



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

Ασφάλεια Πληροφοριακών και Επικοινωνιακών Συστημάτων

**Η Τεχνητή Νοημοσύνη τελικά βελτιώνει ή επιδεινώνει την
κυβερνοασφάλεια;**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Μάρκου Κανιούρα

Επιβλέπων : Κρητικός Κυριάκος

Μέλη εξεταστικής επιτροπής: Κοκολάκης Σπύρος, Στεργιόπουλος Γεώργιος

Σάμος, Νοέμβριος 2023

Η σελίδα αυτή είναι σκόπιμα λευκή

Πρόλογος και ευχαριστίες

Θέλω να εκφράσω τις ειλικρινείς μου ευχαριστίες στον κ. Κρητικό Κυριάκο, επιβλέπων καθηγητή, που με την συμβουλή του και την υποστήριξη του συνέβαλε στην ολοκλήρωση αυτής της διπλωματικής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου, που με στήριξε και με ενέπνευσε καθ' όλη την διάρκεια αυτής της πορείας.

© 2023

του

Μάρκου Κανιούρα

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας περιεχομένων

1	Εισαγωγή	1
1.1	Τεχνητή Νοημοσύνη: Βοήθεια ή απειλή για τη διαδικτυακή ασφάλεια;	1
1.2	Αντικείμενο διπλωματικής	2
1.2.1	Συνεισφορά	3
1.2.2	Σε τι ωφελεί	3
1.3	Δομή της διπλωματικής	4
2	Υπόβαθρο Διπλωματικής	5
2.1	Τεχνητή Νοημοσύνη	5
2.1.1	Τι είναι η Τεχνητή Νοημοσύνη	5
2.1.2	Η εξέλιξη της Τεχνητής Νοημοσύνης και οι κυριότεροι τομείς της	7
2.1.3	Μέθοδοι Τεχνητής Νοημοσύνης	8
2.1.4	Εφαρμογές Τεχνητής Νοημοσύνης	10
2.2	Κυβερνοασφάλεια	11
2.2.1	Ορισμός Κυβερνοασφάλειας	11
2.2.2	Πεδία και τομείς Κυβερνοασφάλειας	11
2.2.3	Απειλές Κυβερνοασφάλειας	14
2.2.4	Χρήση διαδικτύου και ζητήματα κυβερνο-ασφάλειας	15
2.2.5	Αντίμετρα και Μέθοδοι Προστασίας Κυβερνοασφάλειας	17
3	Μεθοδολογία βιβλιογραφικής ανασκόπησης	18
3.1	Κριτήρια επιλογής άρθρων	18
4	Επιθέσεις μέσω Τεχνητής Νοημοσύνης	20
4.1	Στάδια επίθεσης	20
4.2	Τεχνικές που χρησιμοποιούνται	23
4.2.1	Αποφυγή εντοπισμού (evasion)	24
4.2.2	Διήθηση σε δεδομένα (data exfiltration)	25
4.2.3	Απόσπαση κωδικών & δεδομένων εισόδου (password attacks)	27
4.2.4	Δημιουργία κακόβουλου λογισμικού (malware generation)	28
4.2.5	Τεχνική της «αξιοποίησης άγνωστων τρωτών σημείων» (zero-day exploits)	30
4.2.6	Αντίστροφη μηχανική μάθηση (adversarial machine learning)	32
4.2.7	Απάτη μέσω της τεχνικής του «ψαρέματος» (phishing)	34
4.2.8	Αυτοματοποιημένες επιθέσεις (automated attacks)	36
4.2.9	Τεχνική της «υποστηριζόμενης από ΤΝ επίθεσης κυβερνο-ρομπότ» (AI-powered bots)	38
4.2.10	Τεχνική αξιοποίησης υπερρεαλιστικών μέσων (deep fakes)	39
4.3	Συνήθεις στόχοι & κίνητρα των κυβερνοεπιθέσεων	40
5	Αντίμετρα	43
5.1	Αντίμετρα και η συνεισφορά τους	43
5.2	Ορισμός Αντιμέτρων της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια	44
5.3	Αντίμετρα χωρίς αξιοποίηση ΤΝ	47
5.4	Αντίμετρα με αξιοποίηση ΤΝ	52

5.4.1	Εντοπισμός ανωμαλιών (anomaly detection) & αναλυτική συμπεριφοράς (behavior analytics).....	53
5.4.2	Μηχανική μάθηση για εντοπισμό απειλών (machine learning for threat detection) & σε βάθος μάθηση για εντοπισμό εισβολής (deep learning for intrusion detection)	53
5.4.3	Εντοπισμός κακόβουλου λογισμικού (malware detection) & πρακτικών phishing (phishing detection)	54
5.4.4	Προστασία τελικών σημείων (endpoint security) & προστασία δικτύου (network security)	54
5.4.5	Πληροφόρηση Ασφαλείας & Διαχείριση Περιστατικών (Security Information & Event Management, SIEM)	55
5.4.6	Ενορχήστρωση προστασίας & αυτοματισμών (Security Orchestration & Automation)	55
5.4.7	TN για προστασία νέφους (cloud security) & αναλυτική ασφαλείας (security analytics)	56
5.5	Συσχέτιση Αντίμετρων με τις Φάσεις Επιθέσεων που Αντιμετωπίζουν	57
5.6	Θετικά και Αρνητικά Αντιμέτρων Κυβερνοασφάλειας	61
5.7	Προβλήματα & προκλήσεις από τη χρήση των αντιμέτρων TN	63
6	Συμπεράσματα	68
6.1	Συνεισφορά διπλωματικής	68
6.2	Βασικά συμπεράσματα.....	70
	Βιβλιογραφία	73

Λίστα Σχημάτων

Σχήμα 1. Λειτουργίες Τεχνητής Νοημοσύνης	6
Σχήμα 2. Τύποι Τεχνητής Νοημοσύνης	8
Σχήμα 3. Πεδία που καλύπτει η κυβερνο-ασφάλεια	14
Σχήμα 4. Επεξήγηση της τεχνικής της “επίθεσης κωδικών” με τη μορφή της “ωμής βίας” (brute force)	27
Σχήμα 5. Οι πλέον γνωστοί τύπο κακόβουλων λογισμικών	30
Σχήμα 6. Επεξήγηση του εντοπισμού τρωτών σημείων σε επιθέσεις τύπου “zero-day”	32
Σχήμα 7. Διαφορές παραδοσιακής & αντίστροφης μηχανικής μάθησης	33
Σχήμα 8. Αναπαράσταση του τρόπου λειτουργίας μιας επίθεσης τύπου “phishing”	36
Σχήμα 9. Αναπαράσταση του αυτοματοποιημένου τρόπου επιθέσεων μέσω TN	37
Σχήμα 10. Ποσοστά συμβάντων χρηστών που διαπίστωσαν απόπειρες απάτης μέσω χρήσης deepfake πρακτικών (στοιχεία 2022, σε διεθνή κλίμακα)	40

Λίστα Πινάκων

Πίνακας 1. Συγκεντρωτικός πίνακας φάσεων επίθεσης & αντιμέτρων	57
--	----

Ακρωνύμια

AI	Artificial Intelligence
AD	Anomaly Detection
TN	Τεχνητή Νοημοσύνη
NPL	Natural Language Processing
CNNs	Convolutional Neural Networks
RNNs	Recurrent Neural Networks
IDS	Intrusion Detection Systems
VPNs	Virtual Private Networks
GANs	Generative Adversarial Networks
UFL	Unsupervised Feature Learning

Περίληψη

Η Τεχνητή Νοημοσύνη (TN) δεν αποτελεί κάποια νέα τεχνολογία για τον άνθρωπο, ωστόσο δε μετρά ακόμα πολλά χρόνια που αξιοποιείται στην προστασία πληροφοριακών συστημάτων, τάση που σημειώνει διαρκή άνοδο. Όπως όμως προστατεύει, μπορεί να χρησιμοποιηθεί και για τη δημιουργία απειλών, όπως στην περίπτωση κακόβουλου λογισμικού, το οποίο μπορεί να επιτρέψει πρόσβαση σε ευαίσθητες πληροφορίες ή να οδηγεί ακόμη και σε βλάβες των συστημάτων.

Σκοπός, λοιπόν, της παρούσας διπλωματικής, είναι η καταγραφή των πληροφοριών και τεχνικών/μεθόδων που υποστηρίζονται από τη TN και συνδέονται με την προστασία των πληροφοριακών συστημάτων και των δεδομένων, από απειλές του κυβερνοχώρου. Ωστόσο, επειδή η TN αξιοποιείται και για επιθετικές ενέργειες, δε μπορούν να λείπουν από το σύνολο αυτών των πληροφοριών οι πρακτικές επίθεσης στις οποίες αξιοποιείται η TN, αφού καταβλήθηκε προσπάθεια να προσεγγιστεί το ζήτημα αυτό ολιστικά. Η μέθοδος που χρησιμοποιήθηκε είναι αυτή της βιβλιογραφικής ανασκόπησης, έτσι ώστε να μπορέσει η εργασία να συμπεριλάβει μεγάλο αριθμό διαφορετικών πληροφοριών και να μην ασχοληθεί αποκλειστικά με μια από τις δεκάδες πτυχές του ζητήματος. Αξιοποιήθηκαν σύγχρονες πληροφορίες (όχι παλαιότερες της δεκαετίας και ως επί το πλείστον ανήκουσες στην τελευταία εξαετία) από γνωστές πηγές δεδομένων (Google Scholar και Web of Science) και χρησιμοποιήθηκαν συγκεκριμένα λήμματα για τον εντοπισμό των απαραίτητων πληροφοριών, ενώ αποκλείστηκαν άρθρα και πληροφορίες που δεν προέρχονταν είτε από έγκυρα επιστημονικά περιοδικά είτε από επίσημους οργανισμούς, προκειμένου η εργασία να χαρακτηρίζεται από περιεχόμενο υψηλής εγκυρότητας και ποιότητας.

Τα συμπεράσματα στα οποία κατέληξε η παρούσα εργασία, έδειξαν πως η TN όντως μπορεί να συμβάλλει σημαντικά στη βελτίωση της κυβερνο-ασφάλειας, ωστόσο φαίνεται ότι τα καλύτερα αποτελέσματα προκύπτουν κατόπιν συμμετοχής και του ανθρώπου. Επίσης, όπως συμβάλλει στη βελτίωση της ασφάλειας, έτσι η TN μπορεί να βοηθήσει και τους επιτιθέμενους, χάριν σε μεθόδους, όπως οι αντιπαραθετικές επιθέσεις (adversarial attacks) και η μηχανική μάθηση. Η εξαπάτηση, η αποφυγή εντοπισμού και η ανεύρεση τρωτών σημείων, αποτελούν πρακτικές που εφαρμόζονται με έξυπνο και μεθοδικό τρόπο από την TN, ενώ τα αντίμετρα που βασίζονται στην TN περιλαμβάνουν εντοπισμό ανωμαλιών, ανάλυση συμπεριφοράς διαδικτύου και μηχανική μάθηση. Ο ανθρώπινος παράγοντας, ωστόσο, χρειάζεται στην ανάπτυξη αντιμέτρων για τη βαθύτερη κατανόηση του περιβάλλοντος ασφάλειας των οργανισμών, τις αναβαθμίσεις και γενικά μια σειρά απαραίτητων ενεργειών που βελτιώνουν τις επιδόσεις της TN ως προς την κυβερνο-ασφάλεια.

Λέξεις κλειδιά: *Τεχνητή Νοημοσύνη, κυβερνο-ασφάλεια, κίνδυνοι, επιθέσεις, απειλές, αντίμετρα*

Abstract

Artificial Intelligence (AI) does not constitute a new technology to humans; Nevertheless, it does not yet count many years during which it has been used to protect information systems, a trend that is constantly increasing. Though, as it protects, it also can be used to set up threats, like malware, a threat able to allow access to sensitive information or even lead to serious damages to systems.

The present study aims to record information, as well as techniques/methods supported by AI and connected to the protection of data and information systems from cyberthreats. However, due to AI's usage also for offensive acts, the attack practices driven by AI cannot be outside the set of such information, since an effort to approach the issue from every corner has been taken place. The literature review is the method used to fulfill the present dissertation as it can incorporate a vast number of different information, and not just deal exclusively with one of the dozen aspects related to this issue. Up-to-date information has been used (not older than a decade, and mostly within the last six years), discovered in well-known data sources (such as Google Scholar and Web Sciences) where well-suited terms have been used to identify appropriate information. Further, articles and other types of information outside valid scientific journals and official organizations, have been excluded, in order to attribute to the dissertation the highest validity and quality.

Conclusions to which the dissertation ended up, prove that AI does significantly contribute to cybersecurity improvement. However, it seems that the production of best results requires human participation. Additionally, as AI contributes to cybersecurity improvement, it also can assist attackers throughout methods, such as adversarial attacks and machine learning. Deception, evasion and vulnerability identification, constitute practices applied cleverly and methodically by AI, while countermeasures based on AI include anomalies detection, web-behavior analysis and machine learning. The human factor, though, is definitely required in the development of countermeasures to enable the deep understanding of the organizations' security environment, the necessary updates, and in general, a series of acts which can enhance AI's performance in cyber-security.

Keywords: *Artificial Intelligence, cyber security, risks, attacks, threats, countermeasures*

1 Εισαγωγή

1.1 Τεχνητή Νοημοσύνη: Βοήθεια ή απειλή για τη διαδικτυακή ασφάλεια;

Η Τεχνητή Νοημοσύνη (TN), αν και είναι γνωστή σε επιστημονικούς, τεχνολογικούς και ερευνητικούς κύκλους εδώ και δεκαετίες, αποτελεί την πλέον πρόσφατη τάση σε όλα τα προαναφερθέντα πεδία. Όμως, η TN έχει εγείρει αρκετά ερωτήματα για το κατά πόσο πράγματι μπορεί να ωφελήσει τον άνθρωπο και αν μπορεί να προκαλέσει προβλήματα σε μια σειρά ζητημάτων, από την κάλυψη των θέσεων εργασίας έως την ασφάλεια των προσωπικών του δεδομένων (Ardito et al., 2021). Ως εκ τούτου, είναι ένα πεδίο το οποίο χρήζει διαρκούς έρευνας τόσο σε θεωρητικό επίπεδο (για την αναζήτηση ίσως της ηθικής πίσω από την ορθή χρήση της) όσο και σε επίπεδο ποιοτικής και ποσοτικής έρευνας, προκειμένου να εντοπιστούν τα όποια πιθανά μελανά σημεία και να προταθούν εγκαίρως λύσεις, επωφελείς για τον άνθρωπο.

Σκοπός, λοιπόν, της παρούσας διπλωματικής, είναι αρχικά να καταγράψει περιεκτικές και χρήσιμες πληροφορίες σχετικά με τη χρησιμότητα της TN, τις εφαρμογές της στην καθημερινότητα του ανθρώπου (επαγγελματική, καθημερινή διαβίωση, διαχείριση οικονομικών κ.ά.) και τους κινδύνους που αυτή εγκυμονεί εάν δεν ελεγχθεί από αντικειμενικές πηγές. Ο βασικός στόχος, όμως, της διπλωματικής, είναι να καταγράψει τους κινδύνους που ελλοχεύουν σε επίπεδο ασφάλειας σε διαδικτυακά περιβάλλοντα και κατά πόσο η χρήση της TN είναι ικανή να βελτιώσει την ασφάλεια αναφορικά με την ολοένα και αυξανόμενη σύνδεση των καθημερινών υποχρεώσεων των ανθρώπων, με το διαδίκτυο, παράγοντας που συνοδεύεται - σε αρκετές περιπτώσεις - από την αξιοποίηση προσωπικών δεδομένων των χρηστών ή στοιχείων που αφορούν τις οικονομικές τους συναλλαγές. Συνεπώς, η διπλωματική εργασία θα εξετάσει το βαθμό στον οποίο η TN είναι ικανή να βελτιώσει τη διαδικτυακή ασφάλεια (ή κυβερνο-ασφάλεια) ή να αποτελέσει τροχοπέδη για αυτήν. Αξίζει τέλος να σημειωθεί, ότι το αντικείμενο της παρούσας διπλωματικής σχετίζεται με πλήθος υπαρχόντων εργασιών τόσο στο εξωτερικό όσο και στην Ελλάδα. Ωστόσο ο λόγος που την

καθιστά καλύτερη, είναι αρχικά οι πολύ σύγχρονες πληροφορίες που διαθέτει και η συγκέντρωση πλήθους διαφορετικών πτυχών που άπτονται του ζητήματος, γεγονός που την καθιστά περισσότερο συμπεριληπτική.

1.2 Αντικείμενο διπλωματικής

Το αντικείμενο της διπλωματικής είναι πρωτίστως οι τρόποι με τους οποίους μπορεί η TN είτε να συμβάλλει στη βελτίωση της κυβερνο-ασφάλειας είτε να την υποδαυλίζει και κατά δεύτερον, η σε βάθος ενασχόληση με το περιεχόμενο και τις προεκτάσεις της TN στην καθημερινότητα. Η εργασία περιλαμβάνει εκτενή παρουσίαση της TN τόσο από την μεριά των μεθόδων που χρησιμοποιούνται για την αποτελεσματική της λειτουργία και των πρακτικών μέσα από τις οποίες προστατεύονται οι χρήστες του διαδικτύου όσο και από την μεριά των μεθόδων με τις οποίες μπορούν τα δεδομένα χρηστών και οργανισμών να εκτεθούν σε κακόβουλους χρήστες, οδηγώντας έτσι σε ανεπιθύμητες καταστάσεις, όπως οι εκβιασμοί και η δημόσια διαπόμπευση καθώς και οι οικονομικές απάτες.

Ένα από τα κυριότερα προβλήματα που αναμένεται να διαλευκάνει η παρούσα εργασία, είναι η κατανόηση του πλήθους και του είδους των κινδύνων στους οποίους μπορούν να εκτεθούν οι χρήστες, εξαιτίας της ανεξέλεγκτης, κακόβουλης ή λανθασμένης αξιοποίησης της TN σε μεγάλο φάσμα διαδικτυακών δραστηριοτήτων. Ξεκινώντας από την όσο το δυνατόν πιο κατανοητή και απλοϊκή παρουσίαση των όρων που χρησιμοποιούνται στο χώρο της TN και του διαδικτύου, η εργασία θα προχωρήσει στην παρουσίαση των κινδύνων καθώς και των μεθόδων και επιθέσεων που οι διάφοροι επιτήδριοι χρησιμοποιούν, προκειμένου να παραβιάσουν την προστασία των συστημάτων και να έρθουν σε επαφή με πληροφορίες που θα μπορούν να αξιοποιήσουν προς ίδιον όφελος.

Οι λύσεις που θα παρουσιάσει η εργασία, αποτελούν στην ουσία αμυντικούς μηχανισμούς και ενέργειες χάριν στις οποίες οι χρήστες μπορούν να διασφαλίσουν ότι οι πιθανότητες παραβίασης της ασφάλειας των συστημάτων τους ελαχιστοποιούνται. Η παρουσίαση των επιλογών που μπορούν να προστατεύσουν τους χρήστες, σίγουρα δεν αναμένεται να λύσει ολοκληρωτικά το πρόβλημα που γεννάται από την αξιοποίηση της TN σε περιστατικά παραβίασης της κυβερνο-ασφάλειας - δεδομένου ότι οι κακόβουλοι χρήστες συνήθως βρίσκονται ένα βήμα μπροστά από τα μέτρα προστασίας - ωστόσο οι επιλογές αυτές μπορούν να αποτελέσουν ένα χρήσιμο οδηγό προς αυτή την κατεύθυνση, βελτιώνοντας το γνωστικό υπόβαθρο των χρηστών και καθιστώντας τους περισσότερο ενήμερους για τις τεχνικές που χρησιμοποιούνται. Η ενημέρωση ή ο εντοπισμός εξάλλου για καχύποπτες ή κακόβουλες ενέργειες, αποτελεί μια εξαιρετικά σημαντική παράμετρο

που είθισται να ωφελεί τους χρήστες των διαδικτυακών υπηρεσιών, οι οποίοι μπορεί έπειτα να αποφύγουν τις ενέργειες αυτές ή να τις μετριάσουν.

1.2.1 Συνεισφορά

Κρίνοντας λοιπόν από τα παραπάνω, η εργασία αναμένεται να σημειώσει αξιόλογη συνεισφορά προς το αναγνωστικό κοινό. Ο λόγος έγκειται στο γεγονός ότι θα παρουσιάσει μια συγκεντρωτική και περιεκτική ανάλυση τόσο των κινδύνων που αναδύονται σε επίπεδο κυβερνο-ασφάλειας, από την αυξανόμενη χρήση της TN, όσο και των πιθανών λύσεων που θα μπορούσαν να αξιοποιήσουν οι χρήστες με βάση τη TN. Μπορεί, λοιπόν, ερευνητικά η εργασία να μην προσφέρει πρωτογενή στατιστικά δεδομένα όσον αφορά την προβληματική που εκθέτει αλλά αναμφίβολα μπορεί να αποτελέσει έναν οδηγό ίσως και για πιο μικρές ηλικίες, αφού επικεντρώνεται σε συγκεκριμένα ζητήματα, όπως οι πρακτικές βελτίωσης της κυβερνο-ασφάλειας ή η ενημέρωση για μεθόδους προσπάθειας παραβίασής της, για ένα ζήτημα που μονοπωλεί συχνά το ενδιαφέρον των σύγχρονων συζητήσεων.

1.2.2 Σε τι ωφελεί

Συνεπώς, από τα παραπάνω, η παρούσα διπλωματική ενσωματώνει μια σειρά από οφέλη. Για παράδειγμα, βοηθά τους αναγνώστες να έρθουν σε επαφή με έννοιες που ενδεχομένως δεν έχουν συναντήσει ξανά, καθιστώντας τους έτσι γνώστες σύγχρονων πληροφοριών που είναι πολύ πιθανό να τους βοηθήσουν να διατηρήσουν την ασφαλή πλοήγησή τους στο διαδίκτυο και τη χρήση υπηρεσιών που προσφέρει. Πιο συγκεκριμένα, γνωρίζοντας πώς αξιοποιείται η TN στη διεξαγωγή κυβερνο-επιθέσεων, καθώς και τις επιλογές που έχει διαθέσιμες για να προστατευτεί από αυτές, μπορεί να λάβει κατάλληλες αποφάσεις. Επιπλέον, καθίσταται γνώστης ενός ζητήματος που απασχολεί και θα συνεχίσει να απασχολεί την ανθρωπότητα τα επόμενα έτη. Συνεπώς, η κατανόηση του περιεχομένου της διπλωματικής θα βοηθήσει τον αναγνώστη να έρθει σε πιο στενή επαφή με αυτό το ζήτημα. Τέλος, παρουσιάζει μια λίστα επιλογών χάριν στις οποίες μπορεί να προστατευτεί, καλύπτοντας έτσι όλο το φάσμα “*πρόβλημα-αίτια-λύσεις*”.

Η πλειοψηφία των υπολοίπων άρθρων, παρουσιάσεων και αναλύσεων που εντοπίστηκαν είτε με ελεύθερη αναζήτηση στο διαδίκτυο (σε έγκυρες ιστοσελίδες μόνο, οργανισμών ή εταιρειών με αντικείμενο την κυβερνοασφάλεια) είτε σε βάσεις δεδομένων, επικεντρώνονται συνήθως σε ένα ζήτημα γύρω από τις υποστηριζόμενες από την TN κυβερνο-επιθέσεις και τη διάρρηξη της κυβερνο-ασφάλειας, με τις αντίστοιχες πιθανές επιλογές αντιμετώπισης. Η παρούσα διπλωματική

προσπαθεί να καλύψει ένα ευρύ πεδίο και να συμπεριλάβει το σύνολο αυτών των περιπτώσεων, ώστε να προσφέρει στους αναγνώστες της μια πολύ πιο σφαιρική εικόνα τόσο των τεχνικών κυβερνο-επίθεσης όσο και των αντιμέτρων που βασίζονται στην ΤΝ. Έτσι, δίνονται στους χρήστες πολλές περισσότερες επιλογές και παρέχεται μεγαλύτερο εύρος γνώσεων, αυξάνοντας τη μοναδικότητα του περιεχομένου της διπλωματικής.

1.3 Δομή της διπλωματικής

Η παρούσα διπλωματική αποτελείται συνολικά από 7 μέρη, συμπεριλαμβανομένης του τρέχοντος κεφαλαίου (δηλ. του πρώτου) και αυτού της βιβλιογραφίας (έβδομο κεφάλαιο) που αξιοποιήθηκε για την εκπόνησή της. Στο δεύτερο κεφάλαιο παρουσιάζεται η έννοια της Τεχνητής Νοημοσύνης, της κυβερνοασφάλειας και των αντιμέτρων, με μια σειρά χρήσιμων πληροφοριών που άπτονται των τριών αυτών πεδίων, ενώ εισάγει τον αναγνώστη εν συντομία, σε έννοιες που θα συναντήσει πιο αναλυτικά στην πορεία της εργασίας. Το τρίτο κεφάλαιο αποτυπώνει τη μέθοδο που χρησιμοποιήθηκε για τη διάκριση και επιλογή των κατάλληλων πληροφοριών, ενώ στο τέταρτο κεφάλαιο ο αναγνώστης συναντά αναλυτικές πληροφορίες σχετικά με τις επιθέσεις που υποστηρίζονται από την Τεχνητή Νοημοσύνη. Το πέμπτο κεφάλαιο της εργασίας περιλαμβάνει τα αντίμετρα που μπορεί κάποιος να εφαρμόσει, ώστε να αντιμετωπίσει τις επιθέσεις, αντίμετρα που είτε χρησιμοποιούν είτε όχι τις δυνατότητες της Τεχνητής Νοημοσύνης, αφού η ίδια δε μπορεί να αξιοποιηθεί μόνο για επιθετικές ενέργειες αλλά και για αμυντικό σκοπό. Στο έκτο κεφάλαιο καταγράφονται τα συμπεράσματα στα οποία κατάληξε ο συγγραφέας της διπλωματικής.

2 Υπόβαθρο Διπλωματικής

2.1 Τεχνητή Νοημοσύνη

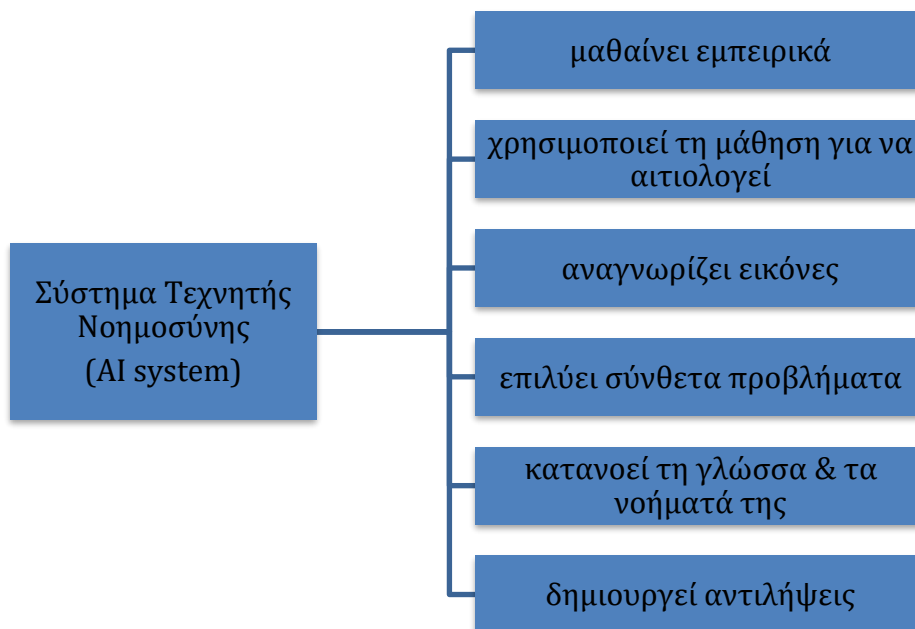
2.1.1 Τι είναι η Τεχνητή Νοημοσύνη

Με τον όρο Τεχνητή Νοημοσύνη (αγγλικά: Artificial Intelligence, AI) υποδηλώνεται το πολυεπιστημονικό πεδίο, το οποίο ως αντικείμενο έχει την αναζήτηση των μεθόδων που θα οδηγήσουν τις μηχανές στην ικανότητα να εκτελούν καθήκοντα για τα οποία μέχρι πρόσφατα απαιτούνταν η παρουσία ή δράση του ανθρώπου (European Union Agency for Network & Information Security, 2015). Στην ουσία, πρόκειται για ένα παρακλάδι της επιστήμης των υπολογιστών (Computer Science) που εξελίσσεται εδώ και δεκαετίες, έχοντας πλέον ενσωματώσει ένα μεγάλο εύρος τεχνικών, μεθόδων και εφαρμογών. Ο υπέρτατος στόχος της ΤΝ, είναι η ανάπτυξη συστημάτων ικανών να μιμούνται τη γνωστική λειτουργία του ανθρώπου, όπως για παράδειγμα η μάθηση (learning), η αιτιολόγηση (reasoning), η επίλυση προβλημάτων (problem solving), η αντίληψη (perception) και η κατανόηση μιας ομιλούσας (ή μη) γλώσσας (language understanding) (Σχήμα 1). Ωστόσο, αυτό που καθιστά ιδιαίτερα ενδιαφέρουσα την ενασχόληση με την ΤΝ, είναι η προσπάθεια που έχει καταβληθεί ώστε οι ικανότητές της να ξεπερνούν τις αντίστοιχες του ανθρώπου (El Hadi, 2023).

Η ΤΝ θεωρείται - δικαιολογημένα πλέον - ως η τεχνολογία που καθόρισε την προηγούμενη δεκαετία και πιθανόν τις επερχόμενες. Από τις πρώτες μέρες της, ο ορισμός της ΤΝ έχει επικεντρωθεί στην ικανότητα μιας μηχανής να μπορεί να συμπεριφερθεί ευφυώς, με διάφορα είδη “Turing test” (μέθοδος προσδιορισμού του βαθμού στον οποίο μια μηχανή μπορεί να προσομοιάσει την ανθρώπινη νοημοσύνη) να διακηρύσσουν τις μηχανές ως ευφυείς (intelligent), αφού σε αντίθεση με αυτές, οι άνθρωποι δε μπορούν να διαφοροποιήσουν τις ενέργειές τους σε σχέση με αυτές που θεωρείται ότι αντικατοπτρίζουν έναν άνθρωπο (Bostrom, 2016). Με άλλα λόγια, οι μηχανές μπορούν να πραγματοποιήσουν ενέργειες που σε πολλές περιπτώσεις είναι πέρα από τις δυνατότητες ενός ανθρώπου, τουλάχιστον σε ταχύτητα και ακρίβεια (Baum, n.d.).

Οι σημερινοί ορισμοί της ΤΝ περιλαμβάνουν συχνά προϋποθέσεις, όπως η αυτονομία, και περιορίζουν την ευφυΐα σε συγκεκριμένα πεδία. Από το 2018, η Ευρωπαϊκή Επιτροπή (European Commission) έχει υιοθετήσει τον ορισμό της ΤΝ ως “η αναφορά στα συστήματα που παρουσιάζουν ευφυή συμπεριφορά, αναλύοντας το περιβάλλον τους και λαμβάνοντας δράση, με ένα βαθμό

αυτονομίας, έτσι ώστε να επιτύχουν συγκεκριμένους στόχους” (Ευρωπαϊκή Επιτροπή, 2018). Ο ορισμός αυτός δε, θέτει περιορισμούς στις μεθόδους που χρησιμοποιούνται για την επίτευξη της ευφυίας. Αντιθέτως, ο όρος TN χρησιμοποιείται ως ομπρέλα κάτω από την οποία τοποθετούνται διάφορες τεχνολογίες και εφαρμογές που διαθέτουν ελάχιστα περισσότερα από τη φαινομενική τους ευφυΐα, δίνοντας ελεύθερο πεδίο σε ερμηνείες. Επιπροσθέτως, η πλειοψηφία των ανθρώπων συζητά για την χρήση της υφιστάμενης TN παράλληλα με αυτή που βρίσκεται υπό ανάπτυξη ή ακόμη και με αυτή που υποτίθεται ότι θα χρησιμοποιείται στο μέλλον. Συνεπώς, γίνεται αντιληπτό ότι ο όρος TN χρησιμοποιείται ευρέως για αναφορές σε οποιαδήποτε τεχνολογία, η οποία είτε χρησιμοποιείται τώρα είτε πιστεύεται ότι θα ακολουθήσει, από τη στιγμή που δηλώνεται ως η έκφραση χαρακτηριστικών τα οποία περιγράφουν κάποια ευφυή λειτουργία (OECD, 2016).



Σχήμα 1. Λειτουργίες Τεχνητής Νοημοσύνης

Πηγή:

<https://www.uc.edu/content/dam/uc/ce/docs/OLLI/Page%20Content/ARTIFICIAL%20INTELLIGENCEr.pdf>

Αυτή η αποκλειστικότητα, όμως, σημειώνει ορισμένες δυσκολίες όσον αφορά την εκτίμηση των επιπτώσεων από την ανάπτυξη της TN, επειδή αναλόγως με την οπτική που την παρατηρεί κάποιος - μεταξύ του μεγάλου εύρους των αντικειμένων που καλύπτει η εν λόγω τεχνολογία - είναι δυνατό να εντοπιστούν διαφορετικά οφέλη και κίνδυνοι. Ως εκ τούτου, η TN αποτελεί ένα πεδίο υψηλού, χαμηλού αλλά και όλων των ενδιάμεσων βαθμίδων, κινδύνου, μιας και από τη στιγμή που αναφέρεται σε πλήθος τεχνολογιών, εφαρμογών και περιεχομένου, απαιτείται μεγαλύτερη ακρίβεια

προκειμένου να προκύψει ουσιαστικός και εποικοδομητικός διάλογος και αντίστοιχα ασφαλή συμπεράσματα. Για παράδειγμα, τα επιχειρήματα σχετικά με τα απλά “εξειδικευμένα συστήματα” που αξιοποιούνται σε συμβουλευτικούς ρόλους, θα πρέπει να διακρίνονται από τα συστήματα που βασίζονται σε πολύπλοκους αλγορίθμους και οδηγούνται αυτομάτως σε λήψη αποφάσεων (Agrawal, Gans & Goldfarb, 2018).

2.1.2 Η εξέλιξη της Τεχνητής Νοημοσύνης και οι κυριότεροι τομείς της

Οι ρίζες της TN πηγαίνουν πολύ πίσω στο χρόνο, σε ιστορίες και μυθοπλασίες με περιεχόμενο μηχανοκίνητες κατασκευές με ανθρώπινα χαρακτηριστικά (Russell & Norving, 2010). Η σύγχρονη εποχή της TN, όμως, ξεκίνησε στα μέσα του 20ου αιώνα, όταν το 1956 στη συνδιάσκεψη του Dartmouth τέθηκαν οι βάσεις του πεδίου της, με τους ερευνητές να συγκεντρώνονται για να εκκινήσουν διεργασίες γύρω από το εν λόγω αντικείμενο. Η πρώτη έρευνα επάνω στην TN επικεντρώθηκε στη συμβολική της μορφή και περιλάμβανε την παρουσίαση γνώσης με τη μορφή συμβόλων και τη χρήση λογικής αιτιολόγησης στη λήψη αποφάσεων (Rid, 2016).

