Future Generation Computer Systems 89 (2018) 349–359

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Detection of advanced persistent threat using machine-learning correlation analysis

Ibrahim Ghafir [a,b,*], Mohammad Hammoudeh [c], Vaclav Prenosil [b], Liangxiu Han [c], Robert Hegarty [c], Khaled Rabie [c], Francisco J. Aparicio-Navarro [d]

[a] Department of Computer Science, Durham University, Durham, UK
[b] Faculty of Informatics, Masaryk University, Brno, Czech Republic
[c] Faculty of Science and Engineering, Manchester Metropolitan University, Manchester, UK
[d] School of Engineering, Newcastle University, Newcastle upon Tyne, UK

ARTICLE INFO

Article history:
Received 30 March 2018
Received in revised form 3 June 2018
Accepted 28 June 2018
Available online 6 July 2018

Keywords:
Cyber attacks
Advanced persistent threat
Malware
Intrusion detection system
Alert correlation
Machine learning

ABSTRACT

As one of the most serious types of cyber attack, Advanced Persistent Threats (APT) have caused major concerns on a global scale. APT refers to a persistent, multi-stage attack with the intention to compromise the system and gain information from the targeted system, which has the potential to cause significant damage and substantial financial loss. The accurate detection and prediction of APT is an ongoing challenge. This work proposes a novel machine learning-based system entitled MLAPT, which can accurately and rapidly detect and predict APT attacks in a systematic way. The MLAPT runs through three main phases: (1) Threat detection, in which eight methods have been developed to detect different techniques used during the various APT steps. The implementation and validation of these methods with real traffic is a significant contribution to the current body of research; (2) Alert correlation, in which a correlation framework is designed to link the outputs of the detection methods, aims to identify alerts that could be related and belong to a single APT scenario; and (3) Attack prediction, in which a machine learning-based prediction module is proposed based on the correlation framework output, to be used by the network security team to determine the probability of the early alerts to develop a complete APT attack. MLAPT is experimentally evaluated and the presented system is able to predict APT in its early steps with a prediction accuracy of 84.8%.

1. Introduction

The volume, complexity and variety of Cyber attacks are continually increasing. This trend is currently being driven by cyber warfare and the emergence of the Internet of Things [1–3]. The annual cost of cyber attacks was $3 trillion in 2015 and it is expected to increase to more than $6 trillion per annum by 2021 [4]. This high cost has brought much interest in research and investment towards developing new cyber attacks defence methods and techniques [5–8]. Although virus scanners, firewalls and intrusion detection and prevention systems (IDPSs) have been able to detect and prevent many of cyber attacks, cyber-criminals in turn have developed more advanced methods and techniques to intrude into the target's network and exploit their resources, targeting both wired and wireless communications [9,10]. In addition, many of the defence approaches against cyber attacks consider those attacks are targeting random networks, so they assume that if the company's network is well protected, the attacker can surrender and move onto an easier target. Nonetheless, according to a technical report by Trend Micro [11], this assumption is no longer valid with the rise of targeted attacks, Advanced Persistent Threats (APTs), in which both cyber-criminals and hackers are targeting selected organizations and persisting until they achieve their goals.

The APT attack is a persistent, targeted attack on a specific organization and is performed through several steps [12]. The main aim of APT is espionage and then data exfiltration. Therefore, APT is considered as a new and more complex version of multi-step attack. These APTs present a challenge for current detection methods as they use advanced techniques and make use of unknown vulnerabilities. Moreover, the economic damage due to a successful APT attack significant. The potential cost of attacks is the major motivation for the investments in intrusion detection and prevention systems [13]. APTs are currently one of the most serious threats to companies and governments [14].

Most of the research in the area of APT detection, has focused on analysing already identified APTs [15–21], or detecting a particular APT that uses a specific piece of malware [22]. Some works have

* Corresponding author.
E-mail address: ibrahim.ghafir@durham.ac.uk (I. Ghafir).

https://doi.org/10.1016/j.future.2018.06.055
0167-739X/

attempted to detect novel APT attacks. However, they face serious shortcomings in achieving real time detection [23], detecting all APT attack steps [23], balance between false positive and false negative rates [22], and correlating of events spanning over a long period of time [24,25]. The existing work is encouraging. However, the accurate and timely detection of APT remains a challenge.

