Lightweight Impact Assessment and Projection of Lateral Movement and Malware Infection

IEEE Conference on Communications and Network Security – Cyber Resilience Workshop (CNS-CRW 2023)

Martin Husák[1,2] **(husakm@ics.muni.cz, martin.husak@utsa.edu)**
Michal Javorník[1] **(javor@ics.muni.cz)**

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

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

October 5, 2023

Introduction

Resilience of ICT Systems

- Ensuring the continuity of ICT services even under attack is a priority
- Various approaches in related works:
  - Asset management, vulnerability scans
  - Dependency detection and modeling, impact assessment
  - Attack projection and prediction, recommender systems

Specific problems

- Cybersecurity teams are understaffed or overloaded
  - The deployment of tools proposed in related work is non-trivial
- High-quality data are required
  - A detailed vulnerability assessment is not feasible in large-scale networks
- The threat landscape and cybersecurity situation are dynamic
  - The related works often rely on a static snapshot of the current situation

Introduction

Approach

- We estimate the next possible targets of an ongoing cyber attack
- Our approach builds upon the **risk score** calculated from the similarity and proximity of devices in the network
- Our work utilizes much simpler measurements and data collection
  - Instead of running deep vulnerability scans and dependency assessments, we only require device fingerprints and network topology

Contributions

- A proposal of a system that is easier to implement, deploy, and use in practice
- A reactive approach that considers lateral movement or malware infection spread

Modeling IT Infrastructures

Key question

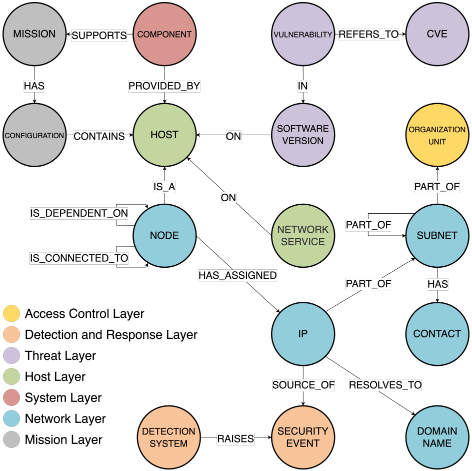
How likely are the other hosts going to be infected if a certain device is already infected?

- The same applies to lateral movement of an attacker
- Infection or exploitation is observed by a antivirus SW or IDS
- The concepts were proposed in previous work[1]

[1] M. Husák. Towards a Data-Driven Recommender System for Handling Ransomware and Similar Incidents. In 2021 IEEE International Conference on Intelligence and Security Informatics (ISI).

Modeling IT Infrastructures: Data Collection

We assume the date collected by the **CRUSOE** toolset[a] via common tools (Nmap, NetFlow, ...) and stored in Neo4j graph database.



[a] 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.

Modeling IT Infrastructures: Data Collection

For each **host** in the network, the **CRUSOE** toolset collects the following:

- Fingerprint of the operating system (via NetFlow or Nmap),
- List of open ports and services, including the name and version of the underlying software (via NetFlow and NBAR2 signatures or Nmap),
- Name and version of a web browser used on the system (via NetFlow),
- Name of the antivirus software on the system and its latest update (via NetFlow),
- List of vulnerabilities (via vulnerability scanner or estimated from fingerprints),
- Location, purpose, contact on administrator or primary user.

Risk Score: Estimating the Probability of Attack Spread

- Formally, the hosts are sorted by their **risk score** (*R*) calculated as a quotient of the similarity (*S*) and distance (*D*) of the two hosts:

$$R = \frac{S}{D} = \frac{s_1 * s_2 * \ldots s_n}{min\{d_1, d_2, \ldots, d_n\}}$$

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 subnet. Alternatively, the two machines can be close to each other if they are controlled by the same users or administrators.

Similarity

Similar software equipment and vulnerabilities, role, profile, or shared history (past security incidents) of the two hosts.

Risk Score: Distance Calculation

Attacker or malware may spread:

- in the same subnet or via shared resources
- to machines used by the same user or within department

Distance

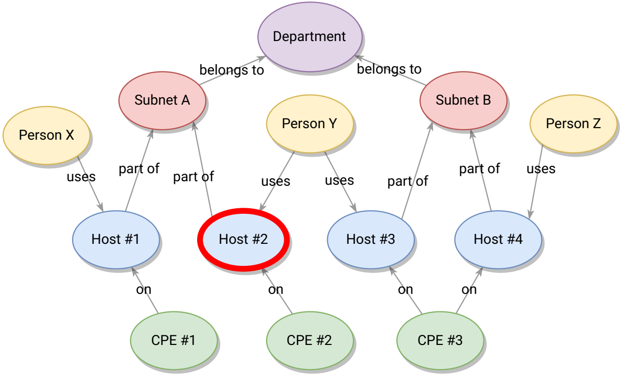
The distance between the two hosts is the minimal value of various distance metrics.

- Breadth-first graph traversal is used to find hosts with minimal distance in any of the distance metrics (in the implementation using graph database).
- The distance in logical network topology is the length of the path in the graph.

Risk Score: 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)



Risk Score: Similarity Calculation

- The malware often uses exploits of specific software or services.
  - If malware uses SSH brute-forcing, then Linux machines with SSH servers are at risk.
- We do not know the exact software equipment and may only assume similarities.
  - If the malware exploits Outlook email client, we shall look up all Windows machines.

