Analisis Kinerja Intrusion Detection System Berbasis Algoritma Random Forest Menggunakan Dataset Honeynet BSSN dan CIC-ToN-IoT = Performance Analysis of Intrusion Detection System Based on Random Forest Algorithm Using Honeynet BSSN and CIC-ToN-IoT Datasets

Kuni Inayah, author

Abstrak

Dengan semakin berkembanganya teknologi dan sistem informasi pada area siber, ancaman siber juga semakin meningkat. Berdasarkan Laporan Honeynet BSSN Tahun 2023, Indonesia menduduki peringkat pertama sebagai negara dengan sumber serangan tertinggi. Untuk mengatasi permasalahan tersebut, IDS dijadikan solusi di berbagai sistem pemerintahan, bekerja sama dengan Honeynet BSSN. Namun, pada sistem IDS ini tidak bekerja secara maksimal untuk melakukan deteksi terhadap anomali atau jenis serangan baru yang belum belum pernah terjadi sebelumnya (zero-day). Solusi untuk meningkatkan performa IDS salah satunya dengan menggunakan machine learning. Beberapa studi sebelumnya membahas tentang perbandingan berbagai algoritma klasifikasi dan didapatkan bahwa algoritma random forest memiliki tingkat akurasi yang tinggi, tingkat false positive yang rendah, dan dalam hal komputasi tidak memerlukan sumber daya yang besar. Oleh karena itu, pada penelitian ini menggunakan algoritma random forest sebagai algoritma klasifikasinya. Dataset yang dipakai menggunakan dataset CIC-ToN-IoT sebagai dataset whitelist dan dataset dari Honeynet BSSN sebagai dataset blacklist. Model diklasifikasikan menjadi 10 (sepuluh) klasifikasi yaitu benign, Information Leak, Malware, Trojan Activity, Information Gathering, APT, Exploit, Web Application Attack, Denial of Service (DoS), dan jenis serangan lainnya (other). Hasil evaluasi menunjukkan bahwa implementasi algoritma random forest dengan dataset CIC-ToN-IoT dan dataset honeynet BSSN memiliki nilai akurasi yang tinggi dalam menganalisis berbagai serangan yang terjadi pada sistem informasi di lingkungan Pemerintah yaitu 99% dan dengan jumlah support data yang besar, model memiliki nilai presisi yang tinggi yaitu 91%.

.....With the increasing development of technology and information systems in the cyber area, cyber threats are also increasing. Based on the 2023 BSSN Honeynet Report, Indonesia is ranked first as the country with the highest source of attacks. To overcome these problems, IDS is used as a solution in various government systems, in collaboration with Honeynet BSSN. However, this IDS system does not work optimally to detect anomalies or new types of attacks that have never happened before (zero-day). One solution to improving IDS performance is by using machine learning. Several previous studies discussed the comparison of various classification algorithms and found that the random forest algorithm had a high level of accuracy, a low false positive rate, and in terms of computing did not require large resources. Therefore, this research uses the random forest algorithm as the classification algorithm. The dataset used uses the CIC-ToN-IoT dataset as a whitelist dataset and a dataset from Honeynet BSSN as a blacklist dataset. The model is classified into 10 (ten) classifications, namely benign, Information Leak, Malware, Trojan Activity, Information Gathering, APT, Exploit, Web Application Attack, Denial of Service (DoS), and other types of attacks. The evaluation results show that the implementation of the random forest algorithm with the CIC-ToN-IoT dataset and the BSSN honeynet dataset has a high accuracy value in analyzing various attacks that

occur on information systems in the government environment, namely 99% and with a large amount of data support, the model has high precision value, namely 91%.