Internship Final Presentation

Lawrence Chang

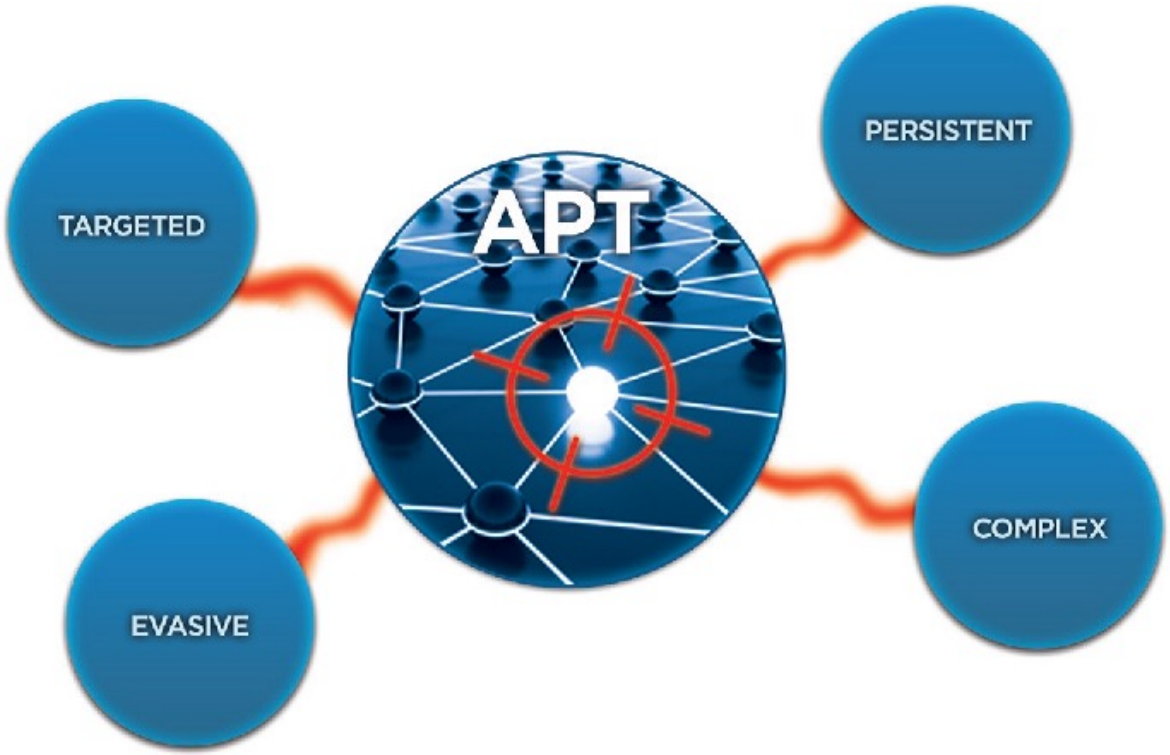
Internship Overview

1. Building Background Knowledge
    1. Develop Cybersecurity background knowledge by reading research paper, doing research online.
    2. First introduce to APT, EDR, OpTC dataset, ANUBIS, AttacKG and ThreatKG

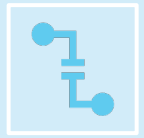
2. Hand-on Visualization and Analyze project
    1. Understanding code from senior engineer.
    2. Visualize behavior graphs and kill chain to analyze malicious activities.
    3. Developed a program to streamline data preprocessing, threat identification, and attack path visualization

Advanced Persistent Threat (APT)

- APT attacks consist of a “kill chain” where attackers perform sequential actions to achieve their goals.
- The goal of APT is to achieve long-term access to the targeted network.
- For APT attacks, hackers typically select high-value targets, such as nation-states and large corporations.



Challenges of APT Detection



APT maintain a very low profile in the compromised system.

APT activities are extremely small percentage of event log data.

Detection of long-running APT campaign is a time consuming task.

