ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑΤΟΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΤΛΗΣΗ ΠΛΗΡΟΦΟΡΙΩΝ ΓΙΑ ΚΥΒΕΡΝΟ-ΑΠΕΙΛΕΣ ΑΠΟ ΤΟ ΣΚΟΤΕΙΝΟ ΔΙΑΔΙΚΤΥΟ

Διπλωματική Εργασία

του

Τουμπόγλου Ιωάννη

Θεσσαλονίκη, Φεβρουάριος 2019

ΑΝΤΛΗΣΗ ΠΛΗΡΟΦΟΡΙΩΝ ΓΙΑ ΚΥΒΕΡΝΟ-ΑΠΕΙΛΕΣ ΑΠΟ ΤΟ ΣΚΟΤΕΙΝΟ ΔΙΑΔΙΚΤΥΟ

Τουμπόγλου Ιωάννης

Πτυχίο Πληροφορικής, ΕΑΠ, 2010

Διπλωματική Εργασία

υποβαλλόμενη για τη μερική εκπλήρωση των απαιτήσεων του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ

Επιβλέπων καθηγητής
Μαυρίδης Ιωάννης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την ___/___/_____

Μαυρίδης Ιωάννης	Φουληράς Παναγιώτης	Ψάννης Κωνσταντίνος
.....	.....	.....

Τουμπόγλου Ιωάννης

.....

Περίληψη

Τα τελευταία χρόνια λαμβάνει μέρος μια εντυπωσιακή ανάπτυξη στους τομείς της πληροφορικής και των τηλεπικοινωνιών, η οποία αποσκοπεί στην παροχή υπηρεσιών που θα βελτιώσουν τη ζωή του σύγχρονου ανθρώπου. Όμως, όπως σε όλους τους τομείς, έτσι και εδώ εντοπίζονται περιστατικά κακής χρήσης, τα οποία δε σέβονται την ηθική και τη νομιμότητα και έχουν αρνητικές επιπτώσεις. Έννοιες όπως το κυβερνο-έγκλημα και οι κυβερνο-επιθέσεις βλέπουν ολοένα και πιο συχνά τα φώτα της δημοσιότητας και τείνουν να αποτελέσουν μέρος της καθημερινότητάς μας.

Οι κακόβουλοι χρήστες κάνουν χρήση εκείνου του μέρους του παγκοσμίου ιστού που αγνοεί ο μέσος χρήστης, αποτελεί τμήμα του βαθύ ιστού και είναι ευρύτερα γνωστό ως σκοτεινό διαδίκτυο. Αυτό τους προσφέρει ελευθερία κινήσεων, αφού χαρακτηρίζεται από ανωνυμία, εξυπηρετώντας την μεταξύ τους οργάνωση, την ανταλλαγή πληροφοριών και την αποφυγή εντοπισμού από τις αρχές. Κύριος εκπρόσωπός του είναι το δίκτυο Tor, το οποίο βασίζεται στην ασφαλή επικοινωνία μέσω κρυπτογραφημένων καναλιών και στην προστασία της ταυτότητας των χρηστών από τρίτους.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η διερεύνηση της δομής και του τρόπου λειτουργίας του Tor και η δημιουργία ενός εργαλείου για την άντληση πληροφοριών γύρω από τα διάφορα είδη κυβερνο-απειλών. Το αποτέλεσμα είναι η δημιουργία ενός web crawler με χρήση της γλώσσας προγραμματισμού Python, ο οποίος εντοπίζει, συλλέγει και αποθηκεύει σε βάση δεδομένων κείμενα, που περιέχονται σε ιστοσελίδες του σκοτεινού διαδικτύου, με στόχο την περαιτέρω επεξεργασία τους. Η υλοποίησή του αποτελείται από τρία τμήματα: τη δημιουργία μιας γενικής λίστας διευθύνσεων onion url, τη χρήση διάφορων τεχνικών αναζήτησης σε ένα υποσύνολο αυτών για τη συλλογή κειμένων, και τέλος, τη χρήση μηχανικής μάθησης χωρίς επίβλεψη για την ταξινόμησή τους. Την λεπτομερή παρουσίαση της αρχιτεκτονικής της εφαρμογής ακολουθούν η ανάλυση των βημάτων εκτέλεσης όλων των λειτουργιών της και τα τελικά αποτελέσματα της έρευνας.

Λέξεις Κλειδιά: Tor, Tor project, σκοτεινό διαδίκτυο, κρυμμένες υπηρεσίες, cyber threat intelligence, onion routing, κυβερνο-απειλή, κυβερνο-επίθεση, κυβερνο-έγκλημα, κακόβουλο λογισμικό, web crawler

Abstract

In recent years, an impressive growth in the fields of information technology and telecommunications occurs, which aims to provide services that will improve the lives of modern people. Unfortunately, cases of misuse are recorded, which do not respect ethics and legitimacy and may have a negative impact in many areas. It has been observed that terms like cyber-crime and cyber-attacks are increasingly used and tend to be part of our everyday life.

Malicious users take advantage of a section of the deep web, the existence of which is ignored by the average user, and is widely known as the dark web. This part of the web provides them with online freedom, as it is characterized by anonymity, helping them to organize themselves, exchange information and avoid detection by the authorities. Its main representative is the Tor network, which utilizes secure communications via encrypted channels, offering identity protection from third party users.

The aim of this thesis is to examine Tor's structure and operation and to create a tool that gathers information concerning various cyber-threat types. The result is a web crawler, developed in the Python programming language, which detects and collects the texts contained in dark web pages, storing them in a database for further processing. Its implementation is divided into three stages: creating a general url address list, collecting texts from a subset of these urls with the use of a variety of searching techniques, and finally, categorizing them using unsupervised machine learning. The detailed presentation of the application's architecture is followed by the analysis of the steps that take place while performing the data collection functions, concluding with the final results of the research.

Keywords: Tor, Tor project, dark web, hidden services, cyber threat intelligence, onion routing, cyber threat, cyber attack, cyber crime, malicious software, web crawler

Πρόλογος – Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μαυρίδη Ιωάννη για την εμπιστοσύνη που μου έδειξε και για την καθοδήγησή του καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα κ. Ράπτη Σωτήριο για τις πολύτιμες συμβουλές και την υποστήριξή του στην προσπάθειά μου. Ένα μεγάλο ευχαριστώ στην οικογένειά μου για την πίστη και την αμέριστη συμπαράστασή τους στην επιδίωξη των προσωπικών μου στόχων όλα αυτά τα χρόνια.

Περιεχόμενα

1. Εισαγωγή 1
1.1 Περιγραφή προβλήματος 1
1.2 Σκοπός - Στόχοι 2
1.3 Διάρθρωση της μελέτης 3
2. Θεωρητικό Υπόβαθρο 4
2.1 Το Διαδίκτυο και ο Παγκόσμιος Ιστός 4
2.2 Ο Βαθύς Ιστός 5
2.3 Ο Σκοτεινός Ιστός 6
2.3.1 Ο Σκοτεινός Ιστός και τα σκοτεινά δίκτυα 6
2.3.2 Onion Routing 7
2.3.2 Tor Project 12
2.3.3 Invisible Internet Project 22
2.3.4 Freenet 23
2.3.5 Συμπεράσματα 25
2.4 Κατηγορίες κυβερνο-επιθέσεων 27
2.5 Τύποι κυβερνο-επιθέσεων 28
2.5.1 Malware (Malicious Software) 28
2.5.2 Phishing 30
2.5.3 Man-in-the-middle (MITM) 31
2.5.4 Denial-of-service (DoS) 32
2.5.5 SQL Injection 34
2.5.6 Zero-day Exploit 34
2.5.7 Cross-site Scripting (XSS) 35
2.5.8 Credential reuse ή stuffing 36
2.6 Cyber Threat Intelligence 36
3. Βιβλιογραφική Επισκόπηση 38
4. Μεθοδολογία 41
4.1 Εισαγωγή 41
4.2 Βιβλιοθήκες 41
4.2.1 Requests 41
4.2.2 BeautifulSoup 41

4.2.3 Urlparse 42
4.2.4 Socks 42
4.2.5 Socket 42
4.2.6 Stem 42
4.2.7 Fake_useragent 42
4.2.8 Sqlite3 43
4.2.9 Sklearn 43
4.3 Σχεδίαση εφαρμογής 43
4.3.1 Δημιουργία λίστας 43
4.3.2 Λειτουργίες συλλογής δεδομένων 44
4.3.3 Κατηγοριοποίηση 45
4.3.4 Δομή βάσης δεδομένων 46
4.3.5 Επιπλέον λειτουργικά χαρακτηριστικά 47
4.4 Βοηθητικό λογισμικό 48
4.4.1 Tor browser 48
4.4.2 VPN 49
5. Επίδειξη Λειτουργίας 51
5.1 Εισαγωγή 51
5.2 Δημιουργία λίστας διευθύνσεων 52
5.2.1 Μεμονωμένη συλλογή συνδέσμων 52
5.2.2 Σειριακή συλλογή συνδέσμων 56
5.3 Συλλογή δεδομένων 58
5.3.1 Τυχαία συλλογή δεδομένων 58
5.3.2 Σειριακή συλλογή δεδομένων 60
5.3.3 Συλλογή δεδομένων κατά πλάτος 61
5.3.4 Εξαγωγή κειμένου 63
5.3.5 Κατηγοριοποίηση κειμένου 65
6. Επίλογος 68
6.1 Σύνοψη και συμπεράσματα 68
6.2 Όρια και περιορισμοί της έρευνας 70
6.3 Μελλοντικές επεκτάσεις 70
7. Βιβλιογραφία 72

Κατάλογος Εικόνων

Εικόνα 2-1: Διακομιστής μεσολάβησης 7
Εικόνα 2-2: Η τοπολογία του onion routing 8
Εικόνα 2-3: Επίπεδα κρυπτογράφησης δεδομένων 10
Εικόνα 2-4: Κίνηση δεδομένων 10
Εικόνα 2-5: Δημιουργία σημείων εισαγωγής 17
Εικόνα 2-6: Δήλωση περιγραφέα υπηρεσίας 18
Εικόνα 2-7: Ανάκτηση περιγραφέα υπηρεσίας και επιλογή σημείου συνάντησης 19
Εικόνα 2-8: Αίτημα σύνδεσης 19
Εικόνα 2-9: Σύνδεση του εξυπηρετητή στο σημείο συνάντησης 20
Εικόνα 2-10: Εγκαθίδρυση σύνδεσης μέσω του σημείου συνάντησης 21
Εικόνα 2-11: Πλήθος χρηστών στο δίκτυο Tor 21
Εικόνα 2-12: Πλήθος κόμβων στο δίκτυο Tor 22
Εικόνα 2-13: Πλήθος διευθύνσεων onion στο δίκτυο Tor 22
Εικόνα 2-14: Αναπαράσταση της δομής του παγκόσμιου ιστού 25
Εικόνα 4-1: Προβολή στοιχείων κυκλώματος 49
Εικόνα 5-1: Αρχεία εφαρμογής 52
Εικόνα 5-2: Αρχικές επιλογές εφαρμογής 52
Εικόνα 5-3: Εκτέλεση συλλογής συνδέσμων 53
Εικόνα 5-4: Επιλογές κατόπιν ολοκλήρωσης συλλογής συνδέσμων 54
Εικόνα 5-5: Συνέχεια εκτέλεσης συλλογής στην ίδια βάση δεδομένων 54
Εικόνα 5-6: Προβολή βάσης δεδομένων 55
Εικόνα 5-7: Παράδειγμα ομάδας σελίδων 56
Εικόνα 5-8: Εύρεση και παράκαμψη συνδέσμου 57
Εικόνα 5-9: Σειριακή συλλογή συνδέσμων 57
Εικόνα 5-10: Τυχαία συλλογή δεδομένων 58
Εικόνα 5-11: Εύρεση νέας σελίδας κατόπιν ελέγχου της βάσης δεδομένων 59
Εικόνα 5-12: Συνέχεια τυχαίας συλλογής δεδομένων 60
Εικόνα 5-13: Σειριακή συλλογή δεδομένων 61
Εικόνα 5-14: Επίπεδα ιστότοπου 62
Εικόνα 5-15: Συλλογή δεδομένων κατά πλάτος 63
Εικόνα 5-16: Εξαγωγή κειμένου 64

Εικόνα 5-17: Παράδειγμα αποθηκευμένου κειμένου 65
Εικόνα 5-18: Συστάδες με λέξεις-κλειδιά 66
Εικόνα 5-19: Κατηγοριοποίηση κειμένου 66
Εικόνα 5-20: Κατανομή κυβερνο-επιθέσεων 67
Εικόνα 6-1: Ιστοσελίδα με μήνυμα περί απάτης 69

1. Εισαγωγή

1.1 Περιγραφή προβλήματος

Τα τελευταία χρόνια είμαστε μάρτυρες της εντυπωσιακής ανάπτυξης στους τομείς της πληροφορικής και των τηλεπικοινωνιών, μεγάλο ποσοστό της οποίας αφορά την εξέλιξη των διαδικτυακών τεχνολογιών. Όλα αυτά αποσκοπούν στην απλοποίηση και την αυτοματοποίηση των καθημερινών διαδικασιών για τη βελτίωση της ζωής του σύγχρονου ανθρώπου. Όπως όμως σε όλους τους τομείς, έτσι και εδώ μπορούν να εντοπιστούν περιστατικά κακής χρήσης, που έχουν ως στόχο το προσωπικό όφελος και το χρηματικό κέρδος, δίχως να λαμβάνεται υπόψη η νομιμότητα και οι επιπτώσεις που θα έχουν τέτοιου είδους ενέργειες στην κοινωνία και στον άνθρωπο. Η ολοένα και μεγαλύτερη εξάρτηση της κοινωνίας από την τεχνολογία της πληροφορίας, σε συνδυασμό με τις απειλές που έχουν κάνει την εμφάνισή τους, οδήγησαν στην εμφάνιση όρων όπως η κυβερνο-επίθεση, η κυβερνο-απειλή και ο κυβερνο-πόλεμος.

Ο αριθμός των περιστατικών, που βλέπουν το φως της δημοσιότητας και έχουν ως θέμα τις κυβερνο-επιθέσεις, αυξάνεται συνεχώς. Μέσα στο 2017, έγινε γνωστό πως δημοσιοποιήθηκε και ήταν διαθέσιμο διαδικτυακά ένα πλήθος διαδικτυακών εργαλείων, τα οποία κατάφερε να υποκλέψει από την Εθνική Υπηρεσία Ασφαλείας των ΗΠΑ (NSA) η ομάδα hackers με το όνομα Shadow Brokers. Αυτά τα εργαλεία αποτελούν απειλή για τα περισσότερα συστήματα με λειτουργικό σύστημα Windows, καθώς και για το τραπεζικό σύστημα SWIFT που λειτουργεί σε μεγάλο αριθμό τραπεζών παγκοσμίως. Τα σενάρια για το ποιος ή ποιοι είναι οι Shadow Brokers είναι πολλά, αφού δεν έχουν εντοπιστεί και δεν έχει γίνει ακόμα γνωστή η ταυτότητά τους [1]. Μεγάλη έκταση δόθηκε τον Μάιο του 2017 στο κακόβουλο λογισμικό WannaCry, το οποίο ήταν τύπου ransomware και απαιτούσε από τους χρήστες την καταβολή του ποσού των $300 σε bitcoins, προκειμένου να αποκρυπτογραφηθούν τα αρχεία τους [2]. Η πιο ευρεία επίθεση DDoS που έχει καταγραφεί πραγματοποιήθηκε τον Μάρτιο του 2018 και είχε ως στόχο τον ιστοτόπο Github, δημιουργώντας στην κορύφωσή της δικτυακή κίνηση 1,35 Terabits ανά δευτερόλεπτο [3].

Το φαινόμενο των κυβερνο-επιθέσεων αποτελεί πλέον καθημερινό φαινόμενο και πρόκειται για τον τύπο εγκλήματος με την ταχύτερη αύξηση, τόσο σε μέγεθος, όσο και σε επιτήδευση. Για αυτό, μέσα στις επόμενες δύο δεκαετίες θα αποτελεί μια από τις μεγαλύτερες προκλήσεις που θα έχει αντιμετωπίσει ποτέ η ανθρωπότητα. Υπολογίζεται

πως μέχρι το 2021, το κόστος που θα επιφέρουν αυτές οι απειλές στον κόσμο θα κυμαίνεται περίπου στα 6 τρισεκατομμύρια δολάρια, ποσό διπλάσιο σε σχέση με εκείνο που δαπανήθηκε το 2015 [4]. Σύμφωνα με τον John Chambers, πρώην διοικητικό στέλεχος της Cisco: "Υπάρχουν δύο κατηγορίες επιχειρήσεων. Είναι εκείνες που έχουν δεχθεί κάποιου είδους κυβερνο-επίθεση και εκείνες που έχουν δεχθεί επίθεση και απλά δεν το γνωρίζουν" [5]. Μπορεί να ακούγεται μη ρεαλιστικό, λαμβάνοντας όμως υπόψη το γεγονός ότι ο αριθμός των επιθέσεων παρουσιάζει αυξητικές τάσεις χρόνο με το χρόνο, τελικά δεν απέχει πάρα πολύ από την πραγματικότητα. Μόνο στις ΗΠΑ, συγκρίνοντας τα στατιστικά μεταξύ των ετών 2016 και 2017, ο αριθμός των παραβιάσεων σε δεδομένα αυξήθηκε κατά 44,5%, ενώ εκείνος των δεδομένων που έχουν εκτεθεί σχεδόν πενταπλασιάστηκε [6].

Όπως μπορεί να γίνει αντιληπτό, οι επιθέσεις που υλοποιούνται με χρήση του Διαδικτύου μπορούν να έχουν σοβαρές συνέπειες και οικονομικές επιπτώσεις, είτε σε ατομικό επίπεδο, απειλώντας τα προσωπικά δεδομένα των χρηστών, είτε σε επίπεδο εταιρειών και οργανισμών, απειλώντας τη λειτουργία και τις υποδομές τους, είτε σε εθνικό επίπεδο, απειλώντας την εθνική ασφάλεια ενός κράτους. Για τους λόγους αυτούς, επιβάλλεται η εύρεση και υλοποίηση μεθόδων για την έγκαιρη και άμεση αντιμετώπισή τους, συμβάλλοντας στον τομέα που είναι γνωστός ως cyber threat intelligence.

1.2 Σκοπός - Στόχοι

Τα στατιστικά στοιχεία δείχνουν πως ο αριθμός των χρηστών του διαδικτύου παρουσιάζει μια διαρκή αύξηση και εκτιμάται πως στα μέσα του 2018 έχει ξεπεράσει τα 4 δισεκατομμύρια [7]. Η ραγδαία αυτή μεταβολή σημαίνει παράλληλη αύξηση του αριθμού των συστημάτων που μπορεί να παρουσιάσουν κενά ασφαλείας, γεγονός που δημιουργεί το πρόσφορο έδαφος για εκμετάλλευση από τους κακόβουλους χρήστες. Έχοντας την κατάλληλη τεχνογνωσία, τόσο γύρω από το υλικό, όσο και από το λογισμικό των συστημάτων, επιδίδονται στην ανάπτυξη τεχνικών, μέσω των οποίων θα καταφέρουν τελικά να παρακάμψουν τα επίπεδα ασφαλείας και να διαπράξουν κυβερνο-εγκλήματα.

Σκοπός της εργασίας είναι η ολοκληρωμένη παρουσίαση του προβλήματος, η οποία είναι αποτέλεσμα της εκτενούς μελέτης της σχετικής βιβλιογραφίας, καθώς και η διερεύνηση και η ανάπτυξη τεχνικών, που θα συμβάλλουν στην προσπάθεια συλλογής πληροφοριών για αυτού του είδους τις απειλές. Σε αυτήν την προσπάθεια, ως πηγή θα

χρησιμοποιηθούν αποκλειστικά ιστοσελίδες που είναι μέρος του σκοτεινού διαδικτύου και συγκεκριμένα του δικτύου Tor, γνωστές ως κρυμμένες υπηρεσίες (hidden services).

1.3 Διάρθρωση της μελέτης

Στο κεφάλαιο 2, αρχικά θα γίνει μια ανασκόπηση των εννοιών που σχετίζονται με τον παγκόσμιο (world wide web), τον βαθύ (deep web) και τον σκοτεινό ιστό (dark web), παρουσιάζοντας το onion routing, πάνω στο οποίο βασίζεται η φιλοσοφία του δικτύου Tor, του αντικειμένου μελέτης της εργασίας. Έπειτα από μια σύντομη αναφορά στα υπόλοιπα δίκτυα ανωνυμίας, τα οποία εξυπηρετούν παρόμοιο σκοπό με το Tor, θα γίνει μια προσπάθεια αποσαφήνισης της έννοιας του cyber threat intelligence, του πεδίου που ασχολείται με τη συλλογή πληροφοριών και έχει ως στόχο τόσο την πρόληψη, όσο και την αντιμετώπιση των απειλών που προέρχονται από το διαδίκτυο. Η βιβλιογραφική αναφορά ολοκληρώνεται με την αναλυτική παρουσίαση των κατηγοριών και των βασικότερων τύπων κυβερνο-επιθέσεων που έχουν καταγραφεί.

Το κεφάλαιο 3 περιλαμβάνει μια αναφορά σε προηγούμενες εργασίες που έχουν συντελεστεί πάνω στη συλλογή πληροφοριών από το σκοτεινό διαδίκτυο, οι οποίες ασχολούνται με τη μελέτη του δικτύου και με την εύρεση μεθοδολογιών, που να την καθιστούν όσο το δυνατόν πιο αποτελεσματική.

Στο κεφάλαιο 4 περιγράφεται η αρχιτεκτονική της εφαρμογής που αναπτύχθηκε στα πλαίσια της εργασίας, με σκοπό τη συλλογή, αποθήκευση και επεξεργασία πληροφοριών από το δίκτυο Tor, ενώ στο κεφάλαιο 5 γίνεται αναλυτική παρουσίαση της λογικής και των λειτουργιών που υλοποιήθηκαν.

Τέλος, το κεφάλαιο 6 θα ολοκληρώσει την έρευνα με την καταγραφή των αποτελεσμάτων και των συμπερασμάτων, προκειμένου να γίνουν προτάσεις για περαιτέρω διερεύνηση του προβλήματος.

2. Θεωρητικό Υπόβαθρο

2.1 Το Διαδίκτυο και ο Παγκόσμιος Ιστός

Το διαδίκτυο (internet) και ο παγκόσμιος ιστός αποτελούν συχνά δύο ταυτόσημες έννοιες, ενώ στην πραγματικότητα αφορούν δύο ξεχωριστά επίπεδα αυτού που ο χρήστης βιώνει ως "διαδίκτυο".

Ξεκινώντας από το πρώτο, το οποίο είναι γνωστό και ως "το δίκτυο των δικτύων", πρόκειται για την παγκόσμια δικτυακή υποδομή, μέσω της οποίας είναι εφικτή η σύνδεση συσκευών για τη μεταφορά πληροφοριών. Τα δίκτυα έχουν διαφορετικά μεγέθη και μορφές, όπως είναι τα τοπικά δίκτυα, τα δίκτυα ευρείας περιοχής ή τα δίκτυα κινητής τηλεφωνίας. παρέχοντας τη δυνατότητα σε μια τεράστια ποικιλία συσκευών, όπως υπολογιστές, τηλεοράσεις τελευταίας τεχνολογίας, αισθητήρες, κοκ. να συνδεθούν μεταξύ τους [8].

Ο παγκόσμιος ιστός αποτελεί μια υπηρεσία του διαδικτύου, η οποία παρέχει στους χρήστες την πρόσβαση σε πληθώρα πληροφοριών και η οποία ακολουθεί το μοντέλο πελάτη-εξυπηρετητή. Στους εξυπηρετητές υπάρχουν αποθηκευμένα έγγραφα υπερκειμένου που διατηρούν συνδέσεις μεταξύ τους και τα οποία περιέχουν πληροφορίες σε διάφορες μορφές, όπως φωτογραφίες, βίντεο, κτλ. Όλα αυτά είναι προσβάσιμα μέσω κατάλληλου λογισμικού, όπως είναι τα προγράμματα περιήγησης, και στόχος είναι η διαθεσιμότητα και η παροχή αυτού του τεράστιου όγκου πληροφοριών σε παγκόσμια κλίμακα [9].

Το μεγαλύτερο ποσοστό των χρηστών θεωρεί ως παγκόσμιο ιστό αυτό στο οποίο έχει άμεση πρόσβαση μέσω ενός απλού περιηγητή με τη χρήση μηχανών αναζήτησης, αγνοώντας το πραγματικό του μέγεθος και την ύπαρξη επιπέδων στη δομή του. Αυτό που ο απλός χρήστης θεωρεί "διαδίκτυο", δεν είναι παρά μόνο ένα πολύ μικρό ποσοστό του συνολικού ιστού και είναι γνωστό ως επιφανειακός ιστός (surface web). Αξίζει να σημειωθεί πως το πλήθος των ιστοσελίδων που τον απαρτίζουν μεταβάλλεται, παρουσιάζοντας μια συνεχή αύξηση και υπολογίζεται ότι ο αριθμός τους εν έτει 2018 πλησιάζει τα δύο δισεκατομμύρια [10].

Υπάρχει όμως ακόμα ένα επίπεδο, μέρος του παγκόσμιου ιστού, το οποίο δεν είναι προσβάσιμο μέσω των κλασσικών μηχανών αναζήτησης, αφού οι σελίδες από τις οποίες αποτελείται δεν είναι καταχωρημένες σε ευρετήριο. Κάποιες εκτιμήσεις θεωρούν

πως μόλις το 10% του παγκόσμιου ιστού αποτελεί τον επιφανειακό ιστό και πως το υπόλοιπο 90% απαιτεί κάποιου είδους ειδικό τρόπο πρόσβασης.

2.2 Ο Βαθύς Ιστός

Ο βαθύς ιστός ή αλλιώς κρυμμένος ιστός (hidden web) είναι το τμήμα του παγκοσμίου ιστού, το οποίο δεν είναι δυνατόν να ανακτηθεί μέσω των γνωστών μηχανών αναζήτησης, με αποτέλεσμα να παραμένει "αόρατο" στον μέσο χρήστη. Το μεγαλύτερο μέρος του περιλαμβάνει αβλαβή στοιχεία, όπως βάσεις δεδομένων, ακαδημαϊκά δεδομένα, επιστημονικές αναφορές, ιατρικά και οικονομικά στοιχεία, εμπιστευτικά έγγραφα. Χρησιμοποιείται κυρίως από δημόσιους και ιδιωτικούς οργανισμούς, κυβερνητικές υπηρεσίες, ερευνητές και δημοσιογράφους. Η διακίνηση πληροφοριών και η χρήση πόρων που να χαρακτηρίζεται από την ιδιωτικότητα, τη διακριτικότητα και την ασφάλεια, αποτελούν μερικούς από τους λόγους που οδήγησαν στη δημιουργία του [11].

Οι κυριότεροι λόγοι, που καθιστούν αδύνατη την καταχώρηση όλου αυτού του όγκου δεδομένων σε ευρετήρια, σχετίζονται με τη φύση των δεδομένων. Δεν είναι εφικτή η καταγραφή του περιεχομένου των βάσεων δεδομένων ή των σελίδων που απαιτούν κάποιον ειδικό τρόπο πρόσβασης, αφού αυτά τα χαρακτηριστικά έρχονται σε αντίθεση με τους κατασκευαστικούς περιορισμούς που διέπουν τις μηχανές αναζήτησης. Οι βάσεις δεδομένων έχουν δυναμική δομή, τα ερωτήματα με τα οποία μπορείς κανείς να ανακτήσει δεδομένα δεν συμβαδίζουν με την αυτοματοποιημένη λειτουργία των μηχανών, οπότε δεν είναι εύκολο να πραγματοποιηθεί καταγραφή τους. Σε αυτό έρχεται να προστεθεί το πλήθος διαδικτυακών υπηρεσιών που, είτε απαιτούν κάποιου είδους συνδρομή προκειμένου να αποκτήσει κανείς πρόσβαση στο περιεχόμενό τους, είτε κάποιου είδους ενέργεια από τον χρήστη, η οποία δεν μπορεί να αυτοματοποιηθεί [12].

Πολλές ιστοσελίδες διαθέτουν πλούσιο περιεχόμενο, που σημαίνει πως αποτελούνται από μεγάλο αριθμό τμημάτων. Οι εταιρείες που διαθέτουν μηχανές αναζήτησης, όπως η Google και η Yahoo!, δεν αφήνουν το λογισμικό τους να ελέγξει όλα αυτά τα τμήματα, αλλά θέτουν κάποιους περιορισμούς σχετικά με το βάθος στο οποίο θα φτάσουν κατά τον έλεγχο, πραγματοποιώντας μία σχετικά επιφανειακή σάρωση. Επιπρόσθετα, είναι πολλές οι σελίδες των οποίων το περιεχόμενο είναι παλαιού τύπου, οπότε συνήθως παραβλέπεται κατά τον έλεγχο και δεν καταγράφεται [12].

Ένα επιπλέον χαρακτηριστικό, το οποίο συνδέεται με τη δομή του δικτύου, αποτελεί το γεγονός ότι είναι πολύ μικρό το ποσοστό των ιστοσελίδων που προσφέρουν κάποιον σύνδεσμο σε άλλες ιστοσελίδες, οπότε κάθε μία από αυτές αποτελεί τελικά μια αυτόνομη οντότητα [13]. Τελικά, αυτό το μέρος του ιστού αποτελείται από ιδιωτικές σελίδες, η ύπαρξή των οποίων παραμένει άγνωστη στις μηχανές αναζήτησης και ο τρόπος να τις επισκεφθεί κανείς είναι μόνον εφόσον έχει εξασφαλίσει κάποιου είδους εξουσιοδοτημένη πρόσβαση.

2.3 Ο Σκοτεινός Ιστός

2.3.1 Ο Σκοτεινός Ιστός και τα σκοτεινά δίκτυα

Μελετώντας τη βιβλιογραφία και τις αναφορές στο σκοτεινό μέρος του παγκοσμίου ιστού, μπορεί να παρατηρηθεί μια σύγχυση γύρω από την ορολογία του σκοτεινού ιστού και των σκοτεινών δικτύων. Αποτελούν λανθασμένα συνυφασμένους όρους, ενώ στην πραγματικότητα πρόκειται για δύο διαφορετικά πεδία.

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, ο βαθύς ιστός αποτελεί ένα υποσύνολο του παγκοσμίου ιστού, με τη διαφορά τους να εντοπίζεται στο γεγονός ότι τα περιεχόμενά του δεν είναι καταχωρημένα σε μηχανές αναζήτησης. Μέσα σε αυτό, λειτουργεί ένας αριθμός δικτύων επικάλυψης, δηλαδή εικονικών δικτύων που λειτουργούν πάνω στο υπάρχον δίκτυο. Αυτά έχουν ως στόχο την απόκρυψη της ταυτότητας των χρηστών και των μεταξύ τους επικοινωνιών και η πρόσβαση μπορεί να πραγματοποιηθεί μόνον μέσω ειδικού λογισμικού. Τα δίκτυα ανωνυμίας είναι γνωστά ως σκοτεινά δίκτυα (darknets), με τα πιο δημοφιλή να είναι το δίκτυο Tor, το Invisible Internet Project (I2P) και το Freenet, η παρουσίαση των οποίων ακολουθεί στις επόμενες ενότητες [14].

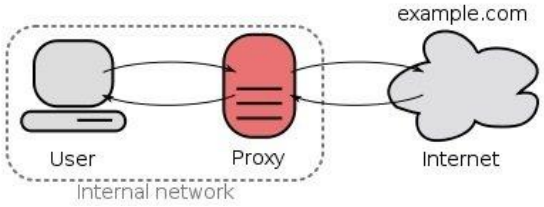
Επιπλέον, υπάρχουν ακόμη δύο δομές που φέρουν τον χαρακτηρισμό του σκοτεινού δικτύου. Η πρώτη από αυτές είναι οι εξυπηρετητές και το λογισμικό που προωθούν την παράνομη διακίνηση υλικού, το οποίο υπόκειται σε πνευματικά δικαιώματα. Η δεύτερη είναι ομάδες εξυπηρετητών, οι οποίοι εκτελούνται σε παθητική λειτουργία, αποφεύγοντας με αυτόν τον τρόπο τον εντοπισμό τους από τους χρήστες. Στόχος είναι η παρατήρηση και η συλλογή πληροφοριών για τις δράσεις των τελευταίων και είναι γνωστές με την ονομασία honeypots [15].

Το σύνολο των σκοτεινών δικτύων και των υπηρεσιών που παρέχονται με τις παραπάνω μεθόδους, αποτελούν το σκοτεινό διαδίκτυο. Πρόκειται για υποσύνολο του βαθύ ιστού και είναι το τμήμα εκείνο, που όντας σκόπιμα κρυμμένο, αποτελεί αντικείμενο εκμετάλλευσης από μέρος του ψηφιακού πληθυσμού για την εκτέλεση παράνομων δραστηριοτήτων, αποφεύγοντας τον εντοπισμό από τις αρχές χάρη στην ανωνυμία που προσφέρει.

2.3.2 Onion Routing

Στο Εργαστήριο Ναυτικών Ερευνών των ΗΠΑ, στα μέσα της δεκαετίας του '90, στην προσπάθεια εξεύρεσης νέων μεθόδων με στόχο την προστασία των πληροφοριών και την ασφάλεια των επικοινωνιών, ξεκίνησαν έρευνες για τη δημιουργία μιας καινούργιας αρχιτεκτονικής. Αυτή χαρακτηρίζονταν από την κρυπτογράφηση των δεδομένων και την αποτελεσματική αντίσταση του καναλιού επικοινωνίας απέναντι στην ανάλυση της κίνησης. Στόχος δεν ήταν η απόκρυψη της ταυτότητας των πλευρών που επιθυμούν να επικοινωνήσουν, αλλά η προστασία της μεταξύ τους επικοινωνίας από τρίτους, όταν γίνεται χρήση του δημοσίου δικτύου. Η αρχιτεκτονική έγινε ευρύτερα γνωστή με την ονομασία onion routing [16].

