Submission length: Long Paper
Submission category: Research Paper

Orchestrating a brighter world NEC

13th USENIX Workshop on Cyber Security Experimentation and Test (CSET '20)

# APTGen: An Approach towards Generating Practical Dataset Labelled with Targeted Attack Sequences

NEC Corporation
Yusuke Takahashi
Shigeyoshi Shima

Yokohama National University
Rui Tanabe
Katsunari Yoshioka

# Background

- The constant threats of targeted cyber attacks are one of the major security challenges in nowadays.

- The computer security incident response team (CSIRT) responds to the incident.

- CSIRT tries to reveal the whole picture of the attack through an incident response cycle.
  - Attack methods that attacker executed in the corporate network
  - Sequence of these attack methods (attack sequence)
  - Attacker's purpose.

# Motivation

- The faster the whole picture of the attack gets revealed, the period between detection and containment or eradication becomes shorter.

- We focus on the process of investigating attack sequences and would like to develop methods to automate or support this process.

- Developing the methods needs various kinds of attack sequence data, network logs and endpoint logs that contain attack traces related to the sequence.

# Our goal

- To the best of our knowledge, these kinds of open dataset are limited.

- We have decided to build the dataset for R&D for incident handling by ourselves.

# Possible Approaches for building the dataset

1. Get the log from organizations victimized by targeted cyber attacks [1, 2].

2. Observe cyber attacks purposely initiated in an observation environment [3].

3. Generate artificial attack sequences, execute them, and obtain logs.

[1] Stevens Le Blond, Adina Uritesc, Cédric Gilbert, Zheng Leong Chua, Prateek Saxena, and Engin Kirda. A look at targeted attacks through the lense of an NGO.In23rd USENIX Security Symposium (USENIX Security14), pages 543–558. USENIX Association, 2014.
[2] Grant Ho, Aashish Sharma, Mobin Javed, Vern Paxson, and David Wagner. Detecting credential spearphishing in enterprise settings. In26th USENIX Security Symposium (USENIX Security 17), pages 469–485. USENIX Association, 2017.
[3] Brown Farinholt, Mohammad Rezaeirad, Paul Pearce, Hitesh Dharmdasani, Haikuo Yin, Stevens Le Blond, Damon McCoy, and Kirill Levchenko. To Catch a Ratter:Monitoring the Behavior of Amateur DarkComet RAT Operators in the Wild. InSecurity and Privacy (SP), 2017 IEEE Symposium On, pages 770–787. IEEE, 2017.

# Possible Approaches for building the dataset

1. Get the log from organizations victimized by targeted cyber attacks.

difficult to infer the attack sequences

2. Observe cyber attacks purposely initiated in an observation environment.

3. Generate artificial attack sequences, execute them, and obtain logs.

We propose APTGen, an approach for generating attack sequences and executing them for building a dataset.

# Requirements

- Requirements for building a dataset for incident handling:
  - Unify the names of attack methods and their scope in attack sequences. (Reproducibility)
  - Clear how realistic generated attack sequences are. (Reality)
  - Generate various attack sequences. (Diversity)

- At present, APTGen meets the requirements of reproducibility and diversity.

- To meet the requirement of reality is one of future works.

# APTGen

APTGen consists of attack sequence generation and execution tools.

APTGen:

- Generates artificial attack sequence from existing security reports based on the attack model defined in MITRE's ATT&CK.
- Executes generated sequence to obtain logs from execution environments.

