Jurnal Infosecure Vol.4 No.1, Maret 2023
p-ISSN: 2087-1023
MTI-UNLA

IMPLEMENTASI CREDAL C45 UNTUK MENURUNKAN FALSE POSITIF PADA HYBIRD INTRUSION DETECTION SYSTEM

Cecep Suwanda[1], Arief Zulianto[2], Toto Suharto[3]

Prodi Magister Teknik Informatika, Pascasarjana, Universitas Langlangbuana[1,2,3]

[1]cecepsuwanda@gmail.com

[2]madzul@gmail.com

[3]tsuharto@gmail.com

Abstrak— Pada penelitian ini kita akan menggunakan NSL-KDD dataset untuk mengevaluasi 2 hybird intrusion detection system. Hybird intrusion detection system yang pertama menggunakan C45 dan 1-class SVM, yang kedua menggunakan Credal C45 dan 1-class SVM. Hasil evaluasi menggunakan data latih menunjukan bahwa Credal C45 memiliki jumlah false positif 9% lebih kecil dari C45, hasil evaluasi menggunakan data test menunjukan bahwa Credal C45 memiliki jumlah false positif 3% lebih kecil dari C45.

Kata kunci— Sistem Deteksi Intrusi, C45,Credal-C45, 1-Class SVM, Noise.

I. PENDAHULUAN

Sistem deteksi intrusi (IDS) adalah sebuah sistem yang digunakan untuk mendeteksi adanya serangan pada sebuah komputer atau server. Secara garis besar, ada dua jenis sistem deteksi intrusi berdasarkan bagaimana cara untuk mendeteksi serangan tersebut, yaitu berbasis aturan (rule-based/signature-based) dan berbasis anomali [1].

IDS yang berbasis aturan akan menggunakan daftar aturan atau daftar pola isi paket yang dimilikinya untuk dicocokkan dengan rangkaian paket data yang melalui sebuah jaringan komputer. Metode ini cukup ampuh untuk serangan-serangan yang sudah dikenali sebelumnya dan metode ini sampai sekarang masih sering digunakan. Kelemahan dari metode berbasis aturan ini adalah jika serangan yang masuk belum pernah ada sebelumnya atau daftar aturan yang digunakan sudah terlalu lama tidak berubah, sehingga tidak bisa mendeteksi jenis-jenis serangan terbaru [1].

IDS yang berbasis anomali memanfaatkan kebiasaan-kebiasaan yang terjadi dalam sebuah sistem dan menganggap apabila ada penyimpangan dari maka itu adalah sebuah aktivitas serangan. Sehingga jenis-jenis serangan yang belum pernah ada sebelumnya dapat lebih mudah ditangkap karena dianggap memiliki pola-pola yang berbeda dari yang umum diterima atau dikirim. Namun metode ini cukup sulit diterapkan di dunia nyata karena akurasinya masih rendah, terutama false positive yang dihasilkan masih cukup tinggi [1]

Hibrid IDS menggabungkan metode berbasis aturan dan metode berbasis anomali, penggabungan ini dimaksudkan untuk mengatasi kelemahan masing-masing [2]. Hibrid IDS ini dilakukan dengan beberapa kombinasi, metode anomali diikuti metode aturan, metode aturan bersamaan dengan metode anomali, metode aturan diikuti metode anomali dan kombinasi lain yang lebih rumit [3].

Pada kombinasi metode anomali diikuti metode aturan, metode anomali pertamakali mendeteksi aktivitas yang mencurigakan, aktivitas yang mencurigakan itu di periksa memakai metode aturan untuk menentukan aktivitas tersebut Normal, alarms atau Unknown Attack [3].

Pada kombinasi metode aturan bersamaan dengan metode anomali, metode anomali dan metode aturan masing-masing mendeteksi aktivitas yang mencurigakan, kedua aktivitas yang mencurigakan tersebut tersebut di analisis oleh corellation system untuk menentukan attack atau bukan [3].

Pada kombinasi metode aturan diikuti metode anomali, metode aturan pertamakali mendeteksi known attack, yang bukan known attack di periksa memakai metode anomali untuk menentukan aktivitas tersebut Normal atau Unknown Attack [3].

Ketiga kombinasi melatih model untuk metode aturan terpisah dengan model untuk metode anomali. Untuk membentuk metode aturan dipakai keseluruhan dataset, untuk membentuk metode anomali digunakan seluruh normal dataset. Seluruh normal dataset memiliki banyak jenis koneksi normal, karena hal ini metode anomali tidak bisa membentuk model yang secara tepat mengenali seluruh jenis koneksi normal, hal ini yang menyebabkan False Positif dari model metode anomali besar [4].