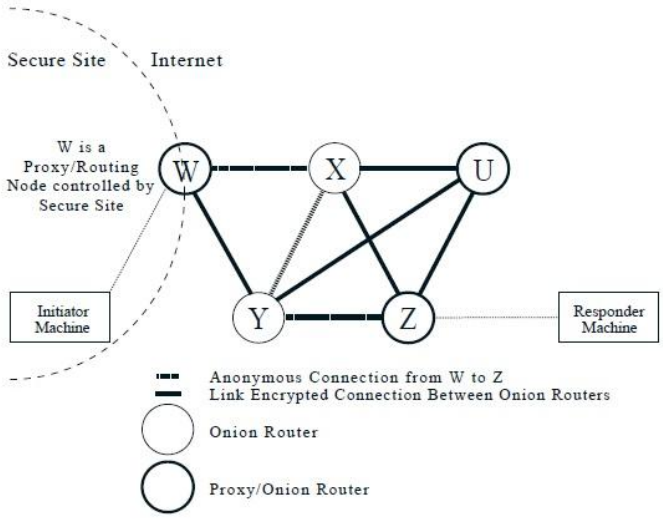
Βασικό συστατικό της αρχιτεκτονικής αποτελούν οι διακομιστές μεσολάβησης (proxy servers), μέσω των οποίων δύο πλευρές επικοινωνούν μεταξύ τους. Αυτού του είδους οι διακομιστές έχουν το ρόλο του ενδιάμεσου συνδέσμου για τους χρήστες, που βρίσκονται σε μια ασφαλή τοποθεσία και συνδέονται στον παγκόσμιο ιστό μέσω ενός τοίχους προστασίας. Ο ρόλος τους είναι να παραλαμβάνουν τα αιτήματα των χρηστών, να τα επεξεργάζονται, να τα αποστέλλουν στον προορισμό τους και στη συνέχεια αφού λάβουν την απάντηση, να την προωθούν πίσω στον αρχικό χρήστη [17]. Τα αιτήματα μπορεί να αφορούν υπηρεσίες περιήγησης στον ιστό για την ανάκτηση πληροφοριών, υπηρεσίες αποθήκευσης, επεξεργασίας και διακίνησης δεδομένων, όπως και υπηρεσίες ηλεκτρονικού ταχυδρομείου και άμεσων μηνυμάτων.



Εικόνα 2-1: Διακομιστής μεσολάβησης

(https://en.wikipedia.org/wiki/File:Forward_proxy_h2g2bob.svg)

Οι διακομιστές μεσολάβησης στα πλαίσια του onion routing, έχοντας ως στόχο την ουδετερότητα λειτουργίας με τις εφαρμογές λογισμικού, αποτελούνται από δύο μέρη, τον πελάτη μεσολάβησης (client proxy) και τον πυρήνα μεσολάβησης (core proxy). Ο πρώτος αποτελεί την ενδιάμεση σύνδεση της εφαρμογής με τον core proxy και φροντίζει για την κατάλληλη προετοιμασία των δεδομένων που αυτή επιθυμεί να αποστείλει, αφαιρώντας όλα τα στοιχεία που περιέχουν πληροφορίες για τους χρήστες. Επίσης αφαιρούνται λειτουργίες που σχετίζονται με cookies, αφού και αυτά μπορεί να περιέχουν ανάλογες πληροφορίες. Ο core proxy, παραλαμβάνοντας τα δεδομένα, πραγματοποιεί έλεγχο για την ύπαρξη των απαιτούμενων πληροφοριών, όπως το πρωτόκολλο επικοινωνίας και η θύρα προορισμού, καθώς και για την τήρηση της ανωνυμίας. Σε περίπτωση απόρριψης αποστέλλει πίσω στον client proxy ανάλογο μήνυμα, διαφορετικά προχωράει στην εγκαθίδρυση σύνδεσης με τον διακομιστή μεσολάβησης που βρίσκεται στην άλλη πλευρά του καναλιού επικοινωνίας, μεταβιβάζοντας πλέον σε αυτόν όλα τα δεδομένα [18].



Εικόνα 2-2: Η τοπολογία του onion routing

(Goldschlag D., Reed M., Syverson P. "Anonymous connections and Onion routing", 1997)

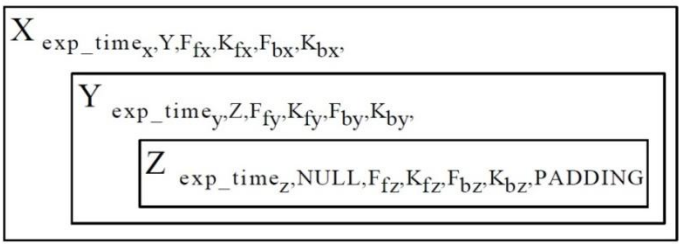
Η ιδιαιτερότητα της αρχιτεκτονικής είναι πως η διαδρομή μεταξύ των διακομιστών μεσολάβησης αποτελείται από ενδιάμεσους κόμβους, οι οποίοι διατηρούν μόνιμα κρυπτογραφημένες συνδέσεις μεταξύ τους και διακινούν τα δεδομένα. Η τοπολογία είναι προκαθορισμένη και όλοι οι κόμβοι γνωρίζουν τους κόμβους που απαρτίζουν το συνολικό δίκτυο, ενώ επίσης γνωρίζουν τα δημόσια κλειδιά τους. Σε μια διαδρομή, οι διακομιστές γνωρίζουν ακριβώς ποιοι κόμβοι θα είναι μέρος αυτής, ενώ οι

ίδιοι οι κόμβοι γνωρίζουν μόνον τον προηγούμενο και τον επόμενο σε αυτούς κόμβο. Λόγω αυτού του περιορισμού, δεν έχουν πλήρη εικόνα της πορείας που θα διαγράψουν τα δεδομένα και απλά τα προωθούν. Πρέπει να σημειωθεί πως ο αρχικός διακομιστής είναι ο πιο ευαίσθητος σε περίπτωση ανάλυσης του δικτύου, διότι είναι αυτός που παραλαμβάνει τα αρχικά δεδομένα προτού εφαρμόσει τη διαδικασία κρυπτογράφησής τους [16].

Για την εγκαθίδρυση της σύνδεσης, ο core proxy επιλέγει τυχαία ένα μονοπάτι προς τον διακομιστή που θα είναι ο τελικός αποδέκτης και στη συνέχεια κατασκευάζει μια δομή δεδομένων, αποτελούμενη από πολλαπλά στρώματα, γνωστή ως "κρεμμύδι" (onion). Πρόκειται για επίπεδα κρυπτογράφησης, το ένα πάνω από το άλλο, ένα για κάθε έναν από τους κόμβους που αποτελούν μέρος της επιλεγμένης διαδρομής, με χρήση του δημοσίου κλειδιού τους (ασύμμετρη κρυπτογραφία). Έπειτα, αποστέλλεται μέσα στο δίκτυο και σε κάθε κόμβο που καταφθάνει, αφαιρείται ένα στρώμα κρυπτογράφησης με χρήση του ιδιωτικού κλειδιού του, φανερώνοντας τα περιεχόμενα που τον αφορούν. Δε γνωρίζει την αρχική πηγή του πακέτου, παρά μόνο τον κόμβο από τον οποίο το παρέλαβε και τον κόμβο στον οποίο πρέπει να το προωθήσει. Επίσης, γνωρίζει την κρυπτογραφική λειτουργία που θα εφαρμοσθεί τόσο προς τη μια κατεύθυνση της επικοινωνίας, όσο και στην αντίθετη φορά. Λόγω του ότι εξυπηρετούνται ταυτόχρονα πολλαπλές συνδέσεις, κάθε κόμβος διατηρεί έναν πίνακα, όπου αποθηκεύονται προσωρινά τα αναγνωριστικά και τα κρυπτογραφικές λειτουργίες που αφορούν κάθε μία από αυτές. Όταν το "κρεμμύδι" φτάσει στον τελικό προορισμό, υπάρχει πλέον ενεργή σύνδεση και ξεκινάει η αποστολή δεδομένων [19].

Η ίδια περίπου μεθοδολογία ακολουθείται στα πακέτα δεδομένων. Στην προς τα εμπρός κατεύθυνση, ο αρχικός διακομιστής, κάνοντας χρήση κρυπτογραφίας μυστικού κλειδιού (συμμετρική), κρυπτογραφεί τις πληροφορίες χρησιμοποιώντας τα κλειδιά με αντίστροφη σειρά, ξεκινώντας με εκείνο του διακομιστή στο τέλος και καταλήγοντας στην αρχή. Στη δομή της εικόνας 3, όπου τη διαδρομή απαρτίζουν τρεις κόμβοι, το εσωτερικό στρώμα αφορά τον τελευταίο (Z), το μεσαίο τον ενδιάμεσο (Y) και το εξωτερικό αφορά τον πρώτο κόμβο (X) που θα συναντήσουν τα πακέτα. Μέσα σε αυτήν περιέχονται η ταυτότητα του επόμενου κόμβου, δύο ζεύγη που αφορούν κρυπτογραφικές λειτουργίες, ένα ανά κατεύθυνση, το φορτίο και ένα χρονόμετρο. Το τελευταίο χρησιμεύει στον εντοπισμό επαναλήψεων μετάδοσης. Πακέτα που είτε έχει λήξει ο χρόνος μετάδοσής τους, είτε έχουν δεχθεί επανάληψη μετάδοσης, μπορεί να έχουν

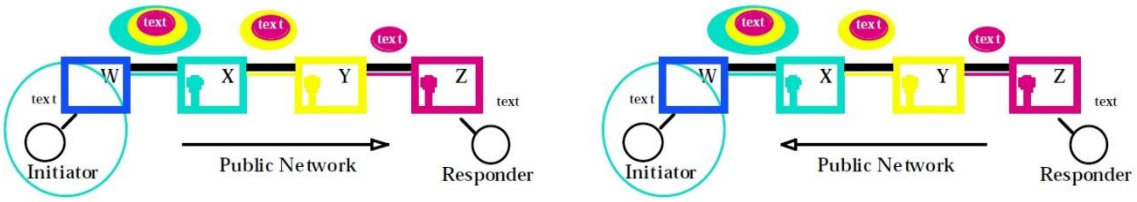
περάσει από κόμβο που έχει παραβιαστεί, οπότε για λόγους ασφαλείας παραβλέπονται [18].



Εικόνα 2-3: Επίπεδα κρυπτογράφησης δεδομένων

(Goldschlag D., Reed M., Syverson P. "Hiding Routing Information", 1996)

Η αντίστροφη διαδικασία ακολουθείται όταν τα δεδομένα κινούνται στην αντίθετη κατεύθυνση. Κάθε κόμβος που δέχεται τα δεδομένα, προσθέτει ένα στρώμα κρυπτογράφησης, οπότε όταν αυτά φτάσουν στον αρχικό διακομιστή, έχουν τον αριθμό επιπέδων κρυπτογράφησης που είχαν κατά την αρχική αποστολή. Αυτός τα αφαιρεί και επιστρέφει στον χρήστη το καθαρό περιεχόμενο [19]. Στην Εικόνα 2-4 φαίνεται η διαδικασία κίνησης των δεδομένων προς τις δύο κατευθύνσεις.



Εικόνα 2-4: Κίνηση δεδομένων

(Goldschlag D., Reed M., Syverson P. "Proxies for Anonymous Routing", 1996)

Οι παράγοντες, που μπορούν να αποτελέσουν πηγή πληροφοριών κατά την ανίχνευση του δικτύου, είναι οι πληροφορίες που αφορούν τους χρήστες, το φορτίο και ο ρυθμός μετάδοσης των δεδομένων. Όπως είναι φυσικό, κάθε φορά που αποκρυπτογραφείται ένα πακέτο δεδομένων και αφαιρείται ένα επίπεδο, μειώνεται το μέγεθός του, πράγμα που το καθιστά ευαίσθητο στην ανίχνευση. Το ίδιο ισχύει και αντίστροφα, κατά την επιστροφή των δεδομένων, όταν δεν έχουν προστεθεί όλα τα επίπεδα κρυπτογράφησης. Για αυτό το λόγο, σε κάθε κρυπτογραφική διαδικασία προστίθεται ένα τμήμα τυχαίων δεδομένων τέτοιου μεγέθους, ώστε το συνολικό μέγεθος να παραμείνει αμετάβλητο. Το επιπλέον τμήμα που προστέθηκε αθροιστικά σε όλους τους κόμβους, είναι γνωστό μόνον στους διακομιστές, οι οποίοι θα το αφαιρέσουν

προκειμένου να ανακτήσουν το καθαρό περιεχόμενο. Έτσι, τα δεδομένα αποκτούν διαφορετική εμφάνιση τόσο για τους ίδιους τους κόμβους, όσο και για έναν εξωτερικό παρατηρητή του δικτύου, οπότε είναι πολύ δύσκολο να παρακολουθήσει κανείς με ακρίβεια την πορεία τους. Η διατήρηση συγκεκριμένου μεγέθους έχει ως στόχο να μη δημιουργούνται διακυμάνσεις, οι οποίες θα μπορούσαν να οδηγήσουν στο σχηματισμό προτύπων. Στην ίδια λογική βασίζεται και το γεγονός ότι όλοι οι κόμβοι που συμμετέχουν στο δίκτυο πρέπει να παρουσιάζουν όμοιους ρυθμούς προώθησης. Αυτό απαιτεί την ενδιάμεση αποστολή πλαστών πακέτων, όταν η κίνηση είναι μειωμένη, προκειμένου να διατηρηθεί σταθερός ο ρυθμός. Με αυτόν τον τρόπο, δε δημιουργούνται πρότυπα σε επίπεδο κίνησης δεδομένων και είναι δύσκολος ο εντοπισμός των αρχικών και τελικών κόμβων. Αυτός είναι ο βασικός λόγος που οι ίδιοι πρέπει να εκτελούν ταυτόχρονα τον ρόλο των ενδιάμεσων κόμβων. Τέλος, ένας ακόμη τρόπος που εξασφαλίζει την ασφάλεια είναι η πολυπλεξία που εκτελείται στους ενδιάμεσους κόμβους, όταν αυτοί προωθούν τα δεδομένα. Δεν τηρείται συγκεκριμένη σειρά εξυπηρέτησής τους, όπως η FIFO, αλλά υπάρχει μια τυχαιότητα [16].

Το κόστος της αρχιτεκτονικής είναι στο σύνολό του μικρό. Χρησιμοποιείται η κρυπτογράφηση δημοσίου κλειδιού κατά την εγκαθίδρυση της σύνδεσης, με το απαιτητικό μέρος της κρυπτογράφησης να πραγματοποιείται στον πρώτο διακομιστή. Οι υπόλοιποι κόμβοι, από τους οποίους θα περάσουν τα πακέτα, μοιράζονται το φορτίο της αποκρυπτογράφησης, καθώς ο καθένας εκτελεί το μερίδιο που του αναλογεί στη διαδικασία. Στην αποστολή δεδομένων, γίνεται μόνο χρήση κρυπτογράφησης μυστικού κλειδιού, η οποία λειτουργεί ταχύτερα. Τελικά, η καθυστέρηση που μπορεί να παρουσιαστεί στην κίνηση των δεδομένων εξαρτάται μόνον από τον αριθμό των κόμβων που θα αποτελέσουν το μονοπάτι της διαδρομής [20].

Στην αρχιτεκτονική του onion routing οι διακομιστές λειτουργούν στο επίπεδο εφαρμογής, προσφέροντας συμβατότητα με τις τεχνολογίες HTTP, SMTP, FTP, RLOGIN και TELNET, γεγονός που προσφέρει ευελιξία και την καθιστά ιδανική για χρήση σε υπηρεσίες διαδικτύου, ηλεκτρονικού ταχυδρομείου, καθώς και στη διακίνηση δεδομένων. Αυτό που ξεκίνησε ως ένα ερευνητικό έργο του ναυτικού, συνδεδεμένο άμεσα με στρατιωτικά θέματα, στη συνέχεια αποτέλεσε την ιδέα ανάπτυξης ενός μεγαλύτερου δικτύου, βασισμένου στην ίδια φιλοσοφία και στο οποίο θα είχε πρόσβαση εύκολα το κοινό. Η υλοποίηση της ιδέας ξεκίνησε το 2006 και είναι γνωστή ως The Onion Routing Project, ή αλλιώς ως Tor Project.

2.3.2 Tor Project

Το Tor αποτελεί τη δεύτερη και σαφώς βελτιωμένη έκδοση του δικτύου onion routing, η οποία καταφέρνει να προσπεράσει κάποιους περιορισμούς που υπήρχαν προηγουμένως και ταυτόχρονα να προσφέρει υπηρεσίες, δίχως να απαιτούνται ειδικά δικαιώματα πρόσβασης ή τροποποιήσεις στα συστήματα των χρηστών. Αποτελεί ένα δίκτυο επικάλυψης που λειτουργεί πάνω στο διαδίκτυο, το οποίο προσφέρει χαρακτηριστικά όπως η ανωνυμία, η αμφίδρομη επικοινωνία και η χαμηλή καθυστέρηση στα κανάλια επικοινωνίας [21].

Κύρια θέματα που απασχόλησαν τους σχεδιαστές του ήταν η ευκολία της υλοποίησής του, διατηρώντας ταυτόχρονα χαμηλά το κόστος, τόσο στο υπολογιστικό κομμάτι, όσο και στο εύρος ζώνης που αυτό θα απαιτούσε κατά τη λειτουργία του. Με αυτόν τον τρόπο θα αυξάνονταν οι πιθανότητες επιτυχούς λειτουργίας του σε πραγματικές συνθήκες. Επίσης, η ευκολία στη χρήση θα αποτελούσε σημαντικό κίνητρο για τους χρήστες ώστε να το δοκιμάσουν, να το χρησιμοποιήσουν και να το στηρίξουν. Μεγαλύτερος αριθμός χρηστών θα σήμαινε και καλύτερο αποτέλεσμα για την ανωνυμία τους. Η ευελιξία και η απλότητα της σχεδίασής του θα βοηθούσαν στην υλοποίηση ενός σταθερού συστήματος, το οποίο θα έχει ως κύριο χαρακτηριστικό την προστασία της ταυτότητας των χρηστών και την αποτελεσματική ασφάλεια, προσφέροντας παράλληλα πρόσφορο έδαφος για μελλοντικές έρευνες [21].

Όπως στην πρώτη έκδοση έτσι και εδώ, τα δεδομένα μεταφέρονται μέσω ενδιάμεσων κόμβων με την ονομασία onion routers και είναι συστήματα χρηστών, τα οποία διατίθενται από τους ίδιους εθελοντικά με στόχο να εξυπηρετήσουν τη λειτουργία του δικτύου. Οι συνδέσεις μεταξύ τους υλοποιούνται με το πρωτόκολλο επικοινωνίας TLS (transport layer security), το οποίο παρέχει αξιοπιστία με τη χρήση κρυπτογραφίας, φροντίζοντας ταυτόχρονα για την ασφάλεια και την ακεραιότητα των δεδομένων [22]. Σε κάθε κόμβο διατηρούνται δύο κλειδιά. Το πρώτο είναι το κλειδί ταυτοποίησης (identity key), το οποίο είναι μακράς διάρκειας και χρησιμοποιείται για την υπογραφή των πιστοποιητικών TLS, του περιγραφέα του κόμβου, που περιέχει στοιχεία όπως τα κλειδιά και την πολιτική εξόδου, καθώς και των καταλόγων, όταν πρόκειται για εξυπηρετητές καταλόγου. Το δεύτερο είναι το κλειδί "κρεμμυδιού" (onion key), το οποίο είναι βραχείας διάρκειας, χρησιμοποιείται για την αποκρυπτογράφηση των εισερχόμενων αιτημάτων σύνδεσης από άλλους κόμβους, αλλά και στη διαπραγμάτευση από την οποία προκύπτουν μικρής διάρκειας κλειδιά. Επιπλέον, το πρωτόκολλο TLS

προβλέπει τη δημιουργία ενός κλειδιού σύνδεσης (link key), που είναι επίσης βραχείας διάρκειας και χρησιμοποιείται στις συνδέσεις μεταξύ των κόμβων. Τα κλειδιά αυτού του είδους εναλλάσσονται ανά τακτά χρονικά διαστήματα, με στόχο την αντιμετώπιση της περίπτωσης εκμετάλλευσής τους από κακόβουλο χρήστη [21].

Ενώ στο onion routing υπήρχε η υλοποίηση των client proxy και core proxy στους διακομιστές μεσολάβησης, στο Tor δεν υπάρχει ειδική υλοποίηση συμβατότητας με τις εφαρμογές, αφού πλέον χρησιμοποιείται η διεπαφή μεσολάβησης SOCKS. Πρόκειται για ένα πρωτόκολλο διαδικτύου, το οποίο χρησιμοποιείται στην ανταλλαγή πακέτων δεδομένων μεταξύ εξυπηρετητή και πελάτη και το οποίο, στην έκδοση SOCKS5, προσφέρει τη δυνατότητα αυθεντικοποίησης, ούτως ώστε μόνον εγκεκριμένοι πελάτες να αποκτούν πρόσβαση σε έναν εξυπηρετητή [23]. Με τη χρήση του πρωτοκόλλου υποστηρίζονται οι διαδικτυακές εφαρμογές, δίχως να απαιτείται κάποια επιπλέον τροποποίηση. Το μόνο που χρειάζεται να κάνει ο χρήστης που επιθυμεί να συνδεθεί στο δίκτυο είναι να εκτελέσει τοπικά ειδικό λογισμικό, με την ονομασία onion proxy. Σε κάθε εκτέλεσή του, πραγματοποιείται η ανάκτηση της τοπολογίας του δικτύου, η εγκαθίδρυση και η διαχείριση συνδέσεων τύπου TCP [21].

Προκειμένου να είναι ικανοποιητική η απόδοση του δικτύου και λαμβάνοντας υπόψη τον ολοένα και μεγαλύτερο αριθμό χρηστών, στο Tor δε χρησιμοποιείται ένα κύκλωμα ανά σύνδεση, αλλά γίνεται πολυπλεξία συνδέσεων σε κάθε ένα από αυτά και είναι κάτι που αναλαμβάνει αυτόματα ο onion proxy. Με στόχο τη βελτίωση της ταχύτητας, ετοιμάζονται προκαταβολικά τέτοιου είδους κυκλώματα δίχως να είναι απαραίτητη η ύπαρξη ενεργού αιτήματος από κάποια εφαρμογή. Νέο κύκλωμα δημιουργείται, είτε λόγω ολοκλήρωσης της χρήσης του παλιού κυκλώματος, είτε διότι υπάρχει ενεργή σύνδεση και πρέπει να γίνει προετοιμασία κυκλώματος για την επόμενη σύνδεση. Επίσης, ανά τακτά χρονικά διαστήματα και συγκεκριμένα κάθε ένα λεπτό, πραγματοποιείται αλλαγή κυκλώματος για καλύτερη διαχείριση του φόρτου. Όταν ζητηθεί η δημιουργία μιας σύνδεσης TCP από τον χρήστη, ο onion proxy επιλέγει το πιο καινούργιο κύκλωμα και τον κατάλληλο κόμβο που θα αποτελέσει τον κόμβο εξόδου, δηλαδή τον τελευταίο κόμβο της διαδρομής. Με την επιβεβαίωση της ύπαρξης σύνδεσης από τον κόμβο, αποστέλλεται στην εφαρμογή μια επιβεβαίωση τύπου SOCKS και πλέον είναι εφικτή η αποστολή δεδομένων. Για τον τερματισμό της σύνδεσης ακολουθείται ανάλογη διαδικασία με εκείνη μιας απλής TCP σύνδεσης, με χρήση διπλής χειραψίας σε περίπτωση κανονικής λειτουργίας και μονής χειραψίας σε περίπτωση σφάλματος [21].

Μια ακόμη βασική διαφορά σε σχέση με το onion routing είναι πως δε χρησιμοποιείται η δομή του "κρεμμυδιού" για την εγκαθίδρυση της σύνδεσης μεταξύ των δύο πλευρών, αλλά η διαδρομή κατασκευάζεται με τηλεσκοπικό τρόπο. Γίνονται διαδοχικές διαπραγματεύσεις με κάθε έναν από τους κόμβους που θα αποτελέσουν τη διαδρομή για τον ορισμό των κρυπτογραφικών κλειδιών που θα χρησιμοποιηθούν. Με αυτόν τον τρόπο κατασκευής του κυκλώματος, δεν είναι απαραίτητος ο έλεγχος επανάληψης μετάδοσης των πακέτων για την αξιοπιστία της σύνδεσης. Όλα τα πακέτα, τα οποία ονομάζονται κελιά (cells), είναι συγκεκριμένου μεγέθους (512 bytes) και διαθέτουν μια κεφαλίδα, όπου περιέχεται το αναγνωριστικό κυκλώματος (circID), μια εντολή διαχείρισης του φορτίου και το φορτίο. Το πρώτο καθορίζει το κύκλωμα που θα χρησιμοποιηθεί για την αποστολή του πακέτου, ενώ με το δεύτερο γίνεται κατηγοριοποίηση των πακέτων, τα οποία είναι είτε πακέτα ελέγχου (control cells), είτε πακέτα αναμετάδοσης (relay cells). Τα πακέτα ελέγχου αναλαμβάνουν τη δημιουργία των συνδέσεων και περιέχουν τις εντολές για την κατασκευή και το κλείσιμό τους. Τα πακέτα αναμετάδοσης, στα οποία εφαρμόζεται κρυπτογραφία τύπου AES 128-bit, περιέχουν μια επιπλέον κεφαλίδα που περιέχει το αναγνωριστικό (streamID) και αφορά τη σύνδεση μέσω της οποίας θα μεταδοθεί το πακέτο, καθώς και τις εντολές που αφορούν τη μετάδοση, τον έλεγχο ακεραιότητας του πακέτου και το μέγεθος του φορτίου [21].

Κατά την εγκαθίδρυση μιας σύνδεσης λαμβάνει μέρος η διαπραγμάτευση για το συμμετρικό κρυπτογραφικό κλειδί που θα χρησιμοποιηθεί με κάθε έναν από τους κόμβους, οι οποίοι θα αποτελέσουν μέρος της διαδρομή των δεδομένων. Η διαδικασία γίνεται ανά κόμβο. Ο αρχικός χρήστης προωθεί στον πρώτο κόμβο της διαδρομής ένα πακέτο, που είναι κρυπτογραφημένο με το δημόσιο κλειδί του και περιέχει στο φορτίο του το πρώτο μισό ενός κλειδιού χειραψίας Diffie-Hellman. Αυτός από την πλευρά του απαντάει με το δεύτερο μισό της DH χειραψίας και ένα παράγωγο, το οποίο προκύπτει από τον κατακερματισμό του τελικού κλειδιού, επιβεβαιώνοντας ταυτόχρονα τη σύνδεση. Έπειτα, ο αρχικός αποστέλλει αίτημα σύνδεσης με τον δεύτερο κατά σειρά κόμβο, μέσω του πρώτου, με ένα κρυπτογραφημένο πακέτο δεδομένων που περιέχει το κλειδί χειραψίας DH που θα χρησιμοποιηθεί μεταξύ τους, ζητώντας ουσιαστικά την επέκταση του κυκλώματος. Ο πρώτος κόμβος προχωράει σε εγκαθίδρυση σύνδεσης με τον δεύτερο με την ίδια διαδικασία, αποστέλλοντας μήνυμα ενημέρωσης στον αρχικό χρήστη με την ολοκλήρωσή της. Η διαδικασία συνεχίζεται μέχρι το σχηματισμό της

διαδρομής προς τον τελικό προορισμό. Μετά από κάθε βήμα, οι κόμβοι ανά δύο είναι σε θέση να ανταλλάξουν πακέτα αναμετάδοσης, τα οποία θα είναι κρυπτογραφημένα με το κλειδί που συμφωνήθηκε κατά τη μεταξύ τους διαπραγμάτευση. Η μονομερής αυθεντικοποίηση των μελών που συμμετέχουν και των κλειδιών που θα χρησιμοποιηθούν καθ' όλη τη διάρκεια που το κύκλωμα θα παραμείνει ανοικτό, πετυχαίνει την πλήρη ανωνυμία τους. Με τηλεσκοπικό τρόπο πραγματοποιείται και η κατάργηση μιας σύνδεσης. Αυτή ζητείται από τον αρχικό χρήστη με την προώθηση του κατάλληλου πακέτου, αφού έχει ολοκληρωθεί η αποστολή δεδομένων. Με την παραλαβή του, κάθε κόμβος κλείνει τη σύνδεση στο συγκεκριμένο κύκλωμα, προωθώντας το αίτημα στον επόμενο κατά σειρά κόμβο, πραγματοποιώντας τελικά μια τηλεσκοπική παύση των συνδέσεων της διαδρομής [21].

Σε αντίθεση με το onion routing όπου δεν υπήρχε κάτι ανάλογο, στο Tor πραγματοποιείται πάντα έλεγχος της ακεραιότητας των δεδομένων προτού αυτά βγουν από το δίκτυο και λαμβάνει μέρος στα άκρα του εκάστοτε κυκλώματος. Δεν θα ήταν αποτελεσματικό να πραγματοποιηθεί σε ενδιάμεσους κόμβους, καθώς εκτός του ότι έτσι θα προκαλούσαν αύξηση του μεγέθους των πακέτων, θα ήταν επιτυχημένος μόνο στα δεδομένα που θα κινούνταν από τον χρήστη προς τον πρώτο κόμβο της διαδρομής. Στο κλειδί που προκύπτει κατά τη διαπραγμάτευση του χρήστη με έναν νέο κόμβο, εφαρμόζεται η συνάρτηση κατακερματισμού του αλγορίθμου SHA-1 (Secure Hash Algorithm 1) και στο αποτέλεσμα που προκύπτει, προστίθενται τα δεδομένα, μαζί με τα τέσσερα πρώτα bytes του αποτελέσματος της εφαρμογής του SHA-1 σε αυτά. Ελέγχοντας τα δεδομένα που παραλαμβάνουν με το παράγωγο, οι δύο πλευρές είναι σε θέση να διαπιστώσουν την ακεραιότητα αυτών. Ανάμεσα στα πρωτόκολλα που κάνουν χρήση του συγκεκριμένου αλγορίθμου είναι τα TLS, SSL, SSH, PGP, S/MIME και IPSec, και όπως αναφέρθηκε παραπάνω, το TLS χρησιμοποιείται για την ασφάλεια των συνδέσμων μεταξύ των κόμβων του δικτύου. Ο SHA-1 χρησιμοποιείται γενικότερα σε κρυπτογραφικές εφαρμογές, όπου υπάρχει ανάγκη για την προστασία της ακεραιότητας των δεδομένων [24], αποτελώντας αποτελεσματική μέθοδο απέναντι στην παραποίησή τους από έναν κακόβουλο χρήστη [21].

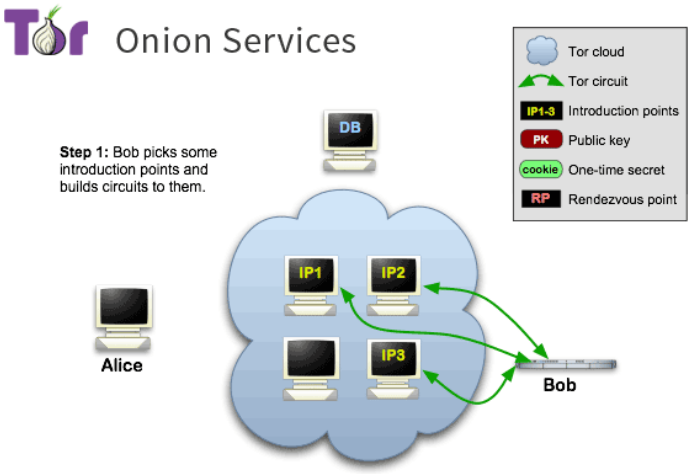
Στη βελτίωση της εικόνας που υπάρχει για το συνολικό δίκτυο συμβάλλει η ύπαρξη εξυπηρετητών καταλόγου (directory servers), οι οποίοι περιέχουν λίστες και ενημερώνουν τους onion proxies για την ακριβή κατάσταση των κόμβων που είναι διαθέσιμοι. Πρόκειται για κόμβους, οι οποίοι λειτουργούν ως εξυπηρετητές HTTP, έχουν

αναλάβει την επίβλεψη του δικτύου και την καταγραφή τυχόν αλλαγών που μπορεί να συντελεστούν, είτε στην τοπολογία, είτε στην κατάσταση των κόμβων. Οι τελευταίοι από την πλευρά τους αποστέλλουν ανά τακτά χρονικά διαστήματα στοιχεία για την κατάστασή τους, συμπεριλαμβάνοντας την υπογραφή τους. Σε περίπτωση που οι εξυπηρετητές λάβουν στοιχεία που δε διαθέτουν υπογραφή με αναγνωρισμένο κλειδί, αυτά παραβλέπονται. Αυτός είναι ο λόγος που κατά την προσθήκη νέων κόμβων στο δίκτυο, είναι αναγκαίο να προηγηθεί η έγκρισή τους από τους διαχειριστές των εξυπηρετητών. Τελικά, αφού ολοκληρωθεί η συλλογή των στοιχείων, εκδίδεται και κοινοποιείται μια περιγραφή του συνολικού δικτύου με τη μορφή καταλόγου. Επειδή αυτός παρέχεται στους χρήστες από τους εξυπηρετητές, πρέπει να είναι ίδιος σε κάθε έναν από αυτούς, να μην υπάρχουν διαφοροποιήσεις, πράγμα που επιτυγχάνεται με τον μεταξύ τους συγχρονισμό [21].

Η εκμετάλλευση των κόμβων εξόδου από κακόβουλους χρήστες μπορεί να τους δώσει τη δυνατότητα ανώνυμης εκτέλεσης παράνομων δραστηριοτήτων, όπως είναι η προώθηση κακόβουλου λογισμικού. Αν και αυτό αποτελεί συχνό φαινόμενο στον επιφανειακό ιστό, είναι λογικό να απασχολεί τους εθελοντές, κατόχους των συστημάτων που αποτελούν κόμβους του δικτύου, οι οποίοι δεν επιθυμούν την εμπλοκή τους με οποιονδήποτε τρόπο σε περιστατικά που βλάπτουν άλλα συστήματα, όπως είναι οι κυβερνο-επιθέσεις. Κάθε κόμβος λοιπόν διαθέτει μια πολιτική εξόδου (exit policy), μια δήλωση που περιγράφει με ακρίβεια τις εξωτερικές διευθύνσεις και τις πόρτες στις οποίες συνδέεται, οπότε καθορίζονται οι τύποι των δεδομένων που θα προωθεί. Τα είδη των κόμβων είναι συνολικά τρία: οι ανοικτοί κόμβοι εξόδου (open exit nodes), οι οποίοι συνδέονται παντού, οι ενδιάμεσοι κόμβοι εξόδου (middleman exit nodes), οι οποίοι απλά αναμεταδίδουν δεδομένα και τέλος, οι ιδιωτικοί κόμβοι εξόδου (private exit nodes), οι οποίοι συνδέονται μόνο με τοπικά συστήματα και δίκτυα. Στο τρίτο είδος παρέχεται η μεγαλύτερη ασφάλεια, αφού είναι εξαιρετικά δύσκολο για έναν κακόβουλο χρήστη να υποκλέψει τα δεδομένα που διακινούνται προς τον τελικό προορισμό. Οι περισσότεροι κόμβοι στο Tor προσφέρουν σύνδεση στον παγκόσμιο ιστό, επιβάλλοντας την ίδια στιγμή περιορισμούς πρόσβασης σε ευάλωτες υπηρεσίες, όπως είναι εκείνη του ηλεκτρονικού ταχυδρομείου. Προστατεύοντας το δίκτυο, οι χρήστες είναι θετικοί στη χρήση του, πράγμα που βοηθάει στη προώθηση και την επέκτασή του. Όσο μεγαλύτερος είναι ο αριθμός των κόμβων, τόσο πιο δύσκολη είναι η παρακολούθησή τους, μειώνοντας έτσι την πιθανότητα επιτυχημένης ανάλυσης της δικτυακής κίνησης [21].

