MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset

Johnny
Faculty of Information Technology
Institut Teknologi Batam
Batam, Indonesia
Email: 1822004@student.iteba.ac.id

***Abstract*—Keamanan jaringan komputer dan mesin virtual sangat penting di era sekarang ini. Berbagai arsitektur telah diusulkan untuk keamanan jaringan atau mencegah akses berbahaya dari pengguna internal atau eksternal. Berbagai sistem yang ada telah dikembangkan untuk mendeteksi aktivitas jahat pada mesin korban; terkadang setiap pengguna eksternal membuat beberapa perilaku jahat dan mendapatkan akses tidak sah dari mesin korban ke sistem perilaku yang dianggap sebagai aktivitas jahat atau Penyusup.Berbagai pembelajaran mesin dan teknik komputasi lunak dirancang untuk mendeteksi aktivitas dalam data audit log jaringan waktu nyata. KKDDCUP99 dan NLSKDD paling banyak menggunakan kumpulan data untuk mendeteksi Penyusup pada kumpulan data benchmark. Dalam makalah ini, kami mengusulkan identifikasi penyusup menggunakan algoritma pembelajaran mesin.Dua teknik berbeda telah diusulkan seperti tanda tangan dengan deteksi dan deteksi berbasis anomali. Dalam analisis eksperimental, menunjukkan SVM, Naive Bayes dan algoritma ANN dengan berbagai set data dan menunjukkan kinerja sistem pada lingkungan jaringan waktu nyata.**

***Keywords*—Intrusion Detection System, Network security, Naïve Bayes, SVM, Artificial Neural Network, KDDCUP99.**

I. PENDAHULUAN

IDS khusus untuk mendeteksi satu jenis serangan, misalnya Contoh atau serangan yang tidak diketahui, serangan DoS atau serangan U2R atau serangan R2L.Kemudian secara berurutan menyebarkan sejumlah subsistem tersebut, satu per satu. Ini memiliki dua tujuan: pertama, hanya sejumlah fitur yang mendeteksi jenis serangan tertentu yang dapat dilatih di setiap sub-fase. Kedua, perangkat sub-ukuran tetap kecil dan karena itu berfungsi. Mirip dengan skema kami, kelemahan umum adalah meningkatkan komunikasi overhead antar modul.

Metode mendefinisikan sistem untuk membuat aturan SVM dalam pekerjaan penelitian ini, berdasarkan prosedur pemilihan perannya, yang bekerja pada HIDS dan NIDS. Algoritma genetika adalah algoritma optimasi yang digunakan untuk mencari solusi terbaik.Untuk semua jenis sub-serangan kelas master, pendekatan ensemble dengan algoritma klasifikasi yang berbeda dapat memberikan deteksi NIDS terbaik.

II. PEMBAHASAN

Bhosale, Karuna S. et al. [1] deep neural network (DNN), Membuat IDS yang terukur dan efisien untuk mengenali dan mengenali serangan cyber yang tidak disengaja dan tidak terduga dipelajari sebagai bentuk sistem pembelajaran mendalam. Perubahan konstan dalam operasi jaringan dan penciptaan serangan yang cepat melibatkan peninjauan kejadian yang sering terjadi selama bertahun-tahun melalui teknik dinamis dan statis.

Chamou et al. [2], Karena penjelasan itulah komunitas sains telah terbiasa dengan kompleksitas dan peningkatan efisiensi sistem deteksi intrusi sehingga sejumlah besar perusahaan di seluruh dunia menjadi sasaran dan terancam oleh munculnya ancaman baru dan yang muncul terus-menerus. Ini adalah alat inovatif untuk mengevaluasi aktivitas mencurigakan di DDoS dan ancaman cyber malware menggunakan model pembelajaran mendalam.

According to [3], sebuah sistem telah dikembangkan untuk secara akurat mendeteksi potensi serangan dengan menggunakan berbagai bebas keputusan, hutan acak, dan KNN. Teknik baru disarankan untuk memperbaiki kekurangan dari pendekatan sebelumnya yang tidak dapat mendeteksi serangan IPv6. Dalam mendeteksi serangan berbasis IPV4, kerangka kerja yang ditetapkan menghasilkan hasil yang mengesankan dan efisien, dengan mempertimbangkan kemungkinan jangkauan—efisiensi dari berbagai algoritme yang telah diukur.

According to [4], mengidentifikasi fenomena baru yang disebut NEC, clustering, dan KDD dapat digunakan secara efektif. Anomali tanpa pengawasan digunakan untuk menghasilkan tingkat deteksi yang tinggi dan tingkat pasif palsu yang lebih sedikit. Ini adalah cara yang efektif untuk memecahkan masalah dan menemukan anomali yang tidak melibatkan pengumpulan data berlabel.

