Poster: Towards Automated and Large-scale Cyber Attack Reconstruction with APT Reports

Zhenyuan Li[1,2], Ahmad Soltani[2], Anis Yusof[2], Aris Cahyadi Risdianto[2],
Kang Huang[2], Jun Zeng[2], Zhenkai Liang[2], Yan Chen[1,3]
[1]Zhejiang University, China, [2]National University of Singapore, [3]Northwestern University, USA

***Abstract*—Attacks by APT are increasingly prevalent and cause a huge impact. Researchers and analysts use CTI reports generated by cybersecurity organizations to learn and gain a better understanding of the current cyberthreat landscape. The diverse nature of CTI reports made them biased towards the organization's viewpoint despite reporting on the same incident. Furthermore, the static information from these reports limits the full potential of understanding the actual APT attacks. We propose an automated process to convert the rich details found in CTI reports to reconstruct a dynamic environment for analysts to use. Learning from an actual simulated environment provides invaluable insights in comparison to static details. By shortening the time of manually reconstructing error-prone and limited environments, we provide an automated platform for researchers and analysts to expedite their understanding and significantly reduce their turnaround time in addressing cyberthreats.**

I. Introduction

Cyber Threat Intelligence (CTI) reports are valuable sources that researchers and analysts seek to have a deeper understanding of the current APT activities and the cyberthreat landscape. These reports are used to obtain insights of vulnerabilities and their associated attack techniques. Attack patterns can also be learned from CTI reports [3], [4]. While CTI largely provides high-level information, fine-grained technical details that are relevant for analysis are generally omitted. In order for security experts and researchers to have an in-depth analysis, they have to bridge the knowledge gap between the real attacks and the CTI reports. This gap can be addressed by having a first-hand practical environment that thoroughly describes the threat's events in the CTI reports. Reproducing cyber attacks in a controlled environment benefits analysts significantly [7], where a high fidelity and live reconstructed environment allows the analysts to directly learn in-depth details that cannot be described in a CTI report. The replaying of the attack as described in the CTI report will enable analysts to apply their newly acquired practical knowledge in addressing current threats.

A. Motivation

CTI reports are generated for a human analyst to better understand the current cyberthreat landscape pertaining to a threat. These reports, however, are largely unstructured and generalized for a wide range of audiences. The static details in CTI report made it a challenge for analysts to have a deeper understanding. To better understand the cyberthreat landscape, analysts may prefer a dynamic and controlled environment where they can interact with the threat and obtain quality insights. Through interaction, they are able to better learn about the challenges and gain an in-depth hands-on knowledge. Manual reconstruction of environment based on CTI reports is a cumbersome process that is ineffective for rapid response. Having an end-to-end simulated and dynamic environment will allow researchers and analysts to gain dynamic insight from the static CTI reports.

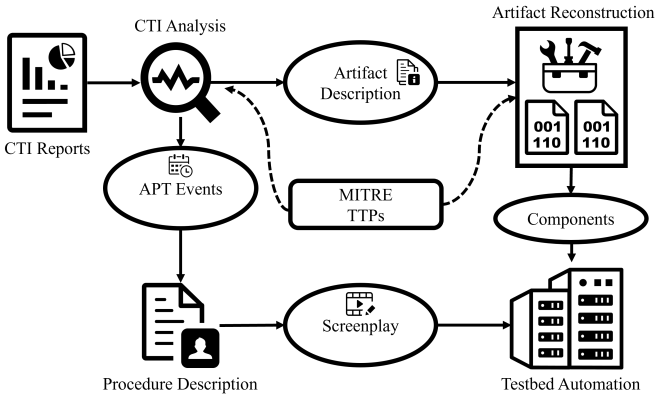


Fig. 1. The overall diagram that depict the process of reconstructing an environment that can be dynamically used by an analyst to gain better insights.

B. Observation

Several challenges have to be addressed in order to achieve environment reconstruction from CTI reports. Given that CTI reports are largely unstructured, the initial challenge is to extract key information from the available APT reports. However, relying only on the extracted information is insufficient as there will be missing details that is necessary to map and link the information. This information has to be bridged with the information that is found in the binaries related to the attack. Using the extracted information as a foundation, the next challenge is to reconstruct the APT kill-chain, with an accurate reconstruction based on a sound foundation. Finally, the challenge is the process of orchestrating an end-to-end transformation from the static details in CTI reports to the reconstruction of a dynamic environment. The comprehensiveness of the simulation will contribute to the usefulness for analysts to maximize their understanding that is reported in the CTI reports.

II. Overall Design

Figure 1 depicts the overview of our solution, reconstructing a live environment based on static CTI reports. A collection of CTI reports will be used as a reliable and rich source of knowledge base. This will be used as a foundation for reconstructing a live environment. It includes the following steps.