Agar metode anomaly bisa mengenali seluruh jenis koneksi normal, Gisum Kim [4] mengusulkan untuk membagi normal dataset menjadi beberapa subdataset normal menggunakan model metode aturan (C45), beberapa subdataset ini digunakan membangun beberapa model aturan anomali (1-class SVM). Model metode anomali yang dihasilkan dengan cara ini menghasilkan False Positif yang lebih kecil. Sekalipun False Positif lebih kecil ukuran subdataset yang dihasilkan oleh C45 masih

ada yang besar dan di dalam subdataset masih ada koneksi lain selain normal dengan kata lain subdataset yang dihasilkan C45 masih mengandung False Positif, Keberadaan False Positif ini dapat menyebabkan model anomali yang terbentuk menghasilkan False Positif.

Masih adanya false positif pada subdata train disebabkan karena adanya noise pada NSL-KDD [5]. Credal C45 dapat digunakan untuk dataset yang ada noise didalamnya [6].

Pada penelitian ini kita akan mengganti C45 pada metode yang diusulkan Gisung Kim, Seungmin Lee dan Sehum Kim (2014) dengan Credal C45. Model yang dihasilkan akan dievaluasi menggunakan dataset NSL-KDD. Hasil evaluasi akan dibandingkan dengan hasil evaluasi C45. Dari perbandingan ini akan ditentukan apakah False Positif Credal C45 lebih kecil dari C45.

II. METODE

A. *Studi Pustaka*

Studi pustaka dilakukan dengan membaca jurnal, buku dan sumber-sumber lainnya yang berhubungan.

B. *Experiment*

Dilakukan percobaan langsung terhadap data untuk menguji apakah solusi menghasilkan hasil yang diharapkan

III. HASIL DAN PEMBAHASAN

A. *Preprocess*

NSL-KDD dataset terdiri dari 2 file, KDDTrain+.TXT dan KDDTest+.TXT. Gabungkan kedua file, pisahkan data dengan label normal dari file gabungan, ganti data dengan label attack pada file gabungan dengan label known dan unknown. Label known untuk attack yang ada di KDDTrain+.TXT dan KDDTest+.TXT, label unknown untuk attack yang ada di KDDTest+.TXT tapi tidak ada di KDDTrain+.TXT.

Pisahkan unknwon dari file gabungan, gabungkan kembali normal ke file gabungan pecah file gabungan menjadi data latih dan data test, satukan unknown ke data test.

B. *Membangun HIDS C45 dan 1-Class SVM*

Data latih digunakan untuk melatih C45, menghasilkan 26 aturan klasifikasi dengan 11 aturan untuk klasifikasi normal dan 15 aturan untuk klasifikasi known.

TABEL I
Confusion Metrik Data Latih C45

Label Asli	Label Prediksi	
	Normal	Known
Normal	38221	263
Known	232	32917

Hasil evaluasi C45 menggunakan data latih ditampilkan pada TABEL I. Hasil evaluasi C45 menggunakan data test ditampilkan pada TABEL I I

TABEL II
Confusion Metrik Data Test C45

Label Asli	Label Prediksi		
	Normal	Known	Unknown
Normal	38177	306	0
Known	260	32889	0
Unknown	2771	1871	0

Aturan klasifikasi normal C45 digunakan untuk memecah data latih dan data test menjadi 11 subdataset. 11 subdata latih digunakan untuk membuat beberapa model 1-Class SVM. TABEL III menunjukan nilai γ dan μ yang dipakai melatih 1-Class SVM.

TABEL III
Nilai γ dan μ 1-Class SVM C45

Subset	γ	μ	Jumlah	Lama Train	SV
1	0,0001	0,01	1520	2,062	50
2	0,0001	0,02	310	2,0133	24
3	0,0001	0,02	193	2,0056	14
4	0,0001	0,02	539	2,0389	80
5	0,0001	0,04	109	2,0049	19
6	0,0001	0,02	100	2,3241	44
7	0,0001	0,03	855	2,026	51
8	0,0001	0,02	33719	38,763	5044
9	0,0001	0,01	299	2,0138	90
10	0,0001	0,01	636	2,0345	215
11	0,0003	0,01	173	2,0102	171

TABEL IV dan TABEL V menampilkan hasil latih dan test dari 1-Class SVM C45

TABEL IV
Confusion Metrik Data Latih 1-Class C45

Label Asli	Label Prediksi		
	Normal	Known	Unknown
Normal	35340	0	2881
Known	180	0	52
Unknown	0	0	0

TABEL V
Confusion Metrik Data Test 1-Class C45

Label Asli	Label Prediksi		
	Normal	Known	Unknown
Normal	32320	0	5857
Known	161	0	99
Unknown	1305	0	1466

C. *Membangun HIDS Credal C45 dan 1-Class SVM*

Data latih digunakan untuk melatih Credal C45, menghasilkan 32 aturan klasifikasi dengan 13 aturan untuk klasifikasi normal dan 20 aturan untuk klasifikasi known.

