Versi Rekaman:https://www.sciencedirect.com/science/article/pii/S0167404823000275
Naskah_eebd6c9bddccb22a516ce8325e680d7a

Mendeteksi Serangan DDoS Menggunakan Jaringan Saraf Adversarial

Ali Mustafa[1]Ridha Khatoun[1]Sherali Zeadally,[2]Fadlallah Chbib[1]Ahmad Fadlallah[1]Walid Fahs,[3]Ali El Attar[1]
Institut Polytechnique de Paris, Telecom Paris (INFRES), LTCI, Prancis[1]
Fakultas Komunikasi dan Informasi, Universitas Kentucky, Lexington, Kentucky, AS[2]
Fakultas Teknik, IUL, Lebanon[3]
(ali.mustapha, rida.khatoun, fadlallah.chbib, ali.elattar)@telecom-paris.fr, szeadally@uky.edu ,
a.fadlallah@usal.edu.lb , walid.fahs@iul.edu.lb

Abstrak—Dalam serangan Distributed Denial of Service (DDoS), jaringan perangkat yang disusupi digunakan untuk membanjiri target dengan banjir permintaan, sehingga target tidak dapat melayani permintaan yang sah. Deteksi serangan ini merupakan masalah yang menantang dalam keamanan siber, yang telah diatasi menggunakan algoritma Machine Learning (ML) dan Deep Learning (DL). Meskipun ML/DL dapat meningkatkan akurasi deteksi, tetapi serangan tersebut masih dapat dihindari - ironisnya - melalui penggunaan teknik ML/DL dalam pembuatan lalu lintas serangan. Secara khusus, Generative Adversarial Networks (GAN) telah membuktikan efisiensinya dalam meniru data yang sah. Kami membahas aspek-aspek di atas dari teknik deteksi dan antideteksi DDoS berbasis ML/DL. Pertama, kami mengusulkan metode deteksi DDoS berdasarkan model Long Short-Term Memory (LSTM), yang merupakan jenis Recurrent Neural Networks (RNN) yang mampu mempelajari dependensi jangka panjang. Skema deteksi menghasilkan tingkat akurasi yang tinggi dalam mendeteksi serangan DDoS. Kedua, kami menguji teknik yang sama terhadap berbagai jenis serangan DDoS adversarial yang dihasilkan menggunakan GAN. Hasilnya menunjukkan ketidakefisienan skema deteksi berbasis LSTM. Terakhir, kami menunjukkan cara meningkatkan skema ini untuk mendeteksi serangan DDoS adversarial. Hasil eksperimen kami menunjukkan bahwa model deteksi kami efisien dan akurat dalam mengidentifikasi lalu lintas DDoS adversarial yang dihasilkan GAN dengan rasio deteksi berkisar antara 91,75% dan 100%.

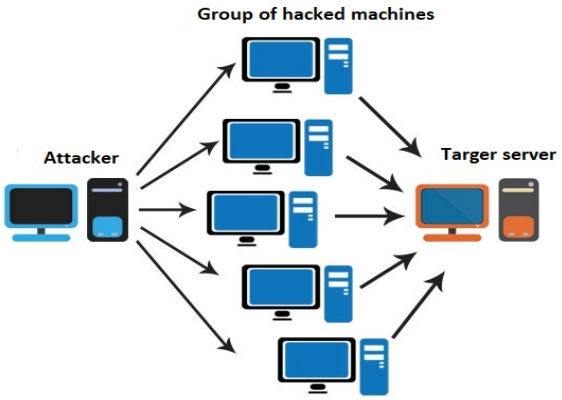
Kata kunci—Penolakan Layanan Terdistribusi (DDoS), Memori Jangka Panjang dan Pendek (LSTM), Jaringan Adversarial Generatif (GAN), Sistem Deteksi Intrusi (IDS), Pembelajaran Mesin (ML)

Aku. AkuPENGANTAR

Menurut Nexusguard [1], pada paruh pertama tahun 2022, jumlah serangan total meningkat sebesar 75,60% dibandingkan dengan paruh kedua tahun 2021. Dalam laporannya [2], Amazon menunjukkan bahwa pada Q1 tahun 2020, tingkat serangan DDoS mencapai 2,3 Tbps. Biaya yang terkait dengan serangan ini juga meningkat. Spamhaus baru-baru ini menerbitkan laporannya yang disebut "Spamhaus Botnet Threat Update" dan menunjukkan statistik tentang botnet dan sumbernya yang diilustrasikan bahwa sebagian besar botnet berasal dari Tiongkok, dengan lebih dari 590.000 bot, AS dengan 376.000 bot, dan India, dengan 350.000 bot [3]. Waktu rata-rata serangan DDoS kurang dari 10 menit menurut laporan Carero [4]. Serangan DDoS dapat merugikan organisasi perusahaan sebesar $50.000 dalam pendapatan yang hilang dari waktu henti dan biaya mitigasi. Sebagai vektor serangan, 71%, banjir SYN, dan serangan DNS tetap menjadi vektor serangan DDoS paling populer di Q3 tahun 2022 menurut Cloudflare. Serangan HTTP DDoS yang bertujuan untuk mengganggu server web, juga merupakan serangan baru yang dapat mengganggu layanan server seperti serangan DDoS lainnya. Sebagai solusi terhadap serangan DDoS, Penyedia Layanan Internet (ISP) dan perusahaan menggunakan banyak solusi dan Pusat Scrubbing seperti Radware DefensePro, Radware Cloud DDoS Protection Service, Cloudflare Layanan Mitigasi DDoS, Akamai Edge DNS, Arbor Cloud, Perlindungan DDoS F5 Silverline, Layanan Mitigasi DDoS Nexusguard, Perlindungan DDoS Oracle Dyn, Perlindungan DDoS Azure, dll. Semua solusi ini menawarkan fitur-fitur seperti perlindungan berlapis-lapis, deteksi ancaman waktu nyata, pelaporan, dan analitik. Namun, serangan DDoS yang sebenarnya lebih cepat dengan tingkat yang belum pernah terjadi sebelumnya, dan lebih canggih. Botnet lebih terdesentralisasi dan sangat aman. Menggunakan pusat scrubbing cerdas adalah arah baru untuk meningkatkan otomatisasi dan presisi pusat. Oleh karena itu, pusat scrubbing berbasis pembelajaran mesin dianggap sebagai Pusat Scrubbing Generasi Berikutnya (NGSC). Layanan berbasis internet dapat memiliki persyaratan keamanan yang berbeda seperti kerahasiaan, integritas, dan ketersediaan. Yang terakhir, khususnya, dapat menjadi sangat penting/kritis untuk layanan tertentu. Penyerang terutama menargetkan ketersediaan layanan melalui serangan Denial of Service (DoS). Serangan DoS terjadi ketika pengguna yang sah tidak dapat mengakses sistem dan sumber daya yang mereka butuhkan karena tindakan jahat dari penyerang cyber [5]. Serangan DoS dapat diluncurkan secara terdistribusi, dalam apa yang disebut serangan Distributed DoS (DDoS), saat beberapa mesin beroperasi bersama untuk menyerang target (Gbr. 1). Serangan DDoS biasanya membanjiri korban dengan sejumlah besar permintaan/paket untuk memenuhi sumber dayanya, sehingga tidak dapat lagi memenuhi permintaan pengguna yang sah.

