ABSTRACT

The rapid advancement of information technology has led to an increasing risk of cyberattacks, particularly malware capable of remotely accessing systems and causing significant damage. This study aims to develop and evaluate a malware detection system using the Decision Tree algorithm through attack simulation and performance analysis. The research process involves simulating attacks using Metasploit with two file scenarios (infected and normal), capturing network traffic via Wireshark, and conducting data preprocessing (labeling, cleaning, normalization, and data splitting). Network traffic data in PCAP format was converted to CSV and classified using Decision Tree with optimal parameter selection via Grid Search (108 combinations).

The best model configuration includes: Criterion 'gini', Max Depth 10, Max Features 15, Min Samples Leaf 500, and Min Samples Split 1000. Evaluation results show excellent performance with an accuracy, precision, recall, and F1-Score of 99.95%, and a ROC-AUC of 100%. Malware detection reached 99.96% accuracy, while normal traffic achieved 99.94%, with a very low false positive rate of just 0.06%. The average training time of 13.40 ± 2.14 seconds per fold indicates high computational efficiency. The Fragment Offset feature contributed the most in detection, with a 91.20% impact. No indication of overfitting was observed, supported by negligible gaps in accuracy and F1-Score.

Overall, this study successfully developed a fast, accurate, and resource-efficient malware detection system using the Decision Tree algorithm, which can be effectively implemented for real-time network protection in both small-scale and enterprise environments.

Keywords: *Decision Tree, malware detection, cybersecurity, machine learning, network traffic analysis*