Machine Learning Survey [5]. Untuk mengakses header paket dan header paket aliran bersih digunakan ke jaringan dan data tingkat kernel untuk proses deteksi intrusi, masalah potensial adalah bahwa penambangan data dan kecerdasan buatan tidak dapat dicapai tanpa database aktual, oleh karena itu, sangat memakan waktu. Kehadiran berbagai statistik dan algoritma pembelajaran mesin dibahas. Makalah penelitian berisi serangkaian kriteria komparatif untuk anomali metodologi untuk penambangan data pembelajaran mesin — deteksi intrusi dependen membantu menemukan, mengevaluasi, dan mengenali penggunaan, replikasi, perubahan, dan penghancuran sistem informasi yang tidak sah.

Based on the [6], deteksi dan pemantauan penyusup menggunakan pembelajaran mesin, peringkat, dan Voronoi, kea-

manan telah ditingkatkan.

According to ADS-B IDS [7], teknik IDS pengawasan-siaran berbasis otomatis diusulkan menggunakan teknik ADS-B. Untuk meningkatkan kinerja kontrol lalu lintas udara, kumpulan data HMAC digunakan. Pendekatan beroperasi dengan overhead minimum. Cakupan mendatang menetapkan bahwa jaraknya dari posisi satu kali yang sesuai berada dalam zona aman untuk lokasi ADS-B agar valid. GPS menggunakan lingkungan fisik-cyber yang divalidasi oleh deteksi serangan untuk menentukan spesifikasi radar di pesawat yang memberi sinyal presisi lokasi yang unggul, karena ADS-B telah muncul sebagai alternatif radio saat ini. Sebuah protokol disarankan untuk menukar kunci yang digunakan untuk algoritma HMAC dengan aman.

According to [8], untuk perangkat listrik yang mendukung data log, laporan tersebut menyatakan bahwa kombinasi IDS menggunakan data mining dibuat menggunakan proses penambangan yang aman. Metode tersebut merupakan pendekatan terpadu untuk merancang prototipe IDS. Ketepatan deteksi, yang mencapai 73 persen, telah menjadi salah satu keunggulan utama. Tetapi pendekatan ini tidak cukup untuk mengumpulkan isu-isu utama seperti log data, meskipun rumit. Kerangka kerja ini memanfaatkan fungsionalitas IDS berbasis karakteristik dan spesifikasi. Metodologi analisis data yang menggabungkan log audit untuk mempelajari rute umum dari perangkat mesin yang berbeda.

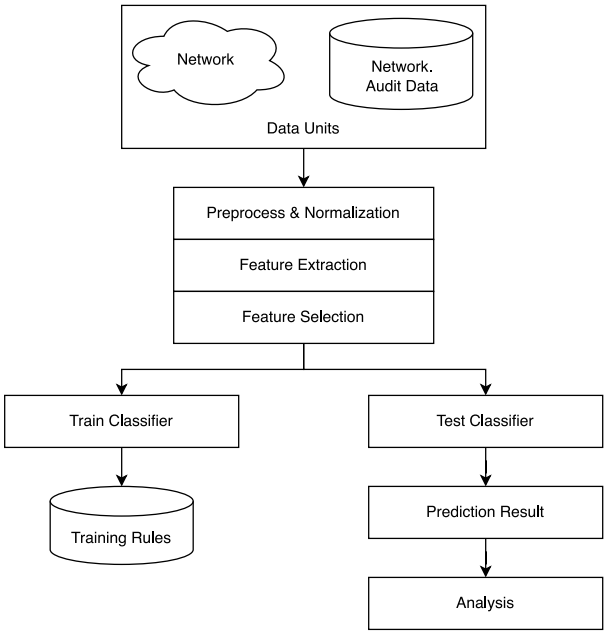
According to Al, Dr. Yogesh Kumar Sharma et. [10] mendefinisikan sistem Akses Jaringan 6G dan Mekanisme Aturan Congestion Edge-Assisted menggunakan Jaringan yang Ditentukan Perangkat Lunak Untuk menghindari buffering arus lalu lintas, kerangka kerja harus menggunakan data frekuensi besar.

