MUNI
C4E

Complex Networks in Cybersecurity: Applications and Challenges

Augmented Complex Networks - Trustworthy Analysis (ACONTA'22)

Martin Husák, husakm@ics.muni.cz

CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence (C4e)

Institute of Computer Science, Masaryk University

December 2, 2022

Presenter's Biography

RDNr. Martin Husák, Ph.D.

- Researcher at Institute of Computer Science, Masaryk University, Czech Republic
- Member of **CSIRT-MU**, university's cybersecurity team (`https://csirt.muni.cz/`)
- Formerly a visiting researcher at Florida Atlantic University, USA
- Contributor to The Honeynet Project

Research Interests

- Network security – traffic monitoring, honeypots, intrusion detection
- Operational security – **incident response**, CSIRT operations
- **Cyber situational awareness** – information sharing, **attack projection**

Outline

Section 1

Graphs and Security

Use Cases

How are graphs used in cybersecurity?

- **Attack graphs** are used for modeling attacks
- Topology graphs are used for modeling the networks we defend
- Connection graphs allow detection of malicious patterns
- Dependency graphs show critical systems and their dependencies
- **Alert correlation** can use graphs
- ... and many other applications

Use Cases

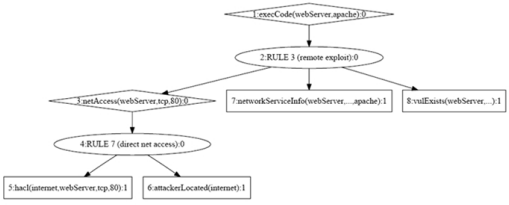
What can we model using the graphs?

- Attacks
  - Attack graphs
- Defenses
  - Network topology graphs
  - Critical missions and dependencies
- Events
  - Network connection graphs
  - Alert correlation
- Combinations of everything
  - Graph-based models for cyber situational awareness

Modeling the Attacks

Attack Graphs

- Models of attacks with many forms and existing extensions
- Useful for security assessment and strategic decisions
- More on that later in this talk



Modeling the Defenses

Network topology graphs

- Very common for networking operations, useful also for security
- Which host is connected where?

Missions and dependencies

- Enterprise missions / business processes and their dependencies
- Which hosts and service in the network are critical for the organization?
- Critical for prioritization of actions and modeling attack impacts

Modeling the Events

Network connections graphs

- Graph-based representation of network communication
- Who talked to whom?
- Useful for anomaly or intrusion detection, e.g., scanning, botnet activity

Graph-based Alert Correlation

- Attacker's action from the perspective of a defender
- Graph-based representation of relationships between alerts from IDS
- More actionable for operational cyber defense

Modeling Everything

Cyber situational awareness

- Perception of the elements in the environment,
- Comprehension of the situation,
- Projection of future state and events

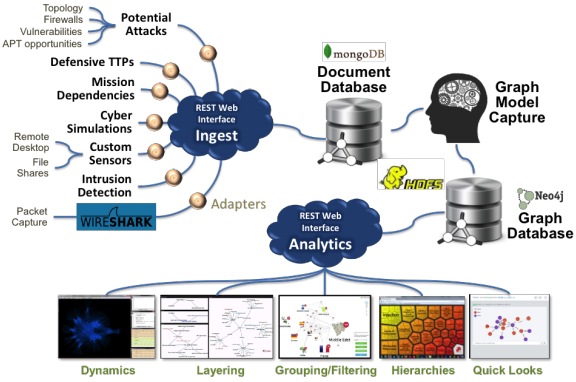
Proposed tools and models

- CyGraph, CAULDRON, ... (MITRE)
- VirtualTerrain (Rochester Institute of Technology)
- CAMUS, M2D2, and many others

