NeDaGen

Dennis van Wijk
Jeroen van Saane

Under Supervision of
Irina Chiscop
Federico Falconieri

08/02/2022

Table of Contents

I

Introduction

Research Question
Problem Statement
Related Work

II

Proposed Solution

Network traffic Data set Generator

III

Development

Architecture
Virtualization
Adversary Simulation

IV

Evaluation & Conclusion

Metrics
Future work

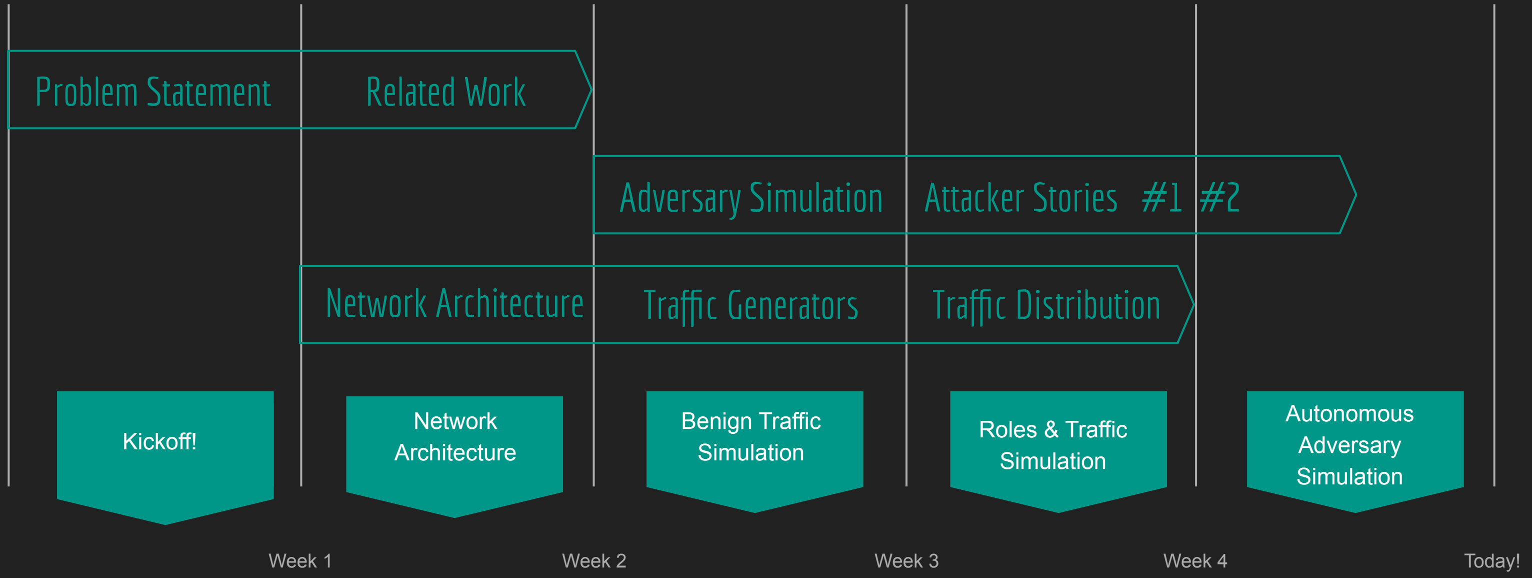
Research Question

"How can a high-quality network-based IDS data set be generated with an adequate attack diversity?"

Contributions

- Three-fold:
  - Provided an insight about NIDS generators;
  - Released a tool for creating flexible and tailored datasets;
  - Practically exemplified the tool’s versatility through adversary simulation.