Behavior-Based Threat Detection

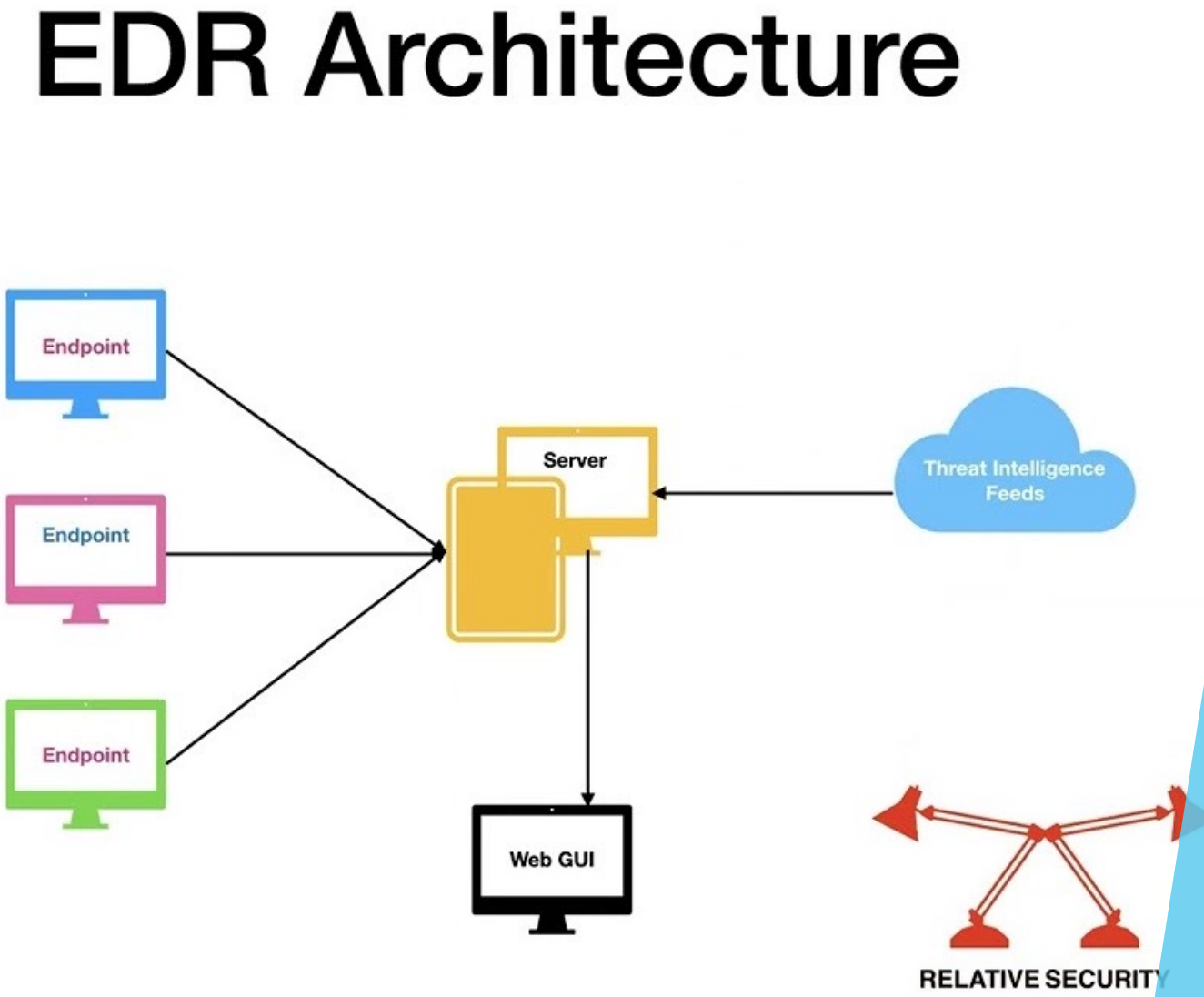
Collecting and analyzing large amounts of data from various sources, such as logs, network traffic, system events, and user behavior.

Applied advanced analytics, machine learning, and statistical modeling to detect suspicious activities.