# APTGen: Generation tool

## Inputs :

- Technique list
- Targeted environment information
- Generation condition

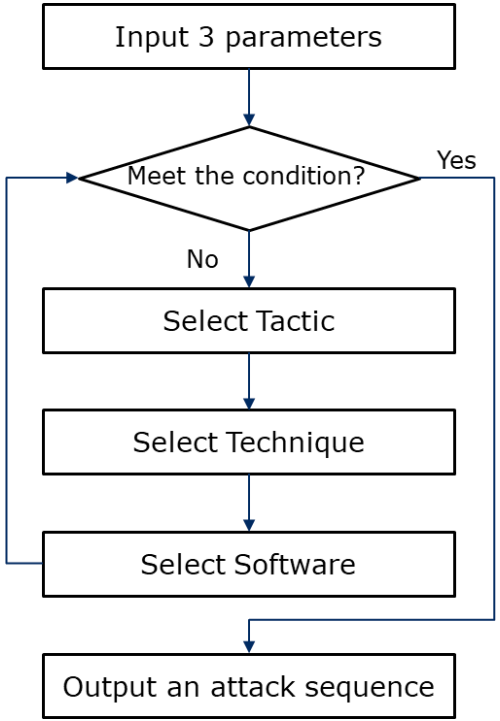
## Output:

- Attack sequences

| Step | Tactic | Technique | Software |
|---|---|---|---|
| 1 | $Ta_1$ | $Te_1$ | $S_1$ |
| ⋮ | ⋮ | ⋮ | ⋮ |
| k | Credential Access | Credential Dumping | Mimikatz |
| ⋮ | ⋮ | ⋮ | ⋮ |
| n | $Ta_n$ | $Te_n$ | $S_n$ |

# APTGen: Generation tool

- The tool selects Tactic/Technique/Software randomly under the following constraints.
  - Selects Tactic/Technique/Software following the ATT&CK design.
  - Selects Tactic/Technique/Software so that an attack sequence is logically completed.
    - e.g.: it is necessary to execute any Techniques in Discovery before executing any Techniques in Lateral Movement.



# Experiment

- The purpose of experiment is to evaluate the diversity of outputted attack sequences.

- We generated 800 different attack sequences based on eight actual security incidents.

| Incident (targeted organization, event, or attacker) |
|---|
| APT29 |
| Bronze Butler |
| Clinton campaign |
| Japan Pension Service |
| National Institute of Advanced Industrial Science and Technology |
| SingHealth |
| South Korean banks and broadcasting organizations |
| Ukrainian electricity distribution companies |

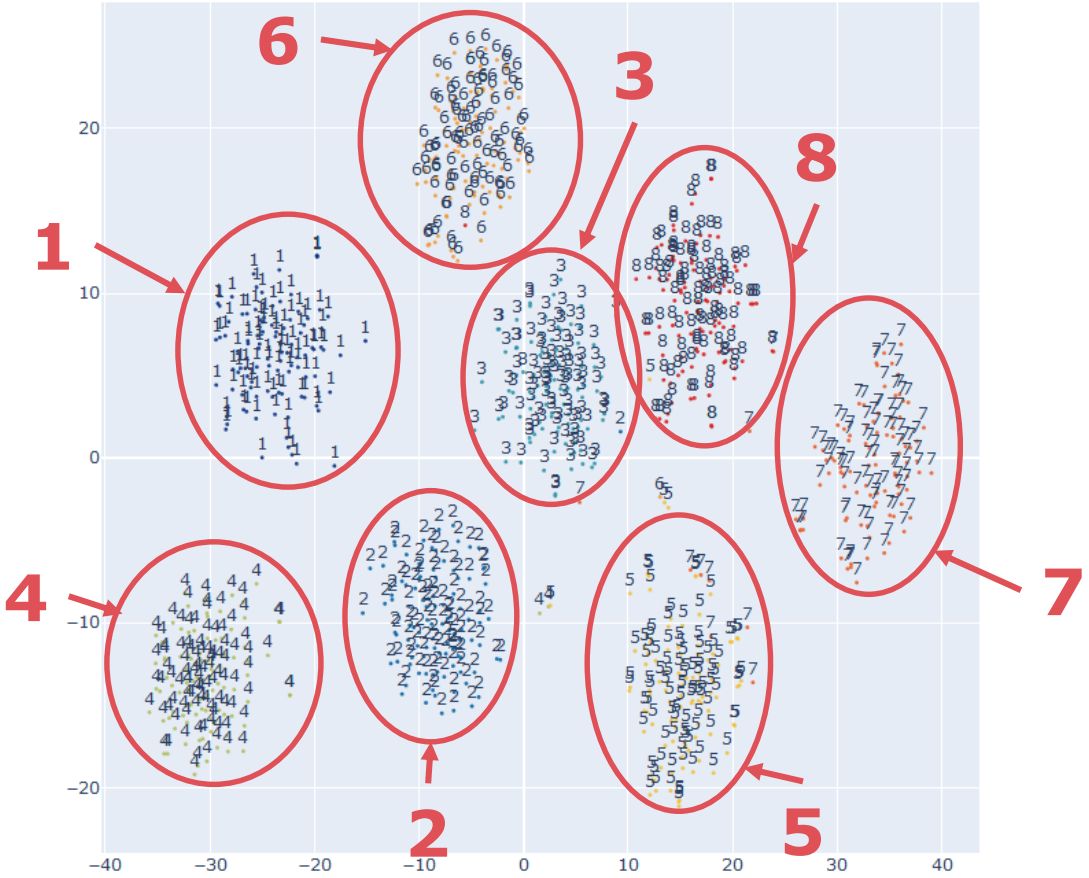
# Experiment: Technique list in Japan Pension Service Incident

