ABSTRAK

Sebagian besar NIDS bergantung pada predefined signature patterns dari serangan yang sudah dikenal, sehingga zero-day exploits dapat lolos. Anomaly-based NIDS membuat profil aktivitas “normal” namun sering menghasilkan false-positive yang tinggi. Traditional classifiers (Naïve Bayes, SVM) memerlukan feature engineering yang ekstensif; LSTM kesulitan menangani long-range dependencies dan tuning; autoencoders cenderung overfit pada pola benign dan kurang memiliki adversarial robustness. Kami memperkenalkan unsupervised Transformer-Encoder yang hanya merekonstruksi benign flows melalui multi-head self-attention dan positional encoding, menggunakan fixed reconstruction-error threshold yang dikalibrasi pada held-out data untuk menandai zero-day intrusions. Pada UNSW-NB15, metode ini mencapai Accuracy 95,09%, Precision 97,31%, Recall 91,63%, dan F1 Score 94,38%. Pada TON-IoT, metode ini mencapai Accuracy 88,75%, Precision 99,25%, Recall 88,73%, dan F1 Score 93,69%, menunjukkan kemampuan deteksi yang kuat dengan false alarm yang lebih rendah.

Kata Kunci — Zero-day attack detection, Network intrusion detection system (NIDS), Anomaly-based detection, Transformer Encoder, Unsupervised learning, UNSW-NB15, TON-IoT