Automatic and on-the-fly Firewalls Configuration

Auteur : Rossetto, Vincent
Promoteur(s) : Donnet, Benoît
Faculté : Faculté des Sciences appliquées
Diplôme : Master en sciences informatiques, à finalité spécialisée en "computer systems security"
Année académique : 2020-2021
URI/URL : http://hdl.handle.net/2268.2/11610

Automatic and on-the-fly firewalls configuration

Master's thesis - summary

- Author: Vincent Rossetto
- Supervisor: Benoit Donnet

- Section: Computer Science (with a professional focus on computer systems security)
- Academic year: 2020-2021

Summary

Automation has the potential to improve reliability and efficiency wherever it takes root. However, firewalls are still configured manually, which is difficult and error prone. As a firewall is essentially a program that determines whether incoming traffic is legitimate or malicious, firewall configuration, given a representative dataset of recorded traffic, can be treated as a classification problem, i.e. a discrete output supervised machine learning problem. The possibility of using a machine learning framework for automatic firewall configuration was studied. A training dataset, composed both of legitimate traffic and attacks, was generated and statistically described both in terms of packets and flows. With the guidance of the statistical observations, per-packet and per-flow features were defined, and the performance of classification algorithms using those features was evaluated. The benefits of a potential feature pre-processing were evaluated. It was found to be useful and even necessary in some cases. A limitation was found in the use of classification algorithms in the form of their lack of interpretability. Practical use of firewalls requires that they can be reconfigured in the case where they make the wrong decision. Such classification errors cannot be avoided completely even with a dataset containing enough examples. This necessity makes the direct use of opaque classifiers for automatic firewall configuration impossible. In conjunction with performance constraints, the most realistic solution consists in configurating classical rule-based firewalls. Rule extraction techniques from opaque classifiers are discussed. The decision tree algorithm was used both to extract rules directly from data and from models built using other classification algorithms. Both approaches yielded a similar classification performance. Finally, rules were extracted from decision trees and translated into a valid format to configure a firewall based on mmb (Modular MiddleBox) using a subset of the per-packet features.