

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Διπλωματική Εργασία

Αξιοποίηση Γραφημάτων Γνώσης (Knowledge Graphs) στην Ανάλυση Δεδομένων
Δικτύου TCP/IP για την Ενίσχυση της Κυβερνοασφάλειας

ΣΑΡΙΔΑΚΗΣ ΓΕΩΡΓΙΟΣ

A.M: 1072478

Επιβλέπων: Παπαϊωάννου Βάιος, Καθηγητής Ε.Δ.Ι.Π. Π.Πατρών

Εξεταστές: Σιούτας Σπυρίδων, Καθηγητής Π.Πατρών

Τσίχλας Κωνσταντίνος, Αναπληρωτής Καθηγητής Π.Πατρών

Πάτρα, Σεπτέμβριος, 2024

Περίληψη

Δεδομένης της αυξανόμενης πολυπλοκότητας και του όγκου των κυβερνοεπιθέσεων, οι παραδοσιακές μέθοδοι ασφάλειας δικτύου κρίνονται συχνά ανεπαρκείς για τον εντοπισμό και την αντιμετώπιση αυτών των απειλών .

Η διπλωματική εργασία εξετάζει την εφαρμογή των γραφημάτων γνώσης στην ανάλυση δεδομένων δικτύου TCP/IP με στόχο τη βελτίωση της κυβερνοασφάλειας. Αρχικά, γίνεται μελέτη του ευρύτερου γνωστικού υπόβαθρου όπως η οντολογία και τα γραφήματα γνώσης. Στην συνέχεια αφότου τεθούν οι στόχοι της εργασίας παρατίθεται η παρουσίαση των δεδομένων που θα χρησιμοποιηθούν για την υλοποίηση της έρευνας.

Τα δεδομένα αυτά σε συνδυασμό με την δημιουργία μιας οντολογίας (Ontology) στο περιβάλλον Protégé εισάγονται σε ένα δυναμικό script σε γλώσσα Python το οποίο δημιουργεί το knowledge graph χρησιμοποιώντας βιβλιοθήκες και πλατφόρμες όπως το Neo4j. Εκεί γίνεται η εκτέλεση κάποιων queries και προκύπτουν αποτελέσματα αποκαλύπτοντας μοτίβα και ανωμαλίες στην κυκλοφορία του δικτύου που διαφορετικά θα μπορούσαν να περάσουν απαρατήρητα αποδεικνύοντας την χρησιμότητα και την αποτελεσματικότητα των γραφημάτων γνώσης για την ενίσχυση της κυβερνοασφάλειας.

Συμπερασματικά αποδεικνύεται ότι το μοντέλο ανάλυσης δεδομένων δικτύων TCP/IP με χρήση γραφημάτων γνώσης (knowledge graph) προσφέρει ένα αξιόπιστο εργαλείο στον αναλυτή για την ανίχνευση απειλών κυβερνοασφάλειας.

Abstract

Given the increasing sophistication and the volume of the cyber-attacks, traditional network security methods are usually considered insufficient to detect and counter these threats.

First, a study of the broader knowledge background such as ontology and knowledge graphs is done. Then, after setting the objectives of the work, the presentation of the data that will be used for the implementation of the research is listed.

These data combined with the creation of an ontology in the Protégé environment are entered into a dynamic script in Python language which creates the knowledge graph using libraries and platforms such as Neo4j. There, some queries are executed, and results emerge, revealing patterns and anomalies in network traffic that might otherwise go unnoticed, proving the usefulness and effectiveness of knowledge graphs in enhancing cybersecurity.

In conclusion, it is proven that the TCP/IP network data analysis model using knowledge graphs offers a reliable tool for the analyst to detect cyber security threats.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Βάιο Παπαϊωάννου κατ' αρχήν για την ανάθεση της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να ασχοληθώ με τον τομέα της κυβερνοασφάλειας σε ένα εξαιρετικά ενδιαφέρον επιστημονικό πεδίο της ανάλυσης δεδομένων διαδικτύου με την χρήση γράφων για εξεύρεση απειλών.

Επίσης για την καθοδήγηση, συνεργασία και επίβλεψη της διπλωματικής εργασίας, δίνοντας μου οδηγίες και υποδείξεις καθόλη τη διάρκεια της εκπόνησης της, επιτρέποντας μου ταυτόχρονα τον χώρο ανάπτυξης δικών μου προτάσεων και ιδεών μοντελοποίησης των δεδομένων ανάλυσης που χρησιμοποιήθηκαν.

Τέλος , θα ήθελα να ευχαριστήσω τους γονείς και τους φίλους μου για την ανιδιοτελή υποστήριξη σε κάθε τομέα των σπουδών μου όλα αυτά τα χρόνια, με αποτέλεσμα να ολοκληρώνω με επιτυχία τον πενταετή κύκλο φοίτησης με την απόκτηση πτυχίου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....	9
1.1 Θέμα.....	9
1.2 Στόχοι Διπλωματικής Εργασίας.....	10
1.3 Σημασία της έρευνας στην κυβερνοασφάλεια.....	11
ΚΕΦΑΛΑΙΟ 2: ΜΕΛΕΤΗ ΠΕΡΙΟΧΗΣ.....	14
2.1 Αξιοποίηση TCP/IP data.....	14
2.2 Ανάπτυξη της τεχνολογίας των ontologies και η σχέση της με την κυβερνοασφάλεια.....	17
2.3 Θεωρητικό υπόβαθρο των Knowledge graphs.....	21
ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΠΡΟΒΛΗΜΑΤΟΣ-ΣΧΕΔΙΑΣΜΟΣ.....	29
3.1 Περιγραφή των δεδομένων που χρησιμοποιήθηκαν.....	30
3.2 Ανάπτυξη του ontology.....	34
3.3 Ανάπτυξη script.....	39
3.4 Κατασκευή του knowledge graph.....	40
3.5 Περιγραφή των queries και τεχνικών που χρησιμοποιήθηκαν για την ανάλυση του knowledge graph.....	42
ΚΕΦΑΛΑΙΟ 4: ΥΛΟΠΟΙΗΣΗ.....	45
4.1 Διαδικασία Ανάλυσης και Προετοιμασίας των δεδομένων.....	45
4.2 Υλοποίηση του script και αντιστοίχιση δεδομένων με βάση το ontology.....	49
4.3 Περιγραφή της διαδικασίας υλοποίησης του Knowledge graph.....	51
4.4 Ανάλυση των αποτελεσμάτων που προέκυψαν από τα queries.....	53
ΚΕΦΑΛΑΙΟ 5: ΑΞΙΟΛΟΓΗΣΗ.....	61
5.1 Συζήτηση για τα ευρήματα και την επίδραση τους στην κυβερνοασφάλεια.....	61
5.2 Συζήτηση για τις προοπτικές και της δυνατότητες βελτίωσης της μεθοδολογίας.....	64
ΚΕΦΑΛΑΙΟ 6: ΕΠΙΛΟΓΟΣ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ.....	67

ΒΙΒΛΙΟΓΡΑΦΙΑ.....	70
ΠΑΡΑΡΤΗΜΑ Α-ΑΡΧΕΙΑ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV.....	74
ΠΑΡΑΡΤΗΜΑ Β – Η ΔΟΜΗ ΤΟΥ ONTOLOGY ΣΕ ΜΟΡΦΗ .TTL.....	79
ΠΑΡΑΡΤΗΜΑ Γ-ΑΚΡΩΝΥΜΙΑ.....	84

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

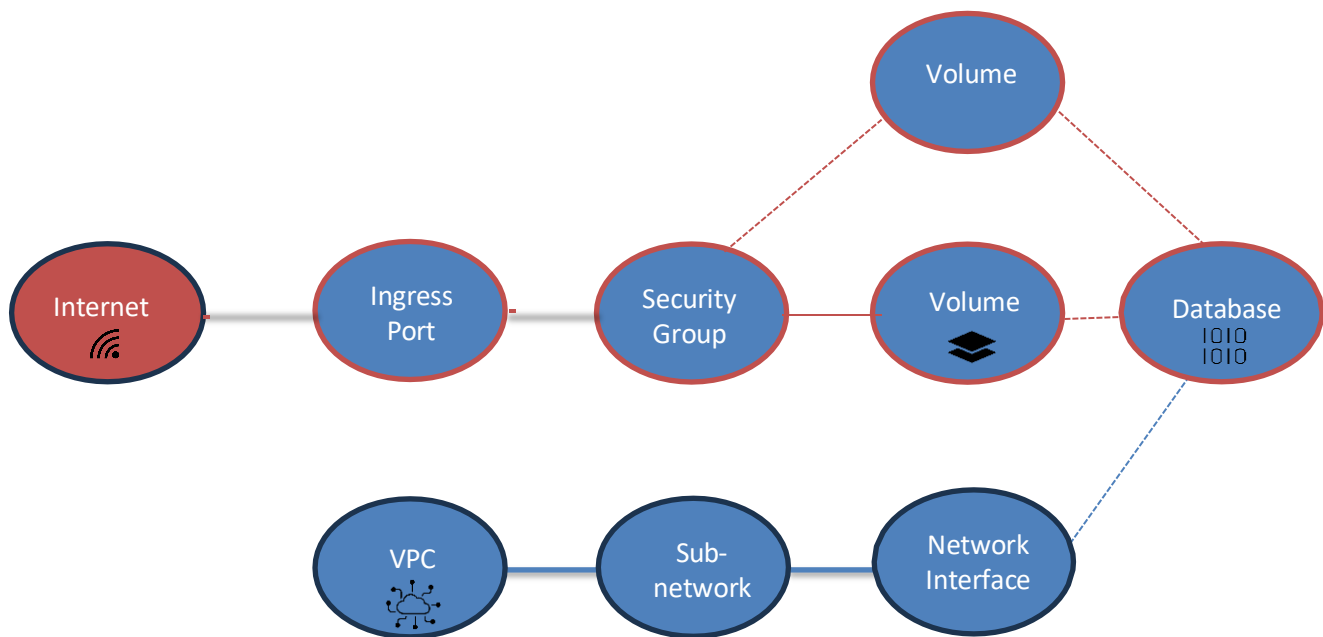
Εικόνα 1-Μοντέλο Security Knowledge Graph [16]	10
Εικόνα 2-TCP/IP Suite	14
Εικόνα 3-Παράδειγμα Ontology	18
Εικόνα 4-Στοιχεία GRC	18
Εικόνα 5-Παράδειγμα γραφήματος γνώσης	22
Εικόνα 6-Παράδειγμα DEL Graph	24
Εικόνα 7-Παράδειγμα Heterogeneous Graph	25
Εικόνα 8-Παράδειγμα Property Graph	26
Εικόνα 9-Παράδειγμα Graph Dataset	27
Εικόνα 10-Σχεδιασμός-Διάγραμμα ροής του προβλήματος	29
Εικόνα 11-Παράδειγμα εταιρικών δεδομένων	31
Εικόνα 12-Παράδειγμα δεδομένων γεωγραφικής θέσης IP	32
Εικόνα 13-Παράδειγμα δεδομένων TCP/IP Traffic	34
Εικόνα 14-Protege Class Browser	35
Εικόνα 15-Περιβάλλον δημιουργίας των Entities	36
Εικόνα 16-Περιβάλλον δημιουργίας των Data Properties	36
Εικόνα 17-Ιδιότητες Data Properties(Geolocation)	37
Εικόνα 18-Ιδιότητες Data Properties(Size)	37
Εικόνα 19-Περιβάλλον δημιουργίας των Object Properties	38
Εικόνα 20-Αντιστοίχιση των Object Properties	38
Εικόνα 21-Το ολοκληρωμένο Ontology	39
Εικόνα 22-Βήματα δημιουργίας του γραφήματος γνώσης	42
Εικόνα 23-Συνάρτηση στατιστικής ανασκόπησης	45
Εικόνα 24-Αποτέλεσμα συνάρτησης στατιστικής ανασκόπησης	46
Εικόνα 25-Συνάρτηση υπολογισμού συσχετίσεων μεταξύ X,Y και αριθμητικών πεδίων	46
Εικόνα 26-Αποτέλεσμα συνάρτησης υπολογισμού συσχετίσεων μεταξύ X,Y και αριθμητικών πεδίων	46
Εικόνα 27-Απόσπασμα διόρθωσης σφάλματος στον κώδικα	47
Εικόνα 28-Παράδειγμα τυποποίησης δεδομένων	47
Εικόνα 29-Παράδειγμα ομαλοποίησης από τον κώδικα	48
Εικόνα 30-Ενοποίηση δεδομένων	48
Εικόνα 31-Παράδειγμα ελέγχου συνέπειας	49

Εικόνα 32-Παράδειγμα ελέγχου πληρότητας.....	49
Εικόνα 33-Συνάρτηση για την εισαγωγή του Ontology στο script	50
Εικόνα 34-Συνάρτηση ελέγχου.....	50
Εικόνα 35-Δημιουργία του connectionid node.....	50
Εικόνα 36-Επιπλέον διαδικασίες annotation.....	51
Εικόνα 37-Σύνδεση με το Neo4j	51
Εικόνα 38-Η γραφική απεικόνιση μετά την δημιουργία του knowledge graph.....	52
Εικόνα 39-Μεγένθυση του knowldege graph	52
Εικόνα 40-Query του 1ου σεναρίου σε Cypher.....	54
Εικόνα 41-Το αποτέλεσμα του 1ου query.....	55
Εικόνα 42-Query του 2ου σεναρίου σε Cypher.....	56
Εικόνα 43-Το αποτέλεσμα του 2ου query.....	57
Εικόνα 44-1ο μέρος του 3ου query σε Cypher	58
Εικόνα 45-2ο μέρος του 3ου query σε Cypher	59
Εικόνα 46-Το αποτέλεσμα του 3ου query.....	59
Εικόνα 47-Το αποτέλεσμα του 3ου query σε μεγέθυνση.....	60

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Θέμα

Είναι γνωστό ότι βρισκόμαστε σε μια ταχέως εξελισσόμενη ψηφιακή εποχή, όπου η ασφάλεια των δικτυωμένων συστημάτων στο διαδίκτυο και όχι μόνο είναι υψίστης σημασίας. Το Πρωτόκολλο Ελέγχου Μετάδοσης/Πρωτόκολλο Διαδικτύου (TCP/IP) αποτελεί την βασική σουίτα πρωτοκόλλων επικοινωνίας που χρησιμοποιούνται στο Διαδίκτυο και είναι εγγενώς πολύπλοκο και επιρρεπές σε ποικίλες απειλές στον κυβερνοχώρο [22]. Καθώς ο όγκος και η πολυπλοκότητα των επιθέσεων στον κυβερνοχώρο αυξάνονται σε συνάρτηση με την ενσωμάτωση όλο και περισσότερων συσκευών στο διαδίκτυο για πιο απλές και καθημερινές χρήσεις, οι παραδοσιακές μέθοδοι ασφάλειας δικτύου είναι συχνά ανεπαρκείς για τον εντοπισμό και τον μετριασμό αυτών των απειλών αποτελεσματικά. Για να χειριστούν τους τεράστιους όγκους ετερογενών δεδομένων σχετικά με τις απειλές στον κυβερνοχώρο, ορισμένες εταιρείες κυβερνοασφάλειας έχουν στραφεί σε γραφήματα γνώσης [24]. Αυτές οι αναγνώσιμες από μηχανή δομές γραφημάτων βοηθούν τις επιχειρήσεις να ενσωματώνουν ομαλά τόσο δομημένα όσο και μη δομημένα δεδομένα σε μια σημασιολογική αναπαράσταση οντοτήτων του πραγματικού κόσμου και των σχέσεων μεταξύ τους. Αξιοποιώντας τη σχεσιακή δομή των γραφημάτων γνώσης, είναι δυνατό να αποκαλυφθούν μοτίβα και ανωμαλίες εντός της κυκλοφορίας του δικτύου που διαφορετικά θα μπορούσαν να περάσουν απαρατήρητες. Αυτή η ικανότητα καθιστά τα γραφήματα γνώσης ένα ισχυρό εργαλείο στον τομέα της κυβερνοασφάλειας, ιδιαίτερα για τον εντοπισμό και την απόκριση σε απειλές που βασίζονται στο δίκτυο[13]. Αυτή η εργασία διερευνά την καινοτόμο χρήση των γραφημάτων γνώσης στην ανάλυση δεδομένων δικτύου TCP/IP για τη βελτίωση των μέτρων ασφάλειας στον κυβερνοχώρο.



Εικόνα 1-Μοντέλο Security Knowledge Graph [16]

1.2 Στόχοι Διπλωματικής Εργασίας

Ο πρωταρχικός σκοπός αυτής της εργασίας είναι να διερευνήσει τις δυνατότητες των γραφημάτων γνώσης για τη βελτίωση της ανάλυσης δεδομένων δικτύου TCP/IP για βελτιωμένα αποτελέσματα στον κυβερνοχώρο. Αυτό περιλαμβάνει πολλούς βασικούς στόχους.

- Διερεύνηση της ενσωμάτωσης γραφημάτων γνώσης με δεδομένα δικτύου. Αυτό θα γίνει μετά από κατανόηση του πώς τα γραφήματα γνώσης μπορούν να κατασκευαστούν αποτελεσματικά από δεδομένα δικτύου TCP/IP.
- Προσπάθεια ανάπτυξης μεθόδων για την ανίχνευση απειλών, δηλαδή να δημιουργήσει και να χρησιμοποιήσει τεχνικές που χρησιμοποιούν τη δομή των γραφημάτων γνώσης για τον εντοπισμό και την πρόβλεψη απειλών στον κυβερνοχώρο με μεγαλύτερη ακρίβεια και ταχύτητα.
- Η πρόταση πρακτικών τρόπων με τους οποίους η ανάλυση δικτύου βάσει γραφημάτων γνώσης μπορεί να ενσωματωθεί στα υπάρχοντα πλαίσια ασφάλειας στον κυβερνοχώρο τονίζοντας και αναδεικνύοντας τα πλεονεκτήματα της χρήσης αυτών μέσα από την ανάλυση δεδομένων δικτύου και όχι μόνο.
- Η προσπάθεια εντοπισμού ανωμαλιών . Η ανάλυση των δεδομένων με την

βοήθεια των γράφων δίνει την δυνατότητα ανακάλυψης ανωμαλιών και δραστηριοτήτων από μη εξουσιοδοτημένους χρήστες.

- Θωράκιση του συστήματος κυβερνοασφάλειας μέσα από τον εντοπισμό διόδων που θα μπορούσαν να εκμεταλλευτούν κακόβουλοι χρήστες και λογισμικά με στόχο να κερδίσουν τον έλεγχο του συστήματος ή να εξάγουν απόρρητά δεδομένα.
- Ως επιμέρους στόχος τίθεται επίσης η αναγνώριση και καταγραφή των προτεινομένων βελτιώσεων για μελλοντική χρήση της εργασίας.

1.3 Σημασία της έρευνας στην κυβερνοασφάλεια

Η IT βιομηχανία έχει εξελιχθεί σε μεγάλο βαθμό τον τελευταίο μισό αιώνα. Η συνεχής εκθετική πρόοδος στην επεξεργαστική ισχύ και τη χωρητικότητα μνήμης έχει κάνει το hardware όχι μόνο πιο γρήγορο αλλά μικρότερο, ελαφρύτερο, φθηνότερο και πιο εύκολο στη χρήση.

Οι ειδικοί και οι υπεύθυνοι χάραξης πολιτικής έχουν εκφράσει αυξανόμενες ανησυχίες σχετικά με την προστασία των συστημάτων από κυβερνοεπιθέσεις- σκόπιμες απόπειρες πρόσβασης από μη εξουσιοδοτημένα άτομα σε συστήματα ICT, συνήθως με στόχο την κλοπή, τη διακοπή, τη ζημιά ή άλλες παράνομες ενέργειες[14]. Πολλοί ειδικοί αναμένουν ότι ο αριθμός και η σοβαρότητα των cyberattacks θα αυξηθούν τα επόμενα αρκετά χρόνια. Η πράξη προστασίας των συστημάτων ΤΠΕ και του περιεχομένου τους έχει γίνει γνωστή ως κυβερνοασφάλεια. Μία ευρεία και αναμφισβήτητη κάπως ασαφή έννοια, η ασφάλεια στον κυβερνοχώρο μπορεί να είναι ένας χρήσιμος όρος, αλλά τείνει να αψηφά έναν ακριβή ορισμό. Συνήθως αναφέρεται σε ένα ή περισσότερα από τρία πράγματα:

- Ένα σύνολο δραστηριοτήτων και άλλων μέτρων που αποσκοπούν στην προστασία—από επίθεση, διακοπή ή άλλες απειλές—υπολογιστές, δίκτυα υπολογιστών, σχετικό hardware και το software συσκευών και τις πληροφορίες που περιέχουν και επικοινωνούν, συμπεριλαμβανομένου λογισμικού και δεδομένων, καθώς και άλλων στοιχείων του κυβερνοχώρου.
- Η κατάσταση ή η ποιότητα της προστασίας από τέτοιες απειλές.

- Το ευρύ πεδίο προσπάθειας που στοχεύει στην εφαρμογή και βελτίωση αυτών των δραστηριοτήτων και ποιότητας αυτών.

Η επιτυχία μιας κυβερνοεπίθεσης μπορεί να είναι καταστροφική για έναν ιδιώτη, μια εταιρεία, ένα κράτος ή έναν διεθνή οργανισμό έχοντας ως αποτέλεσμα ένα ή περισσότερα από τα παρακάτω:

- Ο χρόνος διακοπής λειτουργίας δημιουργεί απώλεια κρίσιμων δυνατοτήτων οργάνωσης, αδυναμία πρόσβασης των πελατών στα συστήματα, η οποία συχνά σημαίνει άμεση απώλεια υπηρεσιών.
- Η βλάβη της φήμης της εταιρείας μετά από μια παραβίαση δεδομένων
- Η αποκατάσταση κάθε είδους παραβίασης συνεπάγεται οικονομικό κόστος. Το μέσο κόστος μιας παραβίασης δεδομένων το 2021 ήταν 4,24 εκατομμύρια δολάρια [5].
- Πολλές χώρες και βιομηχανίες έχουν κανονισμούς που απαιτούν από τους οργανισμούς να προστατεύουν τα προσωπικά δεδομένα που έχουν αποθηκευμένα. Οι παραβιάσεις αυτών των κανονισμών συνοδεύονται από αυστηρά πρόστιμα και ποινές. Οι εταιρείες που δραστηριοποιούνται στην Ευρωπαϊκή Ένωση ενδέχεται να αντιμετωπίσουν πρόστιμα έως και 4 τοις εκατό των ετήσιων ακαθάριστων εσόδων εάν παραβιάζουν το Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR).

Η σημασία αυτής της έρευνας έγκειται :

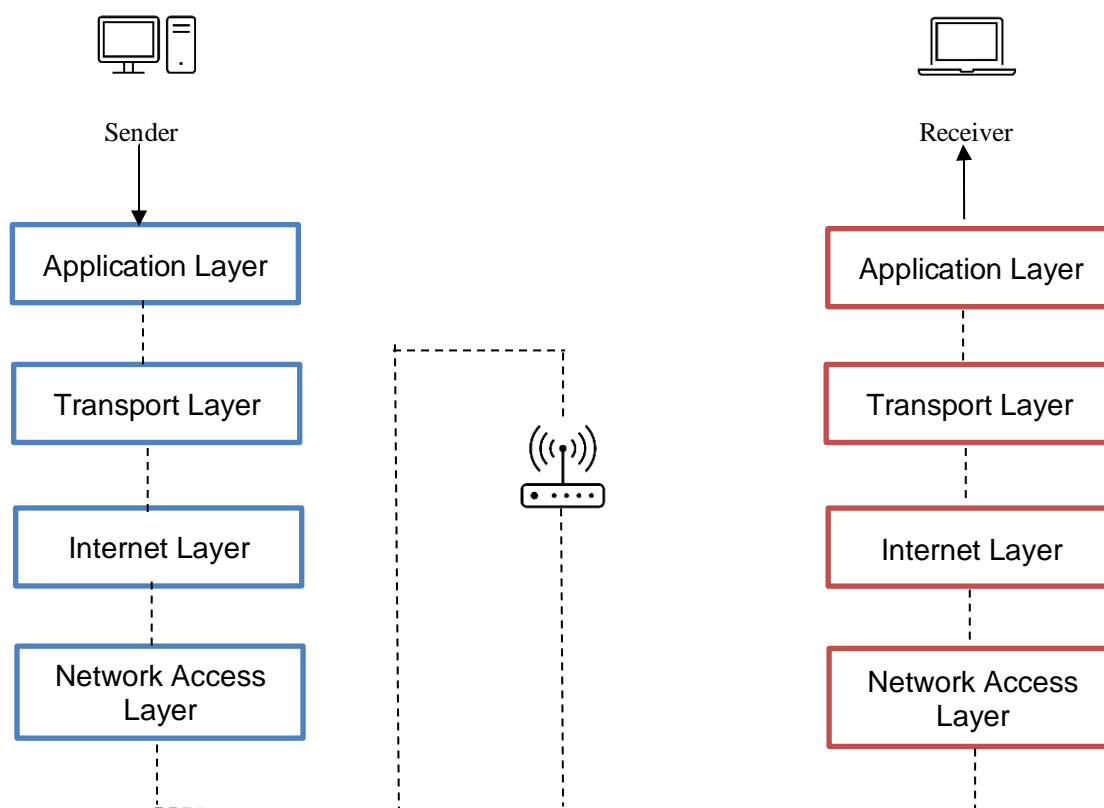
- Στη δυνατότητά της να μεταμορφώσει τον τρόπο με τον οποίο αναλύονται και χρησιμοποιούνται τα δεδομένα δικτύου για την καταπολέμηση των απειλών στον κυβερνοχώρο. Τα τρέχοντα εργαλεία κυβερνοασφάλειας βασίζονται συχνά σε μοντέλα ανίχνευσης βάσει υπογραφών ή μηχανικής μάθησης που έχουν εκπαιδευτεί σε ιστορικά δεδομένα επιθέσεων. Αν και αποτελεσματικές σε κάποιο βαθμό, αυτές οι μέθοδοι δεν μπορούν να αντιμετωπίσουν νέες ή εξελιγμένες επιθέσεις που αποκλίνουν από γνωστά πρότυπα [26].
- Τα γραφήματα γνώσης προσφέρουν μια πιο δυναμική και ευέλικτη προσέγγιση, ικανή να προσαρμόζεται σε νέες απειλές κατανοώντας τις σχέσεις και τα πλαίσια μέσα στα δεδομένα δικτύου [14].