Με στόχο την εξασφάλιση της προστασίας των εξυπηρετητών που παρέχουν διαδικτυακές υπηρεσίες, γνωστές ως υπηρεσίες "κρεμμυδιού" (onion services), έχει υλοποιηθεί ένας μηχανισμός που παρέχει ένα είδος φίλτρου στα αιτήματα που αυτοί δέχονται. Πρόκειται για τα σημεία συνάντησης (rendezvous points), τα οποία ορίζονται κατόπιν διαπραγμάτευσης μεταξύ του πελάτη και του εξυπηρετητή και στα οποία τελικά έρχονται σε επαφή μεταξύ τους τα δύο μέρη. Με τη συγκεκριμένη μέθοδο είναι δυνατή η παροχή υπηρεσιών, δίχως να αποκαλύπτεται η τοποθεσία και η διεύθυνση του εξυπηρετητή, οπότε προσφέρεται προστασία απέναντι σε επιθέσεις μέσω διαδικτύου, όπως είναι οι επιθέσεις άρνησης υπηρεσίας (DoS). Σε αυτήν την απόκρυψη οφείλεται το γεγονός ότι οι εν λόγω υπηρεσίες είναι γνωστές και ως κρυμμένες υπηρεσίες (hidden services) [21].

Πιο αναλυτικά, με την έναρξη λειτουργίας του, ο εξυπηρετητής δηλώνει την παρουσία του σε έναν αριθμό τυχαίων κόμβων, δημιουργώντας μαζί τους κυκλώματα και ενημερώνοντάς τους για το δημόσιο κλειδί του. Οι κόμβοι ονομάζονται σημεία εισαγωγής (introductory points) και στο εξής θα αποτελούν τα σημεία επικοινωνίας στα οποία θα παραλαμβάνονται τα αιτήματα σύνδεσης των χρηστών. Για τη διασφάλιση της αδιάλειπτης παρουσίας του εξυπηρετητή στο δίκτυο, δε χρησιμοποιείται μόνο ένα σημείο, καλύπτοντας έτσι την περίπτωση που κάποιο από αυτά τεθεί εκτός λειτουργίας [21].

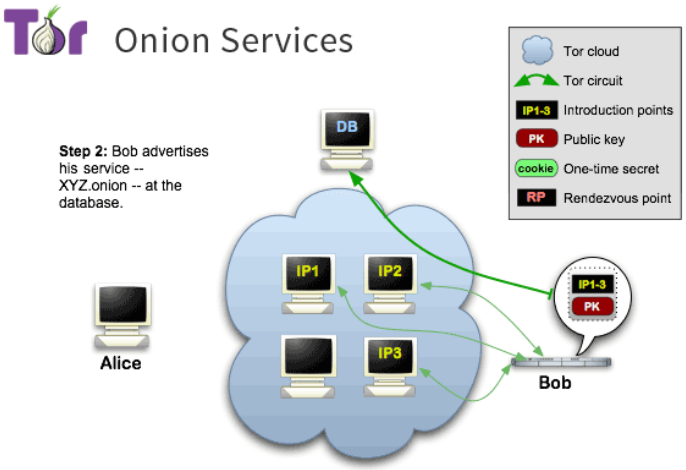


Εικόνα 2-5: Δημιουργία σημείων εισαγωγής

(https://www.torproject.org/docs/onion-services)

Ακολουθεί η αυτόματη δημιουργία ενός περιγραφέα υπηρεσίας (onion service descriptor), όπου περιέχονται το δημόσιο κλειδί και η λίστα με τα σημεία εισαγωγής,

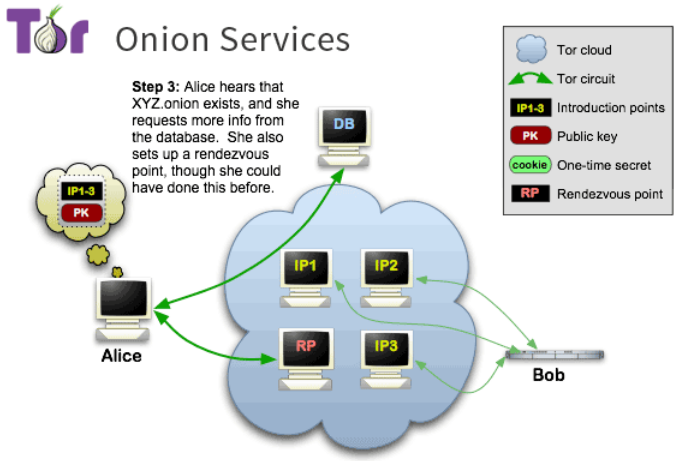
υπογράφοντάς τον με το δημόσιο κλειδί του. Αυτός δημοσιοποιείται σε ένα κατανεμημένο πίνακα κατακερματισμού, αποθηκευμένο σε ειδικούς κόμβους με την ονομασία HSDirs, από όπου μπορούν πλέον οι χρήστες να τον αναζητήσουν με βάση τη διεύθυνσή του. Η διεύθυνση αποτελείται από 16 χαρακτήρες, κατασκευάζεται από την εφαρμογή μιας συνάρτησης κατακερματισμού στο δημόσιο κλειδί του εξυπηρετητή και έχει κατάληξη .onion. Τα ψηφία της αποτελούνται αποκλειστικά από πεζούς, λατινικούς χαρακτήρες και από τους αριθμούς 2 έως 7 [25]. Πρέπει να σημειωθεί πως ενώ χρησιμοποιείται όπως τα urls, δεν είναι καταχωρημένη στο σύστημα DNS του διαδικτύου και είναι προσβάσιμη μόνο μέσω του onion proxy. Η αυτοματοποιημένη παραγωγή του αναγνωριστικού μπορεί να φαίνεται περιοριστική, εξυπηρετεί όμως έναν πολύτιμο σκοπό, που δεν είναι άλλος από την πιστοποίηση του παρόχου της υπηρεσίας [26].



Εικόνα 2-6: Δήλωση περιγραφέα υπηρεσίας

(https://www.torproject.org/docs/onion-services)

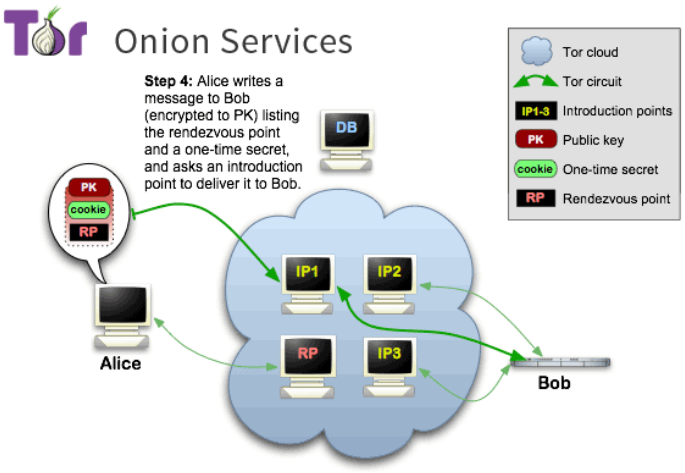
Ο πελάτης που επιθυμεί να επικοινωνήσει με τον εξυπηρετητή, πρέπει πρώτα να ενημερωθεί με κάποιον τρόπο για τη διεύθυνσή του. Αυτό μπορεί να γίνει είτε από τον κάτοχό του, είτε από πληροφορίες που τυχόν βρει σε άλλες ιστοσελίδες ή υπηρεσίες. Με βάση αυτή είναι σε θέση να ανακτήσει την περιγραφή που περιέχεται για αυτόν στον κατανεμημένο πίνακα, γνωρίζοντας έτσι τα σημεία εισαγωγής και το δημόσιο κλειδί που πρέπει να χρησιμοποιηθεί. Ταυτόχρονα επιλέγει έναν τυχαίο κόμβο του δικτύου και του ζητάει να αποτελέσει το σημείο συνάντησης στη συγκεκριμένη σύνδεση, δίνοντάς του παράλληλα ένα κλειδί μιας χρήσης (cookie) [26].



Εικόνα 2-7: Ανάκτηση περιγραφέα υπηρεσίας και επιλογή σημείου συνάντησης

(https://www.torproject.org/docs/onion-services)

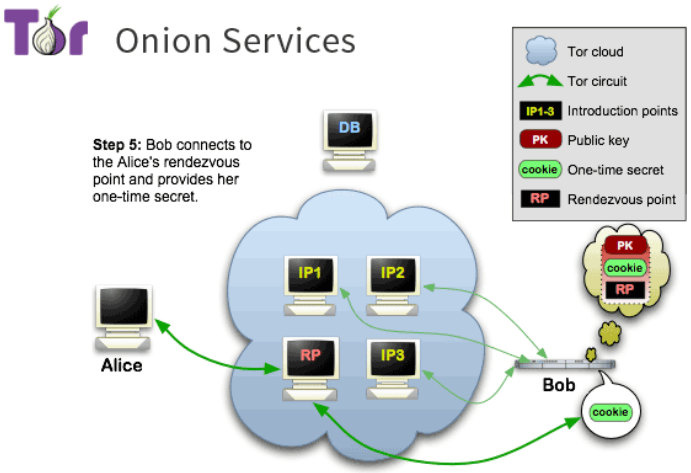
Στη συνέχεια δημιουργεί κύκλωμα με ένα εκ των σημείων εισαγωγής, όπου αφήνει ένα μήνυμα, δηλώνοντας την επιθυμία του για σύνδεση. Στο μήνυμα, το οποίο είναι κρυπτογραφημένο με το δημόσιο κλειδί του εξυπηρετητή, περιέχονται πληροφορίες όπως τα στοιχεία του πελάτη, το επιλεγμένο σημείο συνάντησης, το κλειδί μιας χρήσης και το πρώτο μέρος μιας DH χειραψίας. Έπειτα, το σημείο εισαγωγής αναλαμβάνει την παράδοση του μηνύματος στον εξυπηρετητή και υπάρχει αναμονή σύνδεσης των δύο πλευρών στο σημείο συνάντησης [26].



Εικόνα 2-8: Αίτημα σύνδεσης

(https://www.torproject.org/docs/onion-services)

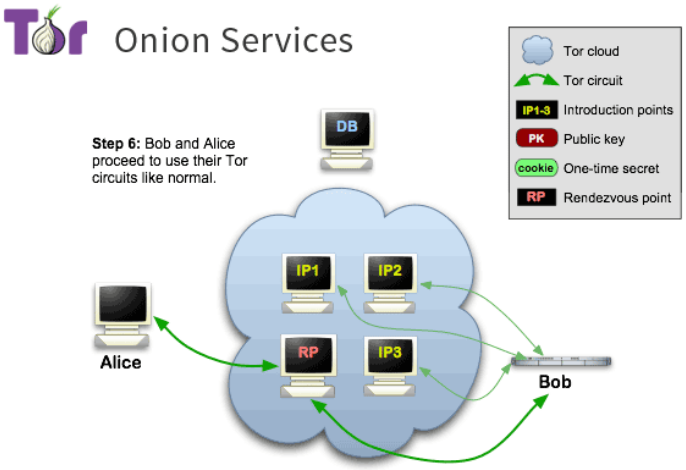
Εφόσον ο εξυπηρετητής δεχθεί το αίτημα, δημιουργεί και αυτός από τη πλευρά του ένα κύκλωμα με το σημείο συνάντησης, στο οποίο αποστέλλει το κλειδί μιας χρήσης, το δεύτερο μέρος της DH χειραψίας και το κλειδί που θα χρησιμοποιηθεί για τη συνεδρία που θα λάβει μέρος μεταξύ τους [21]. Είναι υψίστης σημασίας να χρησιμοποιεί ο εξυπηρετητής τους ίδιους κόμβους εισαγωγής (entry guards) κάθε φορά που δημιουργεί νέα κυκλώματα, προκειμένου να αποφύγει την αποκάλυψη της ταυτότητάς του. Αυτό θα ήταν εφικτό στην περίπτωση που ένας κακόβουλος χρήστης χρησιμοποιήσει δικό του κόμβο και αναγκάσει τον εξυπηρετητή να τον χρησιμοποιήσει ως εισαγωγικό σημείο. Πράττοντας κάτι τέτοιο και μέσω της χρονικής ανάλυσης του κυκλώματος, καθίσταται δυνατή η αποκάλυψη της διεύθυνσης IP του εξυπηρετητή [26].



Εικόνα 2-9: Σύνδεση του εξυπηρετητή στο σημείο συνάντησης

(https://www.torproject.org/docs/onion-services)

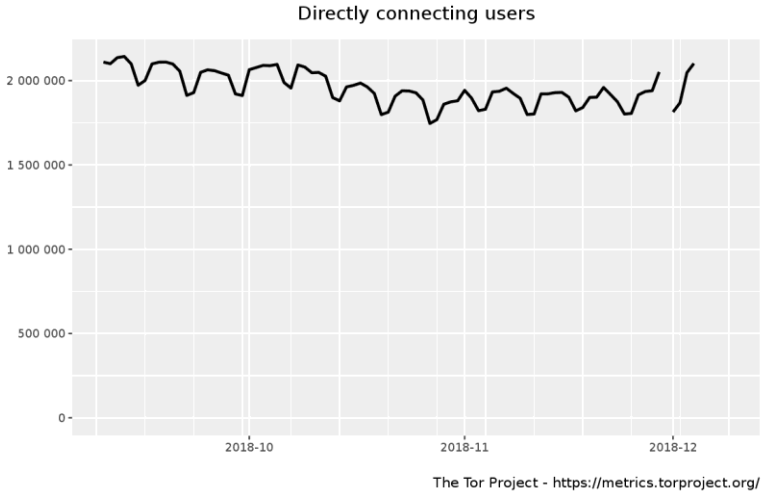
Το σημείο εισαγωγής ενημερώνει τον πελάτη για την επιτυχή δημιουργία της σύνδεσης με τον εξυπηρετητή, οπότε ξεκινάει η μεταξύ τους επικοινωνία. Εννοείται ότι δε γνωρίζει τις ταυτότητες των δύο πλευρών, ούτε έχει πρόσβαση στα δεδομένα που διακινούνται μεταξύ τους, τα οποία απλά αναμεταδίδει. Πρέπει να σημειωθεί πως κάθε κανάλι επικοινωνίας στο δίκτυο Tor αποτελείται από τουλάχιστον τρεις κόμβους. Όταν γίνεται χρήση σημείου συνάντησης, χρησιμοποιούνται συνολικά έξι κόμβοι, οι τρεις αποτελούν επιλογή από την πλευρά του πελάτη, με τον τρίτο να είναι το ίδιο το σημείο συνάντησης, ενώ τα υπόλοιπα ανήκουν στην πλευρά του εξυπηρετητή [26].



Εικόνα 2-10: Εγκαθίδρυση σύνδεσης μέσω του σημείου συνάντησης

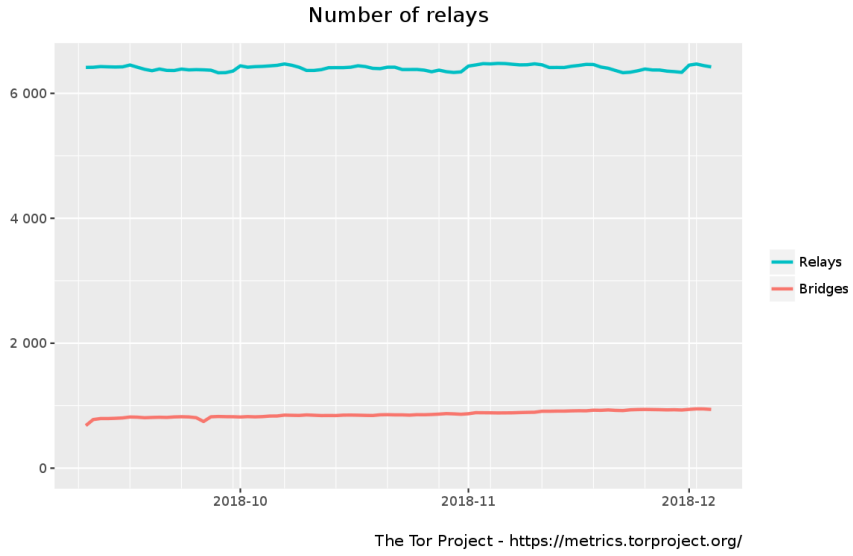
(https://www.torproject.org/docs/onion-services)

Μετρήσεις και στατιστικά του δικτύου είναι διαθέσιμα στην ιστοσελίδα Tor Metrics [27], με στοιχεία όπως το πλήθος των χρηστών, των κόμβων, των onion services, της χωρητικότητας που μπορεί να υποστηρίξει το δίκτυο, κτλ. Σύμφωνα με αυτά, στα τέλη του 2018 οι συνολικοί χρήστες του Tor είναι παραπάνω από 2 εκατομμύρια, οι ιστοσελίδες με κατάληξη onion ανέρχονται περίπου στις 105.000, ενώ οι κόμβοι λίγο πάνω από τους 6.300 και οι γέφυρες (bridges) περίπου στις 1.000. Οι τελευταίες είναι κόμβοι που εσκεμμένα δεν περιλαμβάνονται στη δημόσια λίστα και χρησιμοποιούνται στην περίπτωση που ο πάροχος διαδικτύου του χρήστη εμποδίζει την πρόσβαση στο δίκτυο μέσω των γνωστών κόμβων [28].



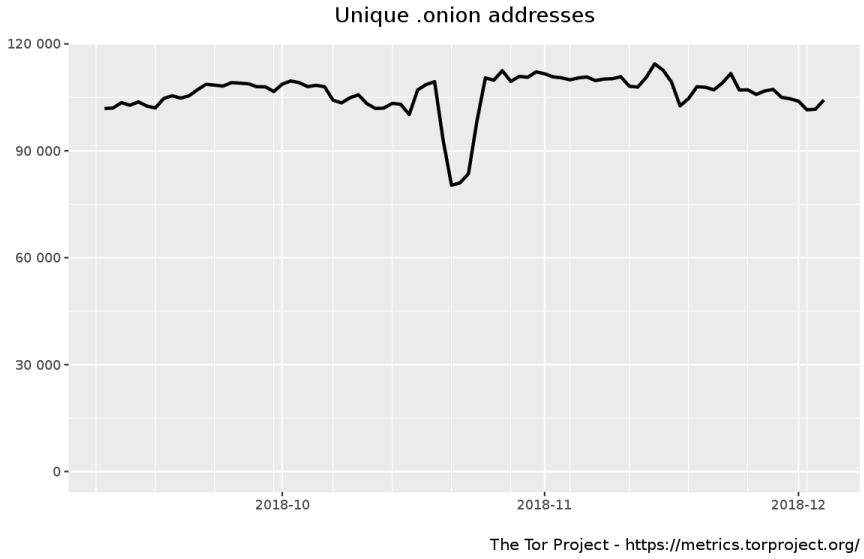
Εικόνα 2-11: Πλήθος χρηστών στο δίκτυο Tor

(https://metrics.torproject.org/userstats-relay-country.html)



Εικόνα 2-12: Πλήθος κόμβων στο δίκτυο Tor

(https://metrics.torproject.org/networksize.html)



Εικόνα 2-13: Πλήθος διευθύνσεων onion στο δίκτυο Tor

(https://metrics.torproject.org/hidserv-dir-onions-seen.html)

2.3.3 Invisible Internet Project

Το Tor δεν είναι το μοναδικό δίκτυο που προσφέρει ανωνυμία στις επικοινωνίες, αφού υπάρχουν και άλλες υλοποιήσεις που το επιτυγχάνουν, προστατεύοντας τόσο την πλευρά των χρηστών, όσο και την πλευρά των εξυπηρετητών. Μια από αυτές, η οποία εξυπηρετεί τους ίδιους στόχους, έχοντας παρόμοια φιλοσοφία, είναι το invisible internet project (I2P) [29].

Πρόκειται για ένα δίκτυο επικάλυψης που λειτουργεί πάνω στο διαδίκτυο, το οποίο είναι ικανό να προσφέρει ανωνυμία σε υπηρεσίες ιστοσελίδων, ηλεκτρονικού ταχυδρομείου, ανταλλαγής μηνυμάτων και διαμοιρασμού αρχείων, μέσα από την υποστήριξη πλήθους πρωτοκόλλων. Το κοινό στοιχείο που έχει με το δίκτυο Tor είναι η υλοποίηση μιας υποδομής παρόμοιας με εκείνη του onion routing, καθώς γίνεται χρήση ενδιάμεσων κόμβων, οι οποίοι διατηρούν μεταξύ τους κρυπτογραφημένες συνδέσεις. Οι κόμβοι διαθέτουν ειδικό λογισμικό με την ονομασία I2P router, με το οποίο δημιουργούνται κανάλια επικοινωνίας (tunnels), το μέγεθος των οποίων ορίζεται από τον χρήστη. Η ιδιαιτερότητα είναι πως υπάρχει δυνατότητα ενθυλάκωσης πολλαπλών μηνυμάτων σε ένα πακέτο, το οποίο είναι κρυπτογραφημένο με το κλειδί του παραλήπτη και λόγω αυτής της μορφής, η μέθοδος έχει πάρει την ονομασία garlic routing [30].

Ενώ στο Tor τα πακέτα δεδομένων είναι συγκεκριμένου μεγέθους, στο I2P υπάρχουν σκόπιμες διακυμάνσεις από κόμβο σε κόμβο, ούτως ώστε να είναι δύσκολη η παρακολούθησή τους μέσα στο δίκτυο. Αυτό επιτυγχάνεται με την προσθήκη επιπλέον δεδομένων στο φορτίο, μη σταθερού μεγέθους. Στην επιτυχία του συστήματος συμβάλλει και η εσκεμμένη καθυστέρηση που δημιουργεί κάθε κόμβος κατά την προώθηση των πακέτων, η οποία δεν είναι σταθερής διάρκειας. Ένα επιπλέον χαρακτηριστικό, που διαφοροποιεί τα δύο δίκτυα, είναι η απουσία εξυπηρετητών καταλόγου. Δεν υπάρχουν συγκεκριμένοι κόμβοι που διατηρούν λεπτομέρειες για την τοπολογία του δικτύου, αλλά εφαρμόζεται ένας τροποποιημένος αλγόριθμος Kademlia, που δίνει στην παροχή των συγκεκριμένων πληροφοριών έναν πλήρως κατανεμημένο χαρακτήρα. Τέλος, θεωρείται ως ένα σύστημα βασισμένο σε μηνύματα (message-based) και όχι βασισμένο σε κυκλώματα (circuit-based), διότι δεν υπάρχει αναμονή του χρήστη για τη δημιουργία διαδρομής μέχρι τον παραλήπτη. Αντιθέτως, τα πακέτα δεδομένων αποστέλλονται απευθείας μέσα στα προκατασκευασμένα κανάλια που συνδέουν τους κόμβους μεταξύ τους και κινούνται προς τον τελικό προορισμό τους [31].

2.3.4 Freenet

Ακόμα μια υλοποίηση που αποσκοπεί στην ελευθερία του λόγου στο διαδίκτυο είναι το Freenet, ένα δίκτυο βασισμένο στη λογική του peer-to-peer, που επιτρέπει την δημοσίευση και ανάκτηση δεδομένων, προστατεύοντας παράλληλα την ταυτότητα των χρηστών. Επίσης, προσφέρει ανωνυμία τόσο στη λειτουργία ιστοσελίδων (freesites), οι οποίες είναι διαθέσιμες μόνον εντός του δικτύου, και στην περιήγηση σε αυτές, όσο και

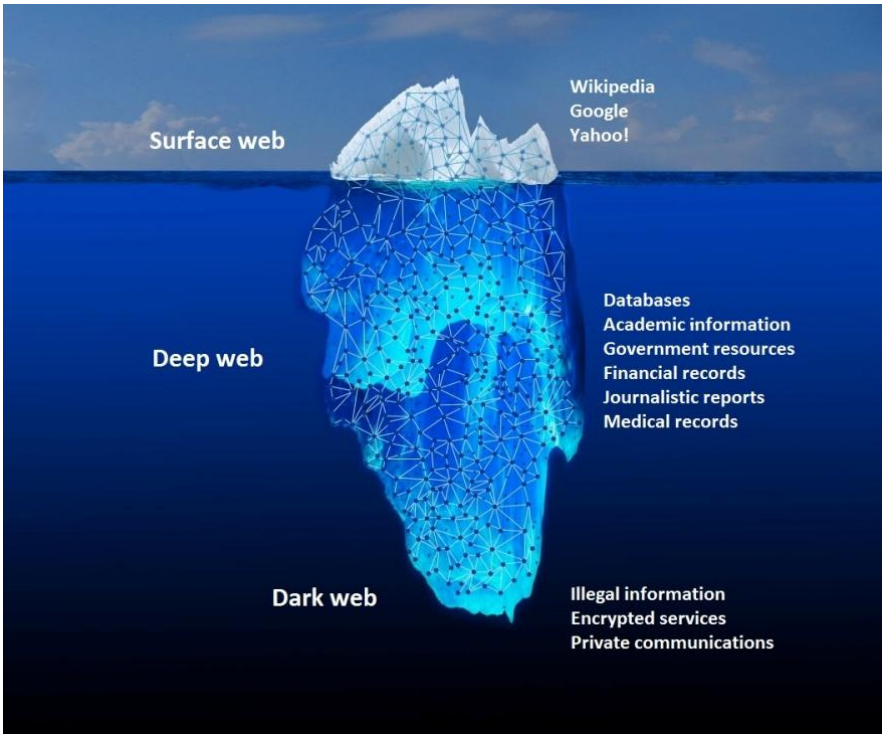
στην επικοινωνία με χρήση ομάδων συζήτησης. Κύριο χαρακτηριστικό των συστημάτων P2P είναι η αποκεντρωμένη αρχιτεκτονική που συνεπάγεται την απουσία εξυπηρετητών για την παροχή υπηρεσιών [32].

Οι κόμβοι που απαρτίζουν το δίκτυο έχουν ισότιμο ρόλο και υποχρεώσεις, μοιράζονται μεταξύ τους πόρους, όπως είναι η επεξεργαστική ισχύς, ο αποθηκευτικός χώρος δεδομένων και το εύρος ζώνης του δικτύου, εξυπηρετώντας συνεργατικά τα αιτήματα αποθήκευσης, επεξεργασίας και αναζήτησης δεδομένων. Κάθε κόμβος διατηρεί το δικό του αποθηκευτικό χώρο, τον οποίο διαθέτει προς χρήση από τους υπόλοιπους κόμβους, ενώ επίσης διαθέτει ένα δυναμικό πίνακα με τις διευθύνσεις και τα κλειδιά τους. Όπως στο δίκτυο Tor, έτσι και εδώ, τα δεδομένα μεταβιβάζονται στον προορισμό τους μέσω ενδιάμεσων κόμβων, οι οποίοι γνωρίζουν μόνον τους γειτονικούς σε αυτούς κόμβους, με τους οποίους διατηρούν κρυπτογραφημένες συνδέσεις. Κάθε αίτημα διαθέτει μηχανισμούς που αποτρέπουν τη δημιουργία ατέρμονων βρόγχων στη διαδρομή που θα ακολουθήσει και ενημερώνει τον αρχικό χρήστη για την επιτυχία ή την αποτυχία αποστολής του. Οι λειτουργίες της εισαγωγής και της ανάκτησης δεδομένων πραγματοποιούνται με βάση ένα κλειδί που παράγεται από τον κατακερματισμό της περιγραφής τους, το οποίο καταχωρείται μέσα στους πίνακες που αναφέρθηκαν προηγουμένως. Με αυτό γίνεται ομαδοποίηση των αρχείων και καθορίζεται σε ποιους κόμβους θα αποθηκευθούν αυτά, αφού πρώτα κρυπτογραφηθούν για λόγους ασφαλείας [33].

Όταν ένας κόμβος δημιουργήσει ένα αίτημα, αυτό δε σημαίνει ότι συνδέεται απευθείας στον κόμβο που θα το εξυπηρετήσει. Τα δεδομένα είναι χωρισμένα σε μικρότερα τμήματα και κατανεμημένα σε διαφορετικούς κόμβους, οπότε όταν ζητηθεί ένα αρχείο, πραγματοποιείται η συλλογή των τμημάτων και η συναρμολόγησή του, δίχως οι κόμβοι που συμμετέχουν στη διαδικασία να γνωρίζουν ποιος κατέθεσε το αίτημα. Οι ίδιοι απλά προσφέρουν το τμήμα που βρίσκεται στην κατοχή τους. Για λόγους ασφαλείας, δημιουργούνται διπλότυπα των τμημάτων σε διαφορετικούς κόμβους, καλύπτοντας την περίπτωση που κάποιος από αυτούς τεθεί εκτός λειτουργίας. Αφού ένας χρήστης πραγματοποιήσει εισαγωγή δεδομένων, αυτά είναι πλέον μέρος του δικτύου, οπότε δεν είναι απαραίτητο να είναι συνδεδεμένος ο ίδιος για να είναι διαθέσιμα [34].

2.3.5 Συμπεράσματα

Η προστασία των επικοινωνιών και η απόκρυψη της ταυτότητας των χρηστών που προσφέρουν τα δίκτυα ανωνυμίας, αποτελεί αντικείμενο εκμετάλλευσης από ένα μέρος του ψηφιακού πληθυσμού, με στόχο ενέργειες που ξεφεύγουν από τα όρια της ηθικής και της νομιμότητας. Δυστυχώς, το συγκεκριμένο υποσύνολο του παγκόσμιου ιστού κάθε άλλο παρά αβλαβές μπορεί να θεωρηθεί. Αυτό δε σημαίνει πως πρέπει να συγχέεται η έννοια του σκοτεινού με εκείνη του βαθέως ιστού, όπως γίνεται συνήθως. Ο πρώτος είναι υποσύνολο του δευτέρου και διαφοροποιούνται ως προς τον σκοπό που εξυπηρετούν.



Εικόνα 2-14: Αναπαράσταση της δομής του παγκόσμιου ιστού

Ο σκοτεινός ιστός αποτελεί ουσιαστικά ένα μικρό μέρος του παγκοσμίου ιστού, που είναι σκόπιμα κρυμμένο και είναι απαραίτητη η χρήση ειδικού λογισμικού προκειμένου να αποκτήσει κανείς πρόσβαση σε αυτό. Λόγω της μικρής πιθανότητας εντοπισμού και γνωστοποίησης της πραγματικής τους ταυτότητας, οι χρήστες του έχουν τη δυνατότητα εκτέλεσης μεγάλου εύρους ενεργειών που απαιτούν ανωνυμία. Χρησιμοποιείται από ακτιβιστές, άτομα που ζουν υπό καθεστώς λογοκρισίας και πληροφοριοδότες, όπου δηλαδή είναι αναγκαία η ελευθερία του λόγου, η παράκαμψη

της λογοκρισίας, δίχως αυτές οι τακτικές να αποτελούν απειλή. Μπορεί όμως, επίσης, να χρησιμοποιηθεί με παράνομο τρόπο, δίνοντας στους σύγχρονους εγκληματίες ένα επιπλέον εργαλείο που διευκολύνει τις δραστηριότητές τους, όπως είναι η διακίνηση και πώληση αγαθών και υπηρεσιών, κυρίως ναρκωτικών, όπλων, πλαστών εγγράφων, κτλ. Επιπρόσθετα, η ψηφιακή φύση του ιστού ωφελεί την ανάπτυξη μιας νέας κατηγορίας εγκλημάτων, βασισμένων στην τεχνολογία, με την υποκλοπή και εκμετάλλευση ευαίσθητων δεδομένων, την παιδική πορνογραφία, τη δημιουργία και διακίνηση κακόβουλου λογισμικού και την εκδήλωση κυβερνο-επιθέσεων να αποτελούν χαρακτηριστικά παραδείγματα [11]. Σε όλα αυτά έρχεται να προστεθεί η ταχύτητα συναλλαγών και η ασφάλεια που προσφέρει στην ταυτότητα των συναλλασσομένων η ύπαρξη κρυπτονομισμάτων, όπως το bitcoin [35], το οποίο χρησιμοποιείται ευρέως ως αμοιβή στις δοσοληψίες.

Ενέργειες όπως αυτές που αναφέρθηκαν, ξεφεύγουν από τα πλαίσια και τον σκοπό για τον οποίο δημιουργήθηκαν αρχικά τα δίκτυα ανωνυμίας και όπως είναι φυσιολογικό, δημιουργούν ανησυχία, έχοντας κινήσει το ενδιαφέρον των ελεγκτικών οργάνων του διαδικτύου και της δικαιοσύνης. Ειδικά τα τελευταία χρόνια, οι συγκεκριμένοι φορείς έχουν επιδοθεί στην εξεύρεση τρόπων αντιμετώπισης αυτού του συνεχώς εξελισσόμενου και ολοένα πιο επικίνδυνου προβλήματος. Μειονέκτημα στην προσπάθειά τους αποτελεί η φύση του δικτύου, που δεν επιτρέπει τον ακριβή προσδιορισμό του μεγέθους του σε σχέση με τον υπόλοιπο ιστό. Αυτό το γεγονός καθιστά πολύ δύσκολη την παρακολούθηση των παράνομων δραστηριοτήτων, τόσο ως προς το πλήθος τους, όσο και ως προς τη μεθοδολογία που ακολουθείται στην εκτέλεσή τους.

Η αποτελεσματική ανταλλαγή πληροφοριών διευκολύνει την οργάνωση, το συντονισμό και τη δράση των κακόβουλων χρηστών, βοηθώντας με αυτόν τον τρόπο την επέκταση των δραστηριοτήτων τους [11]. Όπως και στον επιφανειακό ιστό, έτσι και στον σκοτεινό, μέσα επικοινωνίας και πληροφόρησης μπορούν να αποτελέσουν οι ιστοσελίδες, οι χώροι συζήτησης (forums), τα ιστολόγια (blogs), τα κανάλια ανταλλαγής άμεσων μηνυμάτων (instant messaging) και το ηλεκτρονικό ταχυδρομείο. Τα ηλεκτρονικά καταστήματα (marketplaces) αποτελούν επίσης ένα σημαντικό τμήμα του ηλεκτρονικού εμπορίου παράνομων αγαθών και υπηρεσιών. Κάθε μια από αυτές τις υπηρεσίες προστατεύεται από την ανωνυμία του δικτύου και καταφέρνει να

λειτουργήσει αποφεύγοντας τον έλεγχο, προσφέροντας μεγάλη ποικιλία πληροφοριών σε διάφορες μορφές, όπως απλό κείμενο, εικόνες και πολυμέσα [36].