| Tactic | Technique |
|---|---|
| Collection | Email Collection |
| Collection | Data Staged |
| Credential Access | Account Manipulation |
| Credential Access | Credential Dumping |
| Credential Access | Credentials in Registry |
| Defense Evasion | File Deletion |
| Discovery | System Information Discovery |
| Discovery | System Network Configuration Discovery |
| Discovery | File and Directory Discovery |
| Discovery | Account Discovery |
| Discovery | Permission Groups Discovery |
| Discovery | Network Share Discovery |
| Discovery | Remote System Discovery |
| Exfiltration | Exfiltration Over Command and Control Channel |
| Exfiltration | Data Compressed |
| Lateral Movement | Remote File Copy |
| Lateral Movement | Pass the Hash |
| Persistence | Scheduled Task |

# Experiment: Generation conditions in Japan Pension Service Incident

| ID | Generation Conditions |
|---|---|
| 1 | Sequence length is 8 or more. |
| 2 | Sequence length is 1 or more<br>and last Tactic in sequence is Lateral Movement. |
| 3 | Sequence length is 1 or more<br>and last Technique in sequence is Exfiltration Over Command and Control Channel. |
| 4 | Sequence length is 1 or more,<br>sequence contains Lateral Movement,<br>and last Technique in sequence is File Deletion. |
| 5 | Sequence length is 1 or more,<br>last Technique in sequence is File Deletion,<br>and sequence contains Lateral Movement and Exfiltration Over Command and Control Channel. |

# Analysis



# Dataset

- We publish 800 generated attack sequences and corresponding logs as a dataset.
- We are preparing to provide our generation and execution tools for users who accept our terms of use.
- If you are interested, please see the following URL.

https://ipsr.ynu.ac.jp/aptgen/index.html

# Future work

- Contain benign activity logs in a dataset.
- Make attack sequences and logs more real.
  - Algorithm of generating attack sequences.
  - More exploit codes/tools.
- Test generated attack sequences against detection methods.
- Discuss with experts in CSIRT and have them evaluate the reality of the generated dataset.

# Conclusion

- We presented APTGen, an approach for generating artificial attack sequences of targeted cyber attacks and executing corresponding attacks to generate dataset of targeted attacks.

- We analyzed the relationship between the generated attack sequences by visualization.

- We released a dataset consisting of generated attack sequences and corresponding logs that have been obtained by executing these sequences.

# Thank you!

Please let us know your questions and comments.

ynugr-ylab-aptgen@ynu.ac.jp