DDoS dapat dideteksi dan dimitigasi menggunakan Sistem Deteksi Intrusi (IDS), yang dirancang untuk mendeteksi anomali lalu lintas yang terkait dengan strategi dan implementasi serangan. Meskipun efisien terhadap serangan DDoS "tradisional", IDS "tradisional" gagal mengatasi serangan DDoS yang rumit saat ini. IDS ini umumnya mengikuti pendekatan berbasis tanda tangan yang membuatnya tidak dapat belajar sendiri dan dengan demikian mengambil tindakan yang diperlukan kecuali jika dikonfigurasi untuk aturan/pola yang sesuai dan tindakan yang terkait. Untuk mengatasi keterbatasan tersebut, teknik IDS disempurnakan dengan algoritma Pembelajaran Mesin (ML) atau Pembelajaran Mendalam (DL) [6, 7, 8]. Hal ini saat ini mendapat banyak perhatian di bidang penelitian deteksi DDoS [9, 10, 11, 12].

IDS berbasis ML dapat mengidentifikasi dan bertahan terhadap pola indikatif DDoS yang diketahui. Sistem deteksi intrusi berbasis ML terdiri dari ekstraktor fitur dan model Pembelajaran Mesin yang berfungsi sebagai mesin deteksi. Pengode fitur mengatur data jaringan mentah untuk mengekstrak fitur yang sesuai untuk input model. Mesin deteksi dilatih menggunakan data DDoS dan data jinak agar dapat mengkategorikan sampel dalam lalu lintas waktu nyata. Sejumlah IDS berbasis ML [6, 7, 8, 13, 14] telah diusulkan dalam literatur yang menunjukkan akurasi deteksi yang tinggi untuk serangan DDoS. Namun, banyak algoritma pembelajaran mendalam dan pembelajaran mesin dipelajari berdasarkan satu set data, di mana set pelatihan dan pengujian diambil dari sumber yang sama. Oleh karena itu, jika data input berasal dari sumber eksternal di mana terdapat sedikit perubahan dalam ruang fitur input, kinerja jenis algoritma ini menurun sebagai akibatnya [15]. Aktor jahat mungkin menggunakan masalah generalisasi ini untuk mengarahkan pengklasifikasi untuk mencapai keputusan yang salah. Penyerang



Gambar 1: Serangan DDoS

Bahasa Indonesia: mungkin mengadaptasi serangan mereka untuk mencegah deteksi berbasis tanda tangan, yang membuat IDS berbasis ML rentan terhadap serangan adversarial. Ini disebut sebagai Adversarial Machine Learning (AML), di mana Generative Adversarial Networks (GAN) [16] digunakan untuk mencoba mengelabui detektor berbasis ML dengan menyediakan fitur palsu sebagai input untuk model. GAN adalah paradigma pembelajaran mendalam yang diperkenalkan pada tahun 2014, yang dapat mempelajari pola kumpulan data asli untuk membuat data serupa yang baru. GAN terdiri dari dua model jaringan saraf: generator dan diskriminator, yang telah dilatih untuk bertindak melawan satu sama lain. Serangan adversarial (yang dihasilkan oleh GAN) dapat lebih "ditingkatkan" melalui gangguan. Gangguan adversarial adalah penyesuaian kecil, tidak terlihat tetapi terarah pada input model ML, yang menghasilkan perilaku yang salah. Mereka telah terbukti sangat efektif dalam "menipu" algoritma klasifikasi berbasis ML [17].

Dalam makalah ini, kami menyajikan kerangka kerja berbasis GAN yang dapat menyediakan sampel adversarial DDoS yang kuat. Sampel-sampel ini kemudian dimodifikasi (yaitu, diganggu) dengan mengganti fitur-fiturnya dengan nilai-nilai dari sampel-sampel jinak. Kami mengevaluasi teknik gangguan terhadap IDS presisi tinggi dan mengamati penurunan signifikan dalam akurasi deteksi IDS. Dengan demikian, kami mengusulkan untuk menggunakan GAN guna meningkatkan presisi IDS, dengan melatih model baru berdasarkan sampel adversarial yang dihasilkan agar dapat mendeteksinya nanti. IDS yang diusulkan terdiri dari dua model: model pertama bertanggung jawab untuk memblokir sampel adversarial, dan model kedua membedakan antara DDoS dan contoh-contoh jinak.

Kami menyusun sisa makalah sebagai berikut. Bagian II menyajikan tinjauan umum karya terkait pada IDS berbasis ML dan serangan adversarial terhadap IDS. Bagian III menyajikan model serangan. Bagian IV menjelaskan eksperimen dan hasil pembuatan serangan adversarial, dan gangguan sampel DDoS palsu. Kami membahas peningkatan kinerja IDS di Bagian V. Bagian VI menjelaskan implementasi IDS di penyedia layanan internet (ISP). Dan bagian. Terakhir, bagian VII menyimpulkan makalah dan menyajikan pekerjaan kami di masa mendatang.

II. RGEMBIRAKAMIORK DANRPENELITIANCKONTRIBUSI

Seperti yang telah kami sebutkan sebelumnya, penggunaan machine learning dan deep learning telah menarik banyak perhatian di bidang deteksi DDoS. Bagian ini menyoroti skema deteksi DDoS berbasis ML/DL yang paling dikenal dan menekankan kontribusi utama dari karya ini dibandingkan dengan solusi yang ada.

A. Karya Terkait

Selama dekade terakhir, beberapa upaya penelitian [10, 11, 12, 14, [18, 19, 20] telah melakukan penelitian tentang IDS berbasis ML dengan akurasi tinggi. Dalam [13], penulis mengusulkan metode yang merupakan ansambel pemilihan fitur yang memanfaatkan informasi yang diperoleh menggunakan dataset CIC-IDS2017 [21]. Menurut hasil, metode ansambel ini untuk dataset Jumat pagi memiliki akurasi sebesar 97,86%. Namun, akurasi prediksi untuk file log Jumat sore adalah 73,79% untuk 16 fitur saat memanfaatkan pemilihan fitur berbasis perolehan informasi dan model ML berbasis analisis regresi. Kerugian utama pendekatan mereka adalah kompleksitas komputasinya yang tinggi. Penulis [14] meneliti banyak metode machine learning (Extreme Gradient Boosting (XGBoost), K-Nearest Neighbor (KNN), dan arsitektur deep learning Convolutional Neural Network (CNN) untuk mengidentifikasi dan mengkategorikan serangan DDoS menggunakan dataset CIC-DDoS2019[22]. Hasilnya menunjukkan bahwa XGBoost memperoleh akurasi maksimum sebesar 89,29%, sedangkan CNN dan KNN juga memberikan hasil yang sebanding. Dataset yang digunakan dalam penelitian ini tidak seimbang, dengan persentase data normal yang rendah dibandingkan dengan data serangan, yang dapat menyebabkan kesalahan deteksi dalam beberapa skenario waktu nyata dengan varian data normal.