Όπως ακριβώς μπορούν να ευνοήσουν τις κακόβουλες δραστηριότητες, οι προαναφερθείσες πηγές έχουν τη δυνατότητα να συνδράμουν στην εύρεση και ανάπτυξη μεθοδολογιών και εργαλείων που θα βοηθήσουν στην αντιμετώπιση του κυβερνο-εγκλήματος, όντας κομμάτι του cyber threat intelligence. Είναι κάτι το οποίο αποτελεί επιτακτική ανάγκη, ειδικά αν λάβει κανείς υπόψη του το μέγεθος των επιπτώσεων που μπορεί να έχουν τα ολοένα και πιο εξελιγμένα είδη κυβερνο-απειλών. Οι συγκεκριμένοι χώροι, αν αξιοποιηθούν κατάλληλα, μπορούν να αποτελέσουν πηγές πληροφοριών, αποσκοπώντας στην οργάνωση, την πρόληψη, την πρόβλεψη και την προστασία απέναντι σε τέτοιου είδους απειλές.

2.4 Κατηγορίες κυβερνο-επιθέσεων

Σύμφωνα με τον ορισμό της κυβερνο-απειλής, πρόκειται οτιδήποτε έχει τη δυνατότητα να πλήξει και να προκαλέσει βλάβη σε ένα υπολογιστικό σύστημα. Η υλοποίησή της είναι εφικτή μέσω των κυβερνο-επιθέσεων, δηλαδή των κακόβουλων δραστηριοτήτων, οι οποίες εκτελούνται απέναντι σε πληροφοριακά συστήματα, δίκτυα υπολογιστών, υποδομές και προσωπικές ηλεκτρονικές συσκευές. Βασικός τους στόχος είναι, είτε η διακοπή της ομαλής λειτουργίας, η υποβάθμιση, η καταστροφή και ο κακόβουλος έλεγχος αυτών των συστημάτων, είτε η συλλογή, υποκλοπή, παραποίηση και καταστροφή πληροφοριών [37].

Η κατηγοριοποίηση των επιθέσεων είναι δυνατή με βάση τα μοναδικά τους χαρακτηριστικά και ανάλογα με τις τεχνικές που χρησιμοποιούνται. Έτσι υπάρχει ένας διαχωρισμός σε εσωτερικές (inside attacks) και εξωτερικές επιθέσεις (outside attacks). Στην πρώτη περίπτωση, η επίθεση εκτελείται από μια οντότητα που "ανήκει" στον οργανισμό και η οποία έχει αποκτήσει πρόσβαση μέχρι ένα βαθμό στο δίκτυό του, ενώ στη δεύτερη περίπτωση, η επίθεση εκτελείται από μια μη εγκεκριμένη οντότητα εκτός του οργανισμού [38].

Ένας επιπλέον διαχωρισμός μπορεί να γίνει ανάλογα με τον τρόπο εκτέλεσής τους, οπότε υπάρχουν οι ενεργητικές (active attacks) και οι παθητικές επιθέσεις (passive attacks). Οι ενεργητικές χαρακτηρίζονται από την αμεσότητά τους και κάνουν αισθητή της παρουσία τους, αδιαφορώντας για τον αν θα γίνουν αντιληπτές από το θύμα-στόχο. Σε αυτές, ο επιτιθέμενος συνήθως κλειδώνει τους χρήστες εκτός συστήματος, παίρνει

τον έλεγχο αυτού ή του δικτύου και καταστρέφει δεδομένα. Αντίθετα, στην περίπτωση των παθητικών επιθέσεων, ο επιτιθέμενος δεν γίνεται αντιληπτός από το στόχο και εισβάλλει στο σύστημα ή στο δίκτυο, συλλέγοντας δεδομένα και πληροφορίες, δίχως να διακόπτει την ομαλή λειτουργία του. Σε τέτοιες περιπτώσεις πραγματοποιείται υποκλοπή και εκμετάλλευση των δεδομένων που συλλέχθηκαν, με πιο συχνό φαινόμενο την πώλησή τους είτε στην μαύρη αγορά, είτε στο σκοτεινό διαδίκτυο [39].

2.5 Τύποι κυβερνο-επιθέσεων

Οι διαφορετικοί τύποι επιθέσεων μπορούν να κατηγοριοποιηθούν ανάλογα με τον τρόπο που πλήττουν ένα σύστημα και με τη μεθοδολογία που ακολουθείται., με τους πιο γνωστούς να είναι το λογισμικό τύπου malware, το phishing, η man-in-the-middle, η επίθεση άρνησης υπηρεσίας, η SQL injection, η zero-day exploit, η cross-site scripting και η credential reuse [40], που θα αναλυθούν στη συνέχεια.

2.5.1 **Malware** (**Mal**icious Soft**ware**)

Αυτός ο όρος χρησιμοποιείται για να περιγραφούν τα διάφορα είδη κακόβουλου λογισμικού, όπως ιοί, worms, trojans, ransomware, spyware, adware, rootkits κτλ. Αυτού του είδους το λογισμικό έχει ως στόχο, είτε να βλάψει το σύστημα του θύματος με την υποκλοπή ή καταστροφή ευαίσθητων δεδομένων, είτε την παρακολούθηση των ενεργειών του χρήστη, είτε ακόμα και την λήψη του ελέγχου του συστήματος [41]. Οι μέθοδοι με τις οποίες μπορεί να προσβληθεί ένας υπολογιστής έχουν διάφορες μορφές, αλλά στο τέλος απαιτούν πάντα από τον χρήστη να προβεί σε κάποια ενέργεια, όπως η εκτέλεση και εγκατάσταση λογισμικού. Αυτό μπορεί να γίνει με το κατέβασμα και το άνοιγμα ενός "αθώου" συνημμένου αρχείου, καθώς και με την εκτέλεση κάποιου πρόσθετου που προτείνεται από μια μολυσμένη ιστοσελίδα [42]. Όσον αφορά τα διαφορετικά είδη κακόβουλου λογισμικού, ακολουθεί παρακάτω μια σύντομη περιγραφή των κυριότερων μορφών αυτού.

Ιός: Λογισμικό, το οποίο είναι κρυμμένο μέσα σε κάποιο άλλο, φαινομενικά αβλαβές λογισμικό και δημιουργεί αντίγραφα του εαυτού του. Αυτά είναι σχεδιασμένα με τέτοιο τρόπο ώστε να μεταδίδονται, να εξαπλώνονται και να ενσωματώνονται σε άλλο λογισμικό, περνώντας δικτυακά από τον έναν υπολογιστή στον άλλο. Στόχος είναι η δυσλειτουργία των συστημάτων και η καταστροφή των δεδομένων [43].

Worm: Έχει παρόμοια λογική με τον ιό, αφού και αυτό δημιουργεί αντίγραφα του εαυτού του και έχει ως κύριο στόχο να πλήξει τα συστήματα και να καταστρέψει

δεδομένα. Η διαφορά τους εντοπίζεται στο γεγονός ότι είναι αυτόνομο και δεν απαιτείται η ύπαρξη άλλου λογισμικού. Η εξάπλωση γίνεται μέσω εκμετάλλευσης πιθανών ευπαθειών των συστημάτων ή με τη χρήση τεχνικών social engineering, οπότε ο χρήστης πέφτει στην παγίδα εκτέλεσής του [44].

Trojan: Η μορφή του είναι τέτοια που πείθει τον χρήστη ότι είναι χρήσιμο, προκειμένου αυτός να προχωρήσει στην εγκατάστασή του. Η διαφορά με τον ιό είναι πως δεν έχει ως στόχο την εξάπλωση και την μόλυνση άλλων αρχείων. Σκοπός είναι η υποκλοπή και η διαγραφή αρχείων, καθώς και η δημιουργία ευπαθειών στα συστήματα, όπως backdoors, ώστε ο κακόβουλος χρήστης να έχει μελλοντικά πρόσβαση σε αυτά [44].

Ransomware: Λογισμικό που κρυπτογραφεί τα δεδομένα του χρήστη, μην επιτρέποντας την πρόσβαση σε αυτά, και στη συνέχεια απαιτεί την καταβολή ενός χρηματικού αντιτίμου για την αποκρυπτογράφηση και ανάκτησή τους. Επίσης, σε μερικές περιπτώσεις, εμποδίζει τον χρήστη από το να εισέλθει στο σύστημά του. Ο τρόπος με τον οποίο μεταδίδεται είναι είτε μέσω phishing emails, είτε μέσω ιστοσελίδων που περιέχουν κακόβουλο κώδικα [45].

Spyware: Συνήθως είναι κρυμμένο μέσα σε άλλο λογισμικό και έτσι ο χρήστης το εγκαθιστά δίχως να γίνεται αντιληπτό. Χρησιμοποιείται για τη συλλογή δεδομένων και την αποστολή τους σε κάποια άλλη δικτυακή οντότητα, χωρίς ο ίδιος να το γνωρίζει [44]. Επίσης, καταγράφει τις συνήθειές του και τις τοποθεσίες που έχει επισκεφθεί στο διαδίκτυο. Μια από τις πλέον γνωστές μορφές spyware είναι το keylogger, λογισμικό που παρακολουθεί την πληκτρολόγηση του χρήστη και μπορεί να επιτύχει την καταγραφή ευαίσθητων δεδομένων, όπως είναι οι κωδικοί πρόσβασης [46].

Adware: Λογισμικό που προβάλει διαφημίσεις στον χρήστη, οι οποίες συνήθως εμφανίζονται κατά τη διάρκεια της εγκατάστασης μιας εφαρμογής και μπορεί να έχουν τη μορφή αναδυόμενων παραθύρων. Ο συγκεκριμένος τύπος χαρακτηρίζεται ως κακόβουλο λογισμικό, διότι εγκαθίσταται χωρίς τη συγκατάθεση του χρήστη, έχοντας τη μορφή ενοχλητικών διαφημίσεων ή παραθύρων που δεν είναι δυνατόν να κλείσουν [44].

Rootkit: Πακέτο λογισμικού, το οποίο λειτουργεί βοηθητικά κατά την προσβολή ενός συστήματος από malware. Έχει την ικανότητα να επιτρέπει στο κακόβουλο λογισμικό να παραμένει μη ανιχνεύσιμο σε ελέγχους, λόγω του ότι βρίσκεται πολύ κοντά στον πυρήνα του λειτουργικού συστήματος. Στόχος του είναι η εγκατάσταση των

απαραίτητων εργαλείων, τα οποία θα δώσουν τη δυνατότητα στον κακόβουλο χρήστη να αποκτήσει μελλοντικά απομακρυσμένη πρόσβαση στο σύστημα του θύματός του [47].

Backdoor: Αποτελεί τακτική που ακολουθείται στην ανάπτυξη συστημάτων λογισμικού και δίνει τη δυνατότητα απομακρυσμένης πρόσβασης σε αυτό, από τους δημιουργούς του, για την διενέργεια διαδικασιών επίλυσης προβλημάτων, αναβαθμίσεων και ελέγχων. Αυτές οι δίοδοι πρόσβασης μπορούν να μετατραπούν σε ευπάθειες και αποτελούν στόχο του κακόβουλου χρήστη, ο οποίος προσπαθεί να τις ανακαλύψει με τη χρήση worm ή trojan. Εκμεταλλευόμενος την ύπαρξής τους, αποκτά την ικανότητα να παρακάμπτει τις διαδικασίες αυθεντικοποίησης του συστήματος και να έχει πρόσβαση σε αυτό [48].

2.5.2 Phishing

Λόγω της συνεχούς αύξησης του αριθμού των χρηστών που αντιλαμβάνονται την επικινδυνότητα του να ανοίξουν ένα συνημμένο αρχείο που περιέχεται σε ένα email ή να ακολουθήσουν έναν σύνδεσμο που δεν είναι ασφαλής, οι κακόβουλοι χρήστες καταφεύγουν συχνά στη μέθοδο του phishing. Πρόκειται για μια τακτική που εφαρμόζεται συνήθως με τη χρήση emails, σύμφωνα με την οποία το μήνυμα περιέχει τα στοιχεία ενός αποστολέα που ο χρήστης θα εμπιστευόταν, όπως είναι μια τράπεζα ή ένας επαγγελματικός συνεργάτης. Η εμφάνιση του μηνύματος έχει τη μορφή που θα είχε ένα νόμιμο e-mail και περιλαμβάνει κάποιο συνημμένο αρχείο ή σύνδεσμο. Στην πρώτη περίπτωση επιτυγχάνεται η εγκατάσταση του κακόβουλου λογισμικού όταν ο χρήστης ανοίξει το αρχείο [49]. Στη δεύτερη περίπτωση ο σύνδεσμος οδηγεί σε μία πλαστή ιστοσελίδα, ίδια σε εμφάνιση με εκείνη πχ της τράπεζας, όπου στόχος είναι η υποκλοπή των διαπιστευτηρίων του χρήστη ή των στοιχείων της πιστωτικής του κάρτας. Ανάλογα με την τεχνική που χρησιμοποιείται και τον στόχο του phishing, είναι δυνατός ο διαχωρισμός στις παρακάτω βασικές κατηγορίες.

Deceptive Phishing: Αποτελεί τον πιο συνηθισμένο τύπο phishing, ο οποίος κάνει χρήση των emails, έχει γενικό χαρακτήρα και καλεί το θύμα του να προβεί σε επιβεβαίωση των διαπιστευτηρίων του, ακολουθώντας ένα σύνδεσμο που περιέχεται στο μήνυμα [49].

Spear Phishing: Πρόκειται για στοχευμένη υλοποίηση, η οποία απευθύνεται σε συγκεκριμένα άτομα μιας εταιρείας, κάνοντας χρήση του ονόματος, της θέσης, των στοιχείων επικοινωνίας και οποιασδήποτε άλλης πληροφορίας θα πείσει το θύμα για την

αυθεντικότητα του μηνύματος. Συχνά είναι το πρώτο βήμα στη διαδικασία παράκαμψης της άμυνας ενός εταιρικού στόχου [49].

Whaling: Είναι η συνέχεια της προηγούμενης κατηγορίας και στόχος είναι τα διευθυντικά στελέχη μιας εταιρείας. Απαιτεί μια προεργασία, σύμφωνα με την οποία παρακολουθείται το θύμα για ένα μεγάλο χρονικό διάστημα, κατά το οποίο συλλέγονται πληροφορίες για αυτό, ψάχνοντας ταυτόχρονα την κατάλληλη ευκαιρία υποκλοπής των διαπιστευτηρίων του [49].

Pharming: Σε αυτήν την περίπτωση δεν είναι απαραίτητη η ύπαρξη συνδέσμου που θα οδηγεί στην πλαστή σελίδα. Με την μόλυνση είτε του συστήματος του χρήστη, είτε του DNS sever της ίδιας της σελίδας, όταν αυτός προσπαθήσει να μεταβεί σε αυτήν, θα οδηγηθεί κατευθείαν και με αυτόματο τρόπο σε πλαστή σελίδα [49].

2.5.3 Man-in-the-middle (MITM)

Η επίθεση man-in-the-middle αποτελεί έναν τύπο επίθεσης, όπου υπάρχουν δύο πλευρές σε ένα κανάλι επικοινωνίας και ο κακόβουλος χρήστης παρεμβάλλεται ανάμεσά τους χωρίς να γίνει αντιληπτός. Έτσι, λειτουργεί ως αναμεταδότης των μεταξύ τους μηνυμάτων, ενώ οι ίδιες θεωρούν πως η επικοινωνία τους είναι άμεση. Με την παρεμβολή και την αποκρυπτογράφηση των μηνυμάτων, έχει τη δυνατότητα παρακολούθησης, αλλοίωσης και παραποίησής τους [50]. Βασικό συστατικό της διαδικασίας αποτελεί η εγκατάσταση κακόβουλου λογισμικού στα συστήματα των θυμάτων, το οποίο συνήθως γίνεται με την τεχνική του phishing. Όταν το θύμα εγκαταστήσει το λογισμικό, αυτό ενεργεί μέσω του περιηγητή και καταγράφει τα δεδομένα επικοινωνίας, αποστέλλοντάς τα στη συνέχεια τον κακόβουλο χρήστη [49].

Οι βασικές κατηγορίες αυτού του τύπου επίθεσης είναι οι εξής:

Rogue Access Point: Ο κακόβουλος χρήστης είτε εκμεταλλεύεται κάποιο δημόσιο ή ιδιωτικό ασύρματο δίκτυο με χαμηλό επίπεδο ασφάλειας, είτε παρασύρει τους ανυποψίαστους χρήστες με ένα δικό του, ελεύθερο access point. Όταν συνδεθούν σε αυτό, εκείνος μπορεί πλέον να παρεμβληθεί στην επικοινωνία τους, να εγκαταστήσει στα συστήματά τους κακόβουλο λογισμικό και να χρησιμοποιήσει τα εργαλεία αποκρυπτογράφησης των δεδομένων [51].

ARP Spoofing: Το πρωτόκολλο ARP (address resolution protocol) χρησιμεύει για την αντιστοίχιση της IP διεύθυνσης ενός τερματικού ή μιας συσκευής, με την φυσική διεύθυνσή της (MAC Address). Με αυτόν τον τρόπο υλοποιείται η επικοινωνία μεταξύ

των σταθμών σε ένα τοπικό δίκτυο, οπότε όταν ένας σταθμός επιθυμεί να επικοινωνήσει με έναν άλλο, αναζητεί την MAC που αντιστοιχεί στην IP διεύθυνσή του και σε περίπτωση που αυτή δεν είναι διαθέσιμη, αποστέλλει αίτημα ανάκτησής της. Σε αυτήν την επικοινωνία γίνεται χρήση ενός πίνακα (ARP Table) που περιέχει τις πληροφορίες αντιστοίχισης των IP και MAC διευθύνσεων για όλες τις συσκευές που είναι μέρος του δικτύου [52]. Ο κακόβουλος χρήστης τροποποιεί τον πίνακα και δρομολογεί προς τον ίδιο τα πακέτα δεδομένων που ανταλλάσσουν οι χρήστες μεταξύ τους [53].

DNS Spoofing: Όμοια με την αντιστοίχιση της IP διεύθυνσης με τη φυσική διεύθυνση MAC σε ένα τοπικό δίκτυο, υπάρχει αντιστοίχιση της IP διεύθυνσης με ένα domain name. Το θύμα τροφοδοτείται με λανθασμένα δεδομένα που αφορούν το DNS της ιστοσελίδας που επιθυμεί να επισκεφθεί, οπότε οδηγείται σε μια πλαστή σελίδα, με πιθανό σενάριο την υποκλοπή των ευαίσθητων δεδομένων του [54].

SSL Hijacking: Με τον όρο secure sockets layer (SSL) ή transport layer security (TLS) στην πιο πρόσφατη μορφή του, περιγράφεται το πρωτόκολλο κρυπτογράφησης που χρησιμοποιείται για την ασφαλή επικοινωνία εντός ενός δικτύου. Η επίθεση απέναντι σε τέτοιου είδους κανάλια επικοινωνίας δεν στοχεύει στην αποκρυπτογράφηση των δεδομένων, αλλά στην εκμετάλλευση του σύντομου χρονικού διαστήματος κατά το οποίο ανακατευθύνεται ο περιηγητής του χρήστη από επικοινωνία τύπου HTTP σε HTTPS. Πιο συγκεκριμένα, ο χρήστης μπορεί να αιτηθεί τη σύνδεση σε μια ιστοσελίδα τύπου HTTP και ο εξυπηρετητής να ανακατευθύνει αυτόματα αυτό το αίτημα, πραγματοποιώντας σύνδεση HTTPS. Ο κακόβουλος χρήστης, εκμεταλλευόμενος αυτή τη διαδικασία, παρεμβάλλεται ανάμεσα στον χρήστη και τον εξυπηρετητή, τροφοδοτώντας τον πρώτο με περιεχόμενο HTTP και ανταλλάσσοντας όλα τα απαραίτητα HTTPS δεδομένα με τον εξυπηρετητή [55].

2.5.4 Denial-of-service (DoS)

Η επίθεση άρνησης υπηρεσίας είναι ένα από τα βασικότερα είδη κυβερνο-επιθέσεων και έχει ως σκοπό την παρεμπόδιση των χρηστών από το να χρησιμοποιήσουν ένα πληροφοριακό σύστημα ή έναν διαδικτυακό πόρο, όπως μια ιστοσελίδα ή μια διαδικτυακή υπηρεσία. Αυτό επιτυγχάνεται με τη δημιουργία πολύ μεγάλης κίνησης δεδομένων και την φόρτωση του καναλιού επικοινωνίας του στόχου. Αυτός τελικά αδυνατεί να εξυπηρετήσει τα εισερχόμενα αιτήματα και διακόπτεται έτσι η παροχή υπηρεσιών [56]. Στην ίδια λογική βασίζεται και η κατανεμημένη επίθεση άρνησης

υπηρεσίας (distributed denial-of-service attack - DDoS), όπου πραγματοποιείται μια συντονισμένη επίθεση από μεγάλο αριθμό συστημάτων προς έναν κοινό στόχο [57]. Τα συστήματα αυτά ανήκουν σε ανυποψίαστους χρήστες, έχουν μολυνθεί από λογισμικό τύπου malware, όπως trojans, και βρίσκονται υπό τον έλεγχο του κακόβουλου χρήστη. Σε αυτές τις περιπτώσεις τα συστήματα των θυμάτων ονομάζονται bots, ενώ η ομάδα που σχηματίζουν για την εκτέλεση επιθέσεων ονομάζεται botnet [58].

Οι επιθέσεις DDoS μπορούν να κατηγοριοποιηθούν με βάση κάποιους παράγοντες, όπως είναι οι πόροι που χρησιμοποιούνται κατά την εκτέλεσή τους και το αν είναι άμεσες ή έμμεσες. Στην πρώτη κατηγορία ανήκει η επίθεση SYN Flood, η οποία εκμεταλλεύεται την τριπλή TCP χειραψία που εκτελείται κατά τη δημιουργία σύνδεσης με έναν εξυπηρετητή. Αυτός λαμβάνει πολλαπλά συνεχόμενα αιτήματα, τα οποία κάνουν χρήση πλαστής διεύθυνσης δικτύου, με αποτέλεσμα να μην προλαβαίνει να τα επεξεργαστεί. Κατ' επέκταση δεν εξυπηρετούνται ούτε τα νόμιμα αιτήματα. Σε αυτήν την κατηγορία ανήκει επίσης η κατανεμημένη επίθεση ICMP (Internet Control Message Protocol), γνωστή και ως επίθεση Smurf, όπου αποστέλλονται αιτήματα από έναν αριθμό μολυσμένων συστημάτων σε άλλα συστήματα, εμφανίζοντας ως διεύθυνση αποστολέα εκείνη του εξυπηρετητή-στόχου. Αυτά ανταποκρίνονται αποστέλλοντας απαντητικά πακέτα με παραλήπτη τον στόχο, με αποτέλεσμα να δημιουργείται μεγάλη κίνηση, η οποία καλύπτει όλη η χωρητικότητα του δικτύου του [59].

Στη δεύτερη κατηγορία ανήκει η άμεση επίθεση DDoS, όπου υπάρχουν δύο βαθμίδες μολυσμένων συστημάτων, κατευθυνόμενες από τον κακόβουλο χρήστη. Όταν εκείνος το επιθυμεί, ενεργοποιεί την πρώτη βαθμίδα, της οποίας τα συστήματα ενεργοποιούν με τη σειρά τους τη δεύτερη βαθμίδα και όλα μαζί αποστέλλουν πακέτα δεδομένων στον στόχο τους. Η ύπαρξη αυτής της διαβάθμισης δημιουργεί δυσκολία στον εντοπισμό του κακόβουλου χρήστη. Επιπλέον υπάρχει η ανακλαστική DDoS επίθεση, σύμφωνα με την οποία η δεύτερη βαθμίδα αποστέλλει αιτήματα σε μη μολυσμένα συστήματα, στα οποία εμφανίζεται πάλι ως αποστολέας ο στόχος και αυτά ανταποκρίνονται με την αποστολή πακέτων σε αυτόν. Αυτού του είδους η επίθεση μπορεί εύκολα να εμπλέξει μεγαλύτερο αριθμό συστημάτων σε σχέση με την άμεση DDoS επίθεση, καθιστώντας την πιο αποτελεσματική [59].

2.5.5 SQL Injection

Πρόκειται για μια τεχνική έγχυσης κώδικα που έχει ως στόχο την επίθεση σε συστήματα που χρησιμοποιούν τη γλώσσα SQL. Τέτοια συστήματα είναι οι βάσεις δεδομένων, καθώς και online εφαρμογές ή ιστοσελίδες που είναι συνδεδεμένες με μια βάση δεδομένων. Αυτός ο κώδικας εισάγεται συνήθως σε σημεία της εφαρμογής ή της σελίδας όπου ζητούνται τα διαπιστευτήρια του χρήστη. Στόχος της επίθεσης είναι να αποστείλει ερωτήματα τύπου SQL στη βάση δεδομένων, με τα οποία θα την αναγκάσει να λειτουργήσει με τρόπο που δεν προέβλεψε ο κατασκευαστής της [60]. Εφόσον το σύστημα δε διαθέτει μηχανισμούς ασφαλείας απέναντι σε τέτοιου είδους επιθέσεις, ο κακόβουλος χρήστης έχει τη δυνατότητα να το παραβιάσει και να εκμεταλλευτεί τα περιεχόμενα προς όφελός του.

Αρχικά, μια επίθεση SQL injection μπορεί να έχει ως στόχο απλά την αναγνώριση των ευαίσθητων σημείων και των ευπαθειών της βάσης δεδομένων. Έπειτα, μπορεί να είναι η απόκτηση πληροφοριών σχετικά με τον τύπο και την έκδοσή της, αφού ανάλογα με αυτά διαφοροποιούνται τα ερωτήματα που μπορεί να θέσει κανείς, ο σχεδιασμός και η δομή της. Στόχο αποτελεί επίσης η παράκαμψη των διαπιστευτηρίων εισόδου και τελικά η απόκτηση δικαιωμάτων διαχειριστή. Από αυτό το σημείο και έπειτα είναι εφικτή η ανάκτηση, η προσθήκη, η παραποίηση και η διαγραφή ευαίσθητων δεδομένων. Τέλος, πολλές είναι οι περιπτώσεις που μια επίθεση SQL injection μετατρέπεται σε DoS, εμποδίζοντας τους νόμιμους χρήστες να χρησιμοποιήσουν τη βάση [61].

2.5.6 Zero-day Exploit

Όταν μια κυβερνο-επίθεση πραγματοποιείται την ίδια ημέρα που γνωστοποιείται η ευπάθεια ενός λογισμικού, τότε αυτή έχει την ονομασία zero-day exploit. Αυτό, διότι ο επιτιθέμενος εκμεταλλεύεται τη συγκεκριμένη ευπάθεια πριν την έκδοση της απαραίτητης ενημέρωσης που θα ασφαλίσει το λογισμικό και θα το προστατεύσει..

Συνήθως, όταν ένας χρήστης ανακαλύψει ένα κενό ασφαλείας, το αναφέρει στην κατασκευάστρια εταιρεία προκειμένου εκείνη να το διορθώσει. Επίσης, υπάρχει η πιθανότητα να το αναφέρει σε κοινότητες χρηστών, ούτως ώστε να είναι και εκείνοι ενήμεροι για την ύπαρξή του. Ο κακόβουλος χρήστης προσπαθεί να ενημερωθεί εγκαίρως παρακολουθώντας αυτού του είδους τις ενημερώσεις, με στόχο να καταφέρει

να ενεργήσει πριν από την εταιρεία και εκμεταλλευόμενος το κενό ασφαλείας, να πλήξει συστήματα των χρηστών [62].

Η ύπαρξη ευπάθειας σε μια εφαρμογή μπορεί να δημιουργήσει τις προϋποθέσεις μη εξουσιοδοτημένης πρόσβασης σε ένα σύστημα και να αποτελέσει απειλή για την ασφάλειά του. Εννοείται πως από την πλευρά της η κατασκευάστρια εταιρεία δεν επιθυμεί κάτι τέτοιο για τους πελάτες της, εταιρικούς και λιανικής, αφού πιθανή δυσφήμιση θα επηρεάσει τις πωλήσεις και θα έχει και οικονομικές συνέπειες [63]. Όπως στις προηγούμενες μορφές κυβερνο-απειλών, έτσι και εδώ, η απόκτηση πρόσβασης σε ένα σύστημα μπορεί να έχει ως στόχο ευαίσθητα δεδομένα ή την πρόκληση ζημίας στην υποδομή του και στην ομαλή λειτουργία του.

2.5.7 Cross-site Scripting (XSS)

Το cross-site scripting είναι τύπος επίθεσης που εκμεταλλεύεται τα κενά ασφαλείας σε ιστοσελίδες και διαδικτυακές εφαρμογές. Αυτά επιτρέπουν την ενσωμάτωση κακόβουλου κώδικα στο δυναμικό περιεχόμενο μιας σελίδας, ο οποίος στη συνέχεια θα εκτελεστεί στην πλευρά του χρήστη μέσω του περιηγητή του. Συνήθως εφαρμόζεται σε περιπτώσεις που αυτός καταχωρεί δεδομένα, όπως είναι οι μηχανές αναζήτησης, οι φόρμες εισόδου, όπου πληκτρολογεί τα διαπιστευτήριά του, ή οι πίνακες μηνυμάτων και σχολίων που υπάρχουν σε forums [64].

Η μεθοδολογία που ακολουθείται ξεκινάει με την ανίχνευση ύπαρξης ευπαθειών στην ιστοσελίδα ή την εφαρμογή και συγκεκριμένα στα τμήματά της που δέχονται δεδομένα από τους χρήστες. Σε αυτά είναι που τοποθετείται ο κακόβουλος κώδικας, ο οποίος μπορεί να είναι γραμμένος σε γλώσσες όπως η HTML και η JavaScript, οπότε τελικά αποτελεί μέρος της ιστοσελίδας. Όταν εκτελεστεί η λειτουργία του τμήματος στο οποίο έχει γίνει η ενσωμάτωση, όπως για παράδειγμα η επιστροφή αποτελεσμάτων της μηχανής αναζήτησης, εκτελείται και ο κώδικας. Το αποτέλεσμα μπορεί να είναι κάτι πολύ απλό και ενοχλητικό, όπως η προβολή μιας εικόνας που δεν έχει σχέση με αυτό που επιθυμεί ο χρήστης, μπορεί όμως να είναι η εγκατάσταση επιβλαβούς λογισμικού στο σύστημά του [65].

Με τη συγκεκριμένη τεχνική, ο επιτιθέμενος έχει τη δυνατότητα να εκθέσει ευαίσθητες πληροφορίες, να διαχειριστεί και να υποκλέψει τα cookies που προκύπτουν όταν ο χρήστης επισκεφθεί μια ιστοσελίδα, τα οποία μπορεί να περιέχουν πληροφορίες, όπως τα διαπιστευτήριά του. Επιπλέον, μπορεί να προβεί σε ενέργειες, στις οποίες

εκτελεστής να εμφανίζεται ο ανυποψίαστος χρήστης, πλαστογραφώντας δηλαδή την ταυτότητά του ή να εκτελέσει κακόβουλο κώδικα στο σύστημά του [64].

2.5.8 Credential reuse ή stuffing

Ο συνεχώς αυξανόμενος αριθμός εφαρμογών λογισμικού έχει ως αποτέλεσμα να είναι δύσκολο για τον μέσο χρήστη να απομνημονεύσει και να χρησιμοποιήσει αποτελεσματικά ένα πλήθος διαπιστευτηρίων που αντιστοιχούν σε αυτές. Αυτός είναι ο κύριος λόγος που μεγάλος αριθμός των χρηστών επαναχρησιμοποιεί τα διαπιστευτήριά του σε διαφορετικές εφαρμογές, τακτική που αποτελεί στόχο των κακόβουλων χρηστών. Καταφέρνοντας να αποκτήσουν πρόσβαση στα διαπιστευτήρια μιας εφαρμογής, ιστοσελίδας ή υπηρεσίας, υπάρχουν αυξημένες πιθανότητες να αποκτήσουν πρόσβαση και σε άλλες, των οποίων οι χρήστες-πελάτες είναι κοινοί [66]. Η πιο συνηθισμένη μέθοδος υποκλοπής των στοιχείων είναι μέσω εφαρμογών που έχουν δημιουργηθεί για αυτό το σκοπό, υποσχόμενες κάποιες δελεαστικές και "αθώες" λειτουργίες ώστε να παρασύρουν τους χρήστες. Δεν είναι λίγα τα περιστατικά που τέτοιους είδους δεδομένα πωλούνται στη διαδικτυακή μαύρη αγορά και αφορούν γνωστές εφαρμογές, όπως το Facebook [67].

2.6 Cyber Threat Intelligence

Η συλλογή πληροφοριών που θα βοηθήσει στη λήψη των κατάλληλων μέτρων και αποφάσεων, στοχεύοντας στην πρόληψη και την αντιμετώπιση περιστατικών που αποτελούν απειλή για τα υπολογιστικά συστήματα, είναι γνωστή με την ονομασία cyber threat intelligence (CTI).

Παρ' όλο που χρησιμοποιείται ολοένα και πιο συχνά, ο όρος CTI συνεχίζει ακόμη και σήμερα να αποτελεί μια αφηρημένη έννοια. Έχει παρατηρηθεί η χρήση της για την περιγραφή ενός μεγάλου αριθμού προϊόντων, τα οποία προωθούνται από παρόχους υπηρεσιών ασφαλείας, καθώς και σε μεθοδολογίες που έχουν ως κύριο στόχο την αντιμετώπιση των κυβερνο-απειλών. Οι υπηρεσίες και τα προϊόντα μπορεί να παρουσιάζουν μεγάλη ποικιλομορφία, ανάλογα με τον τρόπο χρήσης και το περιεχόμενό τους. Οι τέσσερις κατηγορίες του CTI που έχουν προταθεί βασίζονται στο ποιος θα είναι καταναλωτής των υπηρεσιών και είναι η στρατηγική (strategic threat intelligence), η λειτουργική (operational threat intelligence), η τακτική (tactical threat intelligence) και η τεχνική (technical threat intelligence) συλλογή πληροφοριών για απειλές.

Η στρατηγική συλλογή πληροφοριών αφορά τις υψηλές θέσεις ενός οργανισμού και πιο συγκεκριμένα τα άτομα που είναι υπεύθυνα για τη λήψη διοικητικών αποφάσεων. Η φύση της δεν είναι τόσο τεχνική, επικεντρώνεται στην ενημέρωση της ύπαρξης και στην κατανόηση της φύσεως των κυβερνο-απειλών, ώστε να ληφθούν οι κατάλληλες διοικητικές αποφάσεις για την αντιμετώπισή τους. Συνήθως έχει τη μορφή αναφορών και συνεδριάσεων [68].

