Accepted Manuscript

Dendron: Genetic Trees driven rule induction for network intrusion detection systems

Dimitrios Papamartzivanos, Félix Gómez Mármol,
Georgios Kambourakis

PII: S0167-739X(16)30546-5
DOI: https://doi.org/10.1016/j.future.2017.09.056
Reference: FUTURE 3712

To appear in: *Future Generation Computer Systems*

Received date : 8 November 2016
Revised date : 26 July 2017
Accepted date : 22 September 2017

Please cite this article as: D. Papamartzivanos, F.G. Mármol, G. Kambourakis, *Dendron*: Genetic Trees driven rule induction for network intrusion detection systems, *Future Generation Computer Systems* (2017), https://doi.org/10.1016/j.future.2017.09.056

Dendron: Genetic Trees driven Rule Induction for Network Intrusion Detection Systems

Dimitrios Papamartzivanos[a,*], Félix Gómez Mármol[b], Georgios Kambourakis[a,c]

[a] *Department of Information & Communication Systems Engineering, University of the Aegean, 83200 Samos, Greece*
[b] *Department of Information and Communications Engineering, University of Murcia, 30100 Murcia, Spain*
[c] *Computer Science Department, George Mason University, Fairfax, VA 22030, USA*

Abstract

Intrusion detection systems (IDSs) are essential entities in a network topology aiming to safeguard the integrity and availability of sensitive assets in the protected systems. In misuse detection systems, which is the topic of the paper at hand, the detection process relies on specific attack signatures (rules) in an effort to distinguish between legitimate and malicious network traffic. Generally, three major challenges are associated with any IDS of this category: identifying patterns of new attacks with high accuracy, ameliorating the human-readability of the detection rules, and rightly designating the category these attacks belong to. To this end, we propose *Dendron*, a methodology for generating new detection rules which are able to classify both common and rare types of attacks. Our methodology takes advantage of both Decision Trees and Genetic Algorithms for the sake of evolving linguistically interpretable and accurate detection rules. It also integrates heuristic methods in the evolutionary process aiming to deal with the challenging nature of the network traffic, which generally biases machine learning techniques to neglect the minority classes of a dataset. The experimental results, using KDDCup'99, NSL-KDD and UNSW-NB15 datasets, reveal that *Dendron* is able to achieve superior results over other state-of-the-art and legacy techniques under several classification metrics, while at the same time is able to significantly detect rare intrusive incidents.

Keywords: Intrusion Detection Systems, Misuse Detection, Decision Trees, Genetic Algorithms, Machine Learning, Information Systems Security.

1. Introduction

Intrusion detection systems (IDSs) aim to protect systems from a variety of attacks threatening their confidentiality, integrity and availability. Internet is an open and active ecosystem which evolves rapidly, whereas new types of attacks emerge as the aggressors become more sophisticated. To make matters worse, new threats and vulnerabilities emerge on a daily basis, increasing in this way the risk for critical infrastructures to become compromised. To battle against intrusive events, an IDS must be always updated to be able to detect and possibly cope with novel attacks.

IDSs can be classified into two major categories, namely *Anomaly Detection Systems* (sometimes known as profile-based detection) and *Misuse Detection Systems* (sometimes referred to as signature-based detection). The former are destined to identify deviations from a *normal* profile behavior in order to detect malicious actions. Even though this kind of systems perform better in detecting previously unseen attacks, they usually suffer from a high False Positive (FP) rate rendering them unpractical solutions for protecting a sensitive infrastructure. This limitation is addressed by Misuse Detection Systems, where the detection process is based on known signatures, that is, detection rules aiming to distinguish legitimate traffic instances from the malicious ones. Nevertheless, while these systems are able to detect known attacks they miss to identify novel attacks or variations of known ones. Thus, the detection ability of a Misuse detection system is directly affected by the freshness of the detection rules it possesses.

Keeping a detection rules database up-to-date is a challenging task that involves system administrators' supervision. Considering the huge traffic volume passing through central network nodes like an IDS, one easily concludes that the rule generation process is necessary to be supported by automated tools able not only to distinguish between legitimate and malicious traffic, but also to infer the specific class of an attack occurring in the target system. Moreover, the set of the detection rules should enable the system to identify attacks with high Attack Detection Rate (ADR) while keeping the False Alarm Rate (FAR) low. Generally, false alarms are a cardinal concern in the field, especially when an IDS is involved in collaborative infrastructures [1], [2] and reputation systems [3],

*Corresponding author
Email addresses: dpapamartz@aegean.gr (Dimitrios Papamartzivanos), felixgm@um.es (Félix Gómez Mármol), gkamb@aegean.gr (Georgios Kambourakis)