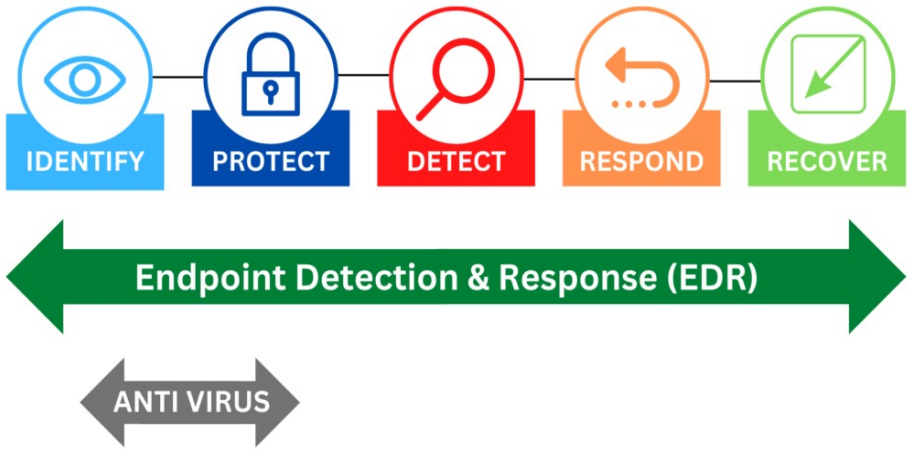
Behavior graph approach can provide early warning signs of a potential security threat.

Possible solution of APT:
Endpoint Detection and Respond Systems (EDR)

- The common solution for combatting APT.
- Behavior-Based Threat Technique
- EDR constantly monitor activities on end hosts and raise threat alerts if malicious behaviors are detected.