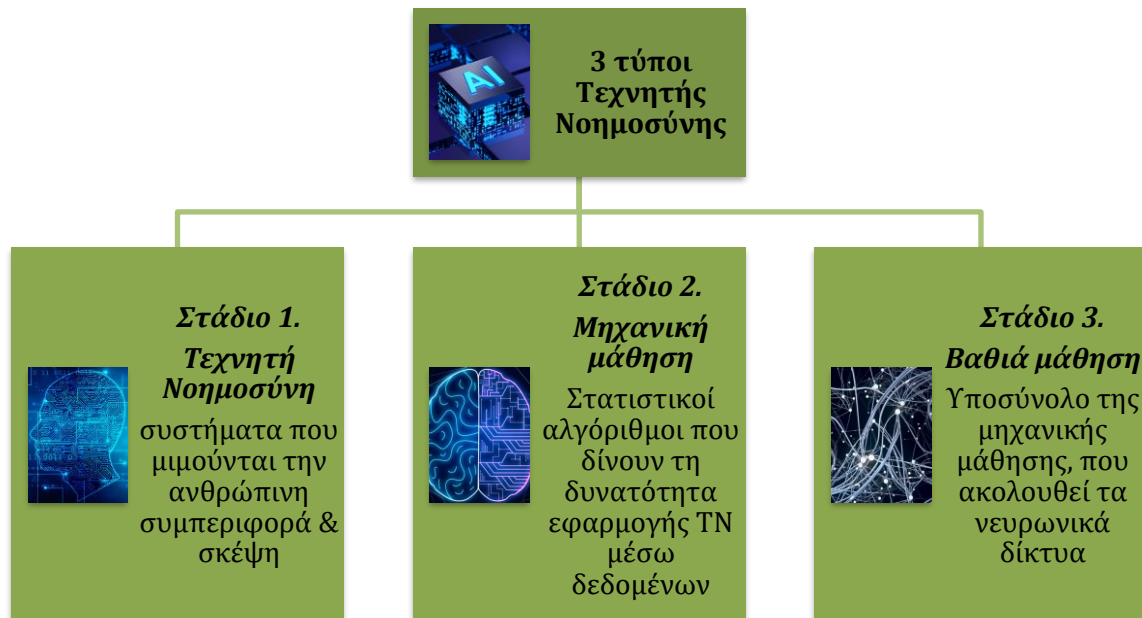
Σημαντικό σημείο στην εξέλιξή της, αποτελεί η έλευση της μηχανικής μάθησης (machine learning). Πρόκειται για ένα υποσύνολο της TN, το οποίο δίνει στα συστήματα τη δυνατότητα να μαθαίνουν από ένα σύνολο δεδομένων και να βελτιώνουν τις επιδόσεις τους σε βάθος χρόνου, χωρίς να λαμβάνουν ξεκάθαρο προγραμματισμό. Επιπροσθέτως, μια εκ των πλέον επιδραστικών εξελίξεων στη μηχανική μάθηση, είναι το αντικείμενο των νευρωνικών δικτύων (neural networks), το οποίο έχει εμπνευστεί από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου (Σχήμα 2).

Όσον αφορά την επεξεργασία φυσικής γλώσσας (αγγλικά: Natural Language Processing, NLP), αυτό επίσης αποτελεί τομέα της TN, το οποίο μάλιστα επικεντρώνεται στο να καταστήσει τα συστήματα ικανά να κατανοούν, να ερμηνεύουν και να δημιουργούν ανθρώπινη γλώσσα. Έχει διαδραματίσει καθοριστικό ρόλο σε εφαρμογές, όπως η μετάφραση, η ανάλυση συναισθημάτων και οι ψηφιακοί βοηθοί. Οι πιο πρόσφατες εξελίξεις στο πεδίο της NLP (Natural Language Processing) (αναγνώριση ονομάτων, διάσπαση κειμένου, ανάλυση συναισθήματος κειμένου κ.ά.) - όπως για παράδειγμα η ανάπτυξη του εργαλείου συζητήσεων “GPT-3” - έχουν βελτιώσει σε σημαντικό βαθμό τη δημιουργία γλώσσας και την κατανόησή της, οδηγώντας σε περισσότερο φυσική αλληλεπίδραση με τα συστήματα TN (ImamGuliyev, 2023).

Ένας ακόμα ουσιαστικός τομέας της TN, είναι η “υπολογιστική όραση” (computer vision), μέσα από το οποίο οι μηχανές καθίσταται εφικτό να ερμηνεύσουν και να κατανοήσουν οπτικές πληροφορίες στις οποίες έχουν πρόσβαση. Το πεδίο έχει οδηγήσει σε επανάσταση στο χώρο της

ανάλυσης εικόνων και βίντεο, την αναγνώριση αντικειμένων, την αναγνώριση προσώπου ή ακόμη και την κατασκευή αυτόνομων τηλεκατευθυνόμενων οχημάτων (drones) και ρομπότ, που μπορούν να αλληλοεπιδράσουν με το περιβάλλον τους (Karn 2021).

Τέλος, η βαθιά μάθηση (deep learning) αποτελεί ένα υποπεδίο της μηχανικής μάθησης, στο οποίο αξιοποιούνται σε βάθος τα νευρωνικά δίκτυα με πολλαπλές διαβαθμίσεις για τη μεταχείριση σύνθετων δεδομένων, οδηγώντας σε αξιοσημείωτα επιτεύγματα στους τομείς της αναγνώρισης εικόνας (image recognition), της επεξεργασίας φυσικής γλώσσας (natural language processing) και αρκετούς άλλους (Agrawal, Gans & Goldfarb, 2018).



Σχήμα 2. Τύποι Τεχνητής Νοημοσύνης

Πηγή: <https://www.bbvaopenmind.com/en/technology/artificial-intelligence/intellectual-abilities-of-artificial-intelligence/>

2.1.3 Μέθοδοι Τεχνητής Νοημοσύνης

Καθώς η τεχνητή νοημοσύνη είναι ένα πεδίο που εξελίσσεται γρήγορα υπάρχουν πολλές μέθοδοι που χρησιμοποιούνται για την επίλυση προβλημάτων. Ανάλογα με τον τύπο του προβλήματος και τον τρόπο που θέλουμε να αντιμετωπίσουμε το συγκεκριμένο τομέα επιλέγονται διάφορες μέθοδοι.

Στην μηχανική μάθηση (Machine Learning) υπάρχουν διάφορες μέθοδοι η οποίες μπορούν να χωριστούν σε τρεις βασικές κατηγορίες. Στην επιτηρούμενη μάθηση (Supervised Learning) στην οποία ο αλγόριθμος εκπαιδεύεται σε ένα σύνολο δεδομένων εισόδου που περιλαμβάνει κάποια χαρακτηριστικά και τα αντίστοιχα δεδομένα εξόδου (αποτελέσματα). Μια ακόμα κατηγορία της μηχανικής μάθησης είναι η μη επιτηρούμενη μάθηση (Unsupervised Learning) στην συγκεκριμένη κατηγορία ο αλγόριθμος δεν λαμβάνει κάποια χαρακτηριστικά εισόδου, στόχος του είναι να αναλύσει τα δεδομένα εισόδου και να εντοπίζει διάφορα μοτίβα και χαρακτηριστικά. Επίσης μια ακόμα κατηγορία είναι η ενισχυτική μάθηση (Reinforcement Learning) στην οποία ο αλγόριθμος εκπαιδεύεται να πραγματοποιεί αποφάσεις μέσω δοκιμασιών και σφαλμάτων λαμβάνοντας τις κατάλληλες ενέργειες για να μεγιστοποιήσει την σωρευτική ανταμοιβή (cumulative reward) (Y.Xin, 2018).

Τα νευρωνικά δίκτυα αποτελούν υπολογιστικά μοντέλα, σχεδιασμένα να επεξεργάζονται και να μαθαίνουν μέσα από μεγάλα σύνολα δεδομένων. Στον πυρήνα τους, τα νευρωνικά δίκτυα αποτελούνται από διασυνδεδεμένους κόμβους - γνωστοί ως “τεχνητοί νευρώνες” ή perceptrons (σε ελεύθερη μετάφραση: αντιλήπτες) - οργανωμένους σε επίπεδα. Τα επίπεδα αυτά διαθέτουν ένα επίπεδο εισόδου (input layer), τουλάχιστον ένα κρυφό επίπεδο και ένα επίπεδο εξόδου (output layer) και κάθε σύνδεση ανάμεσα σε νευρώνες σχετίζεται με ένα συντελεστή βαρύτητας, ενώ οι νευρώνες εφαρμόζουν στις εισόδους τους λειτουργίες ενεργοποίησης, ώστε να παράγουν ένα εξάγόμενο. Ο βασικός σκοπός των νευρωνικών δικτύων είναι να μαθαίνουν μοτίβα και σχέσεις μεταξύ των δεδομένων. Κατά τη διάρκεια της εκπαίδευσης του νευρωνικού δικτύου, αυτό προσαρμόζει τους συντελεστές βαρύτητάς του, ελαχιστοποιώντας τη διαφορά ανάμεσα στο εξάγόμενο που προβλέπεται και στις πραγματικές αξίες-στόχους, μια διαδικασία που ονομάζεται “οπισθοδιάδοση” (back propagation), επιτρέποντάς του να συγκρατεί σύνθετες, μη-γραμμικές σχέσεις δεδομένων, καθιστώντας το ικανό να εκτελεί καθήκοντα, όπως η αναγνώριση εικόνων και η κατάθεση προτάσεων (Nielsen, 2019).

Στα πλαίσια αυτά, διακρίνονται τα “Συνελικτικά Νευρωνικά Δίκτυα” (αγγλικά: Convolutional Neural Networks, CNNs), τα οποία θεωρούνται ιδανικά για την αναγνώριση εικόνων και τα “Επαναλαμβανόμενα Νευρωνικά Δίκτυα” (αγγλικά: Recurrent Neural Networks, RNNs), τα οποία διαπρέπουν στην επεξεργασία δεδομένων αλληλουχίας (sequential data processing) (Shiri et al., 2023).

Τα βαθιά νευρωνικά δίκτυα - γνωστά και ως “μοντέλα βαθιάς μάθησης” (deep learning models) - έχουν συγκεντρώσει τα τελευταία χρόνια την προσοχή της επιστημονικής και επιχειρηματικής κοινότητας του κλάδου, εξαιτίας της ικανότητάς τους να διαχειρίζονται σύνθετα

καθήκοντα. Συνεπώς, έχουν διαδραματίσει καθοριστικό ρόλο στην προώθηση τεχνολογιών σχετικών με την ΤΝ, ιδίως σε τομείς όπως τα αυτόνομα οχήματα, η φροντίδα υγείας και τα χρηματοπιστωτικά (Dignum, 2019).

Η επεξεργασία φυσικής γλώσσας NLP (Natural Language Processing) χρησιμοποιεί τεχνικές και μέθοδοι που την βοηθούν να αλληλοεπιδρά με την ανθρώπινη φυσική γλώσσα κάποιες από αυτές είναι η Συντακτική Ανάλυση (Syntactic Analysis) η οποία αναλύει την γραμματική ανάλυση για να κατανοήσει τη σχέση μεταξύ των λέξεων οι οποίες, ομαδοποιούνται σε φράσεις και οι φράσεις ομαδοποιούνται για να σχηματίσουν προτάσεις σε συντακτικό επίπεδο. Μια άλλη τεχνική είναι η σημασιολογική ανάλυση (Semantic Analysis), η οποία έχει ως καθήκον να προσδιοριστεί το σωστό νόημα μια πρότασης. (Khurana, 2023).

Στις μεθόδους που χρησιμοποιεί η υπολογιστική όραση (computer vision) περιλαμβάνεται η επεξεργασία εικόνων (Image Processing Systems) η οποία μπορεί να χρησιμοποιεί φίλτρα για βελτίωση της ποιότητας και ενίσχυση συγκεκριμένων χαρακτηριστικών. Επίσης μια άλλη μέθοδος που χρησιμοποιείται είναι η ανίχνευση αντικειμένων (Object Detection) σε μια εικόνα ή σε ένα περιβάλλον, παρέχοντας έτσι πληροφορίες σχετικά με την θέση τους και τον τύπο τους. Η ανάλυση κινήσεων στις εικόνες αποτελεί επίσης μια μέθοδο της υπολογιστικής όρασης με την χρήση της περισσότερο να συναντάται στα οχήματα, που περιλαμβάνουν αυτόνομη οδήγηση βοηθώντας τα να κατανοούν το περιβάλλον τους (Colyer, 2018).

2.1.4 Εφαρμογές Τεχνητής Νοημοσύνης

Η ΤΝ έχει βρει εφαρμογές σε ποικίλα πεδία, μεταμορφώνοντας τόσο τον κλάδο της βιομηχανικής παραγωγής όσο και την καθημερινή ζωή. Για παράδειγμα, στον τομέα της υγείας η ΤΝ αξιοποιείται για τη διάγνωση ασθενειών, την εφεύρεση φαρμάκων και το σχεδιασμό πλάνων προσωποποιημένης φροντίδας. Επιπροσθέτως, τα αυτόνομα οχήματα αξιοποιούν σε σημαντικό βαθμό τεχνικές ΤΝ, ιδίως για τη λήψη αποφάσεων οδικής συμπεριφοράς αλλά και για την πλοήγηση του οχήματος.

Ο χρηματοπιστωτικός κλάδος αξιοποιεί πλέον ευρέως την ΤΝ, προκειμένου να εντοπιστούν περιστατικά απάτης, να πραγματοποιηθούν αυτόματες συναλλαγές και να εκτιμηθούν διάφοροι κίνδυνοι, ενώ στην εκπαίδευση, αξιοποιούνται εργαλεία που ενισχύονται από την ΤΝ και συνδράμουν στην προσωποποιημένη μάθηση. Τέλος, οι υπηρεσίες εξυπηρέτησης πελατών πλέον χρησιμοποιούν τις γνωστές “μηχανές συζήτησεων” (chatbots) ή τους ψηφιακούς βοηθούς (digital assistants), προκειμένου να ενισχύσουν την εμπειρία των χρηστών και να επιταχύνουν την εξυπηρέτησή τους (Salas-Pilco & Yang, 2022).

Κλείνοντας ωστόσο την παρούσα ενότητα, η δυναμική εξέλιξη της ΤΝ έχει εγείρει ηθικές και κοινωνικές ανησυχίες. Για παράδειγμα, ζητήματα που σχετίζονται με το βαθμό μιας ενδεχόμενης μεροληψίας των αλγορίθμων προς ομάδες χρηστών/πολιτών, οι παραβιάσεις της ιδιωτικότητας και η αντικατάσταση θέσεων εργασίας εξαιτίας αυτοματισμών, έχουν αρχίσει να αναδύονται στις συζητήσεις. Η ερευνητική κοινότητα και τα κέντρα σχεδιασμού πολιτικών πρέπει να προσπαθήσουν να διασφαλίσουν την υπεύθυνη ανάπτυξη της ΤΝ και την προς όφελος του ανθρώπου, εφαρμογή της (El Hadi, 2023).

2.2 Κυβερνοασφάλεια

2.2.1 Ορισμός Κυβερνοασφάλειας

Ερχόμενος κάποιος στο ζήτημα της ασφαλούς περιήγησης και χρήσης του διαδικτύου, καθώς και των χιλιάδων εφαρμογών, η λειτουργία των οποίων εξαρτάται από αυτό, συναντάται η έννοια της “κυβερνοασφάλειας” (cyber security). Ο όρος υποδηλώνει την περιεκτική και έμπρακτη προστασία των ψηφιακών συστημάτων, των δικτύων και των δεδομένων από μη-εξουσιοδοτημένη πρόσβαση, βλάβες, υποκλοπές ή παρενόχληση. Σύμφωνα με την πολυεθνική εταιρεία Cisco Systems Inc. (2023a), η κυβερνοασφάλεια είναι “η προστασία συστημάτων, δικτύων και προγραμμάτων από ψηφιακές επιθέσεις, οι οποίες συνήθως στοχεύουν στην πρόσβαση, τροποποίηση ή καταστροφή ευαίσθητων πληροφοριών, την απόσπαση χρηματικών ποσών με τη χρήση κακόβουλων λογισμικών (ransomware) ή την παρενόχληση της λειτουργίας των επιχειρήσεων”. Το πεδίο της κυβερνοασφάλειας είναι διεπιστημονικός χώρος και εμπεριέχει ένα ευρύ φάσμα από τεχνολογίες, στρατηγικές και μεθόδους που σχεδιάστηκαν με σκοπό την τήρηση εμπιστευτικότητας, την ακεραιότητα και τη διαθεσιμότητα των ψηφιακών περιουσιακών στοιχείων. Η κυβερνο-ασφάλεια θεωρείται ουσιώδης στο σύγχρονο διασυνδεδεμένο κόσμο, στον οποίο οι πληροφορίες ρέουν διαρκώς και αναζητούνται διαρκώς τα τρωτά σημεία από κακόβουλες πηγές (Canongia & Mandarino, 2014).

2.2.2 Πεδία και τομείς Κυβερνοασφάλειας

Στον πυρήνα της, η κυβερνο-ασφάλεια περιστρέφεται γύρω από τη διατήρηση της τριάδας με το ακρωνύμιο “C-I-A”, δηλαδή της Εμπιστευτικότητας (Confidentiality), της Ακεραιότητας (Integrity) και της Διαθεσιμότητας (Availability) (Popescu, 2011):

- α. Η πρώτη εμπεριέχει τη διασφάλιση των ευαίσθητων δεδομένων, τα οποία πρέπει να προστατεύονται υπό το καθεστώς της ιδιωτικότητας και να είναι προσβάσιμα μόνο σε εξουσιοδοτημένους χρήστες. Αυτό μπορεί να επιτευχθεί μέσω της κρυπτογράφησης των δεδομένων κατά τη μεταφορά και αποθήκευσή τους και τον έλεγχο πρόσβασης, ενώ ύψιστης σημασίας κρίνεται και η προστασία των περιουσιακών στοιχείων ενός οργανισμού (δίκτυα, υπηρεσίες, υπολογιστές κ.ά.) από βλάβες που προκαλούν οι εν λόγω επιθέσεις.
- β. Η δεύτερη σχετίζεται με την εξασφάλιση της παραμονής της अपαράλλακτης μορφής των δεδομένων και της αξιοπιστίας τους. Η ακεραιότητα που εξασφαλίζεται μέσω μηχανισμών, όπως οι ψηφιακές υπογραφές (digital signatures), οι οποίοι θεωρούνται πρακτικές ικανές να συνδράμουν στον εντοπισμό προσπαθειών τροποποίησης των δεδομένων.
- γ. Τέλος, η διαθεσιμότητα δίνει έμφαση στο γεγονός ότι τα δεδομένα και τα συστήματα πρέπει να είναι προσβάσιμα όταν είναι απαραίτητα. Αυτό περιλαμβάνει την εφαρμογή σχεδίων και μηχανισμών backup & recovery, όπως δημιουργίας αντιγράφων (δεδομένων, αντιγράφων εικονικών μηχανών κ.ά.), ανθεκτικότητας ενάντια σε σφάλματα και ανάκτησης (δεδομένων, εικονικών μηχανών κ.ά.).

Η κυβερνο-ασφάλεια δεν αποτελεί ένα πανάρχαιο πεδίο για το οποίο ο άνθρωπος έχει αναπτύξει επαρκή γνώση ούτε μια απλή κατάσταση που επιδέχεται εύκολες λύσεις στα αναδυόμενα προβλήματά της. Αντιθέτως, πρόκειται για ένα σύνθετο οικοσύστημα, στο οποίο συγκαταλέγεται πλήθος εξειδικευμένων πεδίων. Για παράδειγμα, η ασφάλεια δικτύων (network security) αποτελεί μια περιοχή που επικεντρώνεται στην ασφάλεια των διαύλων επικοινωνίας και των υποδομών που τους φιλοξενούν. Ευρέως γνωστές τεχνικές, όπως τα τείχη προστασίας (firewalls), τα συστήματα εντοπισμού εισβολής (Intrusion Detection Systems, IDS) και τα ιδιωτικά δίκτυα (Virtual Private Networks, VPNs), χρησιμοποιούνται για την προστασία των δεδομένων (μεταφορά, αποθήκευση κ.ά.), των υπολογιστών και των συστημάτων. Ομοίως, η ασφάλεια των πληροφοριών (information security) επωμίζεται την προστασία των δεδομένων, με καθοριστικές παραμέτρους να αποτελούν η κρυπτογράφηση, ο έλεγχος πρόσβασης και η ταξινόμησή τους (Taherdoost, 2022).

Η ασφάλεια “τελικού σημείου” (endpoint) επικεντρώνεται στην προστασία μεμονωμένων συσκευών, όπως οι υπολογιστές, οι έξυπνες συσκευές και οι συσκευές IoT (Internet of Things) (Διαδικτύου των Πραγμάτων), σε έναν κόσμο όπου οι συσκευές έχουν πολλαπλασιαστεί και συνδέονται στο διαδίκτυο με σκοπό να διευκολύνουν την καθημερινότητα του ανθρώπου. Τα αντι-ιικά λογισμικά (antivirus), οι λύσεις εντοπισμού επιθέσεων σε τελικά σημεία (endpoint), όπως επίσης η διαχείριση φορητών συσκευών αποτελούν παραδοσιακά εργαλεία για το σκοπό αυτό (Seemba, Nandhini & Sowmiya, 2018). Μάλιστα, για τον όρο “ασφάλεια τελικού σημείου”

(endpoint security) η εταιρεία Cisco Systems Inc. (2023b) επισημαίνει ότι “συνδυάζει προληπτική προστασία μέσω ενός νέου είδους συνεχούς εντοπισμού και ανταπόκρισης, με τα τελικά σημεία να σχεδιάζονται έτσι ώστε να εντοπίζουν γρήγορα, να αναλύουν και να εμποδίσουν επιθέσεις εν εξελίξει”.

Η “διαχείριση ταυτότητας & πρόσβασης” (Identity & Access Management, IAM) περιλαμβάνει τη διαχείριση των ταυτοτήτων των χρηστών και τα δικαιώματα πρόσβασής τους. Εδώ συναντώνται τεχνικές, όπως η πολυπαραγοντική πιστοποίηση (multi-factor authentication) και η ενιαία αυθεντικοποίηση (single sign-on), οι οποίες ενισχύουν την ασφάλεια πρόσβασης και τις διαδικασίες ελέγχου. Ωστόσο, πέρα από τις πρακτικές διασφάλισης της προστασίας των δεδομένων και των εφαρμογών, σημαντικό ρόλο για την επίτευξη υψηλών επιπέδων κυβερνο-ασφάλειας, διαδραματίζει η εκπαίδευση του κοινού και η ευαισθητοποίηση επί αυτών των ζητημάτων. Στους τεχνολογικούς κύκλους, είναι γνωστό ότι ο ανθρώπινος παράγοντας συχνά αποτελεί τον πλέον αδύναμο κρίκο, εξαιτίας της μεγαλύτερης ευκολίας διενέργειας λαθών, της μικρότερης ταχύτητάς του στη διεξαγωγή υπολογισμών ή της πιο εύκολης παράκαμψής του σε θέματα ασφάλειας (με την τεχνική phishing, όπως θα παρουσιαστεί πιο αναλυτικά στη συνέχεια). Συνεπώς, τα προγράμματα εκπαίδευσης μεταδίδουν γνώση σε εργαζομένους (ή απλούς χρήστες) όσον αφορά τις βέλτιστες πρακτικές ασφάλειας στο διαδίκτυο, ελαχιστοποιώντας με αυτό τον τρόπο το ενδεχόμενο ανθρώπινου λάθους (Rabii et al., 2020). Τέλος, θα πρέπει να αναφερθούν τρεις ακόμη τομείς που συνδέονται με την κυβερνο-ασφάλεια και δεν είναι άλλοι από την ανταπόκριση σε συμβάντα (incident response), τη διαχείριση κινδύνων (risk management) και τη συμμόρφωση με κανονισμούς (regulatory compliance) (Σχήμα 3).

Η προετοιμασία και η ανταπόκριση σε περιστατικά διατάραξης της κυβερνο-ασφάλειας, είναι σημαντικά. Τα σχέδια ανταπόκρισης, η ψηφιακή εγκληματολογία (Digital Forensics) και οι τεχνολογίες εντοπισμού συμβάντων (Security Incident Detection), διαδραματίζουν καίριο ρόλο στην ελαχιστοποίηση των ζημιών. Από την άλλη, η διαχείριση των κινδύνων του διαδικτύου περιλαμβάνει την αναγνώριση, εκτίμηση και αντιμετώπισή τους, ενώ βοηθά τους οργανισμούς να θέτουν επενδυτικές προτεραιότητες ασφαλούς λειτουργίας και να τις ευθυγραμμίζουν με τους στόχους που οι ίδιοι θέτουν. Με τη σειρά της, η συμμόρφωση υποδηλώνει πως εφαρμόζονται κανόνες και πρότυπα, όπως το ISO 27001 (διεθνές πρότυπο ασφάλειας πληροφοριών) και το GDPR (Γενικός Κανονισμός για την Προστασία των Δεδομένων στην Ευρωπαϊκή Ένωση), η μη-εφαρμογή των οποίων μπορεί να οδηγήσει σε σημαντικές νομικές ή/και οικονομικές επιπτώσεις (European Union Agency for Network & Information Security, 2015).



Σχήμα 3. Πεδία που καλύπτει η κυβερνο-ασφάλεια

Πηγή: <https://dgrpg.punjab.gov.in/cybersecurity/>

2.2.3 Απειλές Κυβερνοασφάλειας

Αξίζει σε αυτό το σημείο μια αναφορά στην εκπαίδευση του κοινού και την ευαισθητοποίηση επί θεμάτων κυβερνοασφάλειας (security awareness), η οποία αποτελεί ουσιώδες ζήτημα στον σύγχρονο ψηφιακό κόσμο. Πρόκειται στην ουσία για τις γνώσεις, τις δεξιότητες και τη διαρκή επαγρύπνηση που οφείλει να επιδεικνύει το κοινό, στοιχεία απαραίτητα για την προστασία των προσωπικών και οργανωσιακών πληροφοριών, των δικτύων και των συστημάτων γενικότερα, από τις διαρκώς αναδυόμενες κυβερνο-επιθέσεις (Ministry of Digital Governance Greece, 2021). Οι παραβιάσεις δεδομένων, η χρήση λογισμικών ransomware και οι απάτες τύπου phishing αποτελούν συνήθη περιστατικά, με αποτέλεσμα την ανάγκη για κατανόηση των κινδύνων και λήψης μέτρων προστασίας, να θεωρούνται πρωταρχικής σημασίας. Η ευαισθητοποίηση έναντι της κυβερνοασφάλειας, ενισχύει τόσο τα άτομα όσο και τους οργανισμούς, στην αναγνώριση πιθανών κινδύνων, τη λήψη πληροφορημένων αποφάσεων και την υιοθέτηση ασφαλούς συμπεριφοράς στο διαδίκτυο. Ενθαρρύνεται έτσι η χρήση ισχυρών κωδικών πρόσβασης, ενημερώνονται σε συχνή βάση τόσο τα συστήματα όσο και το λογισμικό, ενώ εφαρμόζονται πολυπαραγοντική πιστοποίηση (multi-factor authentication). Κυρίως, όμως, προάγεται μια κουλτούρα ιδιωτικότητας και προστασίας, δίνοντας έμφαση στην ανάγκη εφίστασης της προσοχής όταν διαμοιράζονται

ευαίσθητες πληροφορίες και γίνεται κατανοητή η σημασία του να παραμένει κάποιος πληροφορημένος γύρω από τις πλέον πρόσφατες απειλές (Moore, n.d.).

2.2.4 Χρήση διαδικτύου και ζητήματα κυβερνο-ασφάλειας

Η κυβερνο-ασφάλεια, λοιπόν, διαδραματίζει σοβαρό ρόλο στο σχηματισμό του πλαισίου υπό το οποίο λαμβάνει χώρα η χρήση του διαδικτύου, επηρεάζοντας όλους όσους το αξιοποιούν. Στη σημερινή διασυνδεδεμένη εποχή, ο τρόπος αλληλεπίδρασης του ανθρώπου με το διαδίκτυο δε μπορεί να διαχωριστεί από τα μέτρα που παίρνονται για την ασφαλή πλοήγηση και την προστασία των δεδομένων του. Η κυβερνο-ασφάλεια θεωρείται ύψιστης σημασίας για τη διαφύλαξη των προσωπικών πληροφοριών στο διαδίκτυο, μιας και οι χρήστες εμπιστεύονται τα ευαίσθητα δεδομένα τους (χρηματοοικονομικές λεπτομέρειες, αριθμούς κοινωνικής ασφάλισης, αρχεία υγειονομικού χαρακτήρα κ.ά.) σε διάφορες πλατφόρμες, με τα μέτρα προστασίας να στοχεύουν στην αποσόβηση κακόβουλων ενεργειών και την απόκρυψη πληροφοριών από όσους επιδιώκουν να τα υποκλέψουν. Χωρίς στοιχειώδη έστω επίπεδα κυβερνο-ασφάλειας, οι χρήστες θα απέφευγαν να εμπλακούν σε διαδικτυακές δραστηριότητες, μειώνοντας σημαντικά την χρήση και καταστρέφοντας τη φήμη του διαδικτύου (Petratos, 2021).

Το διαδίκτυο έχει μεταμορφώσει τον τρόπο λειτουργίας των επιχειρήσεων, διευκολύνοντας το ηλεκτρονικό εμπόριο, την εξ' αποστάσεως εργασία και τις παγκόσμιες επικοινωνίες. Η αποδοτική κυβερνο-ασφάλεια θεωρείται ουσιώδης για τις επιχειρήσεις, προκειμένου να διατηρήσουν το επίπεδο εμπιστοσύνης που έχουν με τους καταναλωτές και τους συνεργάτες τους. Οι λεγόμενες “κυβερνο-επιθέσεις” (cyber attacks) μπορούν να οδηγήσουν σε διάρρηξη δεδομένων, οικονομικές απώλειες και ζημιές στη φήμη μιας εταιρείας, συνεπώς η εφαρμογή ισχυρών πρακτικών κυβερνο-ασφάλειας συνδράμουν στην ανενόχλητη ανάπτυξη του ηλεκτρονικού εμπορίου, διασφαλίζοντας ότι οι χρήστες μπορούν να συνεχίσουν να εμπιστεύονται τις διαδικτυακές συναλλαγές και αλληλεπιδράσεις (Wang & Wendy, 2019).

Σοβαρή πτυχή της κυβερνο-ασφάλειας, είναι και η προστασία ζητημάτων εθνικής ασφάλειας (national security) για κάθε κράτος. Είναι γεγονός ότι οι σημερινές κυβερνήσεις έχουν στηριχθεί σε μεγάλο βαθμό στην επικοινωνία μέσω του διαδικτύου, τη συγκέντρωση (αποθήκευση σε σκληρούς δίσκους ή cloud-based πλατφόρμες) απόρρητων πληροφοριών, καθώς επίσης τη διαχείριση κρίσιμων υποδομών (όπως οι στρατιωτικές εγκαταστάσεις και οι εγκαταστάσεις ενέργειας). Η κυβερνο-ασφάλεια, επομένως, θεωρείται ζωτικής σημασίας για την προστασία των στοιχείων εθνικής ασφάλειας (υποδομές ενέργειας, εγκαταστάσεις αμυντικής σημασίας κ.ά.), αφού

οι απειλές όπως οι κυβερνο-επιθέσεις σε κρατικές δομές ή η κυβερνο-τρομοκρατία (cyber terrorism) είναι πιθανό να οδηγήσουν σε σοβαρές επιπτώσεις - έστω και μικρής διάρκειας - όπως η αναποτελεσματική λειτουργία κρατικών οργανισμών ή η μη εξουσιοδοτημένη πρόσβαση σε πληροφορίες υψίστης εθνικής σημασίας. Υπό αυτό το πρίσμα, τα ισχυρά μέτρα κυβερνο-ασφάλειας θεωρούνται ουσιαστικές ενέργειες που μπορούν να διασφαλίσουν τα κυβερνητικά δίκτυα, τις κρίσιμες υποδομές και γενικότερα, οποιαδήποτε ευαίσθητη πληροφορία (Κονάcs, 2018).

Η ελευθερία της έκφρασης και η ιδιωτικότητα, αποτελούν κεκτημένο του ανθρώπου στις σύγχρονες δυτικές κοινωνίες, το οποίο κέρδισε έπειτα από σημαντικούς αγώνες. Η κυβερνο-ασφάλεια, επομένως, δε θα μπορούσε παρά να έρθει σε συνύπαρξη με τις αρχές αυτού του δικαιώματος, αυτή τη φορά στο διαδικτυακό περιβάλλον. Η ισορροπία των ληφθέντων μέτρων διαδικτυακής ασφάλειας, σε σχέση με τα αντίστοιχα μέτρα της προστασίας των δικαιωμάτων των ανθρώπων, αποτελεί μια σοβαρή πρόκληση, αφού τα ακραία μέτρα κυβερνο-ασφάλειας μπορούν να καταπατήσουν την ιδιωτικότητα, οδηγώντας σε προβληματισμούς αναφορικά με φαινόμενα παρακολούθησης και περιορισμών. Για παράδειγμα, με την παρακολούθηση και καταγραφή ενεργειών, γίνονται γνωστές οι προτιμήσεις των χρηστών και για να αντιμετωπίσουν τέτοια περιστατικά οι χρήστες θα πρέπει να χρησιμοποιούν εφαρμογές, όπως τα λογισμικά παρεμπόδισης διαφημίσεων (ad-blockers), να μην αποδέχονται cookies και κυρίως, να εξετάζουν τις πολιτικές απορρήτου που έχει υιοθετήσει η κάθε πλατφόρμα/ιστοσελίδα. Η επίτευξη, λοιπόν, μιας ισορροπίας, μπορεί να διασφαλίσει ότι η χρήση του διαδικτύου παραμένει ανοιχτή και επιδεικνύεται σεβασμός στα δικαιώματα και την ελευθερία των χρηστών (Jung, 2021).

Μέρος του πεδίου της κυβερνο-ασφάλειας αποτελούν, επίσης, η τεχνολογική πρόοδος και οι καινοτομίες. Το διαδίκτυο έχει αποδειχθεί καταλυτικός παράγοντας καινοτομιών και τεχνολογικής προόδου, βοηθώντας στην ανάπτυξη νέων υπηρεσιών και βιομηχανιών. Ωστόσο, οι απειλές σε επίπεδο κυβερνο-ασφάλειας (παραβίαση δεδομένων, πρόκληση ζημιάς στη φήμη, δυσκολίες συμμόρφωσης προς κανονισμούς κ.ά.) μπορούν να καταπνίξουν την καινοτομία, σε περίπτωση που οι επιχειρήσεις αναγκαστούν σταδιακά να μετακινήσουν πόρους για να «αμυνθούν» και να προστατέψουν τα περιουσιακά τους στοιχεία. Τα αποτελεσματικά, όμως, μέτρα κυβερνο-ασφάλειας προάγουν την καινοτομία, δημιουργώντας ένα ασφαλές περιβάλλον για την ανάπτυξη και εφαρμογή νέων τεχνολογιών. Επομένως, απαιτείται ισορροπημένη διαχείριση ανάμεσα στη μετακίνηση πόρων για ανάπτυξη καινοτομιών και την εγκατάσταση μέτρων προστασίας (Antunes et al., 2021).

2.2.5 Αντίμετρα και Μέθοδοι Προστασίας Κυβερνοασφάλειας

Το ζήτημα της ασφάλειας όμως επεκτείνεται και στις εφαρμογές, με την προστασία του λογισμικού να θεωρείται ζωτικής σημασίας για την αποτροπή δημιουργίας, εντοπισμού και εκμετάλλευσης τρωτών σημείων από τους επιτιθέμενους (hackers). Η αναθεώρηση κώδικα (code review), οι έλεγχοι μη-εξουσιοδοτημένης εισβολής (unauthorized intrusion controls) και η χρήση πρακτικών ασφαλούς ανάπτυξης λογισμικού (Software development security) είναι μόνο μερικές από τις στρατηγικές που υιοθετούνται. Μια από τις τεχνολογίες μάλιστα που έχει βοηθήσει σε πολύ μεγάλο βαθμό τον άνθρωπο να αποθηκεύει πληροφορίες ή να αξιοποιεί εφαρμογές χωρίς να τις εγκαθιστά, είναι η νεφο-υπολογιστική (cloud computing), χάριν στην οποία οι οργανισμοί μπορούν να μεταφέρονται για την εκτέλεση των καθηκόντων τους σε περιβάλλοντα πέραν αυτών στα οποία εργάζονται με φυσική παρουσία. Συνεπώς, η διασφάλιση της ασφαλούς χρήσης ενός τέτοιου περιβάλλοντος (cloud-based) κρίνεται καθοριστικής σημασίας. Αυτού του είδους η ασφάλεια διαχειρίζεται θέματα (network security, firewalls, IDS, information security), όπως η ιδιωτικότητα των δεδομένων, η προστασία περιουσιακών στοιχείων, η διασφάλιση εξοπλισμού, η συμμόρφωση των χρηστών και ο έλεγχος πρόσβασης (Taherdoost, Sahibuddin & Jalaliyoon, 2015).