III. METODOLOGI

Metodologi penelitian yang diusulkan melakukan deteksi dan pencegahan intrusi menggunakan teknik pembelajaran mesin. Pelatihan, termasuk pilihan paket untuk anomali dan pemantauan jarak jauh, akan dilakukan oleh lingkungan paket yang dicirikan blok Kemudian akan mengirimkan kumpulan fungsi untuk aktivitas paket tertentu. Jika semuanya baik-baik saja, kirimkan bersama-sama. Untuk mendeteksi serangan individu, sampel perilaku buruk akan diuji untuk pemilihan fitur untuk berbagai kualitas. Ada dua fase dalam sistem yang diusulkan; kami telah mengambil dataset jaringan untuk tujuan pelatihan dan pengujian sistem—komponen kerangka kerja.

Kumpulan data NSL KDD CUP 1999 digunakan untuk eksperimen (kumpulan data KDD, 1999). Dataset KDD CUP 1999 adalah versi perangkat lunak penilaian deteksi intrusi DARPA 1998 asli milik MIT Lincoln Laboratory, yang siap dan terkontrol.Semua jenis kontinu, diskrit,dan variabel simbolik memiliki opsi ini, dengan rentang yang sangat bervariasi jatuh ke dalam empat kategori:

(1) kelas utama terdiri dari opsi intrinsik asosiasi, termasuk opsi kritis untuk menghubungkan protokol kontrol transmisi individu. Berbagai pilihan termasuk panjang afiliasi, jenis protokol (TCP, UDP, dll.), dan unit area akses jaringan (HTTP, telnet, dll.).



Gambar 1: System architecture

(2) Pilihan konten dalam asosiasi yang diberitahukan oleh Unit Area Data Domain, seperti jumlah upaya login yang gagal, tidak akan menentukan muatan paket protokol kontrol transmisi pertama.

(3) Opsi host konstan menganalisis koneksi yang ditentukan yang memiliki host tujuan berkelanjutan dalam 2 detik terakhir karena hubungan saat ini dan mengukur statistik yang terkait dengan perilaku protokol, operasi, dll.

(4) Opsi layanan serupa memeriksa koneksi yang memiliki layanan yang sama dengan tautan saat ini dalam dua detik terakhir.

Pada titik pra-pemrosesan, blok kedua dari Gambar 1 menunjukkan bahwa kita menggunakan packet sniffer yang dirancang dengan perpustakaan winpcap untuk mengekstrak informasi paket jaringan dari setiap paket, termasuk header IP, header TCP, header UDP, dan header ICMP. Setelah itu, dengan mempertimbangkan hubungan antara setiap kombinasi alamat sains (ilmu sumber dan IP tujuan), paket data dipartisi dan dibentuk menjadi catatan dengan menggabungkan setiap informasi.

IV. HASIL DAN DISKUSI

Setelah implementasi sistem berhasil, kami menghitung matriks kebingungan untuk sistem. Tabel 1 dan Tabel 2 menunjukkan klasifikasi dengan algoritma SVM. Angka 2 menunjukkan kinerja klasifikasi pengumpulan data oleh KDDCUP menggunakan pendekatan berbasis kepadatan dari program algoritma pembelajaran mesin Gambar 3 digunakan untuk mengklasifikasikan dan memprediksi presisi sistem menggunakan metode yang berbeda seperti algoritma RNN.

TABLE I: CONFUSION MATRIX CALCULATION USING SVM FOR CLASSIFICATION

Class	Normal	Attack
Normal	1760	19
Attack	9	1640
	1769	1659

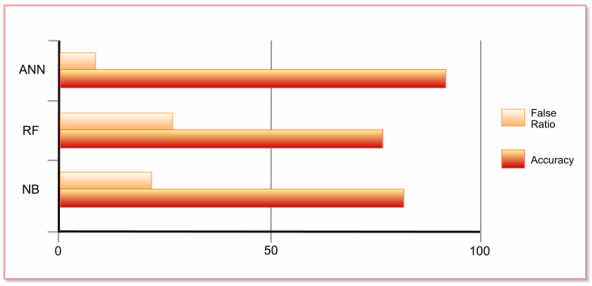
TABLE II: CONFUSION MATRIX CALCULATION USING NB FOR CLASSIFICATION

Class	Normal	Attack
Normal	1830	227
Attack	169	1202
	1999	1429

TABLE III: PERFORMANCE EVALUATION WITH NB AND SVM

	NB	SVM
Accuracy	0.9892	0.9525
Precision	0.9867	0.9797
Recall	0.9933	0.9463
F-Score	0.9899	0.9529

Menurut kedua analisis eksperimen, SVM menunjukkan akurasi klasifikasi yang lebih baik daripada algoritme NB, yang ditunjukkan pada gambar 3. Berdasarkan analisis eksperimen di atas, kita dapat menyimpulkan bahwa sistem menghasilkan akurasi yang lebih baik untuk komputasi kepercayaan di lingkungan in-service IoT. Seluruh penelitian mengikuti beberapa parameter lingkungan simulasi serta kombinasi dari algoritma pembelajaran mesin.