TABEL VI
Confusion Metrik Data Latih Credal C45

Label Asli	Label Prediksi	
	Normal	Known
Normal	38230	254
Known	211	32938

Hasil evaluasi Credal C45 menggunakan data latih ditampilkan pada TABEL VI. Hasil evaluasi Credal C45 menggunakan data test ditampilkan pada TABEL VI I

TABEL VII
Confusion Metrik Data Test Credal C45

Label Asli	Label Prediksi		
	Normal	Known	Unknown
Normal	38179	304	0
Known	213	32936	0
Unknown	2740	1902	0

Aturan klasifikasi normal Credal C45 digunakan untuk memecah data latih dan data test menjadi 13 subdataset. 13 subdata latih digunakan untuk membuat beberapa model 1-Class SVM. TABEL VIII menunjukan nilai γ dan μ yang dipakai melatih 1-Class SVM.

TABEL VIII
Nilai γ dan μ 1-Class SVM Credal C45

Subset	γ	μ	Jumlah	Lama Train	SV
1	0,0001	0,01	1520	2,0583	50
2	0,0001	0,05	77	2,0036	11
3	0,0001	0,03	103	2,0048	10
4	0,0001	0,03	118	2,0047	10
5	0,0001	0,02	193	2,006	14
6	0,0001	0,02	539	2,017	80
7	0,0001	0,04	109	2,0066	19
8	0,0001	0,02	100	2,0066	44
9	0,0001	0,03	855	2,0252	51
10	0,0001	0,02	33719	38,8985	5044
11	0,0001	0,01	299	2,0123	90
12	0,0001	0,01	636	2,0315	215
13	0,0003	0,01	173	2,011	171

TABEL IX dan TABEL X menampilkan hasil latih dan test dari 1-Class SVM Credal C45

TABEL IX
Confusion Metrik Data Latih 1-Class Credal C45

Label Asli	Label Prediksi		
	Normal	Known	Unknown
Normal	35345	0	2885
Known	161	0	50
Unknown	0	0	0

TABEL X
Confusion Metrik Data Test 1-Class Credal C45

Label Asli	Label Prediksi		
	Normal	Known	Unknown
Normal	32305	0	5874
Known	118	0	95
Unknown	1243	0	1497

D. HIDS C45 dan 1-Class SVM VS HIDS Credal C45 dan 1-Class SVM

Berdasarkan perbandingan jumlah False Positif antara TABEL I dan TABEL VI, dapat ditarik kesimpulan bahwa False Positif Credal C45 9% lebih kecil dari C45. Berdasarkan perbandingan False Positif TABEL II dan TABEL VII, False Positif Credal C45 3% lebih kecil dari C45.

IV. SIMPULAN

Credal C45 menghasilkan rule yang lebih banyak dari C45, Credal C45 menghasilkan 33 rule, C45 menghasilkan 26 rule. Credal C45 memiliki jumlah False Positif lebih kecil dari C45, 9% pada data latih dan 3% pada data test. Dari hasil diatas dapat disimpulkan bahwa Credal-C45 dapat menurun jumlah False Positif.

REFERENSI

[1] B. A. Pratomo and R. M. Ijtihadie, "SISTEM DETEKSI INTRUSI MENGGUNAKAN N-GRAM DAN COSINE SIMILARITY," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 14, no. 1, p. 108, Jan. 2016, doi: 10.12962/j24068535.v14i1.a516.

[2] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks," *Expert Syst Appl*, vol. 29, no. 4, pp. 713–722, Nov. 2005, doi: 10.1016/j.eswa.2005.05.002.

[3] J. Zhang and M. Zulkernine, "A hybrid network intrusion detection technique using random forests," *Proceedings - First International Conference on Availability, Reliability and Security, ARES 2006*, vol. 2006, pp. 262–269, 2006, doi: 10.1109/ARES.2006.7.

[4] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Syst Appl*, vol. 41, no. 4 PART 2, pp. 1690–1700, 2014, doi: 10.1016/j.eswa.2013.08.066.

[5] K. M. Al-Gethami, M. T. Al-Akhras, and M. Alawairdhi, "Empirical Evaluation of Noise Influence on Supervised Machine Learning Algorithms Using Intrusion Detection Datasets," *Security and Communication Networks*, vol. 2021, 2021, doi: 10.1155/2021/8836057.

[6] C. J. Mantas and J. Abellán, "Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data," *Expert Syst Appl*, vol. 41, no. 10, pp. 4625–4637, 2014, doi: 10.1016/j.eswa.2014.01.017.