθειών (Vulnerability Management)**

Συγκεκριμένες πρακτικές ασφαλείας βοηθούν στη δραστική μείωση. Η τακτική ενημέρωση (patching), η χρήση τειχών προστασίας (firewalls) και η χρήση ισχυρών κωδικών πρόσβασης μπορούν να αποτρέψουν την κακόβουλη πρόσβαση σε ζωτικά πληροφοριακά συστήματα.

### **3.5.9 Βελτίωση της Ευχρηστίας (Usability)**

Ο όρος ευχρηστία ορίζεται ως ο βαθμός στον οποίο μια υπηρεσία είναι εύκολη στη χρήση. Τα μέτρα ασφαλείας ενός οργανισμού οφείλουν να ενσωματώνονται στα πληροφοριακά συστήματα με τρόπο ο οποίος δε δυσχεραίνει την παραγωγικότητα των χρηστών, αποτρέποντας έτσι την παράκαμψη κρίσιμων ελέγχων ασφαλείας λόγω πολυπλοκότητας.

### **3.5.10 Ισχυρή Αυθεντικοποίηση (Robust Authentication)**

Η επαλήθευση της ταυτότητας ενός χρήστη ή μιας διαδικασίας είναι μια πολύ βασική διαδικασία της κυβερνοασφάλειας. Ανάλογα με το είδος συσκευής, μπορεί να αναπτυχθεί έλεγχος ταυτότητας ενός παράγοντα ή πολλαπλών παραγόντων (MFA) για ενισχυμένη ασφάλεια.

### **3.5.11 Αυτοματοποίηση Διαδικασιών Ασφαλείας (Security Automation)**

Η αυτοματοποίηση διεργασιών οδηγεί σε καλύτερη αποτελεσματικότητα, πρόβλεψη συμπεριφοράς και ταχύτερη εκτέλεση. Η κατάλληλη εφαρμογή της αυτοματοποίησης έχει ως αποτέλεσμα την πρόληψη κυβερνοεπιθέσεων. Ο αυτοματισμός μπορεί να συσχετίσει δεδομένα, να προβεί σε καθοριστικές προληπτικές κινήσεις πιο γρήγορα από την εξάπλωση μιας επίθεσης, καθώς και να εντοπίσει παραβιάσεις δικτύων.

### **3.5.12 Διασφάλιση Διαλειτουργικότητας (Interoperability)**

Η ικανότητα των επιμέρους μηχανισμών και εργαλείων ασφαλείας να επικοινωνούν απροβλημάτιστα μεταξύ τους, καθώς και με εξωτερικά συστήματα, εξασφαλίζει πλήρη εικόνα της ασφάλειας ενός οργανισμού.

### **3.5.13 Καταγραφή και Ανάλυση Δυσμενών Περιστατικών (Incident Logging & Analysis)**

Είναι σημαντικό να αναλύονται δυσμενή γεγονότα στον κυβερνοχώρο, ώστε να εξευρίσκονται λύσεις αποτροπής παραβίασης συστημάτων. Οι πληροφορίες σχετικά με τα αίτια, την έκταση και τον αντίκτυπο αυτών των δυσμενών γεγονότων καταγράφονται με σκοπό τη μελλοντική χρήση για τη θωράκιση και αναβάθμιση πληροφοριακών συστημάτων.

### **3.5.14 Εφαρμογή Πολυεπίπεδων Μέτρων Ασφαλείας (Defense-in-Depth)**

Η συνδυαστική χρήση τεχνικών ελέγχων, όπως η τμηματοποίηση δικτύου (network segmentation), η εγκατάσταση και ρύθμιση τειχών προστασίας, η χρήση ασφαλών εικονικών ιδιωτικών δικτύων (VPNs) και οι αυστηροί έλεγχοι πρόσβασης, δημιουργεί πολλαπλά επίπεδα άμυνας απέναντι σε μεγάλο εύρος κακόβουλων ενεργειών.

### 3.6 Υποδομή κυβερνοασφάλειας και αρχιτεκτονική διαδικτύου (NIST)

Δεδομένου ότι οι κυβερνοεπιθέσεις γίνονται όλο και πιο εξελιγμένες, υπάρχει ανάγκη να εισαχθούν τυποποιημένες πρακτικές για τη διασφάλιση της ασφάλειας. Το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (NIST) ενσωματώνει ορισμένες πολιτικές, πρότυπα, κατευθυντήριες γραμμές και βέλτιστες πρακτικές για την αντιμετώπιση των θεμάτων κυβερνοασφάλειας. Αυτό το πλαίσιο χωρίζεται σε πυρήνα πλαισίου, επίπεδα εφαρμογής και προφίλ. [1]



Σχήμα 5: Πλαίσιο κυβερνοασφάλειας NIST [1]

#### 3.6.1 Πυρήνας πλαισίου

Ο πυρήνας του πλαισίου περιλαμβάνει ορισμένα συστήματα που οδηγούν σε συγκεκριμένα αποτελέσματα. Μπορεί να έχει τη μορφή λειτουργιών, κατηγοριών, υποκατηγοριών και πληροφοριακών αναφορών.

##### 3.6.1.1 Λειτουργίες

Για την ασφάλεια των συστημάτων και την αντιμετώπιση των επιθέσεων, οι πέντε βασικές λειτουργίες είναι ο εντοπισμός, η προστασία, η ανίχνευση, η ανταπόκριση και η ανάκτηση, τις οποίες θα συζητήσουμε στη συνέχεια αυτής της ενότητας.

##### 3.6.1.2 Κατηγορίες

Οι διάφορες λειτουργίες έχουν αντίστοιχες κατηγορίες για τον προσδιορισμό διαφορετικών λειτουργιών και δραστηριοτήτων. Για παράδειγμα, για την προστασία, μπορεί κανείς να κάνει χρήση του ελέγχου πρόσβασης, των ενημερώσεων λογισμικού και των προγραμμάτων κατά του κακόβουλου λογισμικού.

### **3.6.1.3 Υποκατηγορίες**

Οι κατηγορίες με συγκεκριμένους στόχους ονομάζονται υποκατηγορίες. Για παράδειγμα, η διαδικασία ενημέρωσης λογισμικού θα μπορούσε να έχει συγκεκριμένες λειτουργίες όπως η σωστή διαμόρφωση του ή η χειροκίνητη ενημέρωση των μηχανημάτων.

### **3.6.1.4 Ενημερωτικές αναφορές**

Οι ενημερωτικές αναφορές περιλαμβάνουν πολιτικές, πρότυπα και κατευθυντήριες γραμμές. Για παράδειγμα, ορισμένα βήματα που απαιτούν χειροκίνητη ενημέρωση του συστήματος Windows.

### **3.6.2 Επίπεδα υλοποίησης**

Ακολουθούν οι τέσσερις βαθμίδες πληροφοριών

- Βαθμίδα 1 ή Μερική εφαρμογή χειρίζεται τους οργανωτικούς κινδύνους με ασυνέπεια λόγω της ad hoc υποδομής κυβερνοασφάλειας.
- Η υλοποίηση επιπέδου 2 ασχολείται με τους κινδύνους, τα σχέδια και τους πόρους για την προστασία των υποδομών στον κυβερνοχώρο σε βαθύτερο επίπεδο από τη μερική υλοποίηση.
- Η βαθμίδα 3 ή Επαναλαμβανόμενη εφαρμογή μπορεί να τείνει επανειλημμένα σε κρίσεις στον κυβερνοχώρο. Οι πολιτικές μπορούν να εφαρμοστούν στο ίδιο επίπεδο και η ευαισθητοποίηση σε θέματα κυβερνοασφάλειας μπορεί να ελαχιστοποιήσει τους κινδύνους που σχετίζονται με τον κυβερνοχώρο.
- Η βαθμίδα 4 ή προσαρμοστική εφαρμογή είναι υπεύθυνη για τον εντοπισμό απειλών και την πρόβλεψη ζητημάτων σε σχέση με την υποδομή ασφάλειας.

### **3.6.3 Προφίλ**

Ένα πλαίσιο κυβερνοασφάλειας έχει ορισμένους συγκεκριμένους στόχους. Τα προφίλ συνοψίζουν την κατάσταση της κυβερνοασφάλειας ενός οργανισμού. Τα πολλαπλά προφίλ σε ένα πλαίσιο κυβερνοασφάλειας εξασφαλίζουν τον εντοπισμό αρκετών αδύναμων σημείων που αποτελούν μέρος της εφαρμογής της κυβερνοασφάλειας. Μπορούν επίσης να υποστηρίξουν τη σύνδεση μεταξύ των λειτουργιών, των κατηγοριών και των υποκατηγοριών με τους πόρους και την ανοχή κινδύνου των οργανισμών.

Ας ρίξουμε τώρα και μια ματιά στις λειτουργίες του πυρήνα του πλαισίου. Αυτές είναι οι εξής:

### **3.6.4 Αναγνώριση**

Ο προσδιορισμός αναφέρεται στην ανάπτυξη της κατανόησης προκειμένου να διαχειριστεί κινδύνων κυβερνοασφάλειας για τα συστήματα, τα περιουσιακά στοιχεία, τα δεδομένα και τις δυνατότητες. Η λειτουργία εντοπισμού έχει τις ακόλουθες κατηγορίες: Διαχείριση περιουσιακών στοιχείων: Προσδιορισμός δεδομένων, συσκευών και συστημάτων που βοηθούν έναν οργανισμό για επιχειρηματικούς σκοπούς.

#### *3.6.4.1 Επιχειρηματικό περιβάλλον*

Ιεραρχεί τους στόχους, την αποστολή και τις δραστηριότητες ενός οργανισμού. Οι πληροφορίες διαδραματίζουν βασικό ρόλο στη λήψη αποφάσεων για την ασφάλεια στον κυβερνοχώρο ρόλων, των αρμοδιοτήτων και της διαχείρισης κινδύνων.

#### *3.6.4.2 Διακυβέρνηση*

Πρόκειται για τις υποχρεωτικές κατευθυντήριες γραμμές που είναι απαραίτητες για τη διαχείριση του περιβάλλοντος ενός οργανισμού και τον εντοπισμό των κινδύνων κυβερνοασφάλειας.

#### *3.6.4.3 Αξιολόγηση κινδύνων*

Ο οργανισμός πρέπει να είναι σε θέση να συμπεράνει τους κινδύνους που σχετίζονται με την λειτουργίες και την υποδομή του.

#### *3.6.4.4 Στρατηγική διαχείρισης κινδύνων*

Διάφοροι περιορισμοί, ανοχές και παραδοχές επηρεάζουν τις αποφάσεις σχετικά με τους κινδύνους ασφαλείας.

### **3.6.5 Προστασία**

Οι οργανισμοί πρέπει να είναι εξοπλισμένοι με διάφορους μηχανισμούς άμυνας κατά τη διάρκεια μιας συμβάντος κυβερνοασφάλειας. Η εξασφάλιση ελεγχόμενης πρόσβασης, η ευαισθητοποίηση, η εκπαίδευση, η σωστή δικτύωση διαμόρφωση, η ανάπτυξη προστατευτικών εργαλείων και τεχνολογιών, κ.λπ. της υποδομής και των συστημάτων κυβερνοασφάλειας. Ακολουθούν μερικές από τις κατηγορίες για προστασίας:

#### *3.6.5.1 Έλεγχος πρόσβασης*

Οι εξουσιοδοτημένοι χρήστες μπορούν να έχουν πρόσβαση σε συσκευές, διαδικασίες, δραστηριότητες και συναλλαγές.

#### *3.6.5.2 Ευαισθητοποίηση και εκπαίδευση*

Η ευαισθητοποίηση και η κατάρτιση των εργαζομένων του οργανισμού μαζί με τις πολιτικές και τις κατευθυντήριες γραμμές μπορούν να διασφαλίσουν την προστασία από τα ζητήματα του κυβερνοχώρου.

#### *3.6.5.3 Ασφάλεια δεδομένων*

Η εμπιστευτικότητα, η ακεραιότητα και η διαθεσιμότητα των δεδομένων πρέπει να διατηρούνται καθ' όλη τη διάρκεια ζωής τους.

#### *3.6.5.4 Προστασία πληροφοριών*

Οι πολιτικές, οι διαδικασίες, οι κατευθυντήριες γραμμές και οι διεργασίες μπορούν να προστατεύσουν πληροφορίες.

#### *3.6.5.5 Συντήρηση*

Η συντήρηση του συστήματος θα μπορούσε να πραγματοποιηθεί με τη χρήση ορισμένων πολιτικών και διαδικασιών.

#### *3.6.5.6 Προστατευτική τεχνολογία*

Οι πολιτικές, οι διαδικασίες και οι συμφωνίες, μαζί με τις τεχνικές λύσεις, οδηγούν στην ασφάλεια του συστήματος.

### **3.6.6 Εντοπισμός**

Για τον εντοπισμό συμβάντων κυβερνοασφάλειας, οι οργανισμοί πρέπει να λαμβάνουν τα κατάλληλα μέτρα. Η παρακολούθηση και ο εντοπισμός απειλών είναι μερικοί τρόποι για τον εντοπισμό ζητημάτων ασφάλειας στα συστήματα. Οι ακόλουθες κατηγορίες αφορούν την ανίχνευση:

#### *3.6.6.1 Γεγονότα και ανωμαλίες*

Η έγκαιρη ανάλυση των συμβάντων μπορεί να διασφαλίσει τον εντοπισμό ανώμαλων δραστηριοτήτων.

#### *3.6.6.2 Παρακολούθηση*

Η παρακολούθηση του συστήματος σε διάφορα χρονικά διαστήματα μπορεί να βοηθήσει στον εντοπισμό ζητημάτων και επίσης να επαληθεύσει την αποτελεσματικότητα των προτεινόμενων λύσεων.

#### *3.6.6.3 Άλλες τεχνικές ανίχνευσης*

Μπορούν να δοκιμαστούν διάφορες διεργασίες και διαδικασίες για να ανίχνευση κακόβουλης συμπεριφοράς σε ένα σύστημα.

### **3.6.7 Ανταπόκριση**

Σε περίπτωση που συμβούν περιστατικά στον κυβερνοχώρο, οι οργανισμοί πρέπει να είναι σε θέση να περιορίσουν τις επιπτώσεις. Η εκπόνηση σχεδίων αντιμετώπισης, η καθιέρωση γραμμών επικοινωνίας, και η συλλογή και ανάλυση πληροφοριών αποτελούν διάφορες στρατηγικές αντιμετώπισης. Εξετάζουμε μερικές κατηγορίες.

#### *3.6.7.1 Σχεδιασμός αντιμετώπισης*

Εκτέλεση διεργασιών και διαδικασιών για τον εντοπισμό της ασφάλειας στον κυβερνοχώρο συμβάντων.

#### *3.6.7.2 Επικοινωνία*

Οι δραστηριότητες πληροφόρησης και αντιμετώπισης σε ολόκληρο τον οργανισμό μπορούν να συντονίζονται εάν επικοινωνούνται σωστά.

### 3.6.7.3 Ανάλυση

Η ανάλυση των δεδομένων διασφαλίζει τη σωστή ανταπόκριση και ανάκαμψη.

### 3.6.7.4 Μετριάσμος

Αρκετά συμβάντα που έχουν τη δυνατότητα να προκαλέσουν ζητήματα ασφάλειας σε στο σύστημα πρέπει να μετριάσθούν. Τα περιστατικά στον κυβερνοχώρο και τα επακόλουθά τους πρέπει να μετριάσθούν.

### 3.6.7.5 Βελτίωση

Ανάλυση των τρεχουσών και προηγούμενων δραστηριοτήτων ανίχνευσης-απόκρισης μπορεί να βελτιώσει σημαντικά τις μελλοντικές δραστηριότητες αντιμετώπισης.

## 3.6.8 Ανάκτηση

Τα συμβάντα κυβερνοασφάλειας έχουν τη δυνατότητα να επηρεάσουν τις υπηρεσίες και τις ικανότητες των συστημάτων. Ως εκ τούτου, οι οργανισμοί πρέπει να εισάγουν αποτελεσματικές στρατηγικές ή ανάκαμψης για την αποκατάσταση του συστήματος. Ακολουθούν ορισμένες κατηγορίες που κάνουν το ίδιο:

### 3.6.8.1 Σχεδιασμός αποκατάστασης

Η εκτέλεση των διαδικασιών και διαδικασιών αποκατάστασης και η συντήρηση οδηγεί στην έγκαιρη αποκατάσταση των συστημάτων που επηρεάζονται από κακόβουλα γεγονότα.

### 3.6.8.2 Βελτιώσεις

Η αναβάθμιση των υφιστάμενων διαδικασιών και διαδικασιών οδηγεί σε βελτίωση των μελλοντικών συστημάτων για αποκατάσταση.

### 3.6.8.3 Επικοινωνίες

Η επικοινωνία εντός ενός οργανισμού εξασφαλίζει το συντονισμό των δραστηριοτήτων αποκατάστασης. [1]

| Λειτουργία<br>(Function) | Κατηγορίες (Categories)                                                                                                                    | Παραδείγματα<br>Υποκατηγοριών & Δράσεων                                                    |
|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Αναγνώριση<br>(Identify) | Διαχείριση περιουσιακών στοιχείων,<br>Επιχειρηματικό περιβάλλον, Διακυβέρνηση,<br>Αξιολόγηση κινδύνων, Στρατηγική<br>διαχείρισης κινδύνων  | Εντοπισμός συστημάτων και<br>δεδομένων, ανάλυση ρόλων,<br>εκτίμηση ρίσκου                  |
| Προστασία<br>(Protect)   | Έλεγχος πρόσβασης, Ευαισθητοποίηση και<br>εκπαίδευση, Ασφάλεια δεδομένων,<br>Προστασία πληροφοριών, Συντήρηση,<br>Προστατευτική τεχνολογία | Πολιτικές πρόσβασης,<br>εκπαίδευση προσωπικού,<br>κρυπτογράφηση, ενημερώσεις<br>λογισμικού |

|                          |                                                                         |                                                                                        |
|--------------------------|-------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Εντοπισμός<br>(Detect)   | Γεγονότα και ανωμαλίες, Παρακολούθηση,<br>Άλλες τεχνικές ανίχνευσης     | Ανάλυση συμβάντων, συνεχής<br>παρακολούθηση συστημάτων,<br>anomaly detection           |
| Ανταπόκριση<br>(Respond) | Σχεδιασμός αντιμετώπισης, Επικοινωνία,<br>Ανάλυση, Μετριάσμός, Βελτίωση | Δημιουργία σχεδίων δράσης,<br>ανάλυση επιθέσεων, απομόνωση<br>απειλών, lessons learned |
| Ανάκτηση<br>(Recover)    | Σχεδιασμός αποκατάστασης, Βελτιώσεις,<br>Επικοινωνίες                   | Επαναφορά συστημάτων,<br>αναβάθμιση διαδικασιών,<br>συντονισμός ανάκαμψης              |

**Πίνακας 3: Πλαίσιο κυβερνοασφάλειας NIST**

### ***3.7 Απειλές και τρωτά σημεία για την ασφάλεια στον κυβερνοχώρο***

Η κυβερνοασφάλεια βασικά, μπορεί να οριστεί ως η ικανότητα άμυνας και ανάκαμψης από κυβερνοεπιθέσεις. Σύμφωνα με το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (NIST), η κυβερνοασφάλεια είναι η ικανότητα προστασίας ή υπεράσπισης της χρήσης του κυβερνοχώρου από κυβερνοεπιθέσεις. Ολόκληρος ο κυβερνοχώρος αποτελείται από διάφορα αλληλοεξαρτώμενα δίκτυα της υποδομής των πληροφοριακών συστημάτων που μπορεί να είναι το διαδίκτυο, το δίκτυο τηλεπικοινωνιών, ο υπολογιστής συστήματα, ενσωματωμένα συστήματα ή ελεγκτές. Έτσι, η κυβερνοασφάλεια αφορά κρίσιμες υποδομές, την ασφάλεια δικτύων, την ασφάλεια του υπολογιστικού νέφους, την ασφάλεια εφαρμογών, το διαδίκτυο των πραγμάτων και διάφορους άλλους τομείς όπου η ανάγκη διασφάλισης της ασφάλειας είναι υψίστης σημασίας. [6]

#### ***3.7.1 Κρίσιμες υποδομές***

Η ασφάλεια στις κρίσιμες υποδομές αφορά τα κυβερνοφυσικά συστήματα και πραγματικές εφαρμογές. Βιομηχανίες όπως η αυτοματοποίηση, η αεροπορία, η υγειονομική περίθαλψη, φωτεινοί σηματοδότες, δίκτυα ηλεκτρικής ενέργειας κ.λπ. είναι επιρρεπείς σε κυβερνοεπιθέσεις όπως υποκλοπές, επιθέσεις με παραβιασμένο κλειδί, επιθέσεις man-in-the-middle και επιθέσεις άρνησης παροχής υπηρεσιών (DoS).

##### ***3.7.1.1 Υποκλοπή***

Η επίθεση αυτή συμβαίνει σε μη ασφαλές δίκτυο και όταν οι πληροφορίες που μεταδίδονται δεν είναι κρυπτογραφημένες. Έτσι, ένας ξένος μπορεί να κρυφακούσει σε ένα εύαλωτο δίκτυο και να αποκτήσει εύκολα πρόσβαση σε ευαίσθητα δεδομένα. [6]

### 3.7.1.2 *Επιθέσεις με παραβιασμένο κλειδί*

Σε αυτό το τύπο επίθεσης, ένας επιτιθέμενος Ε που έχει πρόσβαση στο μυστικό κλειδί του διακομιστή S υποκλέπτει ένα μήνυμα σύνδεσης από έναν χρήστη  $p_i$  μέσω ενός δημόσιου καναλιού και διαβάζει το περιεχόμενό του. Στη συνέχεια, ο Ε ανακτά τις πληροφορίες ταυτότητας του χρήστη αποκρυπτογραφώντας ένα τμήμα του μηνύματος και χρησιμοποιεί την αποκτηθείσα ταυτότητα για να αποκτήσει πρόσβαση σε λεπτομέρειες που αφορούν συγκεκριμένο χρήστη από έναν κλεμμένο πίνακα εγγραφής της βάσης δεδομένων του διακομιστή.

Χρησιμοποιώντας αυτές τις λεπτομέρειες, ο Ε δημιουργεί νέες τιμές ελέγχου ταυτότητας δημιουργώντας νέα δεδομένα και παραπέμποντας σε γνωστές τιμές από τον διακομιστή και την ταυτότητα του χρήστη. Ο Ε κατασκευάζει ένα νέο μήνυμα σύνδεσης μιμούμενος τη μορφή και τα στοιχεία που αναμένει ο διακομιστής, συμπεριλαμβανομένου ενός έγκυρου χρονοσήμου και όλων των απαραίτητων αναγνωριστικών.

Αυτό το νεοσχηματισμένο μήνυμα σύνδεσης εμφανίζεται νόμιμο στο διακομιστή επειδή περιέχει έγκυρες και ενημερωμένες πληροφορίες. Ως αποτέλεσμα, ο διακομιστής δέχεται το μήνυμα ως προερχόμενο από τον γνήσιο χρήστη  $p_i$ . Αυτό επιτρέπει στον επιτιθέμενο να υποδυθεί με επιτυχία τον χρήστη και να αποκτήσει μη εξουσιοδοτημένη πρόσβαση. [8]

### 3.7.1.3 *Man-in-the-middle(MITM)*

Κατά τη διάρκεια αυτής της επίθεσης, οι συσκευές ομίχλης που χρησιμεύουν ως πύλες μπορεί να παραβιαστούν ή να αντικατασταθούν από πλαστές. Παραδοσιακά αυτές οι επιθέσεις μπορούν να αποτραπούν με τη χρήση μεθόδων κρυπτογράφησης και αποκρυπτογράφησης αλγορίθμων. [6]

### 3.7.1.4 *Επιθέσεις άρνησης παροχής υπηρεσιών (Denial of Service-DoS)*

Η άρνηση παροχής υπηρεσιών είναι μια επίθεση που εμποδίζει την πρόσβαση αυθεντικών χρηστών σε πόρους. Δεδομένου ότι τα κατανεμημένα συστήματα είναι ανοικτά από τη φύση τους, η ανάγκη για προστασία από DoS είναι ακόμη πιο ουσιαστική. Γίνεται πολύ δύσκολο να αποτραπούν ή να διαχειριστούν οι επιθέσεις DoS που εκτελούνται από το από μία/πολλές πηγές για να οργανωθεί μια κατανεμημένη επίθεση άρνησης παροχής υπηρεσιών (DDoS). Η πρόθεση του είναι συνήθως να εγκαταστήσει ένα κακόβουλο λογισμικό στο μηχάνημα του θύματος. Το τείχος προστασίας παίζει σημαντικό ρόλο εδώ στον περιορισμό της κυκλοφορίας σε ένα εσωτερικό δίκτυο από τον εξωτερικό κόσμο με βάση διάφορα φίλτρα ανάλογα με τις κατάλληλες ανάγκες του οργανισμού. [6]

#### 3.7.1.4.1 *Κατανεμημένη επίθεση άρνησης παροχής υπηρεσιών (DDoS)*

Σε αυτόν τον τύπο επίθεσης τίθεται υπό αμφισβήτηση η διαθεσιμότητα του συστήματος-στόχου, καθώς εκατομμύρια πελάτες δεν μπορούν να έχουν πρόσβαση σε ένα διαδικτυακό σύστημα το οποίο είναι υπερφορτωμένο με κίνηση από πολλαπλές πηγές. [6]

### 3.7.2 Ασφάλεια δικτύου

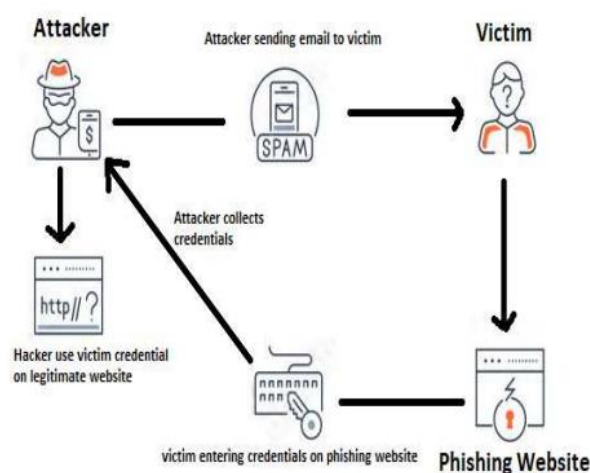
Η ασφάλεια δικτύου ασχολείται με μέτρα και ανησυχίες για την προστασία πληροφοριακών συστημάτων. Προφυλάσσει από μη εξουσιοδοτημένες εισβολές και προστατεύει τη χρηστικότητα και την ακεραιότητα του δικτύου και των δεδομένων. οι κυβερνοεπιθέσεις στα δίκτυα μπορεί να είναι παθητικές όπως η σάρωση θυρών, η υποκλοπή και η κρυπτογράφηση, και ενεργητικές, όπως το phishing, το cross-site scripting(XSS) και επιθέσεις άρνησης παροχής υπηρεσιών (DoS).