Untuk mendeteksi berbagai jenis serangan seperti serangan HTTP Flood, serangan Smurf, dan serangan UDP Flood, penulis [19] menggunakan algoritma machine learning seperti K-Nearest-Neighbors (KNN), Naive Bayes (NB), Decision Tree (DT), Artificial Neural Network (ANN), dan Support Vector Machine (SVM). Mereka menemukan bahwa ANN melampaui teknik lainnya dengan akurasi 98,645%. Keterbatasannya adalah semua classifier tidak dapat mendeteksi serangan DoS kelas Smurf.

Dalam [10], penulis mengusulkan IDS berbasis LSTM untuk mendeteksi serangan DoS. Mereka mengevaluasi kerangka kerja yang diusulkan menggunakan dataset CICIDS-2017 [21] dan NSL-KDS [23]. Hasil yang diperoleh menunjukkan bahwa LSTM dapat secara efektif mendeteksi serangan DoS dengan akurasi 99,2% dengan CICIDS-2017 dan 98,6% dengan NSL-KDD. Kerangka kerja mereka diuji dan dievaluasi hanya pada serangan DOS.

Penulis [8] menggabungkan dataset CIC-DDoS2019 [22] dengan data DDoS yang dihasilkan menggunakan simulator BoNesi dan SlowHTTPTest. Mereka mengevaluasi kinerja detektor DDoS yang diusulkan yaitu model konvolusi dan model berbasis LSTM. Penulis membahas kesesuaian model-model ini dalam jaringan IoT. Hasilnya menunjukkan bahwa akurasi identifikasi LSTM yang diusulkan mencapai 98,9%. Kemudian, ketika mereka mengimplementasikan model LSTM di server edge dengan kapasitas komputasi yang lebih besar daripada komputer pribadi, mereka menemukan bahwa model tersebut memenuhi persyaratan penundaan IoT. Meskipun menggunakan model LSTM, mereka tidak menyelidiki kerentanan IDS berbasis ML terhadap serangan adversarial.

Dalam [11], penulis mengusulkan sistem deteksi intrusi berdasarkan algoritma pembelajaran mendalam. Mereka mengevaluasi berbagai model seperti Deep Neural Network (DNN), Convolutional Neural Networks (CNN), dan Long Short Term Memory (LSTM) menggunakan dataset CIC-DDoS2019 [22]. Hasil yang diperoleh menunjukkan bahwa model CNN memberikan hasil terbaik di antara model yang diusulkan. Penulis memperoleh akurasi 99,99% untuk klasifikasi biner (kelas normal/abnormal) dan 99,30% untuk klasifikasi multikelas. Keterbatasan di sini adalah model tersebut kurang generalisasi karena penggunaan beberapa fitur dengan distribusi yang tidak merata.

Bagian I telah menyoroti, kerentanan IDS berbasis ML terhadap gangguan adversarial telah dieksplorasi dan dievaluasi dalam [24, 25, 26, 27].

Dalam [25], penulis menyelidiki serangan adversarial yang menargetkan IDS berbasis anomali di lingkungan Software Defined Networks (SDN). Ukuran payload, kecepatan paket, dan volume lalu lintas dua arah adalah tiga fitur yang menjadi fokus penulis untuk diganggu. Serangan dilakukan dalam catatan banjir SYN yang diklasifikasikan

dideteksi oleh berbagai model ML. Hasil yang diperoleh menunjukkan efektivitas strategi serangan tersebut dengan mengurangi akurasi deteksi IDS yang ditargetkan dari 100% menjadi 0%. Makalah ini berfokus pada pembuatan serangan baru tanpa menyelidiki model deteksi untuk serangan tersebut.

Dalam [24], penulis membagi fitur-fitur dalam dataset KDDCup99 [28] menjadi fitur-fitur yang tidak dapat dimodifikasi yang diperlukan untuk menjaga fungsi jahat tetap berjalan dan fitur-fitur yang dapat dimodifikasi yang dapat dimodifikasi tanpa mengganggu fungsi jahat tersebut. Mereka menggunakan jaringan saraf berbasis Wasserstein GAN (WGAN) [29] untuk mengendalikan fitur-fitur yang dapat dimodifikasi dari serangan DDoS. Menurut hasil yang dilaporkan, akurasi deteksi IDS menurun hingga 50% yang menunjukkan bahwa IDS berbasis ML rentan terhadap jenis serangan ini.

Untuk menghindari deteksi, penulis [26] merancang kerangka kerja yang menggunakan GAN untuk membangun malware adversarial. Tujuan dari pekerjaan ini adalah untuk menggunakan detektor malware black-box karena penyerang tidak mengetahui teknik deteksi yang digunakan dalam detektor. Daripada langsung menyerang detektor white-box, para peneliti mengembangkan model yang dapat mengumpulkan data dari sistem black-box yang menjadi target. Kemudian, model ini menggunakan komputasi gradien dari GAN untuk menghasilkan data malware yang terganggu. Penulis dapat mencapai akurasi model sekitar 98% menggunakan dataset Drebin [30]. Keterbatasannya adalah penulis tidak menyelidiki stabilitas proses pelatihan model GAN.

Penulis [27] mengusulkan solusi untuk menghindari deteksi file PDF berbahaya. Pendekatan penyerangan dapat menghasilkan vektor fitur yang mirip dengan yang dihasilkan oleh WGAN untuk file PDF yang tidak berbahaya. Memungkinkan fitur file PDF berbahaya untuk mencocokkan fitur file PDF yang tidak berbahaya, untuk menghindari klasifikasi oleh detektor. Penulis mengevaluasi kinerja pendekatan yang diusulkan menggunakan dataset Contagio[31]. Hasil keluaran menunjukkan bahwa sampel adversarial yang dibuat oleh metode ini mencapai tingkat penghindaran 100%. Strategi ini terbatas pada format khusus PDF.

Penulis [32] mengusulkan strategi pertahanan berdasarkan model ensemble ML dan pelatihan adversarial. Lalu lintas jaringan hanya dianggap normal jika semua model setuju pada klasifikasi ini, yang memaksa penyerang untuk mengembangkan sampel yang dapat melewati semua model pada saat yang sama. Selain itu, penulis menggunakan pelatihan adversarial untuk menambahkan sampel adversarial ke set data pelatihan, meningkatkan ketahanan model individual. Penulis mengevaluasi strategi pertahanan pada set data CICIDS2018 [33] dan menemukan bahwa hal itu berhasil menurunkan tingkat keberhasilan serangan adversarial. Menggunakan teknik pembelajaran ensemble memiliki kompleksitas komputasi yang tinggi.