Η λειτουργική συλλογή πληροφοριών αφορά πληροφορίες που σχετίζονται με συγκεκριμένες εισερχόμενες απειλές. Προσδιορίζουν τη φύση της απειλής, το άτομο που είναι υπεύθυνο γι' αυτήν, τις ικανότητές του, καθώς και τα ευαίσθητα σημεία που είναι υποψήφια να πληγούν. Στόχος είναι η μετρίαση των επιπτώσεων που θα έχει η επίθεση μέσω της λήψης των κατάλληλων μέτρων πρόληψης [68].

Η τακτική συλλογή πληροφοριών αφορά τις τακτικές που χρησιμοποιούνται στις κυβερνο-απειλές, όπως τα εργαλεία και τις μεθοδολογίες και για αυτό το λόγο είναι ίσως η πιο σημαντική συλλογή για την προστασία ενός οργανισμού. Στόχος είναι ο εντοπισμός του τρόπου με τον οποίο θα δράσει ο υπεύθυνος της απειλής και αντίστροφα, η εύρεση της ιδανικών μεθόδων και τακτικών που θα σταθούν ικανές να την εντοπίσουν και να την αποτρέψουν [68].

Η τεχνική συλλογή πληροφοριών αφορά τις τεχνικές λεπτομέρειες των εργαλείων που αναφέρθηκαν παραπάνω, καθώς και των καναλιών και της υποδομής που χρησιμοποιούνται σε μία επίθεση. Η διαφορά με την προηγούμενη κατηγορία εντοπίζεται στην υλοποίηση της μεθοδολογίας αντιμετώπισης μιας απειλής. Στην τακτική γίνεται ο εντοπισμός πχ του κακόβουλου λογισμικού που χρησιμοποιείται, ενώ στην τεχνική η χρήση δεικτών και μετρήσεων σε διάφορα σημεία του οργανισμού, όπως το δίκτυο και η ηλεκτρονική αλληλογραφία [66].

Τελικά, ανεξαρτήτως του καταναλωτή των πληροφοριών, μπορούμε να ορίσουμε ως cyber threat intelligence το σύνολο των πληροφοριών που θα βοηθήσουν στη λήψη αποφάσεων. Μέσα από αυτές, θα εκτελεστούν οι απαιτούμενες ενέργειες για την αποτροπή μιας κυβερνο-απειλών, μειώνοντας το χρόνο που απαιτείται για τον εντοπισμό της και καθιστώντας πιο ξεκάθαρο το τοπίο της επικινδυνότητας και του ρίσκου [68].

3. Βιβλιογραφική Επισκόπηση

Στη μελέτη που πραγματοποιήθηκε από τους Fu, Abbasi και Chen [69] γίνεται μια ανάλυση των λειτουργιών που πρέπει να διαθέτει ένας πετυχημένος web crawler και περιγράφεται η υλοποίησή του, με εφαρμογή σε ομάδες συζήτησης του σκοτεινού διαδικτύου. Επιλέγεται η μέθοδος της αυξητικής συλλογής δεδομένων (incremental crawling) και προτείνεται μια ημιαυτόματη προσέγγιση για την είσοδο σε αυτές, με την πολυπλοκότητα της διαδικασίας εισαγωγής διαπιστευτηρίων να καθορίζει το ποσοστό συμμετοχής του ανθρώπινου παράγοντα. Η αναζήτηση μέσα σε κάθε ιστότοπο γίνεται ταυτόχρονα από πολλαπλούς συλλέκτες, πραγματοποιείται κατά πλάτος και κατά βάθος, κάθε φορά με διαφορετικές παραμέτρους, προκειμένου να ελεγχθεί το όριο στο οποίο η διαδικασία θα μπλοκαριστεί από τους διαχειριστές. Το συμπέρασμα είναι πως όσο πιο ήπια είναι η εκτέλεση, τόσο μεγαλύτερη η χρονική διάρκεια της συλλογής δεδομένων. Τέλος, ακολουθεί η στατιστική ανάλυση όσων συλλέχθηκαν και διαπιστώνεται πως υπάρχουν κοινά μέλη σε διαφορετικές ομάδες, δημιουργώντας αμοιβαίες σχέσεις μεταξύ τους.

Στα πλαίσια της έρευνας των Benjamin, Li, Holt και Chen [70] πάνω στην άντληση πληροφοριών για απειλές και ευπάθειες, έγινε μια μελέτη γύρω από τις ομάδες συζήτησης, τις υπηρεσίες άμεσων μηνυμάτων και τα ηλεκτρονικά καταστήματα καρτών. Θεωρείται πως αυτές οι υπηρεσίες, που λειτουργούν στο σκοτεινό διαδίκτυο, μπορούν να αποτελέσουν πηγή για τη συλλογή και ανάλυση δεδομένων, έχοντας αναπτύξει τα κατάλληλα αυτοματοποιημένα εργαλεία.

Η επιλογή των υπηρεσιών βασίστηκε στο γεγονός ότι οι κακόβουλοι χρήστες, γνωστοί ως hackers, τις χρησιμοποιούν καθώς αναπτύσσουν κοινότητες για την ενίσχυση των δραστηριοτήτων τους. Εντός των ομάδων συζήτησης ανταλλάσσουν ιδέες και μεθοδολογίες για μια πληθώρα κυβερνο-επιθέσεων, σε πολλές και διάφορες μορφές, περιλαμβάνοντας οδηγίες που καθιστούν επικίνδυνο ακόμα και έναν άπειρο χρήστη. Πρέπει να σημειωθεί πως σε αυτούς τους χώρους συχνά εφαρμόζονται τεχνικές ασφαλείας από τους διαχειριστές τους, είτε αποκλείοντας την είσοδο σε μη εξουσιοδοτημένους χρήστες, είτε εμποδίζοντας τη λειτουργία λογισμικού που έχει ως στόχο την αυτοματοποιημένη συλλογή δεδομένων. Όσον αφορά τα άμεσα μηνύματα, τα οποία βασίζονται σε διαφορετικό πρωτόκολλο και προσφέρουν επικοινωνία σε πραγματικό χρόνο, υπάρχει δυσκολία στη συλλογή δεδομένων. Αυτή προκύπτει από το

γεγονός πως η διαδικασία πρέπει να γίνει την ίδια στιγμή που λαμβάνει μέρος μια συνομιλία, αφού τα περιεχόμενά της δεν διατηρούνται αποθηκευμένα για μελλοντική χρήση, όπως γίνεται στις συζητήσεις των forums. Τέλος, στα ηλεκτρονικά καταστήματα πωλούνται στοιχεία πιστωτικών καρτών, αποτέλεσμα κυβερνο-επιθέσεων και υποκλοπής προσωπικών δεδομένων. Μέσα από τις περιγραφές που συνοδεύουν τα προϊόντα, μπορούν να προκύψουν πληροφορίες για τους πωλητές και τα θύματά τους.

Το εργαλείο που αναπτύχθηκε για αυτό το σκοπό είναι βασισμένο στο AZSecure framework [71] και πραγματοποιεί επεξεργασία των συζητήσεων που υπάρχουν εντός των καναλιών επικοινωνίας, με τον υπολογισμό του βάρους των λέξεων που περιλαμβάνονται σε αυτές. Με αυτόν τον τρόπο είναι δυνατή η στατιστική ανάλυση και αξιολόγηση του περιεχομένου, εντοπίζοντας έτσι τις τάσεις που επικρατούν σχετικά με τα είδη των κυβερνο-επιθέσεων.

Αντίστοιχη έρευνα πραγματοποιήθηκε και από τους Soska και Christin, οι οποίοι επέλεξαν τη διερεύνηση των ηλεκτρονικών καταστημάτων [72]. Κατάφεραν να συλλέξουν δεδομένα από 16 διαφορετικά καταστήματα σε διάστημα άνω των δύο ετών και να αναλύσουν τις τάσεις της αγοράς, τόσο ως προς τις κατηγορίες των προϊόντων, όσο και προς την αντίστασή της απέναντι στις αρχές. Επιπλέον, μελετήθηκε η συμπεριφορά των εμπόρων με τις τεχνικές απόκρυψης που χρησιμοποιούν, οι οποίες διαρκώς βελτιώνονται.

Επιβεβαιώνεται και εδώ πως με την εκτέλεσή τους ως κρυμμένες υπηρεσίες του Tor, τα καταστήματα αποκρύπτουν την τοποθεσία τους, την ταυτότητα των διαχειριστών και των χρηστών, ενώ η χρήση κρυπτονομισμάτων καθιστά μη ανιχνεύσιμες τις μεταξύ τους συναλλαγές. Κοινά χαρακτηριστικά με την περίπτωση των ομάδων συζήτησης αποτελούν οι δικλείδες ασφαλείας των σελίδων, όπως και το γεγονός ότι εμφανίζονται κοινοί πωλητές σε διαφορετικά καταστήματα. Πέρα από τον ήπιο ρυθμό που πρέπει να έχει η διαδικασία συλλογής δεδομένων, καλό είναι να πραγματοποιείται όσο το δυνατόν ταχύτερα, ούτως ώστε να υπάρχει μικρό χρονικό χάσμα μέχρι την επεξεργασία τους κι έτσι τα συμπεράσματα να είναι πληρέστερα. Για την επεξεργασία των δεδομένων υλοποιήθηκε αλγόριθμος μηχανικής μάθησης, ο οποίος επεξεργάζεται την περιγραφή που παρέχουν οι πωλητές, αποδίδοντας ανάλογα ετικέτες, δημιουργώντας με αυτόν τον τρόπο κατηγορίες για τους πωλητές και τα προϊόντα.

Η βιβλιογραφική επισκόπηση δείχνει πως είναι εφικτή η συλλογή και επεξεργασία δεδομένων χρησιμοποιώντας ως πηγή το σκοτεινό διαδίκτυο. Οι ιδανικές

πηγές προκειμένου να προκύψουν χρήσιμες πληροφορίες όσον αφορά τις κυβερνο-επιθέσεις αποτελούν οι ομάδες συζήτησης, τα ιστολόγια και τα ηλεκτρονικά καταστήματα, λόγω του στατικού περιεχομένου τους και όχι τόσο τα κανάλια ανταλλαγής άμεσων μηνυμάτων. Στην υλοποίηση του λογισμικού που θα χρησιμοποιηθεί σε αυτή τη διαδικασία, πρέπει να ληφθούν υπόψη συγκεκριμένοι παράγοντες, όπως η ασφάλεια που υπάρχει απέναντι σε ανεπιθύμητους επισκέπτες και ο ήπιος ρυθμός ανάκτησης δεδομένων, τόσο για την αποφυγή ανίχνευσης από τους διαχειριστές, όσο και ως προς σεβασμό απέναντι στο εύρος ζώνης του δικτύου Tor.

Κατόπιν μελέτης της υπάρχουσας βιβλιογραφίας, ο στόχος της παρούσας εργασίας είναι να απαντήσει στα εξής ερωτήματα που προκύπτουν:

- Πώς μπορεί να γίνει η εξεύρεση διευθύνσεων url, είτε πρόκειται για ομάδες συζήτησης, είτε για ηλεκτρονικά καταστήματα; Οι διευθύνσεις δεν μπορούν να ανακτηθούν στο σύνολό τους μέσω μηχανών αναζήτησης, οπότε με τις υπάρχουσες μηχανές του σκοτεινού ιστού θα προκύψουν κάποια επιφανειακά αποτελέσματα.
- Ποιες είναι οι διαφορές της διαδικασίας συλλογής δεδομένων από ιστοσελίδες σε σχέση με τον επιφανειακό ιστό;
- Με ποιον επιπλέον τρόπο μπορούν να αξιοποιηθούν τα στοιχεία που θα προκύψουν για την αντιμετώπιση των κυβερνο-επιθέσεων;

4. Μεθοδολογία

4.1 Εισαγωγή

Το εργαλείο που αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας πραγματοποιεί τη συλλογή, αποθήκευση και επεξεργασία δεδομένων που προέρχονται από το δίκτυο Tor. Πιο συγκεκριμένα, πρόκειται για έναν web crawler, ο οποίος είναι σε θέση να συλλέγει και να αποθηκεύει αυτούσιο το HTML περιεχόμενο των σελίδων σε βάση δεδομένων SQL, παρέχοντας επιπλέον τη δυνατότητα εξαγωγής, αποθήκευσης και κατηγοριοποίησης του κειμένου που περιέχεται σε αυτές.

Η υλοποίηση πραγματοποιήθηκε με τη γλώσσα προγραμματισμού Python, η επιλογή της οποίας έγινε με βάση τα χαρακτηριστικά της. Πρόκειται για μια ισχυρή γλώσσα υψηλού επιπέδου, η οποία διαθέτει πολύ μεγάλη κοινότητα για την υποστήριξή της και πλήθος βιβλιοθηκών, που επικεντρώνονται στην επεξεργασία, οπτικοποίηση και στατιστική ανάλυση δεδομένων. Καταγράφεται μια διαρκής αύξηση στη δημοτικότητά της και αποτελεί στις ημέρες μας την πρώτη επιλογή όσων ασχολούνται με την επιστήμη των δεδομένων (data science), την εξόρυξη δεδομένων, τη μηχανική μάθηση και την τεχνητή νοημοσύνη.

4.2 Βιβλιοθήκες

Οι βιβλιοθήκες που χρησιμοποιήθηκαν για την κατασκευή της εφαρμογής είναι οι requests, BeautifulSoup, urlparse, socks, socket, stem, fake_useragent, sqlite3 και sklearn, των οποίων ακολουθεί μια σύντομη παρουσίαση.

4.2.1 Requests

Η γλώσσα, στην οποία είναι γραμμένη, είναι η Python και έχει ως στόχο να υλοποιεί εύκολα αιτήματα τύπου HTTP. Περιέχει λειτουργίες που πραγματοποιούν τη σύνδεση σε μια ιστοσελίδα και την ανάκτηση του HTML περιεχομένου της [73].

4.2.2 BeautifulSoup

Επίσης γραμμένη σε Python, χρησιμοποιείται για την επεξεργασία αρχείων τύπου HTML και XML, περιέχοντας λειτουργίες για την παραγωγή μιας πιο κατανοητής μορφής τους, με στόχο την εξαγωγή δεδομένων [74]. Ο συνδυασμός της με την

βιβλιοθήκη requests χρησιμοποιήθηκαν στον κώδικα που συλλέγει το HTML περιεχόμενο των ιστοσελίδων.

4.2.3 Urlparse

Με την urlparse γίνεται η επεξεργασία των urls και η διάσπασή τους στα συστατικά από τα οποία αποτελούνται. Με αυτόν τον τρόπο είναι δυνατή η απομόνωση τμημάτων τους, όπως είναι το domain ή το path και ο συνδυασμός τους για τη δημιουργία ενός νέου url, ανάλογα με τις εκάστοτε ανάγκες [75]. Στην εφαρμογή, χρησιμοποιήθηκε για την επεξεργασία των συνδέσμων που ανακτώνται σε κάθε ιστοσελίδα.

4.2.4 Socks

Με τη χρήση της βιβλιοθήκης socks παρέχεται η δυνατότητα χρήσης του SOCKS ως πρωτόκολλο επικοινωνίας, το οποίο όπως αναφέρθηκε στο κεφάλαιο 2.3.2 χρησιμοποιείται στο Tor για την ύπαρξη συμβατότητας μεταξύ των εφαρμογών και του onion proxy, αποτελώντας μια ιδανική λύση στο μοντέλο πελάτη-εξυπηρετητή.

4.2.5 Socket

Η socket παρέχει τη δυνατότητα ορισμού της θύρας που θα χρησιμοποιηθεί σε μια σύνδεση. Συνδυάζεται με την socks για τη δημιουργία σύνδεσης στο δίκτυο Tor και την προώθηση αιτημάτων μέσω των κυκλωμάτων του.

4.2.6 Stem

Πρόκειται για ένα API (Application Programming Interface) που χρησιμοποιείται στη διαχείριση των συνδέσεων και έχει δημιουργηθεί ειδικά για το δίκτυο Tor. Με την βιβλιοθήκη stem υλοποιείται η λειτουργία αλλαγής κυκλώματος στην προσπάθεια αποφυγής εντοπισμού κατά τη διάρκεια συλλογής δεδομένων [76].

4.2.7 Fake_useragent

Η fake_useragent διαθέτει τη δυνατότητα διαχείρισης του user agent, της εφαρμογής δηλαδή που εμφανίζεται ως πελάτης προς μια ιστοσελίδα. Ο ρόλος του είναι να ενημερώνει με στοιχεία όπως τον τύπο της εφαρμογής, την έκδοσή της, το

λειτουργικό σύστημα που χρησιμοποιείται, η προεπιλεγμένη γλώσσα, κτλ [77]. Επίσης μέρος της προσπάθειας αποφυγής εντοπισμού της διαδικασίας συλλογής δεδομένων.

4.2.8 Sqlite3

Γραμμένη στη γλώσσα προγραμματισμού C, η sqlite3 αποτελεί τη βιβλιοθήκη για την υλοποίηση σχεσιακών βάσεων δεδομένων SQL και για την εκτέλεση ερωτημάτων, δίχως την ύπαρξη ειδικού εξυπηρετητή ή επιπλέον διεργασιών [78].

4.2.9 Sklearn

Η sklearn είναι μια βιβλιοθήκη γραμμένη στις γλώσσες Python, Cython, C και C++ και περιέχει τα απαραίτητα εργαλεία για την υλοποίηση αλγορίθμων μηχανικής μάθησης. Μέσα από αυτήν, χρησιμοποιήθηκαν τα τμήματα που αφορούν τους αλγόριθμους TF-IDF και K-Means [79]. Με τον πρώτο έγινε επεξεργασία και δόθηκαν βάρη σε λέξεις που περιέχονται σε κείμενα σχετικά με κυβερνο-επιθέσεις, με στόχο την εξεύρεση των πιο σημαντικών όρων μέσα σε αυτά. Με το δεύτερο εφαρμόστηκε ο αλγόριθμος K-Means προκειμένου να γίνει ο διαχωρισμός τους σε συστάδες, μια για κάθε ένα από τα διαφορετικά είδη κυβερνο-απειλών, προκειμένου να είναι δυνατή η κατασκευή ενός εργαλείου κατηγοριοποίησης κειμένου.

4.3 Σχεδίαση εφαρμογής

Η εφαρμογή αποτελείται από τρία μέρη, το πρώτο από τα οποία έχει ως στόχο τη δημιουργία μιας λίστας, η οποία αποτελείται από urls κρυμμένων υπηρεσιών, ελέγχοντας ταυτόχρονα αν αυτά είναι ενεργά. Το δεύτερο μέρος κάνει χρήση των ενεργών urls που προέκυψαν προηγουμένως, συλλέγοντας δεδομένα με τρόπο που επιλέγει ο χρήστης, εξάγοντας χρήσιμες πληροφορίες και τέλος, το τρίτο μέρος πραγματοποιεί την κατηγοριοποίησή τους.

4.3.1 Δημιουργία λίστας

Καθώς η εύρεση σελίδων με χρήση των μηχανών αναζήτησης, που είναι μέρος του σκοτεινού διαδικτύου, έχει γενικά περιορισμένα αποτελέσματα, κάποιες από αυτές χρησιμοποιήθηκαν απλά για την εύρεση ενός αριθμού ιστοσελίδων, οι οποίες στη συνέχεια αποτέλεσαν την πηγή για τη δημιουργία της λίστας. Ομοίως, έγινε χρήση

μηχανής αναζήτησης που ανήκει στον επιφανειακό. Οι σελίδες που προέκυψαν και επιλέχθηκαν, περιέχουν συνδέσμους κρυμμένων υπηρεσιών και αποτελούν αποτέλεσμα μεμονωμένων προσπαθειών για τη διευκόλυνση της περιήγησης των χρηστών στον σκοτεινό ιστό. Η εφαρμογή εντοπίζει όλους τους συνδέσμους σε κάθε μία από αυτές τις σελίδες και τους αποθηκεύει στη βάση δεδομένων. Από τη διαδικασία θα προκύψει μια ενιαία λίστα με urls, η οποία θα χρησιμοποιηθεί για την εύρεση ιστοτόπων με πληροφορίες γύρω από τις κυβερνο-επιθέσεις. Κατά την καταχώρηση ενός συνδέσμου εκτελούνται δύο λειτουργίες, ο έλεγχος για το αν αυτός υπάρχει ήδη μέσα στη βάση, οπότε παραβλέπεται, και ο έλεγχος για το αν είναι ενεργός, αν υπάρχει δηλαδή επικοινωνία με την ιστοσελίδα στην οποία οδηγεί, οπότε καταχωρείται η ανάλογη ένδειξη. Έχοντας απομονώσει τις ενεργές σελίδες, ακολουθεί η επίσκεψη και συλλογή δεδομένων από αυτές.

4.3.2 Λειτουργίες συλλογής δεδομένων

Η εφαρμογή δίνει τη δυνατότητα στον χρήστη να επιλέξει τον τρόπο ανάκτησης των σελίδων, υλοποιώντας συνολικά τρεις λειτουργίες. Η πρώτη είναι η τυχαία αναζήτηση ανάμεσα στις σελίδες ενός ιστότοπου, όπου ζητείται το url της αρχικής σελίδας (seed) και ο επιθυμητός αριθμός σελίδων προς ανάκτηση. Στη συνέχεια, αποθηκεύεται το HTML περιεχόμενό της και δημιουργείται μια λίστα με τους συνδέσμους που περιέχονται σε αυτήν, μέσα από την οποία επιλέγεται με τυχαίο τρόπο η επόμενη προς ανάκτηση σελίδα. Η ίδια διαδικασία επαναλαμβάνεται για κάθε μία σελίδα που επισκέπτεται η εφαρμογή και μέχρι την ολοκλήρωση του επιθυμητού αριθμού σελίδων.

Η δεύτερη επιλογή είναι η σειριακή αναζήτηση. Αυτό είναι ιδιαίτερα χρήσιμο σε περίπτωση που είναι επιθυμητή η αναζήτηση σε forums και ο χρήστης θέλει να ανακτήσει μόνον τις σελίδες που αφορούν μια συζήτηση, αλλά όχι σελίδες που δε σχετίζονται με αυτήν. Λόγω της γενικής ποικιλομορφίας που υπάρχει για τους συνδέσμους που οδηγούν στις επόμενες σελίδες (πχ p=2, page=2, page-2, κοκ), η εφαρμογή δίνει τη δυνατότητα στον χρήστη να πληκτρολογήσει τη μορφή της επόμενης σελίδας. Πληκτρολογείται λοιπόν η αρχική σελίδα, η μορφή του συνδέσμου της επόμενης σελίδας και ο συνολικός αριθμός τους, οπότε η διαδικασία ανάκτησης περιεχομένου εκτελείται μέχρι τη στιγμή της ολοκλήρωσης αυτής της αλυσίδας συνδέσμων.

Τέλος, η τρίτη επιλογή πραγματοποιεί την αναζήτηση κατά πλάτος. Με βάση την αρχική σελίδα και το επιθυμητό βάθος, η εφαρμογή δημιουργεί και εδώ μια λίστα συνδέσμων προς ανάκτηση, ανάλογα με το βάθος που ορίζει ο χρήστης. Εννοείται πως όσο πιο μεγάλο το βάθος που επιλέγεται, τόσο πιο απαιτητική γίνεται η διαδικασία ανάκτησης όσον αφορά τον χρόνο, οπότε για λόγους απόδοσης είναι προτιμότερο να επιλέγεται μικρό βάθος.

Πρέπει να σημειωθεί πως σε κάθε μία από τις παραπάνω περιπτώσεις, εκτελούνται παράλληλα δύο επιπλέον λειτουργίες. Παράλληλα με την αποθήκευση του HTML περιεχομένου, λαμβάνει μέρος η απομόνωση, επεξεργασία και αποθήκευση του κειμένου που περιέχεται σε αυτόν. Εντοπίζονται όλα τα δεδομένα κειμένου και συγκεκριμένα εκείνα που έχουν μέγεθος άνω των 30 χαρακτήρων, οπότε μεμονωμένες λέξεις ή πολύ μικρού μήκους προτάσεις παραβλέπονται. Αποτέλεσμα είναι να αποθηκεύεται στη βάση καθαρό κείμενο με περιγραφές και συζητήσεις χρηστών, από τις οποίες μπορούν να προκύψουν στοιχεία που αφορούν τις κυβερνο-απειλές. Η διατήρηση του κώδικα HTML έχει ως στόχο την παροχή δυνατότητας περαιτέρω επεξεργασίας των δεδομένων, με επιπλέον μεθόδους που τυχόν προκύψουν μελλοντικά.

Όσον αφορά τη δεύτερη λειτουργία, πραγματοποιείται έλεγχος για την ύπαρξη συνδέσμων που οδηγούν σε άλλους ιστότοπους. Όταν αυτό ισχύει, γίνεται έλεγχος και αποθήκευσή τους στη λίστα που περιγράφηκε στην προηγούμενη παράγραφο. Με αυτόν τον τρόπο, πραγματοποιείται η εξεύρεση συνδέσμων που μπορεί να προκύψουν μέσα από την ανταλλαγή πληροφοριών μεταξύ χρηστών.

4.3.3 Κατηγοριοποίηση

Το τελευταίο βήμα στη διαδικασία είναι ο εντοπισμός της κατηγορίας απειλής στην οποία ανήκουν τα δεδομένα κειμένου που έχουν προκύψει από την παραπάνω επεξεργασία, με βάση την κατηγορία της κυβερνο-επίθεσης στην οποία αναφέρονται. Για αυτό το σκοπό, δημιουργήθηκε ένα μοντέλο που περιέχει την χρήση μηχανικής μάθησης, εφαρμόζοντας τον αλγόριθμο των K-μέσων (K-Means).

Ο συγκεκριμένος είναι ένας αλγόριθμος μηχανικής μάθησης χωρίς επίβλεψη, που σημαίνει πως δεν εκπαιδεύεται και μην έχοντας κάποια προηγούμενη εμπειρία, αναζητεί ομοιότητες στα χαρακτηριστικά των δεδομένων εισόδου. Προσπαθεί, δηλαδή, να ανακαλύψει την ύπαρξη προτύπων και με βάση αυτά, να τα ταξινομήσει σε έναν συγκεκριμένο πλήθος συστάδων, ο αριθμός των οποίων έχει οριστεί εκ των προτέρων.

Ξεκινώντας από την επιλογή τυχαίων μέσων, που ονομάζονται κεντροειδή, το πλήθος των οποίων είναι ίσο με εκείνο των επιθυμητών συστάδων, ομαδοποιούνται τα δεδομένα σε αυτές. Έπειτα, γίνεται μετατόπιση αυτών των μέσων και επανάληψη της ομαδοποίησης. Η διαδικασία επαναλαμβάνεται έως ότου η μετατόπιση να είναι μικρότερη μιας τιμής κατωφλίου που έχει δοθεί [80].

Για την υλοποίηση της εφαρμογής και επειδή τα κείμενα που συλλέχθηκαν δεν αφορούσαν αποκλειστικά τις κυβερνο-απειλές, υλοποιήθηκε ένα εργαλείο που δημιουργεί λίστες από λέξεις-κλειδιά μέσω του K-Means, τις οποίες στη συνέχεια χρησιμοποιεί προκειμένου να αποδώσει ετικέτα σε κείμενο που δέχεται ως είσοδο. Για αυτό το λόγο, δόθηκε στον αλγόριθμο ένα πλήθος από διαδικτυακά κείμενα του επιφανειακού ιστού, που έχουν ως θέμα τους διάφορους τύπους κυβερνο-επιθέσεων, και από την εκτέλεσή του προέκυψε μια σειρά από συστάδες με λέξεις-κλειδιά για κάθε μία από αυτές. Οι τύποι που αφορούν είναι εκείνοι της παραγράφου 2.5, μαζί με την εξαπάτηση που αφορά πιστωτικές κάρτες και τα ηλεκτρονικά καταστήματα. Εννοείται, πως όσο πιο πολλά είναι τα δεδομένα που χρησιμοποιούνται για την εκτέλεση του αλγορίθμου, τόσο πιο αποτελεσματική η συσταδοποίησή τους και πιο έγκυρες οι λίστες που θα προκύψουν. Στη συνέχεια, οι λέξεις-κλειδιά αποτελούν το φίλτρο με το οποίο δίδονται ετικέτες στα κείμενα που έχουν αποθηκευτεί στη βάση δεδομένων, επιτυγχάνοντας με αυτόν τον τρόπο την κατηγοριοποίησή τους.

4.3.4 Δομή βάσης δεδομένων

Για τη δημιουργία της βάσης δεδομένων χρησιμοποιήθηκε η βιβλιοθήκη SQLite και δεν απαιτείται η ύπαρξη προεργασίας με ειδικό λογισμικό. Με αυτόν τον τρόπο η εφαρμογή παρουσιάζει φορητότητα, ευκολία στη χρήση και χαμηλές απαιτήσεις σε πόρους συστήματος.

Κατά την πρώτη εκτέλεση της εφαρμογής δημιουργείται αυτόματα το αρχείο της βάσης και σε περίπτωση που υπάρχει ήδη αρχείο, η εφαρμογή ενημερώνει τον χρήστη εμφανίζοντας δύο επιλογές. Η πρώτη δίνει τη δυνατότητα συνέχισης των εργασιών στην υπάρχουσα βάση και η δεύτερη τη δυνατότητα έναρξης νέας βάσης, διαγράφοντας την παλιά. Η βάση αποτελείται από τον πίνακα που προκύπτει από το πρώτο μέρος της εφαρμογής και περιέχει τις διευθύνσεις url με την κατάστασή τους, καθώς και από τον πίνακα που προκύπτει από το δεύτερο μέρος, ο οποίος περιλαμβάνει τον κώδικα HTML των σελίδων και το κείμενο που περιέχεται σε αυτόν.

Για την προβολή του αρχείου μπορεί να χρησιμοποιηθεί η εφαρμογή "DB Browser for SQLite", της οποίας η εγκατάσταση δεν απαιτεί κάποια ειδική παραμετροποίηση και παρέχει έναν μικρό αριθμό εργαλείων διαχείρισης.

4.3.5 Επιπλέον λειτουργικά χαρακτηριστικά

Η πλειοψηφία των ιστοτόπων του παγκοσμίου ιστού διαθέτει το αρχείο με την ονομασία robots.txt, στο οποίο ο διαχειριστής αναφέρει ποια τμήματα επιθυμεί να παραληφθούν από τις μηχανές αναζήτησης και σε ποιους crawlers απαγορεύει την είσοδο. Είναι κάτι που πρέπει κανείς να σεβαστεί, καθώς σε αντίθετη περίπτωση ο διαχειριστής έχει το δικαίωμα να εφαρμόσει αμυντικές τακτικές, όπως τον αποκλεισμό της IP διεύθυνσης του crawler, καθιστώντας αδύνατη τη συλλογή δεδομένων. Η εφαρμογή πάντα ως πρώτο βήμα πραγματοποιεί τον έλεγχο του συγκεκριμένου αρχείου και αποφεύγει όσα τμήματα περιλαμβάνονται σε αυτό.

Καλή τακτική επιτυχημένου crawling ώστε η αναζήτηση να μην τερματιστεί πρόωρα, αποτελεί η τυχαιότητα του τρόπου περιήγησης αποφεύγοντας κάποιο μοτίβο. Αυτός είναι ο λόγος που στη λειτουργία τυχαίας συλλογής δεδομένων που περιγράφηκε, η επόμενη σελίδα επιλέγεται με τυχαίο τρόπο, στοχεύοντας στο να μην γίνει αντιληπτή η εκτέλεση της διαδικασίας από μη ανθρώπινο παράγοντα. Σε αυτό συνεισφέρει επιπλέον η προσθήκη παύσεων στη μετάβαση μεταξύ των σελίδων, διάρκειας λίγων δευτερολέπτων, που φροντίζουν να μην υπάρξει υπερφόρτωση του ιστότοπου. Η ορθή διαχείριση του εύρους ζώνης είναι κάτι που πρέπει να εφαρμόζεται και στη συγκεκριμένη περίπτωση αποτελεί τακτική σεβασμού προς το ίδιο το δίκτυο Tor και τους κόμβους που το απαρτίζουν.

Τέλος, μια ακόμα χρήσιμη λειτουργία είναι οι συχνές αλλαγές της IP διεύθυνσης και του user agent που εμφανίζονται από την πλευρά του crawler. Με αυτόν τον τρόπο δεν εμφανίζεται η πρόσβαση σε μεγάλο αριθμό σελίδων από το ίδιο σημείο. Υπάρχουν σελίδες που αντιλαμβάνονται την αυτοματοποιημένη κίνηση και, είτε αποκλείουν τη διεύθυνση IP του επισκέπτη, είτε τον παγιδεύουν με εγκλωβισμό σε ατέρμονο κύκλο μεταξύ σελίδων. Έχει υλοποιηθεί λειτουργία, όπου κατόπιν αποθήκευσης μιας σελίδας και κατά τη διάρκεια της παύσης, γίνεται αλλαγή του κυκλώματος και του user agent που εμφανίζεται από τον crawler προς τη σελίδα.

Αυτό που πρέπει να λαμβάνεται υπόψη είναι πως στο σκοτεινό διαδίκτυο βασικός στόχος των χρηστών είναι η ανωνυμία των δραστηριοτήτων τους, η εκτέλεσή τους μακριά από τα αδιάκριτα μάτια και όχι τόσο η εφαρμογή κανόνων, οπότε οι

αμυντικές τεχνικές τους δε θα αντιμετωπίσουν πάντα "ευγενικά" την μη εξουσιοδοτημένη πρόσβαση στην περιοχή τους. Παρ' όλο που συμπεριλήφθησαν οι συγκεκριμένες λειτουργίες, υπάρχει πάντοτε η πιθανότητα να γίνει αντιληπτός ο crawler και να τερματιστεί η αναζήτηση. Πιο ακραίο σενάριο αποτελεί το να περιλαμβάνεται ανάμεσα στις τεχνικές η μετάδοση κακόβουλου λογισμικού προς τον επισκέπτη, προκαλώντας ζημία στο σύστημά του, τιμωρώντας τον με αυτόν τον τρόπο για την είσοδό του. Πρέπει λοιπόν οι ενέργειες από την πλευρά του να γίνονται με προσοχή και λαμβάνοντας πρώτα κάποια απαραίτητα μέτρα ασφαλείας.

