A Comparative Analysis of Machine Learning Algorithms for Website Traffic Classification from Network Packets

XXX
YYY
xxx@xxx.com
yyy@yyy.com
William & Mary
Williamsburg, Virginia, USA

ABSTRACT

Accurately Identifying web traffic destination and origins is crucial for the efficiency of a network. This project explores the potential of machine learning in reference to web traffic classification based on the analysis of network packets. We monitored and analyzed web traffic data from ChatGPT, Blackboard, and Linkedin, with the objective of building models which will be able to predict the web traffic origin of a specific packet. The collection of data was performed using wireshark, then the data was reformatted to eliminate bias and get more accurate results. Using the collected data we then trained four models which had varying levels of accuracy, Logistic Regression (56%), K-Nearest Neighbors (77%), Random Forest (78%), and finally a neural network (80%). This project shows the importance of machine learning within the field of network traffic analysis as automaton is much more efficient and precise compared to manually examining web traffic, especially in a scale as large as the internet.

KEYWORDS

Idk what to put for keywords