Poster: APT Detection through Sensitive File Access Monitoring

Wenjia Song  
Virginia Tech  
wenjia7@vt.edu

Danfeng (Daphne) Yao  
Virginia Tech  
danfeng@vt.edu

Abstract—Advanced persistent threats (APTs) have become more destructive in recent years, impacting a wide range of organizations, from government agencies to critical infrastructures, and the daily life of millions of people. There have been many security works focusing on this topic and various solutions have been proposed. However, challenges still exist for the accurate detection of new APT variants. Innovative and strong defenses need to be invented. To achieve this, we plan to systematically analyze recent APT attacks and the current practice of system monitoring of sensitive information, which could lead us to new and secure solutions.

I. INTRODUCTION

Ransomware and zero-day supply-chain trojan, a new form of attack, have posted unprecedented threats to various organizations. $350 worth of cryptocurrency ransom was collected by hackers in 2020.

$4.4 million ransom was paid by Colonial Pipeline Co. in May 2021 and $11 million was paid to the hackers in June 2021 by the meat supplier JBS USA. Zero-day supply-chain trojan leverages software updates from established vendors. The SolarWinds and Kaseya hacks are examples of this new form, threatening the data safety of hundreds of companies. Both of these attacks could be variants of advanced persistent threats (APTs).

As the name implies, advanced persistent threats (APTs) are often achieved by groups of sophisticated hackers. They try to stay present in the system as long as they can and threaten the safety of sensitive data and critical components of targeted organizations. Well-known and documented examples of APTs include APT-28 and APT-38. APTs usually consist of multiple key stages. The life cycle of APTs is shown in Figure 1.

There have been many research studies on the detection of APTs. However, challenges still exist. Innovative and strong security defenses are needed to ensure the safety of sensitive data.

II. RELATED WORK

Various APT detection tools have been proposed in recent years. They can be roughly categorized by the attack stage they focus on (e.g., initial penetration or latent movement) and the approach they use for detection (e.g., rule-based or machine learning-based), as shown in Table 1.

To prevent the malware from entering the victim machine, Mohammad et al. propose a feature-based classification of phishing websites. Chandra et al. use a mathematical model to filter spam emails. APTGuard is a tool to detect spear phishing URLs using decision tree and neural network.

TABLE I

<table>
<thead>
<tr>
<th>Attack Stage</th>
<th>Rule-based Approach</th>
<th>Machine learning-based Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection of initial penetration</td>
<td>Mohammad et al.</td>
<td>APTGuard</td>
</tr>
<tr>
<td>Detection of malicious movement</td>
<td>Poirot</td>
<td>MLAPT</td>
</tr>
</tbody>
</table>

Fig. 1. Life cycle of APTs.

1https://blog.chainalysis.com/reports/ransomware-ecosystem-crypto-crime-2021/
3https://content.fireeye.com/apt/rpt-apt38
learning models to detect malware before installation using static, dynamic, and origin information of executables. On the other hand, researchers have also done works on the detection of malicious movements of APTs after the malware get inside the victim systems. HOLMES [2] finds the behavior pattern connections of key stages of APTs. POIROT [7] analyzes the semantic of causality of system alters for anomalous events. For machine learning-based approaches, MLAPT [8] is a multi-phased machine learning detection framework using network traffic. Another tool, named UNICORN [9], clusters provenance graph sketches and dynamically detects anomaly system changes.

III. CHALLENGES

Despite the research advances, there are still several challenges for detecting sophisticated APTs.

- **Hard to catch new variants.** APTs constantly change and update to disguise themselves from detection. The supply-chain trojan is an example of their new form of entry. Behavior patterns summarized from historical data may be outdated and unable to identify newer versions.

- **Hard to generalize across different hosts and operating systems.** Many anomaly detection models are trained on normal system behaviors. However, normal behavior may vary from system to system. To generalize, a large amount of normal data needs to be collected for each system.

- **False alters and missed detections.** Due to the large amount of system and network data and the lack of clear distinction between malicious and benign behavior, many machine learning-based approaches tend to produce false positives and false negatives.