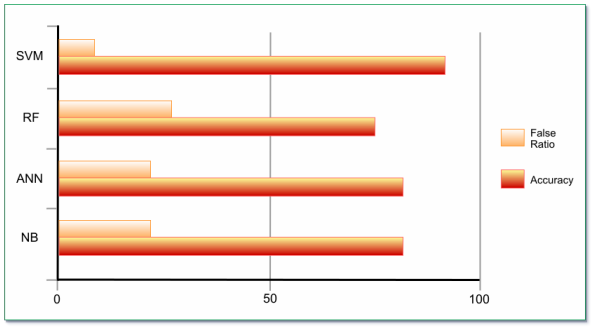
Gambar 2: Detection accuracy for KDD : CUP99 dataset using machine learning

Gambar 2 di atas Menunjukkan akurasi klasifikasi hasil kddCup 99, dengan lima kelas yang berbeda. Rata-rata keluaran perangkat lunak sekitar algoritma untuk pembelajaran mesin 88,50 percent untuk semua kelas.

Gambar 3 menunjukkan efisiensi rata-rata identifikasi di berbagai database, dari (n) kelas yang berbeda. Performa rata-rata sistem dengan algoritme pembelajaran mesin adalah sekitar 95 percent untuk semua (n) kelas

V. KESIMPULAN

Studi ini mengusulkan pendekatan SVM-IDS berdasarkan pembelajaran mendalam untuk menyarankan sistem ID yang



Gambar 3: Detection accuracy various network dataset using propsoed (SVM) vs exisitng

efisien. Untuk menguji akurasi deteksi anomali, kami menggunakan dataset intrusi berbasis sintetis - NSL-KDD. Kami berencana untuk mengimplementasikan IDS ke lingkungan cloud di masa mendatang menggunakan teknik pembelajaran mendalam. Kami juga menganalisis dan membandingkan berbagai metode deep learning, yaitu. Untuk mendeteksi intrusi dalam jaringan, NB ANN, RF, dan SVM pada dataset NSL-KDD; Program ini pada dasarnya bertindak sebagai algoritma kecerdasan buatan dan pengkondisian untuk menentukan contoh yang tidak diketahui selama pemeriksaan data. Struktur aturan yang efisien menghasilkan klasifikasi yang lebih baik dan deteksi kelas tinggi. Berbagai eksperimen menggunakan analisis eksperimental untuk menilai efisiensi algoritme menggunakan beberapa tes dan menyimpulkan bahwa kami mencapai hasil yang memuaskan.

REFERENCES

[1] K. S. Bhosale, M. Nenova, and G. Iliev, "Modified naive bayes intrusion detection system (mnbids)," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*. IEEE, 2018, pp. 291–296.

[2] D. Chamou, P. Toupas, E. Ketzaki, S. Papadopoulos, K. M. Giannoutakis, A. Drosou, and D. Tzovaras, "Intrusion detection system based on network traffic using deep neural networks," in *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, 2019, pp. 1–6.

[3] M. Anbar, R. Abdulah, I. H. Hasbullah, Y.-W. Chong, and O. E. Elejla, "Comparative performance analysis of classification algorithm for internal intrusion detection," in *2016 14th Annual Conference on Privacy Security and Trust (PCT)*, 2016.

[4] W. Chen, F. Kong, F. Mei, G. Yuan, and B. Li, "A novel unsupervised anomaly detection approach for intrusion detection system," in *2017 ieee 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids)*. IEEE, 2017, pp. 69–73.

[5] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications surveys & tutorials*, vol. 18, no. 2, pp. 1153–1176, 2015.

[6] M. S. Koli and M. K. Chavan, "An advanced method for detection of botnet traffic using intrusion detection system," in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2017, pp. 481–485.

[7] T. Kacem, D. Wijesekera, P. Costa, and A. Barreto, "An ads-b intrusion detection system," in *2016 IEEE Trustcom/BigDataSE/ISPA*. IEEE, 2016, pp. 544–551.

[8] S. Pan, T. Morris, and U. Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3104–3113, 2015.

[9] M. Ezzarii, H. Elghazi, H. El Ghazi, and T. Sadiki, "Epigenetic algorithm for performing intrusion detection system," in *2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS)*. IEEE, 2016, pp. 1–6.

[10] P. Borole, Y. K. Sharma, and S. Nemade, "6g network access and edge-assisted congestion rule mechanism using software-defined networking," *International Journal of Future Generation Communication and Networking*, vol. 13, no. 1s, pp. 107–112, 2020.