In this work, we have developed a novel machine learning-based system, called MLAPT, which can accurately, and quickly detect and predict APT attacks in a holistic way, making a significant contribution to the field of intrusion detection systems (IDS). MLAPT runs through three main phases: threat detection, alert correlation and attack prediction, the major contributions of this work include:

- Threat detection: the aim of this first phase is to detect threats during the multi-step APT attack. We have developed eight methods/modules to detect various attacks used in one of the APT attack steps. These include disguised exe file detection (DeFD), malicious file hash detection (MFHD), malicious domain name detection (MDND), malicious IP address detection (MIPD), malicious SSL certificate detection (MSSLD), domain flux detection (DFD), scan detection (SD), and Tor connection detection (TorCD). The output of this phase is alerts, also known as events, triggered by the individual modules. All the methods have been evaluated using real network traffic.
- Alert correlation: this second phase of the alert correlation intends to correlate the alerts produced in the first phase with one APT attack scenario. The main objective of using the correlation framework is to reduce the false positive rate of the MLAPT detection system. The process in this phase undergoes three main steps: alerts filter (AF), to identify redundant or repeated alerts; clustering of alerts (AC), which most likely belong to the same APT attack scenario; and correlation indexing (CI), to evaluate the degree of correlation between alerts of each cluster.
- Attack prediction: in the final phase, a machine-learning-based prediction module (PM) is designed and implemented based on a historical record of the monitored network. This module can be used by the network security team to determine the probability of the early alerts to develop a complete APT attack.
- The proposed MLAPT system is able to process and analyse the network traffic in real time without needing to store data, and make possible the early prediction of APT attacks so that an appropriate and timely response can take place before the attack completes its life cycle.

The remainder of this paper is organized as follows. Section 2 presents the related work to APT detection. The proposed APT detection system and its architecture are described in Section 3. Section 4 explains the implementation of the proposed approach. The evaluation results and the performance comparison with the existing APT detection system are shown in Sections 5 and 6 respectively. Section 7 concludes the paper.

2. Related work

The APT detection has been a challenge for the current Intrusion Detection Systems (IDSs), and much research has been conducted to address this type of multi-stage attack. Table 1 describes current APT detection systems and mesmerizes their limitations.

TerminAPTor, an APT detector, is described in [26]. This detector uses information flow tracking to find the links between the elementary attacks, which are triggered within the APT life cycle. TerminAPTor depends on an agent, which can be a standard intrusion detection system, to detect those elementary attacks. The authors evaluated TerminAPTor by simulating only two APT scenarios and demonstrated that the APT detector needs to be improved by filtering the false positives.

An APT detection system based on C&C domains detection is introduced in [27]. This work analyses the C&C communication and states a new feature that the access to C&C domains is independent while the access to legal domains is correlated. Despite the fact that the detection system achieved significant results when validated on a public dataset, the authors mentioned that the detection can be easily evaded when the infected hosts connect to the C&C domains while users are surfing the Internet. Moreover, missing the detection of C&C domains leads to failure in APT detection since this system depends on detecting only one step of the APT life cycle.

An approach for APT detection based on spear phishing detection is explored in [28]. This approach depends on mathematical and computational analysis to filter spam emails. Tokens, which are considered as a group of words and characters such as (click here, free, Viagra, replica), should be defined for the detection algorithm to separate legitimate and spam emails. However, the spear phishing email might not include any of the tokens which are necessary for the algorithm process. Additionally, depending on one step for APT detection leads the system to fail when missing that step.

A statistical APT detector, similar to TerminAPTor detector, is developed in [29]. This system considers that APT undergoes five states which are delivery, exploit, installation, C&C and actions; and several activities are taken in each state. The generated events in each state are correlated in a statistical manner. This system requires significant expert knowledge to set up and maintain.

An active-learning-based framework for malicious PDFs detection is suggested in [30]. These malicious PDFs can be used in the early steps of APT to get the point of entry. The system collects all PDFs transferred over the network, then all known benign and malicious files are filtered by the "known files module" which depends on white lists, reputation systems and antivirus signature repository. Following this, the remaining files "unknown files" are checked for their compatibility as viable PDF files. This approach detects only one step of the APT life cycle.

An approach based on Data Leakage Prevention (DLP) is proposed in [31]. This approach focuses on detecting the last step of APT which is the data exfiltration. A DLP algorithm is used to process the data traffic to detect data leaks and generate "fingerprints" according to the features of the leak. The proposed system utilizes external cyber counterintelligence (CCI) sensors in order to track the location or path of the leaked data. This approach is limited to detect only one step of APT which is the data exfiltration. In addition, it cannot achieve the real time detection as the CCI analysis unit should wait for the information from the sensors. Moreover, it is not guaranteed that the CCI sensors can provide the required information regarding the leaked data fingerprints. This approach also introduces privacy issues, whereby actors in the CCI have access to the data stored and transferred by all users of the systems.

A working prototype, SPuNge, is presented in [23]. The proposed approach depends on the gathered data on the hosts' side and aims to detect possible APT attacks. SPuNge undergoes two main phases, in the first one, the detected malicious URLs are analysed. Those URLs can be connected by the hosts' computers over HTTP(S) with an Internet browser or by malware installed on the infected machines. The computers which show a similar activity are then determined. This system depends on detecting one activity of the APT attack, which is malicious URL connection, and does not consider the other activities of APT. Meaning, if the detection system misses the malicious URL connection, the whole APT scenario will not be detected. Additionally, the system cannot achieve real time detection.