ANALISIS DAN IMPLEMENTASI ALGORITMA FAST WINNER SEARCH PADA SELF-ORGANIZING MAPS DAN N-GRAM (STUDI KASUS NETWORK INTRUSION DECTION SYSTEM)

Diaz Sastiardi Munarso[1], Imelda Atastina[2], Niken Dwi Wahyu Cahyani[3]

[1]Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Saat ini Internet sudah sangat popular dan sudah menjadi sebagian gaya hidup masyarakat modern. Jadi dibutuhkan adanya proteksi untuk menghalau serangan dari aktivitas yang merugikan pengguna Internet, salah satunya dengan membangun Intrusion Detection System (IDS). IDS komersial yang telah ada kebanyakan bersifat pasif dengan menggunakan informasi yang sudah ada, sedangkan serangan yang baru dan belum ada di data histori, masih akan tetap menyerang pengguna Internet.Untuk mengatasi hal ini maka pada tugas akhir ini dibuat aplikasi IDS berbasis jaringan dengan memanfaatkan Algoritma Fast Winner Search pada Self-Organizing Maps dan N-Gram. Aplikasi ini memberikan informasi mengenai paket intrusi dan akurasi proses deteksi. Dari proses pengujian didapatkan akurasi rata-rata 80% .Semakin banyak paket intrusi dan paket bukan intrusi yang dideteksi dengan benar, maka akurasi makin besar.Namun pada penelitian ini, paket intrusi yang terdeteksi sangat kecil, sedangkan paket bukan intrusi terdeteksi banyak dan benar. Hal ini menunjukkan bahwa aplikasi IDS dengan Algoritma Fast Winner Search pada Self-Organizing Maps dan N-Gram cocok untuk deteksi intrusi, namun masih terdapat kekurangan yang harus diperbaiki.

Kata Kunci : IDS, intrusi, Fast Winner Search, Self-Organizing Maps, N-Gram

Abstract

Nowadays, internet has been part of modern life style. Therefore, it is necessary to have protection to guard internet user from malicious activities, such as building Intrusion Detection System (IDS). Almost commercial IDS are now passive which is using attack information that has been known, while new attack that hasn’t been known will still attack internet user.To overcome this, this final project build IDS application with Fast winner search algorithm on Self-organizing maps and N-Gram. This application gives information about packet intrusion and accuration from detection process. From the testing process, the average accuration is 80% . The more packet intrusion and packet non-intrusion that are true detected, the better of detection accuration. Meanwhile, in fact, packet intrusion that is true detected are few, packet non-intrusion that is true detected are large. It shows that IDS application built by Fast winner search algorithm on Self-organizing maps and N-Gram is suitable for intrusion detection, but it also have some drawbacks that needs to be repaired.

Keywords : IDS, intrusion, Fast Winner Search, Self-Organizing Maps, N-Gram

1. Pendahuluan

1.1 Latar belakang

IDS berbasis deteksi anomali memiliki keuntungan dalam mendeteksi serangan yang sebelumnya belum pernah diketahui (*zero day attack*). IDS akan mengenali trafik jaringan yang "normal" (baca: terbebas dari serangan), untuk kemudian dibentuk model yang dibangun berdasarkan database trafik yang "normal" tersebut. Kemudian IDS akan membangkitkan alarm bila ternyata ada trafik "abnormal", yaitu dasta uji yang memiliki deviasi terhadap model yang secara signifikan melebihi threshold. Namun pada IDS berbasis anomali ini terdapat masalah, yaitu memiliki false negative yang relatif tinggi. Oleh karena itu, untuk memperoleh hasil deteksi yang akurat, dibutuhkan algoritma yang tepat dalam mengklasifikasi kenormalan data trafik.

Self-Organizing Maps (SOM) dapat digunakan untuk mempelajari pola trafik normal dari data historis yang ada. SOM memiliki kemampuan klasifikasi (similaritas intra-cluster dan dissimilaritas inter-cluster) yang tinggi [1][10], mampu memvisualisasi data berdimensi tinggi ke dalam data berdimensi kecil (umumnya dua dimensi) dan juga memiliki performansi real-time [7]. Algoritma SOM dapat juga digunakan untuk proses deteksi anomali, yaitu dengan cara membandingkan quantization error antara input vector terhadap hasil clustering SOM. Meskipun demikian, algoritma SOM mempunyai beberapa kelemahan, antara lain komputasi yang relatif tinggi, membutuhkan *learning time* yang cukup lama dan pencarian Best Matching Unit (BMU) yang relatif lama. Hal ini mempertimbangkan banyaknya kasus penyerangan terhadap banyak *victim* secara cepat,salah satu contohnya adalah kasus worm code red. Metode deteksi dengan SOM juga memiliki kekurangan, karena hasil quantization error-nya dapat dipengaruhi oleh urutan byte payload.