Simple graphs are becoming complex networks

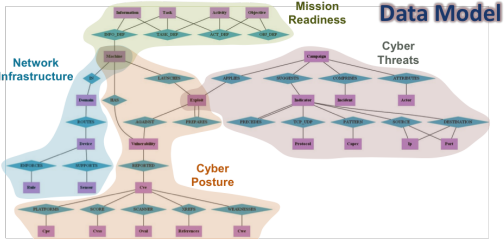
CyGraph

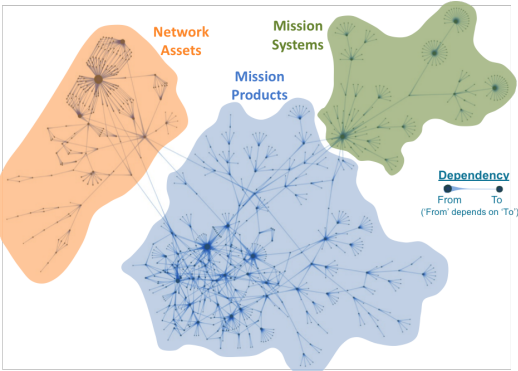
- Graph-based data model for cyber situational awareness
- Detailed representation of almost everything in the network
- Cooperates with other tools by MITRE



S. Noel et al. CyGraph: graph-based analytics and visualization for cybersecurity. In Handbook of Statistics. 2016

CyGraph





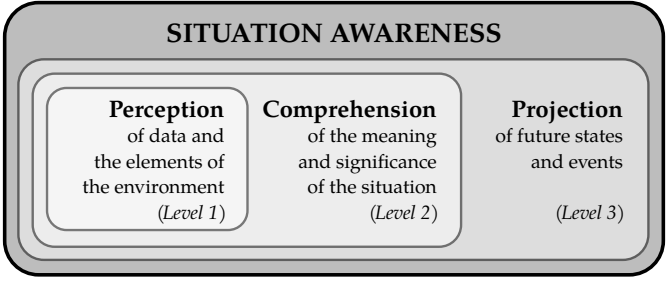
https://neo4j.com/blog/cygraph-cybersecurity-situational-awareness/

Section 2

Cyber Situational Awareness

From SA to CSA

- **Situational Awareness** (SA) is present in everyday life (sports, transport, …)
- SA was first recognized during WWI, studied in military and aviation from 1980's
- **Cyber Situational Awareness** (CSA) is the application of SA into the cyber domain



Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. In: Human Factors. 1995. 37(1).

Cyber Situational Awareness

Specifics of CSA

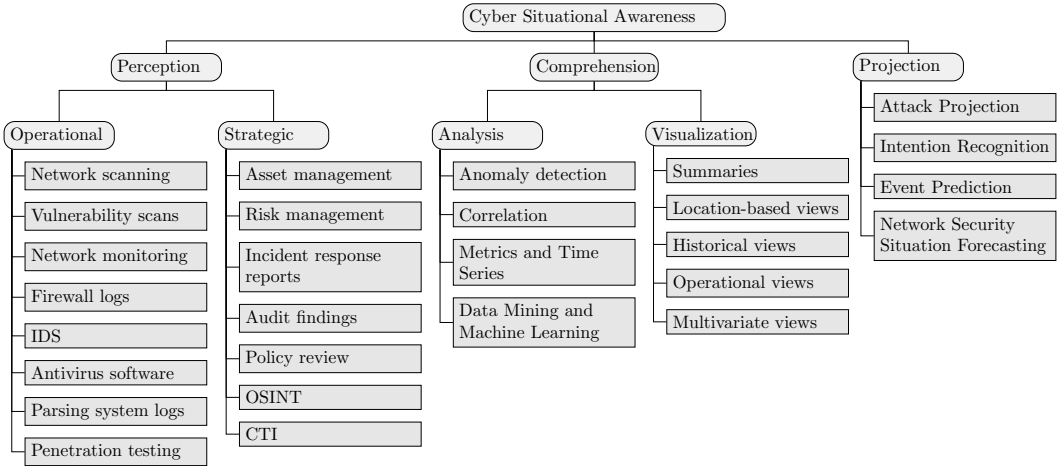
- **Cyber environment** – no borders, scale-free, everything/everywhere
- **Perception** – only sensors, no physical observations
- **Performance** – unbalanced needs of resources, high speed of events
- **Attacker takes the advantage** – in contrary to traditional military doctrine

Entities in CSA

- **Physical entities** – devices, may be characterized by roles
- **Immaterial entities** – programs and services, loosened connection to devices
- **Human entities** – characterized by roles

Husák, M., Jirsík, T., & Yang, S. J. SoK: Contemporary issues and challenges to enable cyber situational awareness for network security. In Proceedings of the 15th International Conference on Availability, Reliability and Security. 2020.