Penulis [34] mengusulkan strategi untuk meningkatkan IDS terhadap serangan DDoS adversarial. Penulis melatih dua model LUCID to citeLUCID untuk mendeteksi serangan banjir SYN dan banjir HTTP GET. Mereka menemukan bahwa model berkinerja tinggi ini rentan terhadap sampel DDoS yang terganggu adversarial. Mereka mengusulkan teknik pertahanan menggunakan model GAN. Kemudian mereka menggunakan data terganggu yang dihasilkan untuk melatih model LUCID di atasnya. Hasilnya, mereka menemukan bahwa model tersebut mencapai skor F1 lebih dari 98% dan bahwa rasio Negatif Palsu menurun hingga kurang dari 1,8% pada lalu lintas DDoS yang terganggu. Keterbatasan pekerjaan mereka adalah bahwa kumpulan data yang dihasilkan yang digunakan untuk meningkatkan model LUCID didasarkan pada teknik gangguan tunggal dengan memodifikasi sejumlah fitur yang terbatas. Tabel III merangkum gagasan utama, hasil, dan batasan skema deteksi DDoS berbasis ML yang dijelaskan di atas.

B. Kesenjangan literatur dan kontribusi penelitian

Seperti yang telah kami jelaskan sebelumnya, model klasifikasi berbasis ML dan DL (misalnya, [10, 11, 13, 14, 19] berkinerja buruk ketika ruang fitur input berubah. Hal ini membuat IDS berbasis ML atau DL rentan terhadap serangan siber yang dihasilkan menggunakan model ML/DL lain seperti GAN dengan beberapa modifikasi fitur seperti yang telah ditunjukkan oleh karya seperti [24, 25, 26, 27].

Baru-baru ini, beberapa upaya penelitian [24, 25, 26, 27] telah difokuskan pada pembuatan serangan adversarial menggunakan GAN dan menyelidiki apakah ini Serangan dapat dideteksi oleh IDS. Ditemukan bahwa sebagian besar upaya ini tidak memprioritaskan pelatihan IDS dengan data adversarial yang dihasilkan oleh GAN dan menguji apakah IDS dapat mengidentifikasi jenis serangan yang sama. Lebih jauh, penggunaan sampel yang diproduksi berdasarkan model korban merupakan kelemahan umum dari pendekatan pelatihan adversarial [32, 34], di mana mereka menggunakan kumpulan data yang dihasilkan yang didasarkan pada teknik gangguan tunggal, atau yang didasarkan pada model korban tunggal. Akibatnya, model serangan mengembangkan kemampuan untuk menghasilkan sampel adversarial yang lemah, dan daripada melatih IDS untuk melindungi terhadap gangguan yang signifikan, ia melatih dari sampel serangan yang lemah, dan tetap rentan terhadap serangan adversarial. Untuk mengatasi kelemahan utama ini, dalam makalah ini, kami mengembangkan metode deteksi baru yang kuat di IDS untuk mendeteksi sampel DDoS secara akurat, terlepas dari apakah fitur serangan telah diganti (yaitu, terganggu) atau tidak.

Kami merangkum kontribusi penelitian utama dari makalah ini sebagai berikut:

- Kami mengembangkan generator model GAN yang mampu menciptakan lalu lintas DDoS yang sangat mirip dengan contoh DDoS dari kumpulan data. Kami memodifikasi nilai yang ada dalam fitur fungsional DDoS dalam lalu lintas DDoS yang dihasilkan agar terlihat mirip dengan contoh yang tidak berbahaya.
- Kami membangun kumpulan data baru berdasarkan kombinasi data yang dihasilkan dan data asli, dengan dua kelas: nyata dan palsu.
- Kami melatih model baru menggunakan kumpulan data baru, untuk dapat mendeteksi data palsu atau data yang dibuat sendiri.
- Kami melatih model lain menggunakan kumpulan data asli yang hanya mencakup fitur fungsional DDoS, untuk dapat membedakan antara sampel DDoS dan sampel normal.

IDS yang diusulkan dalam penelitian ini dapat digunakan oleh organisasi untuk melindungi jaringan mereka dari serangan DDoS. Misalnya, IDS dapat digunakan oleh perusahaan untuk melindungi situs webnya agar tidak ditutup oleh serangan DDoS atau oleh lembaga pemerintah untuk melindungi infrastruktur pentingnya agar tidak terganggu oleh serangan semacam itu. Kemampuan untuk mendeteksi serangan DDoS yang bersifat adversarial yang dihasilkan dengan model WGAN memberikan peningkatan yang berharga dalam keamanan jaringan dan membantu organisasi ini menjaga ketersediaan dan keandalan layanan mereka.

III. SebuahTACKMODEL

Tujuan utama serangan DDoS adalah mengganggu ketersediaan server, sehingga tidak dapat diakses oleh pengguna yang sah. Skema deteksi tradisional (baik yang tertanam dalam IDS atau sebagai modul terpisah) bergantung pada identifikasi tanda tangan serangan DDoS dalam lalu lintas yang dipantau. Sebagian besar skema deteksi DDoS mengusulkan penyerang DDoS "tradisional" dan mesin deteksi DDoS yang disempurnakan dengan ML. Dalam skema kami, kami berasumsi bahwa penyerang yang disempurnakan dengan ML bersama dengan kemungkinan menghasilkan lalu lintas DDoS "tradisional", dapat membuat lalu lintas serangan DDoS yang bermusuhan, di mana fitur serangan DDoS dimanipulasi untuk mengikuti fitur normal guna menghindari deteksi oleh model IDS. Target penyerang adalah jaringan atau server tempat lalu lintas jaringan diamati oleh IDS berbasis ML, tetapi penyerang tidak memiliki akses langsung ke model itu sendiri atau proses pemantauan arsitektur IDS. Dengan kata lain, targetnya adalah IDS berbasis ML kotak hitam. Namun mengingat serangan DDoS telah dipelajari secara ekstensif dalam literatur, penyerang dapat memanfaatkan pengetahuan ini untuk mempelajari fitur lalu lintas mendasar yang digunakan oleh IDS untuk mendeteksi serangan. Fitur lalu lintas dasar ini memungkinkan penyerang untuk mengendalikan nilainya guna meniru distribusi lalu lintas jaringan yang tidak berbahaya, sehingga serangan DDoS tetap valid.

A. Penentuan fitur fungsional DDoS

Model penyerang bertujuan untuk menghasilkan serangan DDoS pada lalu lintas jaringan yang tidak terdeteksi oleh IDS berbasis ML.

