Enhancing Intrusion Detection System using XDP with eBPF

K.Geethika, G.Krishna Prathibha, G.Sneha

Under the esteemed guidance of
Mr. N. Anand
Assistant Professor

VISHNU
UNIVERSAL LEARNING

Bachelor of Technology
Department of Information Technology
BVRIT HYDERABAD College of Engineering for Women

September 13, 2023

Overview

1. Introduction

2. Plan of Action

Introduction

Intrusion Detection System (IDS):
An Intrusion Detection System (IDS) is a network security technology originally built for detecting vulnerability exploits against a target application or computer.

eBPF (extended Berkely Packet Filtering):
eBPF is a revolutionary technology with origins in the Linux kernel that can run sandboxed programs in a privileged context such as the operating system kernel. It is used to safely and efficiently extend the capabilities of the kernel without requiring to change kernel source code or load kernel modules.

XDP (eXpress Data Path) :
XDP is an eBPF-based high-performance data path used to send and receive network packets at high rates by bypassing most of the operating system networking stack.

Kernel-level Packet Filtering :
Packet filtering at the kernel level refers to the process of inspecting incoming and outgoing network packets at the operating system's kernel, which is the core component of the operating system that manages system resources.
This allows the kernel to make decisions about whether to allow or block these packets based on predefined rules, enhancing security by filtering packets at a low level.

Machine Learning Algorithms

Random Forest :

Random Forest is a popular machine learning algorithm that combines multiple decision trees to make robust predictions. It aggregates the results of individual trees, reducing overfitting and improving accuracy in tasks like classification and regression. It's known for its versatility and resistance to overfitting.

Decision Tree :

A Decision Tree is a machine learning algorithm that recursively partitions data into subsets by selecting the most informative features at each node. It's employed for classification and regression tasks, providing interpretability and flexibility but requiring pruning to prevent overfitting.

Support Vector Machine (SVM) :
Support Vector Machine (SVM) is a versatile machine learning algorithm used for classification and regression. It finds an optimal hyperplane to separate data into distinct classes while maximizing the margin between them. SVM's ability to handle high-dimensional and non-linear data makes it valuable in various applications.

Twin Support Vector Machine (Twin SVM) :
Twin Support Vector Machine (Twin SVM) is a variation of the traditional Support Vector Machine (SVM). It's designed for multi-class classification tasks, aiming to find a combination of two hyperplanes that best separate data into multiple classes while maximizing the margin between them. Twin SVM can be particularly useful when dealing with imbalanced datasets or multi-class classification problems.

Plan of Action

The steps involved in this process are:

- We are making use of CIC-IDS-2017 dataset.
- The raw data is pre-processed, that includes data transmission, cleaning (infinity, NULL values, etc.), and reduction (size).
- Specific features are extracted from the pre-processed data using ANOVA F-test method.
- The extracted features are analyzed for intrusion detection using ML algorithms.

  – After training and testing with Random Forest and Decision Tree, SVM and TwinSVM, model parameters are exported.
- write an eBPF program that captures network traffic and utilizes trained model parameters to detect attacks within the kernel
- write a XDP with eBPF program that utilizes trained model parameters to detect attacks (filter packets) faster and accurate within the kernel.

Thank you