- Αξιοποιώντας γραφήματα γνώσης, οι επαγγελματίες της κυβερνοασφάλειας μπορούν να αποκτήσουν βαθύτερες γνώσεις για τη συμπεριφορά των οντοτήτων δικτύου, αποκαλύπτοντας κρυφές σχέσεις και μοτίβα που μπορεί να υποδηλώνουν κακόβουλη δραστηριότητα.
- Στην δημιουργία πιο προληπτικών και ισχυρών μέτρων ασφαλείας, μειώνοντας τον χρόνο εντοπισμού και ανταπόκρισης σε περιστατικά στον κυβερνοχώρο.
- Τελικά, αυτή η έρευνα στοχεύει να συμβάλει στον γενικότερο στόχο για ένα πιο ασφαλές ψηφιακό περιβάλλον, προστατεύοντας ευαίσθητες πληροφορίες και κρίσιμες υποδομές από απειλές στον κυβερνοχώρο.

ΚΕΦΑΛΑΙΟ 2: ΜΕΛΕΤΗ ΠΕΡΙΟΧΗΣ

2.1 Αξιοποίηση TCP/IP data

Ένας βασικός κλάδος της κυβερνοασφάλειας είναι η ασφάλεια της επικοινωνίας μέσω διαδικτύου. Το TCP/IP suite είναι το πιο ευρέως χρησιμοποιούμενο πρωτόκολλο επικοινωνίας και έχει γίνει απαραίτητο πρότυπο για επικοινωνίες που βασίζονται στο Διαδίκτυο. Είναι μια συλλογή που βασίζεται σε δίκτυο πρωτόκολλων επικοινωνίας που παρέχουν και υποστηρίζουν διάφορα είδη υπηρεσιών που εκτελούνται μέσω του δικτύου. Καθιερώνει, διατηρεί και τερματίζει τις συνδέσεις μεταξύ των άκρων σημείων και παρέχει πλήρη αμφίδρομη συνδεσιμότητα από άκρο σε άκρο. Ακόμη μορφοποιεί δεδομένα, διευθύνσεις, δρομολογεί τα πακέτα δεδομένων του δικτύου και διασφαλίζει ότι παραδίδονται στον παραλήπτη. Δύο βασικά στοιχεία του πρωτοκόλλου TCP/IP suite είναι το πρωτόκολλο ελέγχου μετάδοσης TCP και το Πρωτόκολλο IP[25].



Εικόνα 2-TCP/IP Suite

Η αξιοποίηση δεδομένων TCP/IP περιλαμβάνει την εξέταση των πακέτων δεδομένων που ταξιδεύουν σε ένα δίκτυο για την παρακολούθηση και τη βελτίωση της απόδοσης, τον εντοπισμό ανωμαλιών και τη βελτίωση της ασφάλειας[4]. Αυτή η διαδικασία είναι θεμελιώδης για την ασφάλεια στον κυβερνοχώρο, επιτρέποντας την παρακολούθηση σε πραγματικό χρόνο και την ιστορική ανάλυση των δραστηριοτήτων του δικτύου. Η διαδικασία αυτή περιλαμβάνει την ανάλυση επισκεψιμότητας δικτύου (NTA) .

Το NTA είναι μια μέθοδος που χρησιμοποιείται για τη σύλληψη και την επιθεώρηση πακέτων δεδομένων που ρέουν μέσω ενός δικτύου. Υπάρχουν δύο κύριοι τύποι πηγών δεδομένων για το NTA:

- *Δεδομένα ροής(Flow data)*: Λήφθηκαν από συσκευές δικτύου όπως δρομολογητές, οι οποίοι παρέχουν πληροφορίες σχετικά με τον όγκο κίνησης και τις διαδρομές που ακολουθούν τα πακέτα δεδομένων. Αυτό είναι χρήσιμο για την κατανόηση της απόδοσης του δικτύου και τον εντοπισμό μη εξουσιοδοτημένης κίνησης. Αυτό μπορεί να πραγματοποιηθεί με εργαλεία όπως το NetFlow(τεχνολογία της Cisco που συλλέγει πληροφορίες κίνησης IP, επιτρέποντας στους διαχειριστές να παρακολουθούν τη ροή κυκλοφορίας και να εντοπίζουν προβλήματα) και το sFlow(παρέχει δειγματοληπτικά δεδομένα για δίκτυα υψηλής ταχύτητας, χρήσιμα για ανάλυση κυκλοφορίας σε πραγματικό χρόνο και παρακολούθηση δικτύων μεγάλης κλίμακας)
- *Δεδομένα πακέτων(Packet Data)*: Περιλαμβάνει επιθεώρηση πακέτων σε βάθος (DPI) όπου αναλύεται το περιεχόμενο κάθε πακέτου [9]. Αυτή η μέθοδος προσφέρει λεπτομερείς πληροφορίες για τη φύση της κίνησης, καθιστώντας δυνατό τον εντοπισμό συγκεκριμένων απειλών ασφαλείας, όπως κακόβουλο λογισμικό ή μη εξουσιοδοτημένη εξαγωγή δεδομένων.

Η ανάλυση NTA [28] παρέχει ολοκληρωμένες πληροφορίες για όλες τις συσκευές που είναι συνδεδεμένες στο δίκτυο, συμπεριλαμβανομένων των συσκευών IoT και λειτουργικής τεχνολογίας. Αυτή η ορατότητα είναι ζωτικής σημασίας για τον εντοπισμό πιθανών κινδύνων ασφαλείας και την αποτελεσματική διαχείριση των

πόρων του δικτύου. Ακόμη, αναλύοντας μοτίβα κυκλοφορίας και χρησιμοποιώντας μηχανική εκμάθηση, τα εργαλεία NTA μπορούν να ανιχνεύσουν απειλές μηδενικής ημέρας και άλλες ανωμαλίες που μπορεί να μην εντοπιστούν από τα παραδοσιακά μέτρα ασφαλείας. Αυτή η προληπτική προσέγγιση βοηθά στον εντοπισμό περίπλοκων επιθέσεων όπως οι προηγμένες επίμονες απειλές (APT). Και τέλος βοηθά τους οργανισμούς να πληρούν τις κανονιστικές απαιτήσεις και να δημιουργούν λεπτομερείς αναφορές δραστηριοτήτων για ελέγχους και έρευνες. Αυτό είναι ιδιαίτερα σημαντικό για βιομηχανίες με αυστηρές εντολές συμμόρφωσης, όπως η υγειονομική περίθαλψη (HIPAA) και η χρηματοδότηση (PCI-DSS).

Περιπτώσεις αξιοποίησης TCP/IP Traffic δεδομένων:

- Anomaly detection: Προσδιορίζει ασυνήθιστα μοτίβα που μπορεί να υποδεικνύουν συμβάντα ασφαλείας, όπως επιθέσεις DDoS ή δραστηριότητες ransomware. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να το βελτιώσουν κάνοντας διάκριση μεταξύ καλοήθων και κακόβουλων ανωμαλιών. Η ανίχνευση μπορεί να είναι βάσει υπογραφών, η οποία χρησιμοποιεί προκαθορισμένες υπογραφές γνωστών απειλών για τον εντοπισμό κακόβουλων δραστηριοτήτων αλλά και βάση ανάλυσης συμπεριφοράς καθορίζοντας μια βασική γραμμή κανονικής συμπεριφοράς δικτύου για τον εντοπισμό αποκλίσεων.
- Network Performance Monitoring: Βοηθά στη διάγνωση και την επίλυση προβλημάτων απόδοσης δικτύου εντοπίζοντας με ακρίβεια τις πηγές αιχμών εύρους ζώνης και τα σημεία συμφόρησης. Αυτό μπορεί να βελτιώσει τη συνολική αποτελεσματικότητα του δικτύου και την εμπειρία χρήστη.
- Εγκληματολογική ανάλυση (Forensic Analysis): Περιλαμβάνει τη συλλογή και την ανάλυση δεδομένων κίνησης για την κατανόηση της φύσης των προηγούμενων επιθέσεων, κάτι που είναι κρίσιμο για τη βελτίωση των μελλοντικών αμυντικών συστημάτων. Αυτό μπορεί να περιλαμβάνει τον εντοπισμό της προέλευσης των επιθέσεων και την κατανόηση των μεθόδων που χρησιμοποιούνται από τους εισβολείς.

Παρά τα πλεονεκτήματά της, η ανάλυση TCP/IP κίνησης αντιμετωπίζει προκλήσεις όπως :

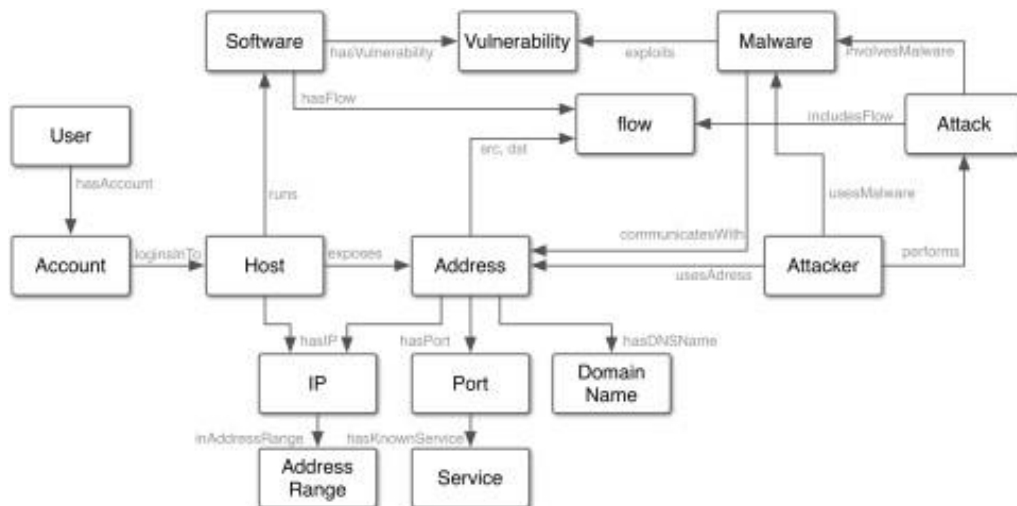
- Κρυπτογραφημένη επισκεψιμότητα(Encrypted Traffic): Η ανάλυση κρυπτογραφημένων πακέτων δεδομένων παραμένει μια σημαντική πρόκληση. Ενώ ορισμένα εργαλεία μπορούν να επιθεωρήσουν μεταδεδωμένα, το περιεχόμενο παραμένει απρόσιτο χωρίς αποκρυπτογράφηση. Οι προηγμένες λύσεις NTA ενδέχεται να χρησιμοποιούν τεχνικές για την ανάλυση μοτίβων και συμπεριφοράς κυκλοφορίας χωρίς να χρειάζεται να αποκρυπτογραφηθεί το περιεχόμενο.
- Όγκος δεδομένων: Ο τεράστιος όγκος δεδομένων που δημιουργείται σε ένα δίκτυο απαιτεί αποτελεσματικές δυνατότητες αποθήκευσης και επεξεργασίας για να αναλυθεί αποτελεσματικά. Αυτό απαιτεί υπολογιστικούς πόρους υψηλής απόδοσης και επεκτάσιμες λύσεις αποθήκευσης και χρήση τεχνολογιών μεγάλων δεδομένων για τη διαχείριση και ανάλυση μεγάλου όγκου δεδομένων κίνησης δικτύου[31].

2.2 Ανάπτυξη της τεχνολογίας των ontologies και η σχέση της με την κυβερνοασφάλεια

Μια **οντολογία** είναι μια ρητή προδιαγραφή μιας εννοιολόγησης, η οποία είναι μια αφηρημένη και απλουστευμένη άποψη του κόσμου που επιθυμούμε να εκπροσωπήσουμε για συγκεκριμένους σκοπούς .

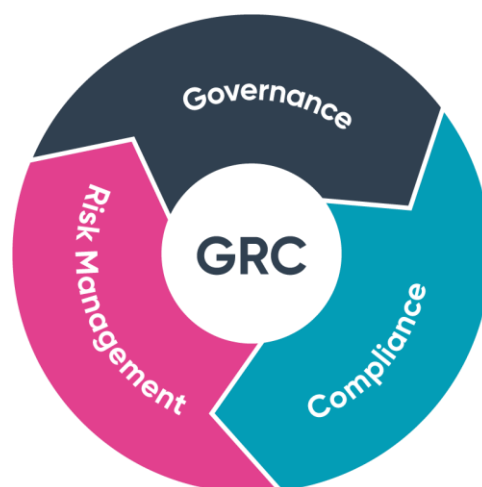
- Δομεί τις πληροφορίες, χρησιμεύει ως βάση μιας αρχιτεκτονικής γνώσης και βοηθά στην ανταλλαγή και την επαναχρησιμοποίηση της γνώσης.
- Μπορεί να παρέχει ένα πλαίσιο για την κοινή χρήση και επαναχρησιμοποίηση τέτοιων πληροφοριών και ορίζει την ορολογία.
- Μπορεί να ενορχηστρώσει τις προδιαγραφές του κλάδου για σχήματα πληροφοριών κυβερνοασφάλειας και να διευκολύνει τη συζήτηση σχετικά με τη δυνατότητα εφαρμογής, την κάλυψη και την αποτελεσματικότητά τους.

Αρκετές οντολογίες έχουν αναπτυχθεί για σκοπούς που σχετίζονται με την ασφάλεια των πληροφοριών. Αν και ήταν καλοσχηματισμένες και μπορούν να προσαρμοστούν με την πάροδο του χρόνου για να αντιπροσωπεύουν καταστάσεις που αλλάζουν ταχέως, κατασκευάστηκαν για διαφορετικά πεδία και στόχους [17].



Εικόνα 3-Παράδειγμα Ontology

Κατά τη δημιουργία ενός ontology, οι οδηγίες για την ασφάλεια στον κυβερνοχώρο είναι χρήσιμες για την κατανόηση διαφόρων πτυχών των επιχειρήσεων. Διάφοροι διεθνείς οργανισμοί προτύπων έχουν δημιουργήσει τέτοιες κατευθυντήριες γραμμές γνωστές ως τομέας GRC.



Εικόνα 4-Στοιχεία GRC

- Το ISO/IEC 27032 παρέχει [12] οδηγίες για κυβερνοασφάλεια: περιγράφει διάφορες έννοιες κυβερνοασφάλειας ,τεχνικούς ελέγχους και παρέχει οδηγίες για πληροφορίες κοινής χρήσης και συντονισμού.
- Η Σύσταση ITU-T E.409 περιγράφει λειτουργίες χειρισμού συμβάντων, ενώ η Σύσταση X.1500 παρέχει μια επισκόπηση ανταλλαγή πληροφοριών στον κυβερνοχώρο.
- Το IETF Request Comments 2350 περιγράφει τις γενικές προσδοκίες του Computer Security Internet Response Teams που εκτελούν περιστατικά επιχειρήσεων απόκρισης.
- Οι Ειδικές Εκδόσεις του NIST στη σειρά 800 είναι ευρείες κατευθυντήριες γραμμές για την ασφάλεια στον κυβερνοχώρο, π.χ. σε βασικές ιδέες για ασφάλεια υπολογιστών , υπηρεσίες ασφαλείας , χειρισμός συμβάντων , εγκληματολογία , δοκιμές και μετρήσεις .

Η ευρεία εξάπλωση του Διαδικτύου ενισχύει την ανάπτυξη μιας **κοινωνίας του κυβερνοχώρου**, στην οποία ποικίλλουν επικοινωνίες , συμπεριλαμβανομένης της ανταλλαγής προσωπικών πληροφοριών και επιχειρηματικών συναλλαγών που πραγματοποιούνται.

- Η αύξηση επίσης του αριθμού των απειλών στον κυβερνοχώρο έχει διαφοροποιήσει τους στόχους . Οι στόχοι ποικίλλουν από άτομα σε ιδιωτικές εταιρείες και μάλιστα κρίσιμες υποδομές όπως ως πυρηνικοί σταθμοί ηλεκτροπαραγωγή, ενώ αποσκοπούν σε χρηματικό όφελος για πολιτικές ενέργειες.
- Αντίστοιχα, η ανάγκη για τις επιχειρήσεις κυβερνοασφάλειας αυξάνεται προκειμένου να μετριάσει αυτές τις απειλές. Σε μια κοινωνία του κυβερνοχώρου, κακόβουλο λογισμικό όπως οι ιοί μπορεί να επιτεθεί σε οποιοδήποτε υπολογιστή και συσκευή πέρα από τα σύνορα της χώρας καταγωγής ή του στόχου του και ένας εισβολέας μπορεί να επιτεθεί σε υπολογιστές σε όλο τον κόσμο τρέχοντας προσυσκευασμένο (pre-packaged) λογισμικό επίθεσης άλλων hacker. Πηγές των απειλών διασχίζουν τα σύνορα χωρών και ακόμη και ηπείρων, και ένας εισβολέας μπορεί να επιτεθεί σε

υπολογιστές στη χώρα Α ελέγχοντας υπολογιστές στη χώρα Β ενώ διαμένουν φυσικά στη χώρα Γ.

- Επιπλέον, η ευπάθεια ενός συστήματος μπορεί να εκτεθεί σε επιτιθέμενους σε όλο τον κόσμο. Ωστόσο, τα αντίμετρα ενάντια σε αυτές τις απειλές για την ασφάλεια στον κυβερνοχώρο εφαρμόζονται συχνότερα από μεμονωμένες οργανώσεις. Κατά συνέπεια, ένας οργανισμός σε μια χώρα μπορεί να δεχθεί επίθεση από κακόβουλο λογισμικό του οποίου τα αντίμετρα είναι ήδη γνωστά και εφαρμόζονται αλλού. Τέτοια περιστατικά συμβαίνουν λόγω έλλειψης ενημέρωσης και ανταλλαγών πληροφοριών μεταξύ οργανισμών. Αν και κάποιοι ατομικοί φορείς εκμετάλλευσης κυβερνοασφάλειας ανταλλάσσουν πληροφορίες τοπικά, οι κύριες μέθοδοι εξακολουθούν να είναι το ηλεκτρονικό ταχυδρομείο, οι τηλεφωνικές κλήσεις, ακόμη και οι συναντήσεις πρόσωπο με πρόσωπο, οι οποίες δεν είναι αποτελεσματικές.

Για την αντιμετώπιση αυτού του ζητήματος, έχουν ξεκινήσει διάφορες οργανώσεις τη δημιουργία μορφών πληροφοριών για την ανταλλαγή πληροφοριών πέραν των συνόρων οργάνωσης. Αυτές είναι χρήσιμες για την ανταλλαγή πληροφοριών για συγκεκριμένους σκοπούς, και οι επιχειρήσεις και όχι μόνο μπορούν να ανταλλάσσουν πληροφορίες σε ένα συγκεκριμένο σχήμα που έχουν συμφωνήσει να χρησιμοποιήσουν πριν από την ανταλλαγή. Παρόλα αυτά, αυτό είναι δύσκολο για αυτούς να ανταλλάξουν πληροφορίες σε άλλα σχήματα. Επιπλέον, μπορεί να μην βρουν ένα κατάλληλο σχήμα για ανταλλαγή πληροφοριών δεδομένου ότι οι υπάρχουσες προδιαγραφές ενδέχεται να μην καλύπτουν ένα επαρκές φάσμα τύπων πληροφοριών και περιπτώσεις χρήσης.

Έτσι, η ανταλλαγή πληροφοριών στον κυβερνοχώρο μεταξύ οργανισμών και η αυτοματοποίησή τους εξακολουθεί να αντιμετωπίζει δυσκολίες στην πραγματικότητα. Επί του παρόντος, δεν υπάρχει βάση για τον καθορισμό της εφαρμογής τους, της κάλυψης και της αποτελεσματικότητας. Πρέπει να υπάρχει μια γενική άποψη για το τι είδη πληροφοριών απαιτούνται και θα πρέπει να ανταλλάσσονται για την διατήρηση της κυβερνοασφάλειας. Για να αντιμετωπίσουμε αυτό το ζήτημα, ακολουθούμε μια

προσέγγιση που εξετάζει ποιος χρησιμοποιεί ποια είδη πληροφοριών για ποιους σκοπούς και δημιουργείται μια οντολογία επιχειρησιακών πληροφοριών για την ασφάλεια στον κυβερνοχώρο.

Γενικά, στα πληροφοριακά συστήματα οι οντολογίες χρησιμοποιούνται κυρίως για τη λήψη και αναπαράσταση πληροφοριών, ανταλλαγή και διαχείριση γνώσης. Ακόμη, οι **οντολογίες ασφάλειας πληροφοριών** διαχωρίζονται σε γενικές οντολογίες ασφάλειας [34] που περιλαμβάνουν όλες (ή τις περισσότερες) από τις έννοιες ασφάλειας και συγκεκριμένες οντολογίες ασφάλειας που σχετίζονται με το μεμονωμένο τμήμα του τομέα ασφάλειας πληροφοριών. Ο στόχος των οντολογιών ασφάλειας είναι να δημιουργήσουν κοινά, μονοσήμαντα σημασιολογικά μοντέλα εννοιών στον τομέα της ασφάλειας που θα χρησιμεύσουν ως βάση για την επικοινωνία μεταξύ ανθρώπων ή πρακτόρων λογισμικού οδηγώντας σε μείωση της γλωσσικής ασάφειας.

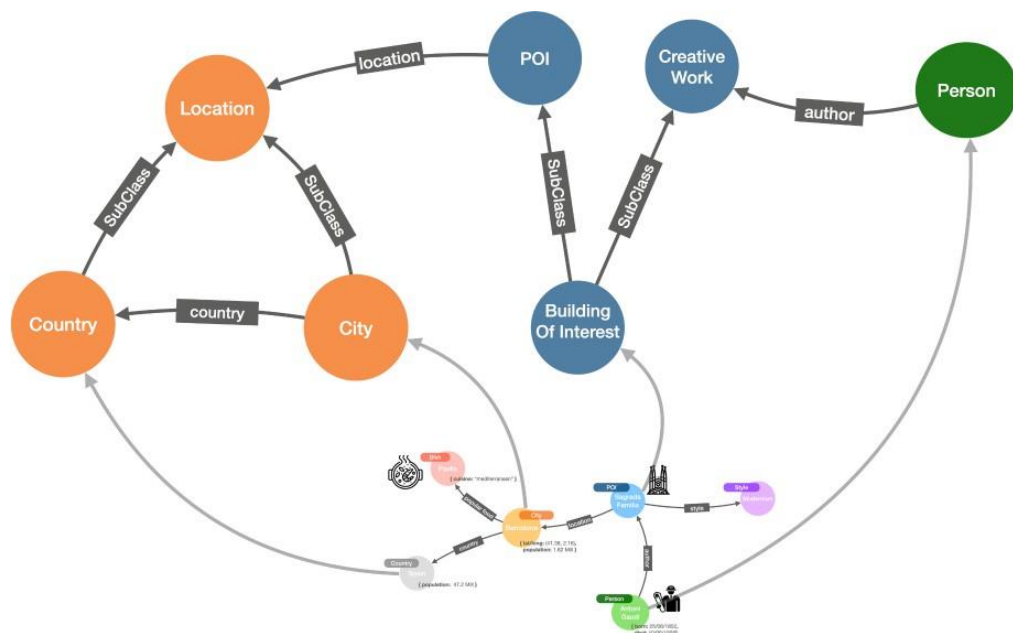
Με βάση τα αποτελέσματα ερευνητών η ταξινόμηση της οντολογίας ασφαλείας αποτελείται από οκτώ ομάδες: αρχικές οντολογίες ασφάλειας, ταξινομίες ασφαλείας, γενικές οντολογίες ασφάλειας, ειδικές οντολογίες ασφάλειας, οντολογίες ασφάλειας που βασίζονται σε κινδύνους, οντολογίες ασφάλειας προσανατολισμένες στον Παγκόσμιο Ιστό, οντολογίες απαιτήσεων ασφαλείας και μοντελοποίηση οντολογιών ασφαλείας. Χρησιμοποιούνται οντολογίες από αυτές τις κατηγορίες σε συνδυασμό με δεδομένα για την κατασκευή ενός πλήρους γραφήματος γνώσης με το οποίο θα αναπαρασταθούν οι οντότητες και οι σχέσεις που επισημαίνονται στο ontology σε γραφική απεικόνιση.