#### 3.7.2.1 Κρυπτογράφηση

Αυτό σημαίνει κωδικοποίηση των δεδομένων με τέτοιο τρόπο ώστε να μην μπορούν να γίνουν κατανοητά απευθείας με την ανάγνωσή τους. [6]

#### 3.7.2.2 Σάρωση θυρών(Port scanning)

Η σάρωση σημαντική τεχνική πληροφοριών. Με στοιχεία σάρωσης σε δίκτυο, οι αμυντικοί να εντοπίσουν Η σάρωση θύρας είναι προσδιορισμού του συγκεκριμένες διαθέσιμες σε έναν ένα δίκτυο, αποκρίσεις σε σύνδεσης. Μια να περιγραφεί ως που αποτελείται από «εχθρικές αναζητήσεις στο Διαδίκτυο για ανοικτές «πόρτες» ή «θύρες», μέσω των οποίων οι εισβολείς αποκτούν πρόσβαση σε υπολογιστές». Οι τεχνικές αυτές συνίστανται στην αποστολή ενός μηνύματος σε μια θύρα και στην ακρόαση της απάντησης. [9]



θυρών, είναι μια συλλογή τεχνικών βάση τα στατιστικά ένα πραγματικό του δικτύου μπορούν κακόβουλες σαρώσεις. μια μέθοδος κατά πόσον υπηρεσίες είναι κεντρικό υπολογιστή ή παρατηρώντας τις προσπάθειες σάρωση θύρας μπορεί

#### 3.7.2.3 Phishing

Το «phishing» ή αλλιώς «ψάρεμα» είναι μια τεχνική κοινωνικής μηχανικής που, μέσω της χρήσης διαφόρων μεθοδολογιών, αποσκοπεί στο να επηρεάσει τον στόχο της επίθεσης ώστε να αποκαλύψει προσωπικές πληροφορίες, όπως διεύθυνση ηλεκτρονικού ταχυδρομείου, όνομα χρήστη, κωδικό πρόσβασης ή οικονομικές πληροφορίες. Οι πληροφορίες αυτές χρησιμοποιούνται στη συνέχεια από τον επιτιθέμενο σε βάρος του θύματος. Ο όρος phishing προέρχεται από τη λέξη «fishing» (ψάρεμα). Η λογική αυτής της ορολογίας είναι ότι ο επιτιθέμενος χρησιμοποιεί «δόλωμα» για να δελεάσει το θύμα και στη συνέχεια «ψαρεύει» τις προσωπικές πληροφορίες που θέλει να κλέψει. [10]

### **Σχήμα 6: Μηχανισμός Phishing [50]**

Στο παραπάνω σχήμα βλέπουμε την διαδικασία του phishing, όπου:

1. Ο επιτιθέμενος (Attacker) στέλνει ένα απατηλό email (spam) στο θύμα.
2. Το θύμα (Victim) ανοίγει το email και ακολουθεί τον σύνδεσμο που οδηγεί σε έναν ψεύτικο ιστότοπο (Phishing Website).
3. Το θύμα εισάγει τα διαπιστευτήριά του (username/κωδικό) νομίζοντας ότι είναι σε νόμιμο site.
4. Ο επιτιθέμενος συλλέγει τα στοιχεία αυτά.
5. Τέλος, ο επιτιθέμενος χρησιμοποιεί τα κλεμμένα στοιχεία για να εισέλθει σε έναν πραγματικό ιστότοπο με την ταυτότητα του θύματος.

#### **3.7.2.4 Cross-site scripting(XSS)**

Το cross-site scripting (XSS) είναι μια ευπάθεια ασφαλείας που επηρεάζει εφαρμογές ιστού. Εμφανίζεται λόγω ακατάλληλης ή ελλιπούς εξυγίανσης των εισόδων του χρήστη. Η ευπάθεια ασφαλείας προκάλεσε πολλά προβλήματα στους χρήστες του και στις εφαρμογές διακομιστή. [11]

#### **3.7.3 Ασφάλεια στο νέφος**

Η ασφάλεια του υπολογιστικού νέφους λαμβάνει υπόψη διάφορες τεχνολογίες και πολιτικές που βασίζονται στον έλεγχο για την προστασία των πληροφοριών, των εφαρμογών δεδομένων και της υποδομής εντός του νέφους. Δεδομένου ότι το νέφος είναι ένας κοινόχρηστος πόρος, οι κυβερνοεπιθέσεις στο νέφος μπορεί να οδηγήσουν σε δεδομένα παραβιάσεις, ευπάθειες συστημάτων, κακόβουλους εσωτερικούς χρήστες, απώλεια δεδομένων και κοινή τεχνολογία. ευπάθειες. Ορισμένες επιθέσεις στο περιβάλλον του υπολογιστικού νέφους είναι το phishing, οι επιθέσεις άρνησης παροχής υπηρεσιών (DoS) και τα παραβιασμένα διαπιστευτήρια χρήστη.

### 3.7.3.1 Παραβιασμένα διαπιστευτήρια χρήση (CUC)

Τα παραβιασμένα διαπιστευτήρια χρήση (CUC) μπορεί να είναι ένας εισβολέας που κλέβει τις λεπτομέρειες σύνδεσης του χρήστη, μπορεί συχνά να διεισδύσει στην ασφάλεια της περιμέτρου για να αποκτήσει προνομιακή πρόσβαση στο περιβάλλον πρόσβασης στη βάση δεδομένων πολλαπλών χρηστών. [20]

### 3.7.4 Ασφάλεια εφαρμογών

Ασφάλεια μιας εφαρμογής εξασφαλίζεται με τον μετριασμό της ασφάλειας τρωτά σημεία. Δεδομένου ότι η ανάπτυξη μιας εφαρμογής έχει διάφορα στάδια, όπως ο σχεδιασμός, η ανάπτυξη, η εγκατάσταση, η αναβάθμιση και η συντήρηση, κάθε στάδιο είναι ευάλωτο σε κυβερνοεπιθέσεις. Οι συνήθεις επιθέσεις που αφορούν την ασφάλεια εφαρμογών ιστού είναι οι cross-site scripting (XSS), η SQL injection, οι υπερχειλίσσεις buffer και οι κατανεμημένες επιθέσεις άρνησης παροχής υπηρεσιών (DDoS). Στις εφαρμογές για κινητά, λαμβάνουν χώρα επιθέσεις όπως spyware, botnets, click fraud και μολύνσεις από κακόβουλο λογισμικό(malware).

#### 3.7.4.1 SQL injection

Όπως υποδηλώνει το όνομα της επίθεσης, πρέπει να υπάρχει ένα ερώτημα SQL το οποίο έχει εισαχθεί ενώ ο χρήστης εισάγει τα δεδομένα- καθώς το ερώτημα εκτελείται στο τέλος του πελάτη, η επίθεση τίθεται σε λειτουργία. Πρόκειται για μια από τις πιο συνηθισμένες τεχνικές hacking στο διαδίκτυο στην οποία εισάγεται κώδικας που μπορεί να δημιουργήσει χάος στη βάση δεδομένων. [6]

#### 3.7.4.2 Υπερχειλίσσεις buffer

Οι υπερχειλίσσεις buffer ήταν η πιο κοινή μορφή ευπάθειας ασφαλείας τα τελευταία δέκα χρόνια. Πιο , οι ευπάθειες υπερχειλίσσης buffer κυριαρχούν στην περιοχή των ευπαθειών απομακρυσμένης διείσδυσης στο δίκτυο, όπου ένας ανώνυμος χρήστης του Διαδικτύου επιδιώκει να αποκτήσει μερικό ή ολικό έλεγχο ενός κεντρικού υπολογιστή. Επειδή αυτού του είδους οι επιθέσεις επιτρέπουν σε οποιονδήποτε να αποκτήσει τον πλήρη έλεγχο ενός κεντρικού υπολογιστή, αντιπροσωπεύουν μια από τις σοβαρότερες κατηγορίες απειλών ασφαλείας. [12]

#### 3.7.4.3 Spyware

Το λογισμικό κατασκοπείας(spyware) συλλέγει λεπτομέρειες σχετικά με έναν οργανισμό ή ένα άτομο και τις αποστέλλει σε κακόβουλους χρήστες . Το keylogger spyware συλλέγει τις κινήσεις πλήκτρων που πατάει ο χρήστης, τις καταγράφει και στέλνει τα δεδομένα σε κακόβουλο λογισμικό επιτιθέμενος. Για την επιβολή της κατάλληλης ασφάλειας δεδομένων είναι απαραίτητη η ανίχνευση και ο εντοπισμός αυτών των επιθέσεων. Η έγκαιρη ανίχνευση των αυτών των επιθέσεων θα βοηθήσει στον έλεγχο της ταχύτητας εξάπλωσης του spyware. [13]

#### 3.7.4.4 Botnets

Botnet ονομάζεται ένα αυτοδιαδιδόμενο, αυτοοργανωμένο και αυτόνομο πλαίσιο, το οποίο σχηματίζεται από μια ομάδα bots κάτω από μια υποδομή διοίκησης και ελέγχου. [14]

#### 3.7.4.5 Απάτη με κλικ (*Click fraud*)

Η απάτη με κλικ είναι η πρακτική του παραπλανητικού κλικ σε διαφημίσεις αναζήτησης με σκοπό είτε να αυξηθούν τα έσοδα του ιστότοπου τρίτων είτε να εξαντληθεί ο προϋπολογισμός ενός διαφημιζόμενου. Οι διαφημιστές αναζήτησης αναγκάζονται να εμπιστεύονται ότι οι μηχανές αναζήτησης ανιχνεύουν και αποτρέπουν την απάτη με κλικ, παρόλο που οι μηχανές πληρώνονται για κάθε μη ανιχνευμένο δόλιο κλικ. Βρίσκουμε συνθήκες υπό τις οποίες είναι προς το συμφέρον μιας μηχανής αναζήτησης να επιτρέψει κάποια απάτη με κλικ. [15]

#### 3.7.4.6 Κακόβουλο λογισμικό (*malware*)

Το κακόβουλο λογισμικό αποτελεί ιδιαίτερα αυξημένη απειλή, επειδή είναι ειδικά σχεδιασμένο για να αποφεύγει τους μηχανισμούς ανίχνευσης, εξαπλώνεται αδρανώς, στη φύση για παρατεταμένες χρονικές περιόδους, συλλέγοντας ευαίσθητες πληροφορίες ή τοποθετώντας τον εαυτό του.

Το κακόβουλο λογισμικό έχει ομαδοποιηθεί κανονικά σε κατηγορίες όπως οι ιοί, τα σκουλήκια, τα Trojans, τα rootkits. [16]

##### 3.7.4.6.1 Ιοί (*viruses*)

Οι ιοί υπολογιστών αναφέρονται γενικά σε προγράμματα που ακούσια εισέρχονται σε υπολογιστές, διαταράσσουν την κανονική λειτουργία και προκαλούν ζημιά σε δεδομένα και προγράμματα. [17]

##### 3.7.4.6.2 Σκουλήκια (*worms*)

Ένα σκουλήκι βρίσκει το δρόμο του σε μητρώα του συστήματος και εξαπλώνεται ως ανεξάρτητο πρόγραμμα. Είναι ένα πρόγραμμα που αυτοδιαδίδεται σε ένα δίκτυο εκμεταλλευόμενο κενά ασφαλείας ή πολιτικής σε ευρέως χρησιμοποιούμενες υπηρεσίες. [17]

##### 3.7.4.6.3 Trojans

Η μέθοδος του Δούρειου Ίππου (Trojan Horse) είναι μια έμμεση προσέγγιση για τον προσδιορισμό της ενεργειακής εξάρτησης των συντελεστών  $S$  των αστροφυσικά σχετικών αντιδράσεων δύο σωμάτων. Αυτό επιτυγχάνεται με τη μελέτη στενά συνδεδεμένων αντιδράσεων τριών σωμάτων υπό συνθήκες οιοιδήποτε ελεύθερης σκέδασης. [18]

##### 3.7.4.6.4 Rootkits

Το rootkit είναι ένας τύπος κακόβουλου λογισμικού που έχει σχεδιαστεί για να αποκτά προνομιακή πρόσβαση σε έναν υπολογιστή, ενώ παράλληλα κρύβεται από τον χρήστη και το λειτουργικό σύστημα, για παράδειγμα, μέσω της διεύθυνσης που θέτει σε κίνδυνο τα κανάλια επικοινωνίας εντός του λειτουργικού συστήματος. Ένα rootkit μπορεί να κρύψει αρχεία, δεδομένα, διεργασίες και θύρες δικτύου και συνήθως μπορεί να επιβιώσει από μια επανεκκίνηση του συστήματος. [19]

### 3.7.5 Ασφάλεια του Διαδικτύου των πραγμάτων

Το διαδίκτυο των πραγμάτων (IoT) αποτελείται από υπολογιστικές, μηχανικές και ψηφιακές συσκευές με μοναδικά αναγνωριστικά που μπορούν να μεταφέρουν δεδομένα μέσω μέσω του δικτύου χωρίς ανθρώπινη παρέμβαση. Η ασφάλεια του IoT διασφαλίζει αυτά τα συνδεδεμένα συσκευές και τα δίκτυα του IoT. Οι επιθέσεις περιλαμβάνουν spyware και botnets. [6]

| Τομέας                    | Είδος Επίθεσης                 | Περιγραφή                                                 |
|---------------------------|--------------------------------|-----------------------------------------------------------|
| <b>Κρίσιμες Υποδομές</b>  | MITM, DoS, DDoS, Υποκλοπή      | Παραβολή ή υπερφόρτωση υπηρεσιών για διακοπή ή υποκλοπή.  |
| <b>Ασφάλεια Δικτύου</b>   | Phishing, XSS, Σάρωση Θυρών    | Απόσπαση στοιχείων ή ανίχνευση ευπαθειών μέσω δικτύου.    |
| <b>Ασφάλεια στο Νέφος</b> | CUC, DoS                       | Κλοπή διαπιστευτηρίων και επιθέσεις σε cloud υποδομές.    |
| <b>Ασφάλεια Εφαρμογών</b> | SQLi, Malware, Buffer Overflow | Εκμετάλλευση ευπαθειών λογισμικού για κακόβουλη πρόσβαση. |
| <b>Ασφάλεια IoT</b>       | Spyware, Botnets               | Παρακολούθηση ή απομακρυσμένος έλεγχος έξυπνων συσκευών.  |

**Πίνακας 4: Απειλές και τρωτά σημεία για την ασφάλεια στον κυβερνοχώρο**

## 3.8 Στρατηγικές κυβερνοασφάλειας

Στον αγώνα κατά των κακόβουλων απειλών, υπήρξε συνεργατική υποστήριξη από εμπειρογνώμονες για τον σχεδιασμό διαφόρων συστημάτων άμυνας στον κυβερνοχώρο. Τόσο οι ερευνητές όσο και οι σχεδιαστές έχουν αντίστοιχα τους ίδιους στόχους: να διατηρήσουν την ιδιωτικότητα, την ακεραιότητα και την προσβασιμότητα των πληροφοριών μέσω των συστημάτων κυβερνοάμυνας ενάντια σε εσωτερικές και εξωτερικές απειλές. Ο κύριος στόχος των συστημάτων κυβερνοάμυνας είναι η καταπολέμηση των απειλών ασφαλείας που προέρχονται από διαδικτυακές πηγές, όπως οι ιοί[17], τα Trojans[18], τα σκουλήκια, το spam και τα botnets[14]. Τα συστήματα αυτά αμύνονται κατά των απειλών κυβερνοασφάλειας τόσο σε επίπεδο δικτύου όσο και σε επίπεδο κεντρικού υπολογιστή. Τα συστήματα άμυνας που βασίζονται στο δίκτυο κάνουν χρήση της ροής του δικτύου, ενώ τα συστήματα άμυνας που βασίζονται στον κεντρικό υπολογιστή ελέγχουν τα επερχόμενα δεδομένα του σταθμού εργασίας μέσω μηχανισμών που έχουν σχεδιαστεί σε τείχη προστασίας, αντικό λογισμικό, η κρυπτογράφηση και συστήματα ανίχνευσης εισβολών (IDS).

Απαιτούνται προηγμένες μέθοδοι για την ανακάλυψη άγνωστων προηγούμενων κυβερνοπαθειών και τεχνικών προς μια πιο αξιόπιστη υποδομή κυβερνοασφάλειας, συμπεριλαμβανομένων αμυντικών και επιθετικών προσεγγίσεων. Οι αμυντικές προσεγγίσεις χρησιμοποιούν αντιδραστικές στρατηγικές που επικεντρώνονται στην πρόληψη, την ανίχνευση και την αντίδραση. Αυτοί οι μηχανισμοί παρακολουθούν, εντοπίζουν και αποκλείουν ιούς και άλλες κακόβουλες κυβερνοεπιθέσεις. Ωστόσο, αυτές οι μέθοδοι δεν εξαλείφουν πλήρως τα τρωτά σημεία, τις απειλές και τις επιθέσεις, επειδή ο σχεδιασμός και η υλοποίηση του λογισμικού και της δικτυακής υποδομής είναι εγγενώς ατελής. Αυτή είναι η πιο παραδοσιακή μέθοδος για να διατηρηθούν τα δίκτυα ασφαλή από τους εγκληματίες του κυβερνοχώρου και απαιτεί ενδελεχή κατανόηση του συστήματος που πρόκειται να διασφαλιστεί. Τα προληπτικά μέτρα αναπτύσσονται από την κατανόηση του συστήματος και των πιθανών αδύναμων σημείων.

Από την άλλη πλευρά, οι επιθετικές προσεγγίσεις αποτελούν τον αντίποδα των αμυντικών μεθόδων και προβλέπουν και απομακρύνουν προληπτικά τις απειλές στο σύστημα χρησιμοποιώντας τεχνικές ηθικού hacking, τις Επιχειρήσεις της Κόκκινης Ομάδας (Red team Operations), την Πληροφορία για απειλές στον κυβερνοχώρο (CTI) καθώς και Honeytraps & Honeynets. [21]

### **3.8.1 Αμυντικά μέτρα προστασίας**

#### *3.8.1.1 Τείχη προστασίας (firewalls)*

Τα τείχη προστασίας, αποτελούν ένα «τείχος» που εμποδίζει την ελεύθερη ροή δεδομένων μεταξύ του εσωτερικού και του εξωτερικού. Χρησιμοποιείται σωστά, ένα τείχος προστασίας μπορεί να προσφέρει σημαντική αύξηση της ασφάλειας των υπολογιστών. [22]

#### *3.8.1.2 Αντιικό λογισμικό (antiviruses-AV)*

Ένα αντιικό λογισμικό (AV) είναι ένα πρόγραμμα που χρησιμοποιείται από τους περισσότερους χρήστες για να διατηρούν τους υπολογιστές τους απαλλαγμένους από κακόβουλο λογισμικό. Το επιθυμητό λογισμικό AV θα πρέπει να εντοπίζει με ακρίβεια και να σταματά το κακόβουλο λογισμικό μόλις αυτό εισέρχεται σε έναν υπολογιστή. Διαφορετικά, θα πρέπει τουλάχιστον να ελέγχει τις ζημιές που οφείλονται στην εκτέλεση του μη εντοπισμένου κακόβουλου λογισμικού, όπως για παράδειγμα την κατάληψη του συστήματος, τη δημιουργία πρόσθετων παραβάσεων ασφαλείας ή την εισαγωγή άλλου κακόβουλου λογισμικού. [23]

#### *3.8.1.3 Κρυπτογράφηση*

Όπως αναφερθήκαμε σε αυτή και παραπάνω (3.7.2.1), η κρυπτογραφία αποτελεί μέρος της επιστήμης που είναι γνωστή ως κρυπτολογία, η οποία αποτελείται από δύο κύριους τομείς: την κρυπτανάλυση και την κρυπτογραφία. Ο όρος κρυπτολογία προέρχεται από την ελληνική λέξη *criptos* (κρυμμένος) και *logos*, που στην προκειμένη περίπτωση σημαίνει επεξεργασία. Η κρυπτανάλυση επικεντρώνεται στον προσδιορισμό του τρόπου με τον οποίο συστήματα πληροφοριών, τα οποία προορίζονται να είναι ασφαλή, μπορούν να δεχθούν επίθεση από ξένους. Με άλλα λόγια, τα συστήματα μπορούν να επηρεαστούν από την κρυπτογραφία ή από

αλγορίθμους που αποκρυπτογραφούνται, δηλαδή, όπου μπορούν να κρυφτούν ευαίσθητες πληροφορίες.

Ευαίσθητες πληροφορίες, όπως οι κωδικοί πρόσβασης, οι κωδικοί πρόσβασης σε δεδομένα και όλες οι είδη ιδιωτικών δεδομένων επηρεάζονται έτσι δυνητικά από εξωτερικές απειλές. Στην πραγματικότητα, από αμνημονεύτων χρόνων, τα ανθρώπινα όντα έχουν εφεύρει κώδικες, όπως η κρυπτογράφηση του Καίσαρα, οι αινιγματικές μηχανές που χρησιμοποιήθηκαν κατά τον Δεύτερο Παγκόσμιο Πόλεμο και η αποκωδικοποίηση από Alan Turing (Άγγλος μαθηματικός, καθηγητής της λογικής, κρυπτογράφος και θεωρητικός βιολόγος. Θεωρείται ο «πατέρας της επιστήμης υπολογιστών»), όπου οι μυστικοί κώδικες που ήταν πολύ δύσκολο να αποκωδικοποιηθούν. Συνεπώς, η κρυπτογραφία αποτελεί μέρος της επιστήμης που κρύβει μηνύματα, καθώς και εμποδίζει την πρόσβαση σε αυτά και καθιστά ορισμένες πληροφορίες να είναι δύσκολο να βρεθούν. [24]

#### *3.8.1.4 Συστήματα ανίχνευσης εισβολών (IDS)*

Ο στόχος της ανίχνευσης εισβολών είναι η παρακολούθηση των περιουσιακών στοιχείων του δικτύου ώστε να ανιχνεύεται ανώμαλη συμπεριφορά και κακή χρήση στο δίκτυο. Η τεχνολογία αυτή υπάρχει εδώ και σχεδόν είκοσι χρόνια, αλλά μόνο πρόσφατα σημείωσε δραματική αύξηση της δημοτικότητας και της ενσωμάτωσης στη συνολική ασφάλεια των πληροφοριακών υποδομών.