Similarity

The similarity is calculated as a product of partial similarities $s_1 * s_2 * \ldots s_n$.
Each partial similarity is a value in the range $< 0, 1 >$.

- The similarity of software equipment and network services are the main features.
- CPE strings represent pieces of software running on a host.

Attack and Malware Spread Projection using Bayesian Network

Constructing BN

- We extend the risk score towards the whole network
- Modeling attack spread in the network using Bayesian network (BN) and inference
- Let's have a set of hosts in the network
    - Each host has a set of features or belongs to certain groups
    - This allows for calculating the RS between all pairs of hosts in the network
- We construct the BN just in time when an attack is detected or suspected.
    - We start with a graph with nodes only (hosts in the network), no edges
    - For each host, we assign a number $< 0, 1 >$
        - 0 means no adversarial control of the host,
        - 1 means confirmed attacker's control over the host,
        - values $(0, 1)$ represent the probability of the attacker's control over the host

Attack and Malware Spread Projection using Bayesian Network

Constructing BN cont'd

- Adding the edges requires caution – we need to ensure the graph is acyclic
  - Only the edges for which the RS is bigger than a threshold can be added to the graph
  - Processing of nodes is repeated until it it not possible to add another edge or there are no isolated nodes in the original graph
- Filling Conditional Probability Tables (CPTs) in nodes with multiple incoming nodes
  - Our assumption is that the attacker may advance simultaneously in all possible ways
  - We construct the CPT of a node using the risk scores of the incoming edges.

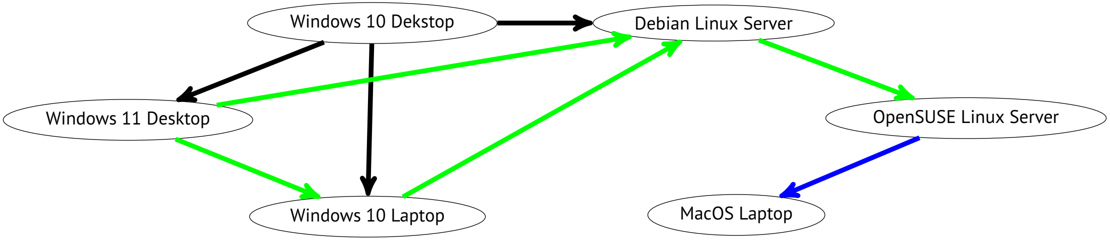
Outputs

- Given a joint probability distribution defined this factorized form, we can efficiently calculate the risk of an individual or even of the group of desired cyber components
- The output is the probability of exploitation of all the devices in the infrastructure

Attack and Malware Spread Projection using Bayesian Network

```
▷ Inputs
H ← list of hosts in the network, h ← already exploited host, t ← risk score threshold value
▷ Variables
G ← new empty directed graph
P ← empty list                                    ▷ nodes added in previous iteration
C ← empty list                                    ▷ nodes added in current iteration
▷ First iteration starting with the already infected host
G ← h, P ← h
▷ Further iterations
while no edge can be added do
    for i in P do
        for j in H do
            if (i,j) would not create cycle in G then
                if (riskScore(i, j) > t) then
                    G ← j
                    G ← (i,j)
                    C ← j
                end if
            end if
        end for
    end for
    P ← C, C ← empty list      ▷ Updating and clearing the lists of newly added hosts for the next iteration
end while
return G
```

Attack and Malware Spread Projection using Bayesian Network



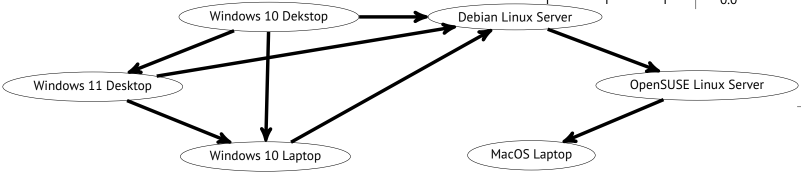
- Construction of a Bayesian network starting with the Windows 10 Desktop
    - Black arrows are added in the first iteration
    - Green arrows are added in the second iteration
    - Blue arrow is added in the final iteration

Attack and Malware Spread Projection using Bayesian Network

P(Win10D)= 0.5

Win10D	P(Win11D)
T	0.4
F	0.0

Win10D	Win11D	Win10L	P(DebLS)
T	T	T	0.271
T	T	F	0.171
T	F	T	0.171
T	F	F	0.81
F	T	T	0.171
F	T	F	0.81
F	F	T	0.81
F	F	F	0.0



DebLS	P(OpenSUSE)
T	0.3
F	0.0

Win11D	Win10D	P(Win10L)
T	T	0.64
T	F	0.24
F	T	0.24
F	F	0.0

OpenSUSE	P(MacOS)
T	0.2
F	0.0

- A sample Bayesian network for the attack spread estimation – final stage with CPTs

Conclusion

Summary

- We proposed a novel, lightweight approach to assess the impact of ongoing attacks
- Projection of the attacker's lateral movement and malware infection spread based on the similarity and closeness of devices in the network
- The required data are easily obtainable with common tools
- Alignment with current incident response practices and operational needs

Future work

- Reference implementation of a functional sample
- Preparing artifical or well-anonymized dataset for laboratory evaluation
  - Followed by evaluation in cybersecurity operations (in collaboration with CSIRT-MU)

MASARYK
UNIVERSITY