Poster: Detecting Monitor Compromise using Evidential Reasoning

Uttam Thakore
University of Illinois at Urbana-Champaign
thakore1@illinois.edu

Ahmed Fawaz
University of Illinois at Urbana-Champaign
afawaz2@illinois.edu

William H. Sanders
University of Illinois at Urbana-Champaign
whs@illinois.edu

ABSTRACT

Stealthy attackers often disable or tamper with system monitors to hide their tracks and evade detection. In this poster, we present a data-driven technique to detect such monitor compromise using evidential reasoning. Leveraging the fact that hiding from multiple, redundant monitors is difficult for an attacker, to identify potential monitor compromise, we combine alerts from different sets of monitors by using Dempster-Shafer theory, and compare the results to find outliers. We describe our ongoing work in this area.

CCS CONCEPTS

• **Security and privacy** → **Intrusion detection systems**; **Distributed systems security**;

KEYWORDS

security, intrusion detection, machine learning, evidential reasoning

1 INTRODUCTION

Over the years, researchers have come up with many novel ways to detect intrusions in enterprise systems that use increasingly diverse sources of system and network data. However, in most such work, there was a fundamental assumption that the data being collected were reliable and uncompromised; indeed, many papers explicitly consider the effects of compromised data to be outside their scope [2, 3]. In light of recent work in adversarial learning [1] and the rise of advanced persistent threats, we find that the assumption of trust in monitor data should be reevaluated, and that detection of monitor compromise should be addressed as an independent problem.

2 MODEL AND METHODOLOGY

Here, we describe how we combine information from redundant, heterogeneous monitors to detect monitor compromise. Let $m_A, m_B, \ldots$ be the monitors available in a system. Each monitor can generate multiple *alerts*, each of which is characterized by an *alert type* (e.g., IDS alert type, log key as defined in [2]), identification features (e.g., timestamp, source IP), and type-specific parameters. For a given monitor $m_{\mathcal{L}}$, let the set of alerts generated by $m_{\mathcal{L}}$ be $\{\ell_1, \ell_2, \ldots\}$.

During a particular *event*, which we define as a distinct action or workflow a user or process might undertake in the system, each monitor will generate alerts from a limited subset of all alerts it could generate. While some alert types may be generated by many types of events, overall distributions of alerts will generally be distinct for each event type, so given a particular log trace, there will likely be a limited set of events consistent with the data.

Let us consider a combination operation $C(\sigma)$ that takes a set of observed alerts σ as input and produces an assignment of likelihood scores $\vec{\mathbf{e}}_\sigma$ for all possible event types e_i in the system. Assuming that only one event takes place during the period and assuming no compromise, we would expect very few elements of $\vec{\mathbf{e}}_\sigma$ to be active (nonzero), and only one to have a high likelihood value.

Furthermore, if there is no monitor compromise, we would reasonably expect that the results of combining alerts from different sets of monitors should not strongly conflict with one another. That is, if we let $\sigma_{\mathcal{M}}$ be the set of alerts within σ generated by all monitors in some set $\mathcal{M}$, then $C(\sigma_{\mathcal{M}}) \cdot C(\sigma_{\mathcal{N}})$ should be close to 1 for any $\mathcal{M}$ and $\mathcal{N}$, provided that $\sigma_{\mathcal{M}} \neq \emptyset$ and $\sigma_{\mathcal{N}} \neq \emptyset$. However, if an adversary were to compromise monitor m and inject or drop alerts to yield $\sigma\prime$, we would expect the results of combination to diverge, causing the values of all $C(\sigma\prime_{\mathcal{M}}) \cdot C(\sigma\prime_{\mathcal{M}\setminus m})$ to decrease.

We use that observation to devise an approach for detection of monitor compromise. Given a large set of event and alert types, we track the values of $C(\sigma_{\mathcal{M}})$ for all $\sigma_{\mathcal{M}}$ in a system over time by combining all observed alerts for a system within a time window $[t, t + \Delta_w)$. For each timestamp, we apply outlier detection techniques to identify candidates for monitor compromise, and consider a monitor to be compromised if some $C(\sigma_{\mathcal{M}})$ deviate significantly from all others for a sufficient duration, where $m \in \mathcal{M}$.

We use Dempster-Shafer theory (DST) [4] for the operator C because of its versatility in combining heterogeneous evidence and its inherent ability to handle uncertainty and conflicting evidence. We train basic belief assignments (BBAs) for each alert type a using labeled event occurrence data during normal activity such that the plausibility for each singleton event e_i is its conditional probability $P(e_i|a)$. To gracefully handle noisy data, which can manifest as conflict during combination, we use Yager's rule for combination [4] and add a small amount of uncertainty (uniform noise) to the BBAs.

3 ONGOING WORK

We are currently evaluating our approach on simulated data and a real dataset, into which we inject monitor compromise based on real and theoretical attacks. Our initial results show that our technique can be used to detect various types of monitor compromise.

REFERENCES

[1] Igino Corona, Giorgio Giacinto, and Fabio Roli. 2013. Adversarial Attacks Against Intrusion Detection Systems: Taxonomy, Solutions and Open Issues. *Inf. Sci.* 239 (Aug. 2013), 201–225. https://doi.org/10.1016/j.ins.2013.03.022

[2] Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar. 2017. DeepLog: Anomaly Detection and Diagnosis from System Logs Through Deep Learning. In *Proc. of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*. ACM, New York, NY, USA, 1285–1298. https://doi.org/10.1145/3133956.3134015

[3] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. 2015. Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence. In *Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. ACM, New York, NY, USA, 554–566. https://doi.org/10.1145/2810103.2813614

[4] Karl Sentz and Scott Ferson. 2002. *Combination of evidence in Dempster-Shafer theory.* Technical Report SAND 2002-0835. Sandia National Laboratories, Albequerque, NM. http://prod.sandia.gov/techlib/access-control.cgi/2002/020835.pdf

HoTSoS '18, April 10–11, 2018, Raleigh, NC, USA
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6455-3/18/04.
https://doi.org/10.1145/3190619.3191693