Ο σκοπός του IDS λοιπόν, είναι να βοηθήσει τα συστήματα υπολογιστών στο πώς να αντιμετωπίζουν τις επιθέσεις. Το IDS συλλέγει πληροφορίες από διάφορες διαφορετικές πηγές εντός των συστημάτων υπολογιστών και δικτύων και συγκρίνει αυτές τις πληροφορίες με προϋπάρχοντα πρότυπα, ώστε να διακρίνει αν υπάρχουν επιθέσεις ή αδυναμίες. [25]

### **3.8.2 Επιθετικά μέτρα προστασίας**

#### *3.8.2.1 Ηθικό hacking(Ethical hacking)*

Το ηθικό hacking, είναι η τεχνική με την οποία άτομα με εξουσιοδότηση και γνώσεις ασφάλειας προσπαθούν να εντοπίσουν και να διορθώσουν ευπάθειες σε υπολογιστικά συστήματα ή δίκτυα, μιμούμενοι τις μεθόδους κακόβουλων εισβολέων, αλλά με σκοπό την ενίσχυση της ασφάλειας. Τους αποκαλούν χάκερς, κράκερς, εισβολείς ή επιτιθέμενους, είναι όλοι παρείσακτοι που προσπαθούν να εισβάλουν στο σύστημα σας. δίκτυα και συστήματα. Κάποιοι το κάνουν για διασκέδαση, κάποιοι το κάνουν για κέρδος, ή κάποιοι απλά το κάνουν για να διαταράξουν τις λειτουργίες σας και να ίσως να κερδίσουν κάποια αναγνώριση. Αν και όλοι έχουν ένα κοινό: προσπαθούν να αποκαλύψουν μια αδυναμία στο το σύστημά σας προκειμένου να το εκμεταλλευτούν. [26]

#### *3.8.2.2 Επιχειρήσεις της Κόκκινης Ομάδας (Red team Operations)*

Οι κόκκινες ομάδες διαδραματίζουν κρίσιμο ρόλο στην αξιολόγηση της ασφάλειας ενός δικτύου. ανιχνεύοντάς το ενεργά για αδυναμίες και ευπάθειες. Η σημασία των δοκιμών κόκκινων ομάδων (red teaming) για τις σύγχρονες επιχειρήσεις δεν μπορεί να υποτιμηθεί. Σε αυτές τις

ασκήσεις, εξωτερικά ή εσωτερικά προερχόμενες «κόκκινες ομάδες» - ομάδες εμπειρογνομόνων ασφαλείας που μιμούνται επιτιθέμενους - επιχειρούν να ελέγξουν όλες τις πτυχές της κατάστασης ασφαλείας ενός οργανισμού, εξαπολύοντας επαναλαμβανόμενες και πολύπλοκες επιθέσεις κατά το επιχειρησιακό δίκτυο υπολογιστών του οργανισμού. [27]

### 3.8.2.3 Πληροφορία για απειλές στον κυβερνοχώρο (CTI)

Σε απάντηση στην αύξηση των παγκόσμιων περιστατικών ασφαλείας και στην αποτυχία των γενικών προσεγγίσεων ετοιμότητας ασφάλειας, οι οργανισμοί έχουν στραφεί όλο και περισσότερο στις προσπάθειες πληροφοριών για τις απειλές στον κυβερνοχώρο. Εν ολίγοις, η Πληροφορία για απειλές στον κυβερνοχώρο, αντιπροσωπεύει αξιοποιήσιμες πληροφορίες για απειλές προσαρμοσμένες σε έναν συγκεκριμένο οργανισμό που απαιτεί προσεκτική προσοχή και πρόληψη. Η Πληροφορία για απειλές στον κυβερνοχώρο αποτελείται από δραστηριότητες που έχουν σχεδιαστεί για την ανάκτηση πληροφοριών απειλής που αφορούν έναν συγκεκριμένο οργανισμό για τη χάραξη ακριβέστερων στρατηγικών άμυνας. Οι πληροφορίες αυτές περιλαμβάνουν τύπους απειλών, πηγές, φορείς, τεχνολογίες και φορείς επίθεσης. [28]

### 3.8.2.4 Honeypots & Honeynets

Μια από τις πιο δημοφιλείς μεθόδους εκμάθησης επιτιθέμενων είναι η χρήση honeypots. Τα honeypots ορίζονται ως ένας πόρος του συστήματος πληροφοριών του οποίου η αξία έγκειται σε ένα μη εξουσιοδοτημένη ή παράνομη χρήση αυτού του πόρου. Μπορεί επίσης να οριστεί ως ένας υπολογιστικός πόρος του οποίου η αξία έγκειται στην ύπαρξη που δέχεται επίθεση. [29]

| Κατηγορία             | Μέτρο / Τεχνική                             | Περιγραφή                                                                                              |
|-----------------------|---------------------------------------------|--------------------------------------------------------------------------------------------------------|
| <b>Αμυντικά Μέτρα</b> | Τείχη προστασίας (Firewalls)                | Ελέγχουν τη ροή δεδομένων μεταξύ εσωτερικών και εξωτερικών δικτύων, λειτουργούν ως «τείχος ασφαλείας». |
|                       | Αντιικό λογισμικό (Antiviruses - AV)        | Εντοπίζουν και αποτρέπουν κακόβουλο λογισμικό πριν μολύνει το σύστημα.                                 |
|                       | Κρυπτογράφηση                               | Διασφαλίζει την ακεραιότητα και την ιδιωτικότητα δεδομένων μέσω κωδικοποίησης.                         |
|                       | Συστήματα Ανίχνευσης Εισβολών (IDS)         | Παρακολουθούν δίκτυα και συστήματα για ανίχνευση ανώμαλης δραστηριότητας ή επιθέσεων.                  |
| <b>Επιθετικά</b>      | Ηθικό Hacking (Ethical Hacking)             | Εξουσιοδοτημένοι ειδικοί προσομοιώνουν επιθέσεις για να εντοπίσουν ευπάθειες.                          |
|                       | Επιχειρήσεις Κόκκινης Ομάδας (Red Team Ops) | Εξειδικευμένες ομάδες πραγματοποιούν οργανωμένες, προσομοιωμένες επιθέσεις σε υποδομές.                |

| Μέτρα | Πληροφορίες Απειλών<br>(Cyber Threat Intelligence -<br>CTI) | Συλλογή και ανάλυση πληροφοριών για απειλές<br>προσαρμοσμένες σε κάθε οργανισμό.                |
|-------|-------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
|       | Honeypots & Honeynets                                       | Ψευδείς πόροι που δελεάζουν τους<br>επιτιθέμενους, βοηθώντας στον εντοπισμό και<br>μελέτη τους. |

**Πίνακας 5: Μέτρα ασφαλείας 2 κατηγοριών**

### 3.9 Τάσεις στην κυβερνοασφάλεια

Ακολουθεί η περιγραφή ορισμένων τάσεων με μεγάλο αντίκτυπο στην κυβερνοασφάλεια:

#### 3.9.1 Η ασφάλεια στον κυβερνοχώρο αλλάζει στις εταιρείες

Η ασφάλεια στον κυβερνοχώρο είναι ένας ταχέως αναπτυσσόμενος τομέας, ο οποίος δημιουργεί νέες θέσεις εργασίας για την επόμενη δεκαετία, καθώς τόσο ο δημόσιος όσο και ο ιδιωτικός τομέας σχεδιάζουν σημαντικές επενδύσεις για την προστασία των δεδομένων και των πληροφοριακών συστημάτων. Η απειλή στον κυβερνοχώρο θεωρείται μία από τις σοβαρότερες προκλήσεις για την οικονομία και την εθνική ασφάλεια, ενώ η οικονομική ευημερία του 21ου αιώνα εξαρτάται σε μεγάλο βαθμό από την αποτελεσματική προστασία ψηφιακών υποδομών. Τα επόμενα χρόνια, χιλιάδες νέοι επαγγελματίες της κυβερνοασφάλειας θα απαιτηθούν για να καλύψουν τις αυξανόμενες ανάγκες της αγοράς. Πολλές από τις εξελίξεις στον τομέα αυτό θα απαιτήσουν νέες δεξιότητες στην επιστήμη των υπολογιστών και στην έρευνα. Η συνεχώς αυξανόμενη ποσότητα δεδομένων καθιστά επίσης απαραίτητη την ανθρώπινη λογική, ενώ η ευελιξία των ικανοτήτων θα αποτελέσει βασικό στοιχείο για το μέλλον της ασφάλειας στον κυβερνοχώρο. [30]

### **3.9.2 Διακομιστές Ιστού (Web servers)**

Ο κίνδυνος επιθέσεων στην εφαρμογή στο διαδίκτυο για την εξαναγκαστική χρήση των δεδομένων για την έγχυση κακόβουλου κώδικα είναι πολύ επίμονος και ο επιτιθέμενος θα διανείμει κακόβουλο κώδικα μέσω ορισμένων νόμιμων διακομιστών στο διαδίκτυο. Έχοντας τα αποκτήματα οι οποίοι δεν είναι εξουσιοδοτημένοι για τα δεδομένα και πολλοί από τους οποίους αποκτούν το προσοχή των μέσων ενημέρωσης είναι εξίσου υψηλός κίνδυνος. Υπάρχει ανάγκη για προστασίας των διακομιστών και όλων των εφαρμογών. [30]

### **3.9.3 Υπολογιστικό νέφος και οι υπηρεσίες του**

Πολλοί οργανισμοί υιοθετούν τη χρήση υπηρεσιών νέφους, και ολόκληρος ο κόσμος προχωρά σταδιακά προς την κατεύθυνση του υπολογιστικού νέφους, η τάση αυτή παρουσιάζει μια τεράστια πρόκληση για την ασφάλεια στον κυβερνοχώρο. Επιπροσθέτως, καθώς το εύρος όλων των εφαρμογών εκεί έξω είναι στο νέφος, η πολιτική του διαδικτύου και των λύσεων νέφους πρέπει επίσης να αυξηθεί ως πρόβλεψη για την απώλεια των απαραίτητων δεδομένων. Θα μπορούσαν οι λύσεις να κάνουν τα ξεχωριστά μοντέλα και πολλά από τα ζητήματα που σημειώνονται επίσης σχετικά με την ασφάλεια, οι υπηρεσίες του νέφους μπορούν να δώσουν τεράστιες ευκαιρίες, και θα πρέπει να σημειωθεί ότι καθώς υπάρχει πρόοδος στο σύννεφο, τα ζητήματα ασφάλειας αυξάνονται επίσης. [30]

### **3.9.4 Προηγμένες επίμονες απειλές**

Το APT, γνωστό ως Advanced Persistent Threat είναι συνήθως μια νέα πλατφόρμα για την καταπολέμηση του εγκλήματος στον κυβερνοχώρο. Για πολλά χρόνια ο υπολογιστής σύστημα δικτύου χρησιμοποιείται ως το κύριο μέρος για την αξιολόγηση του στόχου επιθέσεις. Καθώς πολλές από τις επιθέσεις χρησιμοποιούν πιο προηγμένες μεθόδους, η ασφάλεια του δικτύου θα πρέπει να συνδυαστεί με άλλες λύσεις ασφάλειας για τον έγκαιρο εντοπισμό των επιθέσεων. Γι' αυτό το λόγο πρέπει να προωθήσει τα εργαλεία ασφάλειας για την προστασία του δικτύου από πρόσθετες απειλές. [30]

### **3.9.5 Δίκτυα κινητής τηλεφωνίας**

Τώρα, οι άνθρωποι έχουν τη δυνατότητα να επικοινωνούν με άλλους ανθρώπους σε όλο τον κόσμο και η ασφάλεια του δικτύου αποτελεί μεγάλο μέλημα. Τα διάφορα εργαλεία ασφάλειας δεν είναι αρκετά, καθώς οι άνθρωποι χρησιμοποιούν τα συσκευές υπολογιστών, όπως κινητά τηλέφωνα, φορητούς υπολογιστές, οι οποίοι απαιτούν περισσότερες ασφάλεια εκτός από αυτές που είναι εγκατεστημένες στις συσκευές από τους από προεπιλογή. Υπάρχει ανάγκη να ληφθούν υπόψη οι ανησυχίες αυτών των κινητών τηλεφώνων και επιπλέον, το δίκτυο κινητής τηλεφωνίας κινδυνεύει περισσότερο. σε αυτές τις διαφημίσεις εγκλημάτων στον κυβερνοχώρο, ο χρήστης πρέπει να είναι προσεκτικός στο άνοιγμα διαφημίσεων και συνδέσμων. [30]

### **3.9.6 Κρυπτογράφηση δεδομένων**

Η κρυπτογράφηση δεδομένων, μεταφράζει τις πληροφορίες σε άλλη μορφή, ή κώδικα, έτσι ώστε μόνο τα άτομα που έχουν πρόσβαση σε ένα μυστικό κλειδί (επίσημα που ονομάζεται κλειδί αποκρυπτογράφησης) ή κωδικό πρόσβασης μπορούν να το διαβάσουν. Κρυπτογραφημένα

αρχεία αναφέρονται συνήθως ως κρυπτογραφημένο κείμενο, ενώ οι μη κρυπτογραφημένες στατιστικές ονομάζεται απλό κείμενο. Επί του παρόντος, η κρυπτογράφηση είναι ένα από τα πιο διάσημα και συμφέρουσες στρατηγικές ασφάλειας πληροφοριών που χρησιμοποιούνται στους οργανισμούς. Η κρυπτογράφηση είναι πολύ βασικό επίπεδο για την προστασία των πληροφοριών και της ακεραιότητας των δεδομένων. Αλλά λόγω της μεγαλύτερης χρήσης της, δημιουργούνται περισσότερα ζητήματα στον τομέα της ασφάλειας στον κυβερνοχώρο. Τα δεδομένα που μεταδίδονται είναι προστατεύονται επίσης με τη χρήση της κρυπτογράφησης και, για παράδειγμα, τα δεδομένα που μεταφέρονται μέσω του δικτύου, των ασύρματων μικροφώνων ή των κινητών τηλέφωνα μπορούν επίσης να αντιληφθούν εάν υπάρχει οποιαδήποτε διαρροή των πληροφοριών από την κρυπτογράφηση του κώδικα. [30]

### ***3.9.7 Εφαρμογές για την ασφάλεια δεδομένων***

Υπάρχει επίσης μια νέα ευκαιρία στην ασφάλεια της εφαρμογής αλλά οι περισσότερες εταιρείες δεν την εκμεταλλεύονται επειδή είναι πολύ δαπανηρή. Τώρα ήρθε η ώρα να το καταλάβουμε με τον σωστό τρόπο για αξιολόγηση της αξίας της ασφάλειας και της μεθόδου φωλιάς για την εξήγηση αυτής της στην εταιρεία. [30]

# 4

## *Τεχνητή νοημοσύνη*

### *4.1 Ορισμός και βασικές έννοιες της τεχνητής νοημοσύνης*

Ο ορισμός TN δεν είναι εύκολος – στην πραγματικότητα, δεν υπάρχει ένας γενικά αποδεκτός ορισμός της έννοιας. Χρησιμοποιούνται πολλοί διαφορετικοί ορισμοί και αυτό μπορεί εύκολα να οδηγήσει σε σύγχυση. Επομένως, είναι σημαντικό να αποσαφηνίσουμε τη χρήση του όρου. Ξεκινάμε συζητώντας τους διάφορους ορισμούς της TN και στη συνέχεια εξηγούμε σε ποιον έχουμε καταλήξει. Η μεγάλη ποικιλία των ορισμών που κυκλοφορούν δεν οφείλεται σε απροσεξία, αλλά είναι σύμφυτη με το ίδιο το φαινόμενο της TN.

Στον ευρύτερο ορισμό της, η TN ταυτίζεται με τους αλγορίθμους. Ωστόσο, αυτή δεν είναι μια ιδιαίτερα χρήσιμη προσέγγιση για την ανάλυσή μας. Οι αλγόριθμοι προϋπήρχαν της TN και έχουν χρησιμοποιηθεί ευρέως εκτός αυτού του πεδίου. Ο όρος «αλγόριθμος» προέρχεται από το όνομα του Πέρση μαθηματικού του 9ου αιώνα Mohammed ibn Musa al-Kharizmi και αναφέρεται σε μια συγκεκριμένη οδηγία για την επίλυση ενός προβλήματος ή την εκτέλεση ενός υπολογισμού. Εάν ορίζαμε την TN απλώς ως τη χρήση αλγορίθμων, θα περιλάμβανε πολλές άλλες δραστηριότητες όπως οι λειτουργίες μιας αριθμομηχανής τσέπης ή ακόμη και οι οδηγίες σε ένα βιβλίο μαγειρικής.

Με τον αυστηρότερο ορισμό της, η TN σημαίνει τη μίμηση από τους υπολογιστές της νοημοσύνης που είναι εγγενής στον άνθρωπο. Οι καθαρολόγοι επισημαίνουν ότι πολλές σημερινές εφαρμογές εξακολουθούν να είναι σχετικά απλές και επομένως δεν είναι αληθινή TN. Αυτό καθιστά αυτόν τον ορισμό ακατάλληλο και για την έκθεσή μας – η χρήση του θα σήμαινε ότι η TN δεν υπάρχει επί του παρόντος. Εμείς θα ορίζαμε ουσιαστικά το φαινόμενο εκτός ύπαρξης.

Ένας συνήθης ορισμός της ΤΝ είναι ότι πρόκειται για μια τεχνολογία που επιτρέπει στις μηχανές να μιμούνται διάφορες πολύπλοκες ανθρώπινες δεξιότητες. Αυτό, ωστόσο, δεν μας δίνει πολλά στοιχεία για να προχωρήσουμε. Στην πραγματικότητα, δεν κάνει τίποτα περισσότερο από το να αποδίδει τον όρο «τεχνητή νοημοσύνη» με διάφορες λέξεις. Όσο δεν προσδιορίζονται αυτές οι «σύνθετες ανθρώπινες δεξιότητες», παραμένει ασαφές τι ακριβώς είναι η ΤΝ. Το ίδιο ισχύει και για τον ορισμό της ΤΝ ως της εκτέλεσης από τους υπολογιστές πολύπλοκων εργασιών σε πολύπλοκα περιβάλλοντα.

Άλλοι ορισμοί εξηγούν περαιτέρω αυτές τις δεξιότητες και τα καθήκοντα. Για παράδειγμα, ο επιστήμονας υπολογιστών Nils John Nilsson περιγράφει μια τεχνολογία η οποία «λειτουργεί κατάλληλα και με προνοητικότητα στο περιβάλλον του». Άλλοι μιλούν για την ικανότητα να αντιλαμβάνεται, να επιδιώκει στόχους, να κινεί ενέργειες και να μαθαίνει από ένα βρόχο ανατροφοδότησης. Ένας παρόμοιος ορισμός έχει προταθεί από την ομάδα εμπειρογνομόνων υψηλού επιπέδου για την ΤΝ της Ευρωπαϊκής Επιτροπής: «Συστήματα που επιδεικνύουν ευφυή συμπεριφορά αναλύοντας το περιβάλλον τους και αναλαμβάνοντας δράσεις – με κάποιο βαθμό αυτονομίας – για την επίτευξη συγκεκριμένων στόχων».

Αυτοί οι ορισμοί που βασίζονται σε καθήκοντα μας βοηθούν να κατανοήσουμε καλύτερα τι είναι η ΤΝ. Αλλά εξακολουθούν να έχουν περιορισμούς. Έννοιες όπως «κάποιος βαθμός αυτονομίας» παραμένουν κάπως ασαφείς. Επιπλέον, αυτοί οι ορισμοί εξακολουθούν να φαίνονται υπερβολικά ευρείς, καθώς περιγράφουν φαινόμενα που οι περισσότεροι από εμάς δεν θα ήταν διατεθειμένοι να ομαδοποιήσουν με τον όρο ΤΝ. Για παράδειγμα, ο ορισμός του Nilsson ισχύει και για έναν κλασικό θερμοστάτη. Αυτή η συσκευή είναι επίσης σε θέση να αντιλαμβάνεται (να μετρά τη θερμοκρασία του δωματίου), να επιδιώκει στόχους (την προγραμματισμένη θερμοκρασία), να ξεκινά ενέργειες (να ρυθμίζει τον θερμοστάτη) και να μαθαίνει από έναν βρόχο ανατροφοδότησης (να σταματά μόλις επιτευχθεί η προγραμματισμένη θερμοκρασία). Ακόμα κι έτσι, οι περισσότεροι άνθρωποι δεν θα ήταν διατεθειμένοι να θεωρήσουν έναν θερμοστάτη ως ΤΝ.

Δεν αποτελεί έκπληξη το γεγονός ότι η ΤΝ είναι τόσο δύσκολο να οριστεί με σαφήνεια. Πρόκειται, άλλωστε, για μια απομίμηση ή προσομοίωση κάποιου πράγματος που δεν κατανοούμε ακόμη πλήρως οι ίδιοι: της ανθρώπινης νοημοσύνης. Αυτό αποτελεί εδώ και πολύ καιρό αντικείμενο έρευνας από ψυχολόγους, επιστήμονες της συμπεριφοράς και νευρολόγους, μεταξύ άλλων. Γνωρίζουμε πολλά για τη νοημοσύνη και τον ανθρώπινο εγκέφαλο, αλλά η γνώση αυτή δεν είναι καθόλου πλήρης και δεν υπάρχει καμία συναίσθηση για το τι ακριβώς είναι η ανθρώπινη νοημοσύνη. Μέχρι να γίνει αυτό, είναι αδύνατο να είμαστε ακριβείς για το πώς μπορεί να μιμηθεί τεχνητά αυτή η νοημοσύνη. [31]

Η ΤΝ δεν είναι μια νέα έννοια. Η έννοια των μηχανών να έχουν τις δικές τους τη δική τους νοημοσύνη υπάρχει με τη μία ή την άλλη μορφή εδώ και πολλούς αιώνες. Ωστόσο, η έννοια της ΤΝ θα μπορούσε να ειπωθεί ότι έφτασε κατά τη διάρκεια της στο συνέδριο ΤΝ του Ντάρτμουθ το 1956.

Είναι γνωστό ότι η ΤΝ έχει διάφορες εφαρμογές. Η ΤΝ εξακολουθεί να είναι μια εξελισσόμενη τεχνολογία και το πεδίο εφαρμογής της ΤΝ είναι ποικίλο. Υπάρχουν διάφορες απόψεις σχετικά με την ακριβή δυνατότητα εφαρμογής της ΤΝ. Δεν έχει ακόμη ανακοινωθεί η τελική ετυμηγορία αν η ΤΝ είναι σωτήρας ή καταστροφέας. Θέματα ηθικής εγείρονται σε σχέση με το πόσο εύκολα εφαρμόζεται η ΤΝ. Ωστόσο, σε παγκόσμιο επίπεδο έχει παρατηρηθεί ότι οι ένοπλες δυνάμεις έχουν ήδη αρχίσει να εισάγουν αυτή την τεχνολογία κάπως επιλεκτικά στο συνολικό τους αμυντική τους αρχιτεκτονική. Όπως και διάφορες άλλες στρατιωτικές τεχνολογίες,

η ΤΝ έχει τόσο αμυντικές και επιθετικές διαστάσεις. Ως εκ τούτου, μπορεί να μην είναι πάντα σωστό να εξετάζουμε την ΤΝ μόνο ως μια μέσο για την ενίσχυση της θνησιμότητας. Θα μπορούσαν να υπάρχουν διάφορες πτυχές της ΤΝ που θα μπορούσαν να βοηθήσουν την ανάπτυξη/αναβάθμιση των τεχνολογιών που χρησιμοποιούνται για αμυντικούς σκοπούς. [32]

Η ΤΝ ταυτίζεται συχνά με την τελευταία λέξη της τεχνολογίας. Όπως θα δούμε στη συνέχεια, το ΤΝ έχει αποκτήσει τεράστια δυναμική τα τελευταία χρόνια. Ένας από τους σημαντικότερους μοχλούς αυτής της εξέλιξης ήταν η πρόοδος σε έναν συγκεκριμένο τομέα του πεδίου, τη «μηχανική μάθηση» (ML), όπου η καινοτομία οδήγησε σε αυτό που σήμερα ονομάζεται «βαθιά μάθηση» (DL). Είναι αυτή η τεχνολογία που κρύβεται πίσω από τα πρόσφατα ορόσημα, όπως οι υπολογιστές που μπορούν να αναγνωρίζουν πρόσωπα και να παίζουν παιχνίδια. Σε αντίθεση με τις πιο παραδοσιακές προσεγγίσεις όπου τα υπολογιστικά συστήματα εφαρμόζουν σταθερούς κανόνες, οι αλγόριθμοι ML και DL μπορούν να αναγνωρίσουν μοτίβα στα δεδομένα. Μιλάμε επίσης εδώ για «αλγορίθμους που μαθαίνουν μόνοι τους». Πολλοί άνθρωποι που μιλούν σήμερα για την ΤΝ αναφέρονται στην πραγματικότητα σε αυτούς τους αλγορίθμους, και συχνά συγκεκριμένα σε DL. [32]



Σχήμα 7: Γραφικός κύκλος ζωής της ΤΝ [53]

#### 4.1.1 Μηχανική Μάθηση(Machine Learning ML)

Η μηχανική μάθηση(ML) ασχολείται με το ερώτημα πώς να κατασκευάσουμε υπολογιστές που βελτιώνουν αυτόματα το μέσω της εμπειρίας. Πρόκειται για έναν από τους ταχύτερα αναπτυσσόμενους τεχνικούς τομείς σήμερα, που βρίσκεται στο σημείο τομής της επιστήμης των υπολογιστών και της στατιστικής, και στον πυρήνα της ΤΝ και της επιστήμης των δεδομένων. Η πρόσφατη πρόοδος στη μηχανική μάθηση οφείλεται τόσο στην ανάπτυξη νέων αλγορίθμων μάθησης και θεωρίας όσο και στη συνεχιζόμενη έκρηξη της διαθεσιμότητας διαδικτυακών δεδομένων και υπολογισμών χαμηλού κόστους . Η υιοθέτηση μεθόδων μηχανικής μάθησης έντασης δεδομένων συναντάται σε όλη την επιστήμη, την τεχνολογία και το εμπόριο, οδηγώντας σε πιο τεκμηριωμένη λήψη αποφάσεων σε πολλούς τομείς της ζωής, όπως υγειονομική περίθαλψη, μεταποίηση, εκπαίδευση, χρηματοοικονομική μοντελοποίηση, αστυνόμευση και μάρκετινγκ.

