Perancangan sistem pendeteksi intrusi dengan hadoop 2.0 terintegrasi apache spark menggunakan learning vector quantization (LVQ) dan principal component analysis (PCA) = Intrusion detection system design using hadoop 2.0 integrated with apache spark using learning vector quantization (LVQ) and principal component analysis (PCA)

Pinem, Josua Geovani, author

Abstrak

Keamanan data (data security) sudah menjadi bagian vital didalam suatu organisasi yang menggunakan konsep sistem informasi. Semakin hari ancaman-ancaman yang datang dari Internet menjadi semakin berkembang hingga dapat mengelabuhi firewall maupun perangkat antivirus. Selain itu jumlah serangan yang masuk menjadi lebih besar dan semakin sulit untuk diolah oleh firewall maupun antivirus. Untuk dapat meningkatkan keamanan dari suatu sistem biasanya dilakukan penambahan Intrusion Detection Sistem IDS , baik sistem dengan kemampuan anomaly-based maupun sistem pendeteksi dengan kemampuan signature-based. Untuk dapat mengolah serangan yang jumlahnya besar maka digunakan teknik Big Data. Penelitian yang dilakukan ini menggunakan teknik anomaly-based dengan menggunakan Learning Vector Quantization dalam pendeteksian serangan.
Learning Vector Quantization adalah salah satu jenis neural network yang bisa mempelajari sendiri masukan yang masuk kemudian memberi keluaran sesuai dengan masukan tersebut. Beberapa modifikasi dilakukan untuk meningkatkan akurasi pengujian, antara lain dengan melakukan variasi parameter-parameter uji yang ada pada LVQ. Dengan melakukan variasi pada parameter uji learning rate, epoch dan k-fold cross validation dihasilkan keluaran dengan hasil yang lebih efisien.
Keluaran diperoleh dengan menghitung nilai information retrieval dari tabel confusion matrix tiap- tiap kelas serangan. Untuk meningkatkan kinerja sistem maka digunakan teknik Principal Component Analysis untuk mereduksi ukuran data. Dengan menggunakan 18-Principal Component data berhasil direduksi sebesar 47.3 dengan nilai Recognition Rate terbaik sebesar 96.52 dan efesiensi waktu lebih besar 43.16 daripada tanpa menggunakan PCA.
<hr><i>Data security has become a very serious part of any organizational information system. More and more threats across the Internet has evolved and capable to deceive firewall as well as antivirus software. In addition, the number of attacks become larger and become more dificult to be processed by the firewall or antivirus software. To improve the security of the system is usually done by adding Intrusion Detection System IDS , which divided into anomaly based detection and signature based detection. In this research to process a huge amount of data, Big Data technique is used. Anomaly based detection is proposed using Learning Vector Quantization Algorithm to detect the attacks.
Learning Vector Quantization is a neural network technique that learn the input itself and then give the appropriate output according to the input. Modifications were made to improve test accuracy by varying the test parameters that present in LVQ. Varying the learning rate, epoch and k fold cross validation resulted in a more efficient output.
The output is obtained by calculating the value of information retrieval from the confusion matrix table from each attack classes. Principal Component Analysis technique is used along with Learning Vector

Quantization to improve system performance by reducing the data dimensionality. By using 18 Principal Component, dataset successfully reduced by 47.3 , with the best Recognition Rate of 96.52 and time efficiency improvement up to 43.16.</i>