Αξίζει, τέλος, να τονιστεί και το γεγονός ότι η κυβερνο-ασφάλεια επεκτείνεται έως τα ζητήματα διεθνούς συνεργασίας και εμπιστοσύνης. Σε ένα διασυνδεδεμένο κόσμο, οι διεθνείς συνεργασίες και η εμπιστοσύνη αποτελούν ουσιώδη ζητήματα διαχείρισης των προκλήσεων που άπτονται της κυβερνο-ασφάλειας. Οι διαδικτυακές απειλές ξεπερνούν τα εθνικά σύνορα, καθιστώντας απαραίτητη τη συνεργασία ανάμεσα σε κυβερνήσεις (δημιουργώντας διεθνή κέντρα αντιμετώπισης, διεξάγοντας εκστρατείες ενημέρωσης, εκπαιδύοντας προσωπικό κ.ά.), επιχειρήσεις και ειδικούς στην κυβερνοασφάλεια. Το χτίσιμο εμπιστοσύνης στο διαδικτυακό περιβάλλον, είναι καίριας σημασίας για τη δημιουργία συνεργασιών και τη διασφάλιση ότι το διαδίκτυο παραμένει αξιόπιστο και μια ασφαλή επιλογή για διεθνείς επικοινωνίες, συναλλαγές, εμπορικές συμπράξεις και ανταλλαγή πληροφοριών (Κονάcs, 2018).

3 Μεθοδολογία βιβλιογραφικής ανασκόπησης

3.1 Κριτήρια επιλογής άρθρων

Η παρούσα εργασία αποτελεί μια βιβλιογραφική ανασκόπηση σχετικά με την TN και το βαθμό στον οποίο μπορεί να βελτιώσει ή να δυσχεράνει τα επίπεδα κυβερνο-ασφάλειας, προστατεύοντας προσωπικά δεδομένα των χρηστών και ευαίσθητες πληροφορίες οργανισμών καθώς και τα συστήματά τους. Τα ερωτήματα που καλείται να απαντήσει η παρούσα εργασία, καταγράφονται ως εξής:

- ποιες είναι οι συνήθεις πρακτικές κυβερνο-επιθέσεων με χρήση TN;
- ποια είναι τα συνηθέστερα αντίμετρα στις κυβερνο-επιθέσεις που λαμβάνουν χώρα με χρήση TN;
- υπάρχει εν τέλει βελτίωση ή επιδείνωση της κυβερνο-ασφάλειας λόγω της χρήσης της TN από ειδικούς ασφάλειας και επιτιθέμενους, αντίστοιχα;

Για τη διεκπεραίωση της εργασίας, αναζητήθηκαν άρθρα κυρίως στις πηγές δεδομένων Google Scholar και Web of Science, δηλώνοντας συγκεκριμένα λήμματα και φράσεις σχετικά με το αντικείμενο της διπλωματικής, ενώ εφαρμόστηκαν και ορισμένα χρήσιμα κριτήρια στην επιλογή άρθρων, προκειμένου να διασφαλιστεί τόσο η εγκυρότητα και ποιότητά τους όσο και το γεγονός ότι αποτελούν σύγχρονες πηγές.

Συγκεκριμένα, τα λήμματα και οι φράσεις που χρησιμοποιήθηκαν, προκειμένου η αναζήτηση να επιστρέψει κατάλληλες πληροφορίες, ήταν:

- (Artificial Intelligence OR AI-based) AND (cyberattacks OR risks OR attacks)
- (Artificial Intelligence OR AI-based) AND cybersecurity AND (countermeasures OR protection methods)

Τα κριτήρια (αποδοχής και ποιότητας) σύμφωνα με τα οποία επιλέχθηκαν οι πληροφορίες της παρούσας εργασίας, ήταν τα ακόλουθα:

- γλώσσα: μόνο άρθρα γραμμένα στην Αγγλική ή στην Ελληνική γλώσσα
- χρονολογία: να είναι δημοσιευμένα μετά το 2013

- εγκυρότητα: να προέρχονται από έγκριτα επιστημονικά περιοδικά και συνέδρια και να εντοπίζονται σε γνωστές βάσεις δεδομένων
- θεματικό ταίριασμα: να είναι συναφή με το υπό μελέτη αντικείμενο
- να σημειώνεται συμμετοχή διεθνούς φορέα/οργανισμού (διεθνείς και εθνικοί οργανισμοί, πολυεθνικές εταιρείες, ακαδημαϊκά ιδρύματα) στη συγγραφή τους
- να χρησιμοποιείται κατάλληλη μεθοδολογία και επιστημονική προσέγγιση σε κάθε άρθρο, ώστε να εξασφαλιστεί η αντικειμενικότητα της έρευνας

Από την άλλη, αποκλείστηκαν:

- πληροφορίες που προέρχονταν από ιστοσελίδες τύπου infotainment (Information & Entertainment)
- άρθρα των οποίων η προέλευση δεν ήταν ξεκάθαρη

4 Επιθέσεις μέσω Τεχνητής Νοημοσύνης

4.1 Στάδια επίθεσης

Οι στηριζόμενες στην TN, κυβερνο-επιθέσεις, συνθέτουν μια διαρκώς αυξανόμενη ανησυχία στο πεδίο της κυβερνο-ασφάλειας, μιας και η ίδια η TN - σε συνδυασμό ιδίως με την τεχνολογία της μηχανικής μάθησης - μπορούν να αξιοποιηθούν από κακόβουλες πηγές προκειμένου να ενισχύσουν την αποτελεσματικότητα των ενεργειών τους. Οι επιθέσεις αυτές περιλαμβάνουν διάφορα στάδια, παρόμοια με τις παραδοσιακές κυβερνο-επιθέσεις (παράνομη είσοδο, προσπάθεια διάρρηξης αμυνών, «μόλυνση» αρχείων κ.ά.), με την TN όμως να διαδραματίζει καθοριστικό ρόλο στην ανάληψη μιας σειράς πρωτοβουλιών (Guembe et al., 2021).

Αξίζει σε αυτό το σημείο μια σύντομη αναφορά των διαφορών που παρουσιάζουν τα στάδια των παραδοσιακών και των βασισμένων στην TN κυβερνο-επιθέσεων. Ξεκινώντας από την αναγνώριση, στις παραδοσιακές κυβερνο-επιθέσεις οι επιτιθέμενοι συγκεντρώνουν πληροφορίες εξερευνώντας οι ίδιοι τα δίκτυα, τα συστήματα και αναζητώντας τρωτά σημεία, ενώ στις δεύτερες η TN κάνει ακριβώς το ίδιο, αναλύοντας όμως αυτόματα και με ταχύτητα, μεγάλους όγκους δεδομένων. Επιπλέον, στις παραδοσιακές κυβερνο-επιθέσεις κατασκευάζονται κακόβουλα λογισμικά χρησιμοποιώντας προκαθορισμένους κώδικες, ενώ με την TN δημιουργούνται νέα και εξειδικευμένα κακόβουλα λογισμικά, τα οποία προσαρμόζουν τις τεχνικές επίθεσης αναλόγως των αμυνών που διαθέτουν οι στόχοι και τροποποιούν τον κώδικά τους για να αποφύγουν την αναγνώριση (Yamin et al., 2021). Οι τρόποι παράδοσης δε διαφέρουν, ωστόσο με χρήση TN καθίστανται πιο πειστικές οι μέθοδοι (εξαπάτησης) και αξιοποιείται γνώση σχετική με το μοτίβο συμπεριφοράς του στόχου. Η διαφορά στη φάση της έκθεσης (exploitation) συναντάται στο γεγονός ότι στις παραδοσιακές κυβερνο-επιθέσεις αξιοποιούνται γνωστά τρωτά σημεία, ενώ με χρήση TN εντοπίζονται zero-day αδυναμίες και αποφεύγονται λάθη παλαιότερων προσπαθειών επίθεσης (Wang et al., 2023).

Η αποφυγή στις παραδοσιακές κυβερνο-επιθέσεις γίνεται κατά κύριο λόγο με αποκρυπτογράφηση ή συσκοτίση (obfuscation) (είναι η μετατροπή στοιχείων μιας κυβερνο-επίθεσης σε δύσκολα αναγνωρίσιμα, μεταφράσιμα ή κατανοητά, από τις άμυνες του στόχου), ενώ με χρήση της TN το κακόβουλο λογισμικό μπορεί να «διαβάσει» τις αμυντικές πρακτικές (του στόχου) και να βελτιώσει τις στρατηγικές αποφυγής του. Ιδιαιτερότητα παρουσιάζει η επιμονή (persistence) που μπορούν να επιδείξουν τα δύο είδη κυβερνο-επιθέσεων, με τις παραδοσιακές να

στηρίζονται στη δημιουργία backdoors (δηλαδή σημείων επαν-εισόδου που δεν εντοπίζονται εύκολα) ή την εγκατάσταση ενός κακόβουλου κώδικα και την TN να αυτοματοποιεί τη διαδικασία εντοπίζοντας νέα τρωτά σημεία ή/και προσαρμοζόμενη στις εξελισσόμενες πρακτικές ασφαλείας (Valdes et al., 2023).

Η απομάκρυνση χρήσιμων δεδομένων επίσης ενέχει σημαντικές διαφορές, αφού η παραδοσιακή μέθοδος γίνεται με πιο αργούς ρυθμούς και χωρίς διάκριση των πληροφοριών, ενώ με χρήση της TN μπορούν να επιλεγούν αυτόματα, τα πλέον ευαίσθητα δεδομένα, αξιοποιώντας μεθόδους μηχανικής μάθησης. Τέλος, στο στάδιο μετά την επίθεση, οι όποιες ενέργειες (πώληση εξαχθέντων πληροφοριών, ανάλυση κ.ά.) γίνονται από τους επιτιθέμενους χωρίς να συνοδεύονται απαραίτητα από την αξιολόγηση της πορείας της επίθεσης, ενώ με χρήση της TN κάποιες ενέργειες δε μπορούν να γίνουν αν δεν έχει προηγηθεί εντολή από διαχειριστή (για παράδειγμα η πώληση εξαχθέντων) αλλά μπορούν να γίνουν αρκετές ενέργειες αυτόματα (Zhangyi & Gang, 2022).

Θεωρείται, όμως, σημαντικό να καταστεί σαφές πως οι βασισμένες στην TN, κυβερνο-επιθέσεις, αποτελούν ένα διαρκώς εξελισσόμενο πεδίο. Τα στάδια των βασισμένων στην TN, κυβερνο-επιθέσεων, είναι τα εξής (Finnish Transport & Communication Agency, 2022):

- συλλογή δεδομένων & ενέργειες αναγνώρισης (data collection & reconnaissance): Αναφέρεται στην συλλογή και συγκέντρωση πληροφοριών για εντοπισμό ευπαθειών. Για παράδειγμα, μπορεί να γίνει επεξεργασία λογαριασμών κοινωνικών δικτύων για συγκέντρωση πιο στοχευμένων πληροφοριών σχετικών με άτομα ή οργανισμούς.
- αρχική πρόσβαση (initial access): Αναφέρεται στο στάδιο της επίθεσης στο σύστημα όταν επιτευχθεί η πρόσβαση. Για παράδειγμα, μπορεί να πραγματοποιηθεί ανάλυση συμπεριφοράς στόχου, προτιμήσεων και μοτίβου επικοινωνίας, με σκοπό την κατασκευή στοχευμένων και εξαιρετικά πειστικών μηνυμάτων phishing, για την απόσπαση πληροφοριών εισόδου, όπως ονόματος χρήστη και κωδικού).
- επίμονη (persistence): Αναφέρεται στην δυνατότητα του επιτιθέμενου να παραμένει ενεργός για μεγάλο χρονικό διάστημα στο σύστημα του θύματος χωρίς να ανιχνεύεται. Αυτό θα μπορούσε να επιτευχθεί, για παράδειγμα, μέσα από τη διαρκή ανάλυση δεδομένων, τον εντοπισμό μοτίβων του στόχου και τη διαρκή ανανέωση της λίστας των τρωτών σημείων προς αξιοποίηση.
- αύξηση δικαιωμάτων εισόδου (privilege escalation): Αναφέρεται στη διαδικασία όπου ο επιτιθέμενος αποκτά περισσότερα δικαιώματα σε ένα σύστημα, εφόσον έχει εισβάλει σε αυτό. Για παράδειγμα, ο εντοπισμός ενός αδύναμου κωδικού εισόδου και ο διαμοιρασμός πολλών προσωπικών δεδομένων μέσω κοινωνικών δικτύων, μπορεί να διευκολύνει το

επιτιθέμενο λογισμικό TN, να κατανοήσει τις συνήθειες του χρήστη-στόχου (πχ. ενός διαχειριστή) με αυξημένα δικαιώματα και να αναζητήσει αδυναμίες στον τρόπο εισόδου στο σύστημά του, αποκτώντας έτσι δικαίωμα πρόσβασης.

- πλευρικές κινήσεις (lateral movements): Αναφέρεται στις κινήσεις που πραγματοποιεί ο επιτιθέμενος μέσα στο πληροφοριακό σύστημα, αναζητώντας ευπαθή σημεία για να επεκτείνει την πρόσβαση του. Για παράδειγμα, η TN μπορεί να μιμηθεί τη συμπεριφορά ενός χρήστη και χρησιμοποιώντας μηχανική μάθηση να εντοπίσει αδύναμα σημεία στην άμυνά του, εισχωρώντας στο σύστημα ενός οργανισμού, ενώ στη συνέχεια μπορεί να αυτοματοποιήσει τη διαδικασία και να εξάγει πληροφορίες για το σύνολο του οργανισμού.
- διήθηση σε δεδομένα (data exfiltration): Αναφέρεται στην αφαίρεση, μεταφορά ή αποστολή ευαίσθητων δεδομένων από ένα πληροφοριακό σύστημα ή δίκτυο από τον επιτιθέμενο. Η μεταφορά μπορεί να γίνει εμπεριστατωμένα ώστε να μην γίνει αντιληπτή, κρυπτογραφώντας την αντίστοιχη πληροφορία και προσπαθώντας να ακολουθούνται καθιερωμένα μοτίβα εξερχόμενης κίνησης στο δίκτυο.
- αποφυγή εντοπισμού ή/και απόκρυψη (evasion and/or stealth): Αναφέρεται στην αποφυγή εντοπισμού του επιτιθέμενου στο σύστημα από τα συστήματα ασφαλείας του θύματος. Για παράδειγμα, μπορεί να γίνει ανάπτυξη αντιπαραθετικής τακτικής, με είσοδο πληθώρας μη-κακόβουλων δεδομένων, ή δημιουργία ενός ψηφιακού παραπετάσματος (digital smokescreen) που εμποδίζει τον εντοπισμό της διείσδυσης στο σύστημα και οδηγεί σε σταδιακή παράβλεψη των επιπέδων ασφαλείας για έκθεση των τρωτών σημείων.
- ανάλυση δεδομένων & λήψη αποφάσεων (data analysis & decision making): Αναφέρεται στην ανάλυση των δεδομένων για εξαγωγή συμπερασμάτων ενώ η "λήψη αποφάσεων" αναφέρεται στη διαδικασία χρήσης αυτών των πληροφοριών λαμβάνοντας τις κατάλληλες αποφάσεις.
- διήθηση & ενέργειες "ελέγχου-εντολών" (exfiltration & command-control): Αναφέρεται στη δομή ή το σύστημα που χρησιμοποιεί ο επιτιθέμενος για να διαχειριστεί και να ελέγχει τα συστήματα των επιθέσεών του, εκδίδοντας κατάλληλες εντολές. Για παράδειγμα, μπορεί να εισάγονται στο σύστημα που διαχειρίζεται την ηλεκτρική υποδομή μιας πόλης, πληθώρα από λογισμικά ελεγχόμενα από TN και μεταμφιεσμένα ως αναβαθμίσεις - updates - ή προσθήκες - patches – τα οποία χαρτογραφούν το δίκτυο της πόλης, εντοπίζοντας τρωτά σημεία στην άμυνα του συστήματος και εμποδίζοντας την πρόσβαση στους εργαζομένους, ενώ στη συνέχεια οι επιτιθέμενοι ζητούν μεγάλα χρηματικά για να μην προκαλέσουν βλάβες στο δίκτυο.

- κάλυψη ίχνους (covering tracks): Αναφέρεται στη διαδικασία που ακολουθεί ο επιτιθέμενος, για να μειώσει τα αποτυπώματα της δραστηριότητάς του στο πληροφοριακό σύστημα. Για παράδειγμα, αφού ένα κακόβουλο λογισμικό έχει εισέλθει σε ένα σύστημα, καλύπτει τα ψηφιακά του ίχνη, διαγράφοντας και τροποποιώντας χρονικά δεδομένα, και δημιουργεί ψευδή πληροφόρηση δίνοντας την εντύπωση φυσιολογικής δραστηριότητας στο σύστημα.
- αντίκτυπος & αξιοποίηση (impact & exploitation): Αναφέρεται στις επιπτώσεις ή τις συνέπειες που έχει μια επίθεση σε ένα πληροφοριακό σύστημα, συμπεριλαμβανομένων των ζημιών. Για παράδειγμα - σε ένα ακραία φανταστικό σενάριο - με τη βοήθεια κακόβουλων λογισμικών και αλγορίθμων βαθιάς μάθησης (deep-learning), ξεκινά μια κυβερνο-επίθεση σε ένα σύστημα υγείας, κατά την οποία τα ασθενοφόρα αποστέλλονται σε λάθος τοποθεσίες, με αποτέλεσμα να δημιουργείται χαώδης κατάσταση στην εξυπηρέτηση ασθενών και πανικός στους πολίτες, οι οποίοι αδυνατούν να εντοπίσουν την αιτία του προβλήματος.
- στρατηγική εξόδου (exit strategy): Αναφέρεται στο στη διαδικασία που ακολουθεί ο επιτιθέμενος για να αποχωρήσει από το σύστημα του θύματος χωρίς να αφήνει ψηφιακά ίχνη. Για παράδειγμα, με τη χρήση ενός εξελεγμένου αλγορίθμου κρυπτογράφησης, το κακόβουλο λογισμικό καλύπτει τα ψηφιακά του αποτυπώματα, δημιουργεί σύγχυση στο σύστημα που εισήλθε ώστε να μην εντοπιστεί και αποχωρεί από το σύστημα αφού πρώτα έχει αποσπάσει τις απαραίτητες πληροφορίες.

4.2 Τεχνικές που χρησιμοποιούνται

Περνώντας τώρα σε ένα περισσότερο «πρακτικό» κομμάτι της ΤΝ και των επιθέσεων τις οποίες μπορεί να εξαπολύσει στον κυβερνοχώρο, θα πρέπει να γίνει λόγος για τις τεχνικές που αξιοποιούν τα κακόβουλα λογισμικά προκειμένου να πετύχουν το στόχο τους. Οι βασισμένες στην ΤΝ κυβερνο-επιθέσεις, χρησιμοποιούν τις δυνατότητές της - σε συνδυασμό με την τεχνολογία μηχανικής μάθησης - ως μοχλό ενίσχυσης της αποτελεσματικότητάς τους και εξειδίκευσης των διαδικτυακών απειλών που αποστέλλουν (Petratos, 2021). Συνοπτικά, ορισμένες από τις πλέον γνωστές τεχνικές κυβερνο-επιθέσεων που βασίζονται στην ΤΝ, οι οποίες μπορούν να χρησιμοποιούνται σε ένα από τα προαναφερόμενα στάδια ή σε πολλά (οπότε μπορεί να καλύψουν και μια ολόκληρη κυβερνο-επίθεση), είναι οι εξής (Ardito et al., 2021):

- αξιοποίηση ΤΝ για την αποφυγή εντοπισμού (AI for evasion)
- διήθηση σε δεδομένα (data exfiltration)
- απόσπαση κωδικών & δεδομένων εισόδου (password attacks)

- δημιουργία κακόβουλου λογισμικού (malware generation)
- τεχνική της «αξιοποίησης άγνωστων τρωτών σημείων» (zero-day exploits)
- αντίστροφη μηχανική μάθηση (adversarial machine learning)
- απάτη μέσω της τεχνικής του «ηλεκτρονικού ψαρέματος» (phishing)
- τεχνική επεξεργασίας φυσικής γλώσσας (NLP)
- αυτοματοποιημένες επιθέσεις (automated attacks)
- τεχνική της «υποστηριζόμενης από ΤΝ επίθεσης κυβερνο-ρομπότ» (AI-powered bots)
- τεχνική αξιοποίησης υπερρεαλιστικών μέσων (deepfakes)

4.2.1 Αποφυγή εντοπισμού (evasion)

Η πιο κοινή ίσως πρακτική αξιοποίησης της ΤΝ όσον αφορά την αποφυγή, σχετίζεται με τα πολυμορφικά κακόβουλα λογισμικά (polymorphic malware). Οι αλγόριθμοι ΤΝ μπορούν να δημιουργήσουν αυτόματα κώδικες και στη συνέχεια να τους τροποποιήσουν, με αποτέλεσμα να είναι εφικτό να προκύπτουν νέα, μοναδικά κακόβουλα λογισμικά σε κάθε καινούρια επίθεση. Αυτή η διαρκής μετάλλαξη καθιστά δύσκολο για τα αντικά λογισμικά να εντοπίσουν και να παρεμποδίσουν τέτοιες, συνεχώς εξελισσόμενες απειλές. Επιπροσθέτως, η υποβοηθούμενη από την ΤΝ διαφυγή, είναι ικανή να χειραγωγήσει τα μοτίβα κυκλοφορίας στο διαδίκτυο (network traffic patterns), κάνοντας δύσκολη την αναγνώριση της ύποπτης δραστηριότητας, για τα συστήματα εντοπισμού εισβολών (IDS). Γενικά, οι επιτιθέμενοι χρησιμοποιούν τη μηχανική μάθηση προκειμένου να αναλύουν τη φυσιολογική κίνηση δικτύου στους στόχους που επιτίθενται και να προσαρμόζουν τις ενέργειές τους έτσι ώστε να μιμούνται τη συμπεριφορά των χρηστών που διαθέτουν δικαίωμα πρόσβασης σε ένα σύστημα, αποφεύγοντας με αυτό τον τρόπο τον εντοπισμό (de Witt et al., 2021).

Ένα ακόμα κομβικό σημείο της αποφυγής εντοπισμού με χρήση ΤΝ, είναι η χρήση “αντιπαραθετικών επιθέσεων” (adversarial attacks). Οι επιτιθέμενοι χρησιμοποιούν τεχνικές μηχανικής μάθησης, δημιουργώντας ψεύτικες επιθέσεις, οι οποίες προκαλούν σύγχυση και οδηγούν τα λογισμικά ασφαλείας που υποστηρίζονται από την ΤΝ, σε λανθασμένες αποφάσεις. Αυτές οι ειδικά σχεδιασμένες κατασκευές μπορούν να προκαλέσουν λανθασμένες αντιδράσεις από τα συστήματα ασφαλείας, επιτρέποντας στους επιτιθέμενους να παρακάμψουν τους μηχανισμούς προστασίας. Για παράδειγμα, μια κατασκευασμένη από λογισμικό ΤΝ εικόνα, μπορεί να περιλαμβάνει μικρές τροποποιήσεις, τις οποίες ένα σύστημα αναγνώρισης βασισμένο σε άλλο

λογισμικό TN μπορεί να «παρερμηνεύσει» (δηλαδή να την εκλάβει ως ορθή), επιτρέποντας έτσι σε κακόβουλο περιεχόμενο να διεισδύσει σε ένα σύστημα (Yamin et al., 2021).

Επιπροσθέτως, η TN μπορεί να αξιοποιηθεί για να αυτοματοποιήσει και να βελτιώσει επιθέσεις κοινωνικής μηχανικής (social engineering attacks). Αναλύοντας σύνολα δεδομένων μεγάλου μεγέθους - δηλαδή από σημαντικό πλήθος πιθανών θυμάτων - η TN μπορεί να βοηθήσει τους επιτιθέμενους να δημιουργήσουν εξαιρετικά πειστικά phishing emails ή να εφαρμόσουν άλλες τακτικές κοινωνικής μηχανικής. Φτάνοντας σε πολύ υψηλό επίπεδο προσωποποίησης επί των κοινωνικών επαφών των θυμάτων, αυξάνεται η πιθανότητα επιτυχίας των επιθέσεων, μιας και τα θύματα είναι πιο πιθανό να εμπιστευτούν το μήνυμα και να ανταποκριθούν στο περιεχόμενο και τις απαιτήσεις του. Η TN μπορεί, επίσης, να αξιοποιηθεί με σκοπό τη μίμηση του τρόπου γραφής και της συμπεριφοράς αξιόπιστων επαφών των θυμάτων, καθιστώντας την προσωποποίηση ακόμα πιο πειστική. Έτσι, η TN δεν επιτυγχάνει μόνο την αυτοματοποίηση των επιθέσεων και των προσπαθειών αποφυγής εντοπισμού του επιτιθέμενου αλλά ενισχύει παράλληλα την αποτελεσματικότητα των ανθρωποκεντρικών κυβερνο-επιθέσεων που διεξάγονται (Guembe et al., 2022).

4.2.2 Διήθηση σε δεδομένα (data exfiltration)

Πρόκειται για κυβερνο-επιθέσεις που αποτελούν εξειδικευμένη πρακτική. Χρησιμοποιούν την TN μαζί με μηχανική μάθηση, προκειμένου να αποσπάσουν κρυφά, ευαίσθητα δεδομένα από τα συστήματα-στόχους τους. Συνήθως, μια τέτοια επίθεση εκτυλίσσεται σε τέσσερις φάσεις. Η πρώτη φάση ονομάζεται «αρχική αναγνώριση & διείσδυση» (initial reconnaissance & infiltration), κατά την οποία η επίθεση ξεκινά με την αναγνώριση των δυνητικών στόχων των κυβερνοεγκληματιών και των τρωτών σημείων του συστήματος του οργανισμού. Τα βασιζόμενα σε TN εργαλεία που χρησιμοποιούνται σε αυτή την περίπτωση, αναζητούν αδυναμίες, όπως μη-επιδιορθωμένο/μη-αναβαθμισμένο λογισμικό ή εσφαλμένες ρυθμίσεις ασφάλειας. Από τη στιγμή που εντοπίζεται το τρωτό σημείο εισόδου, οι επιτιθέμενοι χρησιμοποιούν το κακόβουλο λογισμικό τους και αποκτούν μη-εξουσιοδοτημένη πρόσβαση στο σύστημα.

Στην επόμενη φάση (δεύτερη φάση: AI-powered data identification – ενισχυόμενος από τη TN εντοπισμός δεδομένων), ένας αλγόριθμος TN (που έχει «εκπαιδευτεί» και έχει προκύψει αντίστοιχο μοντέλο με βάση το οποίο γίνεται η εξερεύνηση για τον εντοπισμό της πολύτιμης πληροφορίας) εφαρμόζεται με σκοπό την εξερεύνηση και ανάγνωση μεγάλου όγκου δεδομένων, εντός του παραβιασμένου συστήματος. Η αναγνώριση πολύτιμων πληροφοριών πραγματοποιείται αυτόματα από τον αλγόριθμο TN και στοχεύει στον εντοπισμό εμπιστευτικών εγγράφων,

οικονομικών δεδομένων ή ευαίσθητων δεδομένων πελατών, ακόμη και αν βρίσκονται «κρυμμένα» μεταξύ λιγότερο σημαντικών δεδομένων. Τα μοντέλα μηχανικής μάθησης μπορούν να προσαρμόσουν και να βελτιώσουν σε βάθος χρόνου την ακρίβειά τους, καθιστώντας πιο δύσκολο για τα μέτρα ασφαλείας να εντοπίσουν μοτίβα ασυνήθιστης πρόσβασης. Στην συγκεκριμένη περίπτωση, ο αλγόριθμος μπορεί να βελτιώσει την ακρίβεια του από τις ανατροφοδοτήσεις των χρηστών που βρίσκονται ενεργοί στο σύστημα, μέσω της μηχανικής μάθησης που περιέχεται στον κώδικά του. Αυτό έχει ως αποτέλεσμα να παρέχονται ακριβείς πληροφορίες για ένα προϊόν ή σύστημα.

Η ανάκτηση δεδομένων με πρακτικές βασισμένες στην TN, είναι σχεδιασμένη να γίνεται υπό καθεστώς διακριτικότητας και συχνά είναι δύσκολο να εντοπιστεί (τρίτη φάση: *stealthy data exfiltration* – διήθηση κρυφών δεδομένων). Οι επιτιθέμενοι μπορούν να εφαρμόσουν τεχνικές, όπως η “στεγανογραφία”, για την εξαγωγή των δεδομένων κατά την οποία τα κρυπτογραφούν με αλγόριθμους κρυπτογράφησης, καθιστώντας τα δεδομένα δυσνόητα και ακίνδυνα, άρα κάνοντας ακόμα πιο δύσκολο για τα συστήματα ασφαλείας να εγείρουν ειδοποιήσεις που υποδηλώνουν ύποπτη δραστηριότητα. Επιπροσθέτως, η TN μπορεί να χρησιμοποιηθεί για να αποκρύψει την κυκλοφορία του δικτύου, οδηγώντας την να μοιάζει ως φυσιολογική, εκκινώντας έτσι λιγότερους συναγερμούς. Με αυτό τον τρόπο, τα κλεμμένα δεδομένα ανακτώνται σταδιακά και κρυφά σε βάθος χρόνου, χρησιμοποιώντας συχνά κρυπτογραφημένα κανάλια επικοινωνίας, προκειμένου να κρύψουν ακόμα καλύτερα την κακόβουλη δραστηριότητα.

Προκειμένου, τέλος (τέταρτη φάση: *evasion & cover-up* – υπεκφυγή & κάλυψη), να αποφευχθεί ο εντοπισμός και η διαγραφή του «μολυσματικού παράγοντα», οι επιτιθέμενοι προσαρμόζουν διαρκώς τις τακτικές τους, χρησιμοποιώντας τις δυνατότητες της TN. Αυτό περιλαμβάνει την αποφυγή συστημάτων (πχ. IDS) που εντοπίζουν μη εξουσιοδοτημένη πρόσβαση στο σύστημα, την τροποποίηση των μεθόδων διήθησης και ανάκτησης δεδομένων, όπως, επίσης, την κάλυψη των ιχνών, διαγράφοντας αρχεία καταγραφής και παρεμβαίνοντας προσθέτοντας νόμιμα χαρακτηριστικά στο κακόβουλο δείγμα παρακάμπτοντας έτσι την λειτουργία των εργαλείων ασφαλείας. Οι τεχνικές αποφυγής (*evasion techniques*) που υποστηρίζονται από TN, είναι σε θέση να μιμηθούν τη νόμιμη συμπεριφορά των χρηστών, καθιστώντας δύσκολη για τους ειδικούς της κυβερνοασφάλειας, τη διαφοροποίηση μεταξύ γνήσιων και κακόβουλων ενεργειών (King et al., 2020).

4.2.3 Απόσπαση κωδικών & δεδομένων εισόδου (password attacks)

Οι επιθέσεις ξεκινούν και σε αυτή την περίπτωση με τη συλλογή δεδομένων και την αναγνώριση, σχεδόν όπως και οι επιθέσεις τύπου “data exfiltration” (πρώτη φάση: data collection & reconnaissance – συλλογή δεδομένων & αναγνώριση). Συλλέγονται, λοιπόν, μεγάλου όγκου δεδομένα, με τους επιτιθέμενους να αποσπούν πληροφορίες που βρίσκονται διαθέσιμες σε κοινή θέα, όπως σε κοινωνικά δίκτυα (social media) ή βάσεις με δεδομένα που έχουν ήδη παραβιαστεί, προκειμένου να συγκεντρωθούν όσο το δυνατόν περισσότερα προσωπικά δεδομένα, σχετικά με εν δυνάμει στόχους. Τα δεδομένα αυτά ενδέχεται να περιλαμβάνουν ονόματα, ημερομηνίες, διευθύνσεις ηλεκτρονικού ταχυδρομείου και διάφορες άλλες λεπτομέρειες, οι οποίες θα μπορούσαν να χρησιμοποιηθούν με σκοπό το λογισμικό TN να μαντέψει ή να «σπάσει» κωδικούς (Σχήμα 4).

Έτσι, εξοπλισμένος με πληθώρα προσωπικών πληροφοριών, ο αλγόριθμος TN (Alkhwaja et al., 2023) ξεκινά τη δημιουργία ιδιαίτερα στοχευμένων κωδικών (δεύτερη φάση: password guessing & generation – εικασία & παραγωγή κωδικών). Οι αλγόριθμοι αυτοί είναι σε θέση να αναλύουν μοτίβα όσον αφορά τη διαδικτυακή συμπεριφορά των χρηστών, τις προτιμήσεις τους ή γνωστές συνήθειες σχηματισμού κωδικών, προκειμένου να δημιουργήσουν σύνθετες εικασίες (κωδικών). Επιπροσθέτως, η TN μπορεί να δημιουργήσει λεξικά από κωδικούς και να εφαρμόσει τεχνικές, όπως “ωμή βία” (brute force) ή “επιθέσεις λεξικών” (dictionary attacks), έτσι ώστε να δοκιμάζει συστηματικά διάφορους συνδυασμούς μέχρι να εντοπίσει το σωστό κωδικό.

Brute Force Attacks Explained

In a brute force attack, a cybercriminal uses trial and error to try and break into a device, network, or website.



Σχήμα 4. Επεξήγηση της τεχνικής της “επίθεσης κωδικών” με τη μορφή της “ωμής βίας” (brute force)

Πηγή: <https://us.norton.com/blog/emerging-threats/brute-force-attack>

Στην επόμενη φάση συγκαταλέγεται η συμπλήρωση διαπιστευτηρίων και ο «ψεκασμός» κωδικών (τρίτη φάση: credential stuffing & password spraying). Στη συμπλήρωση διαπιστευτηρίων, χρησιμοποιούνται κλεμμένα ονοματεπώνυμα και κωδικοί από παραβιασμένες ιστοσελίδες, επιδιώκοντας πρόσβαση σε άλλους λογαριασμούς χρηστών, στους οποίους ενδέχεται να χρησιμοποιούνται τα ίδια διαπιστευτήρια. Η TN βοηθά στον αυτοματισμό αυτής της διαδικασίας, ώστε να επιτυγχάνεται μεγάλη κλίμακα εφαρμογών, επιτρέποντας τους επιτιθέμενους να δοκιμάζουν γρήγορα, αναρίθμητους λογαριασμούς. Στην πρακτική του «ψεκασμού» κωδικών, οι επιτιθέμενοι δοκιμάζουν ορισμένους ευρέως χρησιμοποιούμενους κωδικούς σε μεγάλο πλήθος λογαριασμών, στηριζόμενοι στην TN προκειμένου να βελτιστοποιήσει την προσέγγισή τους και να μεγιστοποιήσει τις πιθανότητες επιτυχίας τους.

Μια από τις πλέον ανησυχητικές πτυχές των βασισμένων στην TN επιθέσεων, είναι η προσαρμοστικότητα τους. Εδώ η επίθεση υπεισέρχεται στην τέταρτη φάση της, τις “*προσαρμοσμένες & εξελισσόμενες επιθέσεις*” (adaptive & evolving attacks). Οι αλγόριθμοι είναι σχεδιασμένοι να μαθαίνουν από τις επιτυχίες και αποτυχίες τους (πχ. μπορεί να χρησιμοποιούνται αλγόριθμοι μηχανικής μάθησης τύπου reinforcement learning ή self-learning για να προσαρμοστούν και να βελτιωθούν τα μοντέλα που είχαν αρχικά παραχθεί), ανανεώνοντας τις στρατηγικές τους σε βάθος χρόνου, ενώ προσαρμόζονται τόσο σε αλλαγές της συμπεριφοράς των χρηστών όσο και των αντιμέτρων ασφαλείας, καθιστώντας τους ακόμη πιο αποτελεσματικούς και μια πραγματική πρόκληση στο να αντιμετωπιστούν (Martin & Tokutomi, 2012).