Η μηχανική μάθηση είναι ένας κλάδος που επικεντρώνεται σε δύο αλληλένδετα ερωτήματα: Ποιοι είναι οι θεμελιώδεις στατιστικοί-υπολογιστικοί-πληροφοριοθεωρητικοί νόμοι που διέπουν όλα τα συστήματα μάθησης, συμπεριλαμβανομένων των υπολογιστών, ανθρώπων

και οργανισμών; Η μελέτη της μηχανικής μάθησης είναι σημαντική τόσο για την αντιμετώπιση αυτών των θεμελιωδών επιστημονικών και μηχανικών ερωτημάτων όσο και για το εξαιρετικά πρακτικό λογισμικό υπολογιστών που έχει παράγει και εφαρμόζεται σε πολλές εφαρμογές .

Η μηχανική μάθηση έχει προχωρήσει δραματικά τις τελευταίες δύο δεκαετίες, από εργαστηριακή περιέργεια σε μια πρακτική τεχνολογία σε ευρεία εμπορική χρήση . Στο πλαίσιο της ΤΝ, η μηχανική μάθηση έχει αναδειχθεί ως η μέθοδος της επιλογής για την ανάπτυξη πρακτικού λογισμικού για όραση υπολογιστών, αναγνώριση ομιλίας, επεξεργασία φυσικής γλώσσας, έλεγχο ρομπότ και άλλες εφαρμογές. Πολλοί προγραμματιστές συστημάτων ΤΝ αναγνωρίζουν τώρα ότι, για πολλές εφαρμογές, μπορεί να είναι πολύ πιο εύκολο να εκπαιδεύσουν ένα σύστημα δείχνοντάς του παραδείγματα της επιθυμητής συμπεριφοράς εισόδου-εξόδου από το να το προγραμματίσουν χειροκίνητα προβλέποντας την επιθυμητή απόκριση για όλες τις πιθανές εισόδους. Η επίδραση της μηχανικής μάθησης έχει επίσης γίνει ευρέως αισθητή σε όλο το φάσμα της επιστήμης των υπολογιστών και σε μια σειρά από βιομηχανίες που ασχολούνται με θέματα έντασης δεδομένων, όπως οι υπηρεσίες προς τους καταναλωτές, η διάγνωση βλαβών σε πολύπλοκα συστήματα και ο έλεγχος των αλυσίδων εφοδιαστικής . Υπήρξε ένα εξίσου ευρύ φάσμα επιδράσεων στις εμπειρικές επιστήμες, από τη βιολογία έως την κοσμολογία και τις κοινωνικές επιστήμες, καθώς αναπτύχθηκαν μέθοδοι μηχανικής μάθησης για την ανάλυση πειραματικών δεδομένων υψηλής απόδοσης με νέους τρόπους. [33]

#### *4.1.1.1 Τύποι αλγορίθμων για μηχανική μάθηση*

Υπάρχουν πολλοί διαφορετικοί αλγόριθμοι που μπορείτε να χρησιμοποιήσετε στο ML. Το απαιτούμενο αποτέλεσμα είναι αυτό που αποφασίζει ποιος θα χρησιμοποιηθεί. Οι αλγόριθμοι ML ανήκουν χαρακτηριστικά σε έναν από τους δύο τύπους μάθησης: μάθηση με επίβλεψη ή μάθηση χωρίς επίβλεψη. [34]

##### *4.1.1.1.1 Μάθηση με επίβλεψη*

Η «επιβλεπόμενη μάθηση» αναφέρεται στην εργασία με ένα σύνολο δεδομένων εκπαίδευσης με ετικέτες. Για κάθε παράδειγμα στα δεδομένα εκπαίδευσης έχετε ένα αντικείμενο εισόδου και ένα αντικείμενο εξόδου. [34]

##### *4.1.1.1.2 Μάθηση χωρίς επίβλεψη*

Στο αντίθετο άκρο αυτού του φάσματος βρίσκεται η μάθηση χωρίς επίβλεψη, όπου αφήνετε τον αλγόριθμο να βρει ένα κρυμμένο μοτίβο σε ένα φορτίο δεδομένων. Με τη μάθηση χωρίς επίβλεψη δεν υπάρχει σωστή ή λανθασμένη απάντηση- πρόκειται απλώς για την εκτέλεση του αλγορίθμου ML και τη διαπίστωση των μοτίβων και των αποτελεσμάτων που θα προκύψουν. Η μάθηση χωρίς επίβλεψη μπορεί να είναι περισσότερο μια περίπτωση εξόρυξης δεδομένων παρά πραγματικής μάθησης. [34]

#### 4.1.1.2 Χρήσεις της μηχανικής μάθησης

##### 4.1.1.2.1 Λογισμικό

Η ML χρησιμοποιείται στο λογισμικό για βελτίωση της εμπειρίας του χρήστη. Το λογισμικό μαθαίνει από τη συμπεριφορά του χρήστη και προβλέπει τι θέλει να κάνει.

##### 4.1.1.2.2 Ανίχνευση Ανεπιθύμητης Αλληλογραφίας (Spam)

Ταξινομητές μαθαίνουν τι είναι ανεπιθύμητο μήνυμα. Το σύστημα μαθαίνει από την εμπειρία του χρήστη για να αναγνωρίζει καλύτερα τα μελλοντικά μηνύματα.

##### 4.1.1.2.3 Αναγνώριση Φωνής

Η Siri χρησιμοποιεί μοντέλα ML για να αναλύει φράσεις και να εκτελεί ενέργειες. Αν δεν καταλαβαίνει, κάνει αναζήτηση στο Google.

##### 4.1.1.2.4 Χρηματιστηριακές Συναλλαγές

Η ML χρησιμοποιείται για προτάσεις αγοραπωλησίας με βάση ιστορικά δεδομένα. Η αλγοριθμική διαπραγμάτευση πραγματοποιεί χιλιάδες συναλλαγές το δευτερόλεπτο.

##### 4.1.1.2.5 Ρομποτική

Τα ρομπότ μπορούν να μαθαίνουν δεξιότητες όπως το πιάσιμο αντικειμένων και η κίνηση, με αυτόματη ή ανθρώπινη παρέμβαση.

##### 4.1.1.2.6 Ιατρική και Υγειονομική Περίθαλψη

Η ML χρησιμοποιείται στην ανάλυση υγειονομικών δεδομένων και στην υποστήριξη αποφάσεων. Αναλύει ιατρικά δεδομένα για να βοηθήσει τους γιατρούς.

##### 4.1.1.2.7 Διαφήμιση

Οι διαφημιστές χρησιμοποιούν ML για να αναλύουν ενδιαφέροντα και να στέλνουν στοχευμένες διαφημίσεις, βασισμένες σε τοποθεσία και δραστηριότητα.

##### 4.1.1.2.8 Λιανική και Ηλεκτρονικό Εμπόριο

Οι κάρτες πιστότητας και άλλα συστήματα συλλογής δεδομένων χρησιμοποιούν ML για να κατανοούν τις συνήθειες των πελατών.

##### 4.1.1.2.9 Διαδίκτυο των Πραγμάτων (IoT)

Συσκευές συλλέγουν δεδομένα (π.χ. θερμοκρασία) και χρησιμοποιούν ML για να προσαρμόζονται. Υπάρχει ανησυχία για την ασφάλεια, αλλά και δυνατότητες ανάλυσης.

#### 4.1.2 Βαθιά Μάθηση(Deep Learning DL)

Η βαθιά μάθηση είναι ένας κλάδος της μηχανικής μάθησης που προσπαθεί να μοντελοποιήσει υψηλού επιπέδου αφαιρέσεις δεδομένων χρησιμοποιώντας πολλαπλά στρώματα νευρώνων που αποτελούνται από πολύπλοκες δομές ή μετασχηματισμούς μη γραμμικών στοιχείων. Με την αύξηση του όγκου των δεδομένων και της υπολογιστικής ισχύος, τα νευρωνικά δίκτυα με πιο σύνθετες δομές έχουν προσελκύσει ευρεία προσοχή και έχουν εφαρμοστεί σε διάφορους τομείς.

Στον πυρήνα της βαθιάς μάθησης βρίσκονται τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANN), τα οποία αποτελούν τη βασική αρχιτεκτονική πάνω στην οποία βασίζονται όλα τα μοντέλα βαθιάς μάθησης. Τα ANN είναι εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου και επιτρέπουν στα συστήματα να μαθαίνουν και να γενικεύουν μοτίβα μέσα από δεδομένα, [35]

Τα ANN είναι προγράμματα υπολογιστών βιολογικής έμπνευσης που έχουν σχεδιαστεί για να προσομοιώνουν τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος επεξεργάζεται τις πληροφορίες. Τα ANN συλλέγουν τη γνώση τους εντοπίζοντας τα μοτίβα και τις σχέσεις στα δεδομένα και μαθαίνουν (ή εκπαιδεύονται) μέσω της εμπειρίας και όχι μέσω του προγραμματισμού. Ένα ANN σχηματίζεται από εκατοντάδες μεμονωμένες μονάδες, τεχνητούς νευρώνες ή στοιχεία επεξεργασίας, συνδεδεμένες με συντελεστές (βάρη), οι οποίες αποτελούν τη νευρωνική δομή και είναι οργανωμένες σε στρώματα. [36]

##### 4.1.2.1 Αρχιτεκτονικές βαθιάς μάθησης

Η έρευνα της βαθιάς μάθησης επιχειρεί να μοντελοποιήσει δεδομένα μεγάλης κλίμακας με πολλαπλά επίπεδα επεξεργασίας με πολύπλοκες δομές. Έτσι, σε αντίθεση με τις αρχιτεκτονικές της ρηχής μηχανικής μάθησης, οι αρχιτεκτονικές βαθιάς μάθησης αποτελούνται από πολλαπλούς μη γραμμικούς μετασχηματισμούς. Υπάρχει ένας τεράστιος αριθμός παραλλαγών των αρχιτεκτονικών βαθιάς μάθησης και διαφορετικές αρχιτεκτονικές μπορούν να χρησιμοποιηθούν για την αναπαράσταση διαφορετικών πηγών δεδομένων. Για παράδειγμα, το συνελκτικό νευρωνικό δίκτυο (CNN) είναι η πιο δημοφιλής αρχιτεκτονική για την αναγνώριση εικόνων, ενώ το επαναλαμβανόμενο νευρωνικό δίκτυο είναι περισσότερο εφαρμόσιμο σε διαδοχικές εργασίες όπως η αναγνώριση γραφής ή ομιλίας.

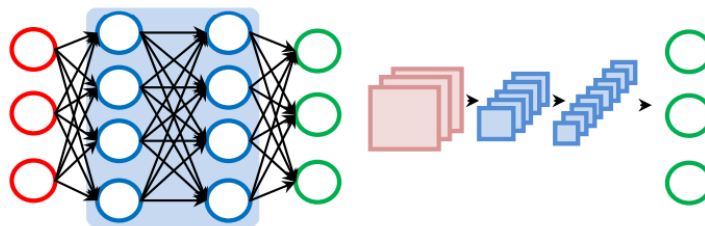
Οι κύριες κατηγορίες είναι: διαφανειακά νευρωνικά δίκτυα (Feedforward Neural Networks - FNN), τα βασικά και απλά δίκτυα όπου η πληροφορία ρέει προς τα εμπρός· συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs), εξειδικευμένα για επεξεργασία εικόνας και βίντεο· επαναληπτικά νευρωνικά δίκτυα (Recurrent Neural Networks - RNNs), που επεξεργάζονται σειρές δεδομένων όπως ομιλία και κείμενο, με παραλλαγές όπως LSTM και GRU. Υπάρχουν επίσης οι μετασχηματιστές (Transformers), που κυριαρχούν σε εφαρμογές επεξεργασίας φυσικής γλώσσας (Native Language Processing - NLP).

Σε αυτή την ενότητα, θα παρουσιάσουμε αυτές τις αρχιτεκτονικές και θα συζητήσουμε γιατί είναι κατάλληλες για τον χειρισμό διαφορετικών δεδομένων.[35]

#### 4.1.2.1.1 Συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNN)

Είναι ένας τύπος τεχνητού νευρωνικού δικτύου με τροφοδότηση προς τα εμπρός, το οποίο χρησιμοποιεί συνέλιξη σε τουλάχιστον ένα από τα στρώματά του. Εμπνεύστηκε από τα βιολογικά νευρωνικά δίκτυα. Το CNN συνδυάζει τα τεχνητά νευρωνικά δίκτυα (και τη διακριτή συνέλιξη για την επεξεργασία εικόνας, η οποία μπορεί να χρησιμοποιηθεί για την αυτόματη εξαγωγή χαρακτηριστικών. Έτσι, είναι ιδιαίτερα σχεδιασμένο για την αναγνώριση δισδιάστατων δεδομένων, όπως οι εικόνες και τα βίντεο. Οι εικόνες μπορούν να χρησιμοποιηθούν απευθείας ως είσοδος του δικτύου, γεγονός που αποφεύγει την πολύπλοκη διαδικασία εξαγωγής χαρακτηριστικών και ανακατασκευής δεδομένων στους παραδοσιακούς αλγορίθμους αναγνώρισης εικόνων.

Τα CNNs μοντελοποιούνται ως συλλογές νευρωνικών μονάδων οι οποίες συνδέονται σε ένα ακυκλικό γράφημα. Στο παρακάτω σχήμα, παρουσιάζεται ένα παράδειγμα ενός CNN με 2 κρυφά στρώματα. Κάθε στρώμα αυτού του δικτύου λαμβάνει ως εισόδους τις εξόδους του του προηγούμενου στρώματος. Επομένως, δεν επιτρέπονται κύκλοι στην αρχιτεκτονική του CNN. Συνήθως, οι μονάδες νευρώνων μεταξύ δύο γειτονικών στρωμάτων είναι. Μόνο για προσωπική χρήση. συνδέονται, αλλά οι νευρώνες εντός ενός στρώματος δεν μοιράζονται καμία σύνδεση. Ένα βαθύ νευρωνικό δίκτυο περιέχει περισσότερα από 2 στρώματα. Μπορεί να υπάρχουν 20 ή περισσότερα στρώματα και κάθε στρώμα μπορεί να είναι λειτουργικά διαφορετικό. [35]



(α) Αρχιτεκτονική CNN

(β) Αναγνώριση εικόνας

**Σχήμα 8: Συνελικτικό νευρωνικό δίκτυο [35]**

Τα CNNs μοντελοποιούνται ως συλλογές νευρωνικών μονάδων που συνδέονται σε ένα ακυκλικό γράφημα. Το Σχήμα 8(α) δείχνει ένα παράδειγμα ενός συνελικτικού νευρωνικού δικτύου με 2 κρυφά στρώματα. Κάθε στρώμα αυτού του δικτύου λαμβάνει τις εξόδους των του προηγούμενου στρώματος ως εισόδους. Επομένως, οι κύκλοι δεν επιτρέπονται στο CNN αρχιτεκτονική. Τυπικά, οι μονάδες νευρώνων μεταξύ δύο γειτονικών στρωμάτων είναι πλήρως συνδεδεμένες, αλλά οι νευρώνες εντός ενός στρώματος δεν έχουν καμία σύνδεση. Ένα βαθύ νευρωνικό δίκτυο περιέχει περισσότερα από 2 στρώματα. Μπορεί να υπάρχουν 20 ή περισσότερα στρώματα και κάθε στρώμα μπορεί να είναι λειτουργικά διαφορετικό.

Ας πάρουμε ως παράδειγμα την αναγνώριση εικόνων, όπου το CNN οργανώνει τους νευρώνες του σε τρία διαστάσεις (πλάτος, ύψος, βάθος), όπως φαίνεται στο Σχήμα 8(β). Η είσοδος του CNN στο Σχ. 8 (β) είναι μια εικόνα, οπότε το πλάτος και το ύψος της είναι οι διαστάσεις της εικόνας και οι βάθος είναι 3 (κόκκινο, πράσινο, μπλε κανάλι). Η έξοδος του CNN

είναι ένα ενιαίο διάνυσμα βαθμολογίες κλάσεων, τοποθετημένο κατά μήκος της διάστασης βάθους. Μεταξύ εισόδου και εξόδου, τρία κύριοι τύποι στρωμάτων χρησιμοποιούνται για την κατασκευή αρχιτεκτονικών CNN, συμπεριλαμβανομένων των συνελκτικών στρώμα, το στρώμα συγκέντρωσης και το πλήρως συνδεδεμένο στρώμα. Και υπάρχει επίσης μια απώλεια συνάρτηση στο τελευταίο στρώμα. Μερικές φορές, άλλοι τύποι στρώματος είναι χρησιμοποιούνται για να βελτιώσουν την απόδοση ενός CNN, όπως το στρώμα εγκατάλειψης που χρησιμοποιείται για να τον έλεγχο του μεγέθους του CNN.[35]

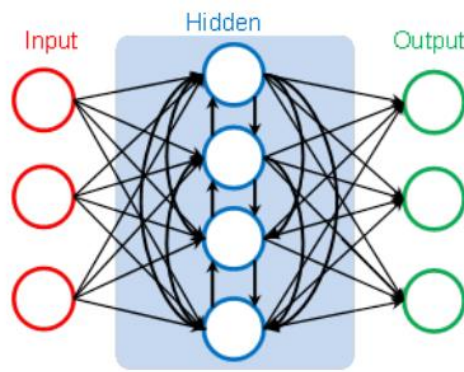
#### 4.1.2.1.2 Αναδρομικά νευρωνικά δίκτυα (Recurrent neural networks - RNN)

Τα RNN αποσκοπούν στην επεξεργασία διαδοχικών δεδομένων. Στο παραδοσιακό μοντέλο του νευρωνικού δικτύου, όπως τα CNN, τα δεδομένα ρέουν από το στρώμα εισόδου στο κρυφό στρώμα και στη συνέχεια στο στρώμα εξόδου. Δεν υπάρχει σύνδεση μεταξύ των νευρώνων στο ίδιο επίπεδο. Ορισμένα προβλήματα δεν μπορούν να αντιμετωπιστούν πλήρως από αυτό το είδος νευρωνικού δικτύου. Αυτή η αρχιτεκτονική δεν μπορεί να λύσει το πρόβλημα όπου τα δεδομένα εισόδου έχουν σχέσεις μεταξύ τους. Για παράδειγμα, χρειαζόμαστε την προηγούμενη λέξη σε μια πρόταση για να προβλέψουμε την επόμενη, επειδή οι λέξεις στην ίδια πρόταση δεν είναι ανεξάρτητες.

Στα RNN, η τρέχουσα έξοδος και η προηγούμενη έξοδος είναι σχετικές. Για να γίνουμε πιο συγκεκριμένοι, η έξοδος του προηγούμενου βήματος αποθηκεύεται και χρησιμοποιείται για τον υπολογισμό της τρέχουσας εξόδου, δηλαδή η είσοδος του δικτύου περιέχει τόσο τα δεδομένα από το στρώμα εισόδου όσο και την έξοδο των κρυφών στρωμάτων από το προηγούμενο βήμα. Αυτό τα καθιστά εφαρμόσιμα σε εργασίες όπως η αναγνώριση γραφής και η αναγνώριση ομιλίας. Η αρχιτεκτονική του RNN παρουσιάζεται στο παρακάτω σχήμα.

Στο σχήμα λοιπόν, εκτός από την ροή δεδομένων από την είσοδο στο κρυφό στρώμα και από το κρυφό στρώμα στην έξοδο, η οποία είναι η ίδια όπως στο CNN, υπάρχει επίσης μεταφορά δεδομένων μεταξύ των νευρώνων στο ίδιο κρυφό στρώμα. Ένα επαναλαμβανόμενο νευρωνικό δίκτυο μπορεί να ξεδιπλωθεί σε ένα άκυκλο νευρωνικό δίκτυο. Για παράδειγμα, ένα ανακυκλούμενο νευρωνικό δίκτυο μιας πρότασης με πέντε λέξεις μπορεί να αναδιπλωθεί σε ένα νευρωνικό δίκτυο πέντε επιπέδων με κάθε επίπεδο να αντιπροσωπεύει μια λέξη.

Είναι κατάλληλα για επεξεργασία ακολουθιακών δεδομένων, όπως κείμενο ή ομιλία, αλλά παρουσιάζουν δυσκολίες στη διατήρηση μακροχρόνιας μνήμης. Για να ξεπεραστούν αυτά τα προβλήματα, αναπτύχθηκαν δύο βελτιωμένες παραλλαγές: τα Long Short-Term Memory (LSTM) και τα Gated Recurrent Units (GRU), τα οποία μπορούν να διαχειριστούν καλύτερα μακροπρόθεσμες εξαρτήσεις στη ροή των δεδομένων. [35]

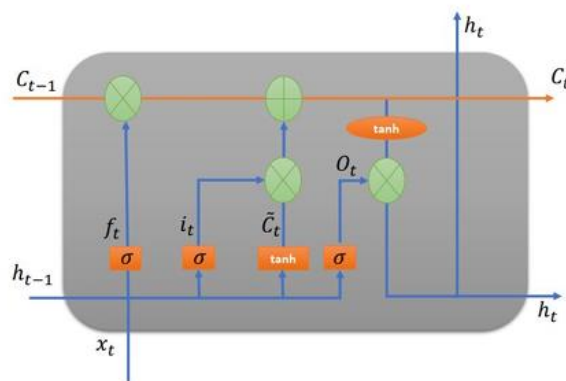


**Σχήμα 9: Ένα επαναλαμβανόμενο νευρωνικό δίκτυο**

Στο Σχήμα 9, εκτός από την ροή δεδομένων από την είσοδο στο κρυφό στρώμα και από το κρυφό στρώμα προς την έξοδο, που είναι η ίδια με το CNN, υπάρχει επίσης μεταφορά δεδομένων μεταξύ των νευρώνων στο ίδιο κρυφό στρώμα.

#### 4.1.2.1.2.1 Μακροχρόνια βραχυπρόθεσμη μνήμη (Long Short Term Memory - LSTM)

Η μακροχρόνια βραχυπρόθεσμη μνήμη (LSTM) είναι πολύ δημοφιλή πρόσφατα. Το RNN μπορεί να αντιμετωπίσει το πρόβλημα όταν υπάρχουν κάποιες σχέσεις ή εξαρτήσεις μεταξύ των δεδομένων εισόδου. Ωστόσο, το RNN μπορεί να χάσει την ικανότητα να μαθαίνει τις συνδέσεις μεταξύ των δεδομένων εισόδου ενώ οι εξαρτήσεις τους είναι πολύ μεγάλες. Για παράδειγμα, δεδομένης μιας πρότασης «τα σύννεφα είναι στο», το RNN μπορεί να προβλέψει την επόμενη λέξη «ο ουρανός», αλλά είναι δύσκολο να προβλέψει την επόμενη λέξη της πρότασης «μεγάλωσα στη Γαλλία». Μιλώ «άπταιστα». Το LSTM προτείνεται για την επίλυση αυτού του προβλήματος. Στο LSTM, οι νευρώνες αντικαθίστανται από μπλοκ. Στο κάθε μπλοκ, υπάρχει ένα κύτταρο που χρησιμοποιείται για να θυμάται τις πληροφορίες από όλα τα δεδομένα εισόδου. [35]

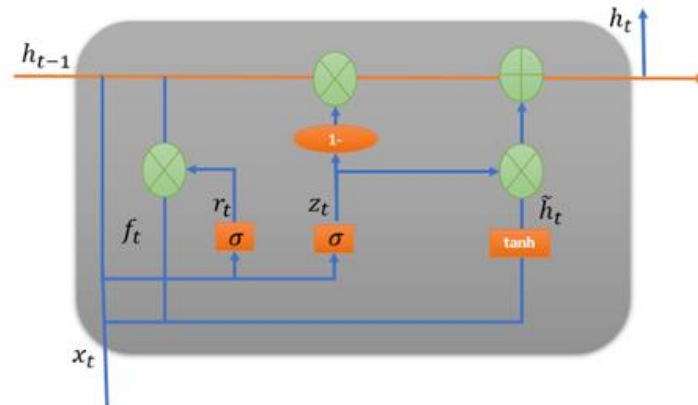


**Σχήμα 10: Η κυτταρική δομή μιας μονάδας μακράς βραχυπρόθεσμης μνήμης (LSTM) [37]**

Το Σχήμα 10 απεικονίζει ένα κύτταρο LSTM, το οποίο χρησιμοποιείται για την επεξεργασία ακολουθιακών δεδομένων, όπως κείμενο ή ήχος, με σκοπό να "θυμάται" και να "ξεχνά" πληροφορίες με έξυπνο τρόπο. Το LSTM φιλτράρει τα δεδομένα, κρατώντας μόνο τις σημαντικές πληροφορίες και απορρίπτοντας τις άχρηστες σε κάθε χρονικό βήμα.