4.4 Βοηθητικό λογισμικό

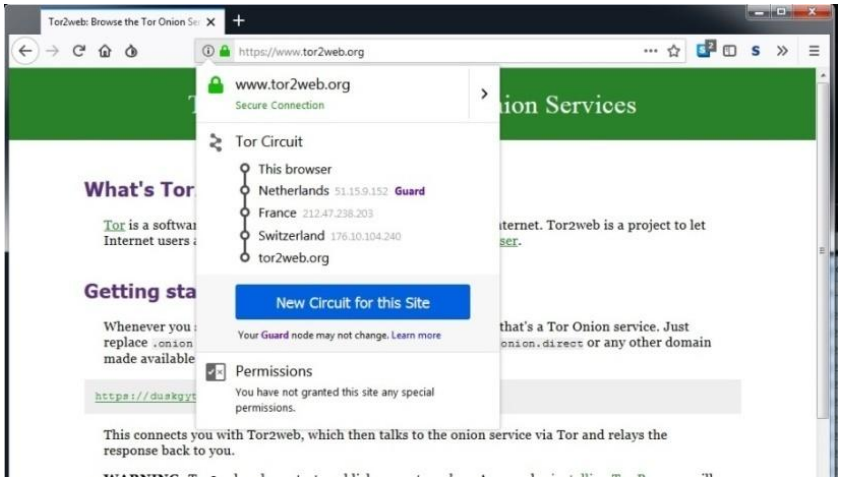
Σε αυτό το σημείο θα γίνει μια αναφορά στο λογισμικό που χρησιμοποιήθηκε παράλληλα με την εκτέλεση της εφαρμογής, καθώς και στους λόγους που κάτι τέτοιο κρίθηκε απαραίτητο.

4.4.1 Tor browser

Η χρήση της εφαρμογής στο δίκτυο Tor απαιτεί την ταυτόχρονη εκτέλεση του Tor browser, ο οποίος είναι διαθέσιμος μέσω της επίσημης ιστοσελίδας του Tor Project και υποστηρίζει τα λειτουργικά συστήματα Windows, Mac και Linux [81]. Πρόκειται για τον onion proxy που αναφέρθηκε στο κεφάλαιο 2.3.2, χρησιμοποιεί ως βάση το πρόγραμμα περιήγησης Mozilla Firefox, προσθέτοντας όλα τα επιπλέον χαρακτηριστικά και τις ρυθμίσεις ασφαλείας για να είναι δυνατή η χρήση του δικτύου, προφυλάσσοντας την ιδιωτικότητα και την ανωνυμία. Έχουν γίνει δοκιμές για συνεργασία και με άλλους περιηγητές, όπως ο Google Chrome, αλλά δυστυχώς οι προσπάθειες των δημιουργών της εφαρμογής δεν στέφθηκαν με επιτυχία [82].

Η παράλληλη χρήση του browser με την εφαρμογή είναι απαραίτητη προκειμένου να ανοιχθεί και να διατηρηθεί η σύνδεση στο δίκτυο Tor. Η συγκεκριμένη υλοποίηση επιλέχθηκε διότι αποτελεί τον πιο ασφαλή τρόπο σύνδεσης σε σχέση με άλλους, που απαιτούν εξειδικευμένη παραμετροποίηση. Ενδεχόμενα λάθη ή παραλήψεις σε αυτήν μπορούν να δημιουργήσουν κενά ασφαλείας, αφήνοντας εκτεθειμένους τόσο τους χρήστες, όσο και τους κόμβους. Αυτός είναι και ο λόγος που προτείνεται η αποκλειστική χρήση του από τους κατασκευαστές, οι οποίοι εφιστούν την προσοχή των χρηστών σχετικά με τις εφαρμογές που χρησιμοποιούν μέσα στο δίκτυο.

Κατά την έναρξή της, η εφαρμογή συλλέγει τη λίστα των ενεργών κόμβων που απαρτίζουν τη συγκεκριμένη χρονική στιγμή το δίκτυο και θα χρησιμοποιηθούν στην κατασκευή κυκλωμάτων. Αφού ολοκληρωθεί η σύνδεση, ο χρήστης μπορεί να περιηγηθεί είτε στον επιφανειακό ιστό, είτε στις κρυμμένες υπηρεσίες, όπως θα έκανε με ένα οποιοδήποτε πρόγραμμα περιήγησης. Υπάρχει η δυνατότητα προβολής των λεπτομερειών του κυκλώματος που έχει σχηματιστεί προς την επιθυμητή ιστοσελίδα, όπως και η αλλαγή κυκλώματος όποτε αυτό κριθεί απαραίτητο.



Εικόνα 4-1: Προβολή στοιχείων κυκλώματος

4.4.2 VPN

Με στόχο την πρόσθετη ασφάλεια, υπάρχει η δυνατότητα συνδυαστικής χρήσης του περιηγητή Tor browser με λογισμικό VPN [83]. Οι τρόποι με τους οποίους μπορεί να γίνει αυτό είναι δύο, οι οποίοι διαφοροποιούνται ανάλογα με το ποιο από τα δύο θα προηγηθεί. Έτσι, θα είναι είτε VPN -> Tor -> διαδίκτυο (Tor-over-VPN), είτε Tor -> VPN -> διαδίκτυο (VPN-over-Tor) και σε κάθε έναν από αυτούς μπορούν να εντοπιστούν κάποια θετικά και αρνητικά στοιχεία.

Στην πρώτη περίπτωση ο τηλεπικοινωνιακός πάροχος (ISP) γνωρίζει για την χρήση του VPN, όχι όμως για την είσοδο στο Tor. Παράλληλα ο κόμβος εισόδου δε γνωρίζει την πραγματική διεύθυνση του χρήστη και υπάρχει κανονικά πρόσβαση στις κρυμμένες υπηρεσίες. Μπορεί αυτή η επιλογή να προσφέρει μεγαλύτερη ασφάλεια, αλλά ο πάροχος VPN γνωρίζει τη χρήση του Tor και ειδικά αν διατηρεί αρχεία καταγραφής, είναι εφικτός ο εντοπισμός και η ταυτοποίηση του χρήστη. Επιπλέον, σε περίπτωση που

για οποιοδήποτε λόγο σταματήσει η λειτουργία της εφαρμογής VPN, αποκαλύπτεται η δραστηριότητα που εκτελείται.

Στη δεύτερη περίπτωση, η χρήση του Tor είναι εμφανής στον ISP, αλλά ο πάροχος VPN δε γνωρίζει την πραγματική διεύθυνση του χρήστη, ούτε έχει πρόσβαση στα δεδομένα που διακινούνται. Είναι δυνατή η επιλογή της τοποθεσίας που θα εμφανίζεται για τον χρήστη, μέσω της επιλογής συγκεκριμένου εξυπηρετητή VPN, ενώ είναι επίσης δυνατή η σύνδεση σε ιστοσελίδες που απαγορεύουν την είσοδο σε συνδέσεις προερχόμενες από το Tor. Ενώ αποτελεί καλύτερη λύση όσον αφορά την ανωνυμία, αρνητικό αποτελεί το γεγονός ότι δεν είναι δυνατή η πρόσβαση στις κρυμμένες υπηρεσίες, καθώς και το ότι ο πάροχος VPN μπορεί να δημιουργήσει προφίλ με βάση τις συνήθειες του χρήστη, αποκαλύπτοντας έτσι την ταυτότητά του. Τέλος, επηρεάζεται ο παράγοντας της απόδοσης, αφού αυτή είναι μειωμένη σε σχέση με την προηγούμενη περίπτωση [84].

Στα πλαίσια των δοκιμών του web crawler επιλέχθηκε η παράλληλη χρήση της υπηρεσίας NordVPN [85], εφαρμόζοντας την πρώτη τακτική για λόγους πρόσθετης ασφάλειας. Οι κατασκευαστές της υποστηρίζουν πως δε διατηρούνται αρχεία καταγραφής, διαθέτει μεγάλο αριθμό εξυπηρετητών παγκοσμίως, επιπλέον λειτουργίες ανάλογα με τις ειδικές ανάγκες κάθε σύνδεσης και γενικά η απόδοσή της σε ταχύτητα είναι ικανοποιητική.

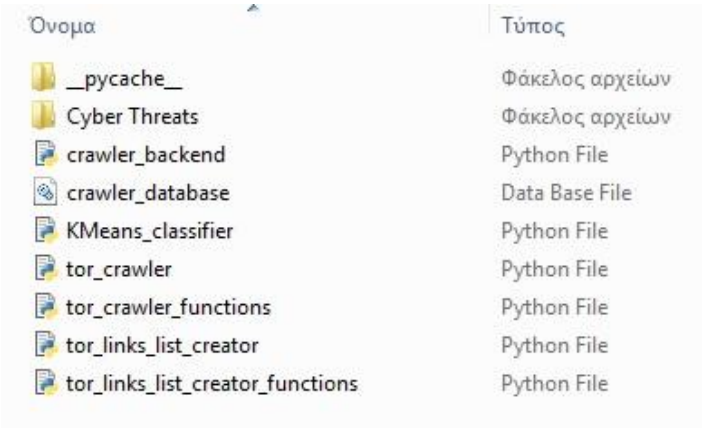
5. Επίδειξη Λειτουργίας

5.1 Εισαγωγή

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, η αναζήτηση ιστοσελίδων με χρήση μηχανών αναζήτησης παρουσιάζει περιορισμούς ως προς τα αποτελέσματα. Αυτός είναι ο λόγος που χρησιμοποιήθηκε η μηχανή αναζήτησης Ahmia [86] που λειτουργεί στο δίκτυο Tor, καθώς και η μηχανή αναζήτησης της Google, για την εύρεση ιστοτόπων που περιέχουν λίστες συνδέσμων κρυμμένων υπηρεσιών. Η εφαρμογή που κατασκευάστηκε για αυτήν την εργασία, χρησιμοποιεί ως πηγή τους παραπάνω ιστότοπους, των οποίων αντλεί το περιεχόμενο προκειμένου να δημιουργήσει μια συνολική λίστα, αποτελούμενη από διευθύνσεις κρυμμένων υπηρεσιών. Συνολικά χρησιμοποιήθηκαν εννέα ιστοσελίδες, οι οποίες ανήκουν στον επιφανειακό και στον σκοτεινό ιστό και αποτελούν το αποτέλεσμα της πρωτοβουλίας κάποιων ομάδων χρηστών, που προσπαθούν να διευκολύνουν και να καταστήσουν πιο εύκολη την επίσκεψη του κοινού στις κρυμμένες υπηρεσίες.

Τα αρχεία της Εικόνας 5-1 αποτελούν την εφαρμογή και είναι τα tor_links_list_creator, tor_links_list_functions, tor_crawler, tor_crawler_functions, crawler_database και KMeans_classifier. Όσο για τις λειτουργίες που επιτελούν, το tor_links_list_creator περιέχει τη διεπαφή χρήστη που αφορά τη δημιουργία της λίστας, ενώ το tor_links_list_functions τις απαραίτητες λειτουργίες για την εκτέλεση των εργασιών. Ομοίως, το αρχείο tor_crawler περιέχει τη διεπαφή για τη συλλογή δεδομένων και το tor_crawler_functions όλες τις απαραίτητες λειτουργίες. Τέλος, το crawler_database περιέχει τις λειτουργίες που αφορούν τη δημιουργία και διαχείριση της βάσης δεδομένων και το KMeans_classifier πραγματοποιεί την κατηγοριοποίηση του κειμένου, με βάση το περιεχόμενό του. Στον φάκελο Cyber Treats υπάρχουν τα αρχεία text που πραγματοποιούν μια μορφή εκπαίδευσης του αλγορίθμου κατηγοριοποίησης.

Το περιβάλλον στο οποίο εκτελούνται τα αρχεία διεπαφών είναι το Python Shell, το οποίο έχει τη μορφή γραμμής εντολών (command line) και στο οποίο ο χρήστης έχει τη δυνατότητα, πέρα από την αλληλεπίδραση με την εφαρμογή, να παρακολουθεί την πορεία εκτέλεσης κάθε λειτουργίας.

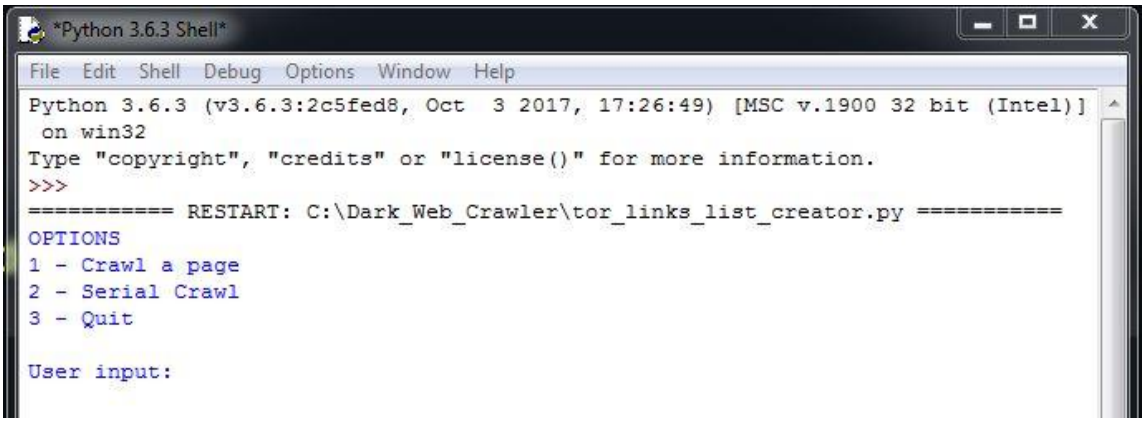


Εικόνα 5-1: Αρχεία εφαρμογής

5.2 Δημιουργία λίστας διευθύνσεων

5.2.1 Μεμονωμένη συλλογή συνδέσμων

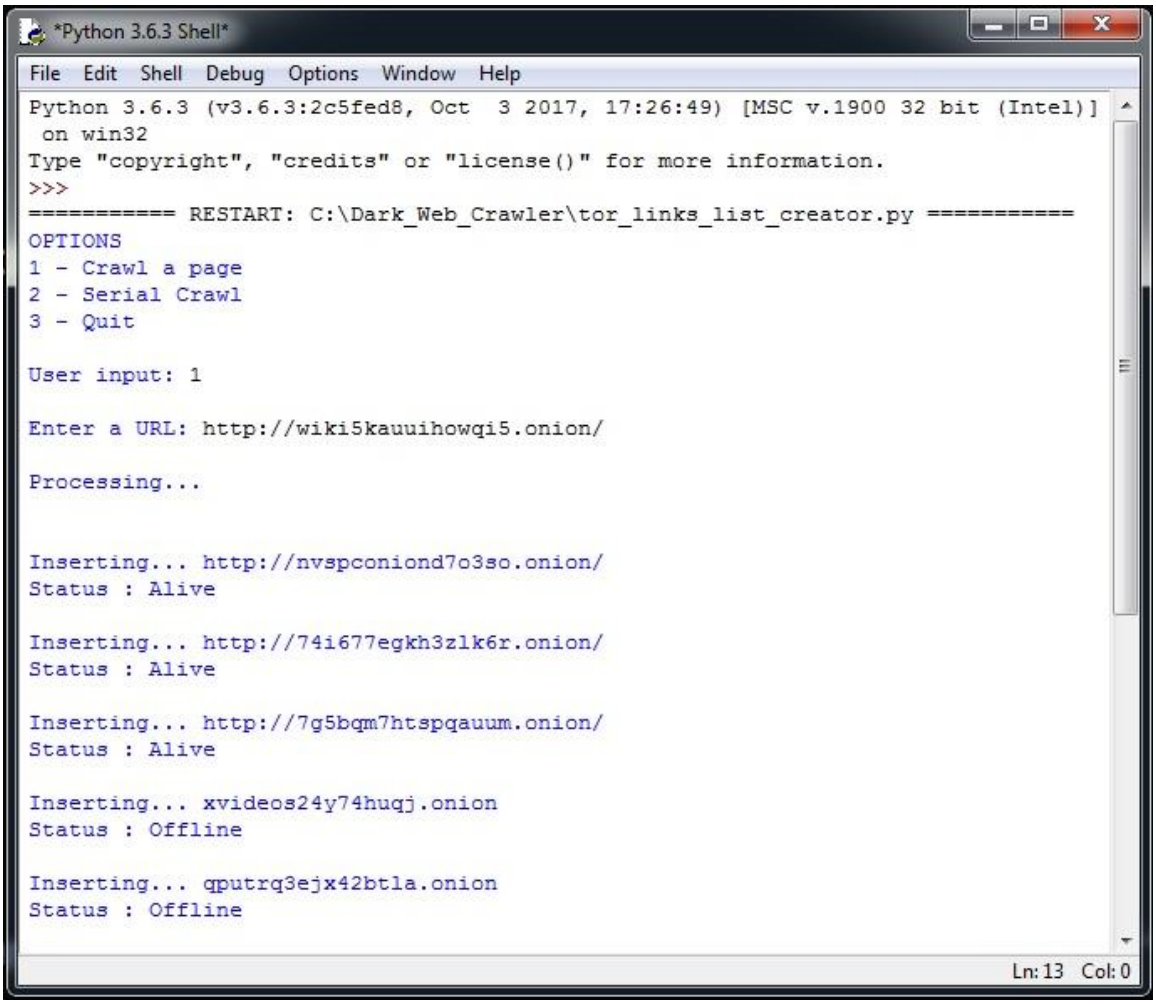
Η εφαρμογή ξεκινάει με την εκτέλεση του αρχείου tor_links_list_creator και τη στιγμή που θα εμφανιστούν οι διαθέσιμες επιλογές στο Python Shell, δημιουργείται το αρχείο της βάσης δεδομένων με ονομασία crawler_database. Μέσα από το μενού ο χρήστης μπορεί είτε να συλλέξει δεδομένα, είτε να τερματίσει την εφαρμογή. Όσον αφορά τη συλλογή, υπάρχουν διαθέσιμες δύο διαφορετικές λειτουργίες, η μία είναι η συλλογή συνδέσμων από μια μεμονωμένη σελίδα και η δεύτερη είναι η σειριακή συλλογή συνδέσμων.



Εικόνα 5-2: Αρχικές επιλογές εφαρμογής

Στη μεμονωμένη συλλογή συνδέσμων, ο χρήστης πληκτρολογεί τη διεύθυνση url της επιθυμητής τοποθεσίας και ακολουθεί η συλλογή όλων των συνδέσμων που οδηγούν

σε κρυμμένες υπηρεσίες και έχουν την κατάληξη .onion. Κατά τη διάρκεια της διαδικασίας εμφανίζονται στην οθόνη τα urls και η κατάστασή τους. Αν είναι ενεργά καταχωρούνται ως "Alive", ενώ στην αντίθετη περίπτωση ως "Offline". Παράλληλα γίνεται ο απαιτούμενος έλεγχος για την αποφυγή δημιουργίας διπλών εγγραφών. Στην περίπτωση που το url υπάρχει ήδη στη βάση δεδομένων αυτό παραβλέπεται, εμφανίζεται το ανάλογο μήνυμα ενημέρωσης και η ροή περνάει στον επόμενο σύνδεσμο.

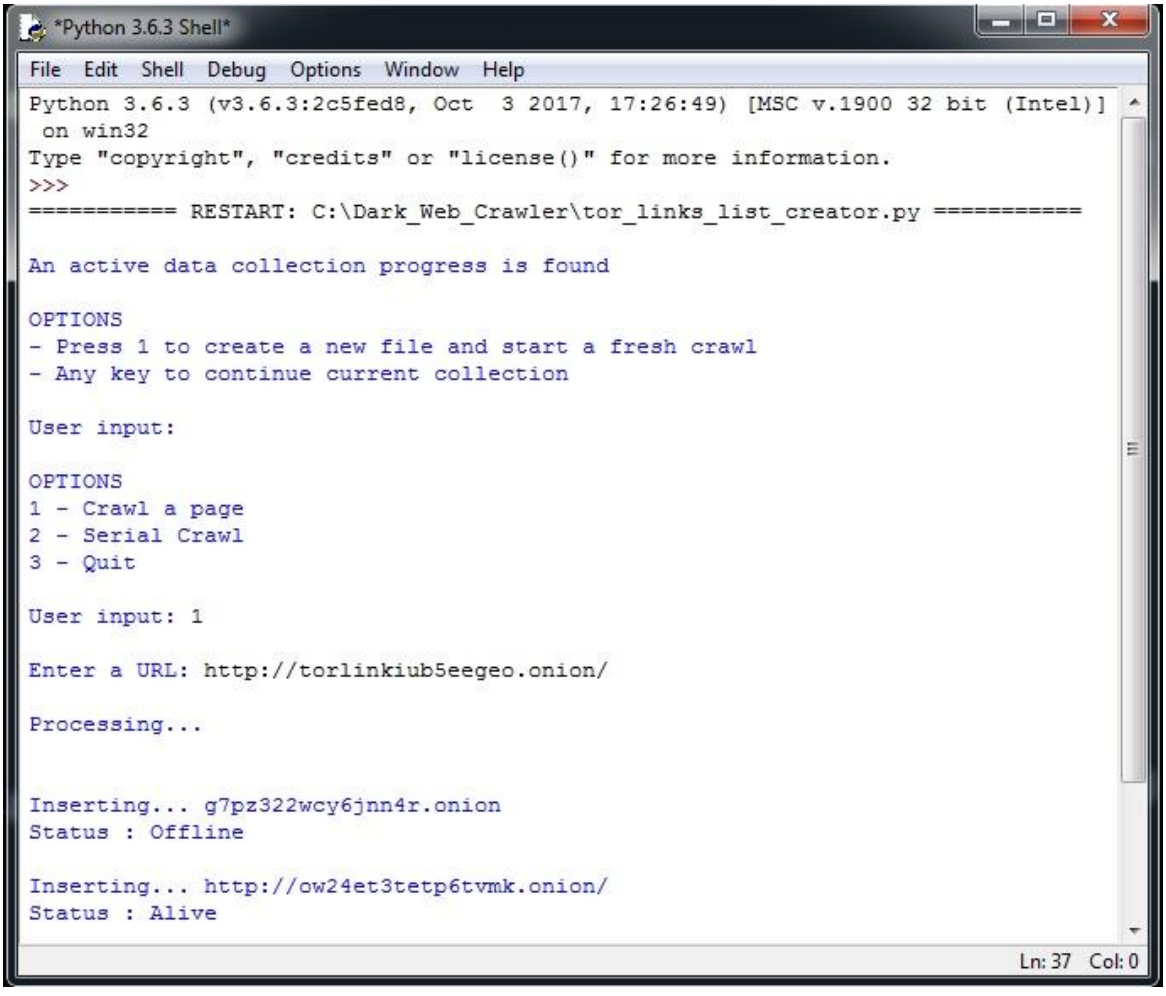


Εικόνα 5-3: Εκτέλεση συλλογής συνδέσμων

Ολοκληρώνοντας την εύρεση και καταχώρηση όλων των συνδέσμων της τοποθεσίας που ζητήθηκε, δίνεται η δυνατότητα συνέχισης της συλλογής πληκτρολογώντας μια νέα διεύθυνση, έχοντας πάλι τις επιλογές των δύο λειτουργιών, καθώς και η επιλογή τερματισμού της εφαρμογής.

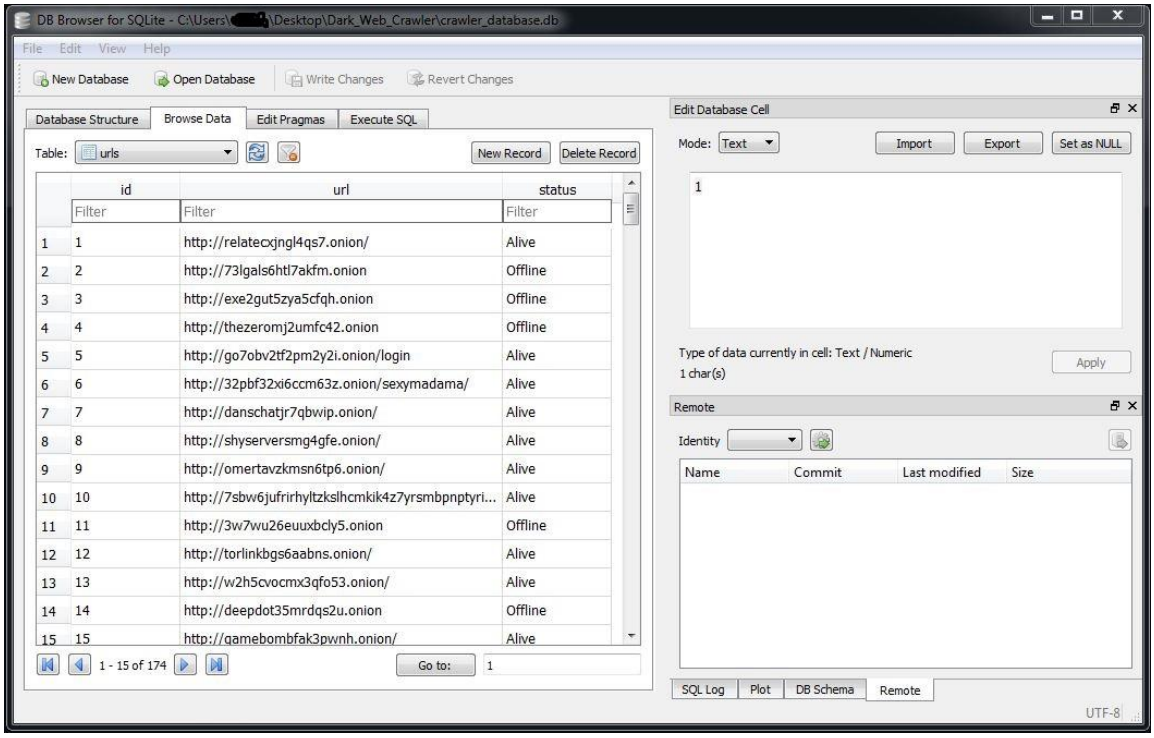
Εικόνα 5-4: Επιλογές κατόπιν ολοκλήρωσης συλλογής συνδέσμων

Με κάθε νέα εκκίνηση της εφαρμογής, εκτελείται ο έλεγχος ύπαρξης βάσης δεδομένων και στην περίπτωση που αυτό ισχύει, επιστρέφεται το ανάλογο μήνυμα με δύο επιλογές. Στην πρώτη, πιέζοντας ο χρήστης το πλήκτρο "1" πραγματοποιείται διαγραφή του παλαιού και δημιουργία νέου αρχείου βάσης, ενώ με οποιοδήποτε άλλο πλήκτρο η διαδικασία συνεχίζεται στο ίδιο αρχείο, όπως φαίνεται στην Εικόνα 5-5.



Εικόνα 5-5: Συνέχεια εκτέλεσης συλλογής στην ίδια βάση δεδομένων

Ο πίνακας urls της βάσης δεδομένων που προκύπτει από την παραπάνω διαδικασία, αποτελείται από τρεις στήλες και κάθε εγγραφή διαθέτει ένα μοναδικό χαρακτηριστικό id, τη διεύθυνση url και την κατάσταση αυτής. Η προβολή του πίνακα μπορεί να γίνει μέσω της εφαρμογής "DB Browser for SQLite". Όπως φαίνεται και στην Εικόνα 5-6, παρέχονται κάποια εργαλεία για τη διαχείριση της βάσης, όπως για παράδειγμα η δυνατότητα εκτέλεσης ερωτημάτων SQL και η εξαγωγή των αποτελεσμάτων τους σε αρχείο. Επίσης, είναι εφικτό με πολύ εύκολο τρόπο να γίνουν τροποποιήσεις, είτε στα δεδομένα, πληκτρολογώντας απλά εντός των κελιών και αποθηκεύοντας τις αλλαγές, είτε στις εγγραφές, προσθέτοντας και αφαιρώντας γραμμές με τη χρήση του ποντικιού και των επιλογών που παρέχει η ίδια η εφαρμογή.



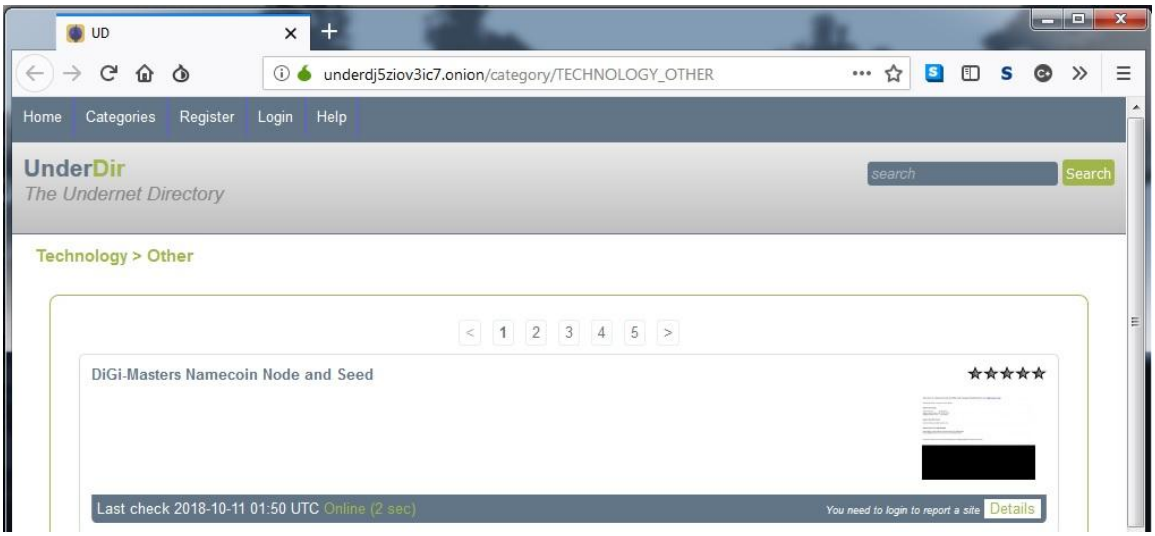
Εικόνα 5-6: Προβολή βάσης δεδομένων

Στην περίπτωση που ο χρήστης επιθυμεί να προχωρήσει στην ανάκτηση συνδέσμων από αλυσίδα σελίδων που περιέχουν το ίδιο θέμα, η αναζήτηση με χρήση μεμονωμένων σελίδων γίνεται απαιτητική σε χρόνο. Για αυτό το λόγο υλοποιήθηκε επιπλέον η λειτουργία του σειριακού τρόπου συλλογής, που αναλύεται στην παράγραφο που ακολουθεί.

5.2.2 Σειριακή συλλογή συνδέσμων

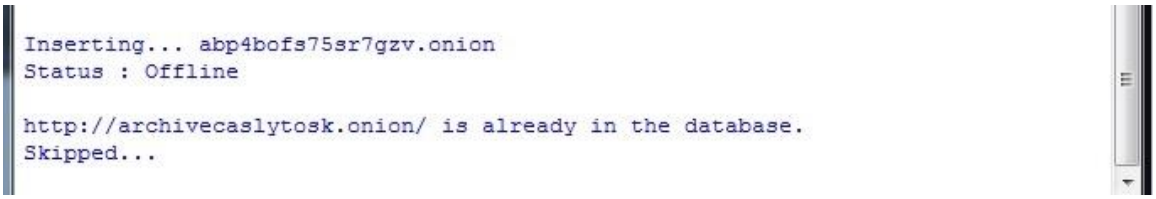
Αποτελεί τη δεύτερη επιλογή, υλοποιεί τη σειριακή συλλογή συνδέσμων και αφορά ομάδες urls που μοιράζονται το ίδιο θέμα, αφού η μεμονωμένη συλλογή δεν αποτελεί πρακτική λύση σε αυτού του είδους τις περιπτώσεις.

Ο χρήστης καλείται να πληκτρολογήσει το url που αποτελεί την πρώτη σελίδα της ομάδας, τη μορφή που θα έχουν οι σύνδεσμοι των επόμενων σελίδων, καθώς και τον συνολικό αριθμό τους. Η μορφή είναι απαραίτητη, διότι η ποικιλομορφία στον τρόπο που ορίζονται τα urls των σελίδων που αποτελούν μέρος της αλυσίδας είναι μεγάλη και εξαρτάται από τον δημιουργό τους. Αυτό έχει ως αποτέλεσμα να μην είναι εφικτή μια γενίκευση που θα μπορεί να χρησιμοποιηθεί σε όλες τις ιστοσελίδες. Από τη στιγμή που η συγκεκριμένη λειτουργία αποτελεί εργαλείο ειδικής χρήσης, υπάρχει ενεργή συμμετοχή του χρήστη με έλεγχο της μορφής και του συνολικού αριθμού σελίδων που απαρτίζουν την ομάδα.



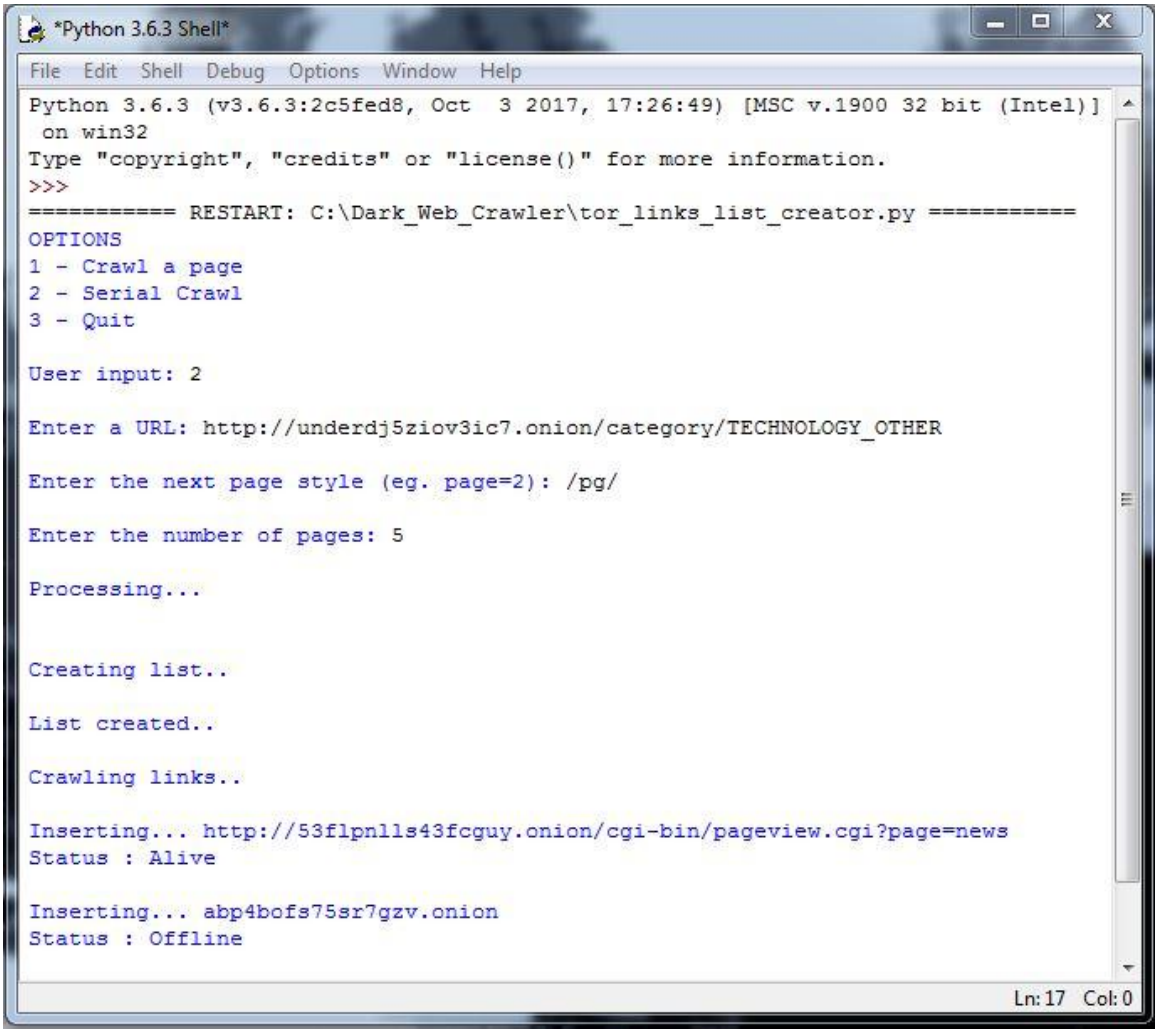
Εικόνα 5-7: Παράδειγμα ομάδας σελίδων

Με τα στοιχεία που εισάγει ο χρήστης σχηματίζεται μια λίστα με τις σελίδες της ομάδας. Στο παράδειγμα της Εικόνας 5-7, η ομάδα αποτελείται από τις σελίδες 1 έως 5. Για κάθε μία από αυτές, πραγματοποιείται η ίδια διαδικασία με εκείνη της μεμονωμένης συλλογής συνδέσμων, με τη λειτουργία να ακολουθεί την ίδια λογική. Κατά τη διάρκεια της εκτέλεσης της εφαρμογής εμφανίζονται στην οθόνη τα urls με την κατάστασή τους και παράλληλα αποφεύγονται οι διπλές εγγραφές κατά την αποθήκευσή τους στη βάση δεδομένων.