Research Timeline



I. Introduction

Problem Statement

- Lack of:
    - Public NIDS Data Sets
    - High-Quality NIDS Data Sets
        - Anonymization (metadata)
        - (Outdated) Attacks
        - Volume of Traffic

I. Introduction

Related Work

- Intrusion Detection Dataset Toolkit

  Corderob et al. (2015) & Vasilomanolakis et al. (2016)

- Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization

  Sharafaldin et al. (2018)

- A survey of network-based intrusion detection data sets

  Ring et al. (2019)

II. Proposed Solution

NeDaGen

A Network Traffic Data Set Generator

Overall Requirements

- Flexible
- Extensible

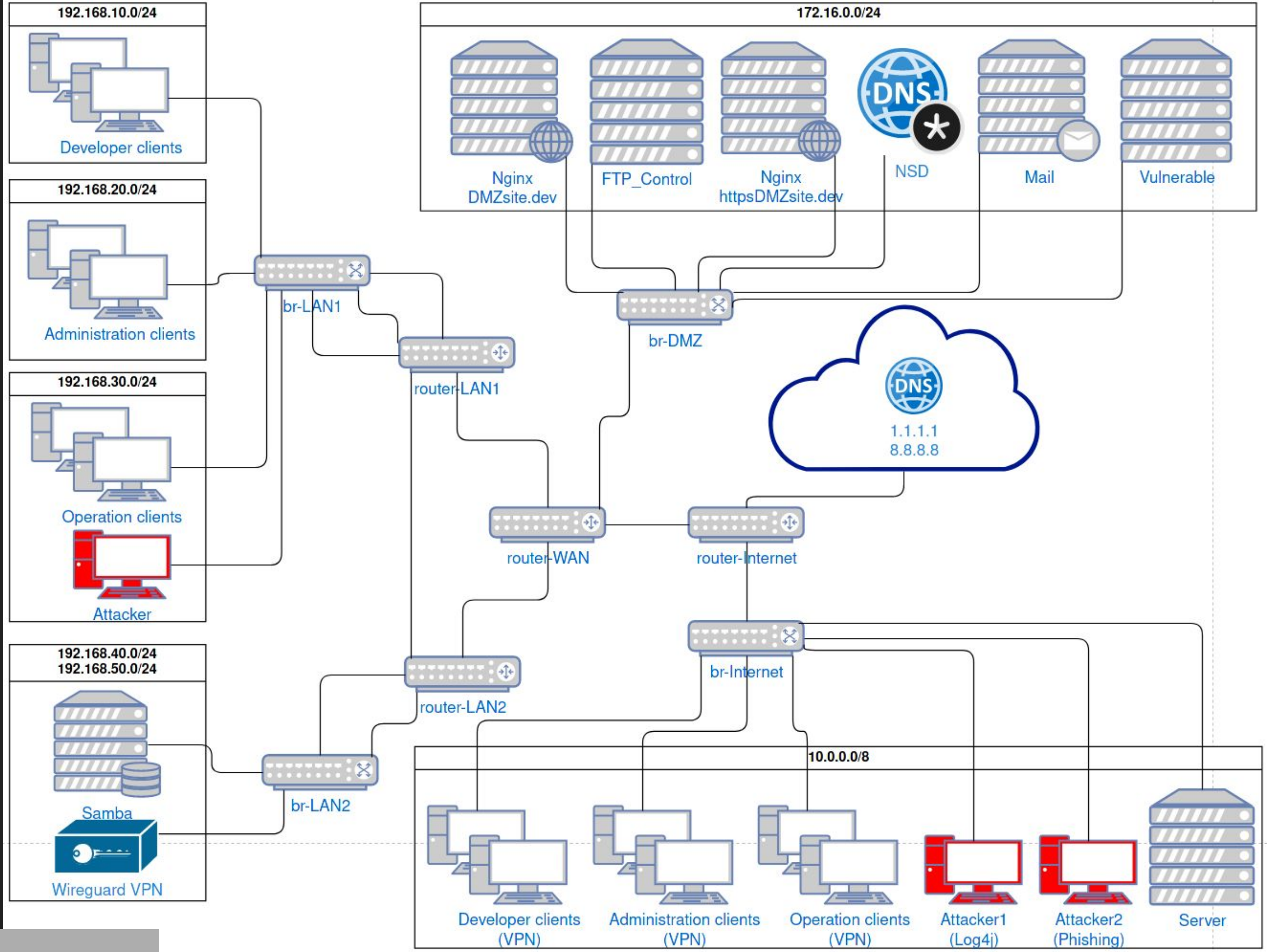
Non-Functional Requirements

- Availability
- Reproducibility
- Interoperability

Network Architecture

Traffic Generator

- Web
- Mail
- SSH
- FTP
- SMB



III. Development

Containerlab

- User-defined, versatile lab topologies
- Containerized Network Operating Systems
- BSD 3-Clause License