Taxonomy and Components of CSA



Based on the taxonomy proposed in Antti Evesti et al.: Cybersecurity Situational Awareness Taxonomy. In 2017 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (CyberSA)

Data Perspective

- Sensors produce massive amount of raw data that bring little understanding
  Data overload – Meaning underload
- Demand for real-time analysis and results delivery
- Emphasis on correct time ordering where **causality** is of interest
- Homogenizing data from different source and of different types
  Central processing in desired, requires proper metrics and thresholds
- Heterogeneous attack behaviors and network environments
  Often **not dependent** on what has been observed in the past

Toolset Perspective

Variety with Veracity and Volatility

- Tension between specialized tools and integrated and unified platforms
- Existing standards and taxonomies differ across organizations and countries
- Shared threat intelligence is vital, yet of low fidelity

Value through Visualization

- High noise-to-signal ratio, low value of information in CSA data
- Major issue is visualization of large-scale, dynamically changing networks
- Anticipatory CSA lacks visualization completely – uncharted scientific challenge

Performance issues

- Scalability can be met by parallelization and cloud computing
- Stream processing reduce delays and incident response times

Summary - CSA

Cyber Situational Awareness (CSA)

- Originally a theoretical concept, now a topic of applied research and development
- More and more applied research and reports from operational environment
- Interest from governments and national strategies

Challenges for CSA and CSA-supporting tools

- Coping with rising volume, variety, and velocity of the data (Big Data)
- Supporting the CSA operators with the **right data** at the **right time**
- Visualizing the data in a meaningful manner
- Maintaining sufficient performance

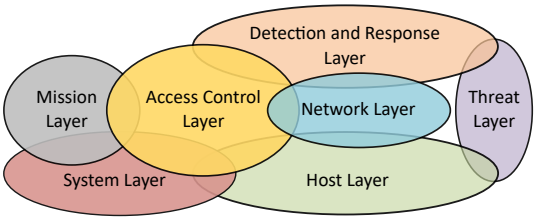
Section 3

Case Study: CRUSOE Project

CRUSOE Project

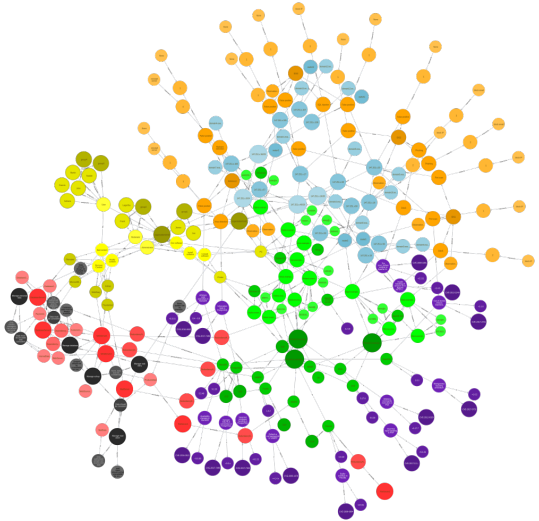
CRUSOE Project at Masaryk University

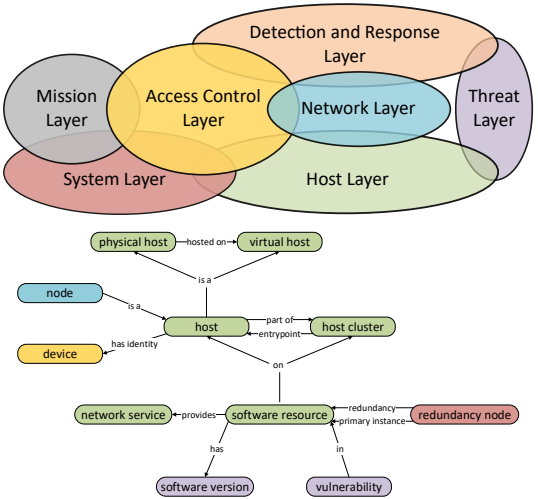
- Development of a toolset for achieving cyber situational awareness
- Inspired by CyGraph, more lightweight and automated
- Similar graph-based data model