Εικόνα 5-8: Εύρεση και παράκαμψη συνδέσμου

Με την ανάκτηση των περιεχομένων και της τελευταίας σελίδας, ολοκληρώνεται η διαδικασία και πλέον ο χρήστης μπορεί να επιλέξει είτε μια νέα σελίδα, την οποία θα συλλέξει μεμονωμένα, είτε μια νέα αλυσίδα σελίδων, είτε να τερματίσει την εφαρμογή. Οι επιλογές σε περίπτωση επανέναρξης είναι οι ίδιες που περιγράφηκαν και προηγουμένως, δηλαδή ο χρήστης μπορεί να ξεκινήσει τη δημιουργία μιας νέας βάσης δεδομένων ή να συνεχίσει τη συλλογή στην ίδια βάση.



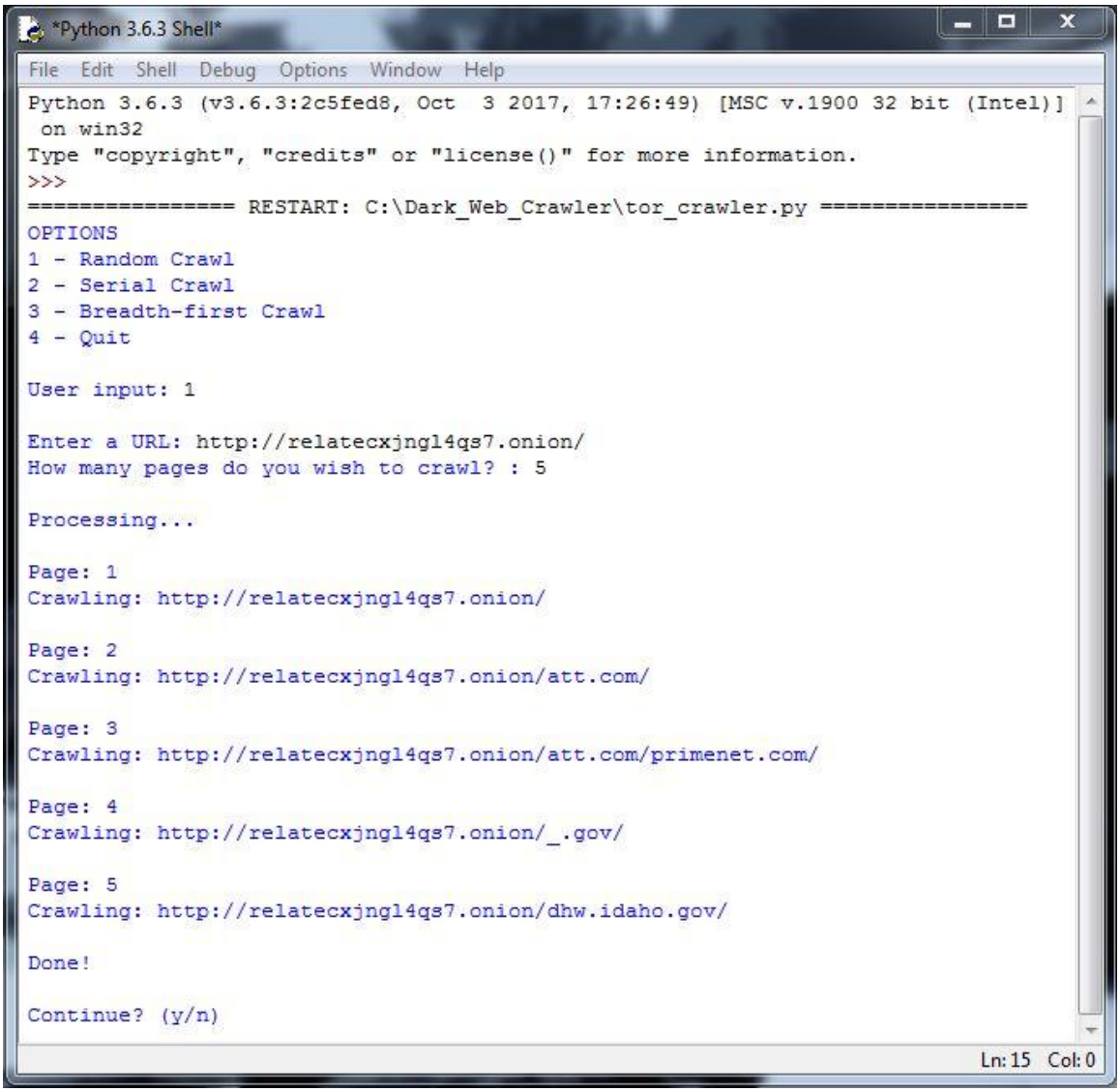
Εικόνα 5-9: Σειριακή συλλογή συνδέσμων

Σε αυτό το σημείο μπορεί εύκολα να γίνει το φιλτράρισμα και ο διαχωρισμός των ενεργών ιστοτόπων, προκειμένου να χρησιμοποιηθούν στη συνέχεια, στο δεύτερο μέρος της εφαρμογής, που είναι η συλλογή δεδομένων.

5.3 Συλλογή δεδομένων

5.3.1 Τυχαία συλλογή δεδομένων

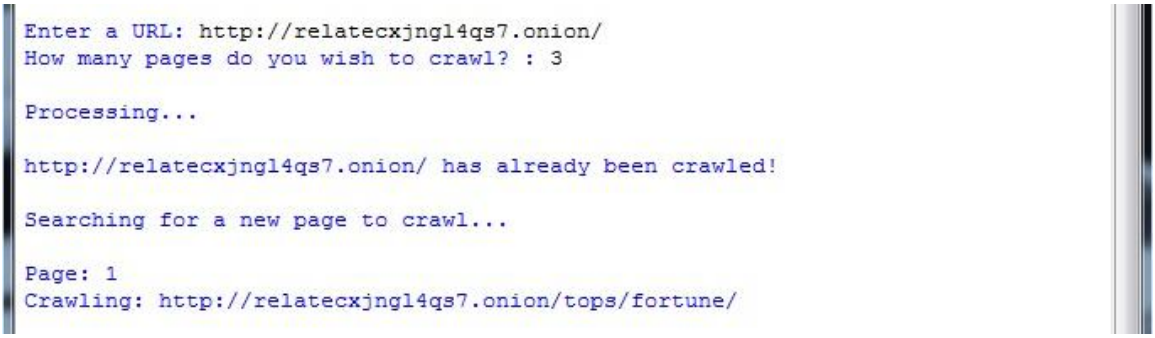
Με την εκτέλεση του αρχείου tor_crawler, ο χρήστης αποκτάει πρόσβαση σε επιλογές σχετικά με τον τρόπο που επιθυμεί να γίνει η αναζήτηση περιεχομένου στις ιστοσελίδες.



Εικόνα 5-10: Τυχαία συλλογή δεδομένων

Κάθε μία από αυτές εκτελεί μια διαφορετική λογική αναζήτησης, με την πρώτη να είναι ο τυχαίος τρόπος, όπου έχοντας ως αρχή μια συγκεκριμένη διεύθυνση url, η εφαρμογή κινείται μέσα στον ιστοτόπο με μη προκαθορισμένο τρόπο, συλλέγοντας δεδομένα. Ο αριθμός των σελίδων προς ανάκτηση ορίζεται από τον χρήστη και όταν αυτός συμπληρωθεί, δίνεται η επιλογή της συνέχισης πληκτρολογώντας νέο αριθμό σελίδων ή αλλιώς του τερματισμού της διαδικασίας.

Σε περίπτωση που μια σελίδα έχει ήδη καταχωρηθεί στη βάση, η αναζήτηση αυτόματα μεταπηδά σε άλλη σελίδα, επίσης με τυχαίο τρόπο.



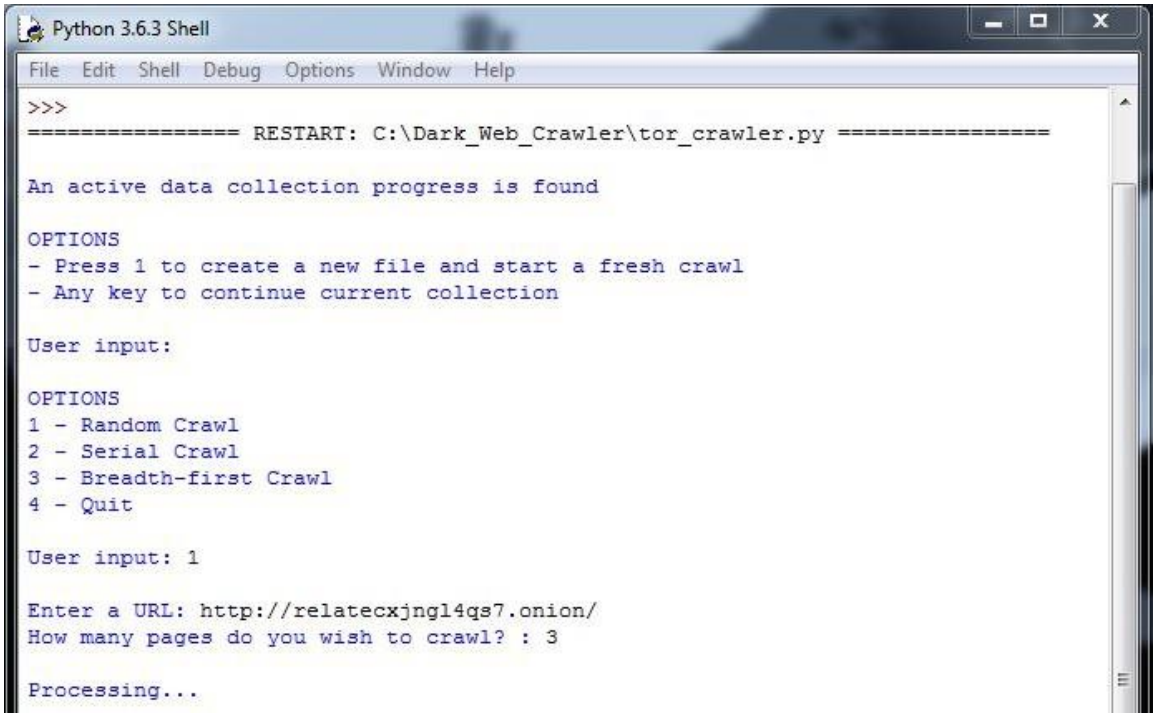
Εικόνα 5-11: Εύρεση νέας σελίδας κατόπιν ελέγχου της βάσης δεδομένων

Παράλληλα με την αποθήκευση του περιεχομένου των σελίδων, η εφαρμογή πραγματοποιεί έλεγχο για την ύπαρξη νέων συνδέσμων μέσα σε αυτό και σε περίπτωση που βρεθούν, ακολουθεί η προσθήκη τους στη λίστα που διατηρείται στον πρώτο πίνακα της βάσης. Ο τρόπος αποθήκευσης των συνδέσμων είναι ίδιος με εκείνον που περιγράφηκε για τη δημιουργία της λίστας, με καταχώρηση του url και της κατάστασης στην οποία αυτό βρίσκεται.

Θα μπορούσε να θεωρηθεί εν μέρη ως μια αναζήτηση κατά βάθος, όμως λόγω της δομής των ιστοτόπων, όπου περιέχονται πολλαπλοί σύνδεσμοι μεταξύ των σελίδων, αλλά και λόγω της τυχαιότητας που εφαρμόζεται στον τρόπο μετάβασης για λόγους ασφαλείας, κάτι τέτοιο δεν ισχύει. Η απουσία δημιουργίας προτύπων στην κίνηση του crawler αποτελεί ένα από τα βασικά συστατικά της λογικής του προκειμένου να αποφεύγει τον εντοπισμό.

Οι επιλογές κατά την επανέναρξη της εφαρμογής είναι οι ίδιες και σε αυτήν την περίπτωση, όσον αφορά το αρχείο της βάσης δεδομένων. Για τη συλλογή δεδομένων γίνεται χρήση του δεύτερου πίνακα της βάσης, ο οποίος έχει την ονομασία page_content και ο οποίος αποτελείται από πέντε στήλες, καθώς πέραν του μοναδικού αναγνωριστικού

id κάθε εγγραφής, περιέχει τη διεύθυνση url, την κατηγορία στην οποία ανήκει η κυβερνο-επίθεση, τον κώδικα HTML και το κείμενο που εξάγεται από αυτόν.

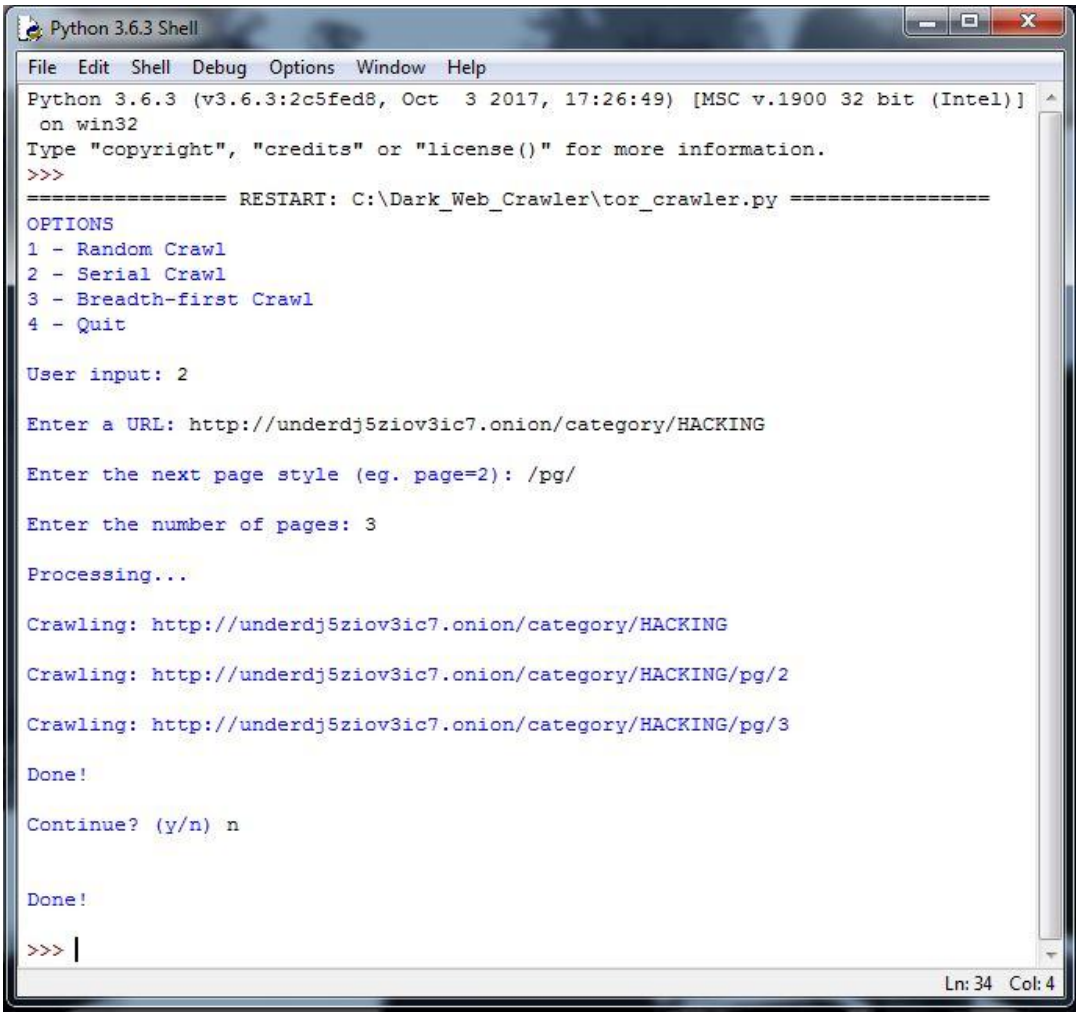


Εικόνα 5-12: Συνέχεια τυχαίας συλλογής δεδομένων

5.3.2 Σειριακή συλλογή δεδομένων

Στη σειριακή αναζήτηση, στόχο αποτελούν οι σελίδες που είναι μέρος μιας αλυσίδας και σχετίζονται με το ίδιο θέμα, όμοια με την περίπτωση της σειριακής συλλογής συνδέσμων για τη δημιουργία της αρχικής λίστας. Η συλλογή συνεχόμενων σελίδων που αποτελούν μια ομάδα έχει νόημα, ειδικά όταν πρόκειται για δεδομένα που περιέχονται σε ομάδες συζητήσεων. Είναι απαραίτητη η συλλογή όλων των μηνυμάτων συζήτησης γύρω από ένα θέμα, καθώς παρουσιάζουν μια συνέχεια και η απουσία κάποιων εξ αυτών μπορεί να επηρεάσει τη χρησιμότητα των πληροφοριών.

Τα δεδομένα που πρέπει να καταχωρήσει ο χρήστης και σε αυτήν την περίπτωση είναι η αρχική διεύθυνση, η μορφή που έχει το url των επόμενων σελίδων και ο συνολικός αριθμός τους. Η διαδικασία συλλογής ολοκληρώνεται μόλις ανακτηθεί το περιεχόμενο του επιθυμητού αριθμού σελίδων, εκτός και αν ο χρήστης επιθυμεί να συνεχίσει με νέα διεύθυνση url.

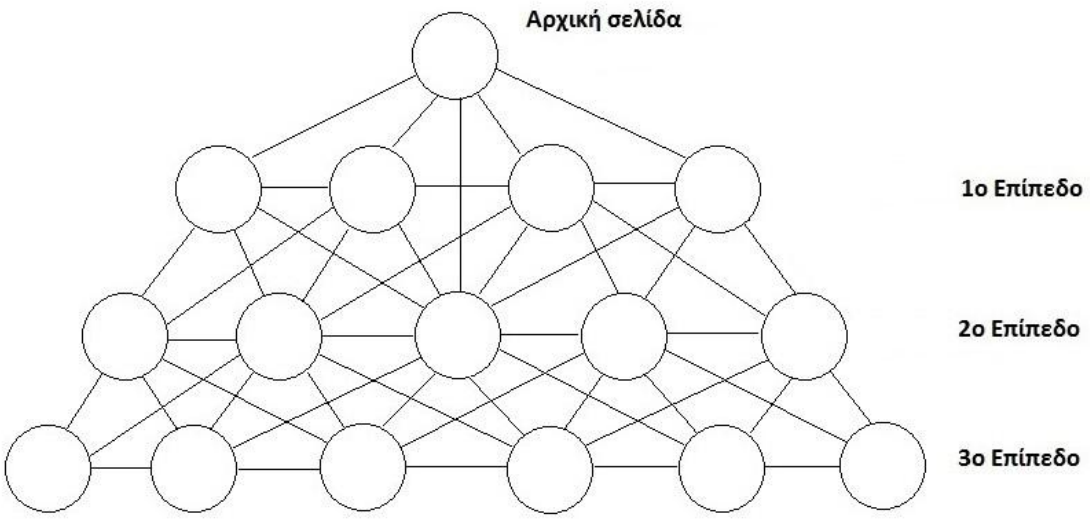


Εικόνα 5-13: Σειριακή συλλογή δεδομένων

Όπως στην τυχαία συλλογή, έτσι και στη σειριακή, παράλληλα με την αποθήκευση του περιεχομένου των σελίδων, η εφαρμογή πραγματοποιεί έλεγχο για την ύπαρξη νέων συνδέσμων και αποθήκευσή τους στον πρώτο πίνακα της βάσης.

5.3.3 Συλλογή δεδομένων κατά πλάτος

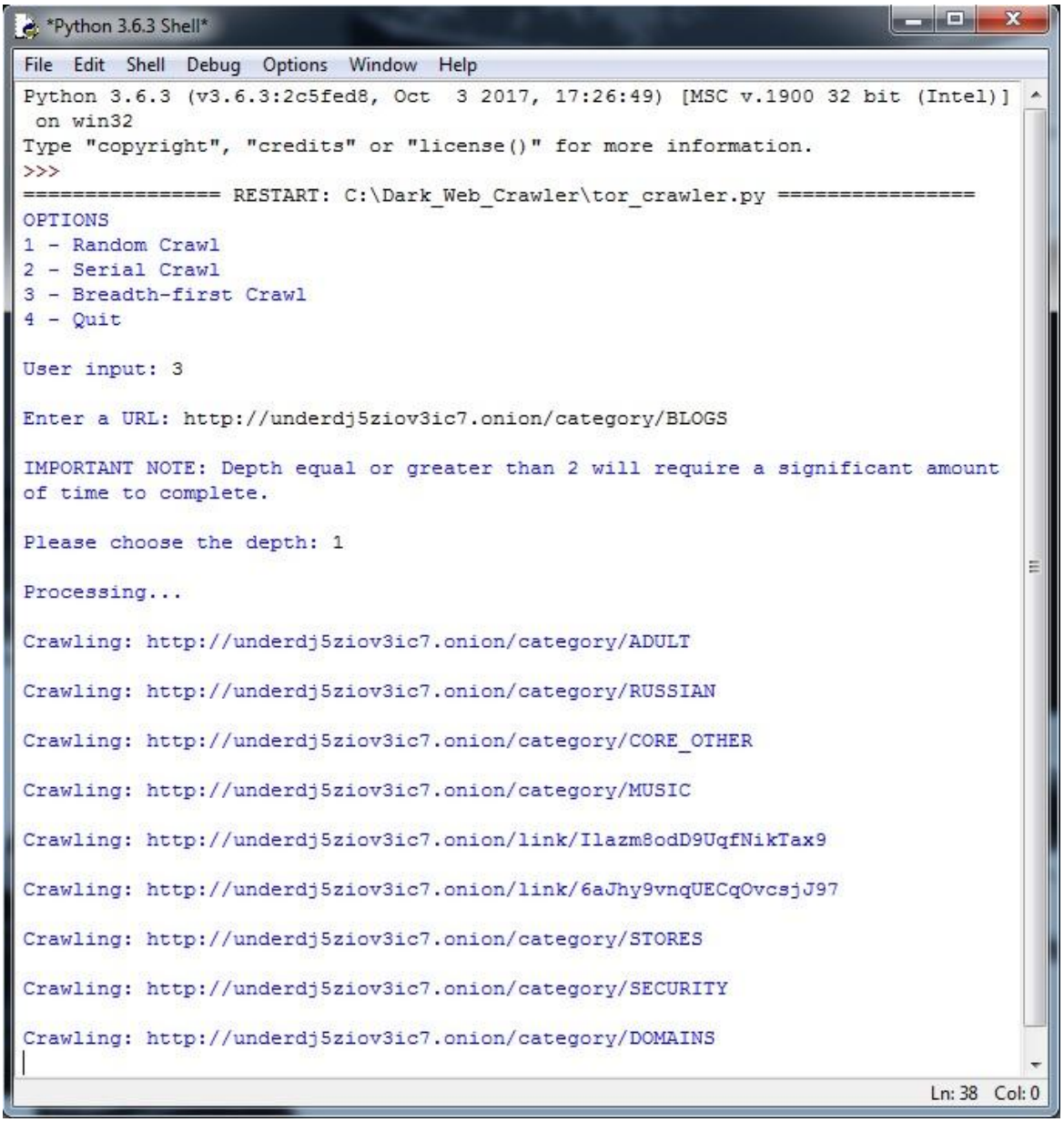
Σε αυτήν την περίπτωση συλλογής δεδομένων, πέραν της αρχικής σελίδας που θα αποτελέσει τη βάση της αναζήτησης σελίδων, ο χρήστης ορίζει το βάθος στο οποίο θα φτάσει αυτή. Αν φανταστεί κανείς τη δομή ενός ιστότοπου ως ένα δέντρο που περιέχει κόμβους που είναι συνδεδεμένοι μεταξύ τους, ως βάθος ορίζονται τα επίπεδα μέσα σε αυτό. Η ρίζα είναι η αρχική σελίδα και αποτελεί το επίπεδο 0. Όσοι σύνδεσμοι περιέχονται σε αυτή τη σελίδα οδηγούν στο πρώτο επίπεδο, οι σύνδεσμοι που περιέχονται στις σελίδες αυτού του επιπέδου οδηγούν στο δεύτερο επίπεδο, κοκ.



Εικόνα 5-14: Επίπεδα ιστότοπου

Όσο για τον τρόπο λειτουργίας, από την αρχική σελίδα συλλέγονται όλοι οι σύνδεσμοι που περιλαμβάνονται σε αυτή, μαζί με το HTML περιεχόμενό τους. Για κάθε έναν από αυτούς, πραγματοποιείται η συλλογή όλων των συνδέσμων για τη δημιουργία μιας συνολικής λίστας, με την παράλληλη αφαίρεση τυχόν διπλότυπων που προκύπτουν από συνδέσμους που οδηγούν πίσω στο προηγούμενο επίπεδο. Έπειτα, αποθηκεύεται στη βάση το περιεχόμενο της κάθε σελίδας που προκύπτει από αυτούς. Να σημειωθεί πως σε περίπτωση που ως βάθος επιλεγεί το μηδέν, η εφαρμογή δε θα προχωρήσει σε κάποιο επίπεδο και θα γίνει η συλλογή μόνον της ρίζας.

Όπως είναι φυσιολογικό, όσο πιο μεγάλο το επιλεγμένο βάθος, τόσο πιο μεγάλη θα είναι η χρονική διάρκεια αναζήτησης. Αυτός είναι ένας από τους λόγους που γενικά, σε τέτοιου είδους εργασίες, προτείνεται η διατήρηση του βάθους σε χαμηλά επίπεδα. Ένας επιπλέον λόγος είναι η αποφυγή δημιουργίας αυξημένης κίνησης στον ιστότοπο που αποτελεί στόχο επεξεργασίας, πρώτα λόγω σεβασμού απέναντι σε αυτόν και έπειτα, λόγω της προσπάθειας αποφυγής εντοπισμού όσο διαρκεί η συλλογή δεδομένων.



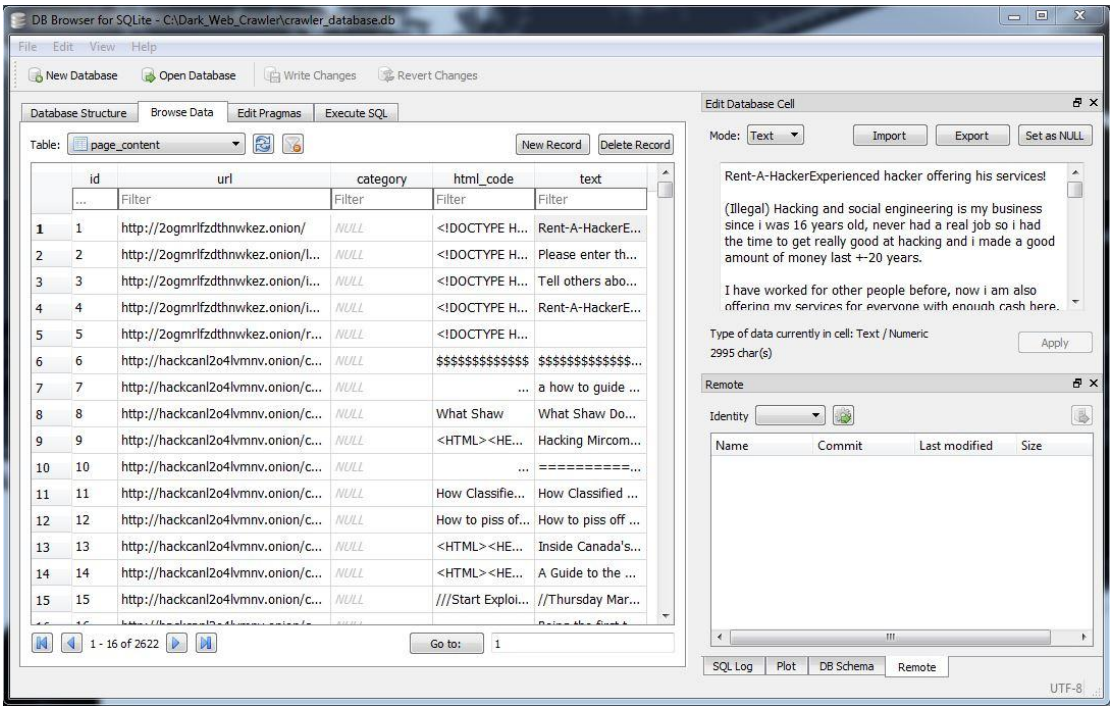
Εικόνα 5-15: Συλλογή δεδομένων κατά πλάτος

5.3.4 Εξαγωγή κειμένου

Σε όλους τους τρόπους ανάκτησης δεδομένων που περιγράφηκαν, παράλληλα πραγματοποιείται η επεξεργασία του περιεχομένου των ιστοσελίδων, ώστε να προκύψουν πληροφορίες που θα χρησιμοποιηθούν στον τομέα του cyber threat intelligence. Πιο συγκεκριμένα, λαμβάνει μέρος η εξαγωγή αυτούσιου του κειμένου που περιέχεται σε αυτές.

Την αποθήκευση του κώδικα HTML στη βάση δεδομένων ακολουθεί η εκτέλεση του τμήματος της εφαρμογής που απομονώνει το κείμενο και το αποθηκεύει στη στήλη

text. Με αυτόν τον τρόπο είναι εφικτή η επεξεργασία για την εξαγωγή χρήσιμων πληροφοριών. Λόγω του ότι κείμενο μικρού μήκους, όπως είναι οι μεμονωμένες λέξεις, δε μπορεί να φανεί ιδιαίτερα χρήσιμο, η εφαρμογή απομονώνει μόνον το κείμενο που έχει μέγεθος μεγαλύτερο των 30 χαρακτήρων. Στις παρακάτω εικόνες φαίνεται ένα παράδειγμα στο περιβάλλον της εφαρμογής "DB Browser for SQLite", με το άνω πλαίσιο στα δεξιά του πίνακα να χρησιμοποιείται για την προβολή του κειμένου.



Εικόνα 5-16: Εξαγωγή κειμένου

Κύριο στόχο αποτελεί η επεξεργασία συζητήσεων και διαλόγων που περιέχονται σε ομάδες συζήτησης, αναρτήσεων που δημοσιεύονται σε ιστολόγια, καθώς και περιγραφών που μπορούν να βρεθούν σε ηλεκτρονικά καταστήματα και αφορούν την πώληση προϊόντων που είναι σχετικά με κυβερνο-επιθέσεις, έναντι κάποιας αμοιβής.

Εικόνα 5-17: Παράδειγμα αποθηκευμένου κειμένου

5.3.5 Κατηγοριοποίηση κειμένου

Το εργαλείο που υλοποιήθηκε για την επεξεργασία των αποθηκευμένων κειμένων, χρησιμοποιεί τον αλγόριθμο μηχανικής μάθησης των K-μέσων (K-Means) για την εφαρμογή ενός φίλτρου, με το οποίο θα πραγματοποιήσει την ταξινόμησή τους.

Έχοντας αποθηκεύσει σε αρχεία text έναν αριθμό από διαδικτυακά κείμενα, που περιέχουν πληροφορίες για τα κυριότερα είδη κυβερνο-απειλών, αυτά τροφοδοτήθηκαν στον αλγόριθμο για τη δημιουργία συστάδων. Κάθε μία από αυτές τις συστάδες αποτελείται από τους σημαντικότερους όρους των κειμένων και αφορά έναν τύπο κυβερνο-επίθεσης. Στη συνέχεια, λαμβάνοντας υπόψη τους 15 πρώτους από αυτούς, η εφαρμογή ελέγχει την ύπαρξή τους στα κείμενα που υπάρχουν μέσα στη βάση δεδομένων και ανάλογα με τους όρους που θα εντοπίσει, αποδίδει μια ετικέτα που καταχωρείται στη στήλη threat_type. Όσα κείμενα δεν είναι δυνατόν να ταξινομηθούν σε κάποια κατηγορία, παίρνουν την ετικέτα undefined. Με αυτόν τον τρόπο είναι δυνατός ο εντοπισμός της κυβερνο-απειλής στην οποία ανήκουν τα κείμενα που συλλέγονται από τον crawler. Στην Εικόνα 5-18 περιέχονται οι συστάδες που προκύπτουν από την εκτέλεση του αλγορίθμου, ενώ στην Εικόνα 5-19 φαίνεται ένα παράδειγμα του αποτελέσματος της διαδικασίας.