2.3 Θεωρητικό υπόβαθρο των Knowledge graphs

Αν και η έννοια «knowledge graph» χρησιμοποιείται στη βιβλιογραφία τουλάχιστον από το 1972, η σύγχρονη ενσάρκωση της φράσης προέρχεται από την ανακοίνωση του 2012 της Google Knowledge Graph ακολουθούμενη από περαιτέρω ανακοινώσεις γραφημάτων γνώσης από Airbnb, Amazon, eBay, Facebook, IBM, LinkedIn, Microsoft, Uber και άλλα. Η αυξανόμενη βιομηχανική υιοθέτηση της έννοιας αποδείχθηκε δύσκολο να αγνοηθεί από τον ακαδημαϊκό κόσμο, με ολοένα και περισσότερη επιστημονική βιβλιογραφία να δημοσιεύεται σε γραφήματα γνώσης τα τελευταία

χρόνια . Τα γραφήματα γνώσης χρησιμοποιούν ένα μοντέλο δεδομένων που βασίζεται σε γράφημα για να συλλάβουν τη γνώση σε σενάρια εφαρμογών που περιλαμβάνουν ενσωμάτωση, διαχείριση και εξαγωγή αξίας από διαφορετικές πηγές δεδομένων σε μεγάλη κλίμακα. Τα γραφήματα παρέχουν μια συνοπτική και διαισθητική εικόνα για μια ποικιλία τομέων, όπου οι ακμές και τα μονοπάτια αποτυπώνουν διαφορετικά, δυνητικά πολύπλοκες σχέσεις μεταξύ των οντοτήτων ενός τομέα .

Τα γραφήματα γνώσης συχνά συγχωνεύουν δεδομένα από διαφορετικές πηγές και μορφές. Οι οντολογίες λειτουργούν ως η γέφυρα που συνδέει αυτά τα ανόμοια στοιχεία, διευκολύνοντας την ενσωμάτωσή τους με συνεκτικό και συνεπή τρόπο. Οι γλώσσες ερωτημάτων γραφήματος υποστηρίζουν όχι μόνο τυπικούς σχεσιακούς τελεστές (ενώσεις, ενώσεις, προβολές κ.λπ.), αλλά και τελεστές πλοήγησης για την εύρεση οντοτήτων που συνδέονται μέσω μονοπατιών αυθαίρετου μήκους. Οντολογίες και κανόνες μπορούν να χρησιμοποιηθούν για να ορίσουν και να αιτιολογήσουν τη σημασιολογία των όρων που χρησιμοποιούνται στον γράφο. Τα κλιμακούμενα πλαίσια για την ανάλυση γραφημάτων μπορούν να αξιοποιηθούν για τον υπολογισμό της κεντρικότητας, της ομαδοποίησης, της σύνοψης και ούτω καθεξής, για να αποκτηθούν γνώσεις σχετικά με τον τομέα που περιγράφεται.



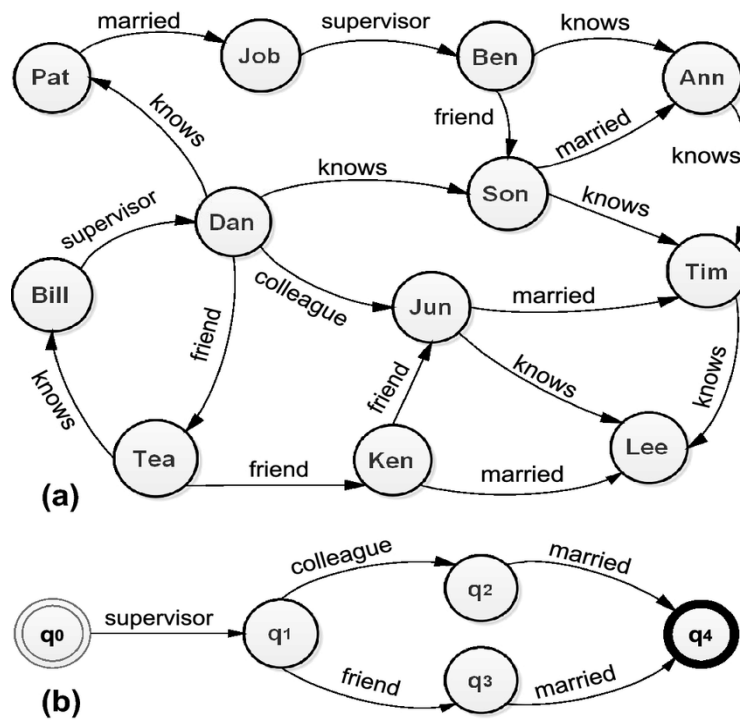
Εικόνα 5-Παράδειγμα γραφήματος γνώσης

Ένα γράφημα σχηματίζεται από κόμβους και σχέσεις. Οποιοδήποτε άτομο, αντικείμενο, τοποθεσία ή συμβάν μπορεί να είναι ένας κόμβος. Οι σχέσεις περιγράφουν

κάθε είδους αλληλεπίδραση μεταξύ κόμβων, για παράδειγμα, ένα συμβάν λαμβάνει χώρα σε μια τοποθεσία, ένα άτομο γνωρίζει ένα άλλο άτομο, κ.λπ. Οι κόμβοι και οι σχέσεις θα σχολιάζονται με τους τύπους τους και θα περιγράφονται από μια συλλογή χαρακτηριστικών που τους χαρακτηρίζουν. Για παράδειγμα, ένας κόμβος που αντιπροσωπεύει μια πόλη θα έχει συνήθως μια ιδιότητα που υποδεικνύει τον τρέχοντα πληθυσμό της ή τη γεωγραφική της θέση, ένα άτομο θα έχει ημερομηνία γέννησης, όνομα και ούτω καθεξής. Από τη στιγμή που τα δεδομένα αναπαρίστανται με συνδεδεμένο τρόπο, η εξερεύνησή τους είναι εύκολη και το πλαίσιο γύρω από ένα θέμα ενδιαφέροντος είναι απλώς η γειτονιά γύρω από αυτό στο γράφημα[22]. Υπάρχουν διάφορες βάσεις δεδομένων γραφημάτων που κατηγοριοποιούνται με βάση τα υποκείμενα μοντέλα δεδομένων γραφημάτων, μεταξύ των οποίων Πλαίσιο Περιγραφής Πόρων με τριπλή αποθήκευση (RDF Triple Store) και Γραφήματα ιδιοκτησίας με ετικέτα (Labeled Property Graphs). Ο σκοπός τόσο των αποθηκών RDF όσο και των γραφημάτων ιδιοκτησίας είναι να αποθηκεύουν γραφικά δομημένα δεδομένα και να προσφέρουν διαφορετικούς τρόπους πλοήγησης στα δεδομένα αυτά.

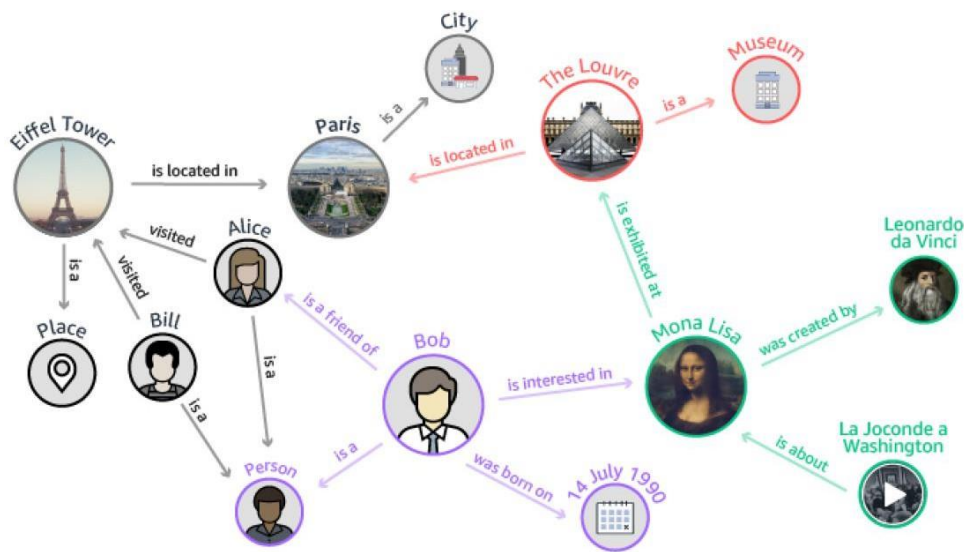
Στην πράξη συναντάμε διάφορα μοντέλα γράφων ιδιοκτησίας:

- **Directed Edge-labelled Graphs:** Ένα κατευθυνόμενο γράφημα με ετικέτα άκρων ή για συντομία γραφικό del(γνωστό και ως πολυσχεσιακό γράφημα) ορίζεται ως ένα σύνολο κόμβων και ένα σύνολο κατευθυνόμενων σημειωμένων ακμών μεταξύ αυτών των κόμβων. Στα γραφήματα γνώσης, οι κόμβοι αντιπροσωπεύουν οντότητες και οι ακμές αντιπροσωπεύουν δυαδικές σχέσεις μεταξύ αυτών ,οντότητες (π.χ. η Santa Lucía βρίσκεται στην πόλη Σαντιάγο). Η μοντελοποίηση δεδομένων με αυτόν τον τρόπο προσφέρει μεγαλύτερη ευελιξία για την ενοποίηση νέων πηγών δεδομένων, σε σύγκριση στο τυπικό σχεσιακό μοντέλο, όπου ένα σχήμα πρέπει να ορίζεται εκ των προτέρων και να ακολουθείται σε κάθε βήμα. Ενώ άλλα μοντέλα δομημένων δεδομένων, όπως δέντρα (XML, JSON, κ.λπ.) θα προσφέρουν παρόμοια ευελιξία, τα γραφήματα δεν απαιτούν ιεραρχική οργάνωση των δεδομένων (θα πρέπει να είναι γονέας, παιδί ή αδερφό τύπου, για παράδειγμα;). Επιτρέπουν επίσης την αναπαράσταση κύκλων και την υποβολή ερωτημάτων.



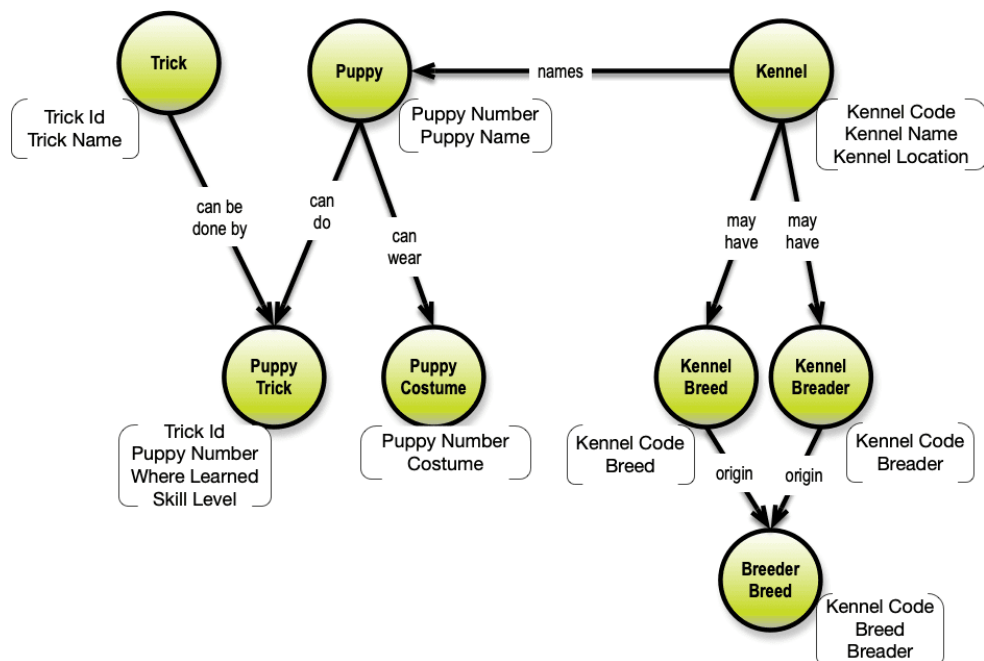
Εικόνα 6-Παράδειγμα DEL Graph

- Heterogeneous Graphs:** Ένα ετερογενές γράφημα [18] (ή ετερογενές δίκτυο πληροφοριών) είναι ένα γράφημα όπου σε κάθε κόμβο και ακμή εκχωρείται ένας τύπος δεδομένων. Τα ετερογενή γραφήματα είναι παρόμοια με τα γραφήματα del - με ετικέτες ακμών που αντιστοιχούν σε τύπους ακμών - αλλά όπου ο τύπος του κόμβου αποτελεί μέρος του ίδιου του μοντέλου γραφήματος, αντί να εκφράζεται ως ειδική σχέση. Μια ακμή ονομάζεται ομοιογενής εάν βρίσκεται μεταξύ δύο κόμβων του ίδιου τύπου (π.χ. περιγράμματα)· αλλιώς λέγεται ετερογενής (π.χ. κεφαλαίο). Τα ετερογενή γραφήματα επιτρέπουν τον διαμερισμό κόμβων ανάλογα με τον τύπο τους, για παράδειγμα, για σκοπούς μηχανικής εκμάθησης. Ωστόσο, σε αντίθεση με τα γραφήματα del, συνήθως υποθέτουν μια σχέση ένα προς ένα μεταξύ κόμβων και τύπων [30].



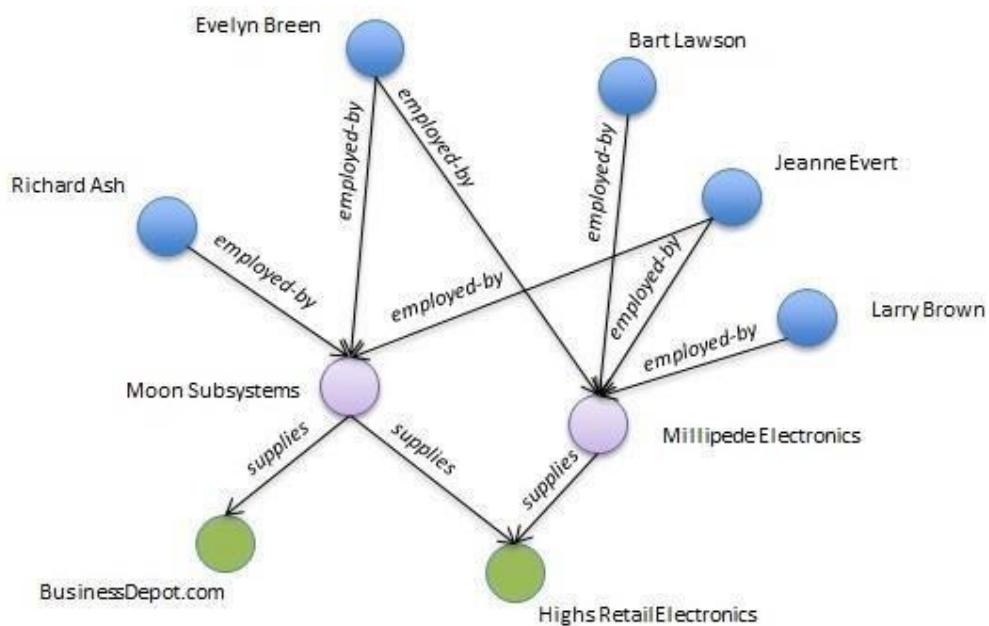
Εικόνα 7-Παράδειγμα Heterogeneous Graph

- Property Graphs:** Ένα γράφημα ιδιοτήτων επιτρέπει ένα σύνολο ζευγών ιδιοτήτων-τιμής και μια ετικέτα να συσχετίζονται με κόμβους και ακμές, προσφέροντας πρόσθετη ευελιξία κατά τη μοντελοποίηση δεδομένων. Στην προσπάθεια, για παράδειγμα, να μοντελοποιηθούν οι αεροπορικές εταιρείες που προσφέρουν πτήσεις. Σε ένα γράφημα *del*, δεν μπορούμε να σχολιάσουμε απευθείας ένα πλεονέκτημα όπως η πτήση Arica στο Σαντιάγο με την εταιρεία, αλλά θα μπορούσαμε να προσθέσουμε έναν νέο κόμβο δηλώνοντας μια πτήση και συνδέοντας την με την πηγή, τον προορισμό, τις εταιρείες και τη λειτουργία. Η εφαρμογή αυτού του μοτίβου σε ένα μεγάλο γράφημα ενδέχεται να απαιτεί σημαντικές αλλαγές. Αν και δεν έχουν ακόμη τυποποιηθεί, χρησιμοποιούνται γραφήματα ιδιοτήτων σε δημοφιλείς βάσεις δεδομένων γραφημάτων, όπως το Neo4j.



Εικόνα 8-Παράδειγμα Property Graph

- Graph Dataset:** Ένα σύνολο δεδομένων γραφήματος επιτρέπει τη διαχείριση πολλών γραφημάτων και αποτελείται από ένα σύνολο ονομασμένων γραφημάτων και ένα προεπιλεγμένο γράφημα. Κάθε γράφημα με όνομα είναι ένα ζεύγος αναγνωριστικού γραφήματος(graph ID) και γραφήματος. Το προεπιλεγμένο γράφημα είναι ένα γράφημα χωρίς id και αναφέρεται "από προεπιλογή" εάν δεν έχει καθοριστεί ένα αναγνωριστικό γραφήματος. Το Σχήμα 4 παρέχει ένα παράδειγμα όπου τα συμβάντα και οι διαδρομές αποθηκεύονται σε δύο επώνυμα γραφήματα και το προεπιλεγμένο γράφημα διαχειρίζεται τα metadata σχετικά με τα γραφήματα που αναφέρονται. Αν και το παράδειγμα χρησιμοποιεί γραφήματα del, τα σύνολα δεδομένων γραφημάτων μπορούν να γενικευθούν σε άλλους τύπους γραφημάτων. Τα σύνολα δεδομένων γραφημάτων είναι χρήσιμα για τη διαχείριση και αναζήτηση δεδομένων από πολλαπλές πηγές, όπου κάθε πηγή μπορεί να διαχειρίζεται ξεχωριστά γραφήματα, επιτρέποντας σε μεμονωμένα γραφήματα να υποβάλλονται ερωτήματα, να ενημερώνονται, να αφαιρούνται και ούτω καθεξής, ανάλογα με τις ανάγκες.



Εικόνα 9-Παράδειγμα Graph Dataset

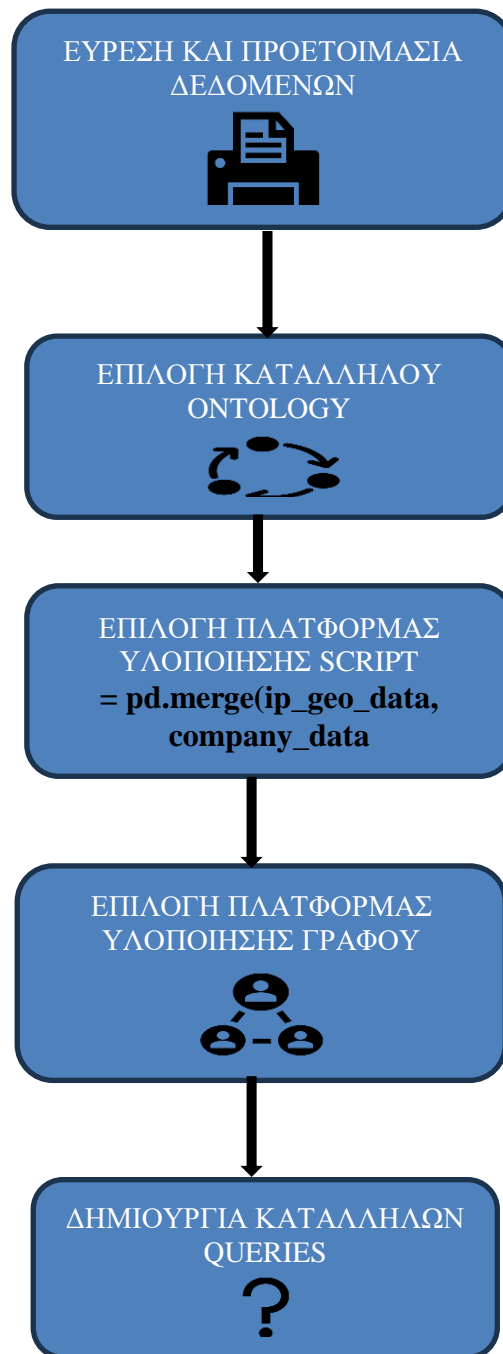
Για τα RDF Triple Store Γραφήματα:

- Από τη δεκαετία του '90, κυκλοφορεί η ιδέα [29] μιας σημασιολογικής παρακολούθησης ιστού των ιστοσελίδων και των συνδέσμων. Ενώ ο σημασιολογικός ιστός δεν έχει εφαρμοστεί ακόμα, η ιδέα πίσω από αυτόν παρέμεινε και έγινε το θεμέλιο του RDF Triple Stores. Το RDF σημαίνει Resource Description Framework, ένα πρότυπο World Wide Web Consortium (W3C) που δημιουργήθηκε αρχικά για τη μοντελοποίηση μεταδεδομένων (metadata). Τα Triple Stores αποθηκεύουν και εκφράζουν πληροφορίες σε μια δομή τριών προτάσεων. Υποκείμενο-Κατηγορούμενο-Αντικείμενο, που συμβολίζεται με δύο κόμβους που συνδέονται με ένα μόνο άκρο. Για π.χ. Στον Τζος αρέσει το ψωμί. Αυτές οι πληροφορίες θα δομηθούν ως τριπλέτα Josh-likes-bread, όπου το υποκείμενο στο Josh, το κατηγορήμα είναι τα likes και το αντικείμενο στο ψωμί.
- Το Υποκείμενο και το Αντικείμενο απεικονίζονται από δύο κόμβους, αρχίζοντας και τελειώνοντας, που αντιπροσωπεύουν οντότητες δεδομένων. Το κατηγορούμενο απεικονίζεται από μια ακμή που συνδέει τους κόμβους υποκειμένου και αντικειμένου, αντιπροσωπεύοντας τη σχέση μεταξύ

οντοτήτων υποκειμένου και αντικειμένου. Το RDF Triple Stores είναι ένα τυποποιημένο μοντέλο όπου κάθε στοιχείο έχει ένα Uniform Resource Identifier (URI), που επιτρέπει στις μηχανές να αναγνωρίζουν κάθε θέμα, κατηγορία και αντικείμενο μοναδικά. Το RDF Triple Stores χρησιμοποιεί μια τυπική γλώσσα ερωτημάτων, τη SPARQL, για την εξαγωγή πληροφοριών από τη βάση δεδομένων. Οι τυπικές μέθοδοι αναπαράστασης και αναζήτησης δεδομένων επιτρέπουν στα RDF Triple Stores να είναι διαλειτουργικά με οποιαδήποτε άλλα γραφήματα γνώσης που βασίζονται σε RDF.

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΠΡΟΒΛΗΜΑΤΟΣ-ΣΧΕΔΙΑΣΜΟΣ

Ο σχεδιασμός του προβλήματος που υλοποιήθηκε φαίνεται στο παρακάτω σχήμα:



Εικόνα 10-Σχεδιασμός-Διάγραμμα ροής του προβλήματος

3.1 Περιγραφή των δεδομένων που χρησιμοποιήθηκαν

Σε αυτό το κεφάλαιο, παρέχεται μια λεπτομερή περιγραφή των διαφόρων συνόλων δεδομένων που χρησιμοποιήθηκαν σε αυτή τη μελέτη. Αρχικά έγινε αναζήτηση για έτοιμες συλλογές δεδομένων οι οποίες θα περιλάμβαναν στοιχεία σχετικά με ονόματα εταιρειών, IP διευθύνσεις τους, περιοχές που βρίσκονται καθώς και έτοιμα δεδομένα κίνησης TCP/IP. Η εύρεση κατάλληλων τέτοιων στοιχείων δεδομένων τα οποία θα μπορούν να συνδυαστούν σε έναν κοινό γράφημα γνώσης αποκόπτοντάς μόνο τις περιττές πληροφορίες δεν βρέθηκαν. Ακόμη ένα άλλο πρόβλημα κατά την εύρεση των δεδομένων αυτών ήταν ο τεράστιος αριθμός δεδομένων κίνησης που βρισκόταν διαθέσιμος δωρεάν καθώς δεν γινόταν να επεξεργαστεί ολόκληρος ή ένα κομμάτι αυτού σε έναν γράφημα. Έτσι τελικά για τα δεδομένα της εργασίας δημιουργήθηκαν 3 αρχεία csv σε συνδυασμό με πληροφορίες από open-source αναζητήσεις αλλά και ρεαλιστικά σενάρια, τα οποία περιγράφουν την δομή της εταιρείας σε τμήματα αλλά και τα τμήματα άλλων εταιρειών που πιθανόν επικοινωνούν σε μορφή ρεαλιστικής κίνησης TCP/IP.