- EDR tools hunt threats by matching system events against a knowledge base. For example, Techniques, and Procedures(TTPs).
- EDR systems use various techniques, including behavior-based analysis, machine learning, and heuristics, to identify suspicious activities on endpoints.



Challenge of EDR



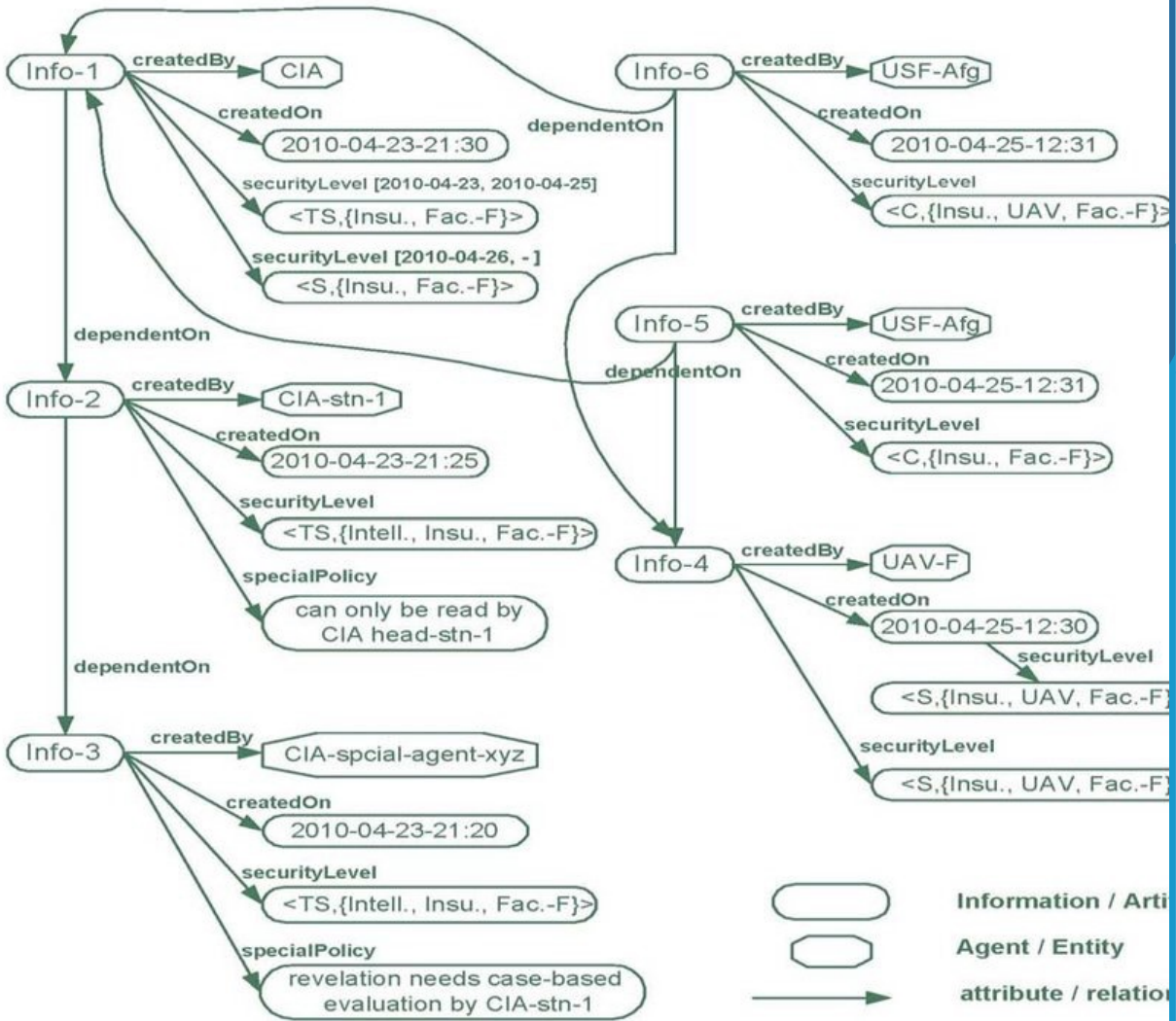
EDR tools generate a high volume of false alarms.