J. Komárková et al. CRUSOE: Data Model for Cyber Situation Awareness. In Proceedings of the 13th International Conference on Availability, Reliability and Security. 2018

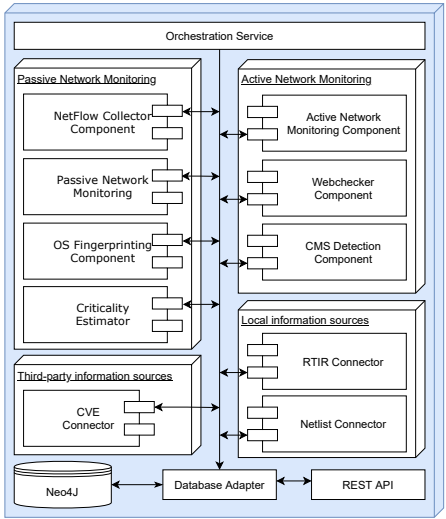
CRUSOE Project





https://github.com/CSIRT-MU/CRUSOE-Data-Model

System Design – Data Collection



Common components

- Orchestration service – Celery
- Database – Neo4j
- Database adapter & REST API

Data collection component

- Grouped by primary data
- Passive network monitoring – adapters to NetFlow monitoring infrastructure
- Active network monitoring – adapters to Nmap and other scanners
- Local and third-party sources – custom adapters to specific data and systems

Passive Network Monitoring

NetFlow collector component

- Connects to NetFlow monitoring infrastructure (collector)
- Queries NetFlow data, downloads records needed by other components

Passive network monitoring component

- OS fingerprinting – uses three methods to identify OS of communicating devices: TCP header, HTTP User-Agent, communication with specific domains (intensive ongoing research – developed separately)
- Service detection using NBAR2 signatures to identify services and software
- Web browser detection via HTTP User-Agent analysis
- Antivirus software detection via communication with specific domains

Active Network Monitoring

Active network monitoring component

- *Nmap*-based, scans 100 top ports for open services and network topology
- Complementary OS and software fingerprinting (CPE-formatted output)
- Time-consuming (16 hours in /16 network), clean-up and resume procedures

Webchecker

- Checks webservers if they provide content on port 80 or 443
- If port 443 is served, the certificate's validity is checked

CMS detection component

- Identification of CMS (*WordPress*, *Drupal*, ...) on previously discovered webservers
- Based on *WhatWeb* tool

Third-party and Local Information Source

CVE connector

- Downloads CVE records from NVD (primary) and vendors' databases (details)
- CVEs are matched with discovered software via CPE: [*CVE*] — (*CPE*) — [*Software*]

RTIR connector

- Downloads history of incidents from *Request Tracker for Incident Response*
- Incident details – timestamps, actors, status, ...

NetList connector

- Local list of network segments, IP ranges, and admin contacts:
  routers,10.0.0.0/24,networkadmin@organization
  servers,10.0.10.0/24,serveradmin@organization