TABEL I: Ringkasan karya terkait

Referensi	Kategori	Kumpulan data	Ide pokok	Pertunjukan evaluasi	Keterbatasan
[11]	IDS berbasis Ml	CICIDS-2017, NSL-KDS	Model LSTM	Ketepatan dari 99,2% dengan CICIDS-2017 dan 98,6% dengan NSL-KDD	Terbatas pada satu jenis serangan yang disebut DoS
[13]	IDS Berbasis ML	CIC-DDoS2019	Mempelajari ensemble pemilihan fitur menggunakan perolehan informasi dan analisis regresi	Akurasi 97,86%	Akurasi prediksi untuk file log Jumat sore adalah 73,79% untuk 16 atribut
[14]	IDS Berbasis ML	CIC-DDoS2019	XGBoost, KNN, CNN, Hutan Acak, SVM, dan ANN	XGBoost memperoleh nilai tinggi ketepatan	Dataset tidak seimbang, dan tidak ada serangan sampelnya lebih sedikit dibandingkan sampel serangan
[11]	IDS Berbasis ML	CIC-DDoS 2019	DNN, CNN, dan LSTM	Model CNN memberikan hasil terbaik yaitu akurasi 99,99% untuk klasifikasi biner dan 99,30% untuk klasifikasi multikelas. klasifikasi	Mereka menggunakan beberapa fitur dengan distribusi yang tidak merata
[19]	IDS Berbasis ML	CIC-DDoS2019	Regresi linier, KNN, Naïve Teluk, Keputusan Pohon, Acak Hutan, SVM, dan ANN	ANN mengungguli sisa metode dengan akurasi 98,645%	Semua pengklasifikasi tidak dapat mendeteksi kelas Smurf
[24]	Serangan yang bersifat permusuhan	Piala KDDC99	Bahasa Indonesia: WGAN	Akurasi menurun sebesar 50%	Sampel yang dihasilkan adalah berdasarkan model korban tunggal
[25]	Serangan yang bersifat permusuhan	CIC-DDoS2019	Menyerang anomali-IDS berbasis SDN dengan mengganggu beberapa fitur menggunakan GAN	Akurasi turun dari 100% hingga 0%	Serangan ini hanya diterapkan pada jejak banjir SYN dan berdasarkan pada satu korban/model yang menjadi target
[26]	Serangan yang bersifat permusuhan	Bahasa Inggris Drebin	Berbasis GAN menyerang model melawan mesin sedang belajar-berdasarkan pengklasifikasi untuk deteksi perangkat lunak jahat	Mengelabui pengklasifikasi hingga 99%	Model GAN yang sulit dilatih
[27]	Serangan yang bersifat permusuhan	Penyakit menular	Menghasilkan sekumpulan vektor fitur yang mirip dengan fitur file PDF jinak menggunakan WGAN	Tingkat serangan penghindaran 100%	Terbatas pada format tertentu yaitu PDF
[32]	Serangan musuh dan strategi pertahanan	CIC-DDoS2019	Pertahanan menggunakan ensemble pemungutan suara berdasarkan tiga model ML yang berbeda	Penurunan tingkat keberhasilan serangan musuh	Model serangan didasarkan pada satu sistem atau model yang ditargetkan
[34]	Serangan musuh dan strategi pertahanan	CIC-DDoS2019	Strategi pertahanan menggunakan Model GAN untuk meningkatkan model LUCID pada IDS	Model mencapai skor F1 sebesar 98%, tingkat negatif palsu turun hingga di bawah 1,8%	Dataset yang dihasilkan digunakan untuk meningkatkan model LU-CID didasarkan pada teknik gangguan tunggal
Usulan kami mendekati	Serangan musuh dan strategi pertahanan	CIC-DDoS2019, CICID-2017	Meningkatkan IDS dengan menambahkan LSTM lainnya model bertanggung jawab atas memblokir sampel yang berlawanan	Model ini mencapai akurasi deteksi sebesar 91,75% terhadap sampel adversarial yang terganggu	Kompleksitas dan memakan waktu algoritma ini karena menggunakan 2 model

Oleh karena itu, model generator harus mampu mengendalikan nilai fitur yang digunakan oleh IDS untuk memutuskan apakah akan memblokir lalu lintas atau meneruskannya ke server. Untuk mencapai hal ini, kita memerlukan kesadaran kualitatif tentang hubungan antara prediksi model dan atribut instans data yang digunakan untuk membuat prediksi tersebut. Kita memerlukan teknik ML yang dapat dijelaskan yang mengklarifikasi beberapa aspek ini. SHapley Additive Explanations (SHAP) [35] adalah salah satu metode ini. SHAP digunakan untuk menjelaskan bagaimana setiap fitur memengaruhi model dan memungkinkan analisis lokal dan global untuk kumpulan data.

SHAP adalah penjelas model-agnostik individual. Asumsi yang dibuat oleh pendekatan model-agnostik adalah bahwa model yang dijelaskan adalah kotak hitam [36], dan tidak diketahui bagaimana model berfungsi secara internal. Akibatnya, pendekatan model-agnostik hanya dapat mengakses data masukan dan prediksi model. Ide di balik pentingnya fitur SHAP adalah bahwa fitur dengan nilai Shapley absolut yang besar [37] adalah penting. Tujuan SHAP adalah untuk menjelaskan prediksi contoh x dengan menghitung kontribusi setiap fitur terhadap prediksi. Nilai Shapley dihitung dengan metode penjelasan SHAP menggunakan teori permainan koalisi. Data

Nilai fitur instance berpartisipasi dalam koalisi sebagai pemain. Kita dapat mendistribusikan "pembayaran" secara adil di antara fitur-fitur dengan menggunakan nilai Shapley. Penjelasan SHAP ditentukan sebagai berikut:

$$G(\text{dari Bahasa Indonesia: } akt_{\text{angka 0}}) = \sum_{j=1}^{N} aku_j \text{dari } j \quad (1)$$

di mana g adalah model penjelasan, $z' \in \{\text{angka 0 Bahasa Indonesia: } 1\}^N$ adalah vektor koalisi, N adalah ukuran koalisi maksimum, dan $aku_j \in$ Radalah atribusi fitur untuk suatu fitur j. Fitur-fitur ini bertindak sebagai fitur fungsional DDoS yang akan dikontrol oleh model serangan untuk menghindari deteksi oleh IDS.

B. Model musuh

Generative Adversarial Network (GAN) adalah paradigma berbasis pembelajaran mesin yang diusulkan pada tahun 2014 [38]. Ide di balik GAN adalah bahwa jaringan saraf dapat digunakan untuk menghasilkan contoh-contoh baru dari distribusi yang ada. Namun, GAN terdiri dari dua jaringan saraf, generator (G), dan diskriminan (D).

Untuk menghasilkan sampel yang mirip dengan data sebenarnya, generator memperkirakan distribusi probabilitas dari sampel sebenarnya. Untuk menentukan apakah sampel berasal dari data sebenarnya dan bukan data yang dihasilkan, diskriminator dilatih. Dalam metode ini, kedua komponen dilatih untuk bersaing satu sama lain. Dengan menyatukan kedua sisi, diskriminator dan generator memainkan permainan minmax. Kita harus meningkatkan fungsi kerugian (Persamaan (2)), di mana x adalah data sebenarnya, z adalah vektor laten (noise), E adalah nilai yang diharapkan, D merupakan diskriminator dan G merupakan generator, dan D dan G keduanya adalah fungsi diferensial yang direpresentasikan oleh multi-layer perceptron. Sementara itu, L(G,D) merupakan nilai fungsi kerugian.

$$\text{menit}_G \text{Maksimal}_C \text{Saya}(G, D\text{Bahasa Indonesia: } \textit{Bahasa Inggris}_{X \in P_{\text{data}}(X)}[\text{catatan}(D(X))] + \textit{Bahasa Inggris}_{\text{dari} \in P_{\text{dari}}(\text{dari})}[\text{catatan}(1 - D(\text{dari}))] \quad (2)$$