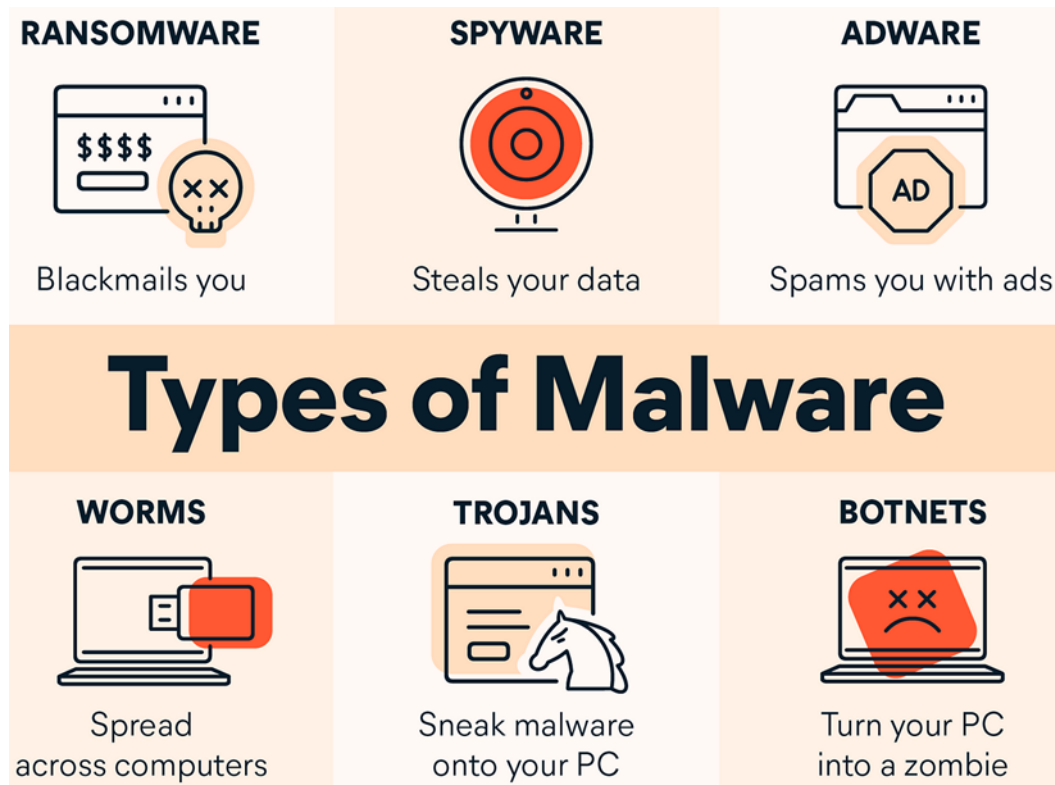
4.2.4 Δημιουργία κακόβουλου λογισμικού (malware generation)

Η πρακτική αυτή ξεκινά με τη συγκέντρωση δεδομένων και την αναγνώριση (πρώτη φάση: data gathering & reconnaissance), όπου συγκεντρώνονται μεγάλες ποσότητες δεδομένων από δυνητικούς στόχους, συμπεριλαμβανομένων στοιχείων εγγραφής δικτύων, συμπεριφορές χρηστών και τρωτά σημεία λογισμικών. Τέτοιες πληροφορίες βοηθούν στην αναγνώριση των πλέον ευπαθών στόχων προς εκμετάλλευση. Σε επόμενη φάση (δεύτερη φάση: AI-powered malware creation – ενισχυόμενη από την TN δημιουργία κακόβουλου λογισμικού) δημιουργείται το κακόβουλο λογισμικό (malware) μέσα από μια ευρεία λίστα επιλογών (Σχήμα 5), με τους επιτιθέμενους να αξιοποιούν αλγορίθμους TN με σκοπό την κατασκευή του λογισμικού ακριβώς βάσει του περιβάλλοντος του στόχου. Το λογισμικό αυτό μπορεί να προσαρμοστεί σε συνθήκες που αλλάζουν, καθιστώντας δύσκολο τον εντοπισμό και την καταπολέμησή του. Η τεχνολογία της

μηχανικής μάθησης μπορεί, επίσης, να χρησιμοποιηθεί με σκοπό τη δημιουργία εξειδικευμένων emails «ψαρέματος» (phishing) ή σχημάτων κοινωνικής μηχανικής (social engineering), ώστε το λογισμικό να παραδοθεί στα υποψήφια θύματα, ενώ μπορούν να αξιοποιηθούν διάφοροι τρόποι εγκατάστασής του, όπως drive-by downloads από κακόβουλες ιστοσελίδες ή ιστοσελίδες που έχουν μολυνθεί, χρήση μολυσμένων μέσων υλικού κ.α.

Το κακόβουλο λογισμικό που κατασκευάζει η TN έχει τη δυνατότητα να αποφεύγει τα παραδοσιακά αντίμετρα κυβερνο-ασφάλειας (τρίτη φάση: evasion & stealth – υπεκφυγή & απόκρυψη). Συγκεκριμένα, μπορεί και αλλάζει τη συμπεριφορά ή τα ίχνη του σε πραγματικό χρόνο, καθιστώντας δύσκολο τον εντοπισμό του από τα αντι-ιικά και τα IDS. Τα κακόβουλα αυτά στελέχη μπορούν να μιμηθούν τη νόμιμη διακίνηση στο διαδίκτυο, να αναμειχθούν με άλλα στοιχεία και να παραμείνουν αδρανή έως ότου εντοπίσουν μια ευκαιρία ενεργοποίησης του κακόβουλου κώδικα που περιλαμβάνουν.

Το πλέον ανησυχητικό με την εν λόγω πρακτική που υποστηρίζεται από την TN, είναι η ικανότητα του κακόβουλου λογισμικού να μαθαίνει και να προσαρμόζεται (τέταρτη φάση: continuous learning & adaptation – συνεχής μάθηση & προσαρμογή). Όσο αλληλοεπιδρά με το περιβάλλον του στόχου του, συγκεντρώνει πληροφορίες και τροποποιεί κατάλληλα τις τακτικές των επιθέσεών του. Αυτή η ικανότητα αυτό-βελτίωσης το καθιστά ακόμα πιο επικίνδυνο και αποτελεσματικό σε βάθος χρόνου, θέτοντας σοβαρές προκλήσεις για τα συστήματα των αμυνομένων (Smith et al., 2021).



Σχήμα 5. Οι πλέον γνωστοί τύποι κακόβουλων λογισμικών

Πηγή: <https://medium.com/@CipherStructor/emerging-threat-next-generation-malware-attacks-48217f4b84af>

4.2.5 Τεχνική της «αξιοποίησης άγνωστων τρωτών σημείων» (zero-day exploits)

Αυτού του είδους οι επιθέσεις, ξεκινάνε με τις προσπάθειες εντοπισμού δυνητικών αδυναμιών του συστήματος-στόχου (πρώτη φάση: discovery & analysis – ανακάλυψη & ανάλυση). Οι αλγόριθμοι TN εξερευνούν ποικίλα είδη λογισμικού και δίκτυα, αναζητώντας μοτίβα και ανωμαλίες που θα μπορούσαν να υποδεικνύουν αδυναμίες. Από τη στιγμή που θα εντοπιστεί μια πιθανή αδυναμία, το σύστημα TN του επιτιθέμενου αρχίζει να την επεξεργάζεται, μια ενέργεια που περιλαμβάνει ανίχνευση του συστήματος-στόχου, τη συγκέντρωση πληροφοριών και την εκτίμηση των δυνατοτήτων που δίνονται για εκμετάλλευσή της.

Τα μοντέλα μηχανικής μάθησης (της TN) μπορούν να αξιολογήσουν τάχιστα το επίπεδο αμυνών του στόχου. Αυτό μπορεί να γίνει είτε με ανίχνευση ανωμαλιών (Anomaly detection) στο σύστημα είτε με εκμάθηση χαρακτηριστικών χωρίς επίβλεψη (unsupervised feature learning). Συγκεκριμένα, μπορούν να εκπαιδευτούν και να ανιχνεύσουν ανωμαλίες στην συμπεριφορά των υπολογιστών ή του δικτύου. Επίσης, για την εκμάθηση των χαρακτηριστικών αντλούνται δεδομένα,

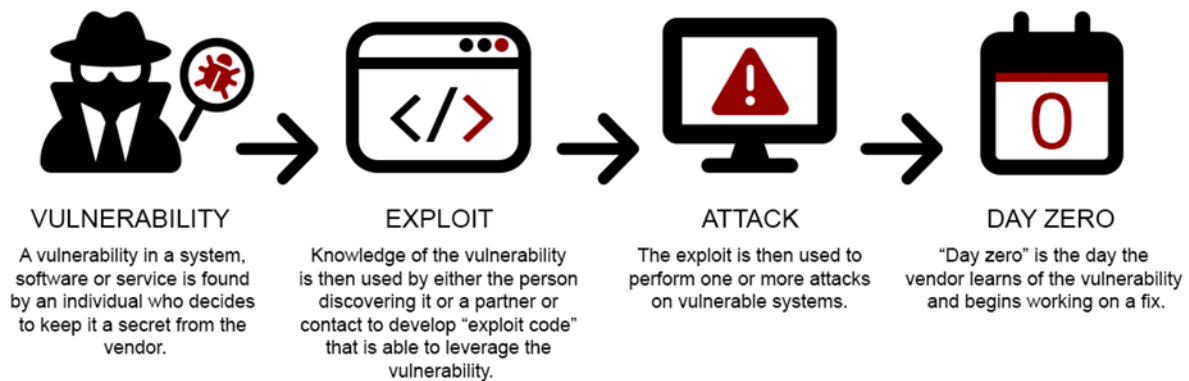
τα οποία είναι χωρίς επισημάνσεις (labels), καθιστώντας δύσκολο για το αμυντικό σύστημα να εντοπίσει τέτοιες πρώιμες προσπάθειες αναγνώρισης (Deldar & Abadi 2023).

Από τη στιγμή που επιβεβαιώνεται η ύπαρξη μιας αδυναμίας και εκτιμάται ως “εκμεταλλεύσιμη”, το σύστημα TN του επιτιθέμενου αναπτύσσει μεθόδους ικανές για το σκοπό αυτό, στις οποίες συγκαταλέγονται η δημιουργία κακόβουλου φορτίου ή κώδικα (code) ειδικά σχεδιασμένου να εκμεταλλεύεται την εντοπισθείσα αδυναμία (δεύτερη φάση: exploitation - εκμετάλλευση). Η TN μπορεί να προσαρμόσει και να τροποποιήσει τα φορτία σε πραγματικό χρόνο (real-time modification), με σκοπό να αποφύγει τον εντοπισμό τουλάχιστον από τα παραδοσιακά αντίμετρα ασφαλείας (όπως τείχη προστασίας και IDS). Το σύστημα TN του επιτιθέμενου, μπορεί να χρησιμοποιήσει επίσης τακτικές αποφυγής, όπως ο πολυμορφικός κώδικας, ώστε να αλλάζουν συνεχώς τη φύση των δραστηριοτήτων που εκτελούν, προκαλώντας δυσκολίες σε λογισμικά προστασίας τόσο στο να τους αναγνωρίσουν όσο και να παρεμποδίσουν την παράνομη αξιοποίηση των δεδομένων στα οποία απέκτησαν πρόσβαση.

Στην επόμενη φάση (τρίτη φάση: delivery & execution – παράδοση & εκτέλεση), το σύστημα TN του επιτιθέμενου αναζητά τρόπους για να προχωρήσει στην εκμετάλλευση της ευπάθειας του συστήματος-στόχου, πρακτική που περιλαμβάνει τακτικές κοινωνικής μηχανικής, emails «*ψαρέματος*» ή την ανάδειξη άλλων αδυναμιών, με απώτερο σκοπό την απόκτηση αρχικής πρόσβασης (Σχήμα 6). Από τη στιγμή, λοιπόν, που το σύστημα TN εισέρχεται στο στόχο, εκμεταλλεύεται την αδυναμία που εντόπισε (μέσω της εκτέλεσης του κακόβουλου κώδικα που εκχύθηκε), επιτρέποντας στον επιτιθέμενο να αποκτήσει μη-εξουσιοδοτημένη πρόσβαση, να υποκλέψει δεδομένα, να εγκαταστήσει άλλο κακόβουλο λογισμικό ή να εκτελέσει διάφορες επιζήμιες ενέργειες. Η συνδρομή της TN καθιστάται αρκετά εξελιγμένη και έτσι είναι πιο δύσκολη να εντοπιστεί, μεγιστοποιώντας τις πιθανότητες επιτυχίας.

Η τελευταία φάση (τέταρτη φάση: evasion & persistence – υπεκφυγή & επιμονή) αφορά τη διατήρηση του ελέγχου και την αποφυγή εντοπισμού, με το λογισμικό αξιοποίησης των τρωτών σημείων να εφαρμόζει συχνά τεχνικές αποφυγής καθ’ όλη τη διάρκεια του κύκλου ζωής της επίθεσης του. Η TN μπορεί να προσαρμοστεί δυναμικά στα αντίμετρα κυβερνο-ασφάλειας με τα οποία θα έρθει αντιμέτωπη, μαθαίνοντας τις πρακτικές τους και τροποποιώντας αναλόγως τις δικές της τακτικές. Επιπροσθέτως, αυτές οι επιθέσεις μπορεί να εγκαταστήσουν μηχανισμούς που εμμένουν, διασφαλίζοντας διαρκή πρόσβαση στο προσβαλλόμενο σύστημα, ακόμη και αν επιδιορθωθεί η αρχική ευπάθεια. Η επίμονη «μόλυνση» μπορεί να ενεργοποιήσει μια διαρκή διαδικασία διήθησης σε δεδομένα, να επιτρέψει τον έλεγχο του συστήματος χωρίς εξουσιοδότηση

ή να μετατραπεί σε πλατφόρμα εκτέλεσης πρόσθετων επιθέσεων στο δίκτυο του συστήματος-στόχου (Vaisla & Saini, 2014).

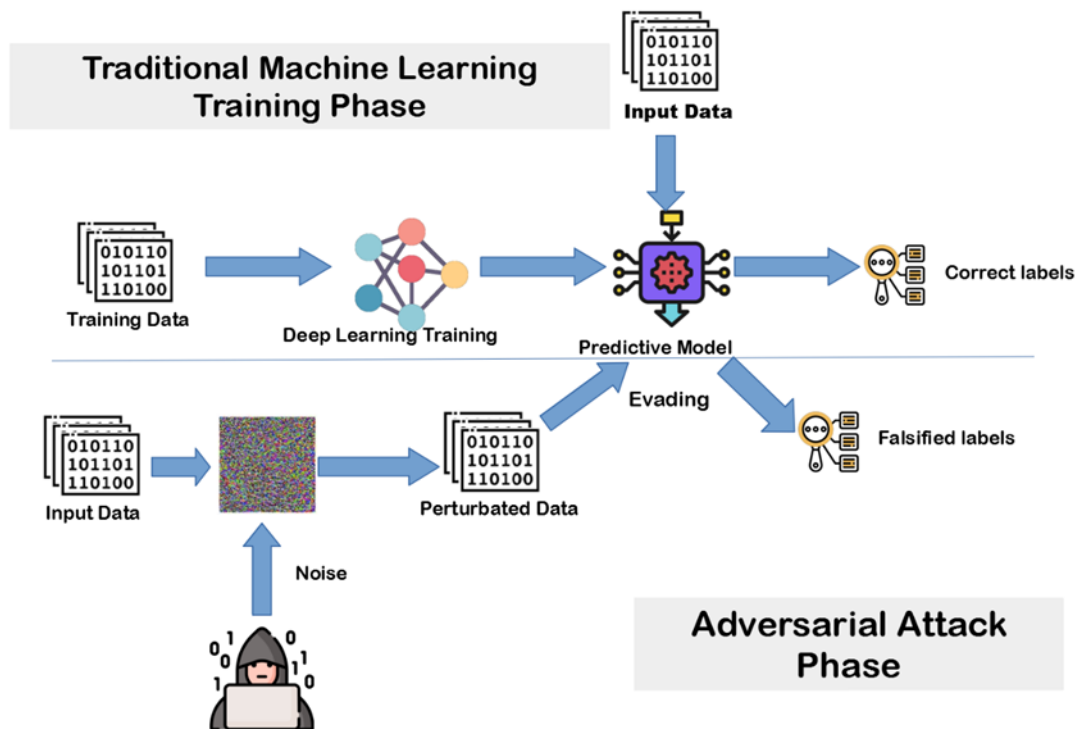


Σχήμα 6. Επεξήγηση του εντοπισμού τρωτών σημείων σε επιθέσεις τύπου "zero-day"

Πηγή: <https://us.norton.com/blog/emerging-threats/brute-force-attack>

4.2.6 Αντίστροφη μηχανική μάθηση (adversarial machine learning)

Στην πρώτη της φάση (data poisoning – δηλητηρίαση δεδομένων), η διαδικασία περιλαμβάνει την «εισαγωγή» κακόβουλων δεδομένων στο σύνολο των εκπαιδευτικών πληροφοριών που χρησιμοποιούνται για την ανάπτυξη του μοντέλου μηχανικής μάθησης του στόχου. Αυτό για να γίνει απαιτεί πρόσβαση στο υπολογιστικό σύστημα. Έτσι, τροποποιούνται ή προστίθενται παραπλανητικά δεδομένα που θα χειραγωγήσουν τη μαθησιακή διαδικασία του μοντέλου. Οι «μολυσμένες» αυτές εισαγωγές είναι σχεδιασμένες με τρόπο που να μπορούν να παραπλανήσουν το μοντέλο κατά τη διάρκεια της διαδικασίας εξαγωγής συμπερασμάτων, οδηγώντας σε λανθασμένες αποφάσεις ή ταξινομήσεις.



Σχήμα 7. Διαφορές παραδοσιακής & αντίστροφης μηχανικής μάθησης

Πηγή: <https://towardsdatascience.com/adversarial-machine-learning-mitigation-adversarial-learning-9ae04133c137>

Από τη στιγμή που το μοντέλο εκπαιδεύεται στο «μολυσμένο» σύνολο δεδομένων, οι επιτιθέμενοι χρησιμοποιούν τεχνικές αποφυγής για να εκμεταλλευτούν τις αδυναμίες που θα εντοπιστούν (δεύτερη φάση: model evasion – υπεκφυγή μοντέλου). Έτσι, κατασκευάζουν εισροές που έχουν σχεδιαστεί με σκοπό να παραπλανούν το μοντέλο ώστε να οδηγείται σε λανθασμένες προβλέψεις, πρακτική που περιλαμβάνει τη δημιουργία εικόνων ή εισροών, οι οποίες φαίνονται φυσιολογικές στους ανθρώπους αλλά ταξινομούνται λανθασμένα από ένα μοντέλο ΤΝ. Συνεπώς, οι αντίστροφες επιθέσεις στοχεύουν στην δημιουργία αδυναμιών, στα όρια λήψης αποφάσεων των μοντέλων. Αυτό μπορεί να έχει σοβαρές επιπτώσεις στην λήψη αποφάσεων όταν πρόκειται για κρίσιμα πεδία, όπως η ιατρική ή και η ασφάλεια. Έτσι, μπορούν να οδηγήσουν σε λανθασμένες προβλέψεις, που οδηγούν σε εσφαλμένη λήψη αποφάσεων, όπως για παράδειγμα τη λανθασμένη διάγνωση από ένα γιατρό για έναν ασθενή, η οποία θα τον οδηγήσει στο χειρουργείο, ενώ αυτό δεν χρειάζεται ή δεν πρέπει να γίνει.

Επιπροσθέτως, οι αντίστροφες επιθέσεις επωφελούνται συχνά από την ικανότητα μεταφοράς (τρίτη φάση: transferability), αφού μια επίθεση που λειτουργεί ενάντια σε ένα μοντέλο μπορεί να είναι εξίσου αποτελεσματική και σε άλλο, παρόμοιο, του ίδιου σχεδιαστή ή της ευρύτερης οικογένειας λογισμικού/κώδικα από όπου προήλθε. Οι επιτιθέμενοι μπορούν να δημιουργήσουν αντίστροφα παραδείγματα σε ένα αρχικό μοντέλο και στη συνέχεια να τα

αξιοποιήσουν ενάντια σε περισσότερο ασφαλή ή πιο ισχυρά προστατευμένα, αυξάνοντας την επίπτωσή τους.

Τέλος (τέταρτη φάση: stealth & persistence – απόκρυψη & επιμονή), προκειμένου να μεγιστοποιήσουν την επίδρασή τους, οι επιτιθέμενοι επιδιώκουν να εκτελέσουν τέτοιου είδους επιθέσεις, κρυφά, αναμιγνύοντας συχνά τα αντίστροφα παραδείγματα με νόμιμα δεδομένα, έτσι ώστε να αποφύγουν τον εντοπισμό. Επιπροσθέτως, προσαρμόζουν συχνά τις επιθετικές τους στρατηγικές με σκοπό να παρακάμπτουν τους εξελισσόμενους αμυντικούς μηχανισμούς, διασφαλίζοντας την επίμονη παρουσία τους και την επιτυχία έναντι των συστημάτων TN, τα οποία «μολύνουν» (Wang et al., 2023).

4.2.7 Απόκτηση μέσω της τεχνικής του «ψαρέματος» (phishing)

Οι επιτιθέμενοι ξεκινούν με τη συγκέντρωση πληροφοριών που αφορούν το στόχο (πρώτη φάση: reconnaissance - αναγνώριση), με τους αλγορίθμους TN να παρακολουθούν λογαριασμούς κοινωνικών δικτύων, δημόσιες βάσεις δεδομένων και διάφορες διαδικτυακές πηγές, προκειμένου να συνθέσουν μια ολοκληρωμένη εικόνα του υποψήφιου θύματος. Δεν αποκλείεται μάλιστα να αναλύουν προγενέστερα μοτίβα επικοινωνίας και συμπεριφορών, με σκοπό να εντοπίσουν τον πλέον αποτελεσματικό τρόπο «ψαρέματος» του θύματος (Σχήμα 8). Η στρατηγική αυτή θεωρείται καθοριστικής σημασίας για την προσωποποίηση της επίθεσης, αφού καθιστά την όλη προσέγγιση πιο πειστική.

Στην επόμενη φάση (δεύτερη φάση: customization - προσωποποίηση), αξιοποιείται η TN με σκοπό να κατασκευάσει ένα εξειδικευμένο / προσωποποιημένο μήνυμα βάσει των αναγκών ή προτιμήσεων του χρήστη. Οι επιτιθέμενοι μπορούν να κατασκευάσουν προσωποποιημένα μηνύματα εξαιρετικής ακρίβειας, emails ή ακόμη και φωνητικές κλήσεις που φαίνεται να προέρχονται από έμπιστες πηγές, όπως συνάδελφοι, φίλοι ή χρηματοπιστωτικά ιδρύματα, αφού η TN είναι ικανή να αναλύει γλωσσικά μοτίβα, τρόπους γραφής και ήχους, με σκοπό να κάνει τα μηνύματα να φαίνονται νόμιμα και να αναγνωρίζεται με δυσκολία η διαφορά τους από την πραγματική πηγή. Εν γένει, ο στόχος είναι η συναισθηματική εκμετάλλευση του θύματος μέσω της δημιουργίας αίσθησης του επείγοντος, έλλειψης της περιέργειας του θύματος, ικανοποίησης αναγκών π.χ. για περισσότερη κοινωνικοποίηση, κ.α.

Από τη στιγμή που το μήνυμα (το κειμενικό περιεχόμενό του) ετοιμαστεί, ο επιτιθέμενος χρησιμοποιεί τεχνικές υποστηριζόμενες από TN, ώστε να επιλέξει την πιο αποτελεσματική στρατηγική προσέγγισης του θύματός του (τρίτη φάση: delivery - παράδοση). Οι δυνητικές

προσεγγίσεις περιλαμβάνουν μηνύματα με επιβλαβή επισυναπτόμενα ή συνδέσμους (links), με τη μέθοδο προσέγγισης να επιλέγεται σύμφωνα με τις προτιμήσεις και τη συνήθη συμπεριφορά του θύματος, αυξάνοντας έτσι τις πιθανότητες επιτυχίας της επίθεσης.

Στην τελευταία φάση (τέταρτη φάση: exploitation - εκμετάλλευση) περιλαμβάνεται η αντίδραση/ανταπόκριση του θύματος. Όταν ο στόχος αλληλοεπιδρά με το μήνυμα «*ψαρέματος*», το εργαλείο TN του επιτιθέμενου μπορεί να αναλύσει αυτές τις αλληλεπιδράσεις σε πραγματικό χρόνο. Εάν το θύμα επιλέξει τον επισυναπτόμενο σύνδεσμο (κακόβουλης σελίδας που έχει κατασκευαστεί εκ των προτέρων) ή ανοίξει το αρχείο, ο επιτιθέμενος εκμεταλλεύεται αυτή την ευκαιρία για να εγκαταστήσει ένα κακόβουλο λογισμικό στο σύστημά του, να υποκλέψει στοιχεία ή να αποκτήσει πρόσβαση σε λογαριασμούς του. Οι αλγόριθμοι TN μπορούν να προσαρμοστούν στις ενέργειες ανταπόκρισης του θύματος, κάνοντας την επίθεση πιο δυναμική αλλά και πιο δύσκολη να εντοπιστεί. Για παράδειγμα, μέσα από την ανταλλαγή μηνυμάτων που φαίνονται σωστά/νομότυπα, ώστε να φθάσει σε ένα τελικό στόχο (όπως είναι η αποκάλυψη κάποιας ιδιωτικής πληροφορίας του χρήστη). Αυτό μοιάζει με τη χρήση chatbots που έχουν παρόμοια συμπεριφορά με έναν κανονικό άνθρωπο (Stivala et al., 2023).

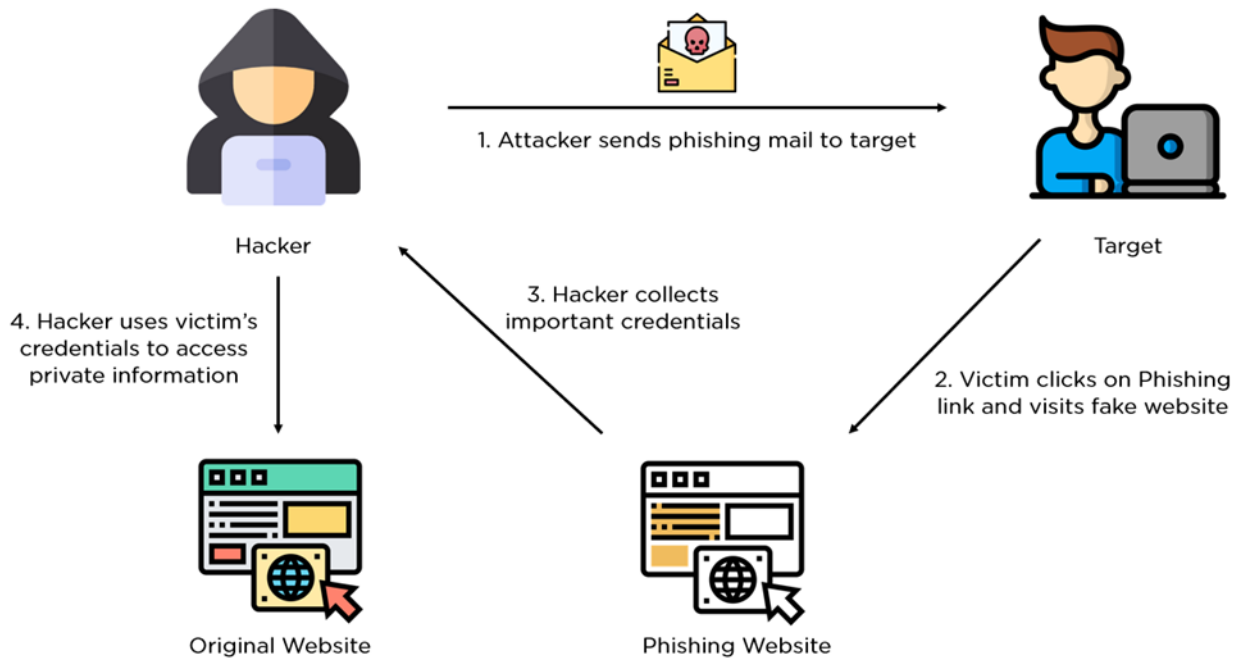
4.2.7.1 Συμπληρωματική χρήση NLP τεχνικής

Συμπληρωματικά, η τεχνική NLP εφαρμόζεται συνδυαστικά με τις δυνατότητες της TN. Αρχικά, λοιπόν, οι καθοδηγούμενες από TN, επιθέσεις NLP (πρώτη φάση: data gathering & reconnaissance), συγκεντρώνουν μεγάλους όγκους δεδομένων από πηγές και οι εξελεγμένοι αλγόριθμοι αναλύουν τα δεδομένα αυτά προκειμένου να δημιουργήσουν λεπτομερή εικόνα των δυνατικών τους στόχων, εντοπίζοντας, φυσικά, τα τρωτά τους σημεία.

Εξοπλισμένοι οι επιτιθέμενοι με πληροφόρηση, χρησιμοποιούν την TN για να δημιουργήσει περιεχόμενο «*ψαρέματος*» μέσα από emails ή άλλα μηνύματα (δεύτερη φάση: phishing & social engineering – ηλ. ψάρεμα και κοινωνική μηχανική), όπου συμπεριλαμβάνονται σχετικές με το στόχο πληροφορίες προκειμένου να τον εξαπατήσουν και να προχωρήσει σε συγκεκριμένες ενέργειες στο διαδίκτυο.

Από τη στιγμή που θα αποκτήσει ο επιτιθέμενος «πάτημα» εντός του συστήματος-στόχου, προχωρά σε διάφορες κακόβουλες ενέργειες, όπως την πρόσβαση σε διάφορων ειδών χρήσιμα δεδομένα/αρχεία (οικονομικές πληροφορίες οικονομικού ενδιαφέροντος κ.ά.), ενώ η TN προσαρμόζει σε πραγματικό χρόνο τις στρατηγικές της, ώστε να μαθαίνει μέσα από τις αλληλεπιδράσεις του χρήστη και να ανανεώνει την προσέγγισή της, χωρίς φυσικά να γίνεται αντιληπτή (τρίτη φάση: automated attacks – αυτοματοποιημένες επιθέσεις).

Οι δυνατότητες των επιθέσεων NLP περιλαμβάνουν την κατασκευή ψευδών ειδήσεων (fake news) ή πληροφοριών που αποπροσανατολίζουν, πέρα από την ικανότητά τους να καλύπτουν τα ίχνη τους και να αναμιγνύονται με τη νόμιμη κυκλοφορία του διαδικτύου, ώστε να καθίσταται δύσκολος ο εντοπισμός τους. Επιπλέον, η NLP μέθοδος θα μπορούσε να χρησιμοποιηθεί για τη μετάδοση της πληροφορίας με κάποιο είδος φωνητικού μηνύματος ή κλήσης (τέταρτη φάση: evasion & persistence – υπεκφυγή & παραμονή) (Ahmed & Uddin, 2020).



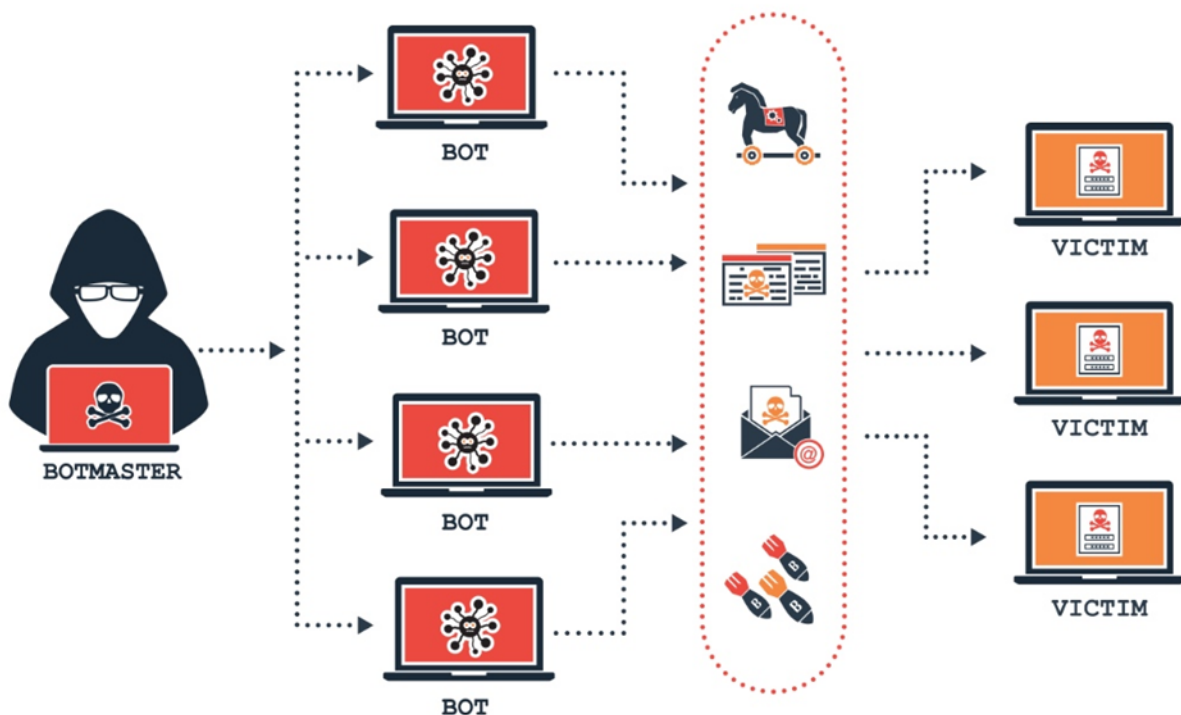
Σχήμα 8. Αναπαράσταση του τρόπου λειτουργίας μιας επίθεσης τύπου “phishing”

Πηγή: <https://www.linkedin.com/pulse/artful-dodger-outsmarting-paypal-phishing-scams-pratik-barahatte>

4.2.8 Αυτοματοποιημένες επιθέσεις (automated attacks)

Η πρώτη φάση της επίθεσης (data gathering & reconnaissance) είναι και εδώ η συγκέντρωση πληροφοριών και η παρακολούθηση υποψηφίων θυμάτων, με τους αλγορίθμους ΤΝ να αναζητούν ενδιαφέρουσες πληροφορίες ή κατάλληλα δεδομένα σε δημόσιες πηγές δεδομένων και να αναλύουν το ψηφιακό αποτύπωμα δυνητικών στόχων. Η φάση αυτή θεωρείται καθοριστικής σημασίας για την κατανόηση τόσο των αδυναμιών όσο και των τρωτών σημείων των θυμάτων, ενώ σε δεύτερο χρόνο αξιοποιείται η ικανότητα των συστημάτων ΤΝ να υπερέχουν των παραδοσιακών τους αντιπάλων (μέτρα ασφάλειας που δεν αξιοποιούν την ΤΝ), σε μεταβαλλόμενες συνθήκες.

Αυτό ακριβώς εκμεταλλεύονται οι επιτιθέμενοι που τα χρησιμοποιούν, με στόχο να συνθέσουν στρατηγικές επίθεσης βασισμένες στη γνώση που έχουν αποκτήσει οι μηχανές ΤΝ από την παρακολούθηση την οποία εμπεριστατωμένα έχουν διεξάγει. Έτσι, επιλέγονται αυτομάτως οι πλέον αποτελεσματικοί τρόποι επίθεσης (δεύτερη φάση: adaptive attack planning – προσαρμόσιμος σχεδιασμός επιθέσεων), καθορίζεται η καταλληλότερη στιγμή της επίθεσης αυτής και χρησιμοποιείται το πιο αποδοτικό - κατά την “άποψη” της ΤΝ - κακόβουλο λογισμικό, σύμφωνα με τις διαθέσιμες άμυνες και τα τρωτά σημεία που εντοπίστηκαν στο θύμα. Επιπροσθέτως, πρέπει να γίνεται και η σύνθεση των τρόπων επίθεσης, δηλαδή η σωστή ροή εκτέλεσής τους, με πιθανή μετάβαση των δεδομένων από τον ένα τρόπο στον άλλο.



Σχήμα 9. Αναπαράσταση του αυτοματοποιημένου τρόπου επιθέσεων μέσω ΤΝ

Πηγή: <https://www.techprevue.com/automated-attacks-threaten-account-security/>

Επιπλέον, οι επιθέσεις που βασίζονται στις ικανότητες της ΤΝ, έχουν (όπως ήδη σημειώθηκε και σε προηγούμενη μορφή επιθέσεων) σχεδιαστεί να παρακάμπτουν τα παραδοσιακά αντίμετρα κυβερνο-επίθεσης, αφού μπορούν να εφαρμόσουν εξελεγχμένες τεχνικές αποφυγής εντοπισμού (τρίτη φάση: execution & evasion – εκτέλεση & υπεκφυγή). Οι τεχνικές αυτές περιλαμβάνουν πολυμορφικό κακόβουλο λογισμικό (όπως στις επιθέσεις zero-day), το οποίο αλλάζει τον κώδικά του ώστε να αποφεύγει τον εντοπισμό που βασίζεται σε υπογραφές (signature-based detection) ή μιμείται νόμιμα μοτίβα διακίνησης της κυκλοφορίας του διαδικτύου, για να

«μπλεχτεί» με αυτά και να μην αναγνωρίζεται ως παράνομο. Ακόμα και σε αυτή την πρακτική επίθεσης, ένα από τα καίριας σημασίας χαρακτηριστικά του πολυμορφικού λογισμικού είναι η ικανότητα που διαθέτει να μαθαίνει και να προσαρμόζεται σε πραγματικό χρόνο. Όσο έρχεται σε επαφή με καινούριες άμυνες ή εμπόδια, το λογισμικό TN μπορεί να τροποποιήσει τις στρατηγικές και τεχνικές του, καθιστώντας το εξαιρετικά ανθεκτικό και επίμονο. Συνεπώς, αυτή η ικανότητα δια βίου μάθησης, το καθιστά ικανό να εκμεταλλεύεται ακόμη και τα πιο ενημερωμένα αντίμετρα από τη στιγμή που θα εντοπίσουν ευπάθειες (Buchanan et al., 2020).