#### 4.1.2.1.2 Πύλη επαναλαμβανόμενης μονάδας (Gated Recurrent Unit - GRU)

Η αναδρομική μονάδα με πύλη είναι ένας ειδικός τύπος βελτιστοποιημένου αναδρομικού νευρωνικού δικτύου βασισμένου σε LSTM. Η εσωτερική μονάδα GRU είναι παρόμοια με την εσωτερική μονάδα LSTM, με τη διαφορά ότι η GRU συνδυάζει τη θύρα εισόδου και τη θύρα λησμονιάς (θύρας που ξεχνάει) στο LSTM σε μια ενιαία θύρα ενημέρωσης.[37]

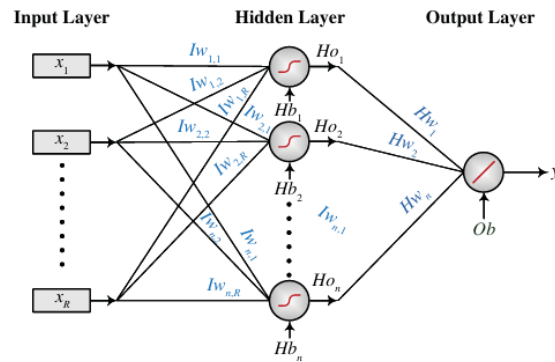


**Σχήμα 11: Η δομή των κυττάρων μιας ελεγχόμενης επαναλαμβανόμενης μονάδας (GRU) [37]**

Το Σχήμα 11, παρουσιάζει το εσωτερικό ενός GRU, ενός «έξυπνου» νευρώνα που μπορεί να διατηρεί πληροφορίες μέσα στον χρόνο, κάτι ιδιαίτερα χρήσιμο για την επεξεργασία κειμένου, ήχου ή σειρών τιμών. Ο σκοπός του GRU είναι να διατηρεί τις σημαντικές πληροφορίες και να αγνοεί τις ασήμαντες, ώστε να μπορεί να κατανοεί τη «ροή» των δεδομένων με μεγαλύτερη ακρίβεια και αποδοτικότητα.

#### 4.1.2.1.3 Νευρωνικά δίκτυα τροφοδότησης (Feedforward Neural Networks – FNN)

Είναι ίσως τα πιο συχνά χρησιμοποιούμενα νευρωνικά δίκτυα και έχουν εφαρμοστεί σε μια ευρεία ποικιλία προβλημάτων, συμπεριλαμβανομένης της προσέγγισης συναρτήσεων, της πρόβλεψης και της προσομοίωσης που προκύπτουν από διάφορους κλάδους, όπως η επιστήμη των υπολογιστών, η μηχανική και η ιατρική. Αν και ένα κρυφό στρώμα είναι επαρκές για να μπορέσουν τα νευρωνικά δίκτυα να προσεγγίσουν οποιαδήποτε δεδομένη συνάρτηση, ορισμένοι ερευνητές υποστήριξαν ότι τα νευρωνικά δίκτυα με περισσότερα από ένα κρυφό στρώμα μπορεί να απαιτούν λιγότερους κρυφούς νευρώνες για να προσεγγίσουν την ίδια συνάρτηση. [38]



**Σχήμα 12:** Νευρωνικό δίκτυο τροφοδότησης ενός κρυφού στρώματος με μία έξοδο

Το Σχήμα 12 δείχνει τη βασική δομή ενός τεχνητού νευρωνικού δικτύου, που αποτελείται από τρία κύρια επίπεδα: το επίπεδο εισόδου (Input Layer), όπου δίνουμε τα αρχικά δεδομένα στο σύστημα, όπως χαρακτηριστικά μιας φωτογραφίας, ηλικία ή βάρος· το κρυφό επίπεδο (Hidden Layer), όπου γίνεται η «σκέψη» του συστήματος, δηλαδή η επεξεργασία των εισόδων μέσω μαθηματικών υπολογισμών· και το επίπεδο εξόδου (Output Layer), το οποίο δίνει το τελικό αποτέλεσμα, όπως η πρόβλεψη αν μια εικόνα δείχνει γάτα ή σκύλο ή πόσο θα κοστίζει ένα σπίτι. Στο εσωτερικό του δικτύου υπάρχουν επίσης βάρη (weights), που καθορίζουν πόση σημασία έχει κάθε είσοδος, και το bias (μετατόπιση), που λειτουργεί σαν μια σταθερή προσθήκη στον υπολογισμό και βοηθά στην καλύτερη απόδοση. Συνοπτικά, δίνουμε δεδομένα στο σύστημα, εκείνο τα επεξεργάζεται μέσω ενδιάμεσων υπολογισμών και μας επιστρέφει μια απόφαση ή πρόβλεψη.

### 4.1.3 Μεγάλα γλωσσικά μοντέλα

#### 4.1.3.1 Επεξεργασία Φυσικής Γλώσσας (Nature language processing - NLP)

Ως σημαντικός κλάδος της ΤΝ, η επεξεργασία φυσικής γλώσσας (NLP) μελετά τις αλληλεπιδράσεις μεταξύ ανθρώπων και υπολογιστών μέσω της φυσικής γλώσσας. Μελετά θεμελιώδεις τεχνολογίες για τις νοηματικές εκφράσεις λέξεων, φράσεων, προτάσεων και εγγράφων, καθώς και για τη συντακτική και σημασιολογική επεξεργασία, όπως η διάσπαση λέξεων, οι συντακτικοί αναλυτές και η σημασιολογική ανάλυση και αναπτύσσει εφαρμογές όπως η μηχανική μετάφραση, η απάντηση ερωτήσεων, η ανάκτηση πληροφοριών, ο διάλογος, η δημιουργία κειμένου και τα συστήματα συστάσεων. Η NLP είναι ζωτικής σημασίας για τις μηχανές αναζήτησης, τα συστήματα υποστήριξης πελατών, την επιχειρηματική νοημοσύνη και τους προφορικούς βοηθούς.

Η NLP είναι ένα υποπεδίο της ΤΝ που επικεντρώνεται στο να επιτρέπει στους υπολογιστές να κατανοούν και να επεξεργάζονται τις ανθρώπινες γλώσσες. Την τελευταία πενταετία, γίναμε μάρτυρες της ραγδαίας ανάπτυξης της NLP σε εργασίες όπως η μηχανική μετάφραση, η απάντηση ερωτήσεων και η μηχανική ανάγνωση κατανόηση με βάση τη βαθιά μάθηση και έναν τεράστιο όγκο σχολιασμένων και μη σχολιασμένων δεδομένων.

Οι τεχνικές NLP αναλύουν τη γλώσσα σε στοιχεία, συμπεριλαμβανομένης της σύνταξης, της σημασιολογίας και της πραγματιστικής, τα οποία μαζί δημιουργούν νόημα. Μέθοδοι όπως το

tokenization, το stemming και η parsing αποτελούν τη βάση της ανάλυσης, επιτρέποντας στις μηχανές να επεξεργάζονται πολύπλοκα γλωσσικά μοτίβα [39]

#### 4.1.3.1.1 *Μετασχηματιστές (Transformers)*

Ο τομέας της Επεξεργασίας Φυσικής Γλώσσας (NLP) έχει υποστεί σημαντική μεταμόρφωση με την εισαγωγή των μετασχηματιστών (transformers). Από την πρώτη εισαγωγή αυτής της τεχνολογίας το 2017, η χρήση των μετασχηματιστών έχει διαδοθεί ευρέως και έχει επηρεάσει βαθιά τον τομέα της NLP.

Ο πυρήνας της αρχιτεκτονικής του μετασχηματιστή, επιτρέπει στο μοντέλο να συλλάβει το πλαίσιο αξιολογώντας το πως η κάθε λέξη σχετίζεται με άλλες, ανεξάρτητα από τη θέση της. Αυτός ο μηχανισμός έχει αποδειχθεί ιδιαίτερα ισχυρός σε εργασίες που απαιτούν διαχείριση εξαρτήσεων μεγάλης εμβέλειας, όπως η ταξινόμηση και η σύνοψη εγγράφων. Μετασχηματιστές όπως το BERT και το GPT έχουν επιτύχει σημαντικές επιδόσεις σε διάφορα σημεία αναφοράς NLP, κυρίως λόγω της ικανότητάς τους να μαθαίνουν από διαφοροποιημένες σχέσεις που υπάρχουν μέσα στη γλώσσα [40].

#### 4.1.3.1.2 *Μοντέλο BERT*

Μετά την επιτυχία των μετασχηματιστών στη μηχανική μετάφραση με την αρχική διατύπωση από ακολουθία σε ακολουθία, οι ερευνητές άρχισαν να διερευνούν τη χρήση τους σε άλλες εργασίες NLP. Μια από τις σημαντικότερες εξελίξεις προς αυτή την κατεύθυνση ήταν η εισαγωγή του BERT (Bidirectional Encoder Representations from Transformers).

Το BERT είναι ένα προ-εκπαιδευμένο μοντέλο μετασχηματιστή που μπορεί να ρυθμιστεί λεπτομερώς για ένα ευρύ φάσμα εργασιών NLP, όπως η ανάλυση συναισθήματος, η αναγνώριση ονομαστικών οντοτήτων και η απάντηση ερωτήσεων. Το BERT εκπαιδεύτηκε χρησιμοποιώντας μια αντιθετική εργασία, όπου του ζητήθηκε να προβλέψει τα ελλείποντα tokens σε μια πρόταση, λαμβάνοντας υπόψη το πλαίσιο των γύρω tokens. Αυτή η προσέγγιση επέτρεψε στην BERT να μάθει πλούσιες αναπαραστάσεις των λέξεων με βάση το περιεχόμενο, καθιστώντας την ιδιαίτερα αποτελεσματική για ένα ευρύ φάσμα εργασιών NLP. [40]

#### 4.1.3.1.3 *Μοντέλο GPT*

Μια άλλη εξέλιξη στον τομέα των μετασχηματιστών ήταν η εισαγωγή του GPT (Generative Pretrained Transformer). Το GPT είναι ένα παραγωγικό μοντέλο που εκπαιδεύεται σε ένα μεγάλο σώμα κειμένου με στόχο την πρόβλεψη του επόμενου σημείου σε μια ακολουθία, λαμβάνοντας υπόψη το πλαίσιο των γύρω σημείων. Το GPT έχει αποδειχθεί ιδιαίτερα αποτελεσματικό για εργασίες όπως η παραγωγή κειμένου, η μοντελοποίηση γλώσσας και η απάντηση ερωτήσεων. Σε αντίθεση με το BERT, το οποίο εκπαιδεύτηκε χρησιμοποιώντας ένα αντιπαραθετικό έργο, το GPT εκπαιδεύτηκε χρησιμοποιώντας ένα γενετικό έργο, επιτρέποντάς του να μάθει μια πιο ποικιλόμορφη και πλήρη αναπαράσταση της γλώσσας.[40]

## **4.2 Τεχνολογίες TN**

Επί του παρόντος, διάφορες σύγχρονες τεχνολογίες TN (έχοντας κατά νου το μέλλον) θα μπορούσαμε να πούμε ότι περιλαμβάνουν τα εξής:

### **4.2.1 Ρομποτική**

Οι μηχανές που εκτελούν εργασίες χειριζόμενες τον φυσικό κόσμο. Πολλές φορές, ένα μεμονωμένο ρομπότ αποτελείται από διάφορα συστήματα που συνεργάζονται μεταξύ τους. Τέτοια συστήματα χρησιμοποιούνται για την εκτέλεση επιχειρήσεων, τη συλλογή πληροφοριών και το ως μη επανδρωμένα συστήματα για την εκτέλεση διαφόρων καθηκόντων, από την παράδοση εμπορικών προϊόντων στους πελάτες μέχρι τη μάχη σε πολέμους.

### **4.2.2 Εξόρυξη δεδομένων**

Ανακάλυψη χρήσιμων πληροφοριών μέσα σε μια βάση δεδομένων που μπορούν στη συνέχεια να χρησιμοποιηθούν για τη βελτίωση των ενεργειών. Στόχος είναι η εξαγωγή μοτίβων από τα δεδομένα και η μετατροπή των δεδομένων σε πληροφορίες. Οι εισροές αυτές έχουν εφαρμογή σε τομείς όπως οι επιχειρήσεις, η επιστημονική έρευνα και η άμυνα.

### **4.2.3 Υπολογιστική Όραση**

Εδώ, η ιδέα είναι ότι οι υπολογιστές θα πρέπει να είναι σε θέση να αντιλαμβάνονται τα αντικείμενα . Η όραση περιλαμβάνει την απόκτηση και την επεξεργασία οπτικών πληροφοριών. Τέτοιες τεχνολογίες με δυνατότητα TN μπορούν να οδηγούν οχήματα ανεξάρτητα, να αναγνωρίζουν και να ερμηνεύουν την ομιλία ή τις εκφράσεις του προσώπου και επίσης να κάνουν αναγνώριση χειρόγραφων εγγράφων . Επιπλέον, οι διαδικασίες όπως η παραγωγή φυσικής γλώσσας βοηθούν την παραγωγή κειμένου από δεδομένα υπολογιστή, ενώ η αναγνώριση ομιλίας επιτρέπει τη μεταφορά γραφής και τη μετατροπή της ανθρώπινης ομιλίας σε μορφή χρήσιμη για εφαρμογές υπολογιστών .

### **4.2.4 Εικονικοί πράκτορες**

Από τα απλά chatbots έως τα προηγμένα συστήματα που μπορούν να δικτυωθούν με ανθρώπους.

### **4.2.5 Υλικό βελτιστοποιημένο για την TN**

Μονάδες επεξεργασίας γραφικών και συσκευές ειδικά σχεδιασμένες και αρχιτεκτονικά σχεδιασμένες για την αποτελεσματική εκτέλεση υπολογιστικών εργασιών προσανατολισμένων στην TN.

### **4.2.6 Διαχείριση αποφάσεων**

Μηχανές που εισάγουν κανόνες και λογική σε συστήματα TN και που χρησιμοποιούνται για την αρχική ρύθμιση/εκπαίδευση και τη συνεχή συντήρηση και ρύθμιση. Αυτό θεωρείται από το ως ώριμη τεχνολογία TN. Χρησιμοποιείται σε μεγάλη ποικιλία επιχειρησιακών εφαρμογών , υποβοηθώντας ή εκτελώντας αυτοματοποιημένη λήψη αποφάσεων.

#### **4.2.7 Πλατφόρμες βαθιάς μάθησης**

Ένας ειδικός τύπος μηχανικής μάθησης που αποτελείται από τεχνητά νευρωνικά δίκτυα με πολλαπλά επίπεδα αφαίρεσης.

#### **4.2.8 Βιομετρία**

Επιτρέπουν πιο φυσικές αλληλεπιδράσεις μεταξύ ανθρώπων και μηχανών, συμπεριλαμβανομένων, μεταξύ άλλων, της αναγνώρισης εικόνας και αφής, της ομιλίας και της γλώσσα του σώματος. Η βιοπληροφορική επιτρέπει τη διαχείριση βιολογικών πληροφοριών. Η ΤΝ προσφέρει αρκετούς ισχυρούς αλγορίθμους και τεχνικές για την επίλυση σημαντικών προβλημάτων στη βιοπληροφορική.

#### **4.2.9 Ανάλυση κειμένου και NLP**

Η επεξεργασία φυσικής γλώσσας (NLP) χρησιμοποιεί και υποστηρίζει ports text analytics διευκολύνοντας την κατανόηση της δομής των προτάσεων και νοήματος, του συναισθήματος και της πρόθεσης μέσω στατιστικών μεθόδων και μεθόδων μηχανικής μάθησης ods. Χρησιμοποιείται επί του παρόντος στην ανίχνευση απάτης και την ασφάλεια, σε ένα ευρύ φάσμα αυτοματοποιημένων βοηθών και σε εφαρμογές για την εξόρυξη αδόμητων δεδομένων.

#### **4.2.10 Αναπαράσταση γνώσης και συλλογιστική**

Δύο έννοιες, κεντρικές στο πεδίο της ΤΝ, περιλαμβάνουν την αναπαράσταση της γνώσης και τις διαδικασίες συλλογισμού που ζωντανεύουν τη γνώση . Η αναπαράσταση της γνώσης αφορά την κωδικοποίηση του πραγματικού κόσμου, κοινής λογικής κ.λπ. σε μια μορφή που είναι τόσο αναγνώσιμη όσο και κατανοητή από τον υπολογιστή . Σε γενικές γραμμές, η μηχανή θα πρέπει να είναι ικανή να σκέφτεται λογικά. [32]

# 5

## *Χρήσεις της TN στην Κυβερνοασφάλεια*

Ο αυξανόμενος αριθμός και η πολυπλοκότητα των επιθέσεων στον κυβερνοχώρο έχουν υπογραμμίσει την επείγουσα ανάγκη για καινοτόμες λύσεις για την ενίσχυση της ασφάλειας των ψηφιακών υποδομών. Μεταξύ αυτών των λύσεων, η TN αναδεικνύεται ως μια πολλά υποσχόμενη τεχνολογία με τη δυνατότητα να βελτιώσει σημαντικά την ασφάλεια στον κυβερνοχώρο και την αντιμετώπιση περιστατικών. Έχει την ικανότητα να βελτιώσει την ταχύτητα και την ακρίβεια της ανίχνευσης απειλών, της αντίδρασης και του μετριασμού, μειώνοντας παράλληλα τον φόρτο εργασίας των επαγγελματιών ασφαλείας. [42]

Η ασφάλεια στον κυβερνοχώρο είναι αναμφισβήτητα ο κλάδος που θα μπορούσε να επωφεληθεί περισσότερο από την εισαγωγή της TN. Όταν τα συμβατικά συστήματα ασφαλείας μπορεί να είναι αργά και ανεπαρκή, οι τεχνικές TN μπορούν να βελτιώσουν τις συνολικές επιδόσεις τους στον τομέα της ασφάλειας και να παρέχουν καλύτερη προστασία από τον αυξανόμενο αριθμό εξελιγμένων απειλών στον κυβερνοχώρο. Εκτός από τις μεγάλες ευκαιρίες που αποδίδονται στην TN στο πλαίσιο της κυβερνοασφάλειας, η χρήση της έχει δικαιολογημένους κινδύνους και ανησυχίες. Για την περαιτέρω αύξηση της ωριμότητας της ασφάλειας στον κυβερνοχώρο, απαιτείται μια ολιστική θεώρηση του κυβερνοπεριβάλλοντος των οργανισμών στην οποία η TN θα συνδυάζεται με την ανθρώπινη διορατικότητα, δεδομένου ότι ούτε οι άνθρωποι ούτε η TN από μόνες τους έχουν αποδείξει συνολική επιτυχία στον τομέα αυτό. Έτσι, η κοινωνικά υπεύθυνη χρήση των τεχνικών TN θα είναι απαραίτητη για τον περαιτέρω μετριασμό των σχετικών κινδύνων και ανησυχιών. [41]

Σε αυτή την ενότητα θα επικεντρωθούμε στο ρόλο της TN σε βασικούς τομείς της ασφάλειας στον κυβερνοχώρο και της αντιμετώπισης περιστατικών, συγκεκριμένα την αξιολόγηση ευπαθειών, την ανίχνευση και πρόληψη εισβολών και την ανάλυση ψηφιακής εγκληματολογίας. Τις κακόβουλες και τις ηθικές χρήσεις της TN με άλλα λόγια.

Η ΤΝ, με τις έμφυτες ικανότητές της, μπορεί να αλλάξει τα δεδομένα, δίνοντας τη δυνατότητα στους οργανισμούς να ανιχνεύουν, να ανταποκρίνονται και να μετριάσουν αποτελεσματικότερα τις απειλές. Ωστόσο, η ΤΝ δεν είναι μια ασημένια σφαίρα για την ασφάλεια στον κυβερνοχώρο. Όπως κάθε τεχνολογία, διαθέτει τους περιορισμούς της και τα πιθανά τρωτά σημεία της.

## ***5.1 Κακόβουλες χρήσεις της ΤΝ στην ασφάλεια στον κυβερνοχώρο***

Η διασταύρωση της ΤΝ και της ασφάλειας στον κυβερνοχώρο φέρνει επανάσταση στην ψηφιακή ασφάλεια με την αυτοματοποιημένη ανίχνευση απειλών, την ανίχνευση ανωμαλιών σε πραγματικό χρόνο και την προγνωστική ανάλυση. Ωστόσο, ενώ η ΤΝ ενισχύει τις άμυνες κατά των εγκληματιών του κυβερνοχώρου, αυτοί οι ίδιοι οι εγκληματίες στρέφουν τώρα τα όπλα τους στην ΤΝ και τη χρησιμοποιούν για να αναπτύξουν αντίπαλες επιθέσεις ΤΝ, καθιστώντας τα ψηφιακά συστήματα πιο ευάλωτα από ποτέ. Η αντίπαλη ΤΝ αναφέρεται στις διάφορες τακτικές που εκτελούνται εναντίον ενός μοντέλου εκμάθησης μηχανής χειραγωγώντας τις εισόδους του, αλλοιώνοντας τα δεδομένα του ή εξαπατώντας ένα σύστημα ασφαλείας που υποστηρίζεται από ΤΝ.

Με τα εργαλεία ασφάλειας ΤΝ, όπως τα συστήματα ανίχνευσης εισβολών, οι αλγόριθμοι ανίχνευσης κακόβουλου λογισμικού και τα αυτοματοποιημένα εργαλεία αντιμετώπισης περιστατικών να θεωρούνται όλο και περισσότερο ως πρότυπο, εξελίσσονται επίσης οι στρατηγικές των επιτιθέμενων για να τα αποφύγουν, να τα αλλοιώσουν ή να τα χειραγωγήσουν. Σε αντίθεση με τις παραδοσιακές απειλές στον κυβερνοχώρο, οι αντίπαλες απειλές ΤΝ είναι τόσο δυναμικές όσο και προσαρμοστικές, ρίχνοντας φως στις επιθέσεις με δυνατότητες που δεν είχαν παρατηρηθεί μέχρι σήμερα.

Η ανάπτυξη των απειλών στον κυβερνοχώρο που βασίζονται στην ΤΝ επιφέρει την ανάγκη επαναπροσδιορισμού των τεχνικών ασφαλείας, επιμένοντας σε προηγμένες τεχνικές αντίπαλης άμυνας, δεοντολογική χρήση της ΤΝ και συνεχή παρακολούθηση της ασφάλειας έναντι αναδυόμενων απειλών. [46]

### ***5.1.1 Αντίπαλη ΤΝ στις απειλές κυβερνοασφάλειας***

Οι επιθέσεις αντίπαλης ΤΝ περιλαμβάνουν γενικά πολλαπλές μορφές, όπου επιθέσεις όπως η αποφυγή ξεγελά ένα σύστημα ανίχνευσης βασισμένο στην ΤΝ ενώ οι επιθέσεις δηλητηρίασης αποσκοπούν στην εισαγωγή κακόβουλων δεδομένων στο μοντέλο. Όλες αυτές οι επιθέσεις εγκυμονούν σοβαρούς επιχειρηματικούς, εθνικούς ή ακόμη και προσωπικούς κινδύνους. Οι τεχνικές αντίπαλης ΤΝ έχουν τελειοποιήσει την εκμετάλλευση των εγγενών αδυναμιών των μοντέλων μηχανικής μάθησης, καθιστώντας δύσκολη την προσαρμογή τους με τη χρήση παραδοσιακών αμυντικών συστημάτων ασφαλείας στον κυβερνοχώρο σε πραγματικό χρόνο. Αυτό αναδεικνύει την επείγουσα ανάγκη για αυτοματοποιημένες λύσεις ασφαλείας στον κυβερνοχώρο με ΤΝ που μπορούν να εντοπίζουν προληπτικά και να μετριάσουν τις αντίπαλες απειλές.

Ακόμη και άνθρωποι με «φτωχή» ικανότητα αντιγραφής υπηρεσιών μπορούν να εξαπολύσουν σήμερα κινδύνους στον κυβερνοχώρο. Το χάος στο εσωτερικό της βιομηχανίας έχει επηρεάσει όχι μόνο εταιρείες και κυβερνητικούς αξιωματούχους αλλά και ιδιώτες – αυτές οι απειλές στον κυβερνοχώρο που βασίζονται στην TN έχουν προκαλέσει χάος σε βιομηχανίες σε όλο τον κόσμο. Αυτές οι διαδικτυακές απειλές κυβερνοασφάλειας δίνουν έμφαση ότι η αντίπαλη TN είναι στην πραγματικότητα μια ασπίδα που ενισχύει το και δοξάζει την αρένα του κυβερνοεγκλήματος όσον αφορά την αποτελεσματικότητα, την αυτοματοποίηση και την πολυπλοκότητα των επιθέσεων. [46]

Η TN ωφελεί τον κόσμο της κυβερνοασφάλειας, ωστόσο οι εγκληματίες του κυβερνοχώρου μπορούν να κάνουν κατάχρηση της δύναμής της για να δημιουργήσουν κινδύνους για όλους . Οι άμυνες κυβερνοασφάλειας χρησιμοποιούν ακριβώς την ίδια τεχνολογία που διαθέτει η TN, την οποία οι κυβερνοεπιτιθέμενοι χρησιμοποιούν ως όπλο για να εξαπολύσουν τις επιθέσεις τους. Οι εγκληματίες του κυβερνοχώρου χρησιμοποιούν επιθέσεις για να τροποποιήσουν τον κώδικα κακόβουλου λογισμικού τους αυτόματα για την αποφυγή της ασφάλειας και χρησιμοποιούν εχθρικές μεθόδους TN για να χειραγωγήσουν τις άμυνες μηχανικής μάθησης. Λόγω αυτού του σκοπού της, η τεχνολογία TN έχει οδηγήσει σε έναν συνεχή ανταγωνισμό μεταξύ όσων εργάζονται στην κυβερνοασφάλεια και των κυβερνοεγκληματιών. Τα εργαλεία κυβερνοασφάλειας με βάση την TN αναπτύσσουν προηγμένες δυνατότητες ωστόσο οι φορείς απειλών αναπτύσσουν ενεργά εξελιγμένες μεθόδους εκβιασμού στον κυβερνοχώρο που χρησιμοποιούν τεχνολογίες TN .

