MUNI

On The Provision of Network-Wide Cyber Situational Awareness via Graph-Based Analytics

Complex Computational Ecosystems 2023 (CCE'23)

Martin Husák[1,2] **(husakm@ics.muni.cz)**
Joseph Khoury[2], **Đorđe Klisura**[2], **Elias Bou-Harb**[2]

[1] Institute of Computer Science, Masaryk University, Czech Republic

[2] The Cyber Center for Security and Analytics, The University of Texas in San Antonio, USA

April 26, 2023

Presenter's Biography

RDNr. Martin Husák, Ph.D.

- Researcher at Institute of Computer Science, Masaryk University, Czech Republic
- Member of Masaryk University's incident response team CSIRT-MU (`https://csirt.muni.cz/`)
- Currently a visiting researcher at The Cyber Center for Security and Analytics, The University of Texas at San Antonio, USA.
- Research interests include network security, incident response, and cyber situational awareness.
- This research was supported by OP JAK "MSCAfellow5_MUNI" (No. CZ.02.01.01/00/22_010/0003229).

Outline

Section 1

Graphs for Cyber Situational Awareness

What can we model in cybersecurity using graphs?

Attack Graphs

- Models of attacks with many forms and existing extensions
- Useful for security assessment and strategic decisions

Network topology graphs

- Very common for networking operations, useful also for security

Missions and dependencies

- Enterprise missions / business processes and their dependencies
- Critical for prioritization of actions and modeling attack impacts

Network connections graphs

- 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

Putting it all together – provisioning CSA

Cyber situational awareness (CSA)

- Perception of the elements in the environment,
- Comprehension of the situation – main focus of this work
- Projection of future state and events

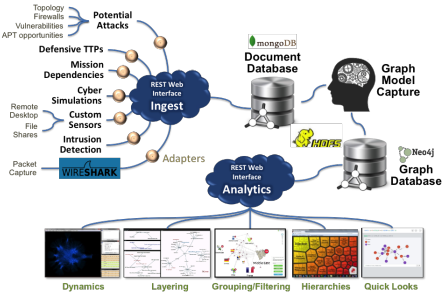
Proposed tools and models

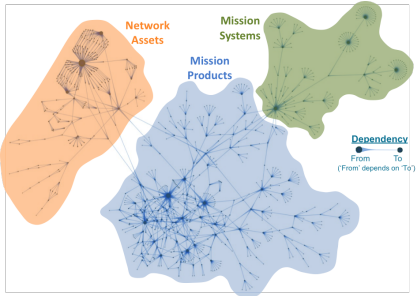
- CAULDRON, CyGraph (MITRE)
- CRUSOE (Masaryk University)
- VirtualTerrain, CAMUS, M2D2, ...

Simple graphs are becoming complex networks

Example – CyGraph (MITRE)

- Graph-based data model for cyber situational awareness
- Detailed representation of the network and security posture

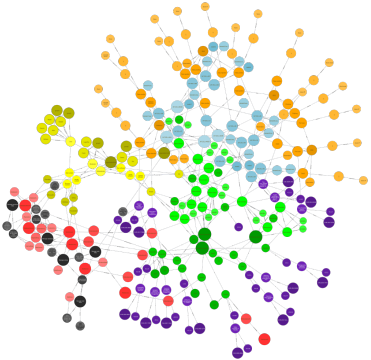


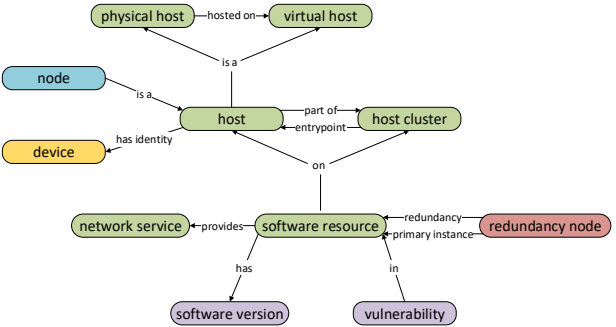


Noel, S., Harley, E., Tam, K. H., Limiero, M., & Share, M. (2016). CyGraph: graph-based analytics and visualization for cybersecurity. In Handbook of Statistics (Vol. 35, pp. 117-167). Elsevier.