III. Development

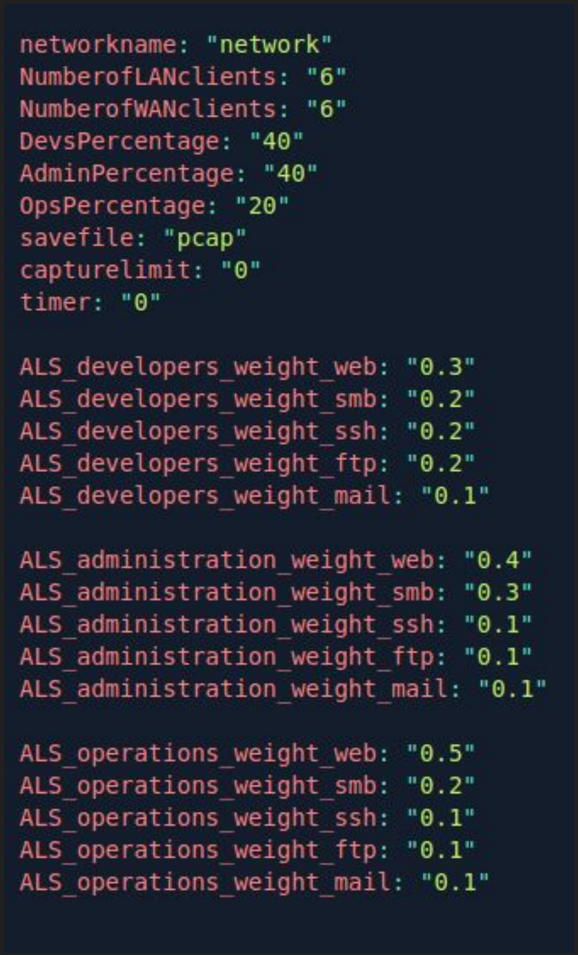
Virtualization

- OS-virtualisation
- Docker
- *Podman, containerd, ignite*

III. Development

Infrastructure as Code (IaC)

- Machine-readable Definition Files
- Jinja Extensible Templating Engine
- Idempotency

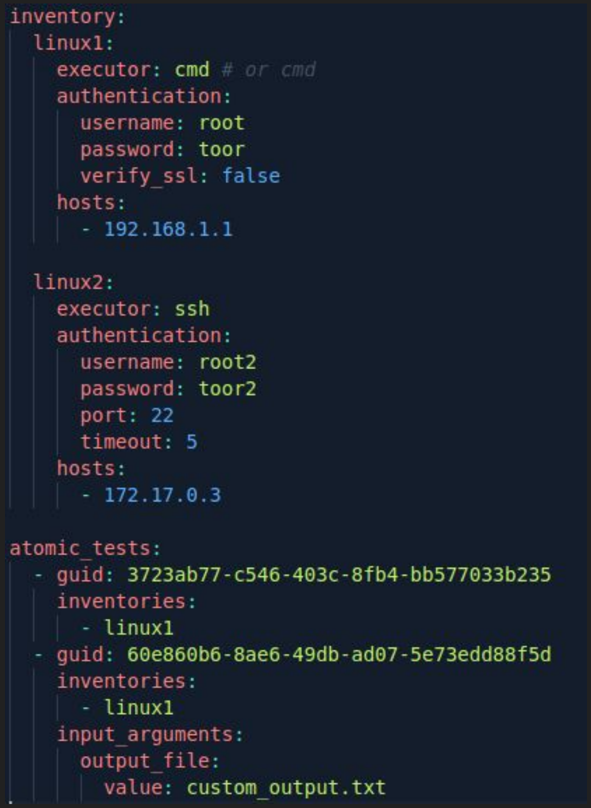


```
networkname: "network"
NumberofLANclients: "6"
NumberofWANclients: "6"
DevsPercentage: "40"
AdminPercentage: "40"
OpsPercentage: "20"
savefile: "pcap"
capturelimit: "0"
timer: "0"

ALS_developers_weight_web: "0.3"
ALS_developers_weight_smb: "0.2"
ALS_developers_weight_ssh: "0.2"
ALS_developers_weight_ftp: "0.2"
ALS_developers_weight_mail: "0.1"

ALS_administration_weight_web: "0.4"
ALS_administration_weight_smb: "0.3"
ALS_administration_weight_ssh: "0.1"
ALS_administration_weight_ftp: "0.1"
ALS_administration_weight_mail: "0.1"

ALS_operations_weight_web: "0.5"
ALS_operations_weight_smb: "0.2"
ALS_operations_weight_ssh: "0.1"
ALS_operations_weight_ftp: "0.1"
ALS_operations_weight_mail: "0.1"
```