4.2.9 Τεχνική της «υποστηριζόμενης από TN επίθεσης κυβερνο-ρομπότ» (AI-powered bots)

Οι αυτοματοποιημένες μηχανές (ή αυτοματοποιημένοι κώδικες λογισμικού), γνωστές και ως “bots” (σύντμηση της αγγλικής λέξης “robots”), χρησιμοποιούνται από κυβερνο-εγκληματίες και έχουν εξελιχθεί σε ιδιαίτερα εξειδικευμένα εργαλεία κυβερνο-επιθέσεων, εφαρμόζοντας εξελιγμένες τεχνικές διείσδυσης και προσβολής των διασυνδεδεμένων συστημάτων και δικτύων. Τα bots, επίσης, αξιοποιούν τη δύναμη των αλγορίθμων μηχανικής μάθησης. Ειδικότερα, σε πρώτη φάση (initial reconnaissance – αρχική αναγνώριση), παρακολουθούν πιθανούς στόχους, συγκεντρώνοντας πληροφορίες, ελέγχοντας το διαδίκτυο για τρωτά σημεία σε συστήματα ή δίκτυα και έπειτα χρησιμοποιούν αυτοματοποιημένα εργαλεία αναγνώρισης πιθανών σημείων εισόδου (δεύτερη φάση: exploitation & infiltration – εκμετάλλευση & διήθηση).

Η συμπεριφορά τους δε διαφέρει από τα προαναφερθέντα είδη επιθέσεων και οι πρακτικές τους περιλαμβάνουν μηνύματα «ψαρέματος» για να έχουν πρόσβαση σε κωδικούς ή άλλα πολύτιμα δεδομένα του θύματος (τρίτη φάση: lateral movement & persistence). Τα bots, όμως, φαίνεται πως εμπλέκονται σε επιθέσεις “κατανεμημένης άρνησης υπηρεσιών” (Distributed Denial-of-Service attacks, DDoS attacks), γεμίζοντας το σύστημα ή τις υπηρεσίες του θύματος με στοιχεία κυκλοφορίας ή με αυξημένο πλήθος αιτήσεων που παρεμποδίζουν την ομαλή του/τους λειτουργία (τέταρτη φάση: data exfiltration & malicious actions – διήθηση δεδομένων & κακόβουλες ενέργειες). Με τη χρήση της μηχανικής μάθησης, καθίσταται εφικτό για τα bots να προσαρμόζουν τις τακτικές τους σύμφωνα με το διαθέσιμο αμυντικό σύστημα του θύματος, θέτοντας σοβαρές προκλήσεις στην ομαλή/αναμενόμενη λειτουργία και του συστήματος ασφαλείας (Guembe et al., 2022).

Αξίζει τέλος να σημειωθεί, ότι μεταξύ των αυτοματοποιημένων επιθέσεων (automated attacks) και των υποστηριζόμενων από την TN επιθέσεων κυβερνο-ρομπότ (AI-powered bots attacks), συναντώνται ορισμένες μικρές διαφορές, παρά τη φαινομενική τους ομοιότητα. Οι πρώτες στηρίζονται σε κώδικες εκ των προτέρων προγραμματισμένους (pre-programmed scripts) ή

εργαλεία που ακολουθούν προκαθορισμένες οδηγίες και είναι λιγότερο εξειδικευμένα, ενώ οι δεύτερες χαρακτηρίζονται από υψηλού βαθμού εξειδίκευση (Buchanan et al., 2020). Επιπλέον, οι πρώτες θεωρούνται στατικές και δεν προσαρμόζονται σε αλλαγές στο περιβάλλον του χρήστη ούτε μπορούν να εξελιχθούν σε βάθος χρόνου μέσω μιας διαδικασίας μάθησης (machine learning), σε αντίθεση με τις δεύτερες που μαθαίνουν από αλληλεπιδράσεις του χρήστη και καθίστανται αποτελεσματικότερες στις επιθέσεις τους, σε βάθος χρόνου. Οι πρώτες επίσης είναι σχεδιασμένες για συγκεκριμένα καθήκοντα, χωρίς υψηλή πολυπλοκότητα, ενώ οι δεύτερες μπορούν να εκτελέσουν πολύπλοκες ενέργειες και να προβούν σε ανάλογη λήψη αποφάσεων. Τέλος, οι πρώτες συνήθως χρησιμοποιούν συχνότερα προσεγγίσεις τύπου “brute force”, στοχεύοντας γνωστά τρωτά σημεία ή αδυναμίες που ισχύουν για περισσότερους του ενός χρήστες, ενώ οι δεύτερες μπορούν να αναγνωρίσουν και να εκθέσουν νέες αδυναμίες, μοναδικές για κάθε χρήστη (Guembe et al., 2022).

4.2.10 Τεχνητή αξιοποίησης υπερρεαλιστικών μέσων (deep fakes)

Η εν λόγω πρακτική έχει περιγραφεί στην ουσία από την επίθεση τύπου phishing και σε κάθε επόμενη που παρουσιάστηκε στη τρέχουσα αναφορά. Εν ολίγοις, στην πρώτη της φάση (data collection & target selection – συλλογή δεδομένων & επιλογή στόχου) συγκεντρώνει σημαντικές ποσότητες δεδομένων για τους στόχους, από πλήθος πηγών, όπως αναρτήσεις σε κοινωνικά δίκτυα, συνεντεύξεις, ομιλίες ή οποιοδήποτε δημοσίως διαθέσιμο ψηφιακό υλικό, παρέχοντας γνώση για το θύμα, συμπεριλαμβανομένων της φωνής του, των κινήσεών του και της γενικότερης συμπεριφοράς. Οι επιτιθέμενοι μάλιστα αναζητούν οποιαδήποτε πληροφορία θα καθιστούσε την επίθεσή τους πιο πειστική, με τις επιθέσεις υπερρεαλιστικών μέσων να αγγίζουν το 66% των περιστατικών που συμμετείχαν σε αντίστοιχη μελέτη εντός του 2022, ενώ τα email τύπου phishing έφτασαν στο 78% των περιστατικών (Σχήμα 10).

Οι δυνατότητες της μηχανικής μάθησης χρησιμοποιούνται και εδώ, ιδίως οι ικανότητες που προσφέρει η βαθιά μάθηση (deep learning), με σκοπό να δημιουργηθεί ένα εξειδικευμένο μοντέλο TN, γνωστό ως “Αντίστροφα Δημιουργικό Δίκτυο” (Generative Adversarial Network, GAN) που αποτελείται από δύο νευρωνικά δίκτυα: το δημιουργό (generator) και το διευκρινιστή (discriminator) (δεύτερη φάση: AI model training). Αυτά λειτουργούν το ένα ενάντια στο άλλο, με το πρώτο να δημιουργεί ψεύτικο περιεχόμενο, ενώ το δεύτερο αξιολογεί το αποτέλεσμα του πρώτου για το επίπεδο αυθεντικότητας που έχει πετύχει. Σε βάθος χρόνου, ο δημιουργός τελειοποιεί τις δημιουργίες του, επιτυγχάνοντας εξαιρετικά ρεαλιστικό, ψευδές περιεχόμενο, αφού μαθαίνει από τα λάθη που του αναγνωρίζει ο διευκρινιστής. Όταν το μοντέλο έχει εν τέλει εκπαιδευτεί επαρκώς,

ο επιτιθέμενος μπορεί να το χρησιμοποιήσει προκειμένου να δημιουργήσει υπερ-ρεαλιστικό υλικό, το οποίο θα χρησιμοποιηθεί ενάντια στον στόχο για να του προκαλέσει κάποια ζημία (τρίτη φάση: content creation & deployment – δημιουργία περιεχόμενου & διάταξη), όπως για παράδειγμα βίντεο του στόχου να μιλά ή να εκτελεί κινήσεις, με τρόπο που εξυπηρετεί κακόβουλα συμφέροντα ή ενέργειες, προωθώντας το έπειτα σε κανάλια επικοινωνίας, όπως τα κοινωνικά δίκτυα.

Το υπερ-ρεαλιστικό υλικό ενδέχεται να σημειώσει πλήθος δυσάρεστων επιπτώσεων (τέταρτη φάση: impact & consequences – επίπτωση & συνέπειες), ανάλογα με τους στόχους του επιτιθέμενου, αφού μπορεί να χρησιμοποιηθεί για να εκβιάσει, να εκκινήσει καμπάνιες παραπληροφόρησης ή να χειραγωγήσει την κοινή γνώμη. Το θύμα, όμως, ενδέχεται να υποστεί σοβαρές συνέπειες στη φήμη του, οικονομική ζημία ή ακόμη και νομικές κυρώσεις εάν το υλικό είναι ιδιαίτερα πειστικό (Frankovits & Mirsky, 2023).



Σχήμα 10. Ποσοστά συμβάντων χρηστών που διαπίστωσαν απόπειρες απάτης μέσω χρήσης deepfake πρακτικών (στοιχεία 2022, σε διεθνή κλίμακα)

Πηγή: <https://securityreviewmag.com/?p=24720>

4.3 Συνήθειες στόχοι & κίνητρα των κυβερνοεπιθέσεων

Το τοπίο στο μέτωπο του κυβερνο-πολέμου είναι ένα διαρκώς εξελισσόμενο πεδίο και η ανάδυση της ΤΝ έχει εισάγει μια νέα διάσταση στο οπλοστάσιο των εγκληματιών που δραστηριοποιούνται σε αυτό, των επιτιθέμενων που υποστηρίζονται από κρατικούς φορείς (state-sponsored hackers) και οποιονδήποτε άλλων διατίθενται να υποκλέψουν ευαίσθητο υλικό ή να διεκπεραιώσουν άλλες κακόβουλες ενέργειες, προς ιδίον όφελος. Οι βασισμένες στην ΤΝ κυβερνο-επιθέσεις παρουσιάζουν τη δυναμική επίτευξης σοβαρών παρενοχλήσεων ανυπολόγιστης κλίμακας σε επίπεδο οικονομίας, της προσβολής κρίσιμων υποδομών και της διάρρηξης δεδομένων που θεωρούνταν ασφαλή. Ως εκ τούτου, κρίνεται σκόπιμο να καταγραφούν οι εν δυνάμει στόχοι αυτών

των επιθέσεων, προκειμένου να γίνει πιο σαφής και ο λόγος για τον οποίο απειλούνται (Comiter, 2019).

Ξεκινώντας με μεμονωμένα πρόσωπα, αυτά θα πρέπει να θεωρούν τους εαυτούς τους στο επίκεντρο των βλέψεων των εγκληματιών του διαδικτύου, μιας και εύκολα μπορούν να τεθούν στο στόχαστρό τους, αφού οι αυτοματοποιημένες επιθέσεις προσπαθούν να εντοπίζουν τρωτά σημεία στην ψηφιακή άμυνα ενός συστήματος και έπειτα να αναζητούν χρήσιμα δεδομένα που θα μπορούσαν να εκμεταλλευτούν, τα οποία μπορεί να ανήκουν ή να αφορούν τον οποιοδήποτε. Όμως, μια διάκριση που συναντάται στις επιθέσεις, αφορά τη στόχευσή τους και την κατάσταση ενός ατόμου, με τις διασημότητες, τους οικονομικά ευκατάστατους, τους πολιτικούς, τους γνωστούς επιχειρηματίες και άλλα πρόσωπα, να κινδυνεύουν πιθανόν περισσότερο εάν στοχοποιηθούν. Κι αυτό διότι ταιριάζουν με τα κίνητρα των επιτιθέμενων (πχ. οικονομικά ή πολιτικά) ενώ υπάρχει περισσότερη πληροφορία για τα άτομα αυτά που μπορεί να συλλεχθεί από πολλές πηγές.

Ειδικότερα, οι προσωπικές συσκευές και οι διαδικτυακοί λογαριασμοί μπορούν να προσβληθούν μέσω των επιθέσεων που υποστηρίζονται από λογισμικά TN, δεδομένου ότι η μεγάλη πλειοψηφία των πολιτών διαθέτει λογαριασμούς σε κοινωνικά δίκτυα, από τα οποία τα συστήματα TN αποσπούν πληροφορίες χάριν στις δυνατότητες της κοινωνικής μηχανικής, της υπερρεαλιστικής αναπαράστασης προσώπων ή της εξειδικευμένης/προσωποποιημένης (καθοδηγούμενης από την TN) κατασκευής μηνυμάτων «ψαρέματος». Τα κίνητρα σε αυτή την περίπτωση περιλαμβάνουν την υποκλοπή στοιχείων ταυτότητας, τις χρηματοοικονομικές απάτες, την κλοπή περιουσιακών στοιχείων ή/και τη διαφόρων τύπων (π.χ. σεξουαλική) παρενόχληση (Biggio & Roli, 2018).

Στη συνέχεια ακολουθούν οι επιχειρήσεις και οι οργανισμοί, ιδιωτικού συμφέροντος, πολλοί από τους οποίους διαχειρίζονται μεγάλα ποσά (πχ. χρηματοπιστωτικά ιδρύματα) και ως εκ τούτου θεωρούνται σοβαροί στόχοι οικονομικού ενδιαφέροντος. Οι υποστηριζόμενες από TN επιθέσεις προς αυτούς τους στόχους, συνθέτουν μια σοβαρή απειλή. Οι επιτιθέμενοι εκτελούν επιθέσεις ιδιαίτερα εξειδικευμένες, όπως ιδιαίτερα εξελιγμένες τακτικές «ψαρέματος», αυτοματοποιημένες πρακτικές αποφυγής εντοπισμού & διενέργειας απάτης καθώς και έκχυση υψηλής ευφύιας κακόβουλου λογισμικού. Τα βασικά κίνητρα περιλαμβάνουν προφανώς το οικονομικό όφελος ή τη διενέργεια πράξεων κατασκοπείας του οργανισμού από ανταγωνιστές, με στόχο την απόκτηση ανταγωνιστικού πλεονεκτήματος. Τέτοιες επιθέσεις μπορεί να έχουν ως αποτέλεσμα τη διάρρηξη βάσεων δεδομένων, την αποκάλυψη προσωπικών ή ευαίσθητων πληροφοριών που οδηγεί στη παραβίαση νομοθεσιών ιδιωτικότητας, την αποκάλυψη εταιρικών μυστικών, σοβαρές οικονομικές απώλειες ή ζημιά στη φήμη του οργανισμού (Dignum, 2019).

Επόμενοι σοβαροί στόχοι αποτελούν τα κράτη και συγκεκριμένα εθνικοί φορείς που πρωτίστως συνδέονται με πολιτικές ενέργειες/αποφάσεις και την άμυνα του έθνους τόσο σε πραγματικό χώρο όσο και στο ψηφιακό περιβάλλον (για παράδειγμα, ορισμένοι δορυφόροι αποτελούν μέρος της εθνικής άμυνας/ασφάλειας αλλά προσεγγίζονται μέσω διαδικτύου). Οι hackers που εργάζονται ή χρηματοδοτούνται από ένα κράτος προκειμένου να εκτελούν ενέργειες παραβίασης σε συστήματα ή δίκτυα εχθρικών κρατών (ή κρατών που επιθυμούν να κατακτήσουν ή να ζημιώσουν), αξιοποιούν την TN με σκοπό να προκαλέσουν προβλήματα ή να αποκτήσουν πρόσβαση σε κυβερνητικούς φορείς ή δίκτυα οργανισμών που σχετίζονται με την άμυνα. Οι λόγοι ποικίλουν και μπορούν να περιλαμβάνουν πολιτική επιρροή, κατασκοπεία, απόκτηση στρατηγικού πλεονεκτήματος σε επίπεδο διεθνών συγκρούσεων κ.α. Η TN μπορεί να χρησιμοποιηθεί για να αυτοματοποιηθούν οι ενέργειες παρακολούθησης, να αναγνωριστούν αδυναμίες και να διεξαχθούν ακριβείς επιθέσεις σε υποδομές κρίσιμης σημασίας, δυνητικά αποσταθεροποιώντας ολόκληρο το κράτος (Matt, Modrák & Zsifkovits, 2020).

Όσον αφορά μάλιστα τις υποδομές κρίσιμης σημασίας, όπως τα δίκτυα ηλεκτρικής ενέργειας, τα συστήματα παροχής πόσιμου ύδατος και τα δίκτυα μεταφορών και επικοινωνίας, οι στηριζόμενες στην TN κυβερνο-επιθέσεις ενδέχεται να προκαλέσουν ένα καταστροφικό αποτέλεσμα. Οι επιτιθέμενοι ενδέχεται να στοχεύουν στην αποδιοργάνωση σημαντικών υπηρεσιών ή να τις θέσουν υπό καθεστώς ομηρίας με σκοπό την πληρωμή χρημάτων. Τα κίνητρα μπορούν να ποικίλουν και εδώ, αφού σε αυτά συγκαταλέγεται το οικονομικό όφελος, η αλλαγή πολιτικής ή ιδεολογικής ατζέντας (ιδίως σε ζητήματα εξωτερικής πολιτικής) και η πιθανή επιρροή στη λήψη αποφάσεων, παράγοντας που αποτελεί ένα σοβαρό ζήτημα εθνικής ασφάλειας (Duddu, 2018).

Ειδική μνεία στο σκέλος των κρίσιμων υποδομών, θα πρέπει να γίνει για αυτές που αφορούν την υγειονομική περίθαλψη και την έρευνα. Ο τομέας υγειονομικής περίθαλψης έχει αποτελέσει με αυξημένη συχνότητα, στόχο κυβερνο-επιθέσεων, εξαιτίας των χρήσιμων & ευαίσθητων δεδομένων που διαφυλάσσει. Οι επιθέσεις που υποστηρίζονται από συστήματα TN, συχνά στοχοποιούν οργανισμούς ιατρικής έρευνας, φαρμακευτικές εταιρείες αλλά και ολόκληρες μονάδες υγείας. Οι επιτιθέμενοι μπορεί να αναζητούν και να υποκλέπτουν δεδομένα ασθενών, να προσβάλλουν βάσεις δεδομένων με ευρήματα ερευνών ή να παρενοχλήσουν την ομαλή πορεία των υπηρεσιών παροχής υγειονομικής περίθαλψης, θέτοντας σε κίνδυνο τις ζωές ασθενών και ανακτώντας πολύτιμα, ευαίσθητα δεδομένα (Biggio & Roli, 2018).

5 Αντίμετρα

5.1 Αντίμετρα και η συνεισφορά τους

Όπως ήδη σημειώθηκε, οι υποστηριζόμενες από TN επιθέσεις, είναι ικανές να προκαλέσουν σοβαρές οικονομικές απώλειες σε άτομα, οργανισμούς και κράτη. Οι επιχειρήσεις που θα προσβληθούν ενδέχεται να υποστούν ανεπανόρθωτες οικονομικές ζημιές, οι κυβερνήσεις να αναγκαστούν να μεταφέρουν πόρους για κάλυψη ζημιών και οι κρίσιμες υποδομές να υποστούν καθυστερήσεις που θα επηρεάσουν την παραγωγικότητά τους και τις συναλλαγές. Η ελαχιστοποίηση των κινδύνων αυτού του είδους, προϋποθέτει από τους οργανισμούς να επενδύσουν σε ιδιαίτερα ισχυρά αντίμετρα, συμπεριλαμβανομένων των εργαλείων εντοπισμού και αποτροπής, τα οποία υποστηρίζονται από συστήματα TN, προκειμένου να διασφαλίσουν ότι παρέχεται ενημέρωση σχετικά με τις αναδυόμενες απειλές και ότι αυτές μπορούν να αντιμετωπιστούν με επιτυχία, στην αρχή τους (Comiter, 2019).

Η εθνική ασφάλεια θεωρείται ο νούμερο ένα χώρος στον οποίο τα κράτη οφείλουν να εγκαθιστούν αξιόπιστα συστήματα κυβερνο-προστασίας από επιθέσεις με TN. Κάθε χώρα θα πρέπει να αναπτύσσει ισχυρές στρατηγικές άμυνας έναντι των επιθέσεων που εξαπολύουν οι χρήστες με συστήματα TN, να επενδύουν σε ανθεκτικότητα στον κυβερνοχώρο και να θεσπίζουν διεθνείς συνεργασίες προκειμένου να ανταπεξέρχονται αποτελεσματικά σε τέτοιες απειλές (El Hadi, 2023).

Διαπιστώνεται επιπλέον, ότι η ιδιωτικότητα των πολιτών βρίσκεται σε κίνδυνο εξαιτίας των επιθέσεων από συστήματα καθοδηγούμενα από TN. Η προστασία των προσωπικών δεδομένων προϋποθέτει τη χρήση αξιόπιστων και ασφαλών λογισμικών που εφαρμόζουν ισχυρή κρυπτογράφηση, αυξάνουν την ευαισθητοποίηση των χρηστών και προάγουν την πρόληψη ενάντια σε εξελιγμένες τεχνικές επιθέσεων TN (Duddu, 2018) ή την αντιμετώπισή τους. Σε κάθε περίπτωση, η χρήση της TN σε επιθέσεις εναντίον ατόμων, εταιρειών και κρατικών φορέων, έχει εγείρει ανησυχίες και ηθικά ερωτήματα. Η ανάπτυξη και εφαρμογή συστημάτων TN θα πρέπει να άπτεται κάποιων ηθικών κατευθυντήριων γραμμών και να ορίζεται από το σύνολο ορισμένων βασικών αρχών, ώστε να διασφαλίζεται πως οι τεχνολογίες αυτές δεν υπόκεινται σε κατάχρηση με σκοπό επιζήμιο προς άλλους. Εν κατακλείδι, η διευθέτηση των προκλήσεων που θέτουν οι κυβερνο-επιθέσεις που υποστηρίζονται από TN, απαιτεί συνεργασία ανάμεσα σε κυβερνήσεις, μεγάλες εταιρείες του τεχνολογικού κλάδου και ειδικούς του χώρου της κυβερνο-ασφάλειας. Οι κανονισμοί

και οι διεθνείς συμφωνίες πιθανόν να αποτελούν ένα αξιόλογο μέτρο καθιέρωσης προτύπων για την ορθή χρήση της ΤΝ στην κυβερνο-ασφάλεια, ισχυροποιώντας την προς όφελος όλων των νόμιμα ενεργών χρηστών (Li, 2018).

Οι υποενότητες που ακολουθούν, διακρίνουν τα αντίμετρα σε αυτά όπου δε χρειάζεται να λαμβάνει απαραίτητα χώρα, αξιοποίηση της ΤΝ και σε αυτά που υποστηρίζονται άμεσα από την ΤΝ.

5.2 Ορισμός Αντιμέτρων της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια

Όταν γίνεται λόγος για τα αντίμετρα (countermeasures) της ΤΝ σε ζητήματα κυβερνο-ασφάλειας (ουσιαστικά υπάρχει η ΤΝ που αντιμετωπίζει κλασικά ζητήματα κυβερνοασφάλειας, καθώς και η ΤΝ που αντιμετωπίζει ζητήματα τα οποία προκύπτουν από τη χρήση της ΤΝ με κακόβουλο τρόπο), στη συζήτηση υπεισέρχονται οι τεχνικές, οι στρατηγικές και τα εργαλεία που χρησιμοποιούνται για την άμυνα ενάντια σε απειλές που προέρχονται από επιθέσεις στο περιβάλλον του διαδικτύου και καθοδηγούνται από συστήματα που βασίζονται στην ΤΝ (αγγλικά: AI-driven cyberattacks). Όσο, λοιπόν, η χρήση της ΤΝ και της μηχανικής μάθησης αυξάνονται, από τα εγκληματικά στοιχεία που δραστηριοποιούνται στο διαδίκτυο (ή φορείς που αποσκοπούν στην παράνομη πρόσβαση σε ευαίσθητα ατομικά, επιχειρηματικά ή εθνικά δεδομένα) θεωρείται σημαντικό να δημιουργούνται διαρκώς καινούρια αντίμετρα από τους επαγγελματίες της κυβερνο-ασφάλειας βασισμένα στην ΤΝ, έτσι ώστε να προστατεύονται τα συστήματα, οι υποδομές και τα δεδομένα, από τις, εκ της ΤΝ καθοδηγούμενες, επιθέσεις (Zeng, 2022).

Η καταγραφή και κατανόηση των πτυχών των αντιμέτρων, αποτελεί την αρχή της εφαρμογής τους. Έτσι, ξεκινώντας με τα βασιζόμενα στη ΤΝ συστήματα εντοπισμού ανωμαλιών, αυτά είναι ικανά να αναγνωρίσουν ασυνήθιστα μοτίβα ή συμπεριφορές, οδηγούμενα/ες από την ΤΝ στην κίνηση του δικτύου, τη δραστηριότητα χρηστών και τη λειτουργία του συστήματος. Ανωμαλίες αυτού του είδους ενδέχεται να υποδεικνύουν μια κυβερνο-επίθεση ή μια «ρωγμή» στα αμυντικά συστήματα, με την ΤΝ να χρησιμοποιείται εναντίον του εαυτού της (ως ένα αντίμετρο ΤΝ εναντίον επίθεσης ΤΝ), βοηθώντας αρχικά στον εντοπισμό και έπειτα στην ειδοποίηση, τη στιγμή που συμβαίνει η οδηγούμενη από την ΤΝ ενέργεια (in real-time) (Fei & Zheng, 2022).

Η ΤΝ, όμως, μπορεί να προχωρήσει και σε συμπεριφορική ανάλυση (behavioral analysis), αξιολογώντας τις ενέργειες των χρηστών και εγκαθιδρύοντας τις βάσεις για τη λεγόμενη “φυσιολογική συμπεριφορά”, αναλόγως της εταιρείας ή του φορέα για τον οποίο λειτουργεί. Πιο συγκεκριμένα, διαφορετική “φυσιολογική συμπεριφορά” λαμβάνεται υπόψιν ότι επιδεικνύει ένας χρήστης πληροφοριών μιας εμπορικής ιστοσελίδας και διαφορετική ένας χρήστης εθνικού οργανισμού που συνδέεται με την άμυνα του κράτους. Για αυτήν την ανάλυση η ΤΝ θα μπορούσε να χρησιμοποιήσει τα συστήματα ανίχνευσης ανωμαλιών (Anomaly Detection Systems) τα οποία επικεντρώνονται στο να εντοπίζουν και να ξεχωρίζουν τις τεχνητές από τις φυσιολογικές συμπεριφορές με βάση αυτά που έχουν μάθει, έτσι όταν εμφανιστεί μια ασυνήθιστη συμπεριφορά να θεωρείται ανεπιθύμητη. Οι παρεκκλίσεις από την προκαθορισμένη αυτή συμπεριφορά (παραβιάσεις λογαριασμών ή εσωτερικοί κίνδυνοι), θα μπορούσαν να θεωρηθούν ως “*εν δυνάμει απειλές*” (Zhangyi & Gang, 2022).

Ενέργειες εντοπισμού, όμως, μπορούν να αντιμετωπίζουν και επιθέσεις που βασίζονται σε μηχανική μάθηση. Τα συστήματα εντοπισμού απειλών από ΤΝ, χρησιμοποιούν αλγορίθμους μηχανικής μάθησης προκειμένου να αναλύσουν μεγάλα σύνολα δεδομένων και να αναγνωρίσουν τόσο γνωστές απειλές όσο και άγνωστες. Αυτού του είδους τα συστήματα, είναι ικανά να μαθαίνουν καινούρια μοτίβα επιθέσεων (new attack patterns) και να προσαρμόζονται σε αυτά, αντιμετωπίζοντάς τα. Ομοίως, μπορεί να χρησιμοποιείται και η “αντίστροφη μηχανική μάθηση” (adversarial machine learning) για την ανάπτυξη μοντέλων ΤΝ αρκετά ισχυρών ώστε να αντιστέκονται σε επιθέσεις που έχουν σχεδιαστεί είτε να τα χειραγωγούν είτε να παραπλανούν. Τέτοιου είδους επιθέσεις περιλαμβάνουν τη «μόλυνση» από κάποιον ιό και τις επιθέσεις διαφυγής (evasion attacks), δηλαδή μεθόδους που εφαρμόζονται από κακόβουλους χρήστες προκειμένου να παρακάμψουν μηχανισμούς ασφάλειας ώστε να αποφύγουν εργαλεία ανάλυσης πιθανών κινδύνων ή εντοπισμού τους. Συμπληρωματικά, τα αντίστροφα δημιουργικά δίκτυα (Generative Adversarial Networks, GANs) ενώ μπορούν να χρησιμοποιηθούν για να δημιουργήσουν ψευδές περιεχόμενο, μπορούν επίσης να αξιοποιηθούν για αμυντικό σκοπό, προκειμένου να δημιουργήσουν συνθετικά δεδομένα για την εκπαίδευση μοντέλων ΤΝ και τη βελτίωση της ανθεκτικότητάς τους (de Azambuja et al., 2023).

Η επεξεργασία φυσικής γλώσσας για τη συγκέντρωση πληροφοριών σχετικών με απειλές, αποτελεί ένα ακόμη αντίμετρο, αφού μπορεί κάποιος να εκμεταλλευτεί όλη αυτή τη γνώση ώστε να αναγνωρίζει απειλές στο σύστημα. Οι τεχνικές επεξεργασίας φυσικής γλώσσας (natural language processing) που υποστηρίζονται από την ΤΝ, μπορούν να αναλύουν και να εξάγουν πολύτιμη γνώση για απειλές που προέρχονται από μεγάλους όγκους δεδομένων ή μη-δομημένες πηγές

προέλευσης, συμπεριλαμβανομένων τον Παγκόσμιο Ιστό (World Wide Web) περιέχοντας κείμενα από δημόσια άρθρα, το περιεχόμενο των οποίων μπορεί να χρησιμοποιηθεί για την εκπαίδευση αυτών. Επιπλέον, βάσεις δεδομένων παρέχουν εταιρικά, επιστημονικά και κυβερνητικά δεδομένα, καταγραφές συνομιλιών που προέρχονται από emails και συνομιλίες μέσω ιστοσελίδων. Ακόμη, δημόσιες πηγές δεδομένων (Wikipedia, Project Gutenberg) παρέχουν δωρεάν κείμενα για την λειτουργία των μοντέλων (Zhangyi & Gang, 2022).

Για την ασφάλεια των τελικών σημείων (endpoints), οι λύσεις που προτείνονται φαίνεται πως αξιοποιούν ολοένα και περισσότερο την TN, προκειμένου να εντοπίζονται και να αντιμετωπίζονται απειλές σε επίπεδο συσκευών. Η TN μπορεί και αναγνωρίζει κακόβουλες ενέργειες επεξεργασίας καθώς και κακόβουλα αρχεία, φακέλους και συμπεριφορές σε μεμονωμένα τελικά σημεία. Η καθοδηγούμενη από την TN ανάλυση δεδομένων ασφαλείας (security analytics) αποτελεί μια διαδικασία ανάλυσης μέσω TN, μεγάλων όγκων δεδομένων σε πραγματικό χρόνο, με σκοπό να υποστηρίζεται το έργο ομάδων που εργάζονται για την ασφάλεια ώστε να αναγνωρίζονται οι απειλές πιο γρήγορα και να αντιμετωπίζονται περισσότερο αποτελεσματικά (Zhangyi & Gang, 2022). Οι ομάδες ασφαλείας μπορούν επίσης να αξιοποιήσουν την TN, με σκοπό να εντοπίσει εκ των προτέρων τις απειλές εντός του δικτύου. Η TN μπορεί να αναγνωρίσει ύποπτες δραστηριότητες και βοηθά αναλυτές να διερευνήσουν δυνητικές απειλές περισσότερο αποτελεσματικά (Abdul-Ghani & Konstantas, 2019).

Μια πρακτική που αποτελεί αντίμετρο σε κυβερνο-επιθέσεις υποστηριζόμενες από TN, είναι η ενορχήστρωση και ο αυτοματισμός της ασφάλειας. Η TN μπορεί να ενσωματωθεί σε πλατφόρμες ενορχήστρωσης και αυτοματισμών ασφαλείας (όπως οι IBM Resilient ή η McAfee MVISION EDR), αφού βοηθά στην αναγνώριση και θέσπιση προτεραιοτήτων όσον αφορά τις ειδοποιήσεις, στην εκτέλεση αυτόματης ανταπόκρισης και στη μείωση του χρόνου απόκρισης. Αντίμετρο, όμως, αποτελούν, επίσης, η συνεχής αναβάθμιση και η παρακολούθηση των αλγορίθμων και των μοντέλων TN προκειμένου να είναι σε θέση να αναβαθμίζονται για την πιο αποτελεσματική αντιμετώπιση εξελισσόμενων απειλών. Οι εταιρείες, επομένως, οφείλουν να τηρούν μια τακτική πορεία αναβαθμίσεων των γενικής χρήσης λογισμικών τους, για λόγους διόρθωσης των όποιων σφαλμάτων και των πιθανών τρωτών σημείων ασφαλείας. Ενώ η TN είναι σαφέστατα ένα ισχυρό εργαλείο διασφάλισης της κυβερνο-ασφάλειας, η ανθρώπινη εξειδίκευση παραμένει καθοριστική. Οι αναλυτές μπορούν να παρέχουν περιεχόμενο, να λαμβάνουν αποφάσεις και να προσαρμόζονται τόσο σε καινούριες όσο και σε απρόβλεπτες απειλές, τις οποίες η TN ίσως δυσκολευτεί να αντιμετωπίσει (de Azambuja et al., 2023).

5.3 Αντίμετρα χωρίς αξιοποίηση TN

Η αντιμετώπιση των υποστηριζόμενων από την TN κυβερνο-επιθέσεων, απαιτεί ένα συνδυασμό προληπτικών μέτρων και στρατηγικών μέσω των οποίων εντοπίζονται και περιορίζονται σε πρώτο χρόνο οι επιθέσεις, ενώ στη συνέχεια αντιμετωπίζονται οι όποιες επιπτώσεις τους, όπως για παράδειγμα μια ενδεχόμενη απώλεια δεδομένων, κάποιες υλικές βλάβες ή οτιδήποτε άλλο θα μπορούσαν να επιφέρουν (King et al., 2020). Σε κάθε χρονική στιγμή, λοιπόν, είναι καλό να υπάρχει ευαισθητοποίηση του κοινού για την κυβερνο-ασφάλεια (cyber-security awareness) και να φροντίζουν τόσο τα άτομα κατ' ιδίαν όσο και οι οργανισμοί, να εκπαιδεύονται (cyber-security training) στον εντοπισμό και την αποτελεσματική αντιμετώπιση των κυβερνο-απειλών. Η εκπαίδευση των ατόμων (μεμονωμένα) και των εργαζομένων, γύρω από τους πιθανούς κινδύνους με τους οποίους συνδέονται οι κυβερνο-επιθέσεις, συμβάλλει καθοριστικά στον έγκαιρο εντοπισμό των απειλών και των προσπαθειών εξαπάτησης, ενώ θεωρείται ικανή συνθήκη τόσο για την έγκαιρη αντιμετώπιση των προβλημάτων/συνεπειών που θα προκύψουν όσο και για τη μείωση της σοβαρότητας των επιπτώσεων (DHH, 2016).

Ένα αντίμετρο με ιδιαίτερα μεγάλη βαρύτητα, είναι η ισχυρή αυθεντικοποίηση και ο έλεγχος πρόσβασης (robust authentication & access control). Η εφαρμογή μεθόδων αυθεντικοποίησης, οι οποίες χαρακτηρίζονται από την αξιοποίηση πολλών παραγόντων (multi-factor authentication) αλλά και η χρήση βιομετρικών στοιχείων (biometric elements) των χρηστών, μπορούν να εμποδίσουν περιπτώσεις μη-εξουσιοδοτημένης πρόσβασης σε δεδομένα, συστήματα και υποδομές. Από την πλευρά του ελέγχου πρόσβασης, μπορεί να γίνεται έλεγχος της τοποθεσίας και της συσκευής που χρησιμοποιείται για είσοδο, αποκλείοντας άγνωστες συσκευές ή μη-δηλωμένες τοποθεσίες – με άλλα λόγια να χρησιμοποιείται το μοντέλο του ελέγχου πρόσβασης με βάση τις ιδιότητες (attribute-based access control – ABAC), ιδιαίτερα τις περιβαλλοντικές. Μπορεί επίσης να αποστέλλεται σε ένα διαχειριστή, ειδοποίηση εισόδου στο σύστημα, σε ασυνήθιστη ώρα, επιτρέποντάς του να απαγορεύσει ακόμη και εξ' αποστάσεως, την πρόσβαση, σε περίπτωση που προκύψει πιθανός κίνδυνος (Fei & Zheng, 2022).