Example – CRUSOE (Masaryk University)

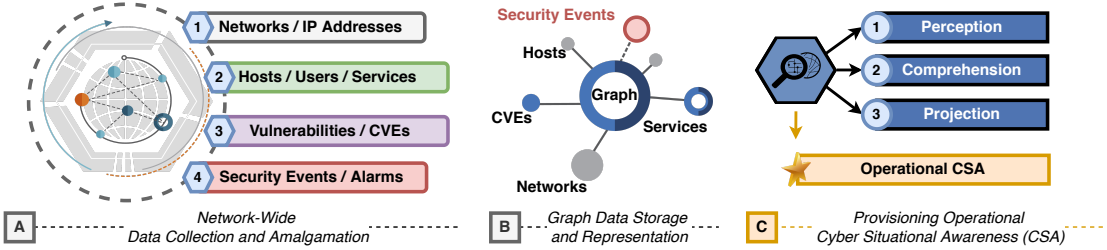
- Development of a toolset for achieving cyber situational awareness
- Inspired by CyGraph, more lightweight and automated





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.

Provisioning CSA via graph-based analytics

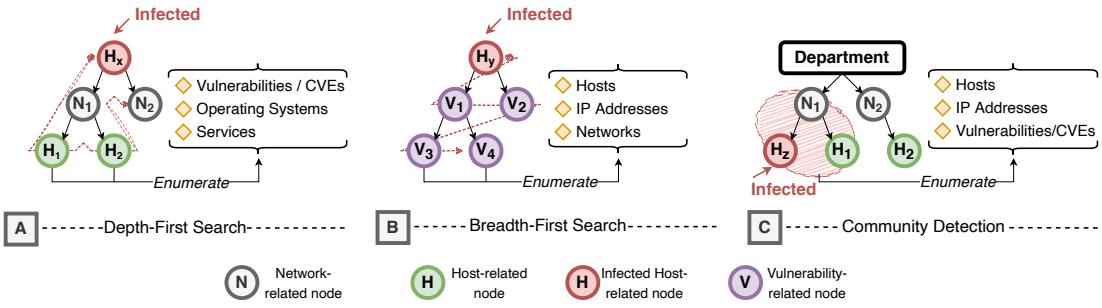


(A) collect and amalgamate network-wide data using heterogeneous tools for computer network monitoring and reconnaissance,
(B) leverage graph-based analytics to store, visualize, and query the data,
(C) leverage this data to provision operational CSA for defensive measures, incident responses, and network forensics.

Section 2

Selected Cyber Security Tasks using Graph-Based Analytics

Graph Traversal



Graph Traversal and Community Detection Algorithms: *(A)* depth-first search, *(B)* breadth-first search, *(C)* community detection algorithms applied to selected cyber security tasks.

Finding similar hosts in close proximity by graph traversal

- **Anticipating infection** – malware spreads in proximity and to similar devices
- However, that requires detailed knowledge of the local environment and collaboration with users and administrators (complicated in large networks).
- If this device is infected, which other devices can be infected or threatened?
- **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.

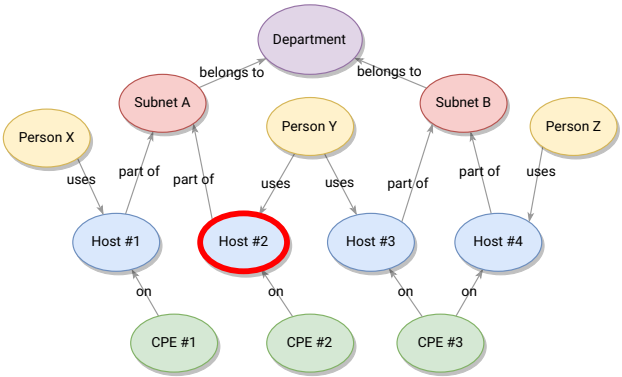
Husák M. Towards a Data-Driven Recommender System for Handling Ransomware and Similar Incidents. 19th Annual IEEE International Conference on Intelligence and Security Informatics (ISI). 2021.

Finding similar hosts in close proximity by graph traversal

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.