Derived Information

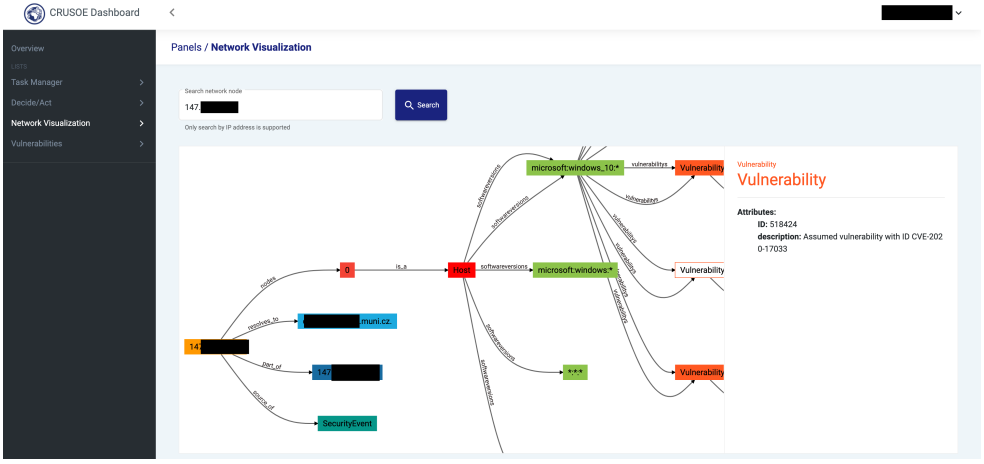
Criticality estimator

- Varying definitions of critical infrastructures, manual enumeration is too laborious
- *Critical host* = *Critical node* in the network topology graph
- *Betweenness score* – how many shortest paths go through a node?
- Nodes with the highest betweenness score are considered critical
- The topic will be expanded in future work

CPE matching

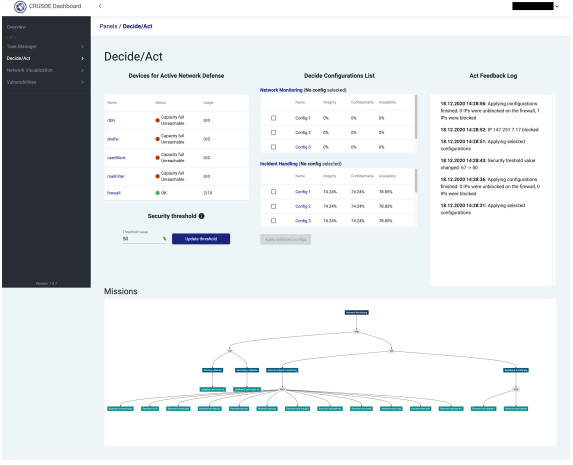
- Matching CVE to software/services is enabled via CPE
- Matches are only partial, vulnerability assessment is not exact
- Vulnerabilities are assumed, not confirmed – still sufficient for CSA

Visualization – CRUSOE Dashboard



Husák, M., Sadlek, L., Špaček, S., Laštovička, M., Javorník, M., & Komárková, J. (2022). CRUSOE: A toolset for cyber situational awareness and decision support in incident handling. Computers & Security, 115, 102609.

Visualization – CRUSOE Dashboard



Husák, M., Sadlek, L., Špaček, S., Laštovička, M., Javorník, M., & Komárková, J. (2022). CRUSOE: A toolset for cyber situational awareness and decision support in incident handling. Computers & Security, 115, 102609.

Section 4

Future Work: Graph Traversal and Target Recommendation

Motivation

Ransomware and similar threat

- The rising complexity and variety of cyberattacks complicate **incident handling**.
- IDS and secure perimeter are bypassed by **social engineering attacks**, e.g., phishing.
- The malware further **spreads in the network**, exploiting surrounding computers.
- There is little chance of mitigating the spread of infection.

Incident handling

- Rapid incident response prevents spread of infection and reduces attack impact.
- Effective **triage and prioritization** of threats and incidents are of utmost importance.
- The behavior of malware can be **anticipated** to some extent.
- Social engineering is difficult to detect – we depend on **user reports**.

Husák M. Towards a Data-Driven Recommender System for Handling Ransomware and Similar Incidents. 19th Annual IEEE International Conference on Intelligence and Security Informatics (ISI). 2021.