Determining the reliability of these threat alerts require labor to handle the overwhelming amount of low-level system logs.

In practice, the system logs describing long-lived attack are often deleted before an investigation is ever initiated.

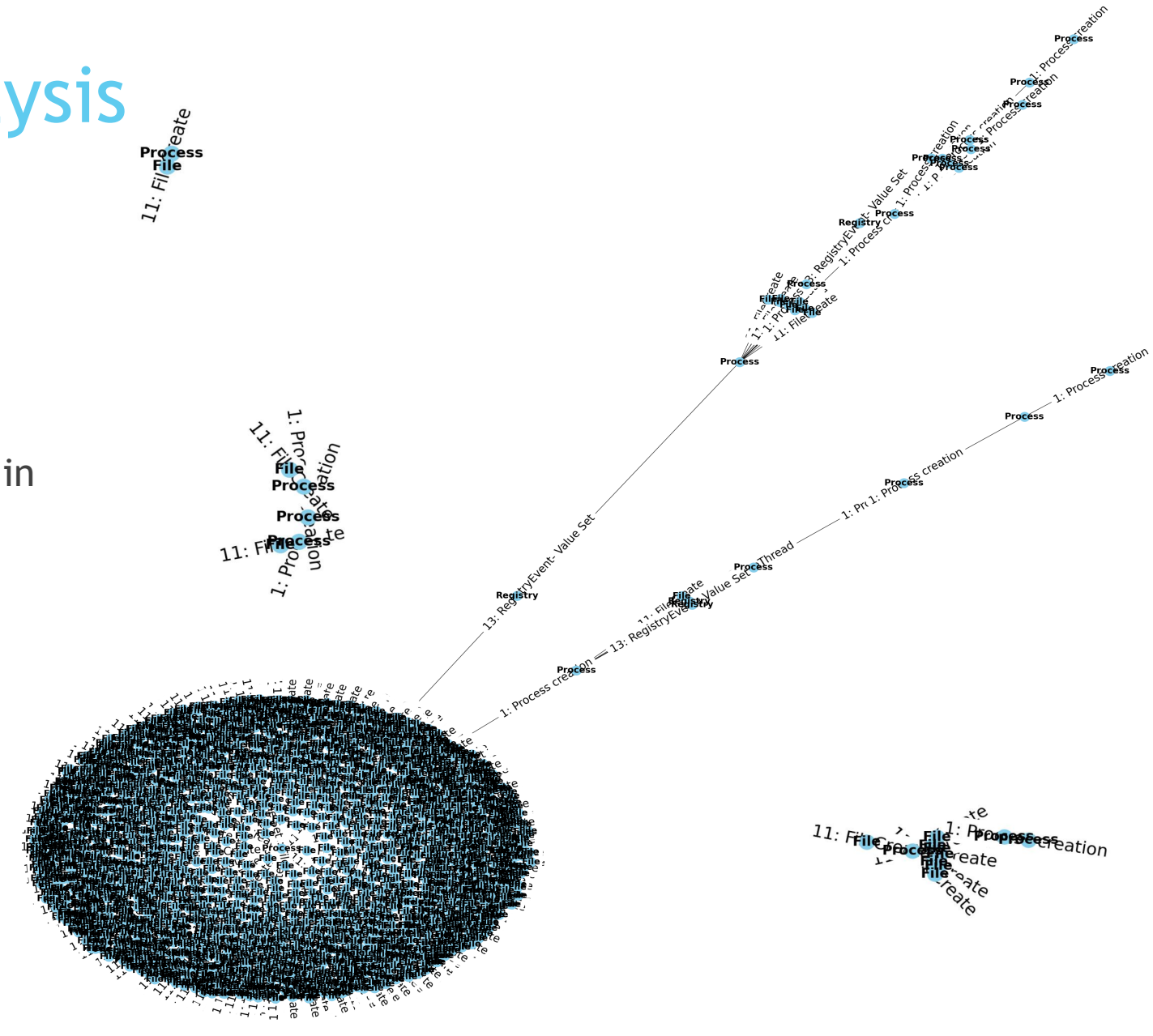
Solution From Recent Papers: provenance graphs