Network segmentation via Community Detection and FSM

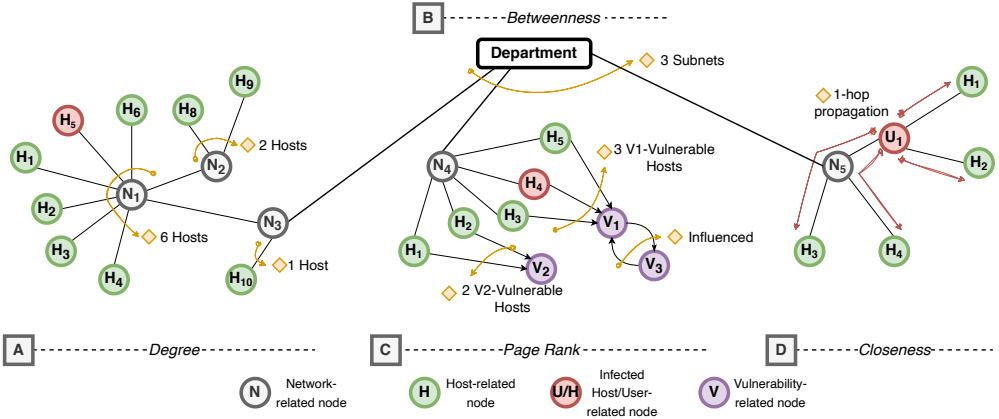
Network segmentation

- Already discussed in literature – Girvan-Newman algorithm
- Use case:
  - Graph algorithm identifies network segments
  - IDS detects an infected host in one of the segment
  - Firewalls are set to mitigate further infection from the segment

Community Detection

- Frequent Subgraph Mining (FSM) – based on DFS and BFS
- Common result – a certain set of nodes and edges occurs more frequently than we would anticipate, such as:
  - a collection of devices connecting to a particular server
  - a collection of devices using having the same vulnerabilities
  - some may even correspond to behavioral patterns of malware, e.g., botnets

Node criticality estimation via graph centrality



Graph Centrality Algorithms: *(A)* degree, *(B)* betweenness, *(C)* page rank, *(D)* closeness algorithms applied to specific cyber security tasks.

Node criticality estimation via graph centrality

How important is a particular machine for the organization?

- Answering requires knowledge of local environment
- Hard problem in large, heterogeneous environment
- Can be well approximated automatically via graph measures

Network topology graph

- Most easily created by actively scanning the network
- Common *Nmap* tool with *traceroute* option
- Several vantage points! (each produces a tree)
- Merging the trees into a topology graph

Node criticality estimation via graph centrality

Graph centrality measures on network topology or communication graphs

- Degree centrality
  - Nodes with high degree centrality are crucial to network's operation, e.g., routers
- Betweenness
  - First, calculate shortest paths between all pairs of nodes
  - Node's score is the number of shortest paths going through it
  - Best approximation of criticality of the hosts in the network
- Page rank
  - More useful in network connection graphs
  - Identifies highly demanded hosts, such as servers, helps criticality estimation
- Closeness centrality
  - Variation on the task of finding similar hosts in close proximity
  - Helps understanding malware spread

Section 3

Open Issues and Challenges

Open Issues and Challenges

Need to learn a new paradigm

- Adoption of new query languages and data processing paradigms can be slow
- Positive attitude towards the graph-based representation of cyber security data
- Similar situation as with the stream-based data analysis in the past decade

Unified ontology

- There is no common language among the adopters of these approaches
- Tools like CyGraph or CRUSOE have their own data models
- Unified ontology like UCO or STUCCO are on the way but not yet here

Open Issues and Challenges

Dataset

- As with many tasks in cybersecurity, there are no good datasets
- Existing datasets obsolete extremely fast, new datasets do not allow comparison
- Need to convert data to a new format – graph databases evolve, too

Application of Graph Neural Networks and Graph AI

- Graph Neural Networks (GNN) – already adopted by researchers
- Link prediction as an interesting technique for network analysis
- Explainability is crucial for cybersecurity applications!
- Graph AI seems to be highly interesting emerging topic, especially in combination with knowledge graphs

Section 4

Conclusion

Conclusion

Conclusion

- Cyber Situational Awareness (CSA) – holistic views on cybersecurity
- Graph databases manage data structuring and CSA provisioning well
- Models grow in size and are becoming complex networks
- Various ways of applying graph theory, including simple graph traversal and graph measures on well-constructed graphs

Challenges and Future Work

- Practitioners need to learn a new data processing paradigm
- Large volumes of cybersecurity data of questionable quality, no datasets
- Several existing ontologies and tools, no unified language
- Excellent opportunities for graph AI (perhaps even more than for graph ML)

MASARYK
UNIVERSITY