Αυτά τα σύνολα δεδομένων περιλαμβάνουν δεδομένα εταιρειών (company_data.csv), δεδομένα γεωγραφικής τοποθεσίας IP (ip_geolocation_data.csv) και δεδομένα κίνησης TCP/IP (tcp_traffic_data.csv). Κάθε σύνολο δεδομένων συμβάλλει μοναδικά στη δημιουργία ενός ολοκληρωμένου γραφήματος γνώσης που στοχεύει στην ενίσχυση της ασφάλειας στον κυβερνοχώρο μέσω της ανάλυσης δεδομένων δικτύου.

1. Εταιρικά δεδομένα

Τα εταιρικά δεδομένα παρέχουν πληροφορίες σχετικά με τα ονόματα των εταιρειών, τα ονόματα των τμημάτων των εταιρειών καθώς και την τοποθεσία αυτών. Αυτά τα δεδομένα είναι ζωτικής σημασίας για τον εντοπισμό των οντοτήτων που εμπλέκονται σε αλληλεπιδράσεις δικτύου και την κατανόηση του πιθανού αντίκτυπού τους στην ασφάλεια στον κυβερνοχώρο.

Πηγή και Μορφή δεδομένων:

- Πηγή: Τα δεδομένα προέρχονται από αξιόπιστους παρόχους επιχειρηματικών πληροφοριών.
- Μορφή: Τα δεδομένα αποθηκεύονται σε ένα αρχείο CSV με το όνομα `company_data.csv`

Βασικά Χαρακτηριστικά:

- `company_name`: Περιγράφουν το όνομα της εταιρείας
- `department`: Το τμήμα της εταιρείας
- `department_location`: Η τοποθεσία του εκάστοτε τμήματος
- `department_lan_address`: Η διεύθυνση LAN που ανήκει το τμήμα της εταιρείας

```
company_name,department,department_location,department_lan_address
TechCorp,TechCorpIT,New York,192.168.0.1
Japanios,JapaniosHR,Japan,198.51.100.1
Brazilero,BrazileroSales,São Paulo,100.64.0.1
BizInc,BizIncFinance,Austin,192.168.1.4
WebSol,WebSolMarketing,Paris,198.18.0.2
Fage,FageHeadquarters,Kifissia,102.38.248.50
```

Εικόνα 11-Παράδειγμα εταιρικών δεδομένων

2. Δεδομένα Γεωγραφικής θέσης IP

Τα δεδομένα γεωγραφικής θέσης IP αντιστοιχούν διευθύνσεις IP σε φυσικές τοποθεσίες. Αυτές οι πληροφορίες είναι απαραίτητες για τον εντοπισμό της προέλευσης της κυκλοφορίας δικτύου και τον εντοπισμό πιθανών γεωγραφικών προτύπων σε απειλές στον κυβερνοχώρο.

Πηγή και Μορφή δεδομένων:

- Πηγή: Αυτά τα δεδομένα προέρχονται από παρόχους υπηρεσιών γεωγραφικής τοποθεσίας.
- Μορφή: Τα δεδομένα αποθηκεύονται σε ένα αρχείο CSV που ονομάζεται `ip_geolocation_data.csv`.

Βασικά Χαρακτηριστικά:

- **IPRangeStart:** Αυτή είναι η πρώτη διεύθυνση IP σε ένα δεδομένο εύρος διευθύνσεων. Βοηθά στον καθορισμό ενός μπλοκ διευθύνσεων IP που ανήκουν στην ίδια γεωγραφική τοποθεσία.
- **IPRangeEnd:** Αυτή είναι η τελευταία διεύθυνση IP σε ένα δεδομένο εύρος διευθύνσεων. Μαζί με το **IPRangeStart**, ορίζει το εύρος του μπλοκ διευθύνσεων IP.
- **Country:** Υποδεικνύει την χώρα που σχετίζεται με το εύρος διευθύνσεων IP, το οποίο είναι κρίσιμο για τον προσδιορισμό της γεωγραφικής κατανομής της κυκλοφορίας δικτύου.
- **Region:** Παρέχει μια πιο λεπτομερή γεωγραφική τοποθεσία σε μια χώρα, όπως μια περιοχή, μια επαρχία ή έναν νομό.
- **City:** Προσφέρει ακόμη πιο λεπτομερείς γεωγραφικές πληροφορίες, προσδιορίζοντας την ακριβή πόλη όπου βρίσκεται η διεύθυνση IP.
- **Latitude:** Αντιπροσωπεύει τη θέση βορρά-νότου της τοποθεσίας, καθοριστικής σημασίας για τη χαρτογράφηση ενός σημείου οπουδήποτε στον κόσμο.
- **Longitude:** Αντιπροσωπεύει τη θέση ανατολής-δύσης της τοποθεσίας, που χρησιμοποιείται σε συνδυασμό με το γεωγραφικό πλάτος(latitude) για τον προσδιορισμό της ακριβούς γεωγραφικής θέσης.
- **ISP:** Προσδιορίζει τον ISP που είναι υπεύθυνος για τις διευθύνσεις IP, παρέχοντας πληροφορίες για την υποδομή δικτύου και τους πιθανούς παρόχους υπηρεσιών που εμπλέκονται.

```
IPRangeStart,IPRangeEnd,Country,Region,City,Latitude,Longitude,ISP  
102.38.248.65,102.38.248.100,Greece,Patra,Achaia,35.65107,96.347015,ISP2
```

Εικόνα 12-Παράδειγμα δεδομένων γεωγραφικής θέσης IP

3. Δεδομένα TCP/IP κίνησης

Τα δεδομένα TCP/IP κίνησης καταγράφουν τη ροή των πακέτων δεδομένων σε ένα δίκτυο. Η ανάλυση αυτών των δεδομένων βοηθά στην κατανόηση των προτύπων επικοινωνίας και στον εντοπισμό πιθανών ανωμαλιών ή απειλών.

Πηγή και Μορφή δεδομένων:

- Πηγή: Τα δεδομένα συλλέγονται από εργαλεία παρακολούθησης δικτύου και αρχεία καταγραφής.
- Μορφή: Τα δεδομένα αποθηκεύονται σε ένα αρχείο CSV που ονομάζεται `tcp_traffic_data.csv`.

Βασικά Χαρακτηριστικά:

- `timestamp`: Η ημερομηνία και η ώρα που καταγράφηκε το πακέτο δεδομένων.
- `connection_id`: Αναγνωριστικό που τοποθετείται σε κάθε σύνδεση για την διαφοροποίηση της.
- `source_ip`: Η διεύθυνση IP από την οποία προήλθε το πακέτο δεδομένων.
- `destination_ip`: Η διεύθυνση IP στην οποία αποστέλλεται το πακέτο δεδομένων.
- `source_port`: Ο αριθμός θύρας στη συσκευή προέλευσης.
- `destination_port`: Ο αριθμός θύρας στη συσκευή προορισμού.
- `protocol`: Το πρωτόκολλο που χρησιμοποιείται για την επικοινωνία (π.χ. TCP, UDP).
- `size`: Το μέγεθος του πακέτου δεδομένων σε byte.

```
timestamp,connection_id,source_ip,destination_ip,source_port,destination_port,protocol,size
2023-05-01
10:00:00,3ace4f4ea,192.168.0.1,102.38.248.50,443,123,HTTPS,128
2023-05-01
10:05:00,974ec5991,102.38.248.54,102.38.248.50,23,80,HTTP,256
2023-05-01
10:10:00,6b53d,102.38.248.50,100.64.0.1,443,80,HTTPS,512
2023-05-01
10:15:00,1b439c9,198.51.100.2,192.168.1.3,53,161,DNS,1024
2023-05-01
```

Εικόνα 13-Παράδειγμα δεδομένων TCP/IP Traffic

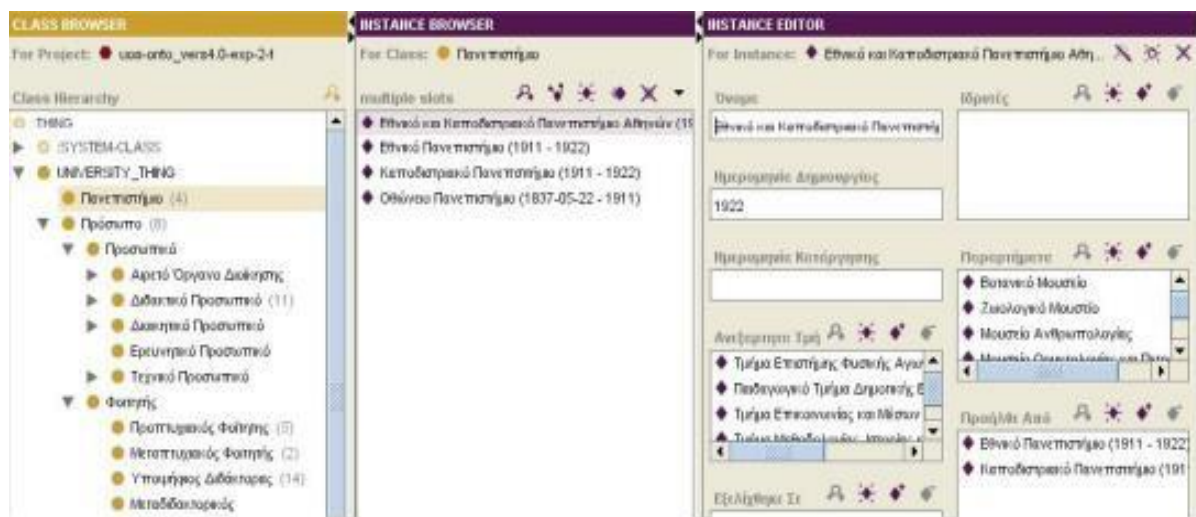
3.2 Ανάπτυξη του ontology

Στην συνέχεια απαιτείται η εύρεση ή δημιουργία μιας οντολογίας η οποία σε συνδυασμό με τα δεδομένα θα καθορίσουν το γράφημα γνώσης. Για την ανάπτυξη του Ontology αρχικά έγινε προσπάθεια της υλοποίησης της εργασίας με ένα έτοιμο ontology το οποίο θα βρισκόταν σε κάποια online διαθέσιμη βάση δεδομένων όπως το UCO [13]. Το έτοιμο αυτό ontology θα βοηθούσε την εργασία καθώς θα υπήρχαν δεδομένα τα οποία θα αντιστοιχούσαν εύκολα σε αυτό και γενικώς πιο ακριβής ανάλυση. Ωστόσο η προσπάθεια αυτή δεν απέφερε αποτέλεσμα καθώς τα έτοιμα αυτά ontologies συνήθως είχαν μια πληθώρα δεδομένων, οντοτήτων και σχέσεων τα οποία δεν χρησίμευαν και κυρίως δεν γινόταν εύκολα ο περιορισμός τους. Έτσι τελικά θα δημιουργηθεί ένα ontology από την αρχή το οποίο θα είναι σχετικό με τα δεδομένα και την δομή που θα επεξεργαστεί η εργασία.

Υπάρχουν αρκετά εργαλεία δημιουργίας ενός Ontology όπως το Fluent Editor, το NeOn Toolkit και το Ontofly επιλέχθηκε το Protégé.

Το *Protégé* είναι ένα πολύ δημοφιλές εργαλείο μοντελοποίησης γνώσης που αναπτύχθηκε στο Πανεπιστήμιο του Stanford. Οι οντολογίες και οι βάσεις γνώσεων

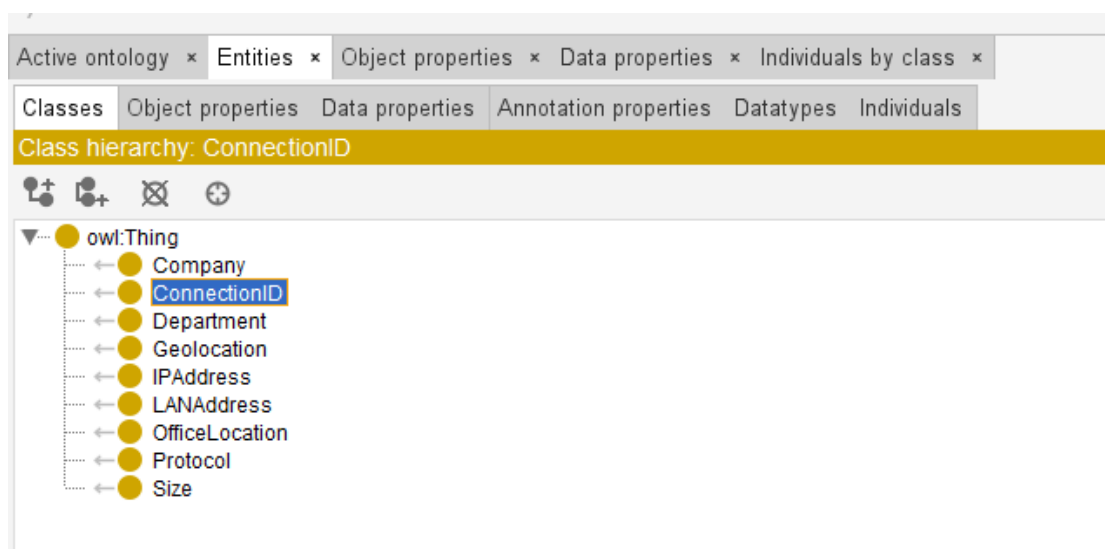
μπορούν να επεξεργαστούν διαδραστικά στο Protégé παρέχοντας πρόσβαση με γραφικό περιβάλλον χρήστη και Java API. Το Protégé μπορεί να επεκταθεί με βυσματικά στοιχεία (pluggable components) για την προσθήκη νέων λειτουργιών και υπηρεσιών. Υπάρχει ένας αυξανόμενος αριθμός plugins που προσφέρουν μια ποικιλία πρόσθετων λειτουργιών, όπως πρόσθετα εργαλεία διαχείρισης οντολογίας, υποστήριξη πολυμέσων, μηχανές αναζήτησης και συλλογισμού, μέθοδοι επίλυσης προβλημάτων κ.λπ. Το Protégé εφαρμόζει ένα πλούσιο σύνολο δομών και δράσεων μοντελοποίησης γνώσης που υποστηρίζουν τη δημιουργία, οπτικοποίηση και τον χειρισμό οντολογιών σε διάφορες μορφές αναπαράστασης. Προσφέρει υποστήριξη για την κατασκευή των οντολογιών που βασίζονται σε πλαίσιο, σύμφωνα με το Ανοιχτό Πρωτόκολλο Συνδεσιμότητας Γνωσιακής Βάσης. Η εκτεταμένη έκδοση του συστήματος που βασίζεται σε πλαίσιο εισήχθη το 2003 για να υποστηρίξει το OWL με ένα πλεονέκτημα της σημασιολογικής έκδοσης Ιστού (semantic web version). Υπάρχουν διάφορες μορφές, όπως RDF(s), OWL και XML Schema στις οποίες η προστατευόμενη οντολογία μπορεί να εξάχθει.



Εικόνα 14-Protege Class Browser

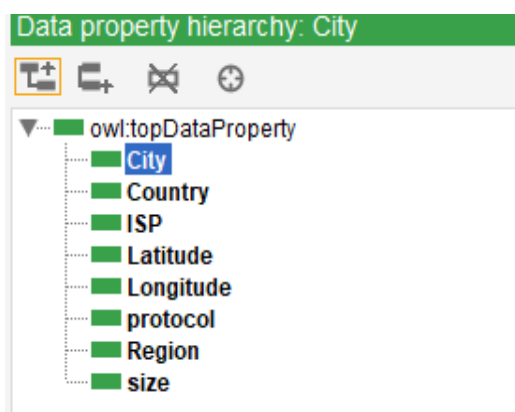
Η διαδικασία δημιουργίας του ontology έχει τα εξής βήματα:

1. Δημιουργία των Entities όπου καθένα θα αντιστοιχεί αργότερα σε ένα node στο γράφημα γνώσης.

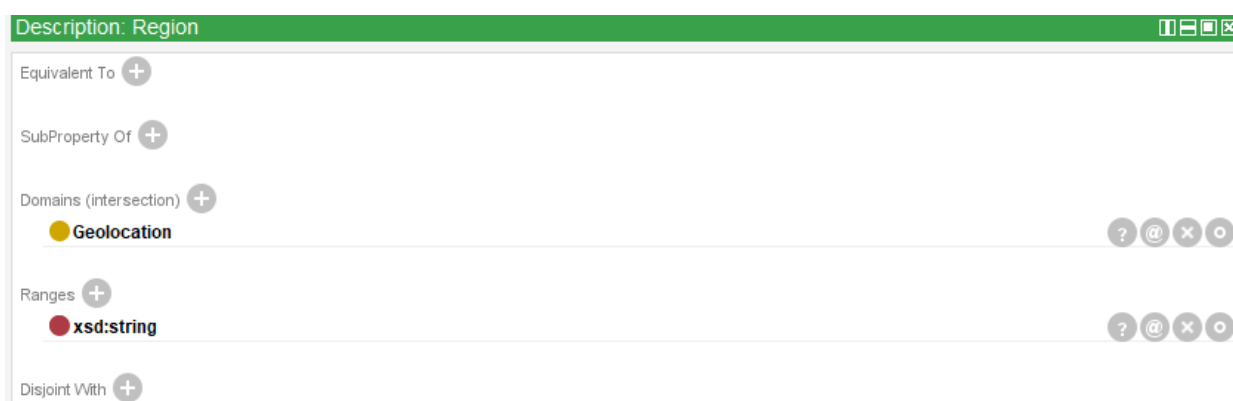


Εικόνα 15-Περιβάλλον δημιουργίας των Entities

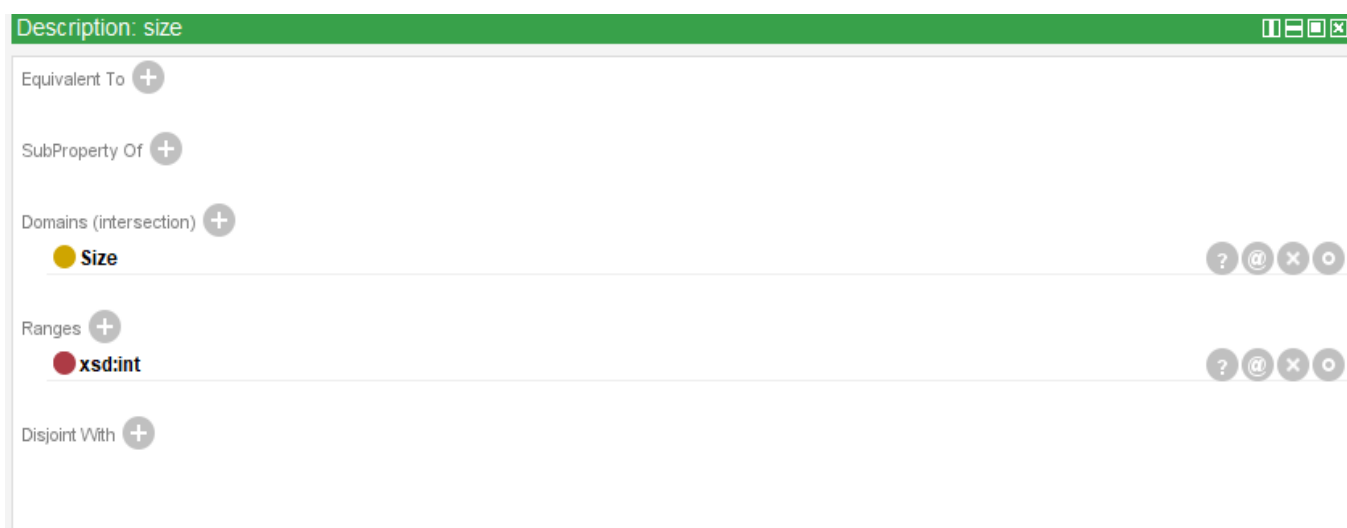
2. Δημιουργία των Data properties όπου καθένα αντιστοιχεί σε ένα ή περισσότερα Entities και καθορίζονται από τον τύπο δεδομένων τους όπως string.



Εικόνα 16-Περιβάλλον δημιουργίας των Data Properties

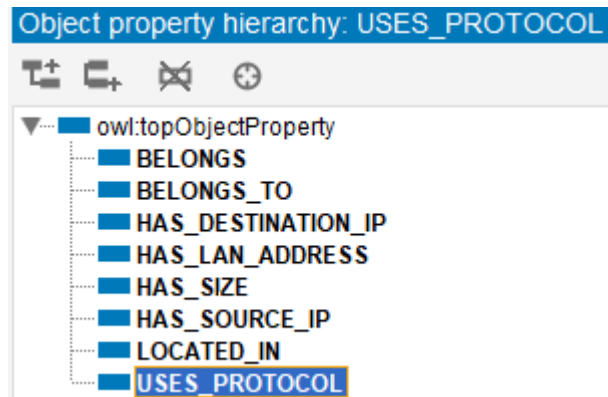


Εικόνα 17-Ιδιότητες Data Properties(Geolocation)

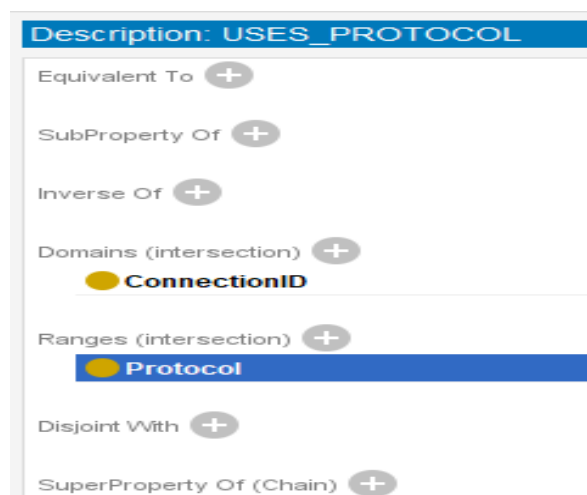


Εικόνα 18-Ιδιότητες Data Properties(Size)

3. Δημιουργία των Object properties, τα οποία είναι οι σχέσεις μεταξύ των Entities. Απαιτείται η αντιστοίχιση τους με τα Entities τα οποία ξεκινούν και καταλήγουν, καθώς θα αποτελέσουν μετά τα relationships στο γράφημα γνώσης.