- Provenance graphs, a logs reduction scheme that can reduce the storage of system logs while preserving causal links between existing and future threat alerts.



Behavior graph Analysis

- Total 2137 data points
- Transfer into Sysmon data model
- Transfer to a Tree structure but store in pandas Data frame.
- Consist of multiple subgraph.
- Each node is subject and object.
- Attacks path account for a small percentage of the entire dataset.
- Interactive graph demonstration.



bg_Lawrence_visualize.py

- Streamline data preprocessing.
- Threat identification.
- Attack path visualization.
- Store the result in csv file
- Work with every Sysmon data model.

To visualize BG from agent logs:

```
python bg_Lawrence_visualize.py –h
```

Usage: bg_Lawrence_visualize.py -i <input_file> -p -b -a
Options:

```
  -h, --help.             Show this help message and exit
  -i, --input             Specify the input .csv file.
  -p, --preprocess        Transfer input data into sysmon data model.
  -b, --behavior_graph    Produce behavior graph of whole data set.
  -a, --attact_graph      Produce attack graph.
                          This graph only display path with malicious activity.
```

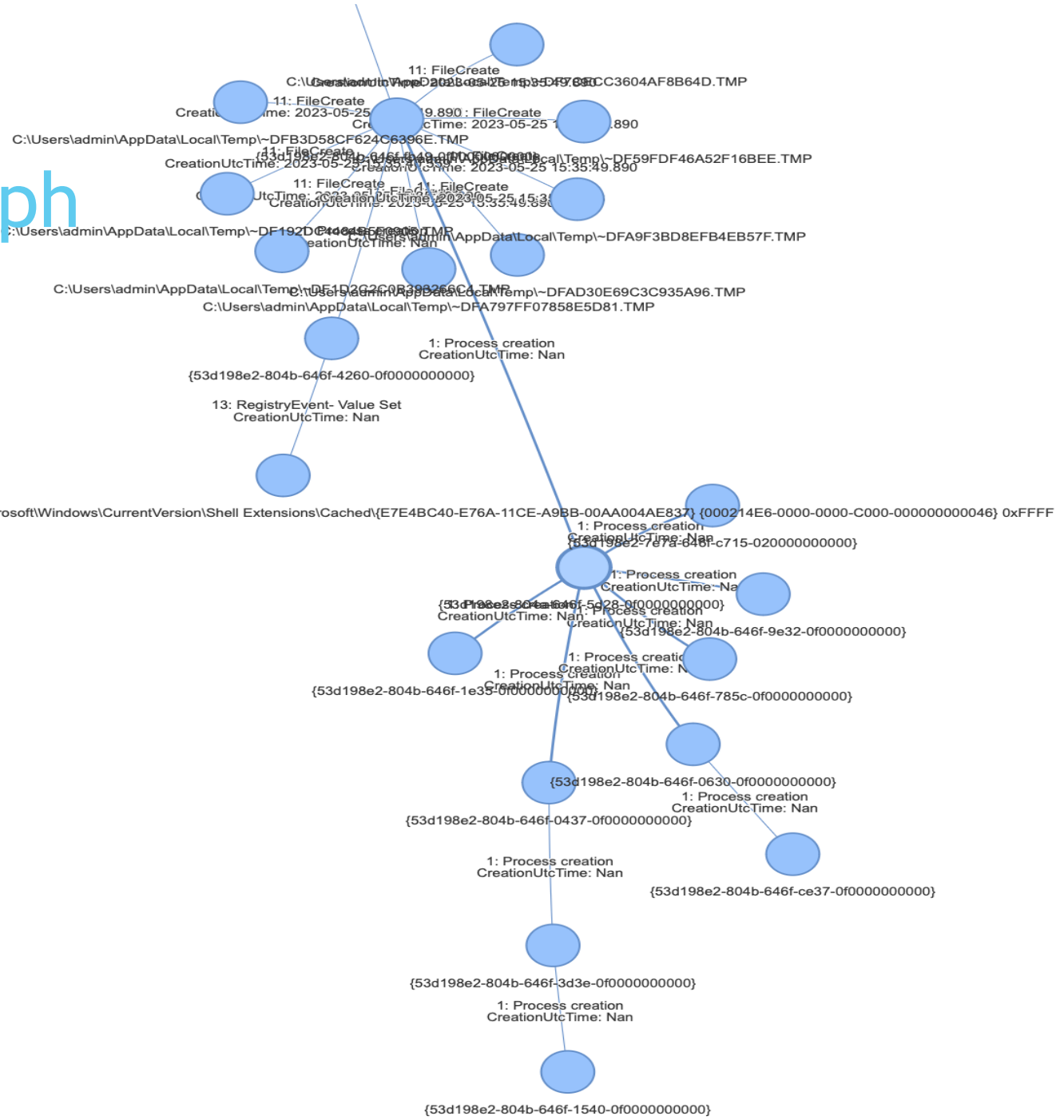
For example:

```
python bg_Lawrence_visualize.py –i input.csv –p –b –a
```

Output:
behavior_graph.JPEG
AttackPath_graph.JPEG
interactive_bg_graph.html
interactive_AttackPath_graph.html
AttackPath_node_data.csv
AttackPath_edge_data.csv
(This interactive graph has to open in browser)

Looks into Behavior Graph

- Node representative subject and object.
- Multiple Process were created in parallel.
- Time range of data set is short, only 10 minutes.

