ABSTRAK

Serangan *SQL Injection* masih menjadi salah satu ancaman utama pada aplikasi berbasis web dan jaringan, sementara ketersediaan dataset jaringan yang realistis dan representatif untuk penelitian keamanan masih terbatas. Dataset yang tersedia umumnya tidak seimbang, kurang bervariasi, serta belum memisahkan karakteristik header jaringan dan *payload SQL Injection* secara eksplisit. Oleh karena itu, penelitian ini bertujuan untuk membangun dataset jaringan *SQL Injection* sintetis menggunakan pendekatan *Generative Adversarial Network* (GAN). Metode yang digunakan memisahkan proses generasi data berdasarkan karakteristiknya. *CTGAN* diterapkan untuk menghasilkan data sintetis header jaringan berbentuk data tabular, sedangkan *CWGAN-GP* digunakan untuk menghasilkan payload SQL Injection sintetis. Data sintetis header dan payload kemudian digabungkan kembali menjadi paket jaringan utuh dan disimpan dalam format PCAP. Evaluasi kualitas data header dilakukan menggunakan *Kolmogorov–Smirnov Test* (KS Test) dan *Jensen–Shannon Divergence* (JSD), sedangkan evaluasi payload dilakukan menggunakan *Mean Squared Error* (MSE) dan *Cosine Similarity.* Hasil penelitian menunjukkan bahwa *CTGAN* mampu mereplikasi distribusi data header jaringan dengan baik, ditunjukkan oleh nilai JSD yang rendah (< 0,02) dan hasil KS Test yang menunjukkan kesesuaian distribusi antara data asli dan sintetis. Pada evaluasi payload, model *CWGAN-GP* menghasilkan *payload SQL Injection* sintetis dengan nilai MSE terendah sebesar 0,0034, *Cosine Similarity* rata-rata sebesar 0,80 hingga 0,88, serta stabilitas hasil yang lebih baik dibandingkan konfigurasi eksperimen lainnya. Dataset hasil penggabungan menghasilkan struktur paket jaringan yang valid dan menyerupai data asli.

Kata kunci: *Data Sintetis, SQL Injection, CTGAN, CWGAN-GP, Keamanan Jaringan*

ABSTRACT

SQL injection attacks remain one of the main threats to web-based applications and networks, while the availability of realistic and representative network datasets for security research is still limited. Available datasets are generally unbalanced, lack variety, and do not explicitly separate network header characteristics and SQL injection payloads. Therefore, this study aims to build a synthetic SQL Injection network dataset using the Generative Adversarial Network (GAN) approach. The method used separates the data generation process based on its characteristics. CTGAN is applied to generate synthetic network header data in tabular form, while CWGAN-GP is used to generate synthetic SQL Injection payloads. The synthetic header and payload data are then recombined into complete network packets and stored in PCAP format. The quality of the header data is evaluated using the Kolmogorov–Smirnov Test (KS Test) and Jensen–Shannon Divergence (JSD), while the payload is evaluated using Mean Squared Error (MSE) and Cosine Similarity. The results show that CTGAN is capable of replicating the distribution of network header data well, as indicated by a low JSD value (< 0.02) and KS Test results that show the distribution between the original and synthetic data is consistent. In the payload evaluation, the CWGAN-GP model produced synthetic SQL Injection payloads with the lowest MSE value of 0.0034, an average Cosine Similarity of 0.99, and better result stability compared to other experimental configurations. The combined dataset produced a valid network packet structure that resembled the original data.

Keywords: *Synthetic Data, SQL Injection, CTGAN, CWGAN-GP, Network Security*