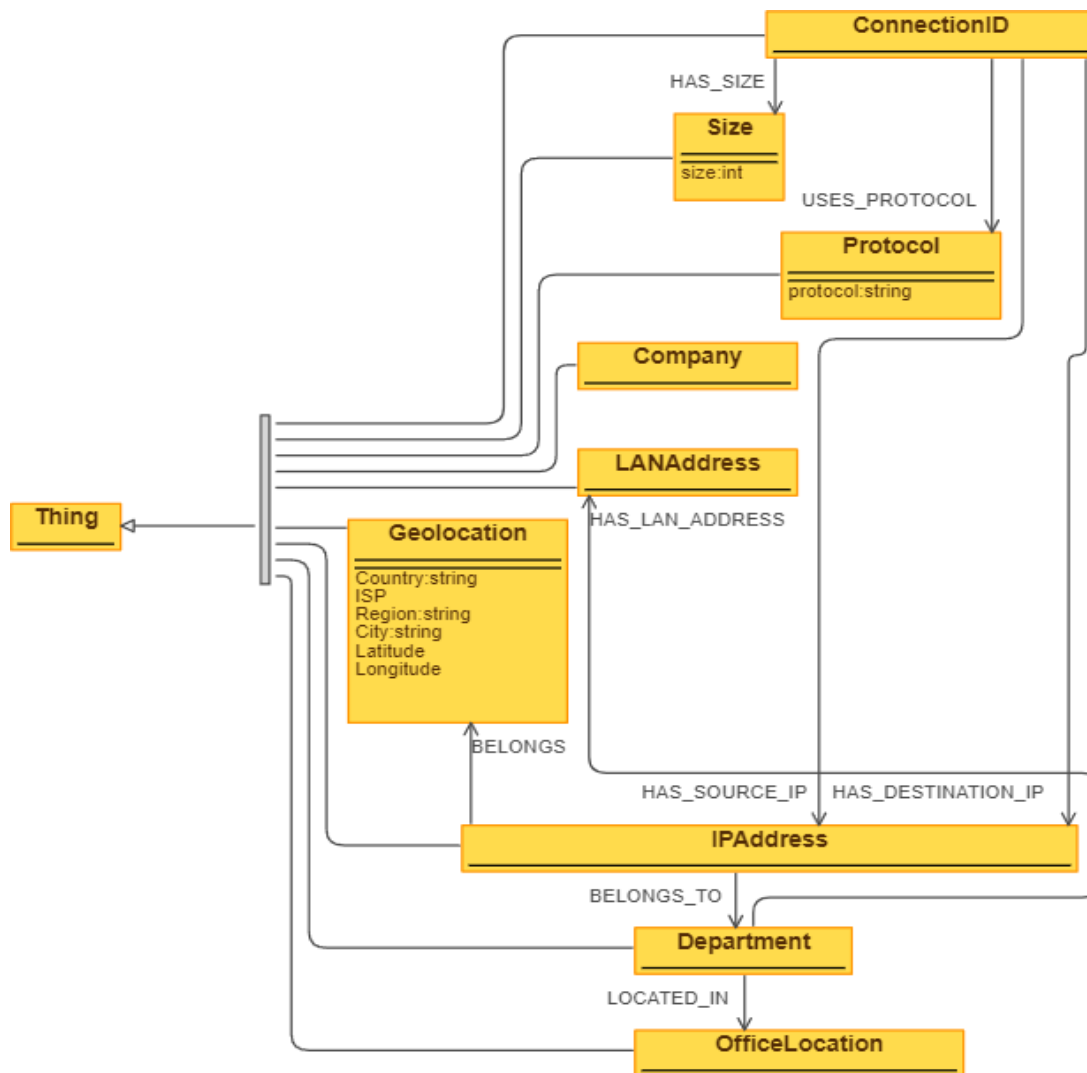


Εικόνα 19-Περιβάλλον δημιουργίας των Object Properties



Εικόνα 20-Αντιστοίχιση των Object Properties

Έτσι δημιουργείται το Ontology :



Εικόνα 21-Το ολοκληρωμένο Ontology

3.3 Ανάπτυξη script

Εφόσον γίνει η επιλογή των κατάλληλων δεδομένων και η δημιουργία ή εύρεση του ontology απαιτείται μια μορφή κώδικα η οποία θα τα συνδυάσει, θα τα αντιστοιχίσει σε οντότητες και σχέσεις και έπειτα θα τα αναπαραστήσει γραφικά μέσω του γραφήματος γνώσης. Η μορφή αυτή κώδικα θα σχεδιαστεί σε περιβάλλον υλοποίησης κώδικα. Μεταξύ άλλων επιλέχθηκε το IntelliJ IDEA. Στην συνέχεια πρέπει να επιλεγθεί η γλώσσα ανάπτυξης του κώδικα. Επιλέχθηκε η γλώσσα Python λόγω της απλότητας και των έτοιμων βιβλιοθηκών και κλάσεων που προσφέρει.

Το δυναμικό script αυτό είναι απαραίτητο για την υλοποίηση του σχεδιασμού καθώς θα προσφέρει :

- Εύκολη τροποποίηση σε διαφορετικά δεδομένα ή τύπους δεδομένων.
- Προσαρμογή σε διαφορετικές οντολογίες.
- Δυναμική αντιμετώπιση και πρόβλεψη καταστάσεων.
- Σύνδεση με πλατφόρμα γραφικής απεικόνισης γράφου.

Στην συνέχεια θα γίνει η επιλογή πλατφόρμας γραφήματος γνώσης το οποίο με απομακρυσμένη σύνδεση στο script θα εισάγει δυναμικά τα annotated δεδομένα σε αυτήν.

3.4 Κατασκευή του knowledge graph

1. Επιλογή Πλατφόρμας Δημιουργίας Γραφημάτων Γνώσης:

- Η επιλογή της κατάλληλης πλατφόρμας είναι κρίσιμη για την υποστήριξη της δημιουργίας του knowledge graph.
- Προσφέρονται διάφορες πλατφόρμες όπως Amazon Neptune, Microsoft Azure Cosmos DB, Neo4j, GraphDB και άλλες.

Στην αρχή έγινε προσπάθεια μέσω του GraphDB ,η οποία απέτυχε καθώς δεν υποστηριζόταν η σύνδεση της βάσης αυτής με δυναμικό script το οποίο θα μπορούσε να έκανε αυτόματα annotate τα δεδομένα και όχι στατικά.

Επιλέχθηκε το Neo4j , διότι μεταξύ άλλων:

- Χρησιμοποιεί Cypher, μια δηλωτική γλώσσα ερωτημάτων παρόμοια με την SQL, αλλά βελτιστοποιημένη για γραφήματα. Τώρα χρησιμοποιείται από άλλες βάσεις δεδομένων όπως το SAP HANA Graph και το γράφημα Redis μέσω του έργου openCypher.
- Εκτελεί συνεχείς διαβάσεις χρόνου σε μεγάλα γραφήματα τόσο για βάθος όσο και για πλάτος χάρη στην αποτελεσματική αναπαράσταση κόμβων και σχέσεων. Επιτρέπει την κλίμακα έως και δισεκατομμύρια κόμβους σε μέτριο υλικό.

- Είναι ένα ευέλικτο σχήμα γραφήματος ιδιοτήτων που μπορεί να προσαρμοστεί με την πάροδο του χρόνου, καθιστώντας δυνατή την υλοποίηση και την προσθήκη νέων σχέσεων αργότερα για τη συντόμευση και την επιτάχυνση των δεδομένων τομέα όταν αλλάζουν οι ανάγκες της επιχείρησης.
- Υποστηρίζει προγράμματα οδήγησης (Drivers) για δημοφιλείς γλώσσες προγραμματισμού, συμπεριλαμβανομένων των Java, JavaScript, .NET, Python και πολλών άλλων.

2. Συλλογή Δεδομένων και Οντολογίας:

- Για την δημιουργία του knowledge graph απαιτούνται δεδομένα και η δομή αυτών όπως ένα ontology.
- Τα στοιχεία αυτά μπορούν αν βρεθούν τόσο σε έτοιμες βάσεις δεδομένων όσο και να δημιουργηθούν εσωτερικά.
- Η οντολογία καθορίζει τις οντότητες, τις ιδιότητες και τις σχέσεις μεταξύ τους, παρέχοντας ένα δομημένο πλαίσιο για την οργάνωση των δεδομένων.

3. Προετοιμασία Δεδομένων:

- Εφόσον, βρεθούν τα δεδομένα απαιτείται η προετοιμασία αυτών όπως αναφέρθηκε σε προηγούμενο κεφάλαιο.
- Χρησιμοποιούνται τεχνικές όπως η αναγνώριση μοτίβων, ο έλεγχος σφαλμάτων και ο έλεγχος διπλότυπων εγγραφών.
- Αυτές οι διαδικασίες εξασφαλίζουν την ποιότητα των δεδομένων, μειώνοντας τον θόρυβο και εξαλείφοντας τυχόν ασυνέπειες.

4. Αντιστοίχιση Δεδομένων(Annotation):

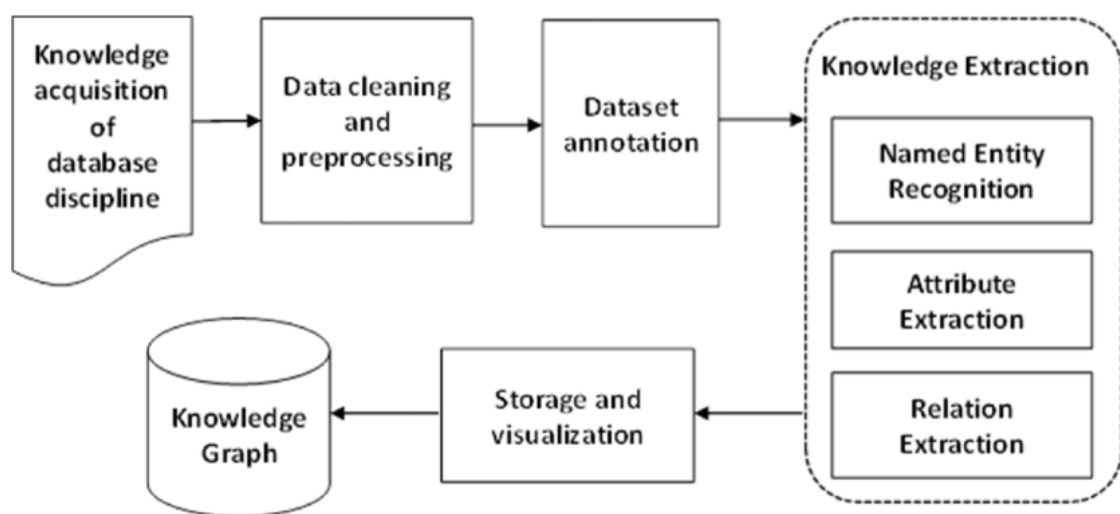
- Πρέπει να γίνει αντιστοίχιση των δεδομένων με τη δομή του γράφου.
- Η διαδικασία αυτή, γνωστή ως annotation, μπορεί να πραγματοποιηθεί στατικά μέσω μιας πλατφόρμας με περιορισμένη δυνατότητα επεξεργασίας ή δυναμικά μέσω της δημιουργίας αντίστοιχων δυναμικών scripts.
- Η αντιστοίχιση εξασφαλίζει ότι τα δεδομένα εισάγονται σωστά στο knowledge graph σύμφωνα με την οντολογία.

5. Κατασκευή του Knowledge Graph:

- Αφού επιλεγθούν η πλατφόρμα, τα δεδομένα και τα εργαλεία υλοποίησης,

ξεκινά η κατασκευή του γράφου.

- Τα δεδομένα εισάγονται στο σύστημα και αντιστοιχίζονται με την οντολογία.
- Η διαδικασία περιλαμβάνει τη δημιουργία κόμβων (nodes) και ακμών (edges), καθώς και την οργάνωση των σχέσεων μεταξύ τους σύμφωνα με το καθορισμένο σχήμα.
- Εξασφαλίζεται η ολοκλήρωση και η συνέπεια του knowledge graph μέσω συνεχών ελέγχων και βελτιώσεων.



Εικόνα 22-Βήματα δημιουργίας του γραφήματος γνώσης

3.5 Περιγραφή των queries και τεχνικών που χρησιμοποιήθηκαν για την ανάλυση του knowledge graph

Καθορισμός Στόχων Ανάλυσης:

- Πριν από την δημιουργία των queries , καθορίστηκαν οι κύριοι στόχοι της ανάλυσης.
- Οι στόχοι αυτοί περιλαμβάνουν τον εντοπισμό ανωμαλιών όπως την επικοινωνία μεταξύ οντοτήτων από τμήματα που ανήκουν σε διαφορετικούς κλάδους και χώρες σε ασυνήθιστες ώρες .
- Ακόμη, την αναγνώριση μοτίβων κυκλοφορίας δικτύου που μπορεί να οδηγήσει στην εύρεση μη εξουσιοδοτημένης πρόσβασης και την ανάλυση σχέσεων

μεταξύ διαφόρων οντοτήτων του δικτύου που μπορεί επίσης να αναγνωρίσει ύποπτη δραστηριότητα στο δίκτυο.

- Έτσι, η ανάλυση αποσκοπεί την βελτίωση της κυβερνοασφάλειας μέσω της αναγνώρισης πιθανών απειλών αλλά και σημείων ευπάθειας του συστήματος.

Δημιουργία Queries:

- Τα queries δημιουργήθηκαν με χρήση της γλώσσας Cypher .
- Η Cypher είναι μια query γλώσσα του εργαλείου Neo4j η οποία προσφέρει την δυνατότητα ανάκτησης δεδομένων από τον γράφο. Σε αντίθεση με την SQL, η οποία ασχολείται με σχεσιακές βάσεις δεδομένων, η Cypher επικεντρώνεται στην έκφραση μοτίβων γραφημάτων. Η αντιστοίχιση προτύπων γραφήματος είναι η βασική τεχνική στο Cypher, που επιτρέπει τη δημιουργία, την πλοήγηση, την περιγραφή και την εξαγωγή δεδομένων από ένα γράφημα με την εφαρμογή δηλωτικών μοτίβων.
- Τα queries σχεδιάστηκαν για να εξάγουν χρήσιμες πληροφορίες από το knowledge graph, όπως συχνά χρησιμοποιούμενες διαδρομές δεδομένων, κεντρικούς κόμβους στο δίκτυο και πιθανές ανωμαλίες.

Εκτέλεση και Επικύρωση των Queries:

- Τα queries εκτελέστηκαν στο knowledge graph και τα αποτελέσματα αναλύθηκαν για την ακρίβεια και τη συνέπεια τους.
- Πραγματοποιήθηκε επικύρωση των αποτελεσμάτων μέσω διασταύρωσης με γνωστά δεδομένα ή μέσω επίλυσης συγκεκριμένων περιπτώσεων χρήσης.
- Εντοπίστηκαν και διορθώθηκαν σφάλματα ή ασυνέπειες στα queries. Για την τελική μορφή του query που θα δώσει το επιθυμητό αποτέλεσμα χρειάζεται επανειλημμένη διόρθωση μεταβλητών και σχέσεων κατά την αναζήτηση.

Παρουσίαση Αποτελεσμάτων:

- Τα αποτελέσματα των queries μπορούν να παρουσιαστούν τόσο με γραφικό τρόπο όσο και με την μορφή tables
- Και με του δύο τρόπους η παρουσίαση των αποτελεσμάτων βοηθά στην κατανόηση των σχέσεων και των μοτίβων μέσα στο δίκτυο και διευκολύνει τη λήψη αποφάσεων για την ενίσχυση της κυβερνοασφάλειας.

- Οι αναλύσεις των αποτελεσμάτων βασίζονται σε ένα δομημένο ontology επιτρέποντας την ερμηνεία των αποτελεσμάτων με βάση τις συσχετίσεις των πληροφοριών.

ΚΕΦΑΛΑΙΟ 4: ΥΛΟΠΟΙΗΣΗ

4.1 Διαδικασία Ανάλυσης και Προετοιμασίας των δεδομένων

Σε αυτή την ενότητα, περιγράφονται οι διαδικασίες και οι μεθοδολογίες που χρησιμοποιούνται για την ανάλυση, τον καθαρισμό και την προετοιμασία των συνόλων δεδομένων για ενσωμάτωση στο γράφημα γνώσης(knowledge graph). Αυτά τα βήματα διασφαλίζουν την ακρίβεια, τη συνέπεια και την καταλληλότητα των δεδομένων για τη βελτίωση της ασφάλειας στον κυβερνοχώρο μέσω ανάλυσης δεδομένων δικτύου και υλοποιούνται μέσω δυναμικού script. Η κίνηση TCP/IP δεδομένων που επεξεργαζόμαστε και δημιουργούν το knowledge graph αφορούν την εταιρεία μας υποθετικά την Fage η οποία έχει 4 τμήματα που εδρεύουν σε διαφορετικές περιοχές της Ελλάδας με IP και στοιχεία που βρίσκονται στα αρχεία δεδομένων csv .

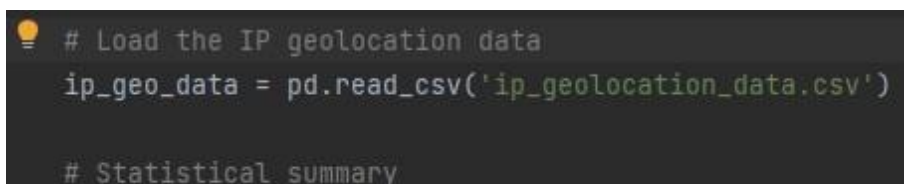
1.Ανάλυση Δεδομένων

Η φάση ανάλυσης δεδομένων περιλαμβάνει την εξέταση των συνόλων δεδομένων για την κατανόηση της δομής τους, τον εντοπισμό προτύπων και τον εντοπισμό ανωμαλιών.

Περιέχει:

- Διερευνητική Ανάλυση Δεδομένων (EDA-Exploratory Data Analysis)

Στόχος είναι η κατανόηση της βασικής δομής και των βασικών χαρακτηριστικών των συνόλων δεδομένων. Χρησιμοποιούνται μέθοδοι όπως η στατιστική ανασκόπηση η οποία υπολογίζει τη μέση, διάμεσο και την τυπική απόκλιση για αριθμητικές στήλες όπως Latitude και Longitude. Ο υπολογισμός αυτός γίνεται μέσω βιβλιοθηκών της Python όπως Pandas και NumPy.



```
# Load the IP geolocation data
ip_geo_data = pd.read_csv('ip_geolocation_data.csv')

# Statistical summary
```

Εικόνα 23-Συνάρτηση στατιστικής ανασκόπησης

	Latitude	Longitude
count	14.000000	14.000000
mean	24.961656	49.683619
std	29.961332	74.455326
min	-33.868800	-118.243700
25%	22.820050	4.659650
50%	35.670285	72.152450
75%	45.514093	96.097015
max	51.507400	151.209300

Εικόνα 24-Αποτέλεσμα συνάρτησης στατιστικής ανασκόπησης

- Αναγνώριση μοτίβου

Στόχος είναι ο προσδιορισμός προτύπων, συσχετίσεων και τάσεων στα δεδομένα με χρήση της μεθόδου Correlation Analysis όπου ελέγχεται η συσχέτιση μεταξύ γεωγραφικού πλάτους, γεωγραφικού μήκους και άλλων αριθμητικών πεδίων.

```
# Select only numeric columns for correlation
numeric_columns = ip_geo_data.select_dtypes(include=[np.number])

# Calculate the correlation matrix
correlation_matrix = numeric_columns.corr()
print(correlation_matrix)
```

Εικόνα 25-Συνάρτηση υπολογισμού συσχετίσεων μεταξύ X,Y και αριθμητικών πεδίων

	Latitude	Longitude
Latitude	1.000000	-0.054723
Longitude	-0.054723	1.000000

Εικόνα 26-Αποτέλεσμα συνάρτησης υπολογισμού συσχετίσεων μεταξύ X,Y και αριθμητικών πεδίων

2. Καθαρισμός Δεδομένων

Ο καθαρισμός δεδομένων περιλαμβάνει τη διόρθωση σφαλμάτων, τον χειρισμό τιμών που λείπουν και την τυποποίηση μορφών δεδομένων για να διασφαλιστεί ότι τα σύνολα δεδομένων είναι ακριβή και συνεπή.

- Διόρθωση σφαλμάτων

Ο στόχος είναι ο εντοπισμός και διόρθωση σφαλμάτων στα δεδομένα διασφαλίζοντας έγκυρες μορφές διευθύνσεων IP σε συγκεκριμένο εύρος διευθύνσεων.

```
# Check if destination_ip matches any LAN address in company_data
for company in company_data:
    if destination_ip == company['department_lan_address']:
        matching_companyd_data = company
        break
for company in company_data:
    if source_ip == company['department_lan_address']:
        matching_companys_data = company
        break
```

Εικόνα 27-Απόσπασμα διόρθωσης σφάλματος στον κώδικα

- Τυποποίηση Δεδομένων:

Ο στόχος είναι η διασφάλιση συνοχής σε μορφές και μονάδες δεδομένων με μεθόδους οι οποίες μετατρέπουν όλες τις ημερομηνίες σε τυπική μορφή, εάν υπάρχει.

```
# Standardize column names ip_geo_data.columns
= [col.lower() for col in ip_geo_data.columns]

# Ensure uniform date format (if applicable)
if 'timestamp' in ip_geo_data.columns:
    ip_geo_data['timestamp'] =
pd.to_datetime(ip_geo_data['timestamp'], format='%Y-%m-%d
%H:%M:%S')
```

Εικόνα 28-Παράδειγμα τυποποίησης δεδομένων

3. Μετασχηματισμός Δεδομένων

Περιλαμβάνει τη μετατροπή των καθαρισμένων δεδομένων σε μια μορφή κατάλληλη για ενσωμάτωση στο γράφημα γνώσης. Αυτό το βήμα διασφαλίζει ότι τα δεδομένα ευθυγραμμίζονται με την οντολογία και το σχήμα που ορίζονται για το γράφημα γνώσης.

- Ομαλοποίηση:

Ο στόχος είναι η κλιμάκωση αριθμητικών δεδομένων σε ένα τυπικό εύρος για ανάλυση. Απαιτείται βιβλιοθήκη Scikit-learn.

```
scaler = MinMaxScaler()
ip_geo_data[['latitude', 'longitude']] =
scaler.fit_transform(ip_geo_data[['latitude', 'longitude']])
```

Εικόνα 29-Παράδειγμα ομαλοποίησης από τον κώδικα

- Ενοποίηση δεδομένων

Ο στόχος είναι η συγχώνευση διαφορετικών συνόλων δεδομένων σε μια ενοποιημένη μορφή μέσω συγχώνευσης δεδομένων γεωγραφικής θέσης IP με δεδομένα εταιρείας με βάση τα σχετικά κλειδιά.

```
# Load company data
company_data = pd.read_csv('company_data.csv')

# Merge with IP geolocation data
merged_data = pd.merge(ip_geo_data, company_data,
left_on='isp', right_on='company name', how='inner')
```

Εικόνα 30-Ενοποίηση δεδομένων

4. Επικύρωση δεδομένων

Διασφαλίζει ότι τα μετασχηματισμένα δεδομένα πληρούν τα απαιτούμενα πρότυπα ποιότητας και είναι έτοιμα για ενσωμάτωση στο γράφημα γνώσης.

- Έλεγχοι συνέπειας

Ο στόχος είναι η διασφάλιση συνοχής δεδομένων μεταξύ διαφορετικών συνόλων δεδομένων με μεθόδους όπως ο έλεγχος διπλότυπων εγγραφών.

```
# Check for duplicates
merged_data.drop_duplicates(inplace=True)

# Referential integrity check (example)
valid_entries = merged_data['company_id'].notnull()
merged_data = merged_data[valid_entries]
```

Εικόνα 31-Παράδειγμα ελέγχου συνέπειας

- Έλεγχοι πληρότητας

Ο στόχος είναι η επιβεβαίωση της ύπαρξης όλων των υπάρχοντων δεδομένων

```
# Ensure all required fields are filled
required_fields = ['iprangestart', 'iprangeend', 'country', 'latitude', 'longitude']
complete_data = merged_data.dropna(subset=required_fields)
```

Εικόνα 32-Παράδειγμα ελέγχου πληρότητας

4.2 Υλοποίηση του script και αντιστοίχιση δεδομένων με βάση το ontology

Μετά την διαδικασία ανάλυσης και προετοιμασίας των δεδομένων γίνεται ταυτόχρονα η διαδικασία annotate αυτών. Πρώτα όμως εισάγεται η δομή της οντολογίας που δημιουργήθηκε και αποθηκεύτηκε στο diplomav4.ttl αρχείο το οποίο και διαβάζεται στο script μέσα από συναρτήσεις και διαδικασίες. Έπειτα καλούνται οι κατάλληλες συναρτήσεις έτσι ώστε να γίνει η κατάλληλη αντιστοίχιση (annotate) των δεδομένων των csv αρχείων μετά την προετοιμασία τους με την δομή του ontology που παρουσιάστηκε στο κεφάλαιο 3.

```
# Function to read CSV files with specified encoding
500+ usages
def read_csv(file_path, encoding='utf-8'):
    with open(file_path, 'r', encoding=encoding) as file:
        reader = csv.DictReader(file) # Use DictReader to read rows as dictionaries
        data = [row for row in reader]
    return data

# Step 1: Parse the Ontology File
ontology_file = r"C:\Users\user\Desktop\diplomav4.ttl"
```

Εικόνα 33-Συνάρτηση για την εισαγωγή του Ontology στο script

```
# Check if destination_ip matches any LAN address in company_data
for company in company_data:
    if destination_ip == company['department_lan_address']:
        matching_company_data = company
        break
```

Εικόνα 34-Συνάρτηση ελέγχου

Για κάθε εγγραφή στο tcp_traffic_data.csv :

- Δημιουργείται node ConnectionID με ορίσματα id και timestamp , το οποίο συνδέεται με node IPAddress μέσω της σχέσης HAS_DESTINATION_IP, συνδέεται με την σχέση HAS_SIZE στο node Size και με τη σχέση USES_PROTOCOL στο node Protocol.

```
f"MERGE (conn:ConnectionID {{id:
'{tcp['connection_id']}', timestamp: '{timestamp}'}})
" f"MERGE (src_ip:IPAddress {{address:
'{source_ip}'}}) " f"MERGE (src_geo:Geolocation
{{city: '{matching_source_ip_data['City']}', "
f"country: '{matching_source_ip_data['Country']}',
region: '{matching_source_ip_data['Region']}', "
f"latitude: {matching_source_ip_data['Latitude']},
longitude: {matching_source_ip_data['Longitude']},
isp: '{matching_source_ip_data['ISP']}'}}) "
f"MERGE (src_ip)-[:BELONGS]->(src_geo) "
f"MERGE (conn)-[:HAS_SOURCE_IP]->(src_ip)
"
f"MERGE (conn)-[:USES_PROTOCOL]->(protocol:Protocol
{{name: '{tcp['protocol']}'}}) "
```

Εικόνα 35-Δημιουργία του connectionid node

- Παράλληλα, η διευθύνσεις IP αυτές καθορίζεται σε ποιο IPRangeStart έως IPRangeEnd του ip_geolocation_data.csv βρίσκονται και δημιουργούνται κόμβοι(nodes) Geolocation που συνδέονται μέσω της σχέσης BELONGS.
- Τέλος, οι διευθύνσεις IP αντιστοιχούνται με βάση τα δεδομένα στο αρχείο company_data.csv πιθανόν σε κάποια εταιρεία και συγκεκριμένα σε τμήμα αυτής καθώς και την τοποθεσία του. Αυτό συμβαίνει μέσω της δημιουργίας σχέσεων IPAddress BELONGS_TO Department node και Department BELONGS_TO Company node.

```
f"MERGE (conn:ConnectionID {{id:
'{tcp['connection_id']}', timestamp: '{timestamp}'}}) "
f"MERGE (comp:Company {{name:
'{matching_companyd_data['company_name']}'}}) "
f"MERGE (dept:Department {{name:
'{matching_companyd_data['department']}'}}) "
f"MERGE (loc:OfficeLocation {{name:
'{matching_companyd_data['department_location']}'}}) "
f"MERGE (lan:LANAddress {{address:
'{matching_companyd_data['department_lan_address']}'}}) "
f"MERGE (dept)-[:BELONGS_TO]->(comp) "
f"MERGE (dept)-[:LOCATED_IN]->(loc) "
f"MERGE (dept)-[:HAS_LAN_ADDRESS]->(lan) "
f"MERGE (dest_ip:IPAddress {{address:
'{destination_ip}'}}) "
f"MERGE (dest_ip)-[:BELONGS_TO]->(dept) "
f"MERGE (conn)-[:HAS_DESTINATION_IP]->(dest_ip) "
```

Εικόνα 36-Επιπλέον διαδικασίες annotation

4.3 Περιγραφή της διαδικασίας υλοποίησης του Knowledge graph

Τέλος, απομένει η κατασκευή του knowledge graph κάτι που απαιτεί τη σύνδεση με το Neo4j. Αφού δημιουργηθεί καινούργιο instance στο Neo4j πρέπει να γίνει σύνδεση μέσα στο script μέσω των στοιχείων πρόσβασης στο instance:

```
# Connect to Neo4j
neo4j_uri = "neo4j+s://9cd266f5.databases.neo4j.io"
neo4j_user = "neo4j"
neo4j_password = "rq7p1nWQRm5QmNcmwNeB_0pgtor9Pi240JYc-nQJc84"
```

Εικόνα 37-Σύνδεση με το Neo4j

Για να γίνει αυτό καλείται η συνάρτηση `main()` του script η οποία καλεί τις συναρτήσεις για ανάγνωση των αρχείων δεδομένων, του αρχείου που περιέχει το ontology , της σύνδεσης με το Neo4j και για annotation των δεδομένων όπως και κατασκευάζεται το knowledge graph. Η γραφική απεικόνιση του γράφου και τα queries φαίνονται στο Neo4j.