Η τμηματοποίηση του δικτύου (network segmentation) θεωρείται ένα ακόμη αποτελεσματικό αντίμετρο. Ο λόγος έγκειται στο γεγονός ότι απομονώνονται μέρη του συστήματος με πολύ μεγάλη σημασία, καθώς και ευαίσθητες πληροφορίες, από ενδεχόμενες απειλές, περιορίζοντας έτσι τις δυνατότητες παράκαμψης των δυνατοτήτων ασφαλείας του συστήματος, από τους επιτιθεμένους. Επιπροσθέτως, οι έλεγχοι διείσδυσης (penetration testing) αποτελούν ενέργειες που καλό είναι να διεξάγονται σε τακτική βάση, ενώ θα πρέπει να ενσωματώνεται στην ανάπτυξη ασφαλών λογισμικών και συστημάτων, δεδομένου ότι αποτελεί σημαντικό χαρακτηριστικό του

κύκλου ζωής τους. Με τον τρόπο αυτό αξιολογείται το επίπεδο ασφάλειας και εντοπίζονται πιθανά καινούρια τρωτά σημεία στο σύστημα, τα οποία οι επιτιθέμενοι που αξιοποιούν κατά κόρον την TN μπορούν να εντοπίσουν και να εκμεταλλευτούν με μεγαλύτερη ευκολία (Security Standards Council, 2017).

Επιπροσθέτως, η εικονικοποίηση του δικτύου (network virtualization) και τα software-defined networks (SDN) (δίκτυα οριζόμενα από το λογισμικό) μπορούν να αποτελέσουν ισχυρά εργαλεία στην ενίσχυση της κυβερνο-ασφάλειας, περιορίζοντας την επίπτωση των κυβερνο-επιθέσεων. Πιο συγκεκριμένα, η πρώτη δημιουργεί διάφορα ψηφιακά δίκτυα σε μια κοινή δομή, επιτρέποντας την απομόνωση (isolation), με αποτέλεσμα εάν «μολυνθεί» κάποιο από αυτά, τα υπόλοιπα να μην μολύνονται. Επιπλέον, τα εικονικά δίκτυα (virtualized networks) συνοδεύονται από ενσωματωμένα εργαλεία παρακολούθησης ασυνήθιστων μοτίβων ή δραστηριοτήτων, οι πολιτικές ασφάλειας που εφαρμόζονται σε αυτά επιτρέπουν πιο εξονυχιστικό έλεγχο ενώ η αυτοματοποίηση των καθηκόντων μειώνει την πιθανότητα ανθρωπίνου λάθους (Bays et al., 2015). Από την άλλη, τα SDN επιτρέπουν στους διαχειριστές να εφαρμόσουν πολιτικές τμηματοποίησης, περιορίζοντας πλευρικές κινήσεις εντός του δικτύου ενώ καθιστούν πιο κεντρικά ελεγχόμενη τη διαχείριση του, παρέχοντας επαρκή οπτική κάθε περιοχής του, άρα βελτιώνοντας την παρακολούθηση όλων των ενεργειών. Επιπλέον, επιτρέπουν την προσαρμογή των πολιτικών ασφαλείας βάσει των υφιστάμενων απειλών, διασφαλίζοντας κατευθυνόμενη και έγκαιρη ανταπόκριση έναντί τους. Οι αυτοματισμοί μάλιστα διαδραματίζουν ρόλο-κλειδί στα SDN, με τις πολιτικές ασφαλείας να εφαρμόζονται αυτόματα, σύμφωνα με πληροφορίες που λαμβάνουν σε πραγματικό χρόνο, μειώνοντας έτσι το χρόνο ανταπόκρισης (Shu et al., 2016).

Δύο ακόμη αξιοσημείωτα αντίμετρα, είναι η υιοθέτηση ενός μοντέλου προστασίας τύπου “zero-trust cyber security model”, σύμφωνα με το οποίο η εμπιστοσύνη σε διάφορες ενέργειες εισόδου δε θεωρείται ποτέ δεδομένη και απαιτείται πάντοτε και από οποιονδήποτε, επιβεβαίωση για την πρόσβασή του σε συγκεκριμένες πληροφορίες ή/και μέρη του συστήματος (World Economic Forum, 2022). Συμπληρωματικά σε αυτό, συστήνεται επίσης η χρήση “διαστρωματοποιημένης αμυντικής προσέγγισης” (layered defense approach), αξιοποιώντας ένα συνδυασμό τεχνολογιών προστασίας, δηλαδή τειχών (firewalls), IDS/IPS, μέτρα προστασίας τελικών σημείων (endpoint protection) κ.ά. Εν γένει, θα μπορούσε κάποιος να υποστηρίξει πως πρέπει να υιοθετείται μια διαστρωματοποιημένη αρχιτεκτονική ασφαλείας, με τη χρήση κατάλληλων μηχανισμών και αντιμέτρων ασφαλείας στο κάθε στρώμα. Επίσης, θα πρέπει να εξασφαλίζεται η διαλειτουργικότητα των αντιμέτρων, καθώς και ο διαμοιρασμός πληροφορίας & δεδομένων για περισσότερη ακρίβεια στον εντοπισμό και την αντιμετώπιση των συμβάντων

κυβερνο-ασφάλειας. Έτσι, δημιουργούνται πολλαπλά εμπόδια σε οποιαδήποτε προσπάθεια παραβίασης από κυβερνο-επιθέσεις (Atkinson & Kress, 2023).

Επιπλέον εργαλεία για την ανίχνευση τρωτών σημείων και γεγονότων ασφαλείας, αποτελούν η καταγραφή (logging) και η ψηφιακή εγκληματολογία (digital forensics), ιδιαίτερα για τις περιπτώσεις επιθέσεων τύπου zero-day, μιας και η δοκιμή διείσδυσης (penetration testing) συνήθως καλύπτει τις περιπτώσεις επιθέσεων που στοχεύουν στον εντοπισμό γνωστών τρωτών σημείων. Το μεν logging περιλαμβάνει την καταγραφή των περιστατικών (αλλαγές, είσοδο χρηστών, περιστατικά ασφαλείας κ.ά.) σε ένα σύστημα ή δίκτυο, με σκοπό τη λεπτομερή παρουσίαση κάθε δραστηριότητας, γεγονός πολύτιμο για τη διερεύνηση περιστατικών απειλής ή διείσδυσης. Η δε ψηφιακή εγκληματολογία περιλαμβάνει τη διερεύνηση και ανάλυση συσκευών και δεδομένων, προκειμένου να αποκαλύπτονται περιστατικά κυβερνο-εγκλήματος ή άλλες παράνομες δραστηριότητες (Choudhary & Sharma, 2016).

Ιδιαίτερη επίσης σημασία θα πρέπει να δίνεται στην προστασία των μοντέλων TN (AI models protection) που χρησιμοποιούνται κατά την κανονική λειτουργία ενός συστήματος. Αυτά συστήνεται να προστατεύονται από τις αντιπαραθετικές επιθέσεις (adversarial attacks) και τις παραβιάσεις, αφενός με την εφαρμογή ασφαλών πρωτοκόλλων επικοινωνίας που επιβάλλουν την αυθεντικοποίηση και των δύο μερών και αφετέρου με τη χρήση αυστηρών πρωτοκόλλων ελέγχου πρόσβασης σε πληροφορίες (μόνο κατόπιν εξουσιοδότησης), ώστε ο επιτιθέμενος να μην είναι σε θέση να καταχωρήσει λάθος πληροφορίες, προκειμένου να ελαχιστοποιείται η έκθεση σε κινδύνους ή αυτοί να εντοπίζονται ή αποτρέπονται εγκαίρως (Neyigapula, 2023). Το ίδιο ισχύει βέβαια και για την κρυπτογράφηση μεμονωμένων ευαίσθητων πληροφοριών ή/και δεδομένων, τόσο όσων παραμένουν σταθερά (at rest) όσο και αυτών που μεταβιβάζονται/μετακινούνται (in transit), ώστε να προστατεύονται αποτελεσματικότερα από επιθέσεις καθοδηγούμενες από την TN (Lauter, 2021). Εν συντομία, αυτό που επιτυγχάνεται με την κρυπτογράφηση, είναι το γεγονός ότι διασφαλίζεται η εμπιστευτικότητα, η ακεραιότητα και η αυθεντικότητα των δεδομένων, προστατεύονται τα κανάλια επικοινωνίας και μετριάζονται οι πιθανοί κίνδυνοι που σχετίζονται με τις κυβερνο-επιθέσεις, συμπεριλαμβανομένων των adversarial attacks και των παραβιάσεων δεδομένων. Παρόλα αυτά, είναι σημαντικό η κρυπτογράφηση να εφαρμόζεται σωστά, τα κλειδιά κρυπτογράφησης να αποτελούν αντικείμενο προσεκτικής διαχείρισης και να λαμβάνουν χώρα τακτικές αναβαθμίσεις των πρωτοκόλλων κρυπτογράφησης & επικοινωνίας, έτσι ώστε τα συστήματα και τα λογισμικά να βρίσκονται μπροστά από τις αναδυόμενες απειλές (Pan et al., 2022). Τα αυστηρά πρωτόκολλα ασφαλείας με τη σειρά τους επιβάλλουν αυστηρούς ελέγχους πρόσβασης, αναγκάζοντας τους χρήστες να αποδεικνύουν την ταυτότητά τους μέσα από βιομετρικά στοιχεία,

κωδικούς ή εξουσιοδότηση δύο (ή παραπάνω) παραγόντων, ενώ μετά την εξουσιοδότηση, οι χρήστες λαμβάνουν δικαιώματα μόνο βάσει των ρόλων και των αρμοδιοτήτων τους (για έναν οργανισμό), διασφαλίζοντας ότι δε θα εισέλθουν σε πόρους που δεν τους είναι απαραίτητοι για τη διεκπεραίωση των καθηκόντων τους (Neyigapula, 2023).

Ένα ακόμη αντίμετρο είναι τα “σχέδια ανταπόκρισης” (incident response plan). Αυτό που συστήνεται στην ουσία, είναι ο σχεδιασμός ενός σχεδίου στο οποίο προσδιορίζεται με λεπτομέρεια και ακρίβεια η ανταπόκριση σε κάθε περίπτωση εντοπισμού απειλής ή προκληθείσας βλάβης. Η ανταπόκριση αυτή περιλαμβάνει συγκεκριμένες ενέργειες που θα λάβουν χώρα με το που γίνει αντιληπτή κάποια κυβερνο-επίθεση (Office of Information Security, 2023). Σε κάθε περίπτωση πάντως, τέτοιου είδους σχέδια θα πρέπει να ανανεώνονται τακτικά (regular patching & updates) (τόσο για την αντιμετώπιση νέων επιθέσεων/απειλών όσο και για την ενσωμάτωση νέων αντιμέτρων), ενώ το ίδιο θα πρέπει να γίνεται στα λογισμικά και τα συστήματα των μεμονωμένων χρηστών ή οργανισμών (ιδιαίτερα αυτών που αντιστοιχούν σε αντίμετρα και μηχανισμούς ασφάλειας), προκειμένου να καταφέρνουν να διαχειρίζονται εντοπισθέντα τρωτά σημεία και να μειώνουν την έκταση των επιθέσεων (King et al., 2020).

Δεδομένης της αύξησης της αξιοποίησης της υπολογιστικής νέφους (cloud computing), θεωρείται χρήσιμο να λαμβάνονται μέτρα προστασίας και για αυτή την κεντριοποιημένη υποδομή που προσφέρει διάφορα είδη υπηρεσιών και εφαρμογών. Άρα τόσο αυτή όσο οι υπηρεσίες και εφαρμογές της, καθώς και τα υποκείμενα δεδομένα των χρηστών, θα πρέπει να προστατευθούν. Ειδικότερα, ο πάροχος της υποδομής θα πρέπει να εφαρμόζει κατάλληλη αρχιτεκτονική ασφάλειας που να περιλαμβάνει τη χρήση κατάλληλων μηχανισμών και υπηρεσιών ασφάλειας, να ακολουθεί βέλτιστες πρακτικές και πρότυπα ασφάλειας και να διατηρεί & ανανεώνει κατάλληλα σχέδια ανταπόκρισης. Από την άλλη μεριά, εάν ένα άτομο ή ένας οργανισμός/επιχείρηση χρησιμοποιεί αυτή την τεχνολογική καινοτομία, δηλ. αποτελώντας ένα πελάτη των υπηρεσιών νέφους, θα πρέπει να εξετάζει και να επιλέγει τις όποιες προτάσεις ασφάλειας υπολογιστικής νέφους συστήνονται από τους παρόχους, προκειμένου να διασφαλίσει τις ευαίσθητες πληροφορίες και τα περιουσιακά στοιχεία (assets) που διαθέτει. Για το λόγο αυτό, θα πρέπει να αξιολογεί τις πρακτικές ασφαλείας των παρόχων και να εξετάζει κατά πόσο αυτές μπορεί να αποτελέσουν ένα ενδεχόμενο τρωτό σημείο σε περίπτωση κυβερνο-επιθέσεων. Επομένως, πρώτο βήμα είναι η επιλογή του πιο ασφαλούς νέφους και έπειτα η εφαρμογή βέλτιστων πρακτικών ασφαλούς χρήσης του καθώς και η επιλογή έξτρα υπηρεσιών/προτάσεων ασφάλειας για να διασφαλιστούν οι υπηρεσίες, εφαρμογές και δεδομένα που μεταναστεύει ένας οργανισμός στο νέφος (Tissir, El Kafhali & Aboutabit, 2021). Συμπληρωματικά, μάλιστα, κρίνεται σκόπιμο ο πελάτης να δημιουργεί back-ups (αντίγραφα

ασφαλείας) για τα κρίσιμα δεδομένα και τις ευαίσθητες πληροφορίες (χρησιμοποιώντας αντίστοιχες, κατάλληλα ελεγμένες υπηρεσίες που προσφέρει ο πάροχος), ενώ συστήνεται να ελέγχει ανά διαστήματα τις διαδικασίες επαναφοράς (ή αποκατάστασης) μετά από ενδεχόμενη πρόκληση βλαβών (restoration processes), προκειμένου να διασφαλίσει μια γρήγορη αποκατάσταση έπειτα από μια σοβαρή κυβερνο-επίθεση με δυσάρεστες επιπτώσεις (King et al., 2020).

Ένα ακόμη από τα αντίμετρα που προτείνονται, είναι οι χρήστες του διαδικτύου να συμμορφώνονται προς τους κανονισμούς (όπως η τακτική αναβάθμιση των λογισμικών ασφαλείας τους και τη χρήση εγκεκριμένων λογισμικών, την εγρήγορση για περιπτώσεις επιθέσεων phishing, τη συχνή ανανέωση των διαπιστευτηρίων ασφαλείας, την ασφάλεια του ασύρματου δικτύου μέσω κρυπτογράφησης WPA3, την ασφαλή περιήγηση στο διαδίκτυο με αποφυγή ύποπτων ιστοσελίδων, τον περιορισμό της πρόσβασης άλλων στο σύστημα κ.ά.) που διέπουν την ασφαλή αξιοποίηση των δυνατοτήτων του. Οι χρήστες οφείλουν να διασφαλίζουν ότι οι πρακτικές ασφάλειας που εφαρμόζουν, ευθυγραμμίζονται προς διάφορες κανονιστικές ρυθμίσεις και πρότυπα (regulations & standards), με στόχο την επίτευξη υψηλότερων επιπέδων ασφαλείας. Επιπροσθέτως, οι χρήστες μπορούν να πραγματοποιούν αναλύσεις συμπεριφοράς (behavior analytics), έτσι ώστε να βρίσκονται σε καλύτερη θέση όσον αφορά τον εντοπισμό ύποπτων δραστηριοτήτων, οι οποίες ενδέχεται να αποτελούν ένδειξη μιας υποστηριζόμενης από TN επίθεσης ή να προδίδουν ενδεχόμενη ύπαρξη εσωτερικής απειλής στο σύστημα, δηλαδή περιστατικό παραβίασης της υφιστάμενης ασφάλειας και διείσδυσης (Antunes et al., 2021).

Ένα πολύ σημαντικό αντίμετρο, θεωρείται η ενσωμάτωση πρακτικών προστασίας και ασφάλειας στην ανάπτυξη λογισμικών και τη διαδικασία εγκατάστασης/εφαρμογής τους, έτσι ώστε να είναι σε θέση να εντοπίζουν και να διαχειρίζονται εγκαίρως τα όποια τρωτά σημεία (που είτε είναι εγγενή ή οφείλονται στο περιβάλλον εκτέλεσής τους ή στη λανθασμένη διαμόρφωσή τους) κατά τον κύκλο ανάπτυξής τους (DevSecOps). Επιπροσθέτως, με τις πολιτικές προληπτικής απενεργοποίησης λογαριασμών, αποτρέπεται η μη-εξουσιοδοτημένη πρόσβαση σε συστήματα και ευαίσθητες πληροφορίες ή μπορεί να απενεργοποιείται ο λογαριασμός ενός χρήστη σε προκαθορισμένο χρονικό σημείο (Valdes et al., 2023).

Ιδιαίτερης σημασίας ενέργεια θεωρείται ο διαμοιρασμός πληροφοριών σχετικών με απειλές (threat intelligence sharing). Η συμμετοχή τόσο των ατόμων μεμονωμένα όσο και των οργανισμών, σε τέτοιου είδους πρωτοβουλίες διαμοιρασμού, τους βοηθά να παραμένουν ενημερωμένοι γύρω από τις απειλές που αναδύονται στον κυβερνοχώρο, ενώ πληροφορούνται επίσης για τις απόψεις της κοινότητας (forums στα οποία συμμετέχουν) και τους τρόπους αντιμετώπισης (Johnson et al.,

2016). Επιπροσθέτως, υπάρχουν λογισμικά δημιουργίας ασφαλών κωδικών (τυχαίες συμβολοσειρές, κρυπτογραφικά κλειδιά και κωδικούς μιας χρήσης), δηλαδή εργαλεία που διασφαλίζουν την αυθεντικοποίηση των χρηστών, διασφαλίζουν τις επικοινωνίες και προστατεύουν τα δεδομένα (Rabii et al., 2020).

5.4 Αντίμετρα με αξιοποίηση TN

Σε αυτή την κατηγορία ανήκει ο υποστηριζόμενος από την TN, εντοπισμός απειλών (AI-enabled threat detection) ως γενική κατηγορία, στην οποία θα μπορούσαν κάλλιστα να ενσωματωθούν ενέργειες όπως ο έλεγχος διείσδυσης, η προστασία δεδομένων, η ενίσχυση της αξιοπιστίας των συστημάτων (για παράδειγμα, με τη χρήση TN για τη δημιουργία πιο ασφαλούς λογισμικού, τη βελτίωση ακριβείας σε μηχανισμούς αυθεντικοποίησης και ελέγχου πρόσβασης), η αυτόματη κυβερνο-ασφάλεια κ.ά. Η αξιοποίηση της TN για προληπτικό, αποτρεπτικό και αμυντικό σκοπό, καθώς και των εργαλείων που βασίζονται στη μηχανική μάθηση (machine learning-based tools), συμβάλλει στον εντοπισμό μη-φυσιολογικής δραστηριότητας στο σύστημα (και το σύνολο του δικτύου) και τον μετέπειτα μετριασμό του αντίκτυπού της. Πιο συγκεκριμένα, αποτελεί ένα μέσο ικανό να εντοπίζει ύποπτες συμπεριφορές, κινήσεις, αρχεία και οτιδήποτε εισερχόμενο (inputs) σε πραγματικό χρόνο, με αποτέλεσμα να εντοπίζονται εγκαίρως οι απειλές, ιδίως όσες επιθετικές ενέργειες καθοδηγούνται από TN (AI-driven), οι οποίες μάλιστα μπορούν πλέον να γίνουν ιδιαίτερα επικίνδυνες όσον αφορά τις δυνατότητες παραβίασης (Karimipour & Derakhshan, 2021).

Το παραπάνω, όμως, ανήκει στην ευρύτερη κατηγορία της καθοδηγούμενης από την TN, κυβερνο-προστασίας. Στην ουσία η τελευταία αφορά λύσεις κυβερνο-ασφάλειας, οι οποίες βασίζονται στην TN και είναι ικανές να προσαρμόζονται και να ανταποκρίνονται σε εξελισσόμενες απειλές περισσότερο αποτελεσματικά από τα παραδοσιακά συστήματα (ή λογισμικά). Συνεπώς, όπως η TN εξελίσσεται σε επίπεδο επιθετικών ενεργειών, μπορεί να αξιοποιηθεί και να εξελιχθεί αντίστοιχα και σε επίπεδο αμυντικών, προληπτικών και αποτρεπτικών ενεργειών, ενώ συστήνεται στους χρήστες της να ενημερώνονται για τις ηθικές και νομικές παραμέτρους που διέπουν την αξιοποίηση TN στην κυβερνο-ασφάλεια, ώστε να διασφαλίζεται πως οι πρακτικές που εφαρμόζονται ευθυγραμμίζονται με τους κανονισμούς και τα πρότυπα των περιοχών όπου δραστηριοποιούνται και των κλάδων στους οποίους εργάζονται (Dignum, 2019).

Στις επόμενες υποενότητες αναλύουμε ορισμένες σημαντικές λύσεις κυβερνο-ασφάλειας βασιζόμενες στην TN.

5.4.1 Εντοπισμός ανωμαλιών (*anomaly detection*) & αναλυτική συμπεριφοράς (*behavior analytics*)

Η ΤΝ μπορεί να αναλύει τη συμπεριφορά που σημειώνουν οι χρήστες και τη γενικότερη συμπεριφορά που παρουσιάζεται σε ένα δίκτυο, με στόχο την αναγνώριση παρεκκλίσεων από θεωρητικά φυσιολογικές συμπεριφορές. Η πρακτική αυτή συμβάλλει στο γρήγορο εντοπισμό ασυνήθιστων δραστηριοτήτων, οι οποίες μπορεί να αποτελούν ένδειξη διενέργειας κυβερνο-επιθέσεων. Συμπληρωματικά, η ΤΝ είναι ικανή να δημιουργεί προφίλ φυσιολογικής συμπεριφοράς τόσο για τους χρήστες όσο και για το σύνολο ενός δικτύου, καθιστώντας ευκολότερο τον εντοπισμό και την ανταπόκριση (*detect & respond*) απέναντι σε αφύσικες ή κακόβουλες δραστηριότητες, σε πραγματικό χρόνο (Jeffrey, Tan & Villar, 2023). Επιπρόσθετα, υπάρχει και η λύση της “αναλυτικής συμπεριφοράς χρήστη & οντοτήτων” (*user & entity behavior analytics, UEBA*). Σε αυτή την περίπτωση, χρησιμοποιείται η ΤΝ με σκοπό να εντοπίζονται εσωτερικές απειλές (*insider threats*) και μη-φυσιολογικές συμπεριφορές χρηστών, μέσα από τη δημιουργία ενός προτύπου τυπικών ενεργειών των χρηστών (Olaniyan, Rakshit & Vajjhala, 2023).

5.4.2 Μηχανική μάθηση για εντοπισμό απειλών (*machine learning for threat detection*) & σε βάθος μάθηση για εντοπισμό εισβολής (*deep learning for intrusion detection*)

Η μηχανική μάθηση είναι αναμενόμενο - λόγω των δυνατοτήτων που τη χαρακτηρίζουν - να θεωρείται πως μπορεί να συνδράμει ως ένα αποτελεσματικό αντίμετρο στις κυβερνο-επιθέσεις. Η αξιοποίησή της μέσα από κατάλληλα υπολογιστικά μοντέλα, τα οποία μπορούν να εντοπίσουν ή να προβλέψουν απειλές (*threat detection*), συμβάλλει καθοριστικά στη βελτίωση των επιπέδων κυβερνο-ασφάλειας, αφού τα υπολογιστικά αυτά μοντέλα είναι ικανά να αναλύουν τεράστιους όγκους δεδομένων και να αναγνωρίζουν πιθανές απειλές τόσο σύμφωνα με συμπεριφορές/ενέργειες σε πραγματικό χρόνο (*real-time behavior analysis*) όσο και σύμφωνα με δεδομένα ιστορικού (*historical data*) του συστήματος, το οποίο προστατεύουν. Αντίστοιχα, οι τεχνικές σε βάθος μάθησης (*deep learning techniques*) μπορούν να χρησιμοποιηθούν από τα συστήματα εντοπισμού εισβολών (*intrusion detection systems, IDS*) για να βελτιωθεί η ακρίβεια στον εντοπισμό των εισβολών που αποπειρώνται να πραγματοποιηθούν στο σύστημα (de Azambuja et al., 2023).

5.4.3 Εντοπισμός κακόβουλου λογισμικού (malware detection) & πρακτικών phishing (phishing detection)

Η ΤΝ θεωρείται πως μπορεί πιο εύκολα να αναγνωρίσει και να επιβεβαιώσει την ύπαρξη ενός κακόβουλου λογισμικού, συμπεριλαμβανομένων των νέων και άγνωστων έως πρόσφατα, απειλών, αναλύοντας κώδικες και μοτίβα συμπεριφοράς, με την πρακτική αυτή να βοηθά στον αποκλεισμό ή/και τον περιορισμό κακόβουλων μολύνσεων. Με τη χρήση της ΤΝ όμως, δε γίνεται ευκολότερη μόνο η αναγνώριση άγνωστων απειλών (χάριν στη μηχανική μάθηση) αλλά εντοπίζονται με μεγαλύτερη ακρίβεια οι εξειδικευμένες επιθέσεις που δεν ακολουθούν γνωστά μοτίβα, αναλύονται με μεγαλύτερη αποτελεσματικότητα (από τα παραδοσιακά αντι-ικά λογισμικά) οι συμπεριφορές χρηστών και η ανταπόκριση σε απειλές είναι ταχύτερη, ενώ μπορούν να επιδοθούν επίσης σε κυνήγι απειλών (threat hunting), λειτουργώντας προληπτικά απέναντι σε κινδύνους, αναλύοντας διαρκώς δεδομένα και μάλιστα μεγάλου όγκου. Από την άλλη, οι αλγόριθμοι ΤΝ είναι ικανοί να αναλύουν ακόμη και το περιεχόμενο των email και των ιστότοπων, προκειμένου να αναγνωρίσουν προσπάθειες phishing, κακόβουλους συνδέσμους (links) ή κακόβουλο περιεχόμενο (πχ. για ιστοσελίδα που μοιάζει με νόμιμη με παρόμοια εμφάνιση και περιεχόμενο). Επιπλέον, οι ικανότητες των αλγορίθμων φτάνουν μέχρι και την επισήμανση των κακόβουλων αποστολών email ή των domains των αντίστοιχων ιστότοπων (Dash et al., 2022).

5.4.4 Προστασία τελικών σημείων (endpoint security) & προστασία δικτύου (network security)

Τα αντίμετρα προστασίας, τα οποία στηρίζονται στην ΤΝ και απευθύνονται σε τελικά σημεία, προστατεύουν τις μεμονωμένες συσκευές (έξυπνες συσκευές, υπολογιστές κ.ά.) παρακολουθώντας και αναλύοντας τις δραστηριότητές τους (endpoint activities), εντοπίζοντας και στη συνέχεια αποτρέποντας τη λειτουργία της απειλής εντός της συσκευής. Παρόμοια πρακτική εφαρμόζεται και σε επίπεδο δικτύων, με την ΤΝ να αξιοποιείται προκειμένου να παρακολουθήσει την κυκλοφορία σε αυτά (network traffic), με σκοπό τον εντοπισμό ασυνήθιστων ή ύποπτων μοτίβων εκτός από τα γνωστά μοτίβα επιθέσεων. Οι ενέργειες της ΤΝ περιλαμβάνουν την αναγνώριση εισβολών, επικοινωνιών ανάμεσα σε κακόβουλα λογισμικά (malware communication), καθώς και διαφόρων άλλων απειλών που βασίζονται στο διαδίκτυο (network-based threats), όπως τη μη-εξουσιοδοτημένη διαμεσολάβηση στην κυκλοφορία του διαδικτύου με σκοπό τη σύλληψη και ανάλυση πακέτων δεδομένων ώστε να εκτεθούν ευαίσθητες πληροφορίες (packet sniffing) ή την

έκθεση τρωτών σημείων σε διαδικτυακές εφαρμογές με σκοπό την «έκχυση» (SQL injection) κακόβουλου κώδικα (structure query language, SQL κώδικας), δίνοντας σε επιτιθέμενους τη δυνατότητα χειραγώγησης βάσεων δεδομένων ή τη μη-εξουσιοδοτημένη παρέμβαση των επικοινωνιών του διαδικτύου με σκοπό την πρόσβαση σε ευαίσθητες πληροφορίες (eavesdropping) (Kaur & Tiwari, 2021).

5.4.5 Πληροφόρηση Ασφαλείας & Διαχείριση Περιστατικών (Security Information & Event Management, SIEM)

Πρόκειται για πλατφόρμες που συνήθως έχουν ενσωματώσει δυνατότητες TN σχετικές με ανάλυση και συσχέτιση δεδομένων από διάφορες πηγές προστασίας και ασφάλειας έναντι κυβερνο-επιθέσεων. Με τον τρόπο αυτό, οι ομάδες παρακολούθησης της διαδικτυακής ασφάλειας καθίστανται ικανές στην αναγνώριση των κινδύνων που απειλούν συστήματα, δίκτυα και δεδομένα, άρα μπορούν να ανταποκριθούν ταχύτερα και πιο αποτελεσματικά για την αντιμετώπισή τους και την ελαχιστοποίηση των πιθανών βλαβών που θα μπορούσαν να προκαλέσουν. Στα πλαίσια αυτά, μάλιστα, εντάσσεται επίσης η διαδικασία εκτίμησης και διαχείρισης των τρωτών σημείων/ευπαθειών (vulnerabilities), με τα εργαλεία που υποστηρίζονται από την TN (AI-driven tools) να είναι ικανά για σάρωση ακόμη και ολόκληρης της τεχνολογικής υποδομής ενός οργανισμού, ώστε να αναγνωρίσουν τέτοιες ευπάθειες και μετέπειτα να θέσουν προτεραιότητες όσον αφορά την επιδιόρθωσή τους. Συμπληρωματικά, επειδή τα συστήματα που υποστηρίζονται από την TN, μπορούν να επεξεργαστούν και να αναλύσουν πιο γρήγορα τους μεγάλους όγκους δεδομένων που αφορούν κυβερνο-απειλές, μπορούν ως εκ τούτου να παρέχουν ανατροφοδότηση και πληροφόρηση γύρω από αυτές τις απειλές, βοηθώντας τους οργανισμούς (και τα μεμονωμένα άτομα φυσικά) να παραμένουν ενήμεροι για τους πιο πρόσφατους κινδύνους που έχουν καταγραφεί στον κυβερνοχώρο (de Azambuja et al., 2023).

5.4.6 Ενорχήστρωση προστασίας & αυτοματισμών (Security Orchestration & Automation)

Η TN μπορεί να αξιοποιηθεί για την αυτοματοποίηση καθηκόντων ασφάλειας, τα οποία θεωρούνται συνηθισμένα και τυπικά (routine security tasks). Οι ομάδες που αναλαμβάνουν τη διατήρηση της ασφάλειας, έχουν έτσι τη δυνατότητα να επικεντρωθούν σε περισσότερο σύνθετα και ενδεχομένως κρίσιμης ζητήματα, αφού αφήνουν την TN να διεξάγει ενέργειες τις οποίες εκτελούσαν παλιότερα

οι ίδιες οι ομάδες, όχι απλά εξοικονομώντας χρόνο αλλά μειώνοντας και την πιθανότητα λάθους εξαιτίας multitasking (δηλ. της παράλληλης εκτέλεσης πολλαπλών ενεργειών). Η TN είναι ικανή να βοηθήσει στην αναγνώριση και θέσπιση προτεραιοτήτων όσον αφορά την ανταπόκριση στα συμβάντα που λαμβάνουν χώρα (incident response), αυτοματοποιώντας τη ροή των ενεργειών ανταπόκρισης και προσφέροντας λύσεις σε πραγματικό χρόνο για τα τρέχοντα περιστατικά απειλής της ασφάλειας (Vast et al., 2021).

5.4.7 TN για προστασία νέφους (cloud security) & αναλυτική ασφαλείας (security analytics)

Δεδομένου ότι ολοένα και περισσότεροι οργανισμοί μεταβαίνουν πλέον στην εποχή της υπολογιστικής νέφους (cloud computing) και των δυνατοτήτων που προσφέρει, η TN χρησιμοποιείται προκειμένου να διασφαλιστεί ότι το περιβάλλον νέφους είναι ασφαλές για την ομαλή λειτουργία των οργανισμών και των υπηρεσιών/εφαρμογών τους. Αυτό επιτυγχάνεται μέσα από τη διαρκή παρακολούθηση, τον έγκαιρο και ακριβή εντοπισμό των κινδύνων, όπως επίσης και τον έλεγχο πρόσβασης, με την προστιθέμενη συνεισφορά της TN να εντοπίζεται στη μεγαλύτερη ταχύτητα διεκπεραίωσης ενεργειών. Σε αυτή την κατεύθυνση συμβάλουν και η αναλυτική ασφαλείας, με τις πλατφόρμες αναλύσεων που υποστηρίζονται από TN να βοηθούν τις ομάδες, που αναλαμβάνουν την προστασία των δικτύων, των συστημάτων και των δεδομένων, να επεξεργαστούν πιο αποτελεσματικά τους μεγάλους όγκους πληροφορίας και να τους αναλύσουν, ανταποκρινόμενες ταχύτερα στις αναδυόμενες απειλές. Γενικότερα, ένα περιβάλλον νέφους είναι αρκετά μεγάλο, οπότε απαιτεί μια ισχυρή πλατφόρμα αναλυτικής ασφαλείας, η οποία να καλύπτει όλα τα επίπεδα αυτού του περιβάλλοντος. Επιπλέον, σε ένα πολύπλοκο σύστημα όπως το υπολογιστικό νέφος, υπάρχει ανάγκη συνδυασμού των αντιμέτρων και της διαλειτουργικότητάς τους (Valdes et al., 2023).

Πιο συγκεκριμένα, οι πάροχοι υπηρεσιών cloud προσφέρουν τη δυνατότητα διαχείρισης της ταυτότητας κάθε χρήστη και της άδειας πρόσβασης (identity & access management), διασφαλίζοντας ότι μόνο εξουσιοδοτημένα άτομα μπορούν να εισέρχονται στο σύστημα. Επιπλέον, χάριν στο μοντέλο του ελέγχου πρόσβασης με βάση τους ρόλους (role-based access control), παρέχεται άδεια σε ρόλους, παρά σε μεμονωμένους χρήστες, που σημαίνει ότι όποιος χρήστης έχει επιφορτιστεί ένα συγκεκριμένο ρόλο ανάλογα με τη δουλειά του (ισχύει κυρίως για οργανισμούς), έχει πρόσβαση, μειώνοντας έτσι τον κίνδυνο μη-εξουσιοδοτημένης πρόσβασης. Η πολυπαραγοντική αυθεντικοποίηση χρησιμοποιείται ως μέτρο ασφαλείας και εδώ, ενώ ισχύει και η «αρχή του ελαχίστου δικαιώματος» (least privilege principle), σύμφωνα με την οποία παρέχεται σε

χρήστες και συστήματα το ελάχιστο επίπεδο πρόσβασης, για να πραγματοποιήσουν τα καθήκοντά τους, ελαχιστοποιώντας έτσι την προοπτική βλαβών μέσω προσβαλλόμενου λογαριασμού (Wang & Wendy, 2019).

