Resumen

En el contexto actual, la clasificación precisa de aplicaciones en el tráfico de red representa un desafío significativo para garantizar el funcionamiento óptimo y seguro de las redes. Este proyecto involucró un estudio exhaustivo sobre la clasificación de aplicaciones en el tráfico de red, adoptando un enfoque de granularidad fina en el conjunto de datos. Se extrajeron flujos de datos utilizando el software CICflowMeter, con dos ajustes de tiempo de espera de flujo diferentes: 120 segundos y 15 segundos. Con el objetivo de lograr una clasificación efectiva, se aplicó la metodología SEMMA y se emplearon cuatro algoritmos de aprendizaje supervisado: Máquina de Vectores de Soporte (Support Vector Machine, SVM), Árbol de Decisión (Decision Tree, DT), Bosque Aleatorio (Random Forest, RF) y K-Vecinos más Cercanos (K-Nearest Neighbors, KNN). Estos algoritmos se utilizaron con dos grupos de características diferentes: uno con 25 características y otro con 15 características. Los resultados obtenidos revelaron que el conjunto de datos con un tiempo de espera de flujo de 15 segundos y el grupo de 15 características lograron el mayor nivel de precisión, con resultados de entrenamiento de precisión en los algoritmos RF (99.99%), DT (99.89%), KNN (99.92%) y SVM (92.06%). Estos hallazgos destacan la notable efectividad del modelo RF en la clasificación de aplicaciones en el tráfico de red.

Palabras clave: Redes Definidas por Software, Aprendizaje Automático, Clasificación de aplicaciones.

Abstract

In the current context, accurate classification of applications in network traffic represents a significant challenge to ensure networks' optimal and secure operation. This project involved a comprehensive study on the classification of applications in network traffic, adopting a fine-grained approach to the data set. Data streams were extracted using CICflowMeter software, with two different flow timeout settings: 120 seconds and 15 seconds. The SEMMA methodology was applied to achieve effective classification, and four supervised learning algorithms were employed: Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN). These algorithms were used with two different feature sets: one with 25 and the other with 15. The results obtained revealed that the dataset with a flow waiting time of 15 seconds and the 15-feature group achieved the highest level of accuracy, with accuracy training results in the RF (99.99%), DT (99.89%), KNN (99.92%) and SVM (92.06%) algorithms. These findings highlight the remarkable effectiveness of the RF model in classifying applications in network traffic.

Keywords: Software-Defined Networking, Machine Learning, Application Classification