Εικόνα 38-Η γραφική απεικόνιση μετά την δημιουργία του knowledge graph

Εικόνα 39-Μεγένθυση του knowledge graph

4.4 Ανάλυση των αποτελεσμάτων που προέκυψαν από τα queries

Τα queries μπορούν να εκτελεστούν τόσο στο περιβάλλον του Neo4j όσο και μέσω κώδικα στο script. Επιλέχθηκε η εκτέλεση μέσω του Neo4j καθώς προσφέρει καλύτερη γραφική απεικόνιση αλλά και πιο αναλυτικά αποτελέσματα χωρίς επιπλέον επεκτάσεις.

Πλεονεκτήματα των Queries:

1. Αυξημένη ακρίβεια και ευελιξία:
 - Εξαγωγή αποτελεσμάτων που βασίζονται σε πλούσια σημασιολογικά μοντέλα δεδομένων.
2. Εμπλουτισμένη ερμηνεία των δεδομένων:
 - Οι αναλύσεις βασίζονται σε ένα καλά δομημένο ontology, που επιτρέπει την ερμηνεία των δεδομένων με βάση τις συσχετίσεις και τα context των πληροφοριών.
3. Βελτιστοποίηση διαδικασιών ανίχνευσης:
 - Οι αναλυτές μπορούν να δημιουργούν σύνθετα queries που ανιχνεύουν ανωμαλίες ή μοτίβα επιθέσεων πιο αποτελεσματικά.
4. Ενοποίηση δεδομένων από πολλαπλές πηγές:
 - Τα knowledge graphs επιτρέπουν την ενοποίηση και την ανάλυση δεδομένων από διαφορετικές πηγές(π.χ. geolocation,tcp traffic) διευκολύνοντας την συνολική ανάλυση.
5. Καλύτερη διαχείριση και οργάνωση δεδομένων:
 - Η χρήση ενός ontology βοηθά στην οργάνωση των δεδομένων σε δομημένη μορφή ,καθιστώντας τα ευκολότερα στην επεξεργασία και ανάλυση.
6. Προσαρμοστικότητα και επεκτασιμότητα:

- Το knowledge graph μπορεί εύκολα να επεκταθεί για συμπεριλάβει νέα δεδομένα και οντότητες, προσφέροντας ευελιξία και προσαρμοστικότητα σε μελλοντικές ανάγκες.

Η αξιοποίηση του knowledge graph φαίνεται στα παρακάτω 3 queries για 3 διαφορετικές περιπτώσεις χρήσης και αξιοποίησης τους.

Σενάριο 1: Εντοπισμός Ασυνήθιστης Δραστηριότητας

Query: Εντοπισμός ασυνήθιστης κίνησης δεδομένων από συγκεκριμένη τοποθεσία σε συγκεκριμένες ώρες της ημέρας.

Π.χ. Κίνηση από Ιαπωνία όπου η εταιρεία δεν έχει επαφές και συγκεκριμένα από τις 10 το βράδυ μέχρι τις 6 το πρωί, ώρες που κατά κύριο λόγο η εταιρεία είναι κλειστή.

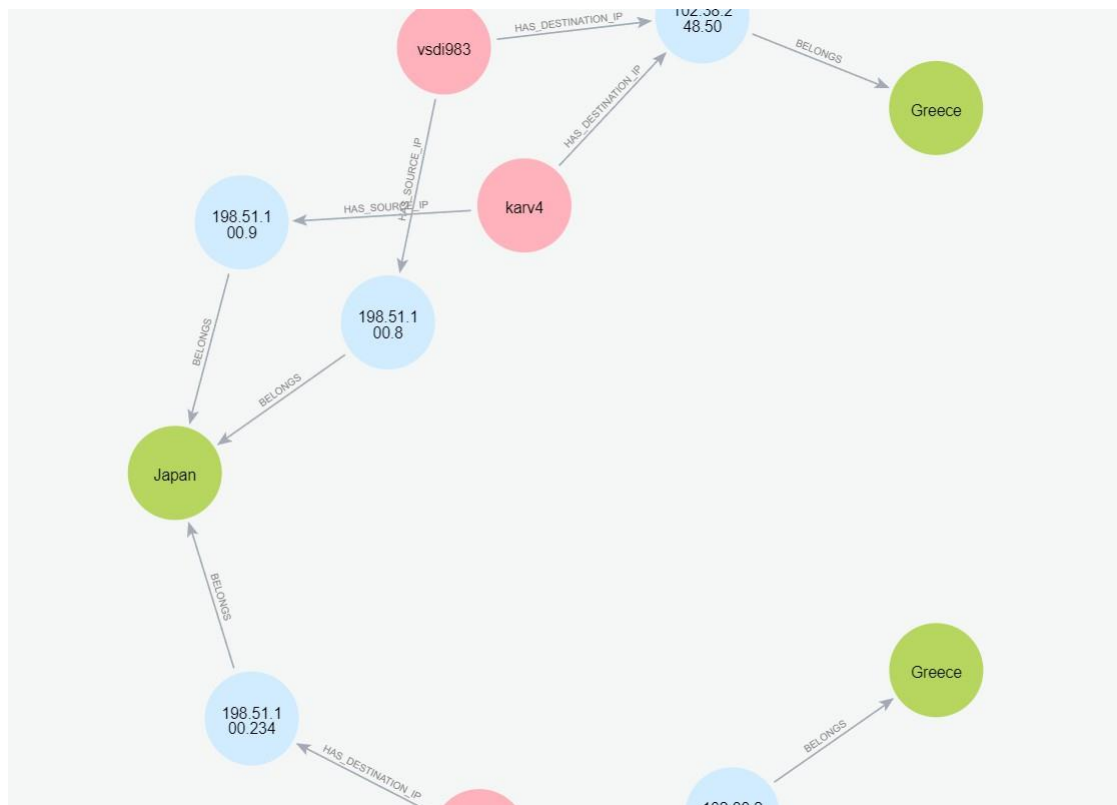
```

1 MATCH (conn:ConnectionID)-[hs:HAS_SOURCE_IP]->(source:IPAddress),
2     (conn:ConnectionID)-[hd:HAS_DESTINATION_IP]->(dest:IPAddress),
3     (source)-[bl:BELONGS]->(source_geo:Geolocation),
4     (dest)-[bb:BELONGS]->(dest_geo:Geolocation)
5 WHERE (source_geo.country = 'Japan' OR dest_geo.country = 'Japan')
6     AND (
7         time(substring(conn.timestamp, 11, 8)).hour >= 21
8         OR time(substring(conn.timestamp, 11, 8)).hour < 6
9     )
10 RETURN conn, hs, source, hd, dest, bl, source_geo, bb, dest_geo
11 ORDER BY conn.timestamp
12

```

Εικόνα 40-Query του 1ου σεναρίου σε Cypher

Κάνουμε αναζήτηση για ConnectionID κάποιας σύνδεσης που συνδέεται με IP διεύθυνση της Ιαπωνίας είτε με σχέση HAS_SOURCE_IP είτε με σχέση HAS_DESTINATION_IP, δηλαδή από η προς την εταιρεία μας. Ταυτόχρονα η ώρα της σύνδεσης αυτής να έγινε από τις 10 το βράδυ μέχρι τις 6 το επόμενο πρωί.



Εικόνα 41-Το αποτέλεσμα του 1ου query

Το αποτέλεσμα του query μας δείχνει ότι:

- Υπήρχαν 3 ύποπτες συνδέσεις δύο προς την IP 102.38.248.50 και μία από την IP 102.38.248.66
- Παρατηρούμε και τις ακριβείς ώρες αυτών.
- Οι IPs αυτές ανήκουν σε 2 τμήματα της εταιρείας
- Με την εκτέλεση του query αυτού εντοπίστηκε ασυνήθιστη δραστηριότητα η περαιτέρω ανάλυση της οποίας μπορεί να οδηγήσει στον εντοπισμό σοβαρής απειλής προς την εταιρεία αλλά και τις εταιρείες με τις οποίες επικοινωνεί η εταιρεία μας.

Έτσι αναδεικνύονται:

- Δυνατότητα εντοπισμού και ανάλυσης ανώμαλης κίνησης δεδομένων σε πραγματικό χρόνο.
- Χρήση χρονικών φίλτρων για την ανίχνευση ύποπτων δραστηριοτήτων σε μη εργάσιμες ώρες.

Σενάριο 2: Συσχέτιση Εσωτερικών και Εξωτερικών IP

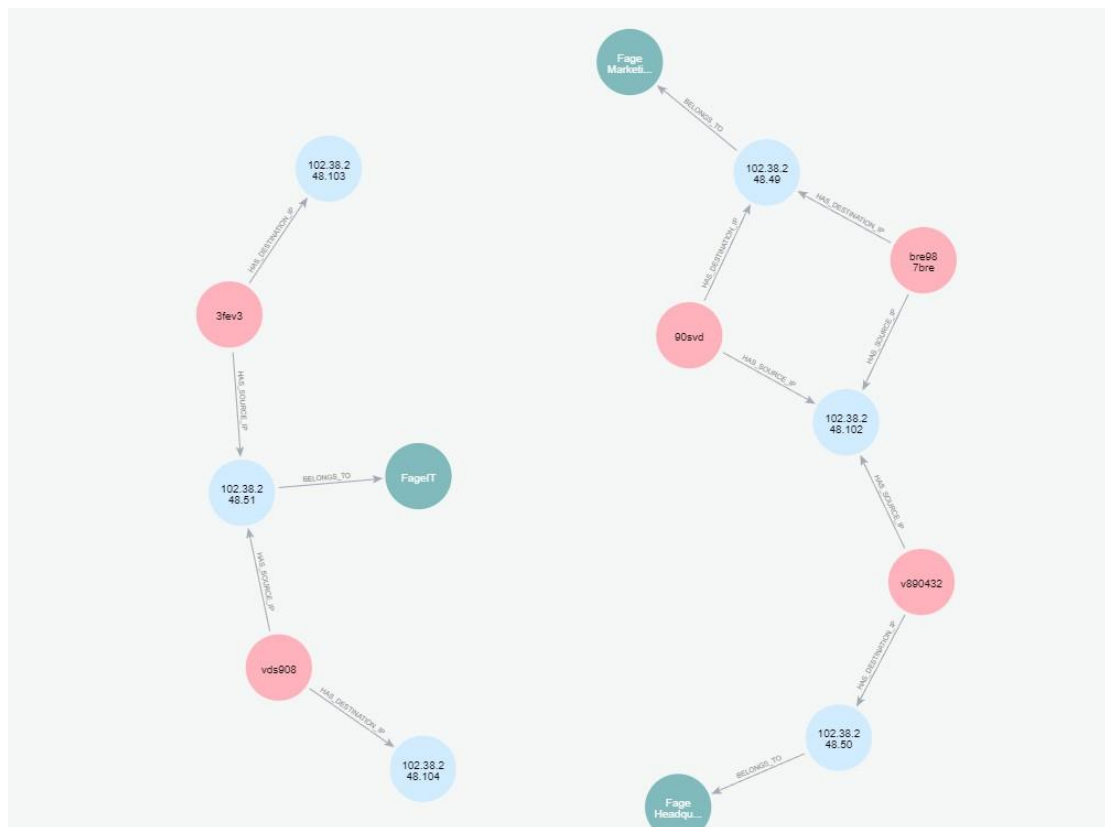
Query: Εύρεση των εσωτερικών IP που έχουν επικοινωνήσει με συγκεκριμένες εξωτερικές IP και συσχέτιση με τμήματα της εταιρείας.

Γίνεται ανάλυση για 3 συγκεκριμένες IP, οι οποίες ανήκουν στην εταιρεία Terkenlis και υποθετικά να φανεί αν υπήρξε επικοινωνία από ή προς την εταιρεία μας καθώς υπήρξε Data Breach σε αυτήν και πρέπει να ελεγχθεί αν μας επηρέασε και σε ποιο τμήμα της εταιρείας συγκεκριμένα.

```
1 MATCH (conn:ConnectionID)-[source_rel:HAS_SOURCE_IP]->(source_ip:IPAddress)
2 WHERE source_ip.address IN ['102.38.248.102', '102.38.248.103', '102.38.248.104']
3 MATCH (conn)-[dest_rel:HAS_DESTINATION_IP]->(dest_ip:IPAddress)-[dept_rel:BELONGS_TO]->(dept:Department)
4 RETURN conn, source_ip, source_rel, dest_ip, dest_rel, dept, dept_rel
5 UNION
6 MATCH (conn:ConnectionID)-[dest_rel:HAS_DESTINATION_IP]->(dest_ip:IPAddress)
7 WHERE dest_ip.address IN ['102.38.248.102', '102.38.248.103', '102.38.248.104']
8 MATCH (conn)-[source_rel:HAS_SOURCE_IP]->(source_ip:IPAddress)-[dept_rel:BELONGS_TO]->(dept:Department)
9 RETURN conn, dest_ip, dest_rel, source_ip, source_rel, dept, dept_rel
10 |
```

Εικόνα 42-Query του 2ου σεναρίου σε Cypher

Εκτελείται το query αναζητώντας σύνδεση που να συνδέεται είτε με την σχέση HAS_SOURCE_IP είτε HAS_DESTINATION_IP με IPs αυτές των 3 τμημάτων της εταιρείας Terkenlis (102.38.248.102, 102.38.248.103, 102.38.248.104) και να επιστραφεί αν υπάρχει το τμήμα της εταιρείας μας που επικοινωνήσε με αυτές.



Εικόνα 43-Το αποτέλεσμα του 2ου query

Το αποτέλεσμα του query μας δείχνει ότι:

- Υπήρξαν 5 διαφορετικές συνδέσεις με τα τμήματα της εταιρείας μας
- Οι δύο είχαν ως Source IP την διεύθυνση του τμήματος της εταιρείας μας FageIT, δύο άλλες είχαν ως Destination IP τη διεύθυνση του τμήματος της εταιρείας μας FageMarketing και η τελευταία είχε Destination IP αυτήν της τμήματος της εταιρείας μας FageHeadquarters.
- Οι συνδέσεις αυτών των τριών τμημάτων της εταιρείας μας πρέπει τώρα να αναλυθούν περαιτέρω για να βρεθεί αν υπήρξε ζημιά στην εταιρεία μας.

Με την χρήση αυτού αλλά και παρεμφερής λογικής παραδειγμάτων αναδεικνύονται:

- Συσχέτιση των IP με συγκεκριμένα τμήματα της εταιρείας για καλύτερη παρακολούθηση και ανάλυση
- Βελτίωση της ικανότητας ανίχνευσης ενδοεταιρικών κινδύνων.

Σενάριο 3: Καταγραφή ιστορικού συνδέσεων

Query: Δημιουργία ιστορικού συνδέσεων για συγκεκριμένη IP με όλες τις σχετικές πληροφορίες.

Εκτελείται το query για την IP '102.38.248.50' η οποία ανήκει στο τμήμα των headquarters της εταιρείας μας για να ελέγξουμε την δραστηριότητα από και προς αυτήν καθώς και να παρατηρήσουμε τυχόν ύποπτη δραστηριότητα μέσω ελέγχου πρωτοκόλλων, μεγεθών πακέτων ή geolocation.

```
1 MATCH (conn:ConnectionID)-[source_rel:HAS_SOURCE_IP]->(ip:IPAddress {address: '102.38.248.50'})
2 MATCH (conn)-[dest_rel:HAS_DESTINATION_IP]->(dest_ip:IPAddress)
3 MATCH (dest_ip)-[b:BELONGS]->(dest_geo:Geolocation)
4 OPTIONAL MATCH (conn)-[size_rel:HAS_SIZE]->(size:Size)
5 OPTIONAL MATCH (conn)-[protocol_rel:USES_PROTOCOL]->(protocol:Protocol)
6 RETURN conn AS connection,
7         ip AS ip_address,
8         dest_ip AS opposite_ip_address,
9         dest_geo AS opposite_ip_geolocation,
10        source_rel AS source_relationship,
11        dest_rel AS destination_relationship,
12        size_rel AS size_relationship,
13        size AS size,
14        protocol_rel AS protocol_relationship,
15        protocol AS protocol,
16        b
17 UNION
18 MATCH (conn:ConnectionID)-[dest_rel:HAS_DESTINATION_IP]->(ip:IPAddress {address: '102.38.248.50'})
19 MATCH (conn)-[source_rel:HAS_SOURCE_IP]->(source_ip:IPAddress)
20 MATCH (source_ip)-[b:BELONGS]->(source_geo:Geolocation)
```

Εικόνα 44-1ο μέρος του 3ου query σε Cypher

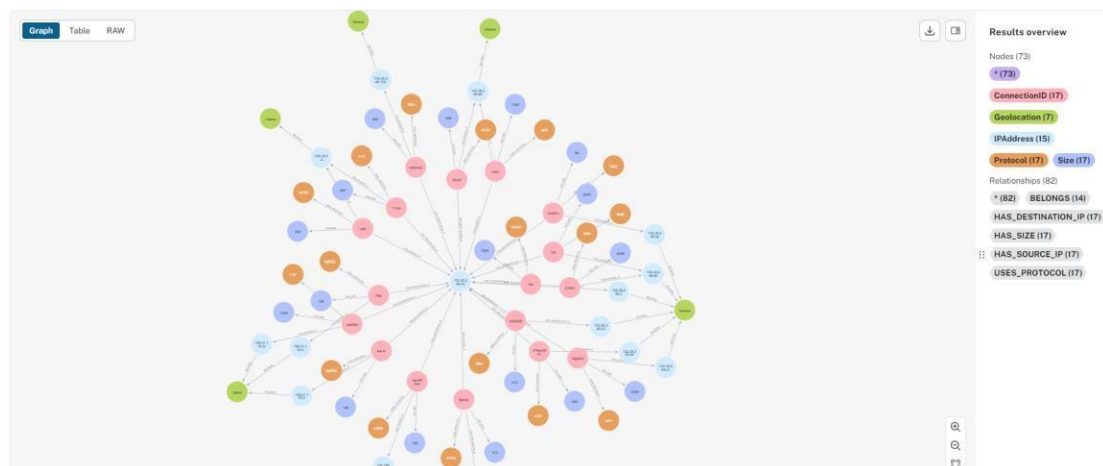
```

20 MATCH (source_ip)-[b:BELONGS]->(source_geo:Geolocation)
21 OPTIONAL MATCH (conn)-[size_rel:HAS_SIZE]->(size:Size)
22 OPTIONAL MATCH (conn)-[protocol_rel:USES_PROTOCOL]->(protocol:Protocol)
23 RETURN conn AS connection,
24         ip AS ip_address,
25         source_ip AS opposite_ip_address,
26         source_geo AS opposite_ip_geolocation,
27         source_rel AS source_relationship,
28         dest_rel AS destination_relationship,
29         size_rel AS size_relationship,
30         size AS size,
31         protocol_rel AS protocol_relationship,
32         protocol AS protocol,
33         b
34

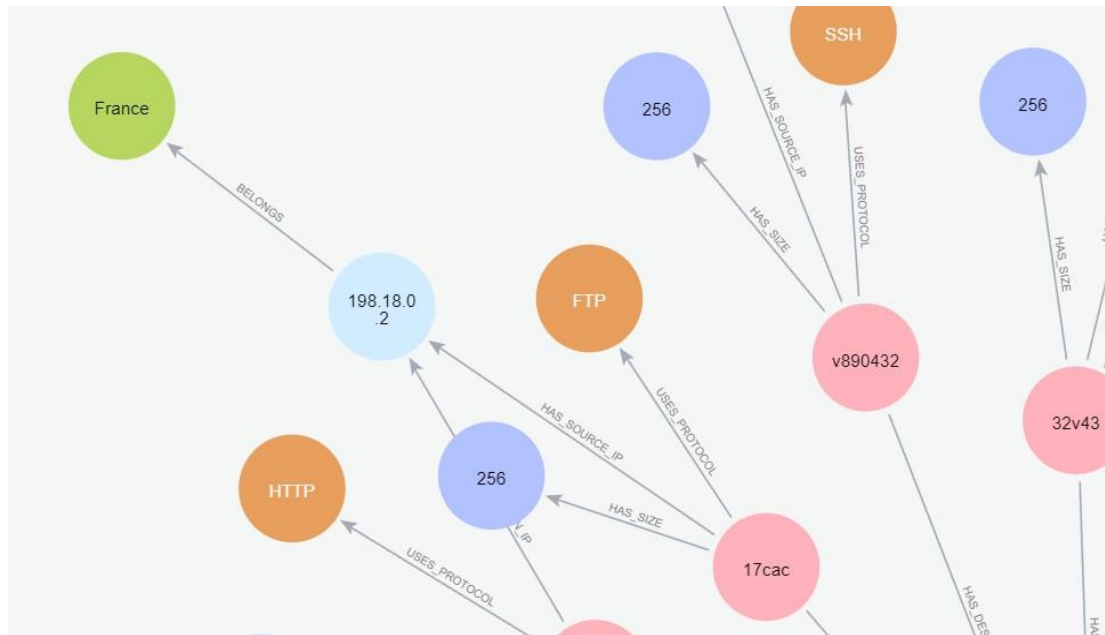
```

Εικόνα 45-2ο μέρος του 3ου query σε Cypher

Εκτελείται το query αναζητώντας τις συνδέσεις που έχουν προορισμό ή ξεκινούν από την διεύθυνση των κεντρικών γραφείων της εταιρείας. Για τις συνδέσεις αυτές επιστρέφονται παράμετροι όπως το πρωτόκολλο που χρησιμοποιήθηκε , το μέγεθος του πακέτου καθώς και η χώρα από ή προς την οποία έγινε η επικοινωνία.