#### *5.1.1.1 Επιθέσεις Ransomware με εξελιγμένη TN*

Οι επιτιθέμενοι αναπτύσσουν αυτοεκπαιδευόμενο ransomware που εξελίσσεται σε πραγματικό χρόνο για να παρακάμψουν τις άμυνες της κυβερνοασφάλειας.[46]

Το Ransomware είναι ένας τύπος κακόβουλου λογισμικού που έχει σχεδιαστεί για να διευκολύνει διάφορες κακόβουλες δραστηριότητες, όπως η παρεμπόδιση της πρόσβασης σε προσωπικά δεδομένα εκτός αν καταβληθούν λύτρα. Αυτά τα λύτρα χρησιμοποιούν συνήθως κρυπτονόμισμα όπως το Bitcoin, το οποίο καθιστά δύσκολο να εντοπίσει τον παραλήπτη της συναλλαγής και είναι ιδανικό για τους επιτιθέμενους να αποφεύγουν τις υπηρεσίες επιβολής του νόμου . [47]

Οι νέες επιθέσεις ransomware δείχνουν ότι οι επιχειρήσεις εκβιασμού στον κυβερνοχώρο ενσωματώνουν πλέον σε μεγάλο βαθμό συστήματα TN στις τακτικές τους. Τα NotPetya και Ryuk και LockBit (τύποι κακόβουλου λογισμικού) αντιπροσωπεύουν τρία διαβόητα στελέχη ransomware που πέτυχαν μέσω αυτοματοποιημένων μεθόδων κρυπτογράφησης να δημιουργήσουν τη μέγιστη δυνατή αναστάτωση .

Αρχικά το NotPetya λειτουργούσε ως ransomware, αλλά οι ειδικοί ασφαλείας επιβεβαίωσαν ότι λειτουργούσε ως κακόβουλο λογισμικό που επιτέθηκε σε χρηματοπιστωτικά ιδρύματα μαζί με κρίσιμες υποδομές.

Το ransomware Ryuk διεξάγει εξειδικευμένες επιθέσεις σε μεγάλες επιχειρήσεις μέσω TN που το καθιστά ικανό να εντοπίζει τους βέλτιστους στόχους χρησιμοποιώντας μηχανική μάθηση παράλληλα με αυτοματοποιημένες μεθόδους συζήτησης λύτρων.

Χρησιμοποιώντας την αυτο-διάδοση το LockBit ransomware επεκτείνει αυτόματα την εξάπλωσή του σε όλα τα περιβάλλοντα δικτύου. Οι επιτιθέμενοι χρησιμοποιούν τεχνολογία deepfake για να παράγουν συνθετικές ηχογραφήσεις φωνής και βίντεο που χρησιμοποιούν εναντίον των θυμάτων με αποτέλεσμα την εκβίαση λύτρων. Οι επιθέσεις ransomware συνεχίζουν να εξελίσσονται σε εξελιγμένες απειλές που απαιτούν προληπτικά αντίμετρα με TN λόγω των υφιστάμενων περιστατικών. [48]

#### *5.1.1.2 Deepfake και παραπληροφόρηση που παράγεται από TN*

Η TN κάνει πλασματική οπτικοακουστική δημιουργία και σύνθεση φωνής που μπορεί να διαστρεβλώσει τις χρηματοπιστωτικές αγορές, παραπληροφορεί τους ανθρώπους και να υποδυθεί περαιτέρω οποιονδήποτε μέσα στο δημόσιο μάτι ή στη σφαίρα.[46]

«Ψευδείς ειδήσεις» είναι οι πληροφορίες που αναπτύσσονται και διαδίδονται για να παραπλανήσουν το κοινό στο οποίο απευθύνονται για κάποιο κοινωνικοπολιτικό ή οικονομικό όφελος. Με την ευρύτερη δυνατή έννοια, θα μπορούσε κανείς να σκεφτεί ότι ο όρος σημαίνει κάτι περισσότερο από το περιεχόμενο που περιέχει οποιονδήποτε βαθμό επιδιωκόμενης πραγματικής ανακρίβεια, που κυμαίνεται από τροποποιημένες λεπτομέρειες έως πλήρη κατασκευή. Υπάρχει κάποιος βαθμός αξίας σε αυτόν τον ευρύ ορισμό στο βαθμό που δίνει έμφαση στην κατευθυνόμενη χειραγώγηση των πληροφοριών. Αλλά όπως έχουν επισημάνει αρκετοί μελετητές, ο όρος χρησιμοποιείται συχνά για να υποδηλώσει τόσες πολλές διακριτά φαινόμενα ώστε να έχει περιορισμένη χρησιμότητα για αυστηρά εμπειρικά ερευνητικά προγράμματα που αποτυγχάνουν να ορίσουν αυστηρές παραμέτρους ορισμού σημειώνουν (ή παραπέμπουν σε παρατηρήσεις άλλων) ότι ο όρος έχει χρησιμοποιηθεί ποικιλοτρόπως από ειδήμονες και μελετητές για να συζητήσουν ένα τεράστιο φάσμα δραστηριοτήτων μόλις τα τελευταία χρόνια, συμπεριλαμβανομένων της σάτιρας, της παρωδίας, της προπαγάνδας, της χειραγωγικής διαφήμισης, της ψευδούς ισορροπίας (εννοώντας ειδήσεις ανάξιες της λαϊκής εστίασης) και disinformata (δηλαδή πληροφορίες που χρησιμοποιούνται για να σπείρουν αμφιβολία), καθώς και δραστηριότητες τόσο διακριτές όσο η παραπληροφόρηση, η παραπληροφόρηση και η κακή πληροφόρηση. Στο εξής, χρησιμοποιώ απλώς τον όρο «παραπληροφόρηση» για να περιγράψω τη δυνητική χρήση μιας ενιαίας νέας τεχνολογίας (π.χ. DeepFakes) για ευρείας κλίμακας χειραγώγηση πληροφοριών. [49]

Τα DeepFakes επιλύουν έναν βασικό ανασταλτικό παράγοντα στην παραγωγή ψηφιακής πλαστογραφίας - την ανάγκη χειροκίνητης παραγωγής και ελέγχου της προέλευσης των πλαστών μέσων. Αντίθετα, το περιεχόμενο εικόνας και βίντεο DeepFake αναπτύσσεται με τη χρήση ενός ζεύγους μοντέλων νευρωνικών δικτύων. Αρχικά, ένα πρόσωπο μελετάται από τον αλγόριθμο και συγκρίνεται με το πρόσωπο ενός άλλου ατόμου που κάνει ανάλογες εκφράσεις και κινήσεις. Όσο περισσότερα δεδομένα εκπαίδευσης έχει να χρησιμοποιήσει ο αλγόριθμος, τόσο καλύτερα είναι τα αποτελέσματα αυτής της διαδικασίας. Για την παραγωγή ακριβών κατασκευών ενός προσώπου ώστε τα χαρακτηριστικά του να μπορούν να εφαρμοστούν στο άλλο, ο αλγόριθμος γενικεύει για κάθε πρόσωπο σε γενικές γραμμές. Με τον τρόπο αυτό επιτρέπει τη δημιουργία πιο αξιόπιστων συνθετικών περιεχομένου προσώπου απ' ό,τι αν το πρόγραμμα προσπαθούσε απλώς να κλωνοποιήσει συγκεκριμένα εκφράσεις σε ένα δεύτερο πρόσωπο. Σε αυτό το σημείο, το δεύτερο νευρωνικό δίκτυο πιάνει δουλειά. Αυτός ο αλγόριθμος δεν είναι παραγωγικός αλλά μάλλον στοχεύει στον εντοπισμό κατασκευών. Η χρήση δύο αλγορίθμων με αυτόν τον τρόπο είναι μια

σχετικά νέα εξέλιξη που ονομάζεται γεννητική αντιθετική μοντελοποίηση. Χρησιμοποιώντας την έξοδο του δικτύου ανίχνευσης πλαστογραφίας, το παραγωγικό τμήμα του προγράμματος δεν παύει να παράγει υψηλότερες και υψηλότερης πιστότητας κατασκευές έως ότου είναι πιθανολογικά σίγουρο ότι παραποιεί το αντίστοιχο πρόγραμμα.

Έτσι, τα DeepFakes δεν είναι απλώς προϊόντα προηγμένων τεχνικών παραγωγής περιεχομένου - είναι φυσικά προσαρμοστικά και ανθεκτικά σε αυτό που θα μπορούσαμε να θεωρήσουμε ως προσπάθειες εμβολιασμού της κοινωνίας με την απλή δημιουργία καλύτερων μεθόδων για την ανίχνευσή τους. Δεδομένου του ότι οι τρέχουσες μέθοδοι για την ανίχνευση των DeepFakes βασίζονται σε μεγάλο βαθμό σε σφάλματα στον τρόπο με τον οποίο οι αλγόριθμοι εκτελούν το έργο τους - για παράδειγμα, εξετάζοντας αποκλίσεις στα μοτίβα ανοιγοκλείσιμου ματιού με την υπόθεση ότι τα τρέχοντα προγράμματα συχνά δεν έχουν αρκετά δεδομένα για να γενικεύσουν τέλεια για μια τόσο διαφοροποιημένη συμπεριφορά - φαίνεται ότι δεν είναι ασφαλής υπόθεση ότι τέτοιες τεχνικές θα μπορούσαν να λειτουργήσουν στο μέλλον. Όπως το έθεσε ένας ερευνητής, «σε κάποιο σημείο είναι πιθανό ότι δεν θα είναι δυνατόν να ανιχνεύσουμε καθόλου [ψεύτικες TN]». [49]

#### 5.1.1.3 *AI-Robust Phishing*

Έχει παρατηρηθεί ότι οι επιτιθέμενοι στα σύγχρονα χρόνια μετατρέπονται σε TN powered phishing, όπου τα προσωπικά μηνύματα ηλεκτρονικού ψαρέματος είναι δυσδιάκριτα από ένα νόμιμο.[46]

Η TN έχει ουσιαστικά μεταμορφώσει τον παραδοσιακό κύκλο ζωής των εκστρατειών phishing, ενσωματώνοντας τις μεθοδολογίες που εμπλέκονται, επινοώντας έτσι νέες προκλήσεις και ευκαιρίες για την ανίχνευση απειλών σε πραγματικό χρόνο. Η TN βελτιώνει την αποτελεσματικότητα και την αποδοτικότητα των εκστρατειών phishing μέσω της αυτοματοποίησης και της βελτιστοποίησης των αντίστοιχων σταδίων επίθεσης, ξεκινώντας από την επιλογή στόχου έως τη δημιουργία και την αποστολή του phishing περιεχομένου. Οι ακόλουθες ενότητες συζητούν με περισσότερες λεπτομέρειες πώς η TN επηρεάζει τον κύκλο ζωής μιας εκστρατείας phishing και τις επιπτώσεις για την ανίχνευση απειλών σε πραγματικό χρόνο.[50]

##### 5.1.1.3.1 *Αυτοματοποίηση των Εκστρατειών Phishing*

Η TN στις εκστρατείες phishing έχει αλλάξει τον τρόπο με τον οποίο γίνονταν στο παρελθόν χειροκίνητα. Για παράδειγμα, η TN μπορεί να αξιοποιηθεί για τη συλλογή δεδομένων από διαδικτυακά προφίλ και κοινωνικές δραστηριότητες πιθανών στόχων. Ως συνέπεια, τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για τη δημιουργία πολύ πειστικών phishing email, τα οποία προσαρμόζονται στις προτιμήσεις του χρήστη και μεγιστοποιούν το ποσοστό κλικ από τα θύματα. Αυτή η αυτοματοποίηση των εργασιών επιτρέπει στους επιτιθέμενους να κλιμακώσουν ευκολότερα τις εκστρατείες τους, επηρεάζοντας ευρύτερο φάσμα στόχων με λιγότερη προσπάθεια. [50]

#### 5.1.1.3.2 Ενισχυμένη Δημιουργία Phishing Περιεχομένου

Αυτό είναι το καλύτερο παράδειγμα για το πώς τα επιθετικά εργαλεία που βασίζονται σε αλγόριθμους TN δημιουργούν phishing περιεχόμενο παρόμοιο με νόμιμη επικοινωνία. Η TN εκπαιδεύεται ώστε να αναγνωρίζει τα μοτίβα, τη γλώσσα και τα σχεδιαστικά στοιχεία σε μια τεράστια συλλογή email (ή ιστοσελίδων ή οποιασδήποτε άλλης μορφής επικοινωνίας) που αντιπροσωπεύουν νόμιμο περιεχόμενο. Αυτό δίνει πρόσβαση στους επιτιθέμενους για τη δημιουργία phishing email και ιστοσελίδων που είναι σχεδόν αδιαχώριστα από τα πραγματικά, αυξάνοντας τις πιθανότητες επιτυχίας. Έτσι, οι παλιές τεχνικές που βασίζονταν αποκλειστικά στις διαφορές μεταξύ νόμιμου και phishing περιεχομένου για την ανίχνευση, χάνουν πλέον την αποτελεσματικότητά τους. [50]

#### 5.1.1.3.3 Δυναμική Προσαρμογή και Μάθηση

Οι δυνατότητες συνεχούς βελτίωσης που διαθέτουν τα συστήματα TN ξεπερνούν ένα ορισμένο όριο μετά την εκτέλεση του αλγορίθμου τους σε νέες εργασίες για κάποιο χρονικό διάστημα. Με την αξιοποίηση εργαλείων phishing βασισμένων σε TN, οι επιτιθέμενοι μπορούν να μαθαίνουν από τις αντιδράσεις σε προηγούμενες επιθέσεις και συνεπώς να εξελίσσονται συνεχώς από τις επιτυχίες και τις αποτυχίες τους, γεγονός που τους βοηθά να προετοιμάσουν καλύτερες στρατηγικές στις επόμενες εκστρατείες. Αν μια συγκεκριμένη μορφή phishing email είναι επιτυχής, η TN μπορεί να την μιμηθεί και σε μελλοντικές επιθέσεις. Αυτή η πτυχή της μάθησης καθιστά τις phishing επιθέσεις αδιαπέραστες για τις συμβατικές τεχνικές ανίχνευσης, καθώς το μοντέλο TN έχει εξελιχθεί και συνεχίζει να βελτιώνεται για να παρακάμπτει νέες μεθόδους φραγής. [50]

#### 5.1.1.3.4 Επιθέσεις Phishing σε Πραγματικό Χρόνο

Phishing σε πραγματικό χρόνο — η αυτοματοποίηση της διαδικασίας δημιουργίας και αποστολής phishing περιεχομένου σε κλίμακα με TN. Η ικανότητα εκτέλεσης ταυτόχρονων, πολλαπλών, εξαιρετικά εξελιγμένων τύπων phishing επιθέσεων αποτελεί σημαντικό πονοκέφαλο για τους δημιουργούς συστημάτων ανίχνευσης απειλών σε πραγματικό χρόνο. Οι εκστρατείες phishing που καθοδηγούνται από TN χαρακτηρίζονται από την ταχύτητα και την κλίμακα της TN, κάτι που επίσης καθιστά απαραίτητο τα συστήματα ανίχνευσης να λειτουργούν πιο γρήγορα και με μεγαλύτερη ακρίβεια από ποτέ [50].

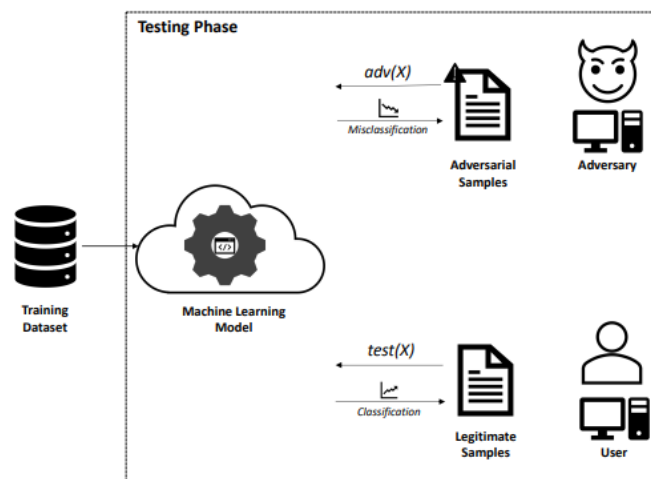
#### 5.1.1.4 Poisoning Attacks

Οι επιθέσεις δηλητηρίασης δεδομένων (data poisoning attacks) αποτελούν σοβαρή απειλή για τα συστήματα μηχανικής μάθησης που χρησιμοποιούνται στην κυβερνοασφάλεια. Σε τέτοιες επιθέσεις, οι επιτιθέμενοι εισάγουν σκόπιμα αλλοιωμένα ή ψευδή δεδομένα στο σύνολο εκπαίδευσης, με στόχο να επηρεάσουν τη συμπεριφορά του μοντέλου και να το οδηγήσουν σε λανθασμένες αποφάσεις. Τεχνικές όπως η αλλοίωση ετικετών (label flipping) και οι επιθέσεις τύπου κρυφής εισόδου επιτρέπουν στους εισβολείς να παρακάμψουν μηχανισμούς ανίχνευσης εισβολών ή malware. Η αυξανόμενη χρήση της μηχανικής μάθησης σε κρίσιμες εφαρμογές απαιτεί πλέον πιο ανθεκτικά μοντέλα και νέες μεθόδους αξιολόγησης που λαμβάνουν υπόψη όχι μόνο την ακρίβεια, αλλά και την αντοχή του μοντέλου απέναντι σε τέτοιες επιθέσεις. [52]

### 5.1.1.5 Επίθεση αποφυγής (Evasion Attack)

Η χρήση της μηχανικής μάθησης στην ανίχνευση εισβολών (IDS) είναι ευρέως διαδεδομένη και έχει επιδείξει αξιοσημείωτη επιδόσεις ως ισχυροί και αποτελεσματικοί μηχανισμοί άμυνας. Ένα IDS παρέχει δυνατότητες ανίχνευσης της κακόβουλης κυκλοφορίας με τη δημιουργία ειδοποιήσεων με αρχεία καταγραφής δικτύου, όπως ώστε να μπορούν να αντληθούν περαιτέρω πληροφορίες όπως και όταν χρειάζεται. Με βάση την τοποθέτησή του στην υποδομή του δικτύου, ένα IDS με βάση το δίκτυο παρακολουθεί την επικοινωνία που ταξιδεύει μέσα και έξω από το δίκτυο, ενώ ένα IDS με βάση τον κεντρικό υπολογιστή σαρώνει έναν συγκεκριμένο κεντρικό υπολογιστή (π.χ. διακομιστή), για να ειδοποιήσει το δίκτυο. διαχειριστή του δικτύου για πιθανές απειλές ασφαλείας.

Με βάση αυτό το σύστημα λοιπόν, στόχος της επίθεσης αποφυγής είναι να προκαλέσει στο μοντέλο μηχανικής μάθησης να ταξινομήσει λανθασμένα τις παρατηρήσεις κατά τη διάρκεια της δοκιμής (όπως φαίνεται στο Σχήμα 13). Εφαρμόζεται σε ένα δικτυακό IDS, ένας αντίπαλος προσπαθεί να αποφύγει την ανίχνευση μεταβάλλοντας το κακόβουλες περιπτώσεις με τέτοιο τρόπο ώστε το IDS να ταξινομεί εσφαλμένα αυτή τη συμπεριφορά ως καλοήγη.[51]



**Σχήμα 13:** Διάγραμμα επίθεσης αποφυγής (evasion attack) εναντίον εκπαιδευμένου μοντέλου μηχανικής μάθησης κατά τη φάση της δοκιμής [51]

Στον Σχήμα 13, βλέπουμε πώς ένας κακόβουλος χρήστης (adversary) μπορεί να ξεγελάσει ένα μοντέλο μηχανικής μάθησης κατά τη φάση δοκιμής (testing phase). Συγκεκριμένα, όταν ένας κανονικός χρήστης παρέχει νόμιμα (legitimate) δεδομένα, το μοντέλο πραγματοποιεί σωστή ταξινόμηση. Ωστόσο, ένας επιτιθέμενος μπορεί να δημιουργήσει παραπλανητικά δείγματα (adversarial samples), δηλαδή να εισάγει πολύ μικρές και δυσδιάκριτες αλλαγές στα αρχικά δεδομένα, οι οποίες οδηγούν το μοντέλο σε λανθασμένη πρόβλεψη (misclassification). Ο σκοπός της εικόνας είναι να καταδείξει ότι το ίδιο μοντέλο μπορεί να λειτουργεί σωστά για νόμιμους χρήστες, αλλά να παραμένει ευάλωτο σε στοχευμένες επιθέσεις από κακόβουλους χρήστες.

| <b>Κατηγορία Απειλής</b>               | <b>Περιγραφή</b>                                                                                            | <b>Παραδείγματα</b>                                    |
|----------------------------------------|-------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| Αντίπαλη TN<br>(Adversarial AI)        | Χειραγώγηση μοντέλων TN μέσω παραποίησης δεδομένων εισόδου ή εκπαίδευσης.                                   | Evasion attacks,<br>Poisoning attacks                  |
| Ransomware με TN                       | Αυτοεκπαιδευόμενο ransomware που εξελίσσεται για να παρακάμπτει άμυνες κυβερνοασφάλειας.                    | NotPetya, Ryuk,<br>LockBit                             |
| Deepfake &<br>Παραπληροφόρηση          | Δημιουργία συνθετικών βίντεο/ήχου για εκβιασμό, απάτες ή χειραγώγηση κοινής γνώμης.                         | Deepfake εκβιαστικά<br>βίντεο, ψευδείς ειδήσεις        |
| AI-Robust Phishing                     | Εξελιγμένες επιθέσεις phishing με χρήση TN για ρεαλιστική προσαρμογή μηνυμάτων.                             | Phishing emails που<br>μοιάζουν με νόμιμα              |
| Αυτοματοποίηση<br>Phishing Εκστρατειών | Χρήση TN για συλλογή δεδομένων στόχων και μαζική αποστολή εξατομικευμένων phishing email.                   | Στοχευμένα spear<br>phishing campaigns                 |
| Ενισχυμένη Δημιουργία<br>Phishing      | Παραγωγή περιεχομένου phishing που είναι αδιαχώριστο από νόμιμο περιεχόμενο.                                | Ρεαλιστικές phishing<br>ιστοσελίδες και<br>μηνύματα    |
| Δυναμική Προσαρμογή<br>& Μάθηση        | Τα συστήματα TN μαθαίνουν και βελτιώνουν τις επιθέσεις τους με βάση τα αποτελέσματα προηγούμενων επιθέσεων. | Βελτιωμένα phishing<br>emails βασισμένα σε<br>feedback |

**Πίνακας 6: Κακόβουλες χρήσεις της TN στην ασφάλεια στον κυβερνοχώρο**

## **5.2 Ηθικές χρήσεις της TN στην κυβερνοασφάλεια**

Με τη χρήση αλγορίθμων αιχμής και τεχνικών μηχανικής μάθησης, οι οργανισμοί μπορούν να βελτιώσουν τη στάση ασφαλείας τους και να προστατεύσουν τα ευαίσθητα δεδομένα από ένα διαρκώς μεταβαλλόμενο τοπίο απειλών. Αυτή η ενότητα διερευνά διάφορες βασικές εφαρμογές της TN στην κυβερνοασφάλεια, συμπεριλαμβανομένων της αυτοματοποιημένης ανάλυσης πληροφοριών για απειλές και , της TN για την αντιμετώπιση και αποκατάσταση περιστατικών και της προγνωστικής ανάλυσης στην ανίχνευση απειλών στον κυβερνοχώρο . Η ευρεία υιοθέτηση της TN στην κυβερνοασφάλεια έχει οδηγήσει στην ανάπτυξη ενός ευρέος φάσματος εφαρμογών που ενισχύουν την ικανότητα εντοπισμού, αντιμετώπισης και μετριασμού των απειλών στον κυβερνοχώρο. [42]

### **5.2.1 Αυτοματοποιημένη ανάλυση, πληροφόρηση και ανίχνευση για απειλές**

Η αυτοματοποίηση των πληροφοριών και της ανάλυσης απειλών είναι μία από τις πιο σημαντικές χρήσεις της TN στην κυβερνοασφάλεια. Οι συμβατικές διαδικασίες πληροφόρησης για απειλές συχνά συνεπάγονται τη χειροκίνητη συλλογή και εξέταση δεδομένων από διάφορες πηγές, συμπεριλαμβανομένων των τροφοδοσιών απειλών, των μέσων κοινωνικής δικτύωσης και της παρακολούθησης του σκοτεινού ιστού. Αυτή η μέθοδος μπορεί να είναι επίπονη και μπορεί να μην είναι σε θέση να συμβαδίσει με το ταχέως εξελισσόμενο τοπίο των απειλών.