- CTI reports will go through a parsing pipeline to extract tactical knowledge and infrastructure information. By using both Natural Language Processing (NLP) [6] and domain-specific graph alignment algorithm [5] as part of the pipeline, critical descriptions are extracted to describe both the APT events and related artifacts which are the building blocks to form the necessary details required for reconstruction. Details from MITRE Tactics, Techniques, and Procedures (TTPs) are added to the extracted details for enrichment which describes the techniques performed by the APT threat actors and associated artifacts. The enrichment details from TTPs provide a bridge between the sequence of events and the related artifacts.
- Related artifacts are reconstructed to supplement the events in an APT attack. The artifacts include files, binaries (e.g., payload, IoT-based binaries & firmware), and configurations. The reconstruction of artifacts is necessary to provide an immersive experience of an actual APT attack where artifacts are dropped and may behave stealthily in order to avoid detection.
- Sometimes, artifacts utilized in the attack rely on external libraries or unknown environments and do not work. Then we will need an alternative solution. Therefore, we involved atomic attack technique scripts, such as Atomic-Red-Team [1].
- Based on the APT events, a procedure description is generated. The description is then translated to an event description language called *Screenplay* which will be used to describe the sequence of events during the reconstruction of the live environment. *Screenplay* outlines the attack events which serve as a temporal framework in the live environment.
- Using both *Screenplay* and related artifacts, an environment is reconstructed on the testbed. The reconstructed environment simulates the APT attack that is described in the CTI reports. Such simulated environment allows analysts to dynamically understand the APT attack.

III. Case Study

We carried out a case study using APT32 (OceanLotus) [2]. The APT operates numerous news websites that stalk and observe users' activities by tracking distinctive data. This allow the group to profile the users and redirect them to phishing websites that contain payload for infection. Understanding the modus operandi would be the key to gain insights that lead to both preventive and defensive solutions.

Detailed network topology, IoCs, and event sequence is extracted based on the CTI reports. This includes details such as website host for delivering fake news, malware dissemination servers, the Cobalt Strike C2 server, and related artifacts.

We are able to build an initial prototype of the APT. The reconstructed environments enable the users to simulate are the victims that visit the malicious website, where they are profiled and the process of malware delivery and execution are observed.

IV. Conclusion

We presented an automated process to reconstruct a live and dynamic environment from static details in CTI reports. The CTI reports are parsed by a pipeline that include NLP and graph alignment algorithm. This produces structured details that is translated to system-specific specifications. The specifications are used to reconstruct a dynamic environment on the testbed. Analysts are able to obtain richer insights by utilizing the reconstructed live environment that reflects the cyberthreat landscape presented in the static reports.

Acknowledgement

This work was supported by the National Research Foundation, Singapore under its NSoE DeST-SCI programme (Grant No. NSoE_DeST-SCI2019-0006) and under its National Cybersecurity R&D Program (Award No. NRF-NCL-P2-0001). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

References

[1] "Meet the Atomic Family — Atomic Red Team." [Online]. Available: https://atomicredteam.io/

[2] A. Dahan, "Operation cobalt kitty: A large-scale apt in asia carried out by the oceanlotus group." [Online]. Available: https://www.cybereason.com/blog/operation-cobalt-kitty-apt

[3] G. Husari, E. Al-Shaer, M. Ahmed, B. Chu, and X. Niu, "Ttpdrill: Automatic and accurate extraction of threat actions from unstructured text of cti sources," in *Proceedings of the 33rd Annual Computer Security Applications Conference*, ser. ACSAC 2017. New York, NY, USA: Association for Computing Machinery, 2017, p. 103–115. [Online]. Available: https://doi.org/10.1145/3134600.3134646

[4] G. Husari, E. Al-Shaer, B. Chu, and R. F. Rahman, "Learning apt chains from cyber threat intelligence," in *Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security*, ser. HotSoS '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3314058.3317728

[5] Z. Li, J. Zeng, Y. Chen, and Z. Liang, "AttacKG: Constructing Technique Knowledge Graph from Cyber Threat Intelligence Reports," nov 2021. [Online]. Available: https://arxiv.org/abs/2111.07093v1

[6] X. Liao, K. Yuan, X. Wang, Z. Li, L. Xing, and R. Beyah, "Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. New York, NY, USA: ACM, 2016. [Online]. Available: http://dx.doi.org/10.1145/2976749.2978315

[7] R. Uetz, C. Hemminghaus, L. Hackländer, P. Schlipper, and M. Henze, "Reproducible and Adaptable Log Data Generation for Sound Cybersecurity Experiments," in *Annual Computer Security Applications Conference*, Nov. 2021, arXiv: 2111.07847. [Online]. Available: http://arxiv.org/abs/2111.07847

Poster: Towards Automated and Large-scale Cyber Attack Reconstruction with APT Reports

Zhenyuan Li[†], Ahmad Soltani[§], Anis Yusof[§], Aris Cahyadi Risdianto[§], Kang Huang[§], Jun Zeng[§], Zhenkai Liang[§], Yan Chen[‡]
[†] Zhejiang University, [§] National University of Singapore, [‡] Northwestern University, USA