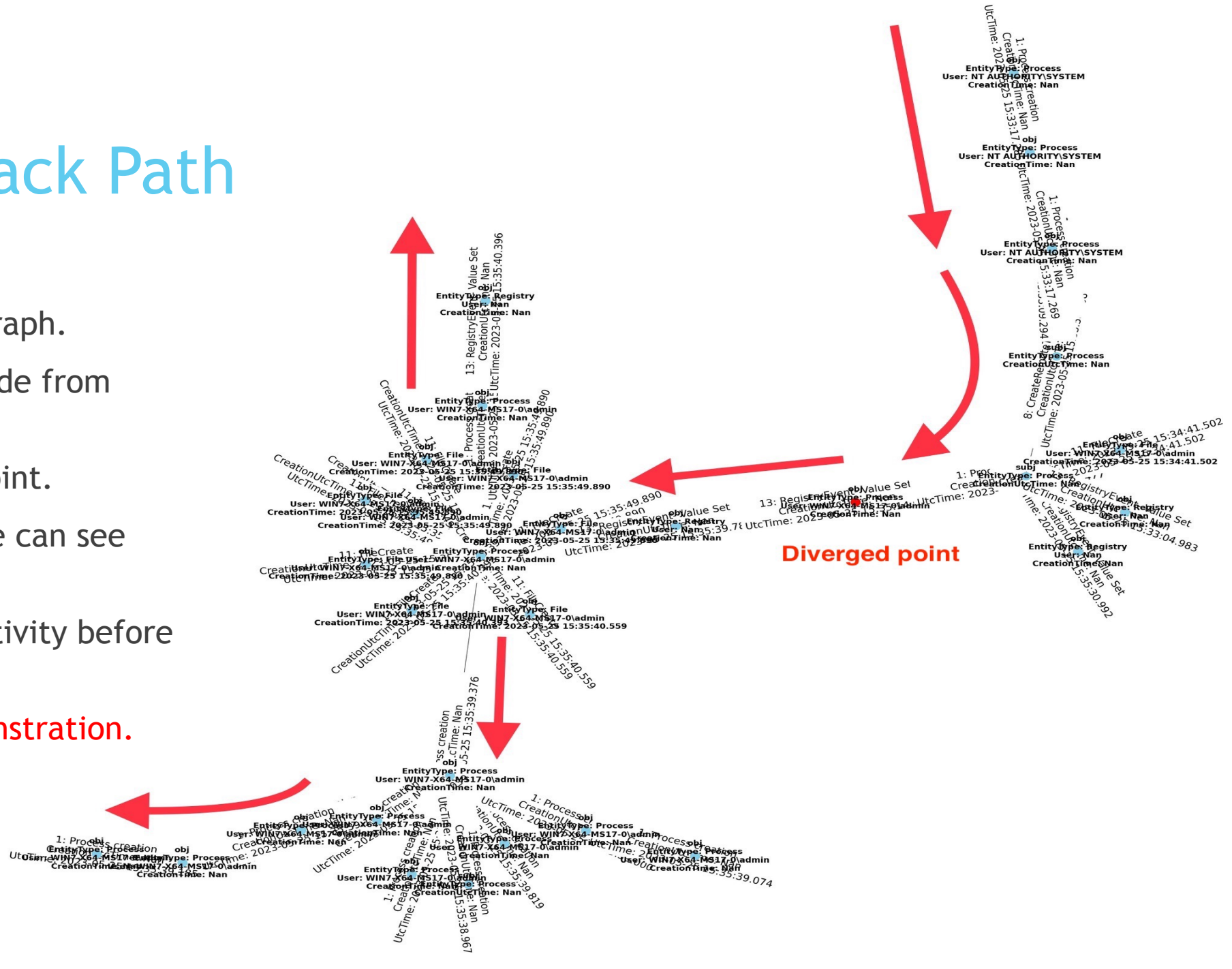


Center Analyze

- All Process create file.
- Action happened at the same time.
- Huge among of files are being encrypted.
- Too late to identify/block the malicious activity.