IV. ONGOING WORK

Although APTs change their form, structure, and behavior to avoid detection, their goal remains the same. Therefore, new approaches focusing on monitoring and protecting sensitive information, which is the target of attacks, are needed. In order to develop such a security defense, we need to understand the recent APT attacks and how the current systems monitor sensitive information. The major tasks of our next steps include:

- Analyze the current practice of how critical and sensitive information is monitored by the system. Different auditing policies generate logs with different types of events. A key step is to find the optimal configuration of security auditing.

- Analyze the recent APT attacks from various sources. The resources include executable malware samples (e.g., real-world APT malware samples summarized in [6]), APT execution logs (e.g., APT-EXE [5]), datasets containing APT traces (e.g., DARPA dataset, KDD-99 [4] and its derivative NSL-KDD [3], DAPT 2020 [7]), and descriptions and documentation of various attacks (e.g., APTNotes [8], MITRE ATT&CK [9]).

- Fine-grain security logs and focus on sensitive information access-related events. Systems generate an overwhelming amount of event logs that contain noise. Focusing on a smaller set of relevant events can potentially help reduce false positives. As shown in Figure 2, events 5379 (i.e., Credential Manager credentials were read) and 4656 (i.e., a handle to an object was requested) are examples of sensitive information-related events that should be handled carefully.

Fig. 2. Example of Windows security event log.

ACKNOWLEDGMENT

This work has been supported by the Office of Naval Research under Grant No. N00014-22-1-2057.

REFERENCES


6http://www.unb.ca/cic/datasets/mlshtml.
7https://gitlab.thothlab.org/Advanced-Persistent-Threat/apt-2020/-tree/master
8https://github.com/kbandia/APTNotes
9https://attack.mitre.org/
Poster: APT Detection through Sensitive File Access Monitoring

Wenjia Song, Danfeng (Daphne) Yao
Department of Computer Science, Virginia Tech, VA, USA
{wenjia7, danfeng} @vt.edu

1. Motivation
The scope and impact of APT attacks, including ransomware and supply-chain trojans, are unprecedented. They affected a wide range of organizations and daily life of millions of people.

$350M ransom was paid in 2020.
$4.4M and $11M were paid by Colonial Pipeline Co. and JBS USA, respectively, in ransom attacks in 2021.

100-280 companies install a trojanized version of the SolarWinds Orion. 800-1500 companies were compromise in the Kaseya hack in July 2021.

2. APT
APTs stands for Advanced Persistent Threats. They are usually highly sophisticated attacks aiming for long-term access.

Reconnaissance
Initial Penetration & Establish Foothold
Maintain Presence
Move Laterally

3. Existing Solutions

Rule-based approaches
- Mohammad et al.
- Chandra et al.
- HOLMES
- Poirot

Machine learning-based approaches
- APTGuard
- Kumar and Somani
- MLAPT
- UNICORE

Focus on detecting initial penetration
Focus on detecting malicious movement

4. Challenges
Hard to detect variants. APT always change its form to disguise itself, e.g., the new supply-chain trojan.

Hard to generalize. Different hosts and systems may different normal behaviors.

False alerts and missed detections. Due to the large amount of system and network data and lack of clear distinction between malicious and benign behavior.

5. Ongoing Work

Task 1: Analyze system monitoring on sensitive information (target of the attacks).

Task 2: Analyze recent attacks from different sources in various formats.

Task 3: Fine grain the system logs and focus on sensitive file and critical component access to reduce false positives.

Keywords

<table>
<thead>
<tr>
<th>Event ID</th>
<th>Task Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>5379</td>
<td>User Account Management</td>
</tr>
<tr>
<td>5379</td>
<td>User Account Management</td>
</tr>
<tr>
<td>4672</td>
<td>Special Logon</td>
</tr>
<tr>
<td>4624</td>
<td>Logon</td>
</tr>
<tr>
<td>4658</td>
<td>Other Object Access Events</td>
</tr>
<tr>
<td>4658</td>
<td>Other Object Access Events</td>
</tr>
<tr>
<td>4656</td>
<td>SAM</td>
</tr>
<tr>
<td>4656</td>
<td>SAM</td>
</tr>
</tbody>
</table>

A read operation is performed on stored credentials.

An access is requested to an object (security account manager).

This work has been supported by the Office of Naval Research under Grant No. N00014-22-1-2057.

Yao Group on Cyber Security
https://yaogroup.cs.vt.edu/