Εικόνα 46-Το αποτέλεσμα του 3ου query



Εικόνα 47-Το αποτέλεσμα του 3ου query σε μεγέθυνση

Τα αποτελέσματα των queries δείχνουν τα εξής:

- Εκτός από τις εγχώριες επικοινωνίες υπήρξαν επικοινωνίες από και προς τα κεντρικά της εταιρείας με χώρες όπως η Ιαπωνία ,η Βραζιλία και οι ΗΠΑ γεγονός που πιθανόν να κρύβουν παράνομη δραστηριότητα.
- Βρέθηκαν συνδέσεις με χρήση πρωτοκόλλου MySQL ,πρωτόκολλο που χρησιμοποιείται συχνά σε SQL Injection επιθέσεις αλλά και SSH πρωτόκολλο που χρησιμοποιείται για ασφαλής απομακρυσμένη πρόσβαση αλλά και αυθεντικοποίηση οντότητας που πιθανόν να μην είναι εταιρική.
- Τέλος, παρατηρούνται συνδέσεις με μέγεθος πακέτων δεδομένων αρκετά πάνω από το σύνηθες κάτι που μπορεί να υποδεικνύει ύποπτη δραστηριότητα.
- Αυτές οι ενδείξεις απαιτούν πιο λεπτομερή ανάλυση και πιθανόν να οδηγήσουν σε εντοπισμό κάποιας παράνομης δραστηριότητας.

Queries παρόμοιας λογικής με το προηγούμενο προσφέρουν:

- Παρακολούθηση και καταγραφή δραστηριότητας μιας συγκεκριμένης IP για την ανάλυση των μοτίβων κυκλοφορίας.
- Εύκολη προσπέλαση στο ιστορικό δεδομένων για ανάλυση περιστατικών ασφαλείας.

ΚΕΦΑΛΑΙΟ 5: ΑΞΙΟΛΟΓΗΣΗ

5.1 Συζήτηση για τα ευρήματα και την επίδραση τους στην κυβερνοασφάλεια

Περίληψη των ευρημάτων

Η έρευνα επικεντρώθηκε στη χρήση γραφημάτων γνώσης (knowledge graphs) για την ανάλυση δεδομένων δικτύου TCP/IP με στόχο την ενίσχυση της κυβερνοασφάλειας. Τα βασικά ευρήματα μέσω των queries (σενάρια) περιλαμβάνουν:

1. Αποτελεσματικότητα των γραφημάτων γνώσης στην ανίχνευση απειλών:

- Τα γραφήματα γνώσης μπορούν να αποκαλύψουν μοτίβα και μη συνηθισμένη κίνηση δεδομένων στην ροή του δικτύου που δεν είναι εύκολα αντιληπτά με τις παραδοσιακές μεθόδους.
- Ενσωματώνοντας τόσο δομημένα όσο και μη δομημένα δεδομένα, τα γραφήματα γνώσης προσφέρουν μια πιο ολιστική και πλούσια αναπαράσταση των δεδομένων του δικτύου μέσα από την γραφική απεικόνιση των γραφημάτων γνώσης.
- Ένα από τα βασικά πλεονεκτήματα της χρήσης γραφημάτων γνώσης είναι ότι διευκολύνουν την αυτοματοποιημένη συλλογιστική ώστε να μπορούν να συναχθούν νέα γεγονότα από ρητές δηλώσεις (υφιστάμενα δεδομένα), και δυναμικά ενημερωμένες πληροφορίες που παρέχονται σχετικά με τις πιο πρόσφατες ευπάθειες και απειλές.
- Τα γραφήματα επικοινωνίας διεργασιών σε επίπεδο κεντρικού υπολογιστή είναι κατάλληλα για την εξαγωγή συμπερασμάτων σύνδεσης δικτύου, τα οποία με τη σειρά τους μπορούν να συγκεντρωθούν σε γραφήματα επικοινωνίας κεντρικού υπολογιστή σε όλο το σύστημα[35].

2. Βελτίωση της απόκρισης σε απειλές:

- Χρησιμοποιώντας τη σχεσιακή δομή των γραφημάτων, είναι δυνατή η ταχύτερη και ακριβέστερη ανίχνευση και απόκριση σε απειλές .
- Τα γραφήματα γνώσης επιτρέπουν τον προληπτικό εντοπισμό πιθανών απειλών πριν αυτές εξελιχθούν σε σοβαρά περιστατικά ασφαλείας εντοπίζοντας ύποπτες δραστηριότητες και πιθανές μη εξουσιοδοτημένες προσβάσεις χρηστών.

- Η αναπαράσταση της ανάλυσης πακέτων με βάση την οντολογία μπορεί να διευκολύνει την αυτοματοποιημένη συλλογιστική για εφαρμογές παρακολούθησης TCP/IP δικτύου. Ο συλλογισμός πάνω από ontologies μπορεί να διευκολύνει την αυτοματοποίηση της ανάλυσης δικτύου για ανίχνευση απειλών για απόκτηση διευθύνσεων IP.

3. Ενσωμάτωση σε ήδη υπάρχοντα πλαίσια κυβερνοασφάλειας:

- Προτάθηκαν πρακτικοί τρόποι ενσωμάτωσης των γραφημάτων γνώσης όπως ο καθαρισμός και η προετοιμασία δεδομένων, η ανακάλυψη μοτίβων και ανωμαλιών.
- Μέσα από την υλοποίηση της εργασίας αποδείχθηκε ο τρόπος με τον οποίο η ανάλυση δεδομένων δικτύου μέσω γραφημάτων γνώσης παίζει καθοριστικό ρόλο στην ασφάλεια ενός συστήματος στον κυβερνοχώρο.
- Η εξόρυξη οντοτήτων από γραφήματα γνώσης στον κυβερνοχώρο μπορεί να βοηθήσει τους αναλυτές για να κατανοήσουν τα δεδομένα απειλών. Τα αρχεία καταγραφής περιστατικών κυβερνοασφάλειας μπορούν να καταγραφούν αποτελεσματικά σε προέλευση που βασίζεται σε γραφήματα, τα οποία μπορούν να χρησιμοποιηθούν για τη δημιουργία γραφημάτων προέλευσης με ειδοποιήσεις και, τελικά, εννοιολογημένα γραφήματα επίθεσης. Αυτό επιτρέπει τον συνδυασμό και την ενσωμάτωση μιας σειράς τεχνικών για τον εντοπισμό απειλών στον κυβερνοχώρο και τη δημιουργία συναγερμών.

4. Εύρεση κενών ασφαλείας:

- Τα παραδείγματα των queries απέδειξαν την πιθανή εύρεση κενών ασφαλείας, διόδων όπου κακόβουλοι χρήστες και λογισμικά θα μπορούσαν να εκμεταλλευτούν αναδεικνύοντας την ανάγκη για επιπλέον θωράκιση του συστήματος.
- Μπορούν να κατασκευαστούν γραφήματα επίθεσης και ως εκ τούτου, οι επιθέσεις ανακατασκευάζονται με αλυσίδες (προς τα πίσω ή προς τα εμπρός) και ερωτήματα γραφημάτων. Τα γραφήματα της γνώσης στον κυβερνοχώρο παρέχουν δεδομένα προέλευσης για ειδοποίηση, τα οποία με τη σειρά τους μπορούν να χρησιμοποιηθούν για τον εντοπισμό μιας πιθανής βασικής αιτίας μιας επίθεσης, με την οποία η βαθμολογία προειδοποίησης αυξάνεται για κάθε προηγούμενη ειδοποίηση στη διαδρομή.
- Τα γραφήματα επίθεσης μπορούν να συνδυαστούν με ένα TCP/IP δίκτυο για να προσδιορίσουν αποτελεσματικά την πιθανότητα μονοπατιών επίθεσης. Γράφοντας [62]

συλλογιστικούς κανόνες για τα τρωτά σημεία (που αναπαρίστανται ως κόμβοι γραφήματος), μπορεί να εκτελεστεί αυτοματοποιημένη συλλογιστική για να συμπεράνει ότι μπορεί να προκαλέσει μια ευπάθεια.

5. Οπτικοποίηση με γραφήματα γνώσης

- Ένας σοβαρός περιορισμός των παραδοσιακών εργαλείων ασφάλειας πληροφοριών είναι ότι οι υπερβολικές πληροφορίες ενδέχεται να εμφανίζονται (από IDS, σαρωτές ευπάθειας, διαχειριστές τείχους προστασίας, εργαλεία SIEM) με πολύ μικρό περιεχόμενο.

- Τα γραφήματα γνώσεων για την κυβερνοασφάλεια παρέχουν τη δυνατότητα αναπαράστασης και οπτικοποίησης των πληροφοριών ασφαλείας, επιτρέποντας έγκαιρα τον εντοπισμό συμβάντων στον κυβερνοχώρο[35].

- Ορισμένα παραδείγματα οπτικοποιήσεων γραφημάτων γνώσης για την ασφάλεια στον κυβερνοχώρο περιλαμβάνουν για παράδειγμα οπτικοποίηση ενός δέντρου επίθεσης με στόχους επίθεσης και υποστόχους, και το αντίστοιχο μέσο επίθεσης για να επιτρέψουν στους αναλυτές ασφαλείας να εξερευνήσουν συγκεντρωτικά δεδομένα καταγραφής μέσω των σχέσεων.

Επίδραση στην κυβερνοασφάλεια

1. Βελτίωση της ικανότητας ανίχνευσης απειλών

- Η χρήση γραφημάτων γνώσης ενισχύει την ικανότητα ανίχνευσης απειλών, ειδικά σε περιπτώσεις που οι επιθέσεις δεν ακολουθούν γνωστά πρότυπα ή υπογραφές. Αυτό συμβάλλει στη μείωση του χρόνου ανίχνευσης και απόκρισης, αυξάνοντας την αποτελεσματικότητα των μέτρων ασφαλείας.

- Τα συστήματα που βασίζονται στη γνώση μπορούν να παρέχουν βοήθεια στους αναλυτές μέσω μερικής αυτοματοποίησης, ανάλυσης και οπτικοποίησης πολύπλοκων δεδομένων κυβερνοασφάλειας[35]. Χρησιμοποιούνται από αναλυτές επιπέδου 1 για την ανίχνευση ενδείξεων πιθανών ανωμαλιών. **Δηλαδή μπορούμε να κατατάξουμε τα γραφήματα γνώσης στην κατηγορία των SIEM συστημάτων.**

2. Ανακάλυψης κρυφών σχέσεων και μοτίβων

- Μέσω της ανάλυσης των σχέσεων και των μοτίβων στα δεδομένα του δικτύου, τα γραφήματα γνώσης επιτρέπουν την αποκάλυψη κρυφών σχέσεων που μπορεί να

[63]

υποδηλώνουν κακόβουλη δραστηριότητα. Αυτή η ικανότητα παρέχει μια πιο βαθιά κατανόηση της συμπεριφοράς των οντοτήτων του δικτύου.

- Με τη μοντελοποίηση της γνώσης υποβάθρου των εισβολέων σε ένα γράφημα γνώσης, οι ευαίσθητες πληροφορίες που δεν έχουν αποκαλυφθεί ακόμη μπορούν να συναχθούν από την υπονοούμενη (κρυφή) γνώση και μπορούν να προσεγγιστούν

3. Ενίσχυση της προληπτικής άμυνας

- Τα γραφήματα γνώσης επιτρέπουν την προληπτική αναγνώριση και αντιμετώπιση απειλών, μειώνοντας τον κίνδυνο επιτυχημένων κυβερνοεπιθέσεων κάτι που φάνηκε μέσω των ευρημάτων των queries. Αυτό συμβάλλει στη δημιουργία πιο ανθεκτικών και ασφαλών δικτύων.

- Οι συλλογιστικοί κανόνες για τις πληροφορίες κυβερνοαπειλής μπορούν να χρησιμοποιηθούν για την παροχή συγκεκριμένων αμυντικών στρατηγικών, σύμφωνα με τις σχέσεις μεταξύ τρωτών σημείων, αδυναμιών, πλατφορμών και μοτίβων επίθεσης. Τα μοτίβα μπορούν να χρησιμοποιηθούν για να συνάγουν αυτόματα μια σειρά χρήσιμων πληροφοριών και απειλών.

4. Οικονομικά οφέλη

Η ταχύτερη ανίχνευση και απόκριση σε απειλές μπορεί να μειώσει τα κόστη που συνδέονται με την αποκατάσταση από παραβιάσεις ασφάλειας. Διεθνώς ο μέσος όρος ζημιάς από παραβίαση δεδομένων ανήλθε σε 7.5 εκατ. ευρώ. Η βελτίωση της ασφάλειας μπορεί επίσης να μειώσει τον κίνδυνο επιβολής προστίμων για μη συμμόρφωση με κανονισμούς προστασίας δεδομένων.

5. Ενσωμάτωση νέων τεχνολογιών

Η έρευνα αυτή προωθεί την ενσωμάτωση νέων τεχνολογιών, όπως η μηχανική μάθηση και η τεχνητή νοημοσύνη, στην ανάλυση δεδομένων δικτύου. Αυτές οι τεχνολογίες μπορούν να ενισχύσουν περαιτέρω την ικανότητα των γραφημάτων γνώσης να ανιχνεύουν και να ανταποκρίνονται σε απειλές.

5.2 Συζήτηση για τις προοπτικές και της δυνατότητες βελτίωσης της μεθοδολογίας

Η μεθοδολογία που εφαρμόστηκε στην εργασία αυτή έχει δείξει σημαντική

αποτελεσματικότητα στην ανάλυση δεδομένων δικτύου TCP/IP με τη χρήση γραφημάτων γνώσης. Ωστόσο, υπάρχουν πάντα περιθώρια για βελτιώσεις και εξελίξεις. Ακολουθούν προοπτικές και δυνατότητες βελτίωσης της μεθοδολογίας:

1. Βελτίωση της προετοιμασίας και καθορισμού των δεδομένων:

Η διαδικασία προετοιμασίας δεδομένων είναι κρίσιμη για την ποιότητα των αποτελεσμάτων. Σύμφωνα με έρευνες η βελτίωση θα προέλθει από την μετάβαση προς μια πιο επεκτάσιμη και αποτελεσματική προσέγγιση στοχεύοντας στην μείωση του ανθρώπινου γενικού κόστους κατά την ανάπτυξη πιο ακριβών και αντιπροσωπευτικών περιπτώσεων δεδομένων.

2. Ενίσχυση της Διαδικασίας αντιστοίχισης δεδομένων(Annotation):

Η αντιστοίχιση δεδομένων με τη δομή του γράφου είναι μια κρίσιμη διαδικασία που μπορεί να βελτιωθεί με την ανάπτυξη πιο αυτοματοποιημένων εργαλείων και πλατφορμών. Η ενσωμάτωση δυναμικών scripts και η χρήση μηχανικής μάθησης [2] μπορεί να μειώσει την ανάγκη για χειροκίνητες διαδικασίες και να αυξήσει την αποδοτικότητα.

3. Επέκταση της οντολογίας:

Η συνεργασία μεταξύ οργανισμών, αναλυτών και επαγγελματιών ως προς την υλοποίηση μιας πιο ολοκληρωμένης και λεπτομερής οντολογίας μπορεί να βελτιώσει την αναπαράσταση δεδομένων καθώς και την ανάλυση των σχέσεων μεταξύ των οντοτήτων.

4. Ενσωμάτωση Προηγμένων Τεχνικών Μηχανικής Μάθησης:

Η μηχανική μάθηση, ιδιαίτερα η βαθιά μάθηση, μπορεί να υιοθετηθεί για την κατασκευή, την ερμηνεία και τον εμπλουτισμό του γραφήματος γνώσης προς άγνωστες οντότητες και σχέσεις [8].

Η κατηγοριοποίηση των αλγορίθμων για την ανίχνευση ανωμαλιών με βάση τα γραφήματα εξαρτάται από την προσέγγιση, εάν είναι χωρίς επίβλεψη ή ημι-επίβλεψη και εάν το γράφημα είναι στατικό ή δυναμικό. Αυτά θα καθορίσουν εάν η ανίχνευση βασίζεται στη δομή, στην κοινότητα, στην ομαδοποίηση, ή στην σχεσιακή μάθηση.

Ο συνδυασμός της διαίσθησης του αναλυτή με τη μηχανική μάθηση είναι ικανός στην εκμάθηση άμυνας από επιθέσεις που δεν είχαν δει προηγουμένως . Η μάθηση

χωρίς επίβλεψη μπορεί να μάθει ένα μοντέλο για τον εντοπισμό ανωμαλιών, όπως ακραία ή σπάνια συμβάντα στον κυβερνοχώρο, τα οποία μπορούν να ταξινομηθούν και βασίζονται σε μια προκαθορισμένη μέτρηση. Το μοντέλο που προκύπτει μπορεί να προβλέψει από χαρακτηριστικά πιθανές επιθέσεις στο εγγύς μέλλον[35].

5. Βελτίωση της Απόδοσης και Κλιμάκωσης:

Η απόδοση των γραφημάτων γνώσης μπορεί να βελτιωθεί μέσω της χρήσης βελτιστοποιημένων αλγορίθμων και τεχνικών αποθήκευσης δεδομένων. Επιπλέον, η κλιμάκωση των λύσεων για την επεξεργασία μεγάλων ποσοτήτων δεδομένων δικτύου είναι κρίσιμη για τη χρήση τους σε πραγματικά περιβάλλοντα.

6. Εφαρμογή Πολυεπίπεδων Προσεγγίσεων ασφαλείας:

Η ενσωμάτωση γραφημάτων γνώσης σε ένα πολυεπίπεδο πλαίσιο ασφάλειας μπορεί να προσφέρει μεγαλύτερη προστασία. Χρησιμοποιώντας γραφήματα γνώσης σε συνδυασμό με άλλες τεχνολογίες ασφαλείας και τα δεδομένα τους, όπως τα συστήματα (IDS) και τα συστήματα ανίχνευσης ανωμαλιών (ADS), μπορεί να δημιουργηθεί ένα πιο ανθεκτικό και ολοκληρωμένο σύστημα ασφαλείας.

7. Συγκέντρωση δεδομένων και συγχώνευση δεδομένων

Τα γραφήματα γνώσης για την κυβερνοασφάλεια έχουν τεράστιες δυνατότητες όσον αφορά τη συγκέντρωση και σύντηξη δεδομένων, που είναι τυπική διαδικασία στους πίνακες ελέγχου SOC και SIEM, για παράδειγμα.

Οι πηγές δεδομένων περιλαμβάνουν, αλλά δεν περιορίζονται σε, τοπολογία δικτύου, IDS, κανόνες τείχους προστασίας, τείχος προστασίας διαχειριστής, δρομολόγησης μηνυμάτων, σαρωτής ευπάθειας, λογισμικού SIEM, πληροφοριών ασφαλείας και δημόσια διαθέσιμα σύνολα δεδομένων[35].

Αυτές οι προοπτικές και δυνατότητες βελτίωσης υπογραμμίζουν τη σημασία της συνεχούς εξέλιξης και αναβάθμισης της μεθοδολογίας που χρησιμοποιείται για την ανάλυση δεδομένων δικτύου και την ενίσχυση της κυβερνοασφάλειας. Η προσαρμογή στις νέες τεχνολογίες και η ενσωμάτωση καινοτόμων προσεγγίσεων μπορεί να οδηγήσει σε πιο αποτελεσματικά και αξιόπιστα συστήματα ανίχνευσης και αντιμετώπισης απειλών.

ΚΕΦΑΛΑΙΟ 6: ΕΠΙΛΟΓΟΣ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Στην παρούσα εργασία, έγινε διερεύνηση της αξιοποίησης των δυνατοτήτων των γραφημάτων γνώσης για τη βελτίωση της ανάλυσης δεδομένων δικτύου TCP/IP στο πλαίσιο ενίσχυσης της κυβερνοασφάλειας. Το εξεταζόμενο αντικείμενο της εργασίας (χρήση γραφημάτων γνώσης στην κυβερνοασφάλεια) δεν τυγχάνει ευρείας ακαδημαϊκής επεξεργασίας και έτσι συναντήθηκαν αρκετές ενδιαφέρουσες προκλήσεις και ευκαιρίες μάθησης στα εξεταζόμενα επιμέρους επιστημονικά πεδία. Η εργασία έχει επιμερισθεί συνολικά σε έξι κεφάλαια και τρία υποστηρικτικά παραρτήματα.

Στο κεφάλαιο 1 τέθηκαν οι στόχοι της διπλωματικής εργασίας για: ανάπτυξη μεθόδου, ανίχνευσης απειλών και εντοπισμού ανωμαλιών στον κυβερνοχώρο με μεγαλύτερη ακρίβεια και ταχύτητα, υλοποίηση και ενσωμάτωση των γραφημάτων γνώσης με δεδομένα δικτύου TCP/IP, ένταξη της προτεινόμενης μεθόδου στο ευρύτερο πλαίσιο της αρχιτεκτονικής κυβερνοασφάλειας που χρησιμοποιείται σήμερα και τέλος την πρόταση μελλοντικών βελτιώσεων της προτεινόμενης μεθόδου.

Στο κεφάλαιο 2 παρουσιάστηκε το θεωρητικό υπόβαθρο της εργασίας. Η αξιοποίηση δεδομένων TCP/IP είναι θεμελιώδης για την ασφάλεια στον κυβερνοχώρο, επιτρέποντας την παρακολούθηση σε πραγματικό χρόνο και την ιστορική ανάλυση των δραστηριοτήτων του δικτύου. Η ανάπτυξη της τεχνολογίας των ontologies στα πληροφοριακά συστήματα επιτρέπει τη χρησιμοποίησή τους στην κυβερνοασφάλεια και κυρίως για αναπαράσταση πηγών πληροφοριών, δημιουργία κίνησης δεδομένων, ανταλλαγή και διαχείριση γνώσης. Τα γραφήματα γνώσης χρησιμοποιούν ένα μοντέλο δεδομένων που βασίζεται σε γράφημα για να συλλάβουν τη γνώση σε σενάρια εφαρμογών που περιλαμβάνουν ενσωμάτωση, διαχείριση και εξαγωγή αξίας από διαφορετικές πηγές δεδομένων σε μεγάλη κλίμακα. Παρουσιάζονται διάφορα μοντέλα γράφων τα οποία συναντώνται στην πράξη.

Στο κεφάλαιο 3 γίνεται ο σχεδιασμός και η ανάλυση της μεθόδου (αρχιτεκτονικής) που χρησιμοποιήθηκε για την ανίχνευση απειλών και εντοπισμού ανωμαλιών στον κυβερνοχώρο. Η αρχιτεκτονική περιλαμβάνει πέντε επιμέρους στοιχεία: Δημιουργία και περιγραφή των δεδομένων, ανάπτυξη του ontology, ανάπτυξη του script για την

[67]

αντιστοίχιση (annotation), κατασκευή των γραφημάτων γνώσης (knowledge graphs) και τέλος περιγραφή των queries για την ανάλυση των γραφημάτων γνώσης. Τα εργαλεία που σχεδιάστηκαν, δοκιμάστηκαν και χρησιμοποιήθηκαν στην εργασία είναι: το Protégé για το Ontology, το IntelliJ IDEA με γλώσσα Python για το Sript annotation, το Neo4j για την κατασκευή του Knowledge Graph και η γλώσσα Cypher για τα queries.

Το κεφάλαιο 4 περιλαμβάνει την υλοποίηση της αρχιτεκτονικής που σχεδιάστηκε σε αντίστοιχη προσέγγιση πέντε επιμέρους στοιχείων, όπως έγινε η σχεδίαση της μεθόδου. Το μοντέλο αφορά μια υποθετική εταιρεία Fage η οποία έχει 4 τμήματα που εδρεύουν σε διαφορετικές περιοχές της Ελλάδας και επικοινωνεί με το εξωτερικό και το εσωτερικό. Έγινε η εισαγωγή, επεξεργασία και επικύρωση των δεδομένων, εν συνεχεία η υλοποίηση του script και η αντιστοίχιση δεδομένων με βάση το δημιουργημένο ontology, η υλοποίηση του knowledge graph και τέλος δημιουργήθηκαν μέσω queries τρία σενάρια ανάλυσης κίνησης. Το πρώτο για εντοπισμό ασυνήθιστων απειλών από εξωτερικές IP, το δεύτερο για συσχέτιση εσωτερικών και εξωτερικών διευθύνσεων IP και το τρίτο για καταγραφή ιστορικού συνδέσεων της κεντρικής IP. Κάθε σενάριο αναλύεται ποσοτικά, οπτικά και ποιοτικά και παρουσιάζονται τα ευρήματα.

Το κεφάλαιο 5 περιλαμβάνει την ποιοτική αξιολόγηση των ευρημάτων και την επίδραση στην κυβερνοασφάλεια. Επιπρόσθετα αναφέρονται οι προοπτικές βελτίωσης της μεθοδολογίας. Από την αξιολόγηση καταγράφονται η επιτυχής επιλογή της αρχιτεκτονικής των πέντε στοιχείων, η αποτελεσματικότητα των γραφημάτων γνώσης στην ανίχνευση απειλών ενός δικτύου επικοινωνίας TCP/IP, η ταχύτερη και ακριβέστερη ανίχνευση και απόκριση σε απειλές στον κυβερνοχώρο, η εύρεση κενών ασφαλείας, η δυνατότητα ενσωμάτωσης ως εργαλείο SIEM σε υπάρχοντα πλαίσια κυβερνοασφάλειας και το πλεονέκτημα της οπτικοποίησης της μεθόδου.