III. Development

Configuration-based Attack Generation

- Atomic-Operator
- MITRE ATT&CK Framework
- Machine-readable Configuration Files



```yaml
inventory:
  linux1:
    executor: cmd # or cmd
    authentication:
      username: root
      password: toor
      verify_ssl: false
    hosts:
      - 192.168.1.1

  linux2:
    executor: ssh
    authentication:
      username: root2
      password: toor2
      port: 22
      timeout: 5
    hosts:
      - 172.17.0.3

atomic_tests:
  - guid: 3723ab77-c546-403c-8fb4-bb577033b235
    inventories:
      - linux1
  - guid: 60e860b6-8ae6-49db-ad07-5e73edd88f5d
    inventories:
      - linux1
    input_arguments:
      output_file:
        value: custom_output.txt
```

III. Development

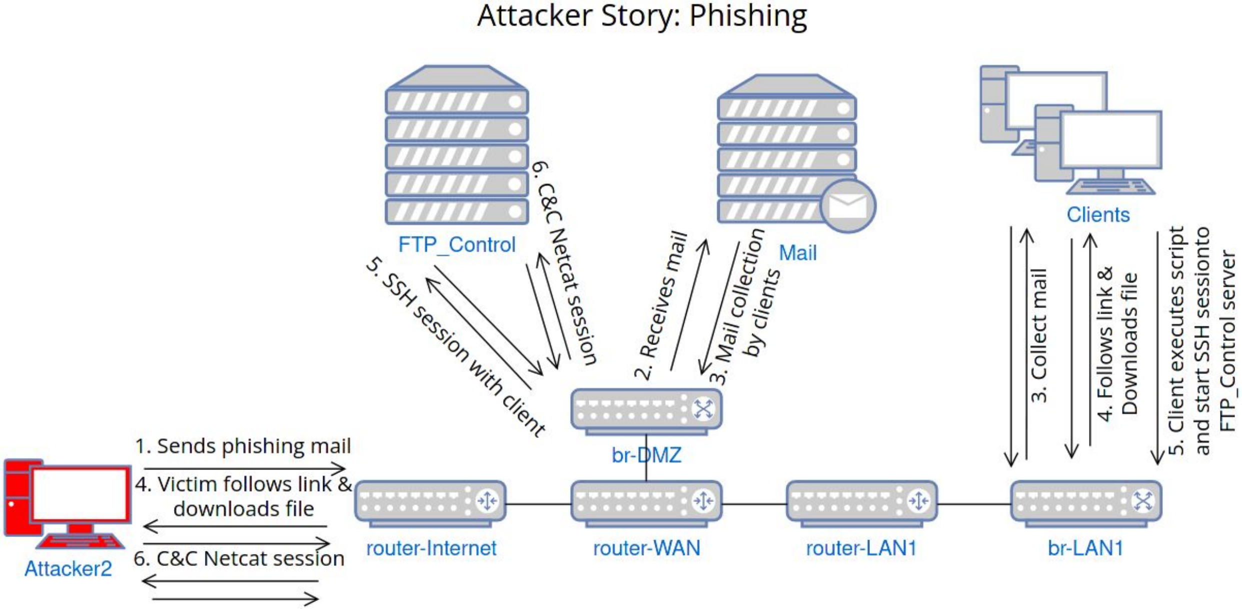
Adversary Simulation

- Importance
- Flexibility
- Autonomous Adversary Simulation (Stories)

