ABSTRACT

System and network security with firewall device installation is not enough. Increased attacks cause data to be analyzed to be very large, existing Internet network security systems have limitations in the ability to adapt large amounts of data and types of new attacks. Intrusion Detection System (IDS) and firewall used into a standard system and network security. The limitations of detecting new types of attacks are increasingly causing the existing intrusion detection system accuracy and performance haved to improved. Several studies have used data mining techniques aim to overcome the problem of IDS attack. The purpose of this study is to test, analyze and classify attacks on the data tested by using the method of data mining classification. Data mining algorithms proposed and compared are Naïve Bayes Classifier, Random Forest and SVM. This study used data from a collection of KDD CUP'99 has 41 attributes. Implementation of feature selection methods to do on the attributes, which is expected to improve the accuracy and performance of the proposed algorithm models. All of the test results, Random Forest algorithms achieve high accuracy that is 98.0545%, but a slight decline accuracy level after feature selection. With application of GA and Cfs*Subset*Eval feature selection on Random Forest there was a reduction of build-model time to 141.24 seconds compared before feature selection is 360.19 seconds. Meaningless reduction in computational size and reduced computational complexity can improve the performance of IDS detection.

Keywords :IDS, data mining, classification , feature selection, Naïve Bayes Classifier, Random Forest, SVM

INTISARI

Keamanan sistem dan jaringan dengan pemasangan perangkat *firewall* tidaklah cukup. Peningkatan serangan menyebabkan data yang harus dianalisis menjadi sangat besar, sistem keamanan jaringan internet yang telah ada memiliki keterbatasan dalam kemampuan beradaptasi sejumlah besar data dan jenis serangan baru. Penggunaan *Intrusion Detection System* (IDS) dan *firewall* menjadi standar keamanan sistem dan jaringan. Keterbatasan mendeteksi jenis-jenis serangan baru yang makin meningkat menyebabkan *Intrusion Detection System* yang ada perlu untuk ditingkatkan tingkat akurasi dan performa kinerjanya. Beberapa penelitian telah menggunakan teknik *data mining* untuk mengatasi masalah serangan IDS. Tujuan dari penelitian ini adalah untuk menguji, menganalisis dan mengklasifikasikan serangan pada data-data yang diujikan dengan menggunakan metoda klasifikasi *data mining*. Algoritme *data mining* yang diusulkan dan dibandingkan adalah *Naïve Bayes Classifier, Random Forest* dan SVM. Dalam penelitian ini digunakan koleksi data dari KDD CUP'99 yang mempunyai 41 atribut. Dimana pada atribut dilakukan penerapan metoda seleksi fitur yang diharapkan dapat meningkatkan tingkat akurasi dan performa model algoritme yang diusulkan. Dari seluruh hasil pengujian, algoritme *Random Forest* mendapatkan tingkat akurasi tertinggi yaitu 98,0545 %, tapi terjadi sedikit penurunan tingkat akurasi setelah seleksi fitur. Dengan penerapan seleksi fitur GA dan *CfsSubsetEval* pada *Random Forest* terjadi pengurangan waktu *build model* menjadi 141,24 detik dibandingkan sebelum seleksi fitur yaitu 360,19 detik. Bermakna terjadi pengurangan ukuran komputasi dan berkurangnya kerumitan komputasi hingga dapat meningkatkan performa dari pendeteksian IDS.

Kata kunci – IDS, *data mining*, klasifikasi, seleksi fitur, *Naïve bayes*, *Random Forest*, SVM.