Untuk mengatasi masalah dan memangkas komputasi algoritma SOM di atas dibutuhkan adanya improvisasi atau pendekatan lain yang mampu meningkatkan performansi tanpa mempengaruhi akurasi algoritma SOM. Algoritma Fast Winner Search merupakan salah satu algoritma yang mampu meningkatkan performansi algoritma SOM melalui pencarian BMU dengan cara menghitung inner product terbesar dari normalisasi input vector terhadap bobot unit pada jaringan yang dibangun algoritma SOM. Sementara itu, Algoritma N-Gram dapat mengatasi masalah deteksi dengan cara menghitung frekuensi kemunculan byte dalam membangun model normal, dan membandingkannya dengan input vector.

1.2Perumusan masalah

Masalah yang dikaji dan diselesaikan dalam tugas akhir ini adalah Analisis dan Implementasi Algoritma Fast Winner Search dan N-Gram pada Network Intrusion Detection System.

Masalah-masalah khusus yang berkaitan dengan masalah ini adalah :

1. Bagaimana mengimplementasikan algoritma Fast Winner Search dan N-Gram pada Intrusion Detection System?
2. Bagaimana mengukur akurasi Algoritma Fast Winner Search dan N-Gram yang dibangun pada Intrusion Detection System?

1.3 Batasan masalah

Untuk menghindari pembahasan yang panjang tetapi dangkal, maka pada tugas akhir ini pembahasan hanya dibatasi pada :

1. Menggunakan data trafik offline yang diperoleh dari dataset DARPA IDS 1999 [3].
2. Asumsi dataset yang bersifat normal (terbebas dari serangan), jauh lebih banyak daripada dataset yang abnormal (berisi serangan).
3. Hanya menganalisis paket inbound TCP dengan alamat tujuan 172.016.0.0/16 dan port tujuan 1-1024.
4. Penggunaan pendekatan Fast Winner Search dilakukan pada pencarian Best Matching Unit (BMU).

1.4 Tujuan

Berdasarkan penjelasan diatas, penulis terdorong untuk melakukan hal-hal yang dirumuskan dalam tujuan pembahasan “Analisis dan Implementasi Algoritma Fast Winner Search dan N-Gram pada Network Intrusion Detection System”.

Secara lengkap, pembahasan dalam tugas akhir ini dilakukan untuk :

1. Menerapkan algoritma Fast Winner Search dan N-Gram pada Intrusion Detection System.
2. Mengukur akurasi algoritma Fast Winner Search dan N-Gram dalam mendeteksi intrusi

Hipotesis :

Algoritma Fast Winner Search pada Self-Organizing Maps dan N-Gram mampu mendeteksi intrusi dengan akurasi yang baik dan waktu yang singkat

1.5 Metodologi penyelesaian masalah

1. Studi literatur
   Merupakan tahapan dalam mempelajari konsep dan teori pendukung untuk memecahkan permasalahan.Dalam tugas akhir ini, studi literatur meliputi pembelajaran Neural network, datamining dan konsep lainnya yang menunjang pembuatan tugas akhir ini.
2. Pengumpulan data
   Pada tahap ini, pengumpulan data hanya melakukan pengunduhan dataset DARPA IDS 1999 pada alamat [3].
3. Analisis dan perancangan sistem
   Pada tahap ini, dianalisis kebutuhan sistem dan dibentuk sebuah *blueprint* yang akan dijadikan sebagai acuan dalam tahap berikutnya, yaitu bagaimana melakukan implementasi sistem dan pengujian.
4. Implementasi dan pembangunan sistem
   Tahap ini merupakan realisasi dari *blueprint* yang telah dbentuk dengan cara *coding* menggunakan bahasa pemrograman MATLAB dalam pengerjaannya. Pertama-tama data trafik jaringan diklasifikasi dengan menggunakan algoritma gabungan Self Organizing Maps dengan Fast Winner Search. Setelah ditemukan *winning neuron*-nya, sistem akan membentuk pola trafik dengan menghitung mean dan standar deviasi berdasarkan winning neuron, port tujuan dan alamat host tujuan kemudian menghitung similaritas antara pola trafik yang telah dilatih dengan data trafik. Jika jarak antar keduanya melebihi treshold secara signifikan, maka payload data trafik tersebut merupakan sebuah *attack.*
5. Pengujian dan analisa hasil
   Tahap ini melakukan pengukuran terhadap hasil akurasi deteksi intrusi yang telah dibangun.
6. Pembuatan Laporan
   Pada tahap ini diambil kesimpulan terhadap setiap informasi yang penting untuk dapat disajikan pada laporan ilmiah.

1.6 Sistematika penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut:

1. Pendahuluan
   Bab ini berisi uraian tugas akhir secara umum, meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan, metodologi penyelesaian masalah, dan sistematika penulisan laporan.