Προοπτικές βελτίωσης και εξέλιξης της εργασίας σαφώς και υφίστανται κυρίως όσον αφορά στη διερεύνηση και ανάπτυξη των εργαλείων που μπορούν να χρησιμοποιηθούν στα επιμέρους στοιχεία (δεδομένα, οντολογίες, script, γραφήματα γνώσης, queries), παρόλο που τα εργαλεία που χρησιμοποιήθηκαν κρίνονται αρκετά

ικανοποιητικά. Όμως ως αρχιτεκτονική παρουσιάζει συνεκτικότητα και βάσιμη αποτελεσματικότητα. Έτσι θα μπορούσαν να εξεταστούν η πραγματική κίνηση δεδομένων (π.χ. Πανεπιστημίου Πατρών), η χρήση εργαλείων όπως το Fluent Editor και το NeOn Toolkit για το ontology και η χρήση άλλων πλατφόρμων για το Knowledge Graph όπως Amazon Neptune, Microsoft Azure Cosmos DB και GraphDB. Επιπρόσθετα μπορούν να δημιουργηθούν περισσότερα σύνθετα σενάρια (queries).

Ακόμη, προτείνεται η εξερεύνηση Νέων αλγορίθμων και μεθόδων Μηχανικής Μάθησης. Η μελλοντική έρευνα αυτή θα μπορούσε να επικεντρωθεί στην ανάπτυξη και δοκιμή νέων αλγορίθμων μηχανικής μάθησης, όπως η βαθιά μάθηση (deep learning) και η ενισχυτική μάθηση (reinforcement learning), για την ενίσχυση της ικανότητας των γραφημάτων γνώσης να ανιχνεύουν και να προβλέπουν κακόβουλες δραστηριότητες στο δίκτυο TCP/IP.

Γενικότερα, επιτεύχθηκαν οι στόχοι που τέθηκαν με σημαντικότερο όλων την σχεδίαση και υλοποίηση της αρχιτεκτονικής και την επιτυχή ενσωμάτωση και υλοποίηση των γραφημάτων γνώσης (Knowledge Graphs) στην Ανάλυση Δεδομένων Δικτύου TCP/IP για την Ενίσχυση της Κυβερνοασφάλειας. Τα γραφήματα γνώσης ενισχύουν την κυβερνοασφάλεια στο επίπεδο της διαχείρισης γεγονότων (event management), οπότε ένα ακόμα SIEM εργαλείο μπορεί να σχεδιασθεί και χρησιμοποιηθεί από τους ειδικούς της κυβερνοασφάλειας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Amirah Alshammari, A. A. (2021, June 14). Ανάκτηση από <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00475-1>
- [2] Ang Liu, D. Z. (2022). *Knowledge graph with machine learning for product design*. Ανάκτηση από <https://www.sciencedirect.com/science/article/abs/pii/S0007850622000208>
- [3] Arne Seeliger, M. P. (2019). *Semantic Web Technologies for Explainable*. Ανάκτηση από https://ceur-ws.org/Vol-2465/semex_paper1.pdf
- [4] Bhardwaj, P. (2022, July 19). *What is Network Traffic Analysis in Cybersecurity?* Ανάκτηση από <https://www.tutorialspoint.com/what-is-network-traffic-analysis-in-cybersecurity>
- [5] Dave Voutila, G. B. (2022, Apr 29). *GRAPHS FOR CYBERSECURITY*. Ανάκτηση από <https://go.neo4j.com/rs/710-RRC-335/images/Neo4j-Graphs-for-Cybersecurity-Whitepaper.pdf>
- [6] David Zhao, I. T. (2013, November). *Botnet detection based on traffic behavior analysis and flow intervals*. Ανάκτηση από <https://www.sciencedirect.com/science/article/abs/pii/S0167404813000837?via%3Dihub>
- [7] *Developing an Ontology for Cyber Security Knowledge Graphs*. (2015, April 1). Ανάκτηση από <https://www.osti.gov/servlets/purl/1424501>
- [8] Ga Young Lee, L. A. (2021, April 13). *A Survey on Data Cleaning Methods for Improved Machine Learning*. Ανάκτηση από <https://arxiv.org/pdf/2109.07127>
- [9] Gabriel Arquela Pimenta Rodrigues, R. D. (2017, October 18). *Cybersecurity and Network Forensics: Analysis of Malicious Traffic towards a Honeynet with Deep Packet Inspection*. Ανάκτηση από <https://www.mdpi.com/2076-3417/7/10/1082>

- [10] Jiaqi Liu, Q. Z. (2019). *Evolving Knowledge Graphs*. Ανάκτηση από https://www.researchgate.net/publication/333848251_Evolving_Knowledge_Graphs
- [11] Jr, J. H. (2012). *Cybersecurity . . . How Important Is It?* . Ανάκτηση από https://www.americanbar.org/content/dam/aba/publications/judges_journal/vol51no4-jj2012-tech.pdf
- [12] Kadobayashi, T. T. (2014, October 8). *Reference Ontology for Cybersecurity*. Ανάκτηση από <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8205615>
- [13] Kai Liu, F. W. (2022, July 22). *Recent Progress of Using Knowledge Graph for Cybersecurity*. Ανάκτηση από <https://www.mdpi.com/2079-9292/11/15/2287>
- [14] Kejriwal, M. (2022, March 23). *Knowledge Graphs: A Practical Review of the Research Landscape*. Ανάκτηση από <https://www.mdpi.com/2078-2489/13/4/161>
- [15] Leo Obrst, P. C. (2012). *Developing an Ontology of the*. Ανάκτηση από https://d1wqtxts1xzle7.cloudfront.net/93768230/STIDS2012_T06_ObrstEtAl_CyberOntology-libre.pdf?1667756195=&response-content-disposition=inline%3B+filename%3DDeveloping_an_Ontology_of_the_Cyber_Secu.pdf&Expires=1722003435&Signature=QRoVT4MHb~Ljo8aj1Ryih9oJ
- [16] Lupsan, S. (2022, June 13). *Creating a Security Knowledge Graph™ Through Integrations*. Ανάκτηση από <https://cyscale.com/blog/security-knowledge-graph-integrations>
- [17] Michael, I. (2015, April). *Developing an Ontology for Cyber Security Knowledge Graphs*. Ανάκτηση από <https://www.osti.gov/servlets/purl/1424501>
- [18] Networks, H. g. (χ.χ.). *David Topps, Corey Wirun, Rachel Ellaway*. Ανάκτηση από <https://olab.ca/heterogeneous-graphs-and-graph-neural-networks/>

- [19] Noemi Scarpato, N. D. (2019). *Reachability Matrix Ontology: A Cybersecurity*.
Ανάκτηση από <https://www.tandfonline.com/doi/pdf/10.1080/08839514.2019.1592344>
- [20] Oltramari, A. (2014). *Building an Ontology of Cyber Security*. Ανάκτηση από <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=35904549cad6196f7fedcb769e24f9eaf900ab35>
- [21] Padia, Z. S. (2016). *UCO: A Unified Cybersecurity Ontology*. Ανάκτηση από <https://cdn.aaai.org/ocs/ws/ws0163/12574-57427-1-PB.pdf>
- [22] Pramatarov, M. (2023, November 22). *TCP (Transmission Control Protocol) – What is it, and how does it work?* . Ανάκτηση από <https://www.cloudns.net/blog/tcp-transmission-control-protocol-what-is-it-and-how-does-it-work/>
- [23] (Cubriilo, 2015)R.Sivakumar. (2011, August). *ONTOLOGY VISUALIZATION PROTÉGÉ TOOLS* —. Ανάκτηση από https://idc-online.com/technical_references/pdfs/information_technology/ONTOLOGY%20VISUALIZATION.pdf
- [24] Qin, Y. (2020, August). Research and Application of Knowledge Graph in Teaching: Take the database course as an example. Ανάκτηση από https://www.researchgate.net/figure/Knowledge-graph-constructionsteps_fig1_343718462
- [25] Vetro, G. F. (2020, February 22). *On the Integration of Knowledge Graphs into Deep Learning Models for a More Comprehensible AI—Three Challenges for Future Research*. Ανάκτηση από <https://www.mdpi.com/2078-2489/11/2/122>
- [26] Weiss, J. (2018, September). *cdw*. Ανάκτηση από <https://www.cdw.com/content/cdw/en/articles/datacenter/why-traditional-security-techniques-are-no-longer-enough.html>
- [27] *What is a graph database?* (χ.χ.). Ανάκτηση από <https://neo4j.com/docs/getting-started/get-started-with-neo4j/graph-database/#:~:text=Neo4j%20is%20a%20native%20graph,on%20top%20of%2>

0another%20technology.

- [28] *What is Network Traffic Analysis (NTA)?* (χ.χ.). Ανάκτηση από <https://www.rapid7.com/fundamentals/network-traffic-analysis/>
- [29] Wiens, C. (2022, August 31). *WHAT IS NETWORK TRAFFIC ANALYSIS? A BEGGINER'S GUIDE*. Ανάκτηση από <https://mixmode.ai/blog/updated-for-2022-what-is-network-traffic-analysis-a-beginners-guide/>
- [30] Wilcke, X. (2017, December 8). *The knowledge graph as the default data model for learning on heterogeneous knowledge*. Ανάκτηση από <https://content.iospress.com/articles/data-science/ds007>
- [31] Yankulov, M. (2020, July 9). *Boosting Cybersecurity Efficiency with Knowledge Graphs*. Ανάκτηση από <https://www.ontotext.com/blog/boosting-cybersecurity-efficiency-with-knowledge-graphs/>
- [32] Zuoxu, L. X. (2021). *Exploiting Knowledge Graphs in Industrial Products and Services*. Ανάκτηση από https://ira.lib.polyu.edu.hk/bitstream/10397/91592/1/Li_Exploiting_knowledge_graphs.pdf
- [33] Μόσχου, Μ. (2022, June 30). *Πασχαλίδης (Neurosoft): Εκρηξη κυβερνοεπιθέσεων στην Ελλάδα*. Ανάκτηση από <https://www.euro2day.gr/specials/interviews/article/2139657/pashalidhs-neurosoft-ekrhxh-kyvernoepitheseon-sthn.html>
- [34] Mirko Cubrilo (2015). *Ontology in Information Security* . Ανάκτηση από <https://hrcak.srce.hr/file/220279>
- [35] Leslie F.Siko (2023). *Cybersecurity knowledge graph* . Ανάκτηση από <https://link.springer.com/article/10.1007/s10115-023-01860-3>

ΠΑΡΑΡΤΗΜΑ Α-ΑΡΧΕΙΑ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV

- Το αρχείο με τα δεδομένα γεωγραφικής θέσης ip_geolocation_data.csv

```
IPRangeStart,IPRangeEnd,Country,Region,City,Latitude,Longitude,ISP
192.168.0.0,192.168.0.255,United States,California,Los Angeles,34.0522,-118.2437,ISP1
102.38.248.0,102.38.248.64,Greece,Athens,Attiki,34.65107,94.347015,ISP2
102.38.248.65,102.38.248.100,Greece,Patra,Achaia,35.65107,96.347015,ISP2
102.38.248.101,102.38.248.160,Greece,Thessaloniki,Thessaloniki,36.65107,95.347015,ISP3
102.38.248.161,102.38.248.255,Greece,Hraklio,Crete,37.65107,97.347015,ISP4
172.16.0.0,172.16.0.255,United Kingdom,England,London,51.5074,-0.1278,ISP3
203.0.113.0,203.0.113.255,Australia,New South Wales,Sydney,-33.8688,151.2093,ISP4
198.51.100.0,198.51.100.255,Japan,Tokyo,Tokyo,35.6895,139.6917,ISP5
192.0.2.0,192.0.2.255,Germany,Bavaria,Munich,48.1351,11.5820,ISP6
192.88.99.0,192.88.99.255,India,Maharashtra,Mumbai,19.0760,72.8777,ISP7
100.64.0.0,100.64.0.255,Brazil,São Paulo,São Paulo,-23.5505,-46.6333,ISP8
198.18.0.0,198.18.0.255,France,Île-de-France,Paris,48.8566,2.3522,ISP9
169.254.0.0,169.254.0.255,South Africa,Gauteng,Johannesburg,-26.2041,28.0473,ISP10
103.106.3.0,103.106.3.255,Kazakhstan,CentralAstana,Astana,51.1655,71.4272,ISP7
```

- Το αρχείο με τα δεδομένα εταιρειών company_data.csv

```
company_name,department,department_location,department_la
n_address
TechCorp,TechCortpIT,New York,192.168.0.1
Japanios,JapaniosHR,Japan,198.51.100.1
Brazilero,BrazileroSales,São Paulo,100.64.0.1
BizInc,BizIncFinance,Austin,192.168.1.4
WebSol,WebSolMarketing,Paris,198.18.0.2
Fage,FageHeadquarters,Kifissia,102.38.248.50
Fage,FageIT,Marousi,102.38.248.51
Fage,FageHR,Galatsi,102.38.248.52
Fage,FageFinance,Aigio,102.38.248.66
Fage,FageMarketing,Athens,102.38.248.49
Terkenlis,TerkenlisMarketing,Thessaloniki,102.38.248.102
Terkenlis,TerkenlisIT,Thessaloniki,102.38.248.103
Terkenlis,TerkenlisFinance,Thessaloniki,102.38.248.104
Atlas,AtlasDatabase,Patissia,102.38.248.1
Atlas,AtlasHeadquarters,Boula,102.38.248.2
Nounou,NounouSales,Gouba,102.38.248.67
Nounou,NounouIT,Aigio,102.38.248.68
Nounou,NounouMarketing,Glyfada,102.38.248.53
Milkers,MilkersOperations
management,Brillisia,102.38.248.54
Milkers,MilkersFinance,Heraklio,102.38.248.162
Milkers,MilkersIT,Athens,102.38.248.55
```

- Το αρχείο με τα δεδομένα κίνησης TCP/IP tcp_traffic_data.csv

timestamp	connection_id	source_ip	destination_ip	source_port	destination_port	protocol	size
2023-05-01							
10:00:00	3ace4f4ea	192.168.0.1	102.38.248.50	443	123	IP	128
2023-05-01							
10:05:00	974ec5991	102.38.248.54	102.38.248.50	23	80	HTTP	256
2023-05-01							
10:10:00	6b53d	102.38.248.50	100.64.0.1	443	80	HTTPS	512
2023-05-01							
10:15:00	1b439c9	198.51.100.2	192.168.1.3	53	161	IP	1024
2023-05-01							
10:20:00	74g00d7	102.38.248.51	102.38.248.50	123	22	NTP	2048
2023-05-01							
10:25:00	7g10d7	102.38.248.51	203.0.113.2	443	22	SSH	4096
2023-05-01							
10:30:00	2345rf	102.38.248.60	102.38.248.50	443	22	SSH	2048
2023-05-01							
10:35:00	f34	102.38.248.60	102.38.248.50	53	22	DNS	2048
2023-05-01							
10:40:00	bre987bre	102.38.248.102	102.38.248.49	53	123	DNS	1024
2023-05-01							
10:45:00	ve34	102.38.248.49	102.38.248.53	80	123	TCP	1024
2023-05-01							
10:50:00	vdsv5	102.38.248.54	102.38.248.49	34	5	HTTPS	512
2023-05-01							
10:55:00	vas2	102.38.248.50	102.38.248.66	123	22	NTP	1024
2023-05-01							
11:00:00	32v43	102.38.248.66	102.38.248.50	443	80	HTTP	256
2023-05-01							
11:05:00	3fev3	102.38.248.51	102.38.248.103	161	22	SNMP	512
2023-05-01							
11:10:00	b66f	102.38.248.67	102.38.248.51	443	22	SSH	1024
2023-05-01							
11:15:00	ldsrf94	102.38.248.49	102.38.248.55	443	22	SSH	256
2023-05-01							
11:20:00	vni2309	102.38.248.52	102.38.248.55	80	22	HTTP	4096
2023-05-01							
11:25:00	vdshj90234	102.38.248.68	102.38.248.49	443	22	SSH	64
2023-05-01							

11:30:00,sdf98h,102.38.248.53,102.38.248.51,161,22,SNMP,256
2023-05-01
11:35:00,vd328,102.38.248.66,102.38.248.2,53,22,DNS,512
2023-05-01
11:40:00,v890432,102.38.248.102,102.38.248.50,443,54,SSH,256
2023-05-01
11:45:00,kl9iv,102.38.248.49,102.38.248.52,443,123,NTP,128
2023-05-01
11:50:00,fsd8932,102.38.248.66,102.38.248.1,443,22,TCP,256
2023-05-01
11:55:00,vds908,102.38.248.51,102.38.248.104,161,22,SNMP,1024
2023-05-01
12:00:00,fwe80v,102.38.248.52,102.38.248.50,443,22,SSH,64
2023-05-01
12:05:00,f12380v,102.38.248.52,102.38.248.54,53,22,DNS,512
2023-05-01
12:10:00,f1ca80v,102.38.248.162,102.38.248.54,443,22,SSH,128
2023-05-01
12:15:00,f1caads0v,102.38.248.55,102.38.248.51,123,22,IP,256
2023-05-01
12:20:00,f219c,102.38.248.2,102.38.248.51,161,22,SNMP,512
2023-05-01
12:25:00,10cv89,169.254.0.1,102.38.248.66,443,22,Kerberos,8192
2023-05-01
12:30:00,10c289,169.254.0.2,102.38.248.66,9444,22,SAML,8192
2023-05-01
12:35:00,10c489,169.254.0.3,102.38.248.66,9444,22,SAML,8192
2023-05-01
12:40:00,11fv4,102.38.248.52,102.38.248.51,123,80,HTTP,256
2023-05-01
12:45:00,54g,198.51.100.5,102.38.248.50,3306,123,MySQL,128
2023-05-01
12:50:00,25f,102.38.248.66,102.38.248.49,25,123,SMTP,1024
2023-05-01
12:55:00,39v,102.38.248.2,102.38.248.50,67,23,DHCP,1024
2023-05-01
13:00:00,09231bf,102.38.248.55,102.38.248.51,443,80,HTTPS,512

```
2023-05-01
13:05:00,17cac,198.18.0.2,102.38.248.50,20,123,IP,256
2023-05-01
13:10:00,vh8,102.38.248.50,198.18.0.2,20,80,TCP,256
2023-05-01
13:15:00,v9023h8,102.38.248.50,102.38.248.53,23,22,SSH,51
2
2023-05-01
13:20:00,czkml,102.38.248.51,102.38.248.1,53,23,DNS,512
2023-05-01
13:25:00,90svd,102.38.248.102,102.38.248.49,161,80,SNMP,5
12
2023-05-01
13:30:00,890vus,102.38.248.66,102.38.248.68,443,25,TLS,25
6
2023-05-02
06:00:00,ast54,103.106.3.4,102.38.248.52,22,123,SSH,512
2023-05-02
06:05:00,ast55,103.106.3.4,102.38.248.52,33893,123,RDP,20
48
2023-05-02
06:10:00,ast56,103.106.3.4,102.38.248.52,21,20,FTP,4096
2023-05-02
02:15:00,vsdi983,198.51.100.8,102.38.248.50,20,123,FTP,10
24
2023-05-01
03:15:00,karv4,198.51.100.9,102.38.248.50,3306,123,MySQL,
128
2023-05-03
23:55:00,kabrrv4,102.38.248.66,198.51.100.234,20,53,DNS,5
12
2023-05-03
13:55:00,kabrvr4,102.38.248.52,102.38.248.162,20,53,DNS,
512
```

ΠΑΡΑΡΤΗΜΑ Β - Η ΔΟΜΗ ΤΟΥ ONTOLOGY ΣΕ ΜΟΡΦΗ .TTL

```
@prefix :  
<http://www.semanticweb.org/user/ontologies/2024/4/diplom  
av3/> .  
@prefix owl: <http://www.w3.org/2002/07/owl#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-  
ns#> .  
@prefix xml: <http://www.w3.org/XML/1998/namespace> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@base  
<http://www.semanticweb.org/user/ontologies/2024/4/diplom  
av3/> .  
  
<http://www.semanticweb.org/user/ontologies/2024/4/diplom  
av3> rdf:type owl:Ontology .  
  
#####  
#####  
#      Object Properties  
#####  
#####  
  
###  
http://www.semanticweb.org/user/ontologies/2024/4/diploma  
v3#BELONGS  
:BELONGS rdf:type owl:ObjectProperty ;  
rdfs:domain :IPAddress ;  
rdfs:range :Geolocation .  
  
###  
http://www.semanticweb.org/user/ontologies/2024/4/diploma  
v3#BELONGS_TO  
:BELONGS_TO rdf:type owl:ObjectProperty ;  
rdfs:domain :IPAddress ;  
rdfs:range :Department .  
  
###  
http://www.semanticweb.org/user/ontologies/2024/4/diploma  
v3#HAS_DESTINATION_IP  
:HAS_DESTINATION_IP rdf:type owl:ObjectProperty ;
```

```

rdfs:domain :ConnectionID ;
rdfs:range :IPAddress .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#HAS_LAN_ADDRESS
:HAS_LAN_ADDRESS rdfs:type owl:ObjectProperty ;
rdfs:domain :Department ;
rdfs:range :LANAddress .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#HAS_SIZE
:HAS_SIZE rdfs:type owl:ObjectProperty ;
rdfs:domain :ConnectionID ;
rdfs:range :Size .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#HAS_SOURCE_IP
:HAS_SOURCE_IP rdfs:type owl:ObjectProperty ;
rdfs:domain :ConnectionID ;
rdfs:range :IPAddress .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#LOCATED_IN
:LOCATED_IN rdfs:type owl:ObjectProperty ;
rdfs:domain :Department ;
rdfs:range :OfficeLocation .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#USES_PROTOCOL
:USES_PROTOCOL rdfs:type owl:ObjectProperty ;
rdfs:domain :ConnectionID ;
rdfs:range :Protocol .

#####
#####
#      Data properties
#####
#####

###

```



```

http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#City
:City rdf:type owl:DatatypeProperty ;
rdfs:domain :Geolocation ;
rdfs:range xsd:string .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Country
:Country rdf:type owl:DatatypeProperty ;
rdfs:domain :Geolocation ;
rdfs:range xsd:string .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#ISP
:ISP rdf:type owl:DatatypeProperty ;
rdfs:domain :Geolocation .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Latitude
:Latitude rdf:type owl:DatatypeProperty ;
rdfs:domain :Geolocation .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Longitude
:Longitude rdf:type owl:DatatypeProperty ;
rdfs:domain :Geolocation .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Region
:Region rdf:type owl:DatatypeProperty ;
rdfs:domain :Geolocation ;
rdfs:range xsd:string .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#protocol
:protocol rdf:type owl:DatatypeProperty ;
rdfs:domain :Protocol ;
rdfs:range xsd:string .

```

```

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#size
:size rdf:type owl:DatatypeProperty ;
rdfs:domain :Size ;
rdfs:range xsd:int .

#####
#####
#      Classes
#####
#####

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Company
:Company rdf:type owl:Class .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#ConnectionID
:ConnectionID rdf:type owl:Class .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Department
:Department rdf:type owl:Class .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#Geolocation
:Geolocation rdf:type owl:Class .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#IPAddress
:IPAddress rdf:type owl:Class .

###
http://www.semanticweb.org/user/ontologies/2024/4/diploma
v3#LANAddress
:LANAddress rdf:type owl:Class .

```

```
###  
http://www.semanticweb.org/user/ontologies/2024/4/diploma  
v3#OfficeLocation  
:OfficeLocation rdf:type owl:Class .
```

```
###  
http://www.semanticweb.org/user/ontologies/2024/4/diploma  
v3#Protocol  
:Protocol rdf:type owl:Class .
```

```
###  
http://www.semanticweb.org/user/ontologies/2024/4/diploma  
v3#Size  
:Size rdf:type owl:Class .
```

ΠΑΡΑΡΤΗΜΑ Γ-ΑΚΡΩΝΥΜΙΑ

ADS - Active Directory Services

API - Application Programming Interface

APT - Advanced Persistent Threat

CSV – Comma Separated Values

DDOS - Distributed Denial Of Service

DPI - Deep Packet Inspection

GDPR - General Data Protection Regulation

GRC - Governance Risk Compliance

HIPAA - Health Insurance Portability and Accountability Act

ICT - Information and Communications Technology

IDS - Intrusion Detection System

IEC - International Electrotechnical Commission

IETF - Internet Engineering Task Force

IOT - Internet Of Things

IP - Internet Protocol

ISO - International Organization for Standardization

ISP - Internet Service Provider

ITU - International Telecommunication Union

LAN - Local Area Network

NIST - National Institute of Standards and Technology

NTA - Network Traffic Analysis

OKBC - Open Knowledge Base Connectivity protocol

OWL - Web Ontology Language

PCI DSS - Payment Card Industry Data Security

RDF - Resource Description Framework

SIEM – Security Information and Event Management

SOC – Security Operation Center

SQL - Structured Query Language

SSH - Secure Shell

TCP - Transmission Control Protocol

UDP - User Datagram Protocol

UCO - Unified Cyber Ontology

XML - Extensible Markup Language