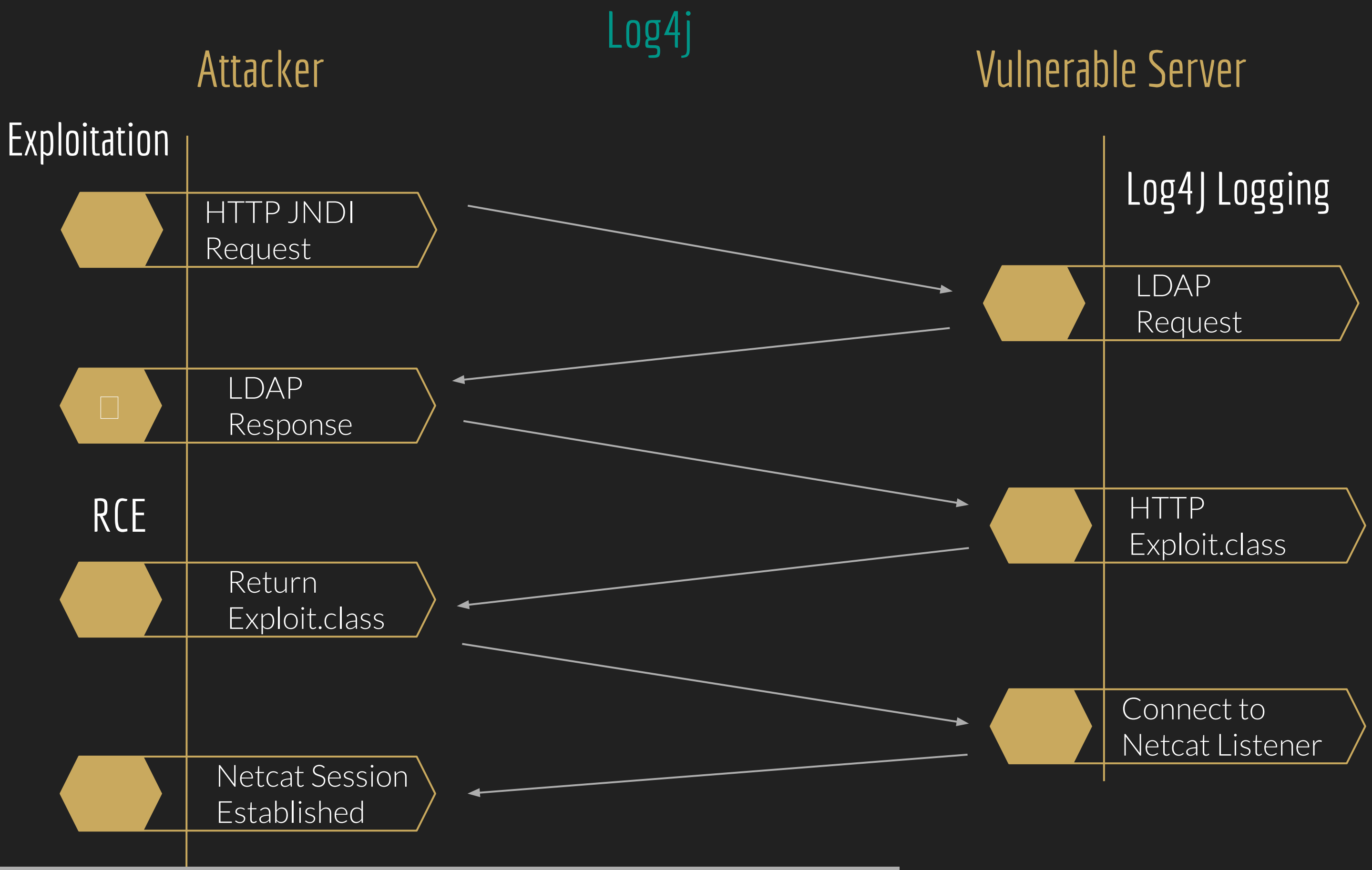
Autonomous Adversary Simulation #1

Spearphishing

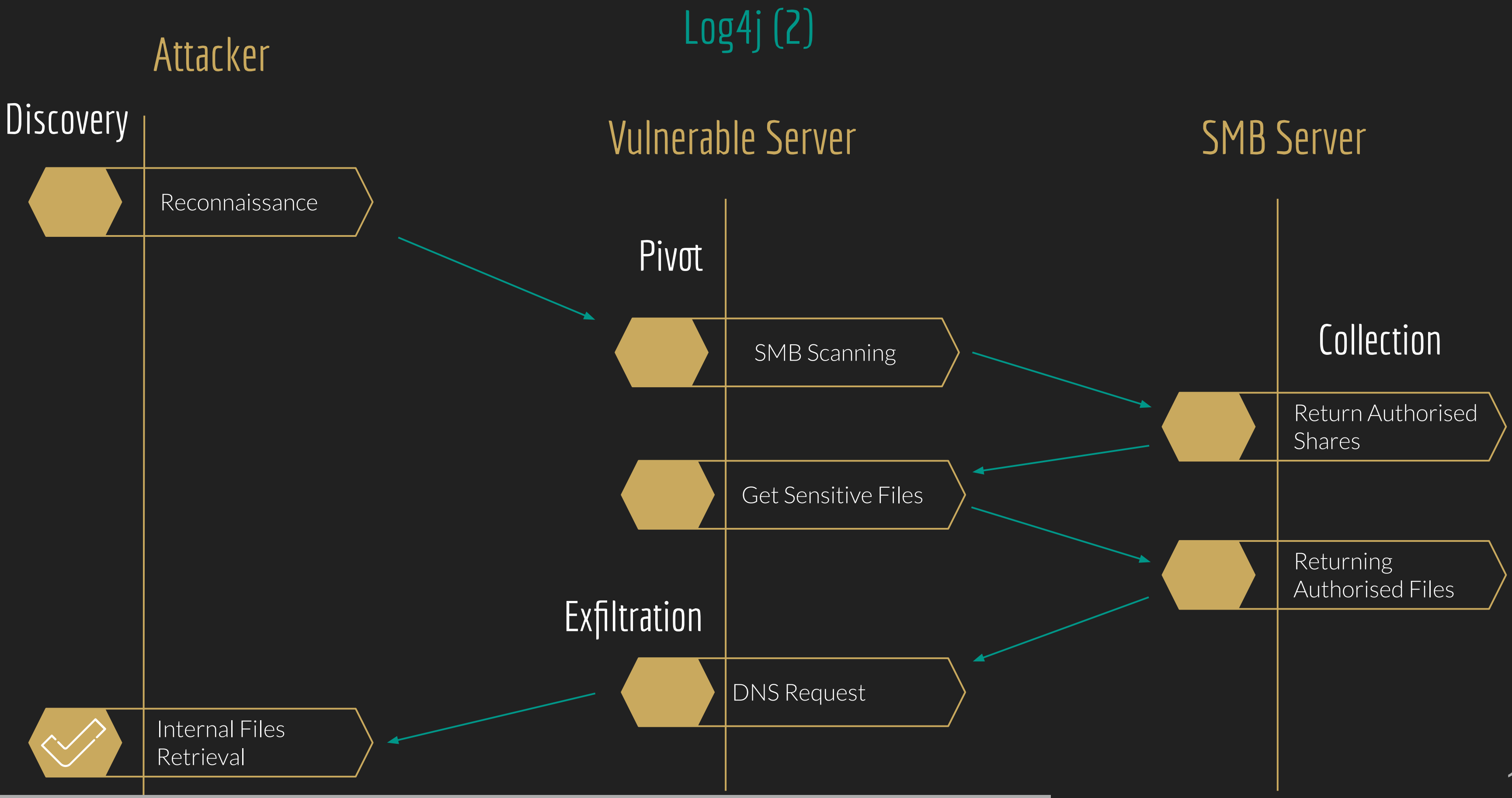
- Abusing Unix Shells
- Defence Evasion
- Privilege Escalation - Credential Access
- Persistence - Task Scheduling
- Lateral Movement - Filesystem Secrets
- Collection - Filesystem Secrets
- Exfiltration - Non-Application layer Protocol Technique