Metode GAN memiliki beberapa keterbatasan dan yang paling umum meliputi [39, 40]:

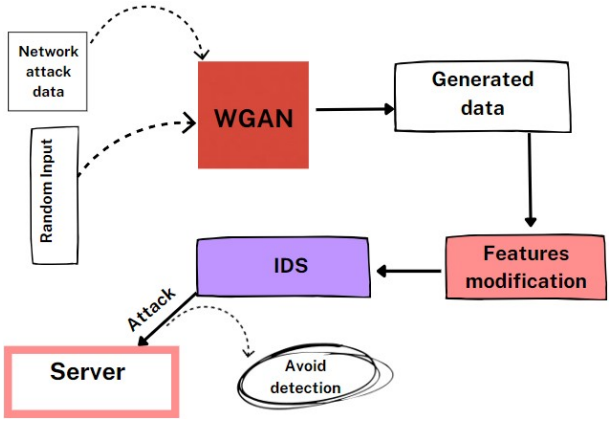
1) Konvergensi: Pada siklus berikutnya, seiring dengan peningkatan Generator, kinerja klasifikasi diskriminator menurun. Masalah konvergensi terjadi saat GAN dilatih dari titik ini karena generator dilatih dari data yang kurang bermakna.
2) Mode collapse: GAN dapat menghasilkan berbagai macam data. Namun, jika generator belajar membuat kumpulan data tertentu, agar diskriminator dapat mengklasifikasikannya sebagai data asli, generator hanya akan menghasilkan kumpulan data ini dan dengan mudah mengelabui diskriminator.

Untuk mengatasi masalah ini, penulis [29] mengusulkan metode Wasserstein GAN (WGAN). WGAN memberikan perkiraan yang lebih baik dari data terdistribusi dalam set pelatihan. WGAN menggunakan generator dan kritikus untuk memberikan skor yang menunjukkan seberapa nyata dan palsu data yang dihasilkan. Fungsi nilai WGAN menggunakan dualitas Kantorovich-Rubinstein adalah sebagai berikut:

$$\text{menit}_G \text{Maksimal}_{C \in A} \textit{Bahasa Inggris}_{X \sim \text{Data } P_r(X)}[C(X)] - \textit{Bahasa Inggris}_{\text{dari} \sim P_z(\text{Bahasa Indonesia: } z)}[C(G(\text{dari}))] \quad (3)$$

di mana (A) adalah himpunan fungsi 1-Lipschitz [41]. Generator (G) membuat titik data sampel dan mengirimkannya ke pengkritik (C) setelah menerima variabel laten (Z) dari distribusi normal multivariat standar. Setelah generator dan pengkritik dilatih bersama, generator pada akhirnya akan menyerupai transformasi deterministik yang menghasilkan data yang mirip dengan data sebenarnya.

Pembuatan serangan DDoS adversarial dibagi menjadi dua tahap. Pada tahap pertama, kami menggunakan WGAN untuk membuat data adversarial yang mengikuti secara dekat kejadian DDoS dari kumpulan data masukan. Kami kemudian mengevaluasi model generator dengan skor kesamaan dan visualisasi jumlah kumulatif per fitur antara dataset yang dihasilkan dan dataset asli. Pada tahap kedua, kami mengganti



Gbr. 2: Menghasilkan serangan DDoS yang bersifat adversarial

Nilai yang disajikan dalam fitur fungsi DDoS dipadukan dengan nilai dari sampel jinak, sedangkan sampel yang terganggu dimasukkan ke model IDS berbasis ML seperti yang ditunjukkan pada Gambar 2.

IV. Bahasa IndonesiaPERCOBAAN: GENERASI ANTARA musuh ATACK

Pada rangkaian percobaan pertama, kami menyelidiki efektivitas model dalam menghasilkan data serangan DDoS yang mengikuti distribusi data sampel dalam kumpulan data. Selain itu, kami mengevaluasi teknik gangguan yang diusulkan dalam pembuatan serangan DDoS adversarial yang bertujuan untuk menghindari deteksi oleh IDS dengan membangun IDS standar dan mengujinya dengan (2) serangan DDoS asli dan adversarial.

A. Kumpulan Data

Dalam penelitian ini, kami menggunakan dua dataset yang berbeda, yaitu dataset CIC-DDoS2019 [22] dan CIC-IDS2017 [21]. Dalam dataset CIC-DDoS2019, jaringan korban adalah infrastruktur keamanan tinggi dengan firewall, router, switch, dan beberapa sistem operasi umum. Serangan jaringan adalah infrastruktur pihak ketiga yang sepenuhnya terpisah yang mengeksekusi berbagai jenis serangan DDoS termasuk bentuk serangan DDoS yang paling umum. CICFlowMeter-V3 [42] digunakan untuk mengekstrak lebih dari 80 fitur lalu lintas seperti port tujuan, durasi aliran, dan jumlah paket per detik dari arus lalu lintas. Dataset yang dihasilkan berisi 251723 sampel serangan dan 3134 sampel lalu lintas jinak. Oleh karena itu, untuk menghindari masalah kelas yang tidak seimbang, kami menggabungkan dataset CIC-DDoS2019 dengan sampel kelas jinak dari dataset CIC-IDS2017 berdasarkan fitur bersama. Dataset yang diperbarui berisi 251723 sampel serangan DDoS dan 251723 sampel lalu lintas jinak dan 80 fitur.

Kami melakukan praproses (Gbr. 3) dataset dengan menghapus nilai null dan tak terbatas, dan membuang beberapa fitur (21 fitur), baik karena tidak digunakan (misalnya, alamat IP sumber, port sumber, stempel waktu), atau karena distribusi data dalam fitur tersebut tidak merata (misalnya, jumlah tanda pengakuan, rasio turun/naik). Selanjutnya, untuk menormalkan dan menskalakan dataset, kami menerapkan penskalaan min-maks (4) untuk menempatkan semua fitur kami dalam skala yang sama. Setelah melakukan langkah praproses, kami membagi dataset menjadi 70% untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian untuk mengevaluasi model.

$$X = \frac{x - \text{menit}(X)}{\text{maks}(X) - \text{menit}(X)} \quad (4)$$

Kami menerapkan analisis SHAP pada dataset, dan Gambar 4 menunjukkan hasil efek global dari fitur-fitur dalam output. Dalam plot ini, sumbu y yang mewakili fitur-fitur yang tercantum dari

pengaruh tertinggi hingga terendah pada prediksi hanya didasarkan pada sumbu x yang mewakili rata-rata nilai SHAP absolut. Jadi, tidak relevan apakah fitur tersebut memiliki dampak positif atau negatif pada output, nilai nilai Shapley absolut per fitur di seluruh data dijelaskan dalam persamaan (5):

$$aku_j = \frac{1\sum^{N}}{N_{j=1}} \; //_{(Saya)} \; {}_{Bahasa\ Indonesia:} \tag{5}$$