Proposed Approach

Anticipating the behavior of the malware

- A typical malware uses a few attack vectors and spreads in close proximity first.
- The lateral movement of an attacker can be observed, traced, and even projected.
- However, that requires detailed knowledge of the local environment and collaboration with users and administrators (complicated in large networks).
- The incident handlers would appreciate any piece of information that would guide them through the network and pinpoint nodes that are immediately threatened.
- The key question of an incident handler is:
  if this device is infected, which other devices can be infected or threatened?

General Idea

The recommendations are based on the **proximity** and **similarity** of the hosts in the network to the host on the input; similar hosts in close proximity are prioritized.

Proximity

Two hosts can be close to each other in physical and logical network topology, e.g., in the same room or in the same IP range. Alternatively, the two machines can be close to each other if they are controlled by the same users or administrators.

Similarity

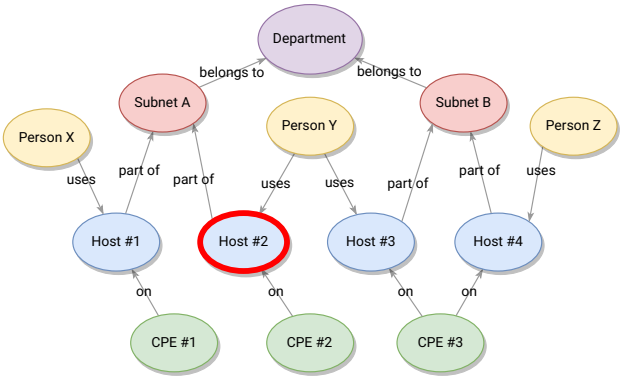
The similarity is based on the similarity in software equipment, role, profile, or shared history of the two hosts. Similarity in software equipment is a prevalent feature due to the fact that the attackers typically exploit certain services or software.

Example – Distance calculation

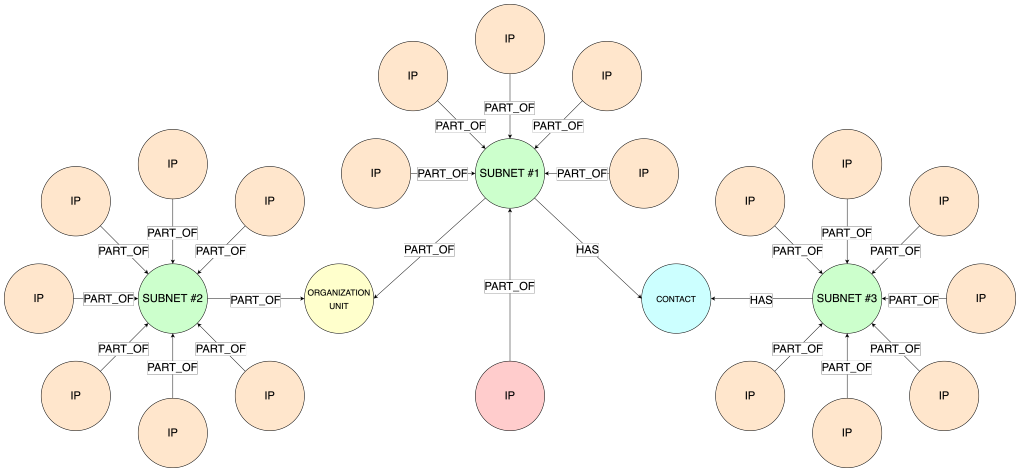
Host #2 is reported to be infected, it's distance to other hosts is:

- 2 to Host #1 (same subnet)
- 2 to Host #3 (same user)
- 4 to Host #4 (subnets belonging to the same department)

Host #4 is too far – Hosts #1 and #3 follows are possible next victims.



Example – Distance calculation



Section 5

Conclusion

Conclusion

Conclusion

- Graph-models can be found almost everywhere in cybersecurity
- Cyber Situational Awareness (CSA) – holistic views on cybersecurity
- Models grow in size and are becoming complex networks

Challenges and Future Work

- Large volumes of cybersecurity data of questionable quality
- Plethora of small tools – orchestration or building larger tools?
- Providing the right data at the right time – and comprehensively!
- Many meaningful queries to graph databases – but are the data OK?

MUNI
C4E

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

C4E.CZ