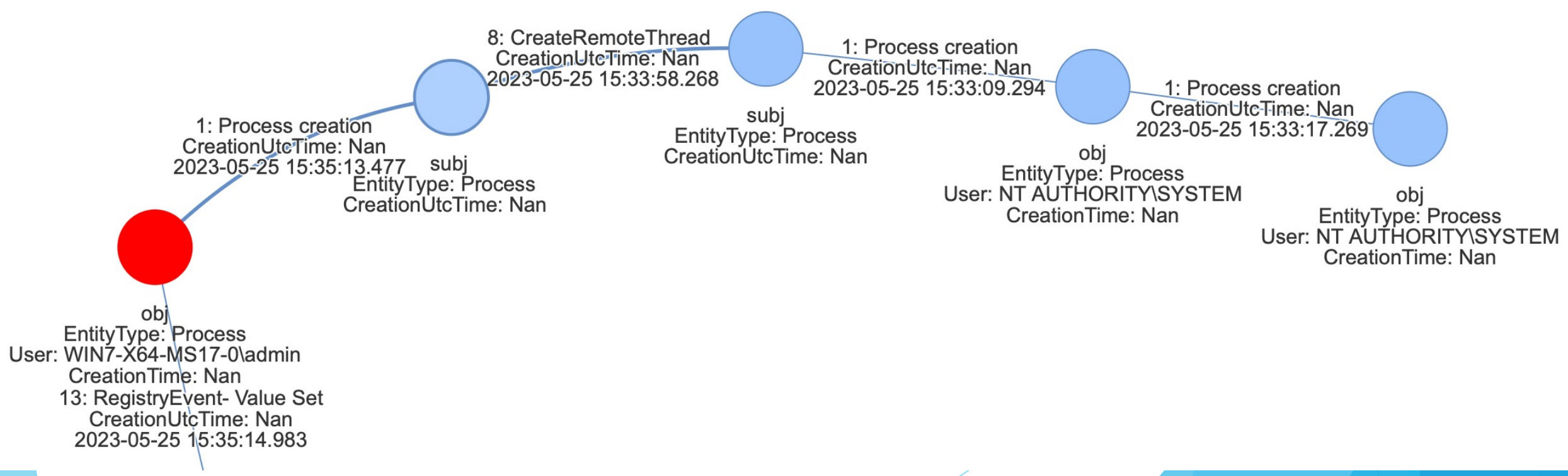
Construct Attack Path

- Remove unrelated subgraph.
- Remove unnecessary node from diverged point.
- Red point is diverged point.
- Base on create time, we can see the path.
- We focus on analysis activity before the diverged point.
- Interactive graph demonstration.



Attack Path visualize & challenge

- Only consist of five steps in attack path.
- Before diverged point, previous activity are regular activity.
- In the program, the tree is not orientation by create time. Therefore, it is hard to do branch pruning.

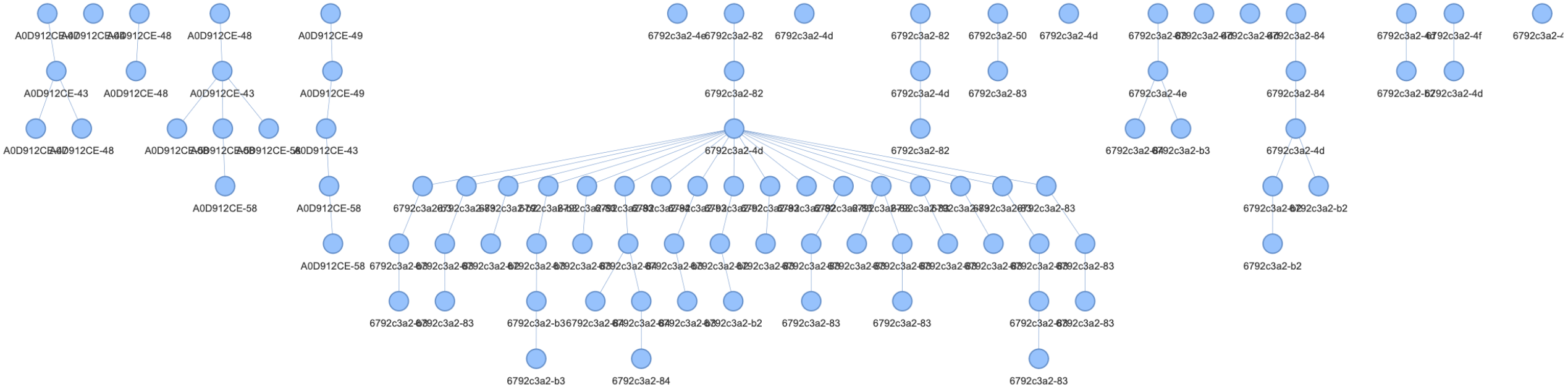


Process Tree of nasa.json data

- Node insertion happened in real time.
- Tree store in RAM
- Match malicious activity by using regular expression .
- Once the malicious activity has been detected, track back to the parent node.

Interactive Process Tree demonstration.

- process_tree_visualize.py



Thank you all