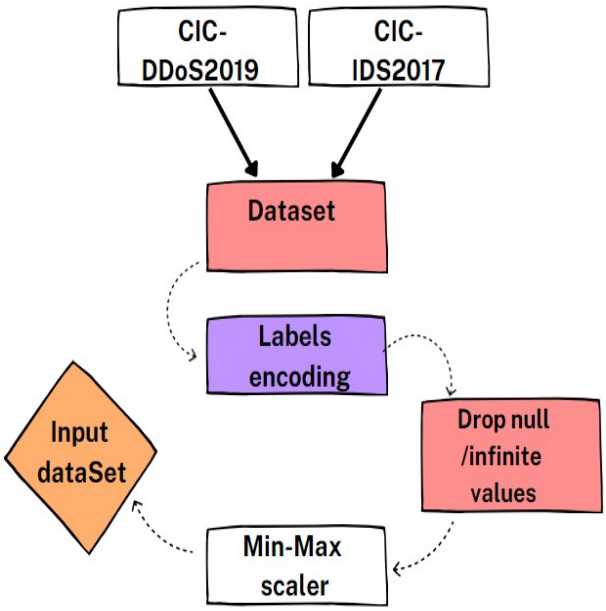
Informasi ini tidak cukup untuk memahami pentingnya data dan dampak dari setiap fitur yang disumbangkan pada hasil. Untuk alasan ini, plot lain dapat digunakan, yaitu plot kawanan lebah Gambar 5. Dalam plot kawanan lebah, fitur-fitur juga diurutkan berdasarkan efeknya pada prediksi, tetapi kita juga dapat melihat bagaimana nilai fitur yang lebih tinggi atau lebih rendah akan memengaruhi hasil. Di sini, sumbu x mewakili nilai SHAP, sumbu y mewakili fitur, dan warna titik menunjukkan apakah observasi itu memiliki nilai yang lebih tinggi atau lebih rendah. Sebagai contoh, nilai yang lebih rendah dari jumlah total byte yang dikirim di jendela awal dalam arah mundur (Init win bytes backward) memiliki dampak positif, tetapi nilai yang lebih tinggi memiliki dampak negatif pada hasil.

Fitur-fitur ini bertindak sebagai fitur DDoS fungsional yang dimodifikasi sedemikian rupa sehingga tampak seperti lalu lintas normal untuk menghindari deteksi oleh IDS.

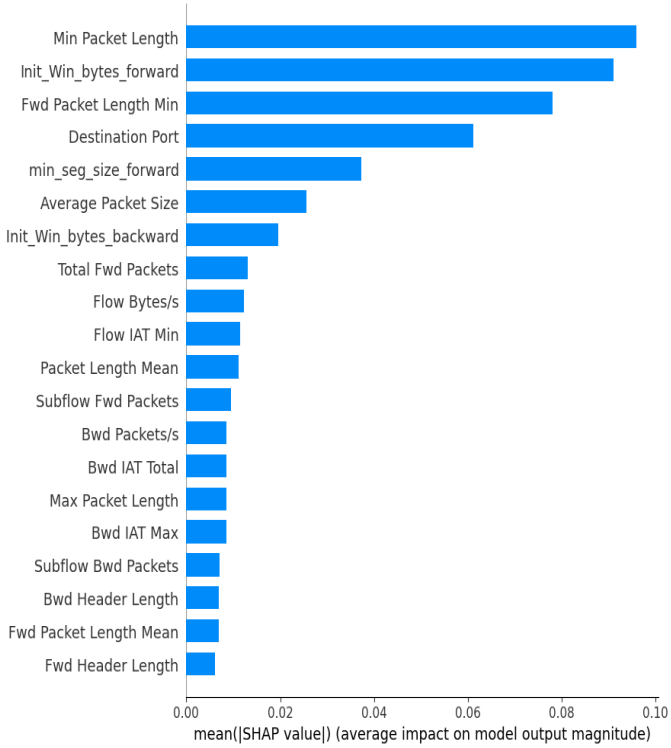
B. Menghasilkan serangan DDoS menggunakan model WGAN

Pada langkah ini, kami menggunakan WGAN untuk menghasilkan data adversarial yang mengikuti distribusi data DDoS dari set data input. Kemudian, kami mengevaluasi model generator berdasarkan skor kesamaan dan visualisasi jumlah kumulatif per fitur antara set data yang dihasilkan dan set data asli.
Model WGAN terdiri dari dua model, generator dan kritik, dan Gambar 6 menunjukkan jaringan yang menghubungkan kedua model ini yang dirancang dengan penalti gradien [43].

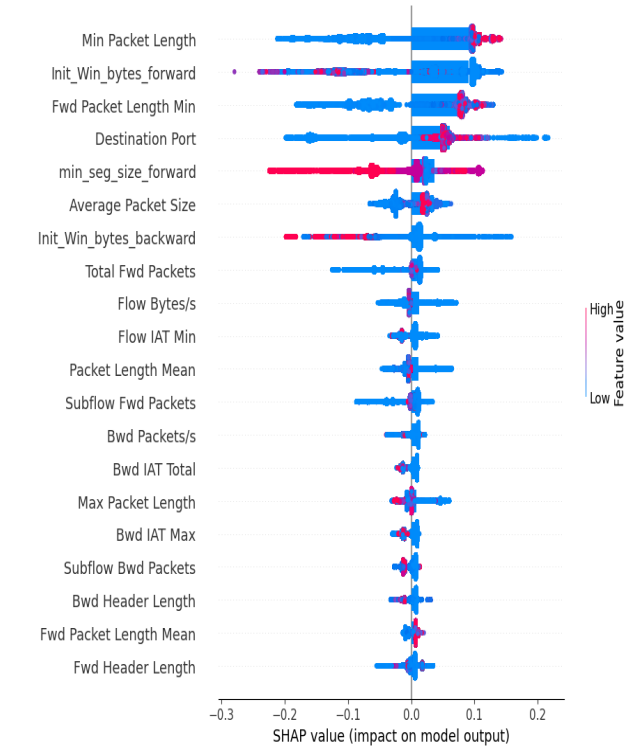
Generator adalah model yang terdiri dari lapisan yang terhubung sepenuhnya *Sepakbola*$_{saya \rightarrow}$ $_{Tidak\ ada}$Bahasa Indonesia:dengan fungsi aktivasi ReLU, dan lapisan tersembunyi dibentuk oleh penggabungan beberapa vektor yang dapat membentuk data yang mirip dengan data asli yang ditransformasikan dengan dimensi yang sama *saya*.



Gambar 3: Pra-pemrosesan dataset



Gbr. 4: Pentingnya fitur menggunakan SHAP



Gbr. 5: Dampak fitur menggunakan SHAP

Arsitektur ini secara formal dijelaskan dalam (IV-B):

$$H0 = \text{Bahasa Indonesia: } Z \text{ (vektor laten)}$$

$$H1 = \text{Kembali ke Lu}(\text{Sepakbola}_{\text{saya}\rightarrow H\text{angka }0})$$

$$H2 = \text{Kembali ke Lu}(\text{Sepakbola}_{\text{saya}\rightarrow H1})$$

Di mana $Sepakbola_{X\rightarrow dari}$ adalah lapisan yang terhubung sepenuhnya dengan ukuran input (X) dan ukuran keluaran *dari*, Dan $Kembali(X)$ menunjukkan penerapan aktivasi Relu pada x.