5.5 Συσχέτιση Αντίμετρων με τις Φάσεις Επιθέσεων που Αντιμετωπίζουν

Πίνακας 1. Συγκεντρωτικός πίνακας φάσεων επίθεσης & αντισμέτρων

Φάσεις επίθεσης	Σύντομη περιγραφή φάσης	Αποτελεσματικότερα αντίμετρα (με ή χωρίς χρήση TN)
Αναγνώριση (reconnaissance)	Αναγνωρίζονται δυνητικοί στόχοι επίθεσης	■ συχνοί έλεγχοι δικτύου & συστήματος (TN) ■ ισχυροί έλεγχοι πρόσβασης (TN) ■ πολιτικές κωδικών (δυσκολία ανάκτησής τους από άλλους) ■ παρακολούθηση & ανάλυση κυκλοφορίας δικτύου (TN)
Διήθηση (infiltration)	Ανάπτυξη λογισμικού για μη-εξουσιοδοτημένη πρόσβαση	■ τείχη προστασίας (TN) ■ σύστημα εντοπισμού & πρόληψης εισβολών (TN) ■ τμηματοποίηση δικτύου ώστε να περιορίζεται η έκταση μιας πιθανής διάρρηξης
Αναγνώριση δεδομένων (data identification)	Εντοπίζονται τα πλέον χρήσιμα δεδομένα	■ κρυπτογράφηση δεδομένων ■ χρήση ασφαλών πρωτοκόλλων επικοινωνίας με αυθεντικοποίηση και των 2 μερών της επικοινωνίας ■ ταξινόμηση δεδομένων
Παράδοση (delivery)	Μεταφορά του κακόβουλου λογισμικού στο στόχο	■ ενημέρωση χρηστών σχετικά με το phishing και την κοινωνική μηχανική ■ ανίχνευση κακόβουλων συνημμένων, σελίδων, τομέων κ.ά. ■ εφαρμογή φίλτρων email για αποτροπή προσπαθειών phishing
Δημιουργία κωδικών (password guessing)	Δημιουργούνται κωδικοί που θα επιτρέψουν πρόσβαση στο στόχο	■ πολιτικές ισχυρών κωδικών ■ πολυπαραγοντική αυθεντικοποίηση (MFA) (TN) ■ πολιτικές προληπτικής απενεργοποίησης λογαριασμών χρηστών (πχ. όταν υπάρχουν διαδοχικές αποτυχημένες απόπειρες ταυτοποίησης)

		■ χρήση λογισμικού για την κατάλληλη παραγωγή ασφαλών κωδικών ■ ενημέρωση χρηστών για τον ασφαλή χειρισμό, τη συχνή ανανέωση κωδικών και τη χρήση διαφορετικών κωδικών ανά υπηρεσία ή εφαρμογή
Επιμονή (persistence)	Διασφαλίζεται η μακροχρόνια παραμονή του κακόβουλου λογισμικού στο στόχο	■ συχνοί έλεγχοι των επιπέδων ασφαλείας (TN) ■ τακτική παρακολούθηση πιθανών ανωμαλιών (TN)
Προσαρμογή & εξέλιξη (adaptation & evolvement)	Ο αλγόριθμος μαθαίνει από αποτυχίες χάριν στη μηχανική μάθηση & προσαρμόζει τις επόμενες επιθέσεις	■ συμπεριφορική ανάλυση (TN) ■ εντοπισμός απειλών βασισμένος στη μηχανική μάθηση (TN) ■ πολιτικές προληπτικής απενεργοποίησης λογαριασμών χρηστών (πχ. όταν υπάρχουν διαδοχικές αποτυχημένες απόπειρες ταυτοποίησης) ■ μη-επιστροφή λεπτομερών μηνυμάτων λάθους για τις αποτυχίες
Άγνωστα τρωτά σημεία (zero-days exploits)	Εντοπίζονται αδυναμίες που θα επιτρέψουν πρόσβαση στο στόχο	■ συχνή αναβάθμιση συστήματος & λογισμικού ■ τμηματοποίηση δικτύου ώστε να περιορίζεται η έκταση μιας πιθανής διάρρηξης ■ εφαρμογή διαστρωματωμένης αρχιτεκτονικής ασφαλείας (ακόμη και να γίνει διείσδυση σε ένα στρώμα, δε θα μπορεί να γίνει το ίδιο στα άλλα) ■ τακτική παρακολούθηση πιθανών ανωμαλιών (TN) (εντοπισμός της διείσδυσης ώστε να μετριαστούν οι επιπτώσεις της) ■ συμπεριφορική ανάλυση (TN)
Εξαγωγή δεδομένων (data exfiltration)	Τα δεδομένα αποστέλλονται σε διαχειριστή εκτός συστήματος, προς αξιοποίηση	■ κρυπτογράφηση δεδομένων που μεταφέρονται ■ ανίχνευση αποστολής ευαίσθητων δεδομένων (TN) ■ χρήση αναχωμάτων ασφαλείας (μπλοκάρισμα εξερχόμενης κίνησης) ■ ασφάλεια τελικών σημείων (TN) (η ασφάλεια τελικών σημείων θα μπορούσε να μπλοκάρει την

		αποστολή αυτή, καθοδηγούμενη από τη TN) ■ ανίχνευση στεγανογραφικού περιεχομένου (αν κάποιο δεδομένο περνάει ως μέρος εικόνας ή αρχείου, χωρίς όμως να κρυπτογραφείται)
Απόκρυψη (stealth)	Αποκρύπτονται τα ίχνη & καθίσταται δύσκολος ο εντοπισμός	■ ασφάλεια τελικών σημείων (TN) ■ εντοπισμός ανωμαλιών (TN) ■ συμπεριφορική ανάλυση (TN) (εντοπισμός ενεργειών απόκρυψης)
Προσβολή δεδομένων (data poisoning)	Τροποποιούνται ή εισάγονται παραπλανητικά δεδομένα που χειραγωγούν τη λειτουργία του συστήματος που προσβάλλουν	■ συχνοί έλεγχοι εγκυρότητας & αξιοπιστίας δεδομένων ■ χρήση αξιόπιστων αναβαθμίσεων στα λογισμικά ■ έλεγχος πρόσβασης με βάση τις ιδιότητες ■ χρήση ασφαλών πρωτοκόλλων επικοινωνίας ■ πιστοποίηση αυθεντικότητας μηνυμάτων / πηγής ■ κρυπτογράφηση δεδομένων (δε θα είναι κατανοητά από τον επιτιθέμενο και έτσι δε θα μπορεί να τα τροποποιήσει με κατάλληλο τρόπο ώστε να πετύχει τους σκοπούς του)
Εξατομίκευση (customization)	Δημιουργία προσωποποιημένων μηνυμάτων ή άλλων απεσταλμένων, με σκοπό την απόσπαση ευαίσθητων πληροφοριών	■ διαμοιρασμός πληροφοριών σχετικά με απειλές ■ ενημέρωση & εκπαίδευση χρηστών ■ ανίχνευση προσπαθειών ηλ. ψαρέματος (TN) (έλεγχος οποιουδήποτε περιεχομένου που αποστέλλεται, ώστε να ανιχνεύεται η καταλληλότητά του, οπότε, αυτό μπορεί να περιλαμβάνει έλεγχο περιεχομένου & ύφους, έλεγχο συνημμένων, έλεγχο χροιάς φωνής σε ηχητικά μηνύματα κ.ά.)
Αποφυγή (evasion)	Αποτροπή εντοπισμού από τους αμυντικούς μηχανισμούς του στόχου	■ εφαρμογή συστήματος εντοπισμού & αποτροπής εισβολών (TN) (θα πρέπει να είναι εξελιγμένο, ώστε να μπορεί να αντιμετωπίζει επιθέσεις οδηγούμενες από τη TN)

		▪ συχνή αναβάθμιση των αναχωμάτων & άλλων αντιμέτρων/μηχανισμών ασφαλείας ▪ έλεγχοι διείσδυσης (penetration testing) για εντοπισμό αδυναμιών (TN) ▪ συμπεριφορική ανάλυση (TN) ▪ ανίχνευση ανωμαλιών (TN)
--	--	--

Αρχικά, θα πρέπει να σημειωθεί ότι σε πολλά είδη επιθέσεων η μορφή του σταδίου της επίθεσης είτε είναι ίδια είτε έχει σημαντικές ομοιότητες. Στον Πίνακα 1 παρουσιάζονται οι φάσεις επίθεσης μια προς μια, συνοδευόμενες από μια σύντομη περιγραφή, ενώ παρουσιάζονται τα καταλληλότερα - κατά την άποψη του συγγραφέα - αντίμετρα, για κάθε μια από αυτές. Πιο συγκεκριμένα, υπάρχουν αντίμετρα των οποίων η διεκπεραίωση θεωρείται αποτελεσματικότερη εάν διενεργηθεί από την TN, κλασικά αντίμετρα καθώς και αντίμετρα που απαιτούν ανθρώπινη παρέμβαση, όπως η εκπαίδευση των χρηστών. Σκοπός είναι να εξετάσουμε την ποσότητα, καθολικότητα και αποτελεσματικότητα των αντιμέτρων που μπορούν να εφαρμοστούν σε κάθε φάση επίθεσης οδηγούμενης από την TN.

Όπως φαίνεται στον Πίνακα 1, κάθε στάδιο αντιμετωπίζεται καλύτερα με διαφορετικού τύπου αντίμετρα ενώ δε φαίνεται να υπάρχει κάποιο αντίμετρο καθολικής αποδοχής και χρήσης. Επίσης, είναι ορατό το γεγονός πως για ορισμένες φάσεις/στάδια επιθέσεων, τα αντίμετρα που μπορούν να προταθούν είναι αρκετά λίγα. Αυτό σημαίνει πως ακόμη και αν τα αντίμετρα αυτά συνδυαστούν, δεν θα μπορέσουν να αντιμετωπίσουν πλήρως τις ενέργειες που επιτελούνται στην αντίστοιχη φάση επίθεσης. Εν γένει, η χρήση διαφορετικών κάθε φορά αντιμέτρων και ο μικρός αριθμός αντιμέτρων σε ορισμένες φάσεις αποδεικνύει τη δυσκολία αντιμετώπισης των κυβερνο-επιθέσεων ή πιο σωστά, την πολυπλοκότητα που τις χαρακτηρίζει. Συνεπώς, συστήνεται οι χρήστες να διαθέτουν διάφορα επίπεδα προστασίας, εφαρμόζοντας μια διαστρωματική αρχιτεκτονική ασφάλειας, και να χρησιμοποιούν διαφορετικούς μηχανισμούς ή αντίμετρα (συνήθως οδηγούμενα από την TN) ανά στρώμα για να προστατεύουν τις ευαίσθητες πληροφορίες τους και τα περιουσιακά στοιχεία που διαθέτουν από επιτιθέμενους. Επίσης, προτείνεται η εστίαση της έρευνας σε εκείνα τα στάδια των επιθέσεων όπου η τρέχουσα συνεισφορά ως προς τα αντίμετρα είναι περιορισμένη ώστε να προταθούν πιο αποτελεσματικά αντίμετρα, τα οποία να είναι σε θέση να αντιμετωπίζουν πλήρως όλες τις πιθανές απειλές που δυνητικά θέτει η κάθε αντίστοιχη φάση επίθεσης. Τέλος, επειδή λόγω της φύσεως των σταδίων επίθεσης δεν είναι δυνατό να δημιουργηθεί ένα μόνο καθολικό αντίμετρο, προτείνεται να πραγματοποιηθεί εστίαση στη διαλειτουργικότητα των αντιμέτρων και τον κατάλληλο διαμοιρασμό της πληροφορίας μεταξύ τους ώστε να λειτουργήσει ενιαία και

αποτελεσματικά η όλη (διαστρωματωμένη) αρχιτεκτονική ασφάλειας και να προκύπτουν πάντοτε τα καλύτερα δυνατά αποτελέσματα από την εφαρμογή των αντιμέτρων που ενσωματώνονται.

5.6 Θετικά και Αρνητικά Αντιμέτρων Κυβερνοασφάλειας

Δεδομένων των παραπάνω πληροφοριών, είναι εύλογο να γεννάται σε πολλούς το ερώτημα για την καλύτερη επιλογή ανάμεσα στη χρήση παραδοσιακών αντιμέτρων και αντιμέτρων TN, προκειμένου να αμυνθούν με το βέλτιστο δυνατό τρόπο ενάντια σε αντίστοιχα υποστηριζόμενες από την TN, κυβερνο-επιθέσεις. Η απάντηση, όμως, σε ένα τέτοιο ζήτημα που άπτεται μιας τεχνολογικής δυνατότητας που διαρκώς αναβαθμίζεται, δε μπορεί να είναι ούτε σύντομη ούτε μονολεκτική. Υπάρχουν, λοιπόν, θετικά (pros) και αρνητικά (cons), τα οποία είναι καλό να αναλογιστεί κάποιος για αμφότερες τις επιλογές και παρατίθενται αναλυτικότερα στη συνέχεια της εργασίας.

Αρχικά, όσον αφορά τα **θετικά** που συνοδεύουν τα αντίμετρα που υποστηρίζονται από την TN, η ταχύτητα και οι αυτοματισμοί είναι ένα σημαντικό ζήτημα που οφείλει κάποιος να αναλογιστεί. Τα συστήματα TN μπορούν να εντοπίσουν και να ανταποκριθούν ιδιαίτερα γρήγορα στις κυβερνο-απειλές σε πραγματικό χρόνο, συχνά πολύ πιο γρήγορα από τις αντίστοιχες ενέργειες των ανθρώπων διαχειριστών ενός συστήματος. Μάλιστα, είναι ικανά να αυτοματοποιούν τυπικά καθήκοντα και μπορούν να ανταποκρίνονται καθ' όλη τη διάρκεια της μέρας, κάθε μέρα (ανταπόκριση 24/7). Στα πλαίσια αυτά, εντάσσεται και το γεγονός ότι η TN μπορεί να επεξεργαστεί αποτελεσματικά και να αναλύσει γρήγορα, μεγάλους όγκους δεδομένων, με σκοπό την αναγνώριση ανωμαλιών και μοτίβων που θα μπορούσαν να θεωρηθούν ενδείξεις μιας (ή και πολλαπλών) κυβερνο-επίθεσης, σε αντίθεση με τις παραδοσιακές μεθόδους (Fei & Zheng, 2022).

Η ικανότητα προσαρμογής των συστημάτων που υποστηρίζονται από την TN (adaptability), είναι μια ακόμη παράμετρος. Πολλά από αυτά τα συστήματα, είναι από κατασκευής τους ικανά να προσαρμόζονται σε εξελισσόμενες τεχνικές και μοτίβα κυβερνο-επιθέσεων. Χάριν στα μοντέλα μηχανικής μάθησης, τα συστήματα αυτά είναι ικανά να μάθουν - με τη βοήθεια νέων δεδομένων που θα ενσωματώσουν - και να βελτιωθούν με την πάροδο του χρόνου, καθιστώντας τα περισσότερο ανθεκτικά ενάντια σε αναδυόμενους κινδύνους. Αυτό συνεπάγεται επίσης μείωση της συμμετοχής του ανθρώπινου παράγοντα και ως εκ τούτου μείωση των δυνητικών λαθών (ή της πιθανότητας λαθών). Η TN, εξάλλου, δεν υποπίπτει σε ανθρώπινα λάθη (αν και αυτό εξαρτάται από τα δεδομένα με τα οποία εκπαιδεύεται, που σημαίνει πως αν αυτά παράγονται από ανθρώπους τότε ενδέχεται να κάνει και αυτή λάθη, εφόσον τα δεδομένα είναι λανθασμένα) ούτε διέπεται από προκαταλήψεις που συνοδεύουν συχνά την ανθρώπινη φύση και τα οποία αμφότερα μπορούν να

θεωρηθούν μειονέκτημα σε επίπεδο κυβερνο-ασφάλειας, όπου τα λάθη είναι πολύ πιθανό να οδηγήσουν σε παράβλεψη τρωτών σημείων και διόγκωση των ευπαθειών έναντι κυβερνο-επιθέσεων (Yamin et al., 2021). Επιπροσθέτως, ο άνθρωπος δε διαθέτει την ίδια ικανότητα εντοπισμού ασυνήθιστων συμπεριφορών ή μοτίβων (behavior-based detection) όπως η TN, με αποτέλεσμα να υπολείπεται αυτής απέναντι σε επιθέσεις που διενεργούνται από ιδιαίτερα υψηλού επιπέδου προγραμματισμένη TN (Zeng, 2022).

Παρά τις ικανότητές της, όμως, δε θα πρέπει να αποκλειστεί πλήρως ότι η TN ενδέχεται να υποπέσει σε **λάθη**. Τα υποστηριζόμενα από TN συστήματα μπορεί να υποπέσουν σε “*λάθος θετικές εκτιμήσεις*” (false positives, δηλαδή να αναγνωρίσουν μια εξουσιοδοτημένη ενέργεια ως κακόβουλη) ή σε “*λάθος αρνητικές εκτιμήσεις*” (false negatives, δηλαδή την αποτυχία αναγνώρισης μιας πραγματικής απειλής). Συνεπώς, τέτοιου είδους σφάλματα μπορούν κάλλιστα να οδηγήσουν είτε σε αχρείαστες ειδοποιήσεις (συναγερμούς) είτε σε μη-εντοπισμένους κινδύνους, προκαλώντας την ανάλογη σύγχυση στους ανθρώπους - χειριστές των συστημάτων και επιτρέποντας ενδεχομένως μελλοντικές παραβιάσεις. Αυτός είναι και ο λόγος που απαιτείται σχεδόν πάντοτε η ανθρώπινη παρουσία και εξειδίκευση, μιας και η ίδια η TN δε διαθέτει την ικανότητα να αντιλαμβάνεται σε κάθε περίπτωση το ευρύτερο περιβάλλον ασφαλείας εντός του οποίου δραστηριοποιείται ένας οργανισμός. Επομένως, ο ανθρώπινος παράγοντας φαίνεται πως παραμένει όχι μόνο χρήσιμος αλλά και απαραίτητος, αφού αυτός μπορεί να ερμηνεύσει ορθώς τις ειδοποιήσεις και τους συναγερμούς που θα ενεργοποιήσει η TN, ενώ δυνητικά θα μπορούσε να αντιληφθεί ένα false negative, συνδυάζοντας διαφορετικά είδη πληροφορίας και γνώσεων που μπορεί να μην είναι διαθέσιμα σε ένα αντίμετρο TN (Fei & Zheng, 2022). Τέλος, η TN μπορεί να αποδειχθεί ιδιαίτερα τρωτή έναντι των αντιπαραθετικών επιθέσεων (adversarial attacks), δηλαδή σειράς ενεργειών κατά τις οποίες οι επιτιθέμενοι προσπαθούν είτε να χειραγωγήσουν είτε να εξαπατήσουν τους αλγορίθμους TN (Wang et al., 2023).

Εξερευνώντας τώρα τα θετικά των παραδοσιακών μεθόδων άμυνας, θα πρέπει αρχικά να σημειωθεί ότι οι πρακτικές αυτές συχνά βασίζονται στην ανθρώπινη εξειδίκευση και την απόκτηση πείρας (expertise & experience), τα οποία δε μπορούν να ενσωματώνονται στα αντίμετρα TN ούτε στο σύνολό τους αλλά ούτε και γρήγορα. Οι πρακτικές βαθιάς μάθησης (deep learning), βέβαια, επιτρέπουν ένα αντίμετρο να αποκτήσει σταδιακά εξειδίκευση. Επομένως, μελλοντικά μπορεί να διευρυνθεί η χρήση τους και προς αυτή την κατεύθυνση. Οι πρακτικές αυτές μπορούν να προσφέρουν σε βάθος κατανόηση είτε των απαιτήσεων ενός μεμονωμένου χρήστη είτε της μοναδικότητας του περιβάλλοντος ασφαλείας εντός του οποίου λειτουργεί ένας οργανισμός, με

όποιες ανάγκες μπορεί να εγείρονται και όσες προκλήσεις ενδέχεται να προκύπτουν (King et al., 2020).

Επιπροσθέτως, ο ανθρώπινος παράγοντας μπορεί να αντιληφθεί με πιο ολιστική ματιά το ευρύτερο πλαίσιο και τις επιπλοκές μιας κυβερνο-επιθέσης, επιτρέποντας μια περισσότερο περιεκτική ανταπόκριση. Επιπροσθέτως, δεν είναι λίγες οι παραδοσιακές μέθοδοι που έχουν εφαρμοστεί για χρόνια και έχουν επιδείξει την αποτελεσματικότητά τους σε ποικίλα διαφορετικά σενάρια κυβερνο-επιθέσεων. Οι μέθοδοι αυτές έχουν παρουσιαστεί σε προηγούμενη ενότητα (Ενότητα 5.3) της εργασίας. (Agrawal, Gans & Goldfarb, 2018).

Φυσικά, όπως με τα αντίμετρα βασισμένα στη τεχνητή νοημοσύνη έτσι και με τις παραδοσιακές μεθόδους, υπάρχουν και αρνητικές πτυχές πέρα από τα θετικά που τις συνοδεύουν. Για παράδειγμα, είναι χαρακτηριστικό το γεγονός ότι οι παραδοσιακές μέθοδοι συνοδεύονται από χαμηλότερες ταχύτητες ανταπόκρισης (slower response rate), αφού είθισται να μην είναι το ίδιο γρήγορες στον εντοπισμό και την αντιμετώπιση των αναδυόμενων κινδύνων, αυξάνοντας έτσι την τρωτότητα των συστημάτων (Ardito et al., 2021).

Επιπροσθέτως, οι παραδοσιακές μέθοδοι είθισται να απαιτούν αρκετές ώρες ενασχόλησης (labor intensive) και ενίοτε απαιτούν αρκετούς πόρους (resource intensive), παράμετρος που συνοδεύεται από αυξημένα κόστη σε σύγκριση με την TN, ιδίως για οργανισμούς οι οποίοι στηρίζονται αποκλειστικά σε αναλυτές για την παρακολούθηση της ασφάλειάς τους από κυβερνο-επιθέσεις. Επιπλέον, ένα ακόμη αρνητικό στοιχείο είναι η περιορισμένη επεκτασιμότητά τους (limited scalability), αφού οι παραδοσιακές μέθοδοι δυσκολεύονται συνήθως να ανταπεξέλθουν στην κλίμακα και την πολυπλοκότητα των σύγχρονων κυβερνο-επιθέσεων, ιδιαίτερα όσο το μέγεθος της επίθεσης παρουσιάζει διεύρυνση (όπως ισχύει πχ. σε επιθέσεις τύπου DDoS). Τέλος, απουσιάζει η αυτόματη προσαρμοστικότητα (automated adaptability), η οποία στις παραδοσιακές μεθόδους προκύπτει χειρωνακτικά, αφού έχει λάβει χώρα μια επίπονη αλλαγή του αντιμέτρου, έτσι ώστε να προκύψει μια νέα, βελτιωμένη έκδοσή του (Charmet et al., 2022).

5.7 Προβλήματα & προκλήσεις από τη χρήση των αντιμέτρων TN

Ολοκληρώνοντας το παρόν κεφάλαιο, αξίζει μια σύντομη αναφορά σε ορισμένα προβλήματα που καλείται να αντιμετωπίσει ο άνθρωπος και πηγάζουν από τη διαρκώς αυξανόμενη χρήση αντιμέτρων τα οποία υποστηρίζονται από την TN, σε ένα εχθρικό διαδικτυακό περιβάλλον, όπου οι επίσης υποστηριζόμενοι από την TN, κίνδυνοι, караδοκούν. Αρχικά, να σημειωθεί ότι στη

βιβλιογραφία αναφέρονται - και παρουσιάστηκαν σε προηγούμενη υποενότητα - δύο ελαττώματα στα οποία μπορεί να υποπέσει η TN: αφενός τις “*false positives*” και αφετέρου τις “*false negatives*” εκτιμήσεις. Το κάθε ένα από αυτά τα δύο ελαττώματα μπορεί να οδηγήσει σε σύγχυση και στην ουσία να μην εξουδετερώσει έναν εν δυνάμει κίνδυνο. Το πρόβλημα αυτό δυσχεραίνεται χάριν στις adversarial επιθέσεις, οι οποίες σχεδιάζονται ώστε να υποσκάψουν την αποτελεσματικότητα ενός συστήματος ασφαλείας οδηγούμενο/βασιζόμενο από TN, κάνοντας «καλά» δεδομένα να φαίνονται κακόβουλα ή το αντίστροφο. Επιπλέον, τα πρώτο από αυτά τα ελαττώματα (false positives) μπορεί να οδηγήσει και σε σπατάλη πολύτιμων πόρων, είτε ανθρώπινων είτε υπολογιστικών, που θα έπρεπε να έχουν διατεθεί για άλλους σκοπούς. Επομένως, είναι επιτακτική η βελτίωση των αντιμέτρων ως προς τα ελαττώματα αυτά. Η βελτίωση αυτή μπορεί να επέλθει μέσα από την ενσωμάτωση στα αντίμετρα TN, εξελιγμένων τεχνικών συμπεριφορικής ανάλυσης ώστε να εντοπίζουν ανωμαλίες (Jeffrey, Tan & Villar, 2023), τη χρήση περισσότερο εξειδικευμένων μοντέλων μηχανικής μάθησης όπως deep learning ή reinforced learning, για τη βελτίωση της ακρίβειας στον εντοπισμό των απειλών (Shiri et al., 2023), τη χρήση προβλεπτικών αναλύσεων (predictive analytics) για την αποσόβηση κυβερνο-απειλών που βασίζονται σε δεδομένα ιστορικού (Agrawal, Gans & Goldfarb, 2018), την ενσωμάτωση πληροφόρησης σχετικής με απειλές (Vast et al., 2021) ή την προσέγγιση τύπου “*human-in-the-loop*”, όπου ειδικοί της κυβερνο-ασφάλειας συνεργάζονται με συστήματα TN (DHH, 2016).

Μια πρόκληση που καλείται να αντιμετωπίσει ο άνθρωπος, σχετίζεται με την αυξανόμενη προσκόλληση στην αξιοποίηση της TN και των εφαρμογών (ή λογισμικών) που υποστηρίζει, για το χτίσιμο των αμυνών. Φαίνεται να γίνεται αντιληπτό στους τεχνολογικούς κύκλους, ότι πολλοί οργανισμοί εξαρτώνται σε υπερβολικά μεγάλο από τα αντίμετρα που βασίζονται στην TN, παραμελώντας άλλες ουσιώδεις πρακτικές και στρατηγικές ασφαλείας, θεωρώντας πως η ενσωμάτωσή της TN στη λειτουργία τους αποτελεί παράγοντα ικανό να αποτρέψει οποιαδήποτε απειλή προερχόμενη από τον κυβερνοχώρο. Αυτή η υπερβολική εξάρτηση μπορεί να δημιουργήσει μικρά σημεία ασφάλειας στο αμυντικό οικοδόμημα, τα οποία χωλαίνουν (τα οποία είναι μικρά και δεν καλύπτουν πλήρως όλο το αμυντικό οικοδόμημα, οπότε από τη στιγμή που θα εντοπιστούν, θεωρείται πιο εύκολο να ξεπεραστούν) και μπορούν να οδηγήσουν σε δυσάρεστα αποτελέσματα εφόσον γίνουν αντιληπτά από κακόβουλους χρήστες (Comiter, 2019) και ξεπεραστούν. Πρόκληση επίσης αποτελεί και η τεράστια ανάγκη που συνοδεύει τα βασισμένα στην TN, συστήματα, για συνεχώς μεγαλύτερη υπολογιστική ισχύ (που απαιτείται για την εκπαίδευση τους με μεγάλα δεδομένα αλλά και για την τάχιστη ανταπόκρισή τους), διαρκώς πιο εξειδικευμένο προσωπικό και φυσικά οικονομικούς πόρους. Οι μικρότερου μεγέθους οργανισμοί είναι πολύ πιθανό να

συναντήσουν σοβαρές δυσκολίες όταν έρθουν αντιμέτωπες με την ικανοποίηση αυτών των αναγκών, αν και πρέπει να σημειωθεί ότι υπάρχει το υπολογιστικό νέφος, το οποίο μπορεί να μετριάσει το πρόβλημα αυτό, λόγω του ευέλικτου μοντέλου χρέωσης που προσφέρει αλλά και των υπηρεσιών εκπαίδευσης μοντέλων MM/TN που παρέχει (de Witt et al., 2021).

Οι ηθικές και νομικές ανησυχίες αποτελούν ένα ζήτημα που φαίνεται να απασχολεί όχι μόνο τους μεμονωμένους χρήστες αλλά και σύνολα κρατών, όπως η Ευρωπαϊκή Ένωση, η οποία πριν μερικά χρόνια εξέδωσε οδηγίες για την προστασία των δεδομένων των πολιτών της (GDPR) (European Union Agency for Network & Information Security, 2015), με σκοπό να προλάβει δυσάρεστες καταστάσεις εξαιτίας της εξέλιξης της τεχνολογίας και των ανάρμοστων πρακτικών μεγάλων εταιρειών. Η χρήση TN στην κυβερνο-ασφάλεια μπορεί να εγείρει ηθικά και νομικά ζητήματα, ιδίως όταν στη συζήτηση υπεισέρχεται η συλλογή και ανάλυση δεδομένων, η ιδιωτικότητα και η λήψη αποφάσεων. Επομένως, θεωρείται κομβικής σημασίας να διασφαλίζεται πως οι άμυνες, οι οποίες βασίζονται στην TN, συμμορφώνονται με τους νομικούς και ηθικούς κανόνες που διέπουν την κάθε περιοχή, ώστε να ελαχιστοποιείται η εύκολη απόκτηση πρόσβασης κακόβουλων χρηστών σε προσωπικά ή ευαίσθητα δεδομένα (Dignum, 2019).

Ένα σοβαρό πρόβλημα που χρήζει παρακολούθησης, είναι το γεγονός ότι παρά τις ικανότητες που παρουσιάζει η TN, είναι πολύ πιθανό ορισμένες φορές να δυσκολευτεί να προσαρμοστεί στις διαρκώς εξελισσόμενες πρακτικές των κυβερνο-επιθέσεων. Η αλήθεια είναι πως η TN στηρίζεται στα δεδομένα και τα μοτίβα στα οποία έχει «εκπαιδευτεί» και ενδέχεται να αδυνατεί να ανταπεξέλθει αποτελεσματικά ή ακόμη και να εντοπίσει εγκαίρως, τις νέες απειλές, αφού ακόμη και με τη βοήθεια της μηχανικής μάθησης, η εξέλιξή της μπορεί να είναι πιο αργή από την αντίστοιχη των κινδύνων που караδοκούν. Επιπλέον, είναι επίσης γεγονός ότι τα βασισμένα στην TN συστήματα που αξιοποιούνται για την κυβερνοασφάλεια, δεν «εκπαιδεύονται» ούτε δοκιμάζονται και αποτιμώνται επαρκώς. Ως εκ τούτου, μπορεί να παρουσιάσουν συμπεριφορά ρατσιστική (πιο απλά, να προβούν σε διακρίσεις όσον αφορά τις επιλογές τους) και να χαρακτηρίζονται από προκαταλήψεις, οδηγώντας ενδεχομένως σε λανθασμένες εκτιμήσεις κινδύνων ή να εξαιρέσουν αδικαιολόγητα, συγκεκριμένες κατηγορίες χρηστών από την πρόσβαση σε πληροφορίες. Τέλος, θα πρέπει να σημειωθεί πως ένα αντίμετρο TN θα μπορούσε να υλοποιηθεί για εσκεμμένη και όχι τυχαία εφαρμογή ρατσιστικής συμπεριφοράς. Δηλαδή η συμπεριφορά αυτή του αντιμέτρου, δεν οφείλεται σε μη-επαρκή εκπαίδευση αλλά σε εσκεμμένη υλοποίηση (Fei & Zheng, 2022).

Σοβαρό πρόβλημα θεωρείται και η πολυπλοκότητα που διέπει τη χρήση αντιμέτρων, βασισμένων στην TN. Η πολύπλοκη διαχείριση την οποία απαιτεί η εφαρμογή τους, γίνεται από

εξειδικευμένο προσωπικό που διαμορφώνει, συντηρεί και προχωρά σε τακτικές αναβαθμίσεις τα εν λόγω συστήματα. Ωστόσο, πολλές εταιρείες ενδέχεται να συναντήσουν σοβαρές δυσκολίες στην απασχόληση τέτοιων εργαζομένων, αν και για αυτές τις περιπτώσεις συστήνονται συνήθως οι λύσεις της επιλογής ενός εξωτερικού παρόχου υπηρεσιών κυβερνο-ασφάλειας (outsourcing), που επίσης συνοδεύονται από συγκεκριμένους κινδύνους (όπως την έλλειψη πλήρους ελέγχου για την ασφάλεια, το ενδεχόμενο μη-συμμόρφωσης με κανονιστικά πλαίσια για την ορθή χρήση των δεδομένων, την εξάρτηση από τρίτη πηγή κ.ά.) αλλά τουλάχιστον είναι οικονομικά βιώσιμες για τις εταιρείες που τις χρειάζονται. Ακόμη όμως και στην περίπτωση του outsourcing, θα πρέπει να γίνεται σωστή διαμόρφωση και χρήση των παρεχόμενων υπηρεσιών, ώστε να επέρχονται τα αναμενόμενα αποτελέσματα. Οπότε, διαπιστώνεται εύκολα ότι το προσωπικό των οργανισμών θα πρέπει να εκπαιδεύεται στη σωστή χρήση τους. Επομένως, πάλι οι οργανισμοί έρχονται αντιμέτωποι με κάποιο κόστος (κόστος χρήσης υπηρεσιών & εκπαίδευσης προσωπικού). (Olaniyan, Rakshit & Vajjhala, 2023)

Ακόμη, αποτελεί πρόκληση η πλήρης κατανόηση του τρόπου με τον οποίο τα βασισμένα στην ΤΝ, συστήματα, λαμβάνουν αποφάσεις, ιδίως μετά από εκτενή έκθεση σε πρακτικές μηχανικής μάθησης. Το γεγονός αυτό όμως οδηγεί σε έλλειψη διαφάνειας (transparency), αναφορικά με τη διαδικασία λήψης αποφάσεων, και μπορεί να αποδειχθεί εν τέλει σοβαρό πρόβλημα σε περιπτώσεις συμμόρφωσης προς κανονιστικές ρυθμίσεις ή λογοδοσίας (Matt, Modrák & Zsifkovits, 2020).

Μια ακόμη σοβαρή πρόκληση, είναι η πολυπλοκότητα των απειλών. Οι σύγχρονες απειλές είναι συχνά σύνθετες και δυναμικές, απαιτώντας ένα συνδυασμός αντιμέτρων για την πιο ολοκληρωμένη και αποτελεσματική αντιμετώπισή τους. Μάλιστα, η επίτευξη διαλειτουργικότητας θα συνέβαλλε στην αποτελεσματικότερη χρήση τους, ταυτοχρόνως σε ένα σύστημα ή δίκτυο, όπως για παράδειγμα το σύνθετο περιβάλλον ενός υπολογιστικού νέφους. Η διαλειτουργικότητα αποτελεί σίγουρα μια πρόκληση που δεν είναι πάντα εύκολο να αντιμετωπιστεί, δεδομένου ότι στο σημερινό διασυνδεδεμένο κόσμο με την υψηλή εξάρτηση συστημάτων και τεχνολογιών, η ανάπτυξη αντιμέτρων από διαφορετικούς κατασκευαστές, με πλήθος προϋποθέσεων συμβατότητας, ενίοτε μπορεί να δυσκολέψει την ταυτόχρονη λειτουργία αντιμέτρων. Συν τοις άλλοις, οι απειλές εξελίσσονται διαρκώς, με αποτέλεσμα να εντοπίζονται καινούρια τρωτά σημεία, γεγονός που σημαίνει πως τα αντίμετρα θα πρέπει να προσαρμόζονται αναλόγως, προκειμένου να παραμείνουν αποτελεσματικά. Τέλος, δεν θα πρέπει να υποτιμηθεί ο παρονομαστής της γνώσης, η οποία θα πρέπει να διαμοιράζεται μεταξύ συνεργαζόμενων αντιμέτρων ώστε αυτά πάντοτε να λειτουργούν με βάση τη μέγιστη δυνατή πληροφόρηση, πράγμα το οποίο εξίσου οδηγεί σε αυξημένη

αποτελεσματικότητα. Και πάλι εδώ η χρήση του υπολογιστικού νέφους θα μπορούσε να βοηθήσει στην κατεύθυνση της τάχιστης διασύνδεσης και ασφαλούς επικοινωνίας μεταξύ των αντιμέτρων.

Προχωρώντας λοιπόν στην κατάθεση μιας σειρά προτάσεων για την αντιμετώπιση των προκλήσεων και την αποτελεσματική καταπολέμηση των κινδύνων που εγείρουν οι κυβερνο-απειλές, η αρχή γίνεται με τη συστηματική και εξονυχιστική αξιολόγηση κινδύνων, η οποία μπορεί να βοηθήσει στην κατανόηση συγκεκριμένων απειλών του διαδικτύου και των τρωτών σημείων που οδηγούν σε έκθεση ευαίσθητων πληροφοριών ή δεδομένων, γεγονός που θα οδηγούσε στη βέλτιστη δυνατή επιλογή αντιμέτρων. Επιπροσθέτως, η χρήση ανοιχτών προτύπων και η καθιέρωση συνεργασίας ανάμεσα σε δημιουργούς αντιμέτρων, με διαρκή ανταλλαγή πληροφορήσης γύρω από τους αναδυόμενους κινδύνους, μπορεί να διασφαλίσει την καθιέρωση κοινών προτύπων αλλά και τη βελτίωση της διαλειτουργικότητάς τους, ενώ τέλος, συστήνεται η επένδυση στην εκπαίδευση του προσωπικού και ενημέρωση χρηστών, δεδομένου ότι ο ανθρώπινος παράγοντας πάντα θα διαδραματίζει ρόλο τόσο στην ευπάθεια των συστημάτων όσο και στην αποτελεσματική αποτροπή των κινδύνων.

