Detection of Distributed Denial Of Service (DDoS) Attacks Using a Hybrid Approach: Recurrent Neural Network (RNN) and Feature Selection Learning Vector Quantization (LVQ)-Autoencoder.

Sigit Pramono

ABSTRACT

The increasing number of internet users in Indonesia and the complexity of network traffic have contributed to the risk of cyber attacks, particularly Distributed Denial of Service (DDoS) attacks, such as the one that occurred on the KPU RI server during the 2024 elections. To address this threat, this study analyzes the impact of Learning Vector Quantization (LVQ) and Autoencoder feature selection techniques on detection efficiency and accuracy by optimizing a Long Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN) model. The dataset used was CICDDoS2019 with stratified sampling of 30% and data balancing through SMOTE. Hybrid feature selection combining LVQ (supervised) and Autoencoder (unsupervised) successfully reduced the dimensions from 85 to 16 main features. The baseline model without feature selection recorded an accuracy of 99.9997% and an F1-score of 99.9399%, but training curve analysis showed indications of overfitting and bias in the majority class (ATTACK). The RNN+LVQ model performed the most balanced, with an accuracy of 99.9726% and high precision (90.3477%), making it reliable with minimal false alarms. In contrast, RNN+Autoencoder experienced a significant decline in performance, with a precision of only 38.8733%. The RNN+Hybrid model exhibits characteristics of an early warning system, with low precision (20.9398%) but very high recall (94.8996%), confirming the trade-off for maximum sensitivity. There are two features that are consistently considered important and selected by all three algorithms, namely Flow Duration and Fwd IAT Total. The presence of both indicates that the duration and totality of a communication session are very fundamental indicators of anomalies. Flow Duration is crucial because DDoS attacks often manipulate session duration, either by making it very short for sudden burst attacks, or very long for slow-rate attacks that aim to slowly tie up server resources. Fwd IAT Total, which is the total amount of time between packets from the attacker's direction, provides an overview of data transmission behavior. This finding concludes that the feature selection method successfully created a more robust and specialized model. The selection of the optimal method proved to be crucial and dependent on the implementation objectives: RNN+LVQ is recommended for systems that demand high reliability with minimal false alarms, while RNN+Hybrid is more suitable for early warning scenarios that demand maximum sensitivity.

Keywords: Cybersecurity, Anomaly Detection, Distributed Denial of Service (DDoS), feature selection, Long Short-Term Memory (LSTM).