2. Landasan Teori
   Bab ini menguraikan teori intrusion detection system (IDS), algoritma Self-organizing Maps, metode Fast Winner Search, metode n-gram dan perhitungan jarak mahalanobis.
3. Analisis Perancangan dan Implementasi
   Bab ini membahas mengenai perancangan sistem untuk melatih model yang menggabungkan algoritma Self-organizing Maps dan Fast Winner Search. Hasil dari perancangan kemudian digunakan dalam proses implementasi sistem.
4. Pengujian dan Analisis Hasil Percobaan
   Bab ini berisi pengujian hasil implementasi yang telah dilakukan pada bab sebelumnya. Pengujian dilakukan dengan mengukur Akurasi, True Positive Rate,True Negative Rate, False Positive Rate dan False Negative Rate. Setelah pengujian selesai, dilakukan analisis terhadap hasil pengujian yang ada.
5. Kesimpulan dan Saran
   Bab ini berisi kesimpulan dari penulisan tugas akhir dan saran – saran untuk pengembangan lebih lanjut.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari hasil pengujian yang telah dilakukan pada Tugas Akhir ini dapat diambil beberapa kesimpulan sebagai berikut:

1. Aplikasi IDS dengan Algoritma Fast Winner Search pada Self-Organizing Maps dan N-Gram cukup baik untuk kasus Network IDS.Namun, meskipun akurasi yang diperoleh cukup tinggi, hasil TPR yang diperoleh sangat kecil.
2. Adanya kondisi dimana nilai threshold tertentu memiliki titik optimal bagi akurasi sistem yaitu pada nilai threshold sama dengan 15, akurasi yang diperoleh tidak jauh dari 80%.
3. Semakin besar nilai threshold, TNR dan FNR semakin meningkat. Sedangkan TPR dan FPR menurun.
4. Semakin banyak data latih, semakin variatif data yang dilatihkan pada model. Namun harus diikuti jumlah iterasi yang memadai untuk tiap paket pada data latih.
5. Pemisahan data latih berdasarkan port memberikan pengaruh terhadap nilai akurasi, TPR, FPR, TNR dan FNR.

5.2 Saran

1. Untuk penelitian selanjutnya dapat dikembangkan dengan bahasa pemrograman lain selain Matlab agar tidak terjadi kasus out-of-memory pada saat pelatihan model, sehingga data latih yang digunakan bisa lebih banyak.

Daftar Pustaka

[1] D. Bolzoni, S. Etalle dan P. Hartel.Poseidon : a 2-tier Anomaly-based Network Intrusion Detection System. *Proc. Of the 4th IEEE International Workshop on Information Assurance*.2006

[2] K. Wang dan S. J. Stolfo. Anomalous Payload-Based Network Intrusion Detection. In E. Jonsson, A. Valdes, and M. Almgren, editors, RAID '04: Proc. 7th Symposium on Recent Advances in Intrusion Detection, volume 3224 of LNCS, pages 203–222. Springer-Verlag, 2004.

[3] Dataset DARPA IDS 1999, http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/1999data.html

[4] M. Tavallaee,E Bagheri dan A.A.Ghorbani.A Detailed Analysis of the KDD CUP 99 Data Set. *Proceedings of IEEE Symposium on Computational Intellegence*.2009

[5] M. Damashek. Gauging similarity with n-grams: language independent categorization of text. Science,267(5199):843—848, 1995.

[6] SANS Institute – Internet Storm Center web site. URL http://isc.sans.org/index.php?on=toptrends.

[7] S. Albayrak et.al. Combining Self-Organizing Map Algorithms for Robust and Scalable Intrusion Detection. *Proc. Of the International Conference on Computational Intellegence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Vol-2*. 2006.

[8] S. Kaski. Fast Winner Search for SOM-Based Monitoring and Retrieval of High-Dimensional Data. *IEEE Conference Publication, volume 2, halaman 239-44*. 2000.

[9] S. Zanero dan S. M. Savaresi. Unsupervised learning techniques for an intrusion detection system. In SAC '04:Proc. 19th Annual ACM Symposium on Applied Computing,pages 412–419. ACM Press, 2004.

[10] T. Kohonen. Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. Springer, 1995. (Second Extended Edition 1997).

[11] P.N.Tan,M.Steinbach,V.Kumar.Introduction to datamining. Addison-Wesley, 2005.

[12] S.Mitra,T.Acharya.Datamining:Multimedia,Soft Computing, and Bioinformatics. Wiley.2003

[13] H.Chen. Data Mining Approaches for Intrusion Detection. 2004

[14] R.Kohavi, F.Provost. Glossary of Terms, Machine Learning.Vol.30,No. 2/3,pp.271-274.1998