Chapter 7
Threat Hunting in Windows Using Big Security Log Data

Mohammad Rasool Fatemi
University of New Brunswick, Canada

Ali A. Ghorbani
University of New Brunswick, Canada

ABSTRACT

System logs are one of the most important sources of information for anomaly and intrusion detection systems. In a general log-based anomaly detection system, network, devices, and host logs are all collected and used together for analysis and the detection of anomalies. However, the ever-increasing volume of logs remains as one of the main challenges that anomaly detection tools face. Based on Sysmon, this chapter proposes a host-based log analysis system that detects anomalies without using network logs to reduce the volume and to show the importance of host-based logs. The authors implement a Sysmon parser to parse and extract features from the logs and use them to perform detection methods on the data. The valuable information is successfully retained after two extensive volume reduction steps. An anomaly detection system is proposed and performed on five different datasets with up to 55,000 events which detects the attacks using the preserved logs. The analysis results demonstrate the significance of host-based logs in auditing, security monitoring, and intrusion detection systems.

INTRODUCTION

One of the results of the continually growing number of devices connected to the Internet is the production of vast amounts of logs and other data. This ever-increasing volume of data, especially security logs, should be stored and can be analyzed for different proposes. Whether it be an attack or a malicious activity, with an excellent and in-depth analysis of the right log data, we can find it. Also, event pattern discovery, one of the most critical tasks in security log management and analysis, can be used to build security log file profiles thus anomalous lines in the log files can be identified. Based on recent researches, despite having notable progress, there are several remaining challenges in log analysis, and there is still a long way to go when it comes to security analysis of big data. Having more than 75% of the share of

DOI: 10.4018/978-1-5225-9742-1.ch007

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

desktop operating systems (Statista, 2019), Windows is considered as the most widely used desktop operating system worldwide, and hence, there is a massive number of logs generated by Windows-operated workstations and laptops every day. Network log analysis has come a long way so far, but system log analysis seems to have been undervalued compared to that. Sysmon could be one possible way to help solve this issue. As a good start, security researchers at Microsoft developed Sysmon to help with Windows auditing process. This logging tool can help address several above issues if properly configured and deployed. There are many system information and event management software (SIEM) that can help with the analysis of these logs and other logs from different sources. In this chapter, the authors aim to show the importance of host-based logs for security analysis. They also list an address related challenges and discuss potential solutions. This chapter will demonstrate how an enormous amount of security logs can be handled and used as the sole source of information for anomaly detection.

Here, the researchers propose a set of tools to help detect anomalies in Windows workstations. The goal of the chapter is to provide a set of simple and easy-to-deploy tools that can be used to track and hunt malicious activities and help with the response after an attack. To achieve this, authors focus on host-based logs generated by Sysmon and aim to show their importance and demonstrate that they can be used as an exclusive source of information for anomaly detection. The researchers take a two-step procedure to reduce the size of the logs produced by Sysmon. First, they define several rules to include relevant information and exclude network logs as well as less informative and noisy events like empty key retraction processes. They create several datasets in two different virtual environments using the two most common versions of Windows. The authors implement a fast and highly configurable Sysmon parser that can clean and parse Sysmon logs which also extracts features based on its configuration. The second step in data reduction is done after parsing. Using clustering techniques, they detect outliers to reduce the size of the data extensively. The outliers are analyzed and flagged in the anomaly detection engine. Here logs are analyzed by all engines available in VirusTotal where the malicious ones are normally detected. A human analyst then checks the few remaining logs that are flagged as *suspicious* or *unknown* in the previous step. All attacks are detected successfully at the end. In summary, the following are the chapter's contributions:

- Implementation of a comprehensive highly configurable Sysmon parser for parsing and extracting features from the logs.
- Proposal of a hierarchical approach to considerably reduce the volume of Windows generated Sysmon logs while preserving actionable intelligence.
- Introduction of an anomaly detection engine that successfully detects attacks and malware on different datasets.
- Creation of 15 different Sysmon events datasets with no modification or anonymization.

The remainder of this chapter is organized as follows. Log-based anomaly detection systems, log clustering, and log parsing tools are discussed in Section Background. After that, challenges in the area are reviewed in the next section. In Section Log Analysis, the authors provide a detailed description of Sysmon and some of its capabilities. The next section discusses the datasets along with the malware and the environment used to create them. Subsequently, details of deploying the system and the proposed anomaly detection engine's architecture are described, followed by the analysis process. Results of the analysis are presented in the next section. The conclusion and potential future work are discussed in the last section.

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/threat-hunting-in-windows-using-big-security-log-data/234810

Related Content

Designing a Secure Cloud Architecture: The SeCA Model
Thijs Baarsand Marco Spruit (2012). *International Journal of Information Security and Privacy (pp. 14-32).*
www.irma-international.org/article/designing-secure-cloud-architecture/64344

Applying Blockchain Security for Agricultural Supply Chain Management
Amarsinh V. Vidhate, Chitra Ramesh Saraf, Mrunal Anil Wani, Sweta Siddarth Waghmareand Teresa Edgar (2023). *Research Anthology on Convergence of Blockchain, Internet of Things, and Security (pp. 1229-1239).*
www.irma-international.org/chapter/applying-blockchain-security-for-agricultural-supply-chain-management/310505

Applying Enterprise Risk Management on a Fiber Board Manufacturing Industrial Case
Syed Aftab Hayat (2014). *International Journal of Risk and Contingency Management (pp. 51-66).*
www.irma-international.org/article/applying-enterprise-risk-management-on-a-fiber-board-manufacturing-industrial-case/120557

Disassociations in Security Policy Lifecycles
Michael Lapkeand Gurpreet Dhillon (2015). *International Journal of Information Security and Privacy (pp. 62-77).*
www.irma-international.org/article/disassociations-in-security-policy-lifecycles/145410

Incident Preparedness and Response: Devlopming a Security Policy
Warren Wylupski, David R. Championand Zachary Grant (2008). *Information Security and Ethics: Concepts, Methodologies, Tools, and Applications (pp. 2366-2387).*
www.irma-international.org/chapter/incident-preparedness-response/23227