CTI Report

OceanLotus,https://unit42.paloaltonetworks.com/tracking-oceanlotus-new-downloader-kerrdown/,2019

Background & Insight

- Cyber threat intelligence (CTI) reports are widely available & reliable information source that describe APTs attacks. However, CTI reports are static and incomplete
- Researchers & analysts require technical fine-grained details while manual reconstruction of environment is costly & error-prone
- Interactive environment is a preferred method for investigators to obtain deeper insight

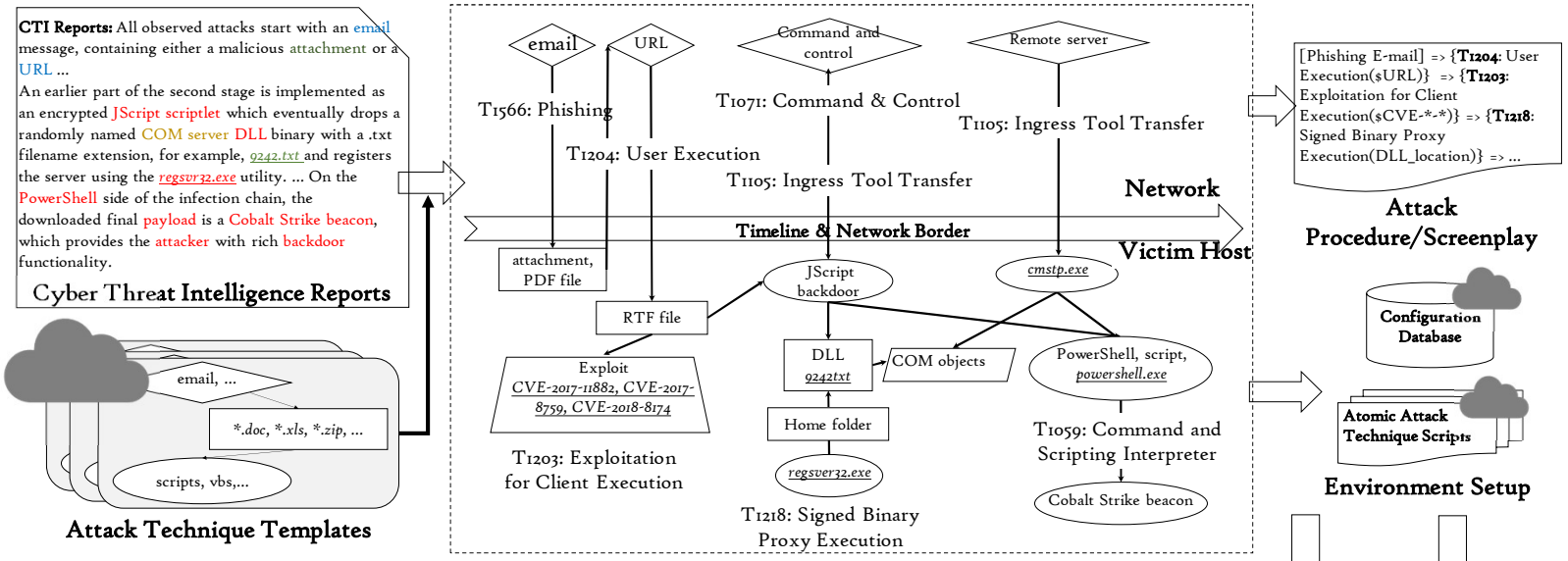
Challenge 1: It is difficult to extract useful and structure information from reports.

- CTI reports written in natural language are unstructured.
- The knowledge we want may scatter across multiple reports.

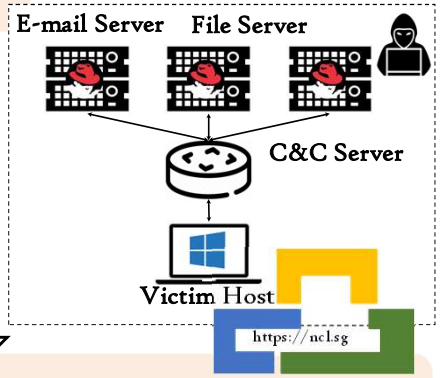
Challenge 2: Reconstructing attack in simulated environment requires lot of extra knowledge.

- Understanding the end-to-end cyber attack scenario.
- Recovering artifacts (e.g., binaries, IoT-based artifacts).

System Architecture & Methodology



Step 1: Analyzing CTI Reports of an APT ⇨ **Step 2: Reconstruct Attack Artifacts** ⇨ **Step3: APT Event Description**



Step 4: Reconstruct Dynamic Environment for Live Analysis

Case Study : APT32 Cyber Espionage

Network Topology, Configuration & Attack Sequences

- phishing web server
- malware host CDN server
- C&C server

Attack Screenplay

Schema that describe temporal activities and list of components

Attacker Toolkit & Artifacts

- Archived payload - Adobe_Flash_Install.rar
- Genuine installation file - Flash_Adobe_Install.exe
- DLL file – goopdate.dll
- BEACON binary returned by C2 server