Marketplace	SQL Injection	Zero-day	MITM	Carding
market	sql	zero	mitm	carding
markets	injection	day	middle	card
darknet	database	vulnerability	ssl	credit
dream	query	software	man	cards
alphabay	application	security	https	fraud
olympus	input	patch	tls	carder
silk	data	fix	attack	stolen
road	queries	exploit	spoofing	address
pirate	sqli	vulnerabilities	hackers	merchant
bitcoin	web	attacks	arp	payment

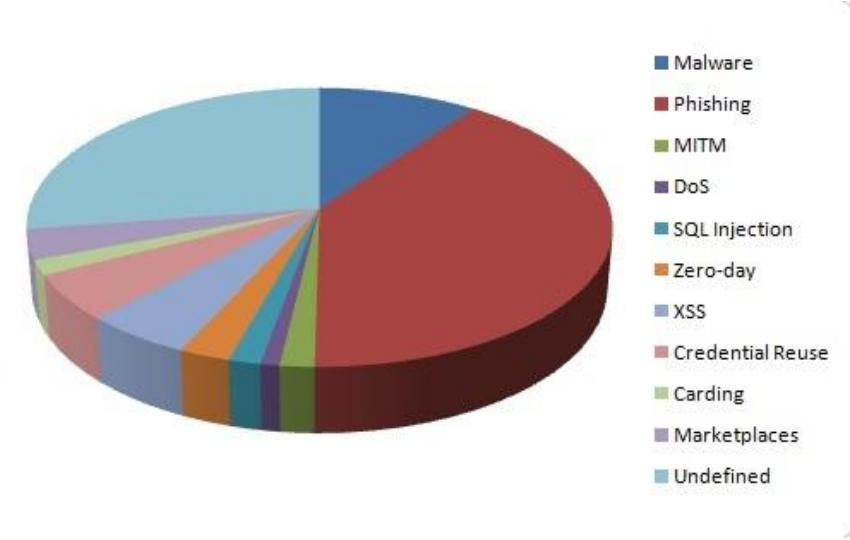
XSS	Malware	Cred. Stuffing	Phishing	DoS
xss	malware	stuffing	phishing	dos
scripting	computer	credential	email	attack
cross	adware	credentials	spear	ddos
user	viruses	password	information	attacks
reflected	worms	passwords	message	denial
site	software	stolen	attack	service
attacker	spyware	accounts	legitimate	traffic
dom	program	login	link	network
data	virus	account	users	server
page	files	data	recipient	flood

Εικόνα 5-18: Συστάδες με λέξεις-κλειδιά

id	url	threat_type	html_code
...	Filter	Filter	Filter
1602	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=8523	Undefined	<!DOCTYPE html P...
1603	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=7601	Zero Day	<!DOCTYPE html P...
1604	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=10346	Man-in-the-middle	<!DOCTYPE html P...
1605	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=10300	Phishing	<!DOCTYPE html P...
1606	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=7931	Undefined	<!DOCTYPE html P...
1607	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=1172	Phishing	<!DOCTYPE html P...
1608	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=7665	Malware	<!DOCTYPE html P...
1609	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=7749	Undefined	<!DOCTYPE html P...
1610	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=1937	Undefined	<!DOCTYPE html P...
1611	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=1578	Zero Day	<!DOCTYPE html P...
1612	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=2138	Malware	<!DOCTYPE html P...
1613	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=7108	Malware	<!DOCTYPE html P...
1614	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=1872	XSS	<!DOCTYPE html P...
1615	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=2226	Undefined	<!DOCTYPE html P...
1616	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=1970	Undefined	<!DOCTYPE html P...
1617	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=2115	Phishing	<!DOCTYPE html P...
1618	http://rrcc5uuudhh4oz3c.onion/?cmd=topic&id=2048	Malware	<!DOCTYPE html P...

Εικόνα 5-19: Κατηγοριοποίηση κειμένου

Με την ολοκλήρωση της κατηγοριοποίησης συνολικά 2.622 αποθηκευμένων εγγραφών, δόθηκε επιτυχώς ετικέτα στις 1.905, ποσοστό 72.65%. Από αυτές, 263 αφορούν malware, 1.054 phishing, 44 man-in-the-middle, 24 denial-of-service, 41 SQL injection, 65 zero-day exploit, 137 cross-site scripting 140 credential reuse, 45 carding και 92 αφορούν ηλεκτρονικά καταστήματα.



Εικόνα 5-20: Κατανομή κυβερνο-επιθέσεων

Στην ανάλυση των παραπάνω στοιχείων φαίνεται πως το μεγαλύτερο ποσοστό στις σελίδες που συλλέχθηκαν αφορά επιθέσεις τύπου phishing, ενώ η αμέσως επόμενη απειλή είναι το κακόβουλο λογισμικό. Το ποσοστό των μη ταξινομημένων εγγραφών είναι αρκετά μεγάλο, δίχως όμως να σημαίνει απαραίτητα αυτό ότι τα συγκεκριμένα κείμενα δε σχετίζονται με κυβερνο-απειλές. Η αποτελεσματικότητα της κατηγοριοποίησης συνδέεται σε μεγάλο βαθμό με τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου. Όσο πιο πολλά είναι αυτά, τόσο μεγαλύτερη η ακρίβεια λειτουργίας του. Αυτός είναι ο λόγος που η καλύτερη λύση, σε τέτοιου είδους εργασίες, είναι η εύρεση και χρήση μεγάλων datasets, προκειμένου να εκπαιδευτεί ένας αλγόριθμος μηχανικής μάθησης με επίβλεψη για την υλοποίηση ενός ταξινομητή.

6. Επίλογος

6.1 Σύνοψη και συμπεράσματα

Γενικό συμπέρασμα της έρευνας αποτελεί το γεγονός ότι οι διαδικασίες ανάκτησης, αποθήκευσης και επεξεργασίας δεδομένων από σελίδες του σκοτεινού διαδικτύου, δεν παρουσιάζουν σημαντικές διαφορές σε σχέση με την εκτέλεσή τους στον επιφανειακό ιστό.

Κατόπιν χρήσης του λογισμικού που δημιουργήθηκε στα πλαίσια της παρούσας εργασίας, προέκυψε μια συλλογή με διευθύνσεις url και ταυτόχρονα πραγματοποιήθηκε έλεγχος για το ποιες από αυτές είναι ενεργές και ποιες όχι. Από το σύνολο των ενεργών, έγινε επιλογή κάποιων εξ αυτών και αποθηκεύτηκε σε βάση δεδομένων SQL το HTML περιεχόμενό τους και το κείμενο που υπήρχε μέσα σε αυτόν, με στόχο την άντληση πληροφοριών που θα φανούν χρήσιμες στην αντιμετώπιση κυβερνο-απειλών.

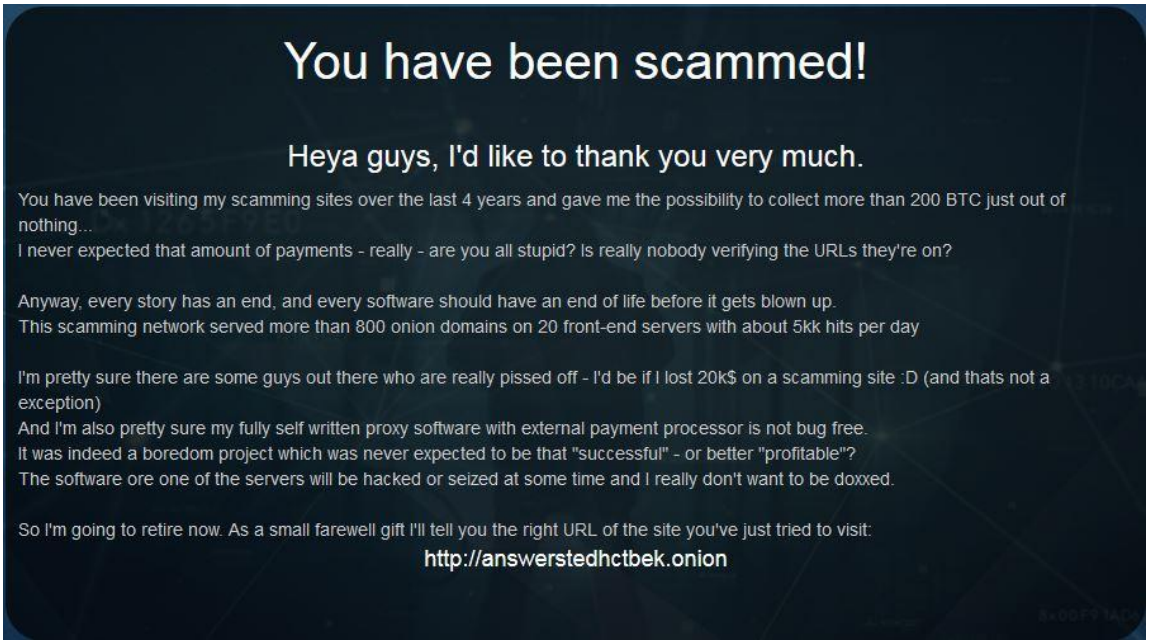
Ένα από τα βασικά συμπεράσματα που προκύπτουν από την έρευνα είναι το πολύ μεγάλο ποσοστό των ανενεργών ιστοτόπων. Από τις 11.964 σελίδες, μόλις οι 2.634 είναι ενεργές, ποσοστό δηλαδή περίπου 22%. Αιτία του φαινομένου αποτελεί ο σύντομος χρόνος ζωής των κρυμμένων υπηρεσιών, είτε γιατί έχουν ολοκληρώσει το σκοπό για τον οποίο δημιουργήθηκαν, είτε γιατί οι αρχές κατάφεραν να τις εντοπίσουν και να τις θέσουν εκτός λειτουργίας. Σημαντικούς παράγοντες, λοιπόν, στην εύρεση και άντληση πληροφοριών από αυτό το κομμάτι του διαδικτύου, αποτελούν η ταχύτητα εντοπισμού των υπηρεσιών και η όσο το δυνατόν αμεσότερη εκμετάλλευση του περιεχομένου τους.

Επιπλέον, διαπιστώθηκε το γεγονός πως πρόκειται για ένα τμήμα του διαδικτύου με πραγματικά ασύνδετη δομή, αφού είναι μικρός ο αριθμός των σελίδων που προσφέρουν συνδέσμους που οδηγούν σε άλλες σελίδες. Οι χρήστες των υπηρεσιών ανακαλύπτουν την ύπαρξη των υπηρεσιών, είτε μέσω ενημέρωσης που είχαν σε διαλόγους εντός ομάδων συζήτησης, είτε από σελίδες όπως αυτές που χρησιμοποιήθηκαν στα πλαίσια της έρευνα, όπου μπορεί κανείς να βρει λίστες από urls, είτε μέσω του στενού τους κύκλου.

Ο σκοτεινός ιστός, έστω και αν αποτελεί ένα σχετικά μικρό υποσύνολο του βαθύ ιστού, διαθέτει έναν μεγάλο αριθμό ιστοτόπων που εξυπηρετούν τα συμφέροντα των κακόβουλων χρηστών και αποτελούν το μέσο οργάνωσής τους, για την επιτυχημένη εκτέλεση παράνομων δραστηριοτήτων. Ο αριθμός των χρηστών που γίνονται δεκτοί

μέσα στις κλειστές ομάδες συζήτησης είναι μικρός, ενώ τυχόν νέα μέλη πρέπει πρώτα να κερδίσουν την εμπιστοσύνη τους, ώστε να συμμετέχουν στις δραστηριότητές τους. Στις ομάδες συζήτησης που έχουν ανοικτό χαρακτήρα, οι έμπειροι χρήστες είναι πρόθυμοι να μοιραστούν τις γνώσεις τους, να καθοδηγήσουν και να βοηθήσουν άλλους χρήστες, οι οποίοι θέτουν ερωτήματα γύρω από υλοποιήσεις κυβερνο-απειλών. Παρατηρήθηκε πως, είτε τους συμβουλεύουν άμεσα, είτε παρέχουν τους απαραίτητους συνδέσμους ιστοσελίδων με το ανάλογο περιεχόμενο. Το συμπέρασμα είναι πως το πλούσιο υλικό που παρέχεται σε διάφορες μορφές, μπορεί να χρησιμοποιηθεί εύκολα ακόμα και από άπειρους χρήστες, πράγμα που τους καθιστά, αν μη τι άλλο, επικίνδυνους.

Βέβαια, δεν είναι λίγοι εκείνοι που υποστηρίζουν ότι παρέχουν παράνομες υπηρεσίες μέσα από τις ιστοσελίδες τους και έχουν ως μοναδικό στόχο να εξαπατήσουν τους χρήστες, αποσπώντας σημαντικά χρηματικά ποσά. Όντας προστατευμένοι από την ανωνυμία του δικτύου, έχουν τη δυνατότητα να αποφύγουν τον εντοπισμό. Τελικά, λόγω του ότι η αγορά τέτοιου είδους υπηρεσιών αποτελεί από μόνη της αδίκημα, οι χρήστες πρέπει να είναι έτοιμοι για την πιθανότητα να πέσουν ανά πάσα στιγμή θύματα απάτης, γνωρίζοντας ταυτόχρονα ότι δε θα είναι καλυμμένοι νομικά.



Εικόνα 6-1: Ιστοσελίδα με μήνυμα περί απάτης

6.2 Όρια και περιορισμοί της έρευνας

Η λειτουργία εύρεσης urls και της ανάκτησης πληροφοριών από αυτά πραγματοποιήθηκε με περιορισμούς που σχετίζονται με τους πόρους συστήματος. Λόγω αυτών, δεν ήταν δυνατή η εκτέλεση πολλαπλών στιγμιοτύπων της εφαρμογής, η οποία θα είχε ως αποτέλεσμα την ταχύτερη εξεύρεση νέων συνδέσμων και την παράλληλη αναζήτηση σε μεγάλο αριθμό ιστοσελίδων. Αυτός ήταν ένας από τους λόγους που προτιμήθηκε μια υλοποίηση της βάσης δεδομένων που να μην έχει μεγάλες απαιτήσεις σε υπολογιστική ισχύ.

Επιπλέον, λόγω της έλλειψης datasets που να περιέχουν πληροφορίες για τα διάφορα είδη κυβερνο-απειλών, δεν ήταν δυνατές οι δοκιμές μέσω της υλοποίησης και εφαρμογής αλγορίθμων μηχανικής μάθησης με επίβλεψη. Τέτοιου είδους αλγόριθμοι μπορούν να χρησιμοποιηθούν στην κατηγοριοποίηση δεδομένων και βασίζονται στην ύπαρξη μεγάλων datasets για την εκπαίδευσή τους. Οι πιο γνωστοί από αυτούς, που θα μπορούσαν να εφαρμοστούν στην κατηγοριοποίηση κειμένου είναι οι Naive Bayes, KNN (K-Nearest Neighbors) και Support Vector Machine (SVM).

6.3 Μελλοντικές επεκτάσεις

Μια από τις μελλοντικές επεκτάσεις γύρω από το θέμα της διπλωματικής είναι η δημιουργία ενός μοντέλου που θα χρησιμοποιεί μεθόδους μηχανικής μάθησης με επίβλεψη ή βαθιάς μάθησης (deep learning), με στόχο την κατηγοριοποίηση των δεδομένων που θα αποτελούν αποτέλεσμα συλλογής. Κάτι τέτοιο απαιτεί την ύπαρξη datasets για την εκπαίδευση των αλγορίθμων, τα οποία θα διαθέτουν πληθώρα πληροφοριών γύρω από τις κατηγορίες των κυβερνο-επιθέσεων. Όσο πιο μεγάλα είναι τα datasets, τόσο πιο μεγάλη η ακρίβεια και η αποτελεσματικότητα της συγκεκριμένης μεθοδολογίας. Πηγές πληροφοριών μπορούν να αποτελέσουν ιστοσελίδες του επιφανειακού και του σκοτεινού ιστού. Η επίτευξη της κατηγοριοποίησης θα αποτελέσει ένα χρήσιμο εργαλείο που θα συνεισφέρει στην υλοποίηση αμυντικών τεχνικών, στην πρόληψη και την προστασία συστημάτων που θα αποτελούν στόχο κυβερνο-επιθέσεων.

Μια ακόμη επέκταση είναι η απρόσκοπτη λειτουργία του συστήματος που αναπτύχθηκε στα πλαίσια της εργασίας, που θα έχει ως αποτέλεσμα τη συνεχή ανανέωση και ενημέρωση των δεδομένων που συλλέγονται. Με αυτόν τον τρόπο είναι εφικτή η διατήρηση μιας λίστας που θα εμπλουτίζεται με νέα urls, τα οποία θα κάνουν μελλοντικά την εμφάνισή τους, και ταυτόχρονα θα ενημερώνεται για την κατάσταση των urls που

ήδη αποτελούν μέρος της συλλογής. Αυτή η διαδικασία είναι εφικτή με την εκτέλεση της εφαρμογής σε σύστημα, όπως είναι ένας εξυπηρετητής, που θα χαρακτηρίζεται από την αδιάλειπτη λειτουργία και τη συνεχή σύνδεση στο διαδίκτυο, καθώς και από τον επαρκή αποθηκευτικό χώρο που θα καλύπτει το συνεχώς αυξανόμενο μέγεθος της βάσης δεδομένων.

Τέλος, μια αλλαγή που θα μπορούσε να γίνει στο υπάρχον σύστημα είναι η χρήση της MongoDB [87] για την αποθήκευση των δεδομένων. Πρόκειται για ένα δωρεάν σύστημα βάσης δεδομένων noSQL, ανοικτού κώδικα, το οποίο κάνει χρήση εγγράφων τύπου JSON και είναι ιδανικό για την αποθήκευση μεγάλου όγκου δεδομένων. Παρέχει μεγάλες ταχύτητες στην αναζήτηση, ανάκτηση και επεξεργασία δεδομένων και για αυτό το λόγο κερδίζει διαρκώς έδαφος στον τομέα της επιστήμης των δεδομένων και συγκεκριμένα στα Μεγάλα Δεδομένα (Big Data). Σε συνέχεια της προηγούμενης πρότασης, ένα σύστημα που θα είναι διαρκώς συνδεδεμένο στο διαδίκτυο συλλέγοντας δεδομένα, μακροπρόθεσμα θα έχει μεγαλύτερες απαιτήσεις όσον αφορά τον αποθηκευτικό χώρο και την ταχύτητα εκτέλεσης διεργασιών.

Η συνεχής και έγκαιρη άντληση πληροφοριών που αφορούν τις κυβερνο-επιθέσεις θα οδηγήσει στην ταχύτερη εξεύρεση νέων ειδών κυβερνο-απειλών. Αυτό θα προσφέρει σημαντική βοήθεια στην ανάπτυξη νέων μεθόδων για την πιο αποτελεσματική αντιμετώπισή τους, συμβάλλοντας στη δημιουργία των απαιτούμενων εργαλείων για την ασφάλεια των συστημάτων, διαδικασία που αποτελεί τον πρωταρχικό στόχο του τομέα του cyber threat intelligence.

7. Βιβλιογραφία

[1] Goodin D. "NSA-leaking Shadow Brokers just dumped its most damaging release yet" Ars Technica, 2017 [Online]. Available: https://arstechnica.com/information-technology/2017/04/nsa-leaking-shadow-brokers-just-dumped-its-most-damaging-release-yet/ [Accessed: 17 August 2018]

[2] Symantec, Website "Ransom.Wannacry", 2017 [Online]. https://www.symantec.com/security-center/writeup/2017-051310-3522-99 [Accessed: 17 August 2018]

[3] Kan M. "Biggest DDoS attack on record hits GitHub" PC Mag, 2018 [Online]. Available: https://www.pcmag.com/news/359610/biggest-ddos-attack-on- record-hits-github [Accessed: 17 August 2018]

[4] Morgan S. "Hackerpocalypse: A Cybercrime Revelation" Cybersecurity Ventures, 2016 [Online]. Available: https://cybersecurityventures.com/hackerpocalypse-original-cybercrime-report-2016/ [Accessed: 09 September 2018]

[5] Chambers J. "What does the Internet of Everything mean for security" World Economic Forum, 2017 [Online]. Available: https://www.weforum.org/agenda/2015/01/companies-fighting- cyber-crime/ [Accessed: 06 September 2018]

[6] Statista, Website "Annual number of data breaches and exposed records in the United States from 2005 to 2018 (in millions)" [Online] Available: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ [Accessed: 09 September 2018]

[7] Statista, Website "Global digital population as of July 2018" [Online] Available: https://www.statista.com/statistics/617136/digital-population-worldwide/ [Accessed: 09 October 2018]

[8] Encyclopedia.com, Website "Network of Networks" [Online] Available: https://www.encyclopedia.com/computing/news-wires-white-papers-and-books/network-networks [Accessed: 08 October 2018]

[9] Encyclopedia Britannica, Website "World Wide Web" [Online] Available: https://www.britannica.com/topic/World-Wide-Web [Accessed: 09 October 2018]

[10] Internet Live Stats, Website "Total number of Websites" [Online] Available: http://www.internetlivestats.com/total-number-of-websites/ [Accessed: 14 October 2018]

[11] Finklea K. "Dark Web", 2015 Available: http://aquadoc.typepad.com/files/crs_dark_web_10march2017.pdf

[12] Devine J., Egger-Sider F. "Going beyond Google: The Invisible Web in learning and teaching", 2009

[13] Wikipedia, Website "Deep Web" [Online] Available: https://en.wikipedia.org/wiki/Deep_web [Accessed: 02 November 2018]

[14] Tiwari A. "What is the difference between deep web, darknet and the dark web", Fossbytes, 2017 [Online] Available: https://fossbytes.com/difference-deep-web-darknet-dark-web/ [Accessed: 13 January 2019]

[15] Fachkha C., Debbabi M. "Darknet as a source for cyber threat intelligence : Survey, taxonomy and characterization", 2015 Available: https://www.researchgate.net/profile/Claude_Fachkha/publication/283827224_Darknet_as_a_Source_of_Cyber_Intelligence_Survey_Taxonomy_and_Characterization/links/5656265a08ae1ef92979db33.pdf

[16] Goldschlag D., Reed M., Syverson P. "Hiding Routing Information", 1996 Available: http://www.cs.jhu.edu/~fabian/courses/CS600.424/course_papers/goldschlag96hiding.pdf

[17] Luotonen A., "World Wide Web Proxies", 1994 Available: http://www.andrew.cmu.edu/course/15749/READINGS/optional/lutonen94.pdf

[18] Goldschlag D., Reed M., Syverson P. "Anonymous connections and Onion Routing", 1997 Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a465335.pdf

[19] Goldschlag D., Reed M., Syverson P. "Proxies for Anonymous Routing", 1996 Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a465331.pdf

[20] Goldschlag D., Reed M., Syverson P. "Onion Routing for anonymous and private internet connections", 1999 Available: https://www.onion-router.net/Publications/CACM-1999.pdf

[21] Dingledine R., Mathewson N., Syverson P. "Tor: The second-generation Onion Router", 2004 Available: https://svn.torproject.org/svn/projects/design-paper/tor-design.pdf

[22] Dierks T., Rescorla E. "The Transport Layer Security (TLS) protocol Version 1.2", 2008 Available: https://www.rfc-editor.org/rfc/pdfrfc/rfc5246.txt.pdf

[23] Wikipedia, Website "SOCKS" [Online] Available: https://en.wikipedia.org/wiki/SOCKS [Accessed: 03 November 2018]

[24] Techopedia, Website "Secure Hash Algorithm 1 (SHA-1)" [Online] Available: https://www.techopedia.com/definition/30570/secure-hash-algorithm-1-sha-1 [Accessed: 23 November 2018]

[25] Wikipedia, Website ".onion" [Online] Available: https://en.wikipedia.org/wiki/.onion [Accessed: 07 December 2018]

[26] The Tor Project, Website "Tor: Onion Service Protocol" [Online] Available: https://www.torproject.org/docs/onion-services [Accessed: 22 November 2018]

[27] Tor Metrics, Website [Online] Available: metrics.torproject.org [Accessed: 09 December 2018]

[28] The Tor Project, Website "Tor Bridges" [Online] Available: https://www.torproject.org/docs/bridges.html.en [Accessed: 09 December 2018]

[29] The Invisible Internet Project, Website [Online] Available: https://geti2p.net/el/ [Accessed: 25 November 2018]

[30] Timpanaro J.P., Chrisment I., Festor O. "A bird's eye view on the I2P anonymous file-sharing enviroment", 2012 Available: https://hal.inria.fr/file/index/docid/744919/filename/A_Birda_s_Eye_View_on_the_I2P_Anonymous_0AFile-sharing_Environment_0A.pdf

[31] B. Zantout, R. Haraty "I2P data communication system", 2011
Available: http://csm.beirut.lau.edu.lb/~rharaty/pdf/IC15.pdf

[32] Freenet, Website [Online]
Available: https://freenetproject.org/pages/about.html
[Accessed: 27 November 2018]

[33] Clarke I., Sandberg O., Wiley B., Hong T., "Freenet: A distributed anonymous information storage and retrieval system", 2000
Available: http://bourbon.usc.edu/cs694-s09/papers/freenet.pdf

[34] Wikipedia, Website "Freenet" [Online]
Available: https://en.wikipedia.org/wiki/Freenet
[Accessed: 27 November 2018]

[35] Bitcoin, Website "Frequently asked questions" [Online]
Available: https://bitcoin.org/en/faq#general

[36] Chen H., "Dark Web : Exploring and Data Mining the Dark Side of the Web", 2011

[37] CNSS (Committee on National Security Systems) Instruction No 4009, Revised: April 6 2015
Available:
https://www.cnss.gov/CNSS/openDoc.cfm?kJ0C+j6a2og3WDTJUukgNg==

[38] Symantec, SebastianZ "Security 1:1 - Part 3 - Various types of Network Attacks" Available: https://www.symantec.com/connect/articles/security-11-part-3- various-types- network-attacks [Accessed: 06 September 2018]

[39] DiGiacomo J. "Active vs passive cyber attacks explained" Revision Legal, 2017 [Online]. Available: https://revisionlegal.com/cyber-security/active-passive-cyber- attacks-explained/ [Accessed: 06 September 2018]

[40] Cisco, Website "What are the most common cyberattacks?" [Online]
Available: https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html [Accessed: 09 September 2018]

[41] Nash T. "An undirected attack against critical infrastucture", 2005 Available: https://ics-cert.us-cert.gov/sites/default/files/recommended_practices/CaseStudy-002.pdf

[42] Provos N., McNamee D., Mavrommatis P., Wang K., Modadugu N. "The Ghost in the browser analysis of web-based malware", 2007 Available: https://www.usenix.org/legacy/event/hotbots07/tech/full_papers/provos/provos.pdf

[43] Symantec, Website "What is a computer virus?" [Online] Available: https://us.norton.com/internetsecurity-malware-what-is-a-computer-virus.html [Accessed: 09 September 2018]

[44] Cisco, Website "What is the difference: Viruses, Worms, Trojans and Bots", 2018 [Online] Available: https://www.cisco.com/c/en/us/about/security-center/virus-differences.html [Accessed: 09 September 2018]

[45] Kaspersky, Website "Ransomware and Cyber Blackmail" [Online] Available: https://usa.kaspersky.com/resource-center/threats/ransomware [Accessed: 09 September 2018]

[46] Kaspersky, Website "What is a keylogger?" [Online] Available: https://www.kaspersky.com/resource-center/definitions/keylogger

[47] Kaspersky, Website "What is a rootkit and how to remove it", 2013 [Online] Available: https://www.kaspersky.com/blog/rootkit/1508/ [Accessed: 09 September 2018]

[48] Study.com, Website "What is a Backdoor virus? Definition, removal & example [Online] Available: https://study.com/academy/lesson/what-is-a-backdoor-virus-definition-removal-example.html [Accessed: 09 September 2018]

[49] Cisco, Website "What is Phishing?" [Online] Available: https://www.cisco.com/c/en/us/products/security/email- security/what-is-phishing.html [Accessed: 11 September 2018]

[50] ENISA, Website "Man-in-the-middle" [Online] Available: https://www.enisa.europa.eu/topics/csirts-in-europe/glossary/man-in-the-middle [Accessed: 11 September 2018]

[51] Norton.com, Website "What is a man-in-the-middle attack?" [Online] Available: https://us.norton.com/internetsecurity-wifi-what-is-a-man-in-the-middle-attack.html [Accessed: 11 September 2018]

[52] Techopedia, Website "Address Resolution Protocol (ARP)" [Online] Available: https://www.techopedia.com/definition/5493/address-resolution-protocol-arp [Accessed: 11 September 2018]

[53] Techopedia, Website "Address Resolution Protocol (ARP) Poisoning" [Online] Available: https://www.techopedia.com/definition/27471/address-resolution-protocol-poisoning-arp-poisoning [Accessed: 11 September 2018]

[54] Sanders C. "Understanding man-in-the-middle attacks: Part 2 - DNS Spoofing", 2010 [Online] Available: http://techgenix.com/Understanding-Man-in-the-Middle-Attacks-ARP-Part2/ [Accessed: 11 September 2018]

[55] Sanders C. "Understanding man-in-the-middle attacks: Part 4: SSL Hijacking", 2010 [Online] Available: http://techgenix.com/understanding- man-in-the-middle-attacks-arp-part4/ [Accessed: 20 September 2018]

[56] NCCIC "Security Tip (ST04-015) Understanding Denial-of-Service Attacks" US-CERTH, 2018 [Online] Available: https://www.us-cert.gov/ncas/tips/ST04-015 [Accessed: 20 September 2018]

[57] Specht S., Lee R. "Distributed Denial of Service: Taxonomies of Attacks, Tools and Countermeasures" Available: https://www.researchgate.net/profile/Ruby_Lee/publication/220922510_Distributed_Denial_of_Service_Taxonomies_of_Attacks_Tools_and_Countermeasures/links/0fcfd50bd9fda81b51000000.pdf

[58] Techopedia, Website "Botnet" [Online] Available: https://www.techopedia.com/definition/384/botnet [Accessed: 22 September 2018]

[59] Sharma S., Garg S., Karodiya A., H. Gupta "Distributed Denial of Service Attack", 2015 Available: http://www.ijser.in/archives/v4i11/IJSER151057.pdf

[60] Buehrer G., Weide B., Sivilotti P. "Using parse tree validation to prevent SQL injection attacks", 2005 Available: https://www.researchgate.net/profile/Bruce_Weide/publication/221215947_Using_parse_tree_validation_to_prevent_SQL_injection_attacks/links/02bfe5117c849676a6000000/Using-parse-tree-validation-to-prevent-SQL-injection-attacks.pdf

[61] Halfond W., Viegas J., Orso A. "A Classification of SQL Injection Attacks and Coutermeasures", 2006 Available: www.cc.gatech.edu/fac/Alex.Orso/papers/halfond.viegas.orso.ISSSE06.pdf

[62] Kaspersky, Website "What is a Zero-day Exploit?" [Online] Available: https://www.kaspersky.com/resource-center/definitions/zero-day-exploit [Accessed: 22 September 2018]

[63] ENISA, Website "Zero Day" [Online] Available: https://www.enisa.europa.eu/topics/csirts-in-europe/glossary/zero-day [Accessed: 22 September 2018]

[64] Spett K. "Cross-Site Scripting - Are your web applications vulnerable?", SPI Dynamics, 2005 Available: http://people.cs.ksu.edu/~hankley/d764/Topics/SPIcross-sitescripting.pdf

[65] Trend Micro, Website "Cross-site scripting" [Online] Available: https://www.trendmicro.com/vinfo/us/security/definition/cross-site-scripting-(xss) [Accessed: 23 September 2018]

[66] Techopedia, Website "Credential stuffing" [Online] Available: https://www.techopedia.com/definition/32586/credential-stuffing

[67] Keach S. "Facebook hacked - attackers logged into 50 million profile, with access to posts, photos and messages in security breach disaster", 2017 [Online] Available: https://www.thesun.co.uk/tech/7373518/facebook-hack-profiles-computer-security-password-change/ [Accessed: 04 December 2018]

[68] Chismon D. Ruks M. "Threat Intelligence: Collecting, Analyzing, Evaluating", 2015 https://www.ncsc.gov.uk/content/files/protected_files/guidance_files/MWR_Threat_Intelligence_whitepaper-2015.pdf

[69] Fu T., Abbasi A., Chen H. "A focused crawler for dark web forums", 2010 Available: https://www.researchgate.net/profile/Ahmed_Abbasi4/publication/220432724_A_Focused_Crawler_for_Dark_Web_Forums/links/5a9f4757a6fdcc22e2cb521f/A-Focused-Crawler-for-Dark-Web-Forums.pdf

[70] Benjamin V., Li W., Holt T., Chen H. "Exploring threats and vulnerabilities in hacker web: Forums, IRC and Carding Shops", 2015
Available:
https://ai.arizona.edu/sites/ai/files/AILabCybersecurityPapers/benjamin_et_al_2015_exploring_threats_and_vulnerabilities_in_hacker_web_forums_irc_and_carding_shops.pdf

[71] Samtani S., Chinn K., Larson K., Chen H. "AZSecure hacker assets portal: Cyber threat intelligence and malware analysis", 2016 Available:
https://ai.arizona.edu/sites/ai/files/AILabCybersecurityPapers/samtani_et_al_2016_azsecure_hacker_assets_portal.pdf

[72] Soska K., Christin N. "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem", 2015 Available:
https://www.usenix.org/system/files/conference/usenixsecurity15/sec15-paper-soska.pdf

[73] Requests, Website "Requests: HTTP for humans" [Online]
Available: https://github.com/requests/requests
[Accessed: 09 December 2018]

[74] Crummy: The Site, Website "BeautifulSoup documentation" [Online]
Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/
[Accessed: 09 December 2018]

[75] Python.org, Website "20.16 urlparse - Parse URLs into components" [Online]
Available: https://docs.python.org/2/library/urlparse.html
[Accessed: 09 December 2018]

[76] The Tor Project, Website "Stem docs" [Online]
Available: https://stem.torproject.org/
[Accessed: 09 December 2018]

[77] pypi.org, Website "fake-useragent 0.1.11" [Online]
Available: https://pypi.org/project/fake-useragent/
[Accessed: 09 December 2018]

[78] Python.org, Website "11.13. sqlite3 - DB-API 2.0 interface for SQLite databases" [Online] Available: https://docs.python.org/2/library/sqlite3.html
[Accessed: 09 December 2018]

[79] Scikit-learn, Website [Online] Available: https://scikit-learn.org/stable/ [Accessed: 02 January 2019]

[80] Wikipedia, Website "K-Means clustering" [Online] Available: https://en.wikipedia.org/wiki/K-means_clustering [Accessed: 31 December 2018]

[81] The Tor Project, Website [Online] Available: https://www.torproject.org/download/download-easy.html.en [Accessed: 12 December 2018]

[82] The Tor Project, Website "Tor Project: FAQ" [Online] Available: https://www.torproject.org/docs/faq.html.en#TBBOtherBrowser [Accessed: 12 December 2018]

[83] Wikipedia, Website "Virtual Private Network" [Online] Available: https://en.wikipedia.org/wiki/Virtual_private_network [Accessed: 10 December 2018]

[84] BolehVPN, Website "TOR over VPN & VPN over TOR: Which is better?" [Online] Available: https://blog.bolehvpn.net/tor-over-vpn-vpn-over-tor-which-is-better/ [Accessed: 10 December 2018]

[85] NordVPN, Website [Online] Available: https://nordvpn.com/ [Accessed: 10 December 2018]

[86] Ahmia, Website [Online] Available: http://www.msydqstlz2kzerdg.onion [Accessed: 04 December 2018]

[87] MongoDB, Website "What is MongoDB?" [Online] Available: https://www.mongodb.com/what-is-mongodb [Accessed: 04 January 2019]