Autonomous Adversary Simulation #2



Autonomous Adversary Simulation #2

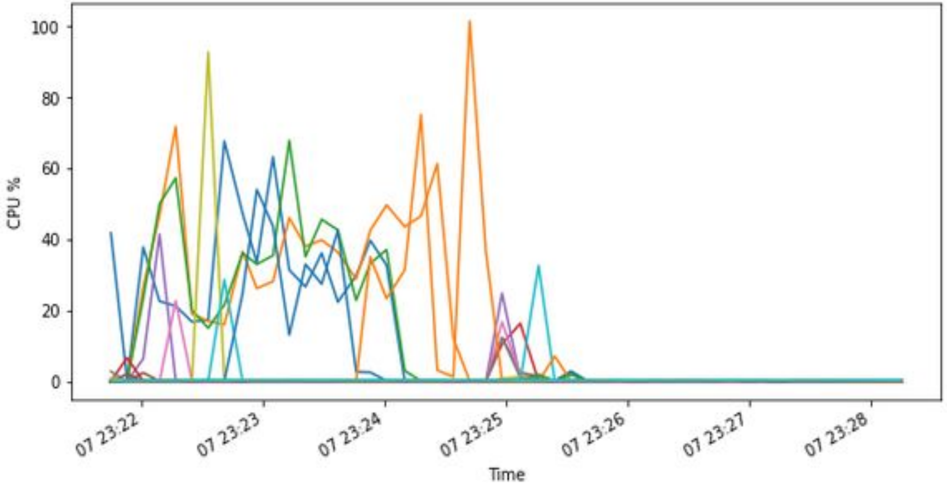


Autonomous Adversary Simulation #2

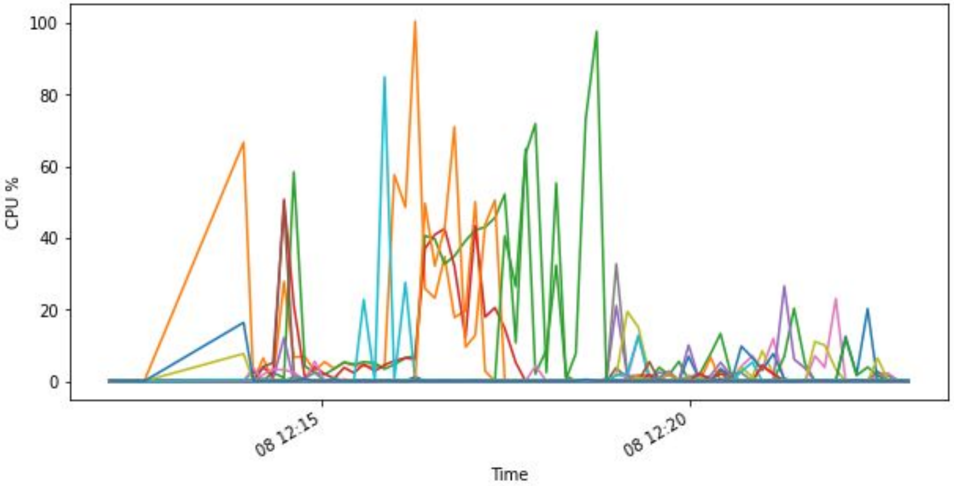
Autonomous Exploitation

IV. Evaluation

30 Nodes

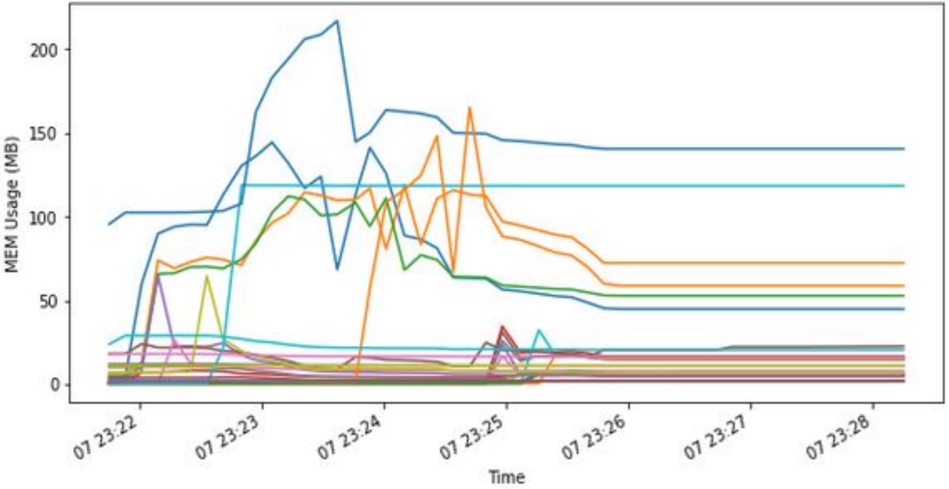


60 Nodes

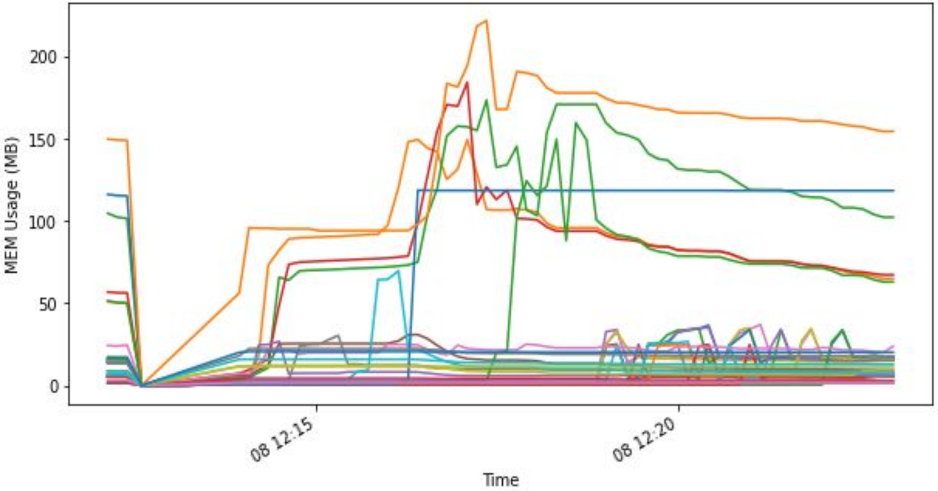


IV. Evaluation

30 Nodes



60 Nodes



IV. Evaluation

Data Set Requirements

Ring et al. (2019)

General Information

- ✓ Normal Traffic
- ✓ Attack Traffic

Nature of the Data

- ✓ Metadata
- ✓ Format

Data Volume

- ✓ Count
- ✓ Duration

Recording Environment

- ✓ Kind of Traffic
- ✓ Network Type

Evaluation

- ✓ Balanced
- % Labelled

IV. Conclusion

NeDaGen

- Extensible
- Customizable

IV. Conclusion

Future Work

- Probabilistic User Activity Traffic Distribution
- Automated IP Address Replacement
- Expand!

Questions?

Contact:
denvanwijk@gmail.com
contact@jeroenvansaane.com