Μέσω της χρήσης NLP για την εξαγωγή σχετικών πληροφοριών από αδόμητα δεδομένα, όπως εκθέσεις, ιστολόγια και συζητήσεις στα μέσα κοινωνικής δικτύωσης, οι πλατφόρμες πληροφοριών απειλών με βάση την TN μπορούν να αυτοματοποιήσουν τη διαδικασία αυτόματης συγκέντρωσης, συσχέτισης και ανάλυσης τεράστιων όγκων δεδομένων σε πραγματικό χρόνο. Αυτά τα συστήματα μπορούν να εφαρμόζουν αλγορίθμους μηχανικής μάθησης για τον εντοπισμό αναδυόμενων απειλών και μοτίβων, επιτρέποντας στους οργανισμούς να λαμβάνουν προληπτικά μέτρα για την αντιμετώπιση πιθανών ευπαθειών.

Αυτή η προληπτική προσέγγιση των απειλών πληροφοριών όχι μόνο βελτιώνει την επίγνωση της κατάστασης, αλλά βοηθά επίσης τους οργανισμούς να κατανέμουν τους πόρους πιο αποτελεσματικά, δίνοντας προτεραιότητα στις πιο πειστικές απειλές. Επιπλέον, οι αυτοματοποιημένες πληροφορίες απειλών μπορούν να βελτιώσουν τη συνεργασία μεταξύ των ομάδων ασφαλείας. Τα εργαλεία TN μπορούν να παρέχουν ειδοποιήσεις και πληροφορίες σε πραγματικό χρόνο, επιτρέποντας στις ομάδες να λαμβάνουν γρήγορα ενημερωμένες αποφάσεις.[42]

Για την ανίχνευση απειλών, χρησιμοποιούνται σήμερα διάφορα εργαλεία και τεχνολογίες TN. Η ασφάλεια στον κυβερνοχώρο κάνει εκτεταμένη χρήση των μοντέλων μηχανικής μάθησης, τα οποία κατασκευάζονται για να μαθαίνουν από τις τάσεις των δεδομένων και να προσαρμόζονται σε αυτές. Για τον εντοπισμό παρόμοιων μοτίβων σε νέα δεδομένα, τα μοντέλα μάθησης με επίβλεψη μπορεί να εκπαιδευτούν σε επισημασμένα σύνολα δεδομένων που περιέχουν γνωστούς κινδύνους.

Αντίθετα, τα μοντέλα μάθησης χωρίς επίβλεψη είναι πολύ χρήσιμα για την ανίχνευση ανωμαλιών, καθώς δεν χρειάζονται επισημασμένα δεδομένα και μπορούν να εντοπίσουν

ανωμαλίες στη συμπεριφορά του συστήματος που μπορεί να υποδεικνύουν κινδύνους που δεν είχαν εντοπιστεί προηγουμένως.

Μια άλλη πτυχή της μηχανικής μάθησης που έχει αποδειχθεί χρήσιμη στην ανίχνευση απειλών είναι η βαθιά μάθηση. Τα μοντέλα βαθιάς μάθησης, όπως τα νευρωνικά δίκτυα, μπορούν να επεξεργάζονται πολύπλοκες δομές δεδομένων και να αναγνωρίζουν περίπλοκα μοτίβα, καθιστώντας τα ιδιαίτερα αποτελεσματικά στην ανίχνευση εξελιγμένων απειλών όπως το κακόβουλο λογισμικό και οι επιθέσεις phishing.

Τα παραδοσιακά συστήματα κυβερνοασφάλειας βασίζονται σε προκαθορισμένους κανόνες και υπογραφές για τον εντοπισμό απειλών. Αν και είναι αποτελεσματικά έναντι γνωστών επιθέσεων, τα συστήματα αυτά δυσκολεύονται να συμβαδίσουν με τη συνεχώς εξελισσόμενη φύση των απειλών στον κυβερνοχώρο.

Αντίθετα, τα συστήματα με TN προσφέρουν μια δυναμική προσέγγιση στην ανίχνευση απειλών. Αξιοποιώντας αλγορίθμους μηχανικής μάθησης και βαθιάς μάθησης, τα συστήματα TN μπορούν να μαθαίνουν συνεχώς από τα δεδομένα, να προσαρμόζονται σε νέες απειλές και να βελτιώνουν την ακρίβεια ανίχνευσης με την πάροδο του χρόνου.

Παρά τα σαφή πλεονεκτήματα της TN στην κυβερνοασφάλεια, παραμένουν προκλήσεις, όπως η ανάγκη για δεδομένα υψηλής ποιότητας, ο κίνδυνος της αντίπαλης TN και οι ανησυχίες σχετικά με τη διαφάνεια και την ερμηνευσιμότητα των αποφάσεων της TN.

Οι οργανισμοί μπορούν πλέον να εντοπίζουν και να σταματούν πολύπλοκες επιθέσεις όπως οι APTs και οι ευπάθειες μηδενικής ημέρας χάρη στις σημαντικά βελτιωμένες δεξιότητες ανίχνευσης απειλών της TN. [44]

### **5.2.2 Αξιολόγηση ευπαθειών(Vulnerability assessment)**

Στο σύγχρονο διασυνδεδεμένο και ψηφιοποιημένο κόσμο, οι οργανισμοί αντιμετωπίζουν πλέον μια πληθώρα κινδύνων που θέτουν οι εγκληματίες του κυβερνοχώρου, οι οποίοι εκμεταλλεύονται τρωτά σημεία στα συστήματα και τα δίκτυά τους, με στόχο την παράνομη πρόσβαση σε ευαίσθητες πληροφορίες, τη διατάραξη των λειτουργιών ή την πρόκληση οικονομικών ζημιών.

Για να αντιμετωπίσουν και να μετριάσουν αυτούς τους κινδύνους, οι οργανισμοί εφαρμόζουν μια σειρά μέτρων ασφαλείας, μεταξύ των οποίων η αξιολόγηση των ευπαθειών αναδεικνύεται σε κεντρικό στοιχείο των ολοκληρωμένων στρατηγικών τους για την ασφάλεια στον κυβερνοχώρο και την αντιμετώπιση περιστατικών. Αναμφίβολα, η αξιολόγηση τρωτότητας αποκτά ύψιστη σημασία στη σφαίρα της διαχείρισης της ασφάλειας στον κυβερνοχώρο.

Περιλαμβάνει τον λεπτομερή εντοπισμό των ευπαθειών που υπάρχουν σε λογισμικό και συστήματα, αποτελώντας μια προληπτική διαδικασία σάρωσης και εξέτασης των πιθανών στόχων και των αναδυόμενων απειλών με στόχο την αποτροπή κακόβουλων επιθέσεων. Ο τομέας της αξιολόγησης τρωτότητας έχει φθάσει σε σημαντικό επίπεδο ωριμότητας - ωστόσο, το να συμβαδίζει με το ευρύ φάσμα των υπολογιστικών και ψηφιακών συσκευών που απαιτούν έλεγχο αποτελεί σημαντική πρόκληση.

Η πρακτική αυτή περιστρέφεται γύρω από μια μεθοδική προσέγγιση για τον εντοπισμό και την αξιολόγηση των ευπαθειών που υπάρχουν στην υποδομή, τις εφαρμογές και τα συστήματα

πληροφορικής ενός οργανισμού. Περιλαμβάνει την προληπτική σάρωση, τον έλεγχο και την ανάλυση πιθανών αδυναμιών που μπορεί να εκμεταλλευτούν κακόβουλα άτομα.

Οι συνήθεις παραδοσιακές προσεγγίσεις για την αξιολόγηση των ευπαθειών έχουν βασιστεί κυρίως σε χειροκίνητες τεχνικές και στατικά συστήματα που βασίζονται σε κανόνες, τα οποία συχνά δυσκολεύονται να ανταποκριθούν στον ρυθμό του εξελισσόμενου τοπίου απειλών και στην αδιάκοπη αύξηση του όγκου και της πολυπλοκότητας των ευπαθειών.

Η έλευση της TN έχει επιφέρει μια μετασχηματιστική στροφή στον τομέα της ασφάλειας στον κυβερνοχώρο, που περιλαμβάνει ζωτικές πτυχές όπως η αξιολόγηση των ευπαθειών και η αντιμετώπιση περιστατικών. Η TN εισάγει νέες δυνατότητες και αποδοτικότητες που έχουν τη δυνατότητα να βελτιώσουν σημαντικά την αποτελεσματικότητα και την αποδοτικότητα αυτών των καίριων διαδικασιών ασφάλειας.

Μέσω της αξιοποίησης των αλγορίθμων μηχανικής μάθησης, της NLP και της μεθόδου βαθιάς μάθησης, η αξιολόγηση ευπαθειών με TN δίνει τη δυνατότητα στους οργανισμούς να εντοπίζουν, να αναλύουν και να αντιμετωπίζουν τις ευπάθειες με πιο προληπτικό, ακριβή και έγκαιρο τρόπο. Χρησιμοποιώντας αυτοματοποιημένους μηχανισμούς για να βοηθήσουν στην ανίχνευση ευπαθειών πριν από την ανάπτυξη του συστήματος, οι ομάδες προϊόντων μπορούν να αφιερώσουν περισσότερη προσοχή στην ανάπτυξη χαρακτηριστικών και στη βελτίωση της απόδοσης. Ο πολλαπλασιασμός των συσκευών και των εφαρμογών που αναπτύσσονται σήμερα όχι μόνο ενισχύει τους κινδύνους που συνδέονται με τα δικτυωμένα συστήματα, αλλά παρέχει επίσης ένα πλούσιο απόθεμα δεδομένων εκπαίδευσης για χρήση σε συνδυασμό με τεχνικές TN. Ο ρόλος της TN στην αξιολόγηση της τρωτότητας αποκτά πολύπλευρο χαρακτήρα.[45]

### **5.2.3 TN για την αντίδραση σε συμβάντα και την αποκατάσταση**

Η TN είναι απαραίτητη για τη βελτίωση των διαδικασιών αντιμετώπισης συμβάντων και αποκατάστασης, εκτός από τη νοημοσύνη απειλών. Τα περιστατικά στον κυβερνοχώρο συχνά απαιτούν γρήγορη δράση για τον περιορισμό των ζημιών και την επανέναρξη των κανονικών λειτουργιών. Η TN μπορεί να αυτοματοποιήσει και να εκσυγχρονίσει πολλές διαδικασίες απόκρισης σε συμβάντα, μειώνοντας σημαντικά τον χρόνο που απαιτείται για τον εντοπισμό και την αντιμετώπιση απειλών. Οι ομάδες ασφαλείας μπορούν να επικεντρωθούν σε πιο περίπλοκες και στρατηγικές εργασίες, όπως η ανάλυση της αιτίας των περιστατικών και η ενίσχυση της μελλοντικής άμυνας, χρησιμοποιώντας πλατφόρμες ενορχήστρωσης, αυτοματοποίησης και απόκρισης ασφάλειας (SOAR) με βάση την TN, οι οποίες, για παράδειγμα, μπορούν να αυτοματοποιήσουν τις συνήθεις εργασίες που εμπλέκονται στην απόκριση περιστατικών. Όταν εντοπίζεται μια πιθανή απειλή, τα συστήματα αυτά μπορούν να συλλέγουν αυτόματα τα σχετικά δεδομένα, να απομονώνουν τα επηρεαζόμενα συστήματα και να δρομολογούν προκαθορισμένα πρωτόκολλα αντιμετώπισης.

Επιπλέον, η TN μπορεί να βελτιώσει τις διαδικασίες αποκατάστασης συμβάντων αξιολογώντας τις επιπτώσεις μιας επίθεσης και εντοπίζοντας τους τομείς που απαιτούν βελτίωση. Για παράδειγμα, οι αλγόριθμοι TN μπορούν να αξιολογήσουν την αποτελεσματικότητα των σχεδίων αντιμετώπισης, βοηθώντας τους οργανισμούς να βελτιώσουν τα σχέδια αντιμετώπισης περιστατικών και να δημιουργήσουν ισχυρότερες στρατηγικές ασφάλειας. Η TN μπορεί επίσης να

βελτιώνει συνεχώς την ικανότητα ενός οργανισμού να ανταποκρίνεται σε μελλοντικές απειλές, μαθαίνοντας από προηγούμενα περιστατικά.[42]

#### 5.2.4 Χρήση της Προγνωστικής ανάλυσης (Predictive Analytics)

Η προγνωστική ανάλυση, η οποία προβλέπει πιθανές απειλές στον κυβερνοχώρο χρησιμοποιώντας ιστορικά δεδομένα και εξελιγμένους αλγορίθμους, είναι μια άλλη ουσιαστική χρήση της ΤΝ στην κυβερνοασφάλεια. Με τον εντοπισμό των ευπαθειών και την πρόβλεψη πιθανών φορέων επίθεσης πριν από την εκμετάλλευσή τους, η προγνωστική ανάλυση μπορεί να βοηθήσει τους οργανισμούς να παραμείνουν μπροστά από τους επιτιθέμενους. Μεγάλοι όγκοι δεδομένων από διάφορες πηγές, όπως αρχεία καταγραφής δικτύου, συμπεριφορά χρηστών και τροφοδοσίες πληροφοριών για απειλές, μπορούν να αναλυθούν από εργαλεία προγνωστικής ανάλυσης με ΤΝ. Με το να εντοπίζουν μοτίβα και τάσεις σε αυτά τα δεδομένα, αυτά τα εργαλεία μπορούν να προβλέπουν με ακρίβεια πιθανές απειλές και ευπάθειες. Για παράδειγμα, εάν ένα μοντέλο μηχανικής μάθησης ανιχνεύσει μια έξαρση σε ανεπιτυχείς προσπάθειες σύνδεσης από μια συγκεκριμένη διεύθυνση IP , μπορεί να επισημάνει αυτή τη δραστηριότητα ως πιθανή επίθεση brute-force, προκαλώντας τις ομάδες ασφαλείας να το εξετάσουν περαιτέρω. Η επίθεση που χρησιμοποιεί τη μέθοδο «Δοκιμή και σφάλμα» μαντεύοντας κωδικούς πρόσβασης ονομάζεται επίθεση brute-force. Ο επιτιθέμενος συλλέγει πρώτα τις βασικές πληροφορίες για τον χρήστη. Για παράδειγμα, το ονοματεπώνυμο του χρήστη, τον αριθμό δωματίου, τον αριθμό οχήματος , τα ονόματα των παιδιών κ.λπ. Ο επιτιθέμενος δοκιμάζει συνεχώς τυχαίους κωδικούς πρόσβασης με βάση τις προσωπικές πληροφορίες του χρήστη. Ο επιτιθέμενος προσπαθεί αυτό μέχρι να επιτύχει. Αυτό μπορεί να διαρκέσει ώρες, ημέρες, μήνες ή και χρόνια. [43]

Βοηθώντας τους οργανισμούς να ιεραρχήσουν τις ευπάθειες ανάλογα με τις πιθανές επιπτώσεις τους, η προγνωστική ανάλυση μπορεί να βελτιώσει τη διαχείριση κινδύνων. Τα συστήματα ΤΝ είναι ικανά να αποδίδουν βαθμολογίες κινδύνου στις ευπάθειες που ανακαλύπτονται , γεγονός που επιτρέπει στις ομάδες ασφαλείας να επικεντρωθούν στην επίλυση των πιο κρίσιμων ζητημάτων πρώτα. Αυτού του είδους η ιεράρχηση προτεραιοτήτων είναι ιδιαίτερα κρίσιμη σε περιβάλλοντα με περιορισμένους πόρους, καθώς βοηθά τους οργανισμούς να κατανέμουν καλύτερα τους πόρους τους . [42]

#### 5.2.5 Εφαρμογές στον πραγματικό κόσμο και μελέτες περίπτωσης

Πολλά χρηματοπιστωτικά ιδρύματα χρησιμοποιούν ΤΝ για την παρακολούθηση των συναλλαγών για ενδείξεις δόλιας δραστηριότητας. Αναλύοντας μοτίβα στα δεδομένα συναλλαγών, τα συστήματα αυτά μπορούν να επισημάνουν ύποπτη συμπεριφορά για περαιτέρω διερεύνηση, συμβάλλοντας στην πρόληψη οικονομικών απωλειών. Αρκετοί οργανισμοί έχουν εφαρμόσει με επιτυχία λύσεις με βάση την ΤΝ για να ενισχύσουν τις προσπάθειες κυβερνοασφάλειας.

Παρόμοια, οι εταιρείες υγειονομικής περίθαλψης αρχίζουν να χρησιμοποιούν τεχνολογίες ΤΝ για την προστασία ευαίσθητων δεδομένων ασθενών. Οι εταιρείες αυτές μπορούν να συμμορφωθούν με τους κανονισμούς και να προστατεύσουν την ιδιωτικότητα των ασθενών, χρησιμοποιώντας εργαλεία με γνώμονα την ΤΝ που αναλύουν τη συμπεριφορά των χρηστών και

τα αρχεία καταγραφής πρόσβασης για τον εντοπισμό πιθανών εσωτερικών απειλών και προσπαθειών μη εξουσιοδοτημένης πρόσβασης.

Καθώς το τοπίο των απειλών στον κυβερνοχώρο συνεχίζει να αυξάνεται σε πολυπλοκότητα, η ενσωμάτωση της TN στις στρατηγικές κυβερνοασφάλειας θα είναι ζωτικής σημασίας για τους οργανισμούς που επιδιώκουν να μετριάσουν τους κινδύνους και να ανταποκριθούν στις αναδυόμενες προκλήσεις σε πραγματικό χρόνο. Με την υιοθέτηση αυτών των εφαρμογών που βασίζονται στην TN από το, οι οργανισμοί μπορούν να ενισχύσουν την ανθεκτικότητά τους έναντι των απειλών στον κυβερνοχώρο.

Οι εφαρμογές της TN στην ασφάλεια στον κυβερνοχώρο είναι ποικίλες και εξελίσσονται συνεχώς, αντανακλώντας την ανάγκη των οργανισμών να παραμείνουν μπροστά από τις εξελιγμένες απειλές στον κυβερνοχώρο. Από την αυτοματοποιημένη νοημοσύνη και ανάλυση απειλών έως την ενισχυμένη αντιμετώπιση περιστατικών και την προγνωστική ανάλυση, οι τεχνολογίες TN προσφέρουν ισχυρά εργαλεία που δίνουν τη δυνατότητα στις ομάδες ασφαλείας να προστατεύουν αποτελεσματικότερα τα ψηφιακά τους περιουσιακά στοιχεία. [42]

| Κατηγορία Χρήσης TN                            | Περιγραφή                                                                                                     |
|------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Αυτοματοποιημένη Ανάλυση και Ανίχνευση Απειλών | NLP και ML εντοπίζουν απειλές σε πραγματικό χρόνο, αναλύοντας μεγάλους όγκους αδόμητων δεδομένων.             |
| Αξιολόγηση Ευπαθειών                           | Με TN, οι οργανισμοί εντοπίζουν και αντιμετωπίζουν ευπάθειες προληπτικά και με μεγαλύτερη ακρίβεια.           |
| Αντίδραση σε Συμβάντα & Αποκατάσταση           | Τα συστήματα SOAR αυτοματοποιούν την απόκριση σε απειλές και βελτιώνουν τη διαχείριση περιστατικών.           |
| Προγνωστική Ανάλυση                            | Προβλέπει μελλοντικές επιθέσεις, βοηθώντας στην ιεράρχηση των κινδύνων και στη διαχείριση πόρων.              |
| Εφαρμογές στον Πραγματικό Κόσμο                | Χρησιμοποιείται σε τράπεζες και υγειονομικούς φορείς για ανίχνευση απάτης και προστασία προσωπικών δεδομένων. |

Πίνακας 6: Ηθικές χρήσεις της TN στην κυβερνοασφάλεια

## 5.3 Προκλήσεις & Περιορισμοί

### 5.3.1 Ευπάθεια στον Χειρισμό και την Κακόβουλη Χρήση

Μία από τις πιο σημαντικές περιοριστικές δυνατότητες της TN είναι ότι είναι απλώς ένας υπολογιστικός κώδικας που έχει προγραμματιστεί ώστε να διασφαλίζει ότι έχουν τηρηθεί τα πρωτόκολλα και ότι μπορεί να αναπτυχθεί σε περίπτωση ανάγκης. Αυτή η περίπτωση μπορεί να ακούγεται εντάξει, καθώς μπορεί να αναπτυχθεί σε περίπτωση ανάγκης. Ωστόσο, το σύστημα είναι εξ ολοκλήρου προγραμματισμένο· επομένως, οποιοσδήποτε μπορεί να το πάρει υπό τον έλεγχό του και να το χειραγωγήσει και να το χρησιμοποιήσει ως όπλο. Αρκούν λίγες γραμμές κώδικα να επεξεργαστούν, και τότε οι πολλές ώρες εργασίας μπορεί να μετατραπούν σε όπλο που θα χρησιμοποιηθεί εναντίον του ίδιου του συστήματος. Συνεπώς, με την κατάλληλη ικανότητα

και γνώση, η τεχνολογία ΤΝ μπορεί να χρησιμοποιηθεί ως όπλο που θα χρησιμοποιηθεί για να καταστρέψει αυτό που προορίζεται να προστατεύει. Αυτός ο παράγοντας είναι ένας από τους σημαντικότερους περιορισμούς της ΤΝ στην κυβερνοασφάλεια. Οι προγραμματιστές και οι επιστήμονες των υπολογιστών θα πρέπει να το λάβουν αυτό υπόψη, καθώς κατανοούν τη δυνατότητα της τεχνολογίας ΤΝ. [53]

### **5.3.2 Περιορισμοί στην Ανίχνευση Απειλών**

Τα συστήματα ΤΝ μπορούν επίσης να εκπαιδευτούν ώστε να εντοπίζουν κυβερνοαπειλές και κακόβουλα λογισμικά, καθιστώντας τα έτσι πιο αποτελεσματικά στην κυβερνοασφάλεια. Ο αυξανόμενος αριθμός επιθέσεων στην κυβερνοασφάλεια έχει οδηγήσει στην υιοθέτηση της ΤΝ στην κυβερνοασφάλεια. Ολόκληρη η διαδικασία αποσκοπεί στη διασφάλιση της αποτελεσματικότητας και της ακρίβειας. Ωστόσο, η ΤΝ έχει περιορισμούς και δεν μπορεί να αντικαταστήσει τους ανθρώπους, αφού της δίνεται εντολή να εκτελέσει μια συγκεκριμένη εργασία. Κάποιες φορές δεν μπορεί να εντοπίσει σχεδόν αδιακρίτως απειλές και έτσι αντιμετωπίζει προβλήματα, καθώς μοιάζει με το πραγματικό μήνυμα. Η ΤΝ μπορεί επίσης να δυσκολεύεται να εντοπίσει απειλές λόγω των εξελισσόμενων κυβερνοαπειλών. Οι ιοί και τα κακόβουλα λογισμικά βελτιώνονται διαρκώς, και έτσι θα πρέπει και το σύστημα ΤΝ να βελτιώνεται για μεγαλύτερη αποτελεσματικότητα. Επίσης, η πρακτική της κυβερνοασφάλειας είναι πολλαπλή σε σχέση με τους κυβερνοεγκληματίες, οι οποίοι τείνουν να αποκτούν περισσότερες πληροφορίες για το χακάρισμα. Επομένως, οι κυβερνοεγκληματίες μπορούν να δημιουργήσουν μια καλύτερη απειλή που η ΤΝ δεν θα εντοπίσει εύκολα. Παρόλο που η ΤΝ εξοικονομεί χρόνο για την ομάδα ασφάλειας, απαιτεί επίσης ανθρώπινους ειδικούς για δημιουργικότητα, διευκολύνοντας έτσι τη δουλειά τους. Ο περιορισμός αυτός καλεί τους προγραμματιστές να εξασφαλίσουν ότι έχουν εξοπλίσει την τεχνολογία με πολλαπλές δυνατότητες ώστε να αντιμετωπίσουν οποιοδήποτε έγκλημα προκύψει από τους περιορισμούς τους. [53]

### **5.3.3 Δυνατότητα Κακόβουλης Χρήσης από Τρίτους**

Από την άλλη πλευρά, μπορούμε να διαπιστώσουμε ότι η τεχνολογία ΤΝ δεν χρησιμοποιείται εξ ολοκλήρου για την προστασία δεδομένων και τη διασφάλιση της ασφάλειας των δεδομένων. Μπορούμε επίσης να έχουμε τεχνολογία ΤΝ που έχει αναπτυχθεί ώστε να δημιουργεί και να παράγει ιούς υπολογιστών. Η πολυπλοκότητα που περιλαμβάνεται στην τεχνολογία καθιστά δύσκολο για ένα άτομο να ανταγωνιστεί το μηχάνημα, οδηγώντας στην παραβίαση των δεδομένων. Με τόσες πολλές δυνατότητες που παράγονται στους κώδικες που έχουν δημιουργηθεί από ΤΝ, οι οποίοι χρησιμοποιούνται για την ανάπτυξη ενός ιού υπολογιστή, καθίσταται απολύτως δυνατό να είναι εξαιρετικά εύκολο να διαφθαρεί μια βάση δεδομένων και να χειραγωγηθούν τα δεδομένα της. Αυτός είναι ένας άλλος περιορισμός που ουσιαστικά δεν αφορά την τεχνολογία ΤΝ ως φρουρό του κυβερνοεγκλήματος αλλά ως συμμετόχο στο κυβερνοέγκλημα. Αυτός ο περιορισμός αναδεικνύει μια άλλη τεράστια επίδραση της ΤΝ στην κυβερνοασφάλεια. [53]

### **5.3.4 Περιορισμένη Κατανόηση και Χρήση από την Κοινωνία**

Η πολυπλοκότητα της τεχνολογίας αποτελεί περιορισμό της τεχνολογίας ΤΝ, καθώς δεν γνωρίζουν όλοι στην κοινωνία την τεχνολογία. Υπάρχει επίσης το γεγονός ότι δεν είναι εύκολη υπόθεση να κατανοηθούν τα διάφορα μοντέλα που εμπλέκονται στην τεχνολογία. Η τεχνολογία, όντας τόσο δύσκολη στη χρήση και στην υλοποίηση στο μέγιστο των δυνατοτήτων της, μπορεί να δώσει στους εγκληματίες την ευκαιρία να καταλάβουν το σύστημα, αφού εμείς δεν μπορούμε να το χειριστούμε στη μέγιστη ικανότητά του. Η τεχνολογία απαιτεί πολλές πληροφορίες σχετικά με τη λειτουργία της, τις οποίες πολλοί άνθρωποι δεν έχουν. Συνεπώς, οι οργανισμοί εξακολουθούν να βρίσκονται σε κίνδυνο, καθώς δεν μπορούν να κάνουν το σύστημα να λειτουργεί στο καλύτερο δυνατό.

