

**Scenario Independent Feature Extraction for**

**Detecting Intrusions over TCP/IP connections**

*Report submitted to the SASTRA Deemed to be University*

*As the requirement for the course*

**CSE 302: COMPUTER NETWORKS**

*Submitted by*

**Sankara Narayanan S**

**(Reg. No.: 124156079, B. Tech, CSE spl. Artificial intelligence and data science)**

**NOVEMBER 2022**

**Abstract:**

Work inspired by E. Viegas, A. Santin and V. Abreu's paper:

"Enabling Anomaly-based Intrusion Detection Through Model Generalization".

The goal is to recreate an Intrusion Detection System (IDS) by training a **Machine Learning** model based on the traffic recreated within a virtual environment. The traffic generated is difficult to use to train machine learning models as it is scenario dependent, so it would lead to models trained for that specific scenario. To solve this problem, it is necessary to treat the generated traffic to be independent of the simulated session (virtual or real environment).

The traffic generated (HTTP, SMTP, SMNP, SSH) is listened to using **tcpdump**; The generated .dump file is converted into a file called totaltraffic.c containing array C using **wireshark**; Featuresextractor.py containing **python** code is launched;

In the end, 50 features independent of the scenario are obtained and can be used for model training.

**KEY WORDS**: Machine Learning,tcpdump,wireshark,python