Όσον αφορά τώρα ορισμένες τεχνικές με τις οποίες θα μπορούσαν να αντιμετωπιστούν προκλήσεις στο θέμα της διαλειτουργικότητας των αντιμέτρων και της αποτελεσματικότερης καταπολέμησης των κυβερνο-απειλών, υπάρχουν παραδείγματα όπως οι μηχανές Tsetlin, οι οποίες αποτελούν έναν τύπο μοντέλου μηχανικής μάθησης που μπορεί να συμβάλει προς αυτή την κατεύθυνση. Αυτού του είδους τα εργαλεία έχουν σχεδιαστεί για να είναι προσαρμόσιμα, καθιστώντας τα κατάλληλα για δυναμικά περιβάλλοντα, στα οποία το πεδίο των απειλών αλλάζει διαρκώς και αυτά έχουν τη δυνατότητα να αναβαθμίζονται και να βελτιώνονται ώστε να αντιμετωπίζουν αποτελεσματικά, νέες απειλές. Είναι επίσης ερμηνεύσιμα (interpretable), κάτι που λείπει από τα παραδοσιακά μοντέλα μηχανικής μάθησης, μια ιδιότητα που επιτρέπει τους ειδικούς της κυβερνο-ασφάλειας να κατανοήσουν πώς το μοντέλο καταλήγει στη λήψη αποφάσεων, με αποτέλεσμα να καθίσταται ευκολότερη η ενσωμάτωση σε αυτά, υφιστάμενων μέτρων και πολιτικών ασφαλείας. Τέλος, θεωρούνται ιδιαίτερα εύρωστα (robust), μια ιδιότητα που επιτρέπει τη διαχείριση θορυβωδών δεδομένων (noisy data), η οποία ικανότητα στο χώρο της κυβερνο-ασφάλειας θεωρείται ιδιαίτερα ωφέλιμη για την αποτελεσματική αντιμετώπιση των σύνθετων κυβερνο-απειλών (Granmo, 2021).

6 Συμπεράσματα

6.1 Συνεισφορά διπλωματικής

Η παρούσα εργασία, λοιπόν, έχει ασχοληθεί με ένα ιδιαίτερα σύγχρονο αλλά και σοβαρό ζήτημα της εποχής, αυτό της κυβερνο-ασφάλειας και των κινδύνων εξαιτίας των οποίων απειλούνται προσωπικά δεδομένα, ευαίσθητες πληροφορίες ατόμων και οργανισμών αλλά ακόμη και η καλή λειτουργία δικτύων και συστημάτων. Πιο συγκεκριμένα, όμως, οι πτυχές αυτές διέρχονται μέσα από το πρίσμα της TN και τον τρόπο με τον οποίο αυτή είναι ικανή να συμβάλλει είτε στην αναβάθμιση των απειλών υπέρ του συμφέροντος των κακόβουλων χρηστών είτε στη βελτίωση της ασφάλειας των χρηστών του διαδικτύου.

Η συνεισφορά της παρούσας εργασίας συναντάται στο γεγονός ότι καλύπτει ένα πολύ μεγάλο κομμάτι αυτού του ζητήματος, παρουσιάζοντας πληθώρα πληροφοριών τόσο γύρω από τον τρόπο που λαμβάνουν χώρα οι επιθέσεις που υποστηρίζονται από την TN (AI-based attacks) όσο και τις επιλογές που έχουν οι χρήστες για τη βελτίωση της άμυνάς τους που επίσης υποστηρίζονται από την TN (AI-based countermeasures). Επιπλέον, η εργασία συνεισφέρει και με έναν ακόμη τρόπο στα ακαδημαϊκά δρώμενα. Διαπιστώθηκε πως στην Ελλάδα δεν έχουν λάβει χώρα παρά ελάχιστες εργασίες που να καλύπτουν επαρκώς το ζήτημα των απειλών και των αντιμέτρων που βασίζονται στην TN, ενώ όσες έχουν εκπονηθεί, περιλαμβάνουν συνήθως ένα γενικό πλαίσιο της ασφάλειας στο διαδίκτυο, πώς μπορεί αυτή να επιτευχθεί και ποια δεδομένα χρηστών ή οργανισμών, βρίσκονται σε κίνδυνο. Συνεπώς, η παρούσα εργασία - όπως εύκολα μπορεί κάποιος να διαπιστώσει από τον πίνακα περιεχομένων αλλά και το ίδιο το περιεχόμενό της αναλυτικά - περιλαμβάνει πληθώρα αυτών των πληροφοριών και μάλιστα με παραδείγματα, τα οποία βοηθούν τον αναγνώστη να κατανοήσει πολλές πτυχές τόσο της λειτουργίας των επιθέσεων που λαμβάνουν χώρα μέσω TN όσο και των υποστηριζόμενων από την TN, πρακτικών άμυνας που είναι ικανές να τις αποκρούσουν.

Τα πλέον θετικά στοιχεία, λοιπόν, που χαρακτηρίζουν την παρούσα διπλωματική, εντοπίζονται αρχικά στο γεγονός ότι αναλύονται οι φάσεις των επιθέσεων οδηγούμενων από την TN, άρα γίνεται σε βάθος κατανοητός ο τρόπος με τον οποίο αυτές λειτουργούν αναφορικά με την αναγνώριση τρωτών σημείων, τη διάκριση σημαντικών δεδομένων και την αποφυγή εντοπισμού του κακόβουλου λογισμικού που έχουν εκχύσει, το οποίο «μόλυνε» το σύστημα ή δίκτυο. Επιπροσθέτως, θετικό είναι και το γεγονός ότι παρουσιάζονται εξίσου αναλυτικά, οι αμυντικοί

μηχανισμοί βασιζόμενοι στην TN που μπορούν να αξιοποιήσουν οι χρήστες, καθώς και με ποιο τρόπο προστατεύουν τα ευαίσθητα δεδομένα και τα περιουσιακά τους στοιχεία, από υποκλοπή και διαρροή. Μάλιστα, με τον συγκεντρωτικό πίνακα στον οποίο παρουσιάζονται οι διάφορες φάσεις των κυβερνο-επιθέσεων και τα πλέον κατάλληλα εργαλεία αποτροπής τους, δημιουργείται στον αναγνώστη μια συνολική εικόνα αυτού του πολύ σημαντικού ζητήματος. Η γνώση που αυτός ο πίνακας προσφέρει μπορεί να αποδειχθεί πολύτιμη για έναν οργανισμό, ο οποίος επιθυμεί να επενδύσει σε κατάλληλα αντίμετρα για την αντιμετώπιση επιθέσεων κυβερνο-ασφάλειας που στοχεύουν τα συστήματά του. Ειδικότερα, γνωρίζοντας τις πιθανές επιθέσεις που μπορούν να συμβούν και τις φάσεις που τις απαρτίζουν, ο οργανισμός μπορεί να επιλέξει ένα κατάλληλο υπο-σύνολο από όλα τα δυνατά αντίμετρα που μπορούν να τις αντιμετωπίσουν.

Η TN, λοιπόν, έχει μετατραπεί σε αναπόσπαστο κομμάτι του σύγχρονου ψηφιακού τοπίου, επιτυγχάνοντας αξιοσημείωτη συμβολή σε επίπεδο κυβερνο-ασφάλειας αλλά εγείροντας και ορισμένες προκλήσεις, οι οποίες δεν πρέπει να παραβλέπονται έτσι ώστε να αντιμετωπίζονται καταλλήλως. Η TN, έχει οδηγήσει σε επανάσταση το σκέλος της κυβερνο-ασφάλειας - τόσο από τη σκοπιά των αμυνόμενων όσο και από τη σκοπιά των επιτιθέμενων - αφού ενισχύει τις πρακτικές εντοπισμού και ανταπόκρισης έναντι απειλών, ενώ ταυτόχρονα διευκολύνει τις προσπάθειες εξάπλησης των συστημάτων ασφαλείας και παραβίασης ισχυρών προστατευτικών εργαλείων.

Για τους αμυνόμενους, οι αλγόριθμοι μηχανικής μάθησης και η ικανότητα προβλεπτικής ανάλυσης (predictive analytics) εφαρμόζονται με σκοπό να αναλυθούν δεδομένα μεγάλου όγκου και να αναγνωριστούν μοτίβα που υποδηλώνουν κυβερνο-επίθεση, γεγονός που έχει οδηγήσει στην ανάπτυξη συστημάτων εντοπισμού εισβολών (intrusion detection systems) υποστηριζόμενων από την TN, τα οποία είναι ικανά να παρακολουθούν την κινητικότητα του δικτύου και να αναγνωρίζουν μη-φυσιολογικές κινήσεις (ή ανωμαλίες), με εκπληκτική ακρίβεια. Επιπροσθέτως, η ικανότητα της TN να επεξεργάζεται και να κατηγοριοποιεί ογκώδη σύνολα δεδομένων, έχει βελτιώσει τη δυνατότητα αναγνώρισης από τους οργανισμούς (αλλά και τα άτομα) των δικών τους τρωτών σημείων, βοηθώντας τα να λαμβάνουν μέτρα πρόληψης των συστημάτων και των περιουσιακών τους στοιχείων γενικότερα.

Για τους επιτιθέμενους - δεδομένου ότι η TN που ενισχύει την κυβερνο-ασφάλεια μπορεί και να την απειλήσει - έχουν βελτιωθεί οι δυνατότητες κατασκευής κακόβουλων λογισμικών, με αποτέλεσμα οι επιθέσεις που υποστηρίζονται από την TN να έχουν μετατραπεί σε σοβαρές προκλήσεις για τους αμυνόμενους. Οι κυβερνο-εγκληματίες αλλά και όσοι υποστηρίζονται από κρατικές υπηρεσίες με απώτερο στόχο να λειτουργήσουν απειλητικά προς άλλα κράτη (κυρίως επιδιώκοντας την πρόσβαση σε ευαίσθητες πληροφορίες), είναι σε θέση να εφαρμόσουν εργαλεία

που καθοδηγούνται από την TN (AI-driven) προκειμένου να διενεργήσουν εξειδικευμένης φύσεως και στοχευμένες επιθέσεις, να αυτοματοποιούν τεχνικές κοινωνικής μηχανικής και να αναπτύσσουν κακόβουλα λογισμικά, τα οποία έχουν την ικανότητα προσαρμογής και εξέλιξης σε πραγματικό χρόνο (real-time adapt & evolve), ώστε να αποφεύγουν τον εντοπισμό. Η αντιπαραθετική μηχανική μάθηση (adversarial machine learning) με τη σειρά της, αποτελεί μια τεχνική όπου τα μοντέλα TN αποτελούν αντικείμενο χειραγώγησης προκειμένου να οδηγηθούν σε λανθασμένες προβλέψεις. Ως εκ τούτου, συνθέτει μια σημαντική απειλή για την αξιοπιστία των αμυνών κυβερνο-ασφάλειας που βασίζονται στην TN.

Σε αυτές τις απειλές, οι επαγγελματίες του χώρου της κυβερνο-ασφάλειας και η επιστημονική κοινότητα γενικότερα, αναπτύσσει αντίμετρα βασισμένα στην TN, ώστε να εντοπίζουν τις αντίστοιχα βασισμένες στην TN, επιθέσεις και να μειώνουν τον αντίκτυπό τους. Τα αντίμετρα αυτά περιλαμβάνουν μοντέλα TN, τα οποία έχουν σχεδιαστεί συγκεκριμένα ώστε να αναγνωρίζουν ευκολότερα τις αντιπαραθετικές επιθέσεις (adversarial attacks) και σε πραγματικό χρόνο να τις αποκρούουν. Επιπροσθέτως, το αποκαλούμενο «κυνήγι κυβερνο-απειλών υποστηριζόμενων από την TN» (AI-powered threat hunting) και οι πλατφόρμες πληροφόρησης κατά των κυβερνο-απειλών αξιοποιούνται προληπτικά, ώστε να αναγνωρίζονται οι αναδυόμενοι κίνδυνοι και τα πιθανά τρωτά σημεία. Η συνεργασία ανάμεσα στην ανθρώπινη εξειδίκευση και τις ικανότητες ανάλυσης που προσφέρει η TN, πιστεύεται πως αποτελεί τον καταλληλότερο συνδυασμό στη συνεχιζόμενη μάχη ενάντια στις κυβερνο-επιθέσεις που υποστηρίζονται από TN.

6.2 Βασικά συμπεράσματα

Εν κατακλείδι, ενσωματώνοντας τα βασισμένα στην TN αντίμετρα, στις στρατηγικές κυβερνοασφάλειας, αυτό μπορεί να βοηθήσει στο να παραμένουν τα συστήματα ένα βήμα μπροστά από τις εξελιγμένες απειλές στον κυβερνοχώρο, ακόμη και αυτές που μπορεί να οδηγούνται από την TN. Ωστόσο, είναι σημαντικό να ισορροπούν οι ικανότητες της TN με την ανθρώπινη επίβλεψη και να διατηρείται προληπτική και προσαρμοστική κουλτούρα ασφάλειας. Όσο εξελίσσεται η τεχνολογία TN, τόσο οι αμυνόμενοι όσο και οι επιτιθέμενοι θα συνεχίσουν να αναπτύσσουν νέες τακτικές, δημιουργώντας έναν αέναο κύκλο προκλήσεων που αφορά την προστασία των ψηφιακών στοιχείων και δεδομένων.

Θα μπορούσε κάποιος να ισχυριστεί πως στο ερώτημα “*κατά πόσο η TN θεωρείται καλύτερη πρακτική από τις παραδοσιακές μεθόδους άμυνας έναντι των κυβερνο-επιθέσεων*” δεν υπάρχει μια σύντομη και περιεκτική απάντηση, η οποία να ικανοποιεί πλήρως και να καλύπτει όλες τις πτυχές

του ερωτήματος. Είναι σαφές πως η TN πράγματι μπορεί να βοηθήσει σημαντικά στη βελτίωση των επιπέδων άμυνας ενός πληροφοριακού συστήματος και να προστατέψει σε μεγαλύτερο βαθμό τα δεδομένα και τα περιουσιακά στοιχεία των χρηστών, από τις λεγόμενες παραδοσιακές πρακτικές προστασίας. Ωστόσο, η παρουσία του ανθρώπινου παράγοντα φαίνεται πως είναι εξίσου καθοριστική για την αποτελεσματικότερη ανταπόκριση της TN στις κυβερνο-επιθέσεις, μιας και μπορεί να υποπέσει μέχρι και αυτή σε σφάλματα ή να σημειώσει χαμηλή κατανόηση του περιβάλλοντος ασφαλείας στο οποίο ένας οργανισμός επιδιώκει να λειτουργεί.

Επιπλέον, φαίνεται πως η βέλτιστη αμυντική πρακτική περιλαμβάνει το συνδυασμό των δυνάμεων της TN και των παραδοσιακών μεθόδων προστασίας. Η υβριδική αυτή προσέγγιση, αξιοποιεί τους αυτοματισμούς και την ακρίβεια που μπορεί να προσφέρει η TN στον εντοπισμό των διαδικτυακών κινδύνων και στην αντιμετώπισή τους, βελτιώνοντας αντίστοιχες παραδοσιακές μεθόδους, ενώ ταυτόχρονα η εξειδίκευση που προσφέρει ο ανθρώπινος παράγοντας συνοδεύεται από καλύτερη ικανότητα επίβλεψης και κατανόησης των ευρύτερων αναγκών, του πλαισίου λειτουργίας (του οργανισμού ή ενός μεμονωμένου ιδιωτικού συστήματος) και των βέλτιστων αποφάσεων που πρέπει να ληφθούν. Συνεπώς, η προσέγγιση αυτή μπορεί να συμβάλει στην ελαχιστοποίηση των αδυναμιών κάθε προαναφερθείσας μεθόδου και άρα της βελτίωσής της.

Ένα άλλο σημαντικό συμπέρασμα είναι πως οι πρακτικές που συνήθως συναντώνται στα αντίμετρα, έχουν τη βάση τους στη λογική του εντοπισμού μη-φυσιολογικών ενεργειών ή καταστάσεων (anomalies), εντοπισμού των προσπαθειών διάρρηξης των τειχών προστασίας και εντοπισμού των προσπαθειών εξαπάτησης των χρηστών. Για το λόγο αυτό εξάλλου υιοθετείται συχνά η πρακτική της ανάλυσης συμπεριφορών στο διαδίκτυο (behavior analytics), ενώ η μηχανική μάθηση έχει την τιμητική της και στην περίπτωση της άμυνας.

Δύο πρακτικές οδηγούμενες από τη TN φάνηκε να θεωρούνται ιδιαίτερα σημαντικές για την αντιμετώπιση κυβερνοεπιθέσεων. Η μια εξ αυτών στηρίζεται στην αποτελεσματική διαχείριση περιστατικών απειλής σε συνδυασμό με μια ευρεία πληροφόρηση γύρω από ζητήματα ασφάλειας (SIEM) και η δεύτερη αφορά την αποτελεσματική ενορχήστρωση της προστασίας σε συνδυασμό με καθιέρωση αυτοματισμών (orchestration & automations).

Η συνεργασία (μεταξύ υπηρεσιών πληροφόρησης των κρατών ή των εταιρειών προστασίας έναντι κινδύνων στον κυβερνοχώρο) και ο διαμοιρασμός πληροφοριών, θεωρούνται πλέον δύο ζωτικά συστατικά των αντιμέτρων που βασίζονται στην TN. Τόσο οι οργανισμοί όσο και τα κράτη, αναγνωρίζουν την ανάγκη συγκέντρωσης πόρων για περαιτέρω έρευνα ή διάθεσής τους για αναβαθμίσεις υλικού (software & hardware), του διαμοιρασμού πληροφοριών γύρω από τους κινδύνους και της ανάπτυξης τυποποιημένων στρατηγικών άμυνας ενάντια στις απειλές που

στηρίζονται στην ΤΝ. Πρωτοβουλίες, όπως η συνεργασία ανάμεσα στον ιδιωτικό και δημόσιο τομέα, κρίνονται στοιχεία ικανά να αντιμετωπίσουν αποτελεσματικά τις αναδυόμενες απειλές στο πεδίο αυτό. Η χρήση της ΤΝ σε επίπεδο κυβερνοασφάλειας, τόσο ως εργαλείο ενίσχυσης των αμυνών όσο και ως πιθανό όπλο για τους επιτιθέμενους, από μόνη της υπογραμμίζει την ανάγκη για διαρκή εξέλιξη και καινοτομία, θέσπιση κανονιστικών προτύπων και επαγρύπνησης στη σύγχρονη, ψηφιακή εποχή.

Εν τέλει, αξίζει να σημειωθεί - αν κάποιος παρατηρήσει μεμονωμένα το σκέλος της προστασίας από τις απειλές του διαδικτύου - πως η ΤΝ πράγματι μπορεί να βελτιώσει την προστασία ενός συστήματος (ατομικό ή ολόκληρου οργανισμού) από κακόβουλες επιθέσεις, οι οποίες σκοπό έχουν την πρόσβαση σε ευαίσθητες πληροφορίες ή την πρόκληση ζημιών σε περιουσιακά στοιχεία. Οι ικανότητες της ΤΝ ξεπερνούν σε πολλά ζητήματα τις αντίστοιχες των ανθρώπων, ιδίως στο σκέλος της ταχύτητας και της ακρίβειας, μειώνοντας συνήθως τα σφάλματα στα οποία θα μπορούσε πολύ πιο εύκολα να υποπέσει ο άνθρωπος. Επομένως, η ΤΝ είναι ένα εργαλείο που μπορεί να ενισχύσει σημαντικά τα επίπεδα ασφάλειας έναντι των συνεχώς αυξανόμενων απειλών του κυβερνοχώρου, ωστόσο υπάρχει μια λεπτομέρεια που δε θα πρέπει να παραβλέπεται και αυτή περιλαμβάνει τον άνθρωπο. Η συμμετοχή του κρίνεται απαραίτητη και μπορεί να καταστήσει αποδοτικότερη την προστασία των δικτύων, των συστημάτων και των πληροφοριών, αφού όσο και αν υπολείπεται της ΤΝ σε επίπεδο ικανοτήτων, δεν παύει να είναι αυτός που θα αποτρέψει τα δικά της πιθανά σφάλματα.

Βιβλιογραφία

1. Agrawal, A., Gans, J. & Goldfarb, A. (2018). *Prediction machines: The simple economics of Artificial Intelligence*. New York, NY: Harvard Business Press.
2. Ahmed, M. & Uddin, M.N. (2020). Cyber-attack detection method based on NLP and ensemble learning approach (*Proceedings of the 23rd International Conference on Computer & Information*). Dec 19-20, 2020, Dhaka, Bangladesh.
3. Antunes, M., Maximiano, M., Gomes, R. & Pinto, D. (2021). Information Security and Cybersecurity Management: A case study with SMEs in Portugal. *Journal of Cybersecurity & Privacy*, 1: 219-238. DOI: 10.3390/jcp1020012
4. Ardito, C., di Noia, T., di Sciascio, E. et al. (2021). An Artificial Intelligence Cyberattack Detection System to improve threat reaction in e-Health (*Proceedings of the Italian Conference on Cybersecurity*). April 07-09, 2021, Digital event.
5. Atkinson, M.P. & Kress, M. (2023). Resource allocation in two-layered cyber-defense. *Naval Research Logistics*, 70 (6): 574-583.
6. Baum, H. (n.d.). An introduction to Artificial Intelligence. University of Cincinnati. Available at: <https://www.uc.edu/content/dam/uc/ce/docs/OLLI/Page%20Content/ARTIFICIAL%20INTELLIGENCEr.pdf> (Access: Oct 12, 2023).
7. Biggio, B. & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84: 317-331.
8. Bostrom, N. (2016). *Superintelligence: Paths, dangers, strategies*. Oxford, UK: University Press.
9. Buchanan, B., Bansemer, J., Cary, D. et al. (2020). *Automating cyber-attacks: Hype and reality*. Center for Security and Emerging Technology, University of Georgetown. Available at: <https://cset.georgetown.edu/wp-content/uploads/CSET-Automating-Cyber-Attacks.pdf>
10. Canongia, C. & Mandarino, R. (2014). Cybersecurity: The New Challenge of the Information Society (In: *Information Resources Management Association, Crisis Management: Concepts, methodologies, tools and applications: 60-80*). Hershey, PA: IGI Global. <http://dx.doi.org/10.4018/978-1-4666-4707-7.ch003>
11. Charmet, F., Tanuwidjaja, H.C., Ayoubi, S. et al. (2022). Explainable artificial intelligence for cybersecurity: A literature survey. *Annals of Telecommunications*, 77: 789-812.

12. Cisco Systems Inc. (2023a). What is cybersecurity? Available at: <https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html> (Access: Oct 16, 2023).
13. Cisco Systems Inc. (2023b). What is Endpoint Security? Available at: <https://www.cisco.com/c/en/us/products/security/endpoint-security/index.html> (Access: Oct 16, 2023).
14. Colyer, S.L., Evans, M., Cosker, D.P. et al. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Med - Open* 4, 24 (2018). <https://doi.org/10.1186/s40798-018-0139-y>
15. Comiter, M. (2019). *Attacking Artificial Intelligence: AI's security vulnerability and what policymakers can do about it*. Cambridge, MA: University Press.
16. Dash et al., (2022). Threats and opportunities with AI-based cyber security intrusion detection: A review. *International Journal of Software Engineering & Applications*, 13 (5): 13-21. DOI: 10.5121/ijsea.2022.13502
17. de Azambuja, A.J.G., Plesker, C., Schützer, K. et al. (2023). Artificial Intelligence-based cyber security in the context of industry 4.0 - A survey. *Electronics*, 12, 1920. <https://doi.org/10.3390/electronics12081920>
18. Abdul-Ghani, H.A. & Konstantas, D. (2019). A comprehensive study of Security and Privacy Guidelines, threats, and countermeasures: An IoT perspective. *Journal of Sensor & Actuator Networks*, 8 (2): 22.
19. de Witt, C.S., Huang, Y., Torr, P.H.S. & Strohmeier, M. (2021). Fixed points in cyber space: Rethinking optimal evasion attacks in the age of AI-NIDS. *arXiv*, <https://doi.org/10.48550/arXiv.2111.12197>
20. DHH (Department of Health & Human Services) (2016). *Cybersecurity Awareness Training*. Available at: <https://www.hhs.gov/sites/default/files/ocio/securityprivacy/awarenesstraining/cybersecurity-awareness.pdf> (Access: Oct 24, 2023).
21. Dignum, V. (2019). *Responsible Artificial Intelligence: How to develop and use AI in a responsible way*. Berlin: Springer.
22. Khurana, D., Koli, A., Khatte, K. et al. Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl* 82, 3713–3744 (2023). <https://doi.org/10.1007/s11042-022-13428-4>

23. Duddu, V. (2018). A survey of adversarial machine learning in cyber warfare. *Defense Science Journal*, 68 (4): 356-366.
24. El Hadi, M.M. (2023). Artificial Intelligence background, definitions, challenges and benefits. *CompuNet*, 31 (2023): 5-13.
25. European Union Agency for Network & Information Security (2015). *Definition of Cybersecurity - Gaps and overlaps in standardization*. Brussels: ENISA.
26. Fateneh Deldar , Mahdi Abadi (2023). Deep Learning for Zero-day Malware Detection and Classification: A Survey. *ACM Computing Surveys* Volume 56 Issue 2 Article No.: 36, pp 1–37 <https://dl.acm.org/doi/10.1145/3605775>
27. Fei, G. & Zheng, Y. (2022). Security threats and countermeasures of artificial intelligence empowered network attack. *Network Security Technology & Application*, 04 (2022): 18-19.
28. Finnish Transport & Communication Agency (2022). *The security threat of AI-enabled cyberattacks*. Available at: https://www.traficom.fi/sites/default/files/media/publication/TRAFICOM_The_security_threat_of_AI-enabled_cyberattacks%202022-12-12_en_web.pdf (Access: Sep 22, 2023).
29. Frankovits, G. & Mirsky, Y. (2023). Discussion Paper: The threat of real time deepfakes. *arXiv*, <https://arxiv.org/pdf/2306.02487.pdf>
30. Guembe, B., Azeta, A., Misra, S. et al. (2022). The emerging threat of AI-driven cyber-attacks: A review. *Applied Artificial Intelligence*, 36: 1. DOI: 10.1080/08839514.2022.2037254
31. Imamguluyev, R. (2023). The rise of GPT-3: Implications for Natural Language Processing and beyond. *International Journal of Research Publication & Reviews*, 4 (3): 4893-4903.
32. Jeffrey, N., Tan, Q. & Villar, J.R. (2023). A review of anomaly detection strategies to detect threats to cyber-physical systems. *Electronics*, 12, 3283. <https://doi.org/10.3390/electronics12153283>
33. Johnson, C., Badger, L., Waltermire, D. et al. (2016). *Guide to cyber threat information sharing*. Available at: <https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-150.pdf> (Access: Oct 29, 2023).
34. Jung, K. (2021). Extreme data breach losses: An alternative approach to estimating probable maximum loss for data breach risk. *North American Actuarial Journal*, 25 (4): 580-603.
35. Karimipour, H. & Derakhshan, F. (2021). *AI-enabled threat detection and security analysis for industrial IoT*. Berlin: Springer.

36. Karn, A. (2021). Artificial Intelligence in computer vision. *International Journal of Engineering Applied Sciences & Technology*, 6 (1): 249-254.
37. Kaur, H. & Tiwari, R. (2021). Endpoint detection and response using machine learning. *Journal of Physics: Conference Series*, 2062 (2021): 1-8.
38. King, J., Bendiab, G., Savage, N. & Shiaeles, S. (2020). Data exfiltration: Methods and detection countermeasures (*Project funded from the European Union's "Horizon 2020" research and innovation programme under grant agreement no. 833673*). Available at: https://pure.port.ac.uk/ws/portalfiles/portal/42722347/Data_Exfiltration.pdf (Access: Sep 22, 2023).
39. Kovács, L. (2018). National Cyber Security as the cornerstone of National Security. *Land Forces Academy Review*, 23 (11): 120-123. DOI: 10.2478/raft-2018-0013
40. Lauter, K.E. (2021). *Private AI: Machine Learning on encrypted data*. Available at: <https://eprint.iacr.org/2021/324.pdf> (Access: Oct 21, 2023).
41. Li, J.H. (2018). Cyber security meets artificial intelligence: A survey. *Frontiers of Information Technology & Electronic Engineering*, 19 (12): 1462–1474.
42. Martin, S. & Tokutomi, M. (2012). *Password cracking*. Available at: <https://www2.cs.arizona.edu/~collberg/Teaching/466-566/2012/Resources/presentations/topic7-final/report.pdf> (Access: Sep 22, 2023).
43. Matt, D.T., Modrák, V. & Zsifkovits, H. (2020). *Industry 4.0 for SMEs: Challenges, opportunities and requirements*. Cham: Springer.
44. Ministry of Digital Governance Greece (2021). *Cybersecurity handbook*. Available at: <https://mindigital.gr/wp-content/uploads/2022/09/Cybersecurity-Handbook-English-version.pdf> (Access: Oct 14, 2023).
45. Moore, R. (n.d.). *Cyber Security Awareness Seminar*. University of Cincinnati. Available at: https://www.uc.edu/content/dam/uc/ce/docs/OLLI/Page%20Content/OCRI_Cyber%20Security%20Awareness%20Seminari%20Presentation.pdf (Access: Oct 20, 2023).
46. Neyigapula, B.S. (2023). Secure AI model sharing: A cryptographic approach for encrypted model exchange. *Research Square*. DOI: <https://doi.org/10.21203/rs.3.rs-3210577/v1>
47. Nielsen, M. (2019). *How the backpropagation algorithm works*. Available at: (Access: Sep 05, 2023).
48. OECD (Organization for Economic Cooperation & Development) (2016). *OECD Science, Technology and Innovation Outlook*. Paris: OECD.

49. Office of Information Security (2023). *Cybersecurity Incident Response Plans*. Available at: <https://www.hhs.gov/sites/default/files/cybersecurity-incident-response-plans.pdf> (Access: Oct 26, 2023).
50. Olaniyan, R., Rakshit, S. & Vajjhala, N.R. (2023). Application of user and entity behavioral analytics (UEBA) in the detection of cyber threats and vulnerabilities management (*Proceedings of the 9th International Conference on Computational Science & Technology*). 27-28 August, Johor Bahru, Malaysia.
51. Pan, Q., Dong, M., Ota, K. & Wu, J. (2022). Device-bind key-storageless hardware AI model IP Protection: A PUF and permute-diffusion encryption-enabled approach. *arXiv*. <https://doi.org/10.48550/arXiv.2212.11133>
52. Petratos, P.N. (2021). Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons*, 64 (6): 763-774.
53. Popescul, D. (2011). The Confidentiality - Integrity - Accessibility Triad into the knowledge security. A reassessment from the point of view of the knowledge contribution to innovation (*Proceedings of the 16th International Business Information Management Association Conference*). June 29-30, 2011, Kuala Lumpur, Malaysia.
54. Rabii, A., Assoul, S., Touhami, K.O. & Roudies, O. (2020). Information and cyber security maturity models: A systematic literature review. *Information & Computer Security*, 28 (4): 627-644.
55. Rid, T. (2016). *Rise of the machines: A cybernetic history*. London, UK: WW Norton & Co.
56. Russell, S.J. & Norving, P. (2010). *Artificial Intelligence: A modern approach* (3rd edn). Upper Saddle River, NJ: Prentice Hall.
57. Salas-Pilco, S.Z. & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 21 (2022). <https://doi.org/10.1186/s41239-022-00326-w>
58. Security Standards Council (2017). *Information Supplement: Penetration Testing Guidance*. Available at: https://listings.pcisecuritystandards.org/documents/Penetration-Testing-Guidance-v1_1.pdf (Access: Oct 26, 2023).
59. Seemma, P.S., Nandhini, S. & Sowmiya, M. (2018). Overview of Cyber Security. *International Journal of Advanced Research in Computer & Communication Engineering*, 7 (11): 125-128. DOI: 10.17148/IJARCCE.2018.71127

60. Shiri, F.M., Perumal, T., Mustapha, N. & Mohamed, R. (2023). A comprehensive overview and comparative analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. *ArXiv*, DOI: 10.48550/arXiv.2305.17473
61. Smith, M.R., Verzi, S.J., Johnson, T.N. et al. (2021). Malware Generation with specific behaviors to improve machine learning-based detection (*Proceedings of the 2021 IEEE International Conference on Big Data*). Dec 15-18, 2021, Orlando, FL, USA. DOI: 10.1109/BigData52589.2021.9671886
62. Stivala, G., Abdelnabi, S., Mengascini, A. et al. (2023). A large-scale study of phishing PDF documents. *arXiv*, <https://doi.org/10.48550/arXiv.2308.01273>
63. Taherdoost, H. (2022). Cybersecurity vs. Information Security. *Procedia Computer Science*, 215 (2022): 483-487.
64. Taherdoost, H., Sahibuddin, S. & Jalaliyoon, N. (2015). How security issues can influence on usage of electronic services. *Advances in Information Science & Computer Engineering*, 310-316.
65. Vaisla, K.S. & Saini, R. (2014). Analyzing of zero-day attack and its identification techniques (*Proceedings of the 1st International Conference on Advances in Computing & Communication Engineering*). Oct 25-26, 2014, Dwarahat, India.
66. Valdes, Y., Hochstetter, J., Diaz-Arancibia, J. & Martinez, R.C. (2023). Towards the integration of Security Practices in agile software development: A systematic mapping review. *Applied Sciences*, 13 (7): 4578. DOI: 10.3390/app13074578
67. Tissir, N., El Kafhali, S. & Aboutabit, N. (2021). Cybersecurity management in cloud computing: semantic literature review and conceptual framework proposal. *Journal of Reliable Intelligent Environments*, 7: 69-84. <https://doi.org/10.1007/s40860-020-00115-0>
68. Vast, R., Sawant, S., Thorbole, A. & Badgujar, V.S. (2021). Artificial Intelligence based Security Orchestration, Automation and Response System (*Proceedings of the 6th International Conference for Convergence in Technology*). 02-04 April, Mumbai, India.
69. Wang, G. & Wendy, K. (2019). Measuring information security and cybersecurity on private cloud computing. *Journal of Theoretical & Applied Information Technology*, 96 (1): 156-168.
70. Wang, Y., Sun, T., Yuan, X. et al. (2023). Adversarial attacks and defenses in machine learning-powered networks: A contemporary survey. Available at: <https://arxiv.org/pdf/2303.06302.pdf> (Access: Sep 29, 2023).

70. World Economic Forum (2022). *The “zero trust” model in cybersecurity: Towards understanding and deployment (Community Paper)*. Available at: https://www3.weforum.org/docs/WEF_The_Zero_Trust_Model_in_Cybersecurity_2022.pdf (Access: Oct 29, 2023).
71. Yamin, M.M., Ullah, M., Ullah, M. & Katt, B. (2021). Weaponized AI for cyber-attacks. *Journal of Information Security & Applications*, 57. <https://doi.org/10.1016/j.jisa.2020.102722>
72. Y. Xin *et al.*, "Machine Learning and Deep Learning Methods for Cybersecurity," in *IEEE Access*, vol. 6, pp. 35365-35381, 2018, doi: 10.1109/ACCESS.2018.2836950.
73. Zeng, Y. (2022). AI empowers security threats and strategies for cyber-attacks. *Procedia Computer Science*, 208 (2022): 170-175.
74. Zhangyi, L. & Gang, L. (2022). Risk prevention and control of network ideology empowered by artificial intelligence: background, direction and approach. *Contemporary Communication*, 03 (2022): 56-58.
75. Ευρωπαϊκή Επιτροπή (2018). *Ανακοίνωση της επιτροπής προς το Ευρωπαϊκό Κοινοβούλιο, το Ευρωπαϊκό Συμβούλιο, το Συμβούλιο, την Ευρωπαϊκή Οικονομική και Κοινωνική Επιτροπή και την Επιτροπή των Περιφερειών*. Διαθέσιμο στο: <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52018DC0237> (Πρόσβαση: 01 Σεπ 2023).