### **5.3.5 Υψηλό Κόστος Υλοποίησης**

Επίσης, λόγω της πολυπλοκότητας του συστήματος, καταλαβαίνουμε ότι η τεχνολογία θα κοστίσει πολύ. Επομένως, το κόστος υλοποίησης της τεχνολογίας είναι πολύ πιο ακριβό. Συνεπώς, δεν θα έχουν όλοι οι οργανισμοί στον κόσμο τη δυνατότητα πρόσβασης στην τεχνολογία και τη διασφάλιση της ασφάλειας των δεδομένων τους. Επομένως, το κόστος της τεχνολογίας αποτελεί επίσης έναν περιορισμό για την υλοποίηση. Παρόλο που οι πληροφορίες του οργανισμού είναι απαραίτητες για κάθε οργανισμό, το κόστος υλοποίησης της τεχνολογίας ΤΝ είναι πολύ υψηλότερο, περιορίζοντας τον αριθμό των ατόμων που θα χρησιμοποιήσουν την τεχνολογία για την ασφάλεια των δεδομένων και των πληροφοριών τους. Το κόστος του συστήματος έχει ως αποτέλεσμα λίγα μέλη και οργανισμούς να χρησιμοποιούν την τεχνολογία, καθιστώντας δύσκολη την εκτίμηση της ικανότητας της τεχνολογίας. [53]

# 6

## *Συμπεράσματα και Προοπτικές για το μέλλον*

### **6.1 Σύνοψη**

Συνοψίζοντας, στην παραπάνω εργασία εστιάζουμε στο πώς η τεχνητή νοημοσύνη (TN) μετασχηματίζει ριζικά την ασφάλεια στον κυβερνοχώρο, λειτουργώντας τόσο ως ισχυρό εργαλείο ενίσχυσης των αμυντικών μηχανισμών όσο και ως πιθανό όπλο στα χέρια κακόβουλων επιτιθέμενων. Στο αρχικό μέρος, η εργασία εισάγει τον αναγνώστη στις βασικές έννοιες και την ιστορική εξέλιξη της κυβερνοασφάλειας, αναλύοντας τα θεμελιώδη θεωρητικά μοντέλα που τη διέπουν, όπως η τριάδα Εμπιστευτικότητας, Ακεραιότητας και Διαθεσιμότητας (CIA) και το πλαίσιο Αυθεντικοποίησης, Εξουσιοδότησης και Λογιστικής (AAA). Παράλληλα, δίνεται έμφαση σε κρίσιμους όρους και πρακτικές που αποτελούν τη βάση για την κατανόηση των σύγχρονων προκλήσεων στον τομέα.

Στη συνέχεια, εμβαθύνουμε στην ανάλυση των τεχνολογιών TN, εξετάζοντας πώς μέθοδοι όπως το machine learning, το deep learning και η επεξεργασία φυσικής γλώσσας αξιοποιούνται για την ανίχνευση εισβολών, την ανάλυση κακόβουλου λογισμικού και την αυτοματοποίηση της αντίδρασης σε επιθέσεις. Παρουσιάζονται πρακτικά παραδείγματα εφαρμογών της TN που ενισχύουν τη δυνατότητα άμεσης απόκρισης και περιορισμού των επιπτώσεων από κυβερνοεπιθέσεις. Ταυτόχρονα, αναδεικνύονται οι περιορισμοί και οι ηθικές προκλήσεις που συνοδεύουν τη χρήση αυτών των τεχνολογιών, όπως η προκατάληψη στους αλγόριθμους, η έλλειψη διαφάνειας στις αποφάσεις των μοντέλων και ο κίνδυνος για νέες μορφές επιθέσεων βασισμένων στην TN.

Η εργασία δεν παραλείπει να επισημάνει τη διπλή φύση της TN στον κυβερνοχώρο: από τη μία πλευρά προσφέρει προηγμένα μέσα προστασίας, ενώ από την άλλη δημιουργεί νέους κινδύνους και απαιτεί ιδιαίτερη προσοχή και υπευθυνότητα κατά την ανάπτυξη και εφαρμογή της.

Στην τελική ενότητα, παρέχονται προτάσεις για τη διαμόρφωση ασφαλών και ηθικά ορθών πρακτικών χρήσης της ΤΝ στην κυβερνοασφάλεια, και εντοπίζονται οι τάσεις που αναμένεται να κυριαρχήσουν στο μέλλον, όπως η ανάπτυξη αλγορίθμων.

## 6.2 Συμπεράσματα

### *1. Η Τεχνητή Νοημοσύνη (TN) αποτελεί καταλύτη στην ενίσχυση της κυβερνοασφάλειας*

Η αξιοποίηση τεχνολογιών ΤΝ, όπως η μηχανική μάθηση και η βαθιά μάθηση, έχει αυξήσει σημαντικά την αποτελεσματικότητα των συστημάτων ανίχνευσης απειλών και αυτοματοποιημένης αντίδρασης. Οι οργανισμοί που ενσωματώνουν ΤΝ στις άμυνες τους είναι σε θέση να εντοπίζουν και να αποκρούουν επιθέσεις με ταχύτερο και πιο ακριβή τρόπο.

### *2. Η ΤΝ ενισχύει, αλλά δεν αντικαθιστά την ανθρώπινη παρουσία στην ασφάλεια*

Παρά τις σημαντικές δυνατότητες αυτοματοποίησης, η ανθρώπινη εποπτεία παραμένει κρίσιμη, τόσο για την ερμηνεία των ευρημάτων όσο και για τη διαχείριση πολύπλοκων περιστατικών ασφαλείας. Το μέλλον της κυβερνοασφάλειας βρίσκεται στη συνεργασία ανθρώπου και μηχανής, και όχι στην πλήρη αυτοματοποίηση.

### *3. Η διπλή φύση της ΤΝ δημιουργεί νέες απειλές*

Ενώ οι αλγόριθμοι ΤΝ βελτιώνουν την άμυνα, ταυτόχρονα αποτελούν εργαλείο στα χέρια κακόβουλων επιτιθέμενων, οι οποίοι χρησιμοποιούν ΤΝ για πιο εξελιγμένες επιθέσεις (π.χ. AI-generated phishing, adversarial attacks). Η κακόβουλη χρήση της ΤΝ αποτελεί μείζονα κίνδυνο για το μέλλον της κυβερνοασφάλειας.

### *4. Οι υπάρχουσες δομές ασφαλείας χρειάζονται αναθεώρηση υπό το πρίσμα της ΤΝ*

Τα παραδοσιακά πλαίσια κυβερνοασφάλειας, όπως το NIST framework, πρέπει να προσαρμοστούν ώστε να ενσωματώνουν και να αξιοποιούν τις δυνατότητες της ΤΝ, όχι μόνο για άμυνα αλλά και για προληπτική ανάλυση και στρατηγικό σχεδιασμό.

### *5. Η εκπαίδευση και η ανάπτυξη νέων δεξιοτήτων είναι κρίσιμη*

Οι επαγγελματίες του κλάδου πρέπει να αναπτύξουν νέες δεξιότητες που συνδυάζουν γνώση κυβερνοασφάλειας και κατανόηση των αρχών της ΤΝ, προκειμένου να αξιοποιήσουν αποτελεσματικά τα νέα εργαλεία και να ανταπεξέλθουν στις εξελισσόμενες απειλές.

### **6.3 Μελλοντικές κατευθύνσεις**

Η μελλοντική πορεία της κυβερνοασφάλειας διαγράφεται άρρηκτα συνδεδεμένη με την εξέλιξη και την ενσωμάτωση της Τεχνητής Νοημοσύνης (TN). Όπως αναλύεται στο κείμενο, η TN προσφέρει πρωτοφανείς δυνατότητες στον εντοπισμό και την αποτροπή κυβερνοαπειλών, ξεπερνώντας τα όρια των παραδοσιακών μέσων ασφαλείας. Στο άμεσο μέλλον, οι επιχειρήσεις και οι κυβερνητικοί οργανισμοί θα στραφούν ακόμη περισσότερο προς την εφαρμογή συστημάτων TN για να ενισχύσουν την αποτελεσματικότητα της ψηφιακής άμυνας. Τα συστήματα αυτά μπορούν να αναλύουν τεράστιους όγκους δομημένων και μη δομημένων δεδομένων σε πραγματικό χρόνο, εντοπίζοντας μοτίβα που διαφεύγουν από την ανθρώπινη αντίληψη. Παράλληλα, η TN μπορεί να αναγνωρίζει συμπεριφορικές ανωμαλίες, όπως ασυνήθιστες ώρες σύνδεσης ή μοτίβα πληκτρολόγησης, εντοπίζοντας πιθανές παραβιάσεις πριν αυτές εξελιχθούν σε κρίσεις.

Επιπλέον, το μέλλον της κυβερνοασφάλειας θα βασίζεται σε «έξυπνα» αμυντικά συστήματα που θα μαθαίνουν και θα προσαρμόζονται συνεχώς. Αυτό σημαίνει ότι οι μηχανισμοί ασφαλείας δεν θα είναι στατικοί, αλλά δυναμικοί, ικανοί να εξελίσσονται όσο εξελίσσονται και οι τεχνικές των κυβερνοεγκληματιών. Η TN θα αξιοποιηθεί για να ενισχύσει υπάρχουσες λύσεις, όπως τα firewalls, τα IDS/IPS, τα antivirus και τις πλατφόρμες διαχείρισης απειλών, δημιουργώντας ένα πολυεπίπεδο πλέγμα προστασίας.

Ωστόσο, η μελλοντική χρήση της TN στην κυβερνοασφάλεια δεν θα είναι χωρίς προκλήσεις. Η τεχνολογία αυτή απαιτεί σημαντικούς πόρους, τόσο σε υπολογιστική ισχύ όσο και σε ανθρώπινο δυναμικό με εξειδίκευση. Παράλληλα, οι ίδιοι οι εγκληματίες του κυβερνοχώρου θα αρχίσουν να αξιοποιούν την TN για να σχεδιάζουν πιο πολύπλοκες και αποτελεσματικές επιθέσεις. Το μέλλον, λοιπόν, φέρνει μαζί του ένα νέο είδος «ψηφιακού πολέμου» με την TN και από τις δύο πλευρές – των αμυνόμενων και των επιτιθέμενων. Παρ' όλα αυτά, η TN αποτελεί ισχυρό σύμμαχο στον αγώνα για την προστασία των ψηφιακών μας πόρων και δεδομένων. Η συνεχής επιτήρηση, η ταχύτατη ανάλυση και αντίδραση, και η δυνατότητα ενίσχυσης των ανθρώπινων ικανοτήτων με προηγμένα αλγοριθμικά συστήματα, καταδεικνύουν πως το μέλλον της κυβερνοασφάλειας θα καθορίζεται όλο και περισσότερο από την εξέλιξη της TN. [54]

## **Βιβλιογραφία**

- [1] Kaur, Ramanpreet, et al. “Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions.” *Information Fusion*, vol. 97, Sept. 2023, p. 101804, <https://doi.org/10.1016/j.inffus.2023.101804>.
- [2] Craigen, Dan, et al. “Defining Cybersecurity.” *Technology Innovation Management Review*, Oct. 2014, pp. 13–21, [www.timreview.ca](http://www.timreview.ca).
- [3] Peslak, Alan, and D. Scott Hunsinger. “WHAT IS CYBERSECURITY and WHAT CYBERSECURITY SKILLS ARE EMPLOYERS SEEKING? .” *Issues in Information Systems*, Penn State University , Appalachian State University, 2019, pp. 62–72.
- [4] Morgan, Jacob. *5/4/2018 a Simple Explanation of “the Internet of Things” a Simple Explanation of “the Internet of Things.”* 2014.
- [5] Warner, Michael. “Cybersecurity: A Pre-History.” *Intelligence and National Security*, vol. 27, no. 5, Oct. 2012, pp. 781–799, <https://doi.org/10.1080/02684527.2012.708530>.
- [6] Le, Dac-Nhuong, et al. *Cyber Security in Parallel and Distributed Computing : Concepts, Techniques, Applications and Case Studies*. Hoboken, Nj, Usa, John Wiley & Sons, Inc. ; Beverly, Ma, 2019.
- [7] Obiora Nweke, Livinus . “Using the CIA and AAA Models to Explain Cybersecurity Activities.” *PM World Journal*, Dec. 2017, [www.pmworldjournal.net](http://www.pmworldjournal.net).
- [8] Kumari, Saru, et al. *Questioning Key Compromise Attack on Ostad- Sharif et Al.'S Authentication and Session Key Generation Scheme for Healthcare Applications*. pp. 2169–3536, <https://doi.org/10.1109/ACCESS.2017>.

- [9] Bhuyan, M. H., et al. “Surveying Port Scans and Their Detection Methodologies.” *The Computer Journal*, vol. 54, no. 10, 20 Apr. 2011, pp. 1565–1581, <https://doi.org/10.1093/comjnl/bxr035>.
- [10] Alabdan, Rana. “Phishing Attacks Survey: Types, Vectors, and Technical Approaches.” *Future Internet*, vol. 12, no. 10, 30 Sept. 2020, p. 168, <https://doi.org/10.3390/fi12100168>.
- [11] Hydera, Isatou, et al. “Current State of Research on Cross-Site Scripting (XSS) – a Systematic Literature Review.” *Information and Software Technology*, vol. 58, Feb. 2015, pp. 170–186, <https://doi.org/10.1016/j.infsof.2014.07.010>.
- [12] Cowan, Crispin, et al. *Buffer Overflows: Attacks and Defenses for the Vulnerability of the Decade*.
- [13] Mallikarajunan, Narasima, et al. *DETECTION of SPYWARE in SOFTWARE USING VIRTUAL ENVIRONMENT*. Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore, 2019.
- [14] Liu, Jing, et al. “Botnet: Classification, Attacks, Detection, Tracing, and Preventive Measures.” *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, 13 Sept. 2009, <https://doi.org/10.1155/2009/692654>.
- [15] Wilbur, Kenneth, and Yi Zhu. *Click Fraud Full Terms and Conditions of Use: Http://Pubsonline.informs.org/Page/Terms-And-Conditions Click Fraud*. Vol. 28, no. 2, 2009, <https://doi.org/10.1287/mksc.1080.0397>. Accessed 9 May 2025.
- [16] Rudd, Ethan M., et al. “A Survey of Stealth Malware Attacks, Mitigation Measures, and Steps toward Autonomous Open World Solutions.” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, 2017, pp. 1145–1172, <https://doi.org/10.1109/comst.2016.2636078>.
- [17] Weaver, Nicholas, et al. *A Taxonomy of Computer Worms* \*. 2003.
- [18] Typel, S., and G. Baur. “Theory of the Trojan–Horse Method.” *Annals of Physics*, vol. 305, no. 2, June 2003, pp. 228–265, [https://doi.org/10.1016/s0003-4916\(03\)00060-5](https://doi.org/10.1016/s0003-4916(03)00060-5).
- [19] Sheng, Victor. *Sensitivity of Different Machine Learning Algorithms to Noise*. 2015.

- [20] Shah, Saleh, et al. "Compromised User Credentials Detection in a Digital Enterprise Using Behavioral Analytics." *Future Generation Computer Systems*, vol. 93, Apr. 2019, pp. 407–417, <https://doi.org/10.1016/j.future.2018.09.064>.
- [21] Aiyanyo, Imatitikua D., et al. "A Systematic Review of Defensive and Offensive Cybersecurity with Machine Learning." *Applied Sciences*, vol. 10, no. 17, 22 Aug. 2020, p. 5811, <https://doi.org/10.3390/app10175811>.
- [22] Bellovin, S.M., and W.R. Cheswick. "Network Firewalls." *IEEE Communications Magazine*, vol. 32, no. 9, Sept. 1994, pp. 50–57, <https://doi.org/10.1109/35.312843>.
- [23] Sukwong, Orathai, et al. *Commercial Antivirus Software Effectiveness: An Empirical Study*. IEEE Computer Society, Mar. 2011.
- [24] Sanchez, Javier, et al. *Encryption Techniques: A Theoretical Overview and Future Proposals*.
- [25] Shaker Ashoor, Asmaa , and Sharad Gore. *Importance of Intrusion Detection System (IDS)*. International Journal of Scientific & Engineering Research, vol. 2, no. 1, Jan. 2011.
- [26] Sahare, Bhawana, et al. "Study of Ethical Hacking." *International Journal of Computer Science Trends and Technology*, vol. 2, 2014.
- [27] Applebaum, Andy, et al. "Intelligent, Automated Red Team Emulation." *Proceedings of the 32nd Annual Conference on Computer Security Applications*, 5 Dec. 2016, pp. 363–373, <https://doi.org/10.1145/2991079.2991111>.
- [28] Shin, Bongsik, and Paul Benjamin Lowry. "A Review and Theoretical Explanation of the "Cyberthreat-Intelligence (CTI) Capability" That Needs to Be Fostered in Information Security Practitioners and How This Can Be Accomplished." *Computers & Security*, vol. 92, May 2020, p. 101761, <https://doi.org/10.1016/j.cose.2020.101761>.
- [29] Sokol, Pavol, et al. "Honeypots and Honeynets: Issues of Privacy." *EURASIP Journal on Information Security*, vol. 2017, no. 1, 28 Feb. 2017, <https://doi.org/10.1186/s13635-017-0057-4>.
- [30] Ali, Md Liakat, et al. "Challenges of Cyber Security and the Emerging Trends." *Proceedings of the 2019 ACM International Symposium on Blockchain and Secure Critical Infrastructure*, 2 July 2019, pp. 107–112, <https://doi.org/10.1145/3327960.3332393>.
- [31] Sheikh, Haroon, et al. "Artificial Intelligence: Definition and Background." *Research for Policy*, 2023, pp. 15–41, [https://doi.org/10.1007/978-3-031-21448-6\\_2](https://doi.org/10.1007/978-3-031-21448-6_2).

- [32] Lele, Ajey. “Artificial Intelligence (AI).” *Smart Innovation, Systems and Technologies*, 29 Dec. 2018, pp. 139–154, [https://doi.org/10.1007/978-981-13-3384-2\\_8](https://doi.org/10.1007/978-981-13-3384-2_8).
- [33] Horvitz, Eric, and Deirdre Mulligan. “Data, Privacy, and the Greater Good.” *Science*, vol. 349, no. 6245, 17 July 2015, pp. 253–255, <https://doi.org/10.1126/science.aac4520>.
- [34] Bell, Jason . *How Machines Learn Section III*. John Wiley & Sons Ltd, 2022. Machine Learning and the City: Applications in Architecture and Urban Design,.
- [35] Hao, Xing, et al. “Deep Learning.” *International Journal of Semantic Computing*, vol. 10, no. 03, Sept. 2016, pp. 417–439, <https://doi.org/10.1142/s1793351x16500045>.
- [36] Agatonovic-Kustrin, S, and R Beresford. “Basic Concepts of Artificial Neural Network (ANN) Modeling and Its Application in Pharmaceutical Research.” *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, June 2000, pp. 717–727, [https://doi.org/10.1016/s0731-7085\(99\)00272-1](https://doi.org/10.1016/s0731-7085(99)00272-1).
- [37] Mateus, Balduino César, et al. “Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press.” *Energies*, vol. 14, no. 21, 22 Oct. 2021, p. 6958, <https://doi.org/10.3390/en14216958>.
- [38] Razavi, S., and B. A. Tolson. “A New Formulation for Feedforward Neural Networks.” *IEEE Transactions on Neural Networks*, vol. 22, no. 10, Oct. 2011, pp. 1588–1598, <https://doi.org/10.1109/tnn.2011.2163169>.
- [39] Zhou, Ming, et al. “Progress in Neural NLP: Modeling, Learning, and Reasoning.” *Engineering*, vol. 6, no. 3, Mar. 2020, pp. 275–290, <https://doi.org/10.1016/j.eng.2019.12.014>.
- [40] Patwardhan, Narendra, et al. “Transformers in the Real World: A Survey on NLP Applications.” *Information*, vol. 14, no. 4, 17 Apr. 2023, p. 242, <https://doi.org/10.3390/info14040242>.
- [41] Wirkuttis , Nadine , and Hadas Klein. “Artificial Intelligence in Cybersecurity.” *Cyber, Intelligence, and Security*, vol. 1, no. 1, 2017, p. 103.
- [42] Khan, Muhammad Ismaeel, Aftab, Arif, Ali, Raza A Khan<sup>3</sup>et al. *The Most Recent Advances and Uses of AI in Cybersecurity*. Jurnal Multidisiplin Ilmu , 2024.

- [43] Truptiben Dave, Konark. *Brute-Force Attack “Seeking but Distressing.”* International Journal of Innovations in Engineering and Technology (IJET), June 2013.
- [44] Folorunso, Adebola, et al. “Impact of AI on Cybersecurity and Security Compliance.” *Global Journal of Engineering and Technology Advances*, vol. 21, no. 1, 30 Oct. 2024, pp. 167–184, <https://doi.org/10.30574/gjeta.2024.21.1.0193>.
- [45] Khurram Hassan, Syed , and Asif Ibrahim. “Electronic Crime Investigation.” *International Journal for Electronic Crime Investigation*, vol. 7, no. 2, 15 June 2023, <https://doi.org/10.54692/ijeci.2023.0702154>.
- [46] Shoeb Syed, Ali. *Adversarial AI and Cybersecurity: Defending against AI-Powered Cyber Threats*. ICONIC RESEARCH AND ENGINEERING JOURNALS, vol. 8, no. 9, Mar. 2025.
- [47] Beaman, Craig, et al. “Ransomware: Recent Advances, Analysis, Challenges and Future Research Directions.” *Computers & Security*, vol. 111, Dec. 2021, p. 102490, <https://doi.org/10.1016/j.cose.2021.102490>.
- [48] Blake, Harrison. *AI-Powered Cyber Extortion: How Artificial Intelligence Is Transforming Ransomware Attacks*.
- [49] Whyte, Christopher. “Deepfake News: AI-Enabled Disinformation as a Multi-Level Public Policy Challenge.” *Journal of Cyber Policy*, vol. 5, no. 2, 3 May 2020, pp. 199–217, <https://doi.org/10.1080/23738871.2020.1797135>.
- [50] Liaqat, Muhammad, et al. “Exploring Phishing Attacks in the AI Age: A Comprehensive Literature Review.” *Journal of Computing & Biomedical Informatics* , vol. 7, no. 2, <https://doi.org/10.56979/702/2024>.
- [51] Ayub, Md. Ahsan, et al. “Model Evasion Attack on Intrusion Detection Systems Using Adversarial Machine Learning.” *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2020, pp. 1–6, <https://doi.org/10.1109/ciss48834.2020.1570617116>.
- [52] Yerlikaya, Fahri Anıl, and Şerif Bahtiyar. “Data Poisoning Attacks against Machine Learning Algorithms.” *Expert Systems with Applications*, vol. 208, Dec. 2022, p. 118101, <https://doi.org/10.1016/j.eswa.2022.118101>.

[53] Ansari, Meraj Farheen, et al. “The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review.” *IJARCCCE*, vol. 11, no. 9, 30 Sept. 2022, <https://doi.org/10.17148/ijarcce.2022.11912>.

[54] Das, Rammanohar, and Raghav Sandhane. “Artificial Intelligence in Cyber Security.” *Journal of Physics: Conference Series*, vol. 1964, no. 4, 1 July 2021, p. 042072, <https://doi.org/10.1088/1742-6596/1964/4/042072>.