Kritik terdiri dari dua lapisan yang terhubung sepenuhnya dengan fungsi aktivasi LeakyRelu [44] dan dijelaskan dalam Equatilooks(6):

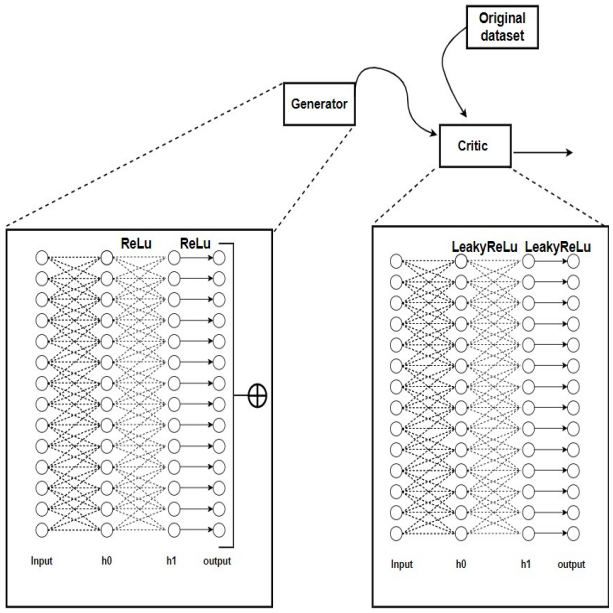
$$H0 = \text{keluaran generator}$$

$$H1 = \text{BocorReLu}_{\text{angka }0.01}(\text{Sepakbola}_{\text{saya}\rightarrow H\text{angka }0}) \quad (6) \text{Di mana}$$

$$H2 = \text{BocorReLu}_{\text{angka }0.01}(\text{Sepakbola}_{\text{saya}\rightarrow H1}) \text{kasus nd}$$

$\text{BocorReLu}_R(X)$ menerapkan fungsi LeakyRelfeaturetion dengan slop R pada X.

Setelah membangun model generator dan diskriminator, kami menginisialisasi parameter pelatihan. Kami menggunakan ukuran batch 256, pengoptimal Adam[45] dengan *sebuah* = 0,0002, *sebuah*1 =0,5, dan *sebuah*2 = 0,999. Kami memuat dataset, dan melatih model selama 100 periode. Model generator terkadang dapat menjadi terlalu kuat dan mengeksploitasi model Critic, yang menyebabkan masalah gradien menghilang dan menghasilkan sampel adversarial yang tidak valid. Untuk mencegah hal ini terjadi, kami telah melatih model Critic untuk periode yang lebih lama guna membangun model yang lebih kuat yang lebih mampu mengukur jarak antara sampel asli dan palsu. Ini membantu memastikan bahwa sampel adversarial yang dihasilkan valid dan secara akurat mencerminkan sifat statistik data lalu lintas yang sebenarnya. Dengan melatih model Critic untuk periode yang lebih lama, kami dapat meningkatkan kemampuannya untuk membedakan antara sampel asli dan palsu serta mengurangi kemungkinan menghasilkan sampel adversarial yang tidak valid. Untuk tujuan ini, untuk setiap periode, kami melatih Critic sebanyak 4 kali dan generator sebanyak 1 kali. Dalam pelatihan Critic, setiap kali kami menggunakan model generator untuk menghasilkan sampel palsu dari input noise. Kami secara acak memilih sampel dari dataset asli, dan berdasarkan prediksi Critic untuk kedua kelas sampel ini, kami memperbarui bobot Critic dengan mengurangi gradien. Di dalam generator



Gambar 6: Arsitektur model WGAN

pelatihan, kami menghasilkan sampel palsu, dan menggunakannya sebagai masukan untuk model kritik, dan berdasarkan keluaran, kami memperbarui bobot model generator dengan mengurangi gradien.

Ketika proses pelatihan berakhir, kami menggunakan model generator untuk menghasilkan kumpulan data palsu baru. Untuk mengevaluasi validitas sampel adversarial yang dihasilkan, kami mengikuti pendekatan evaluasi yang digunakan dalam literatur dengan membandingkan jumlah kumulatif per fitur. Kami membandingkan jumlah kumulatif sampel adversarial dengan sampel nyata, dan kami menggambar grafik yang menunjukkan perbedaan antara kumpulan data yang dihasilkan dan data asli. Gambar 7 menyajikan hasil perbandingan ini yang menunjukkan bahwa jumlah kumulatif sampel adversarial serupa dengan sampel nyata dan sebagian besar fitur sangat cocok dengan fitur data nyata. Ini menunjukkan bahwa sampel adversarial "valid" dan mengikuti distribusi statistik sampel nyata. Selain itu, di bagian berikutnya, kami membandingkan kinerja berbagai algoritme pembelajaran mesin pada data yang dihasilkan yang membantu kami menentukan apakah model yang dilatih pada data asli dapat digeneralisasi ke data yang dihasilkan dan memberikan wawasan tentang apakah distribusi data yang dihasilkan serupa dengan data asli. Dengan melakukan perbandingan ini, kami dapat menilai kualitas data yang dihasilkan.

C. Arsitektur Model IDS Standar

Bahasa Indonesia: Setelah mengevaluasi pembuatan sampel DDoS dan membandingkannya dengan data asli, tujuan kami berikutnya adalah mengevaluasi teknik perturbasi pada IDS berbasis ML berkinerja tinggi. Untuk memenuhi tujuan ini, kami perlu membangun IDS. Kami menggunakan berbagai algoritma pembelajaran mesin yang banyak digunakan dalam literatur yang menghasilkan akurasi tinggi dalam mendeteksi serangan DDoS. Algoritma tersebut meliputi pengklasifikasi pohon keputusan [46], peningkatan gradien ekstrem (XGBoost), [47], pengklasifikasi multilayer perceptron (MLPClassifier) [48], pengklasifikasi hutan acak [49], dan algoritma pembelajaran mendalam yaitu LSTM [50]. Kami menggunakan LSTM dalam evaluasi kami dan terdiri dari model LSTM Sederhana dengan lapisan DNN. Kami menggunakan fungsi aktivasi ReLU di semua lapisan yang ditambahkan ke fungsi aktivasi Sigmoid di lapisan terakhir, dan kami menggunakan entropi silang biner sebagai fungsi kerugian dengan pengoptimal ADAM.
Model terbaik dipilih untuk bertindak sebagai detektor dasar dalam mesin deteksi. Kami melatih model-model ini menggunakan data asli dan kami melakukan pengujian untuk mengevaluasinya. Selain itu, kami menguji semua model ini dengan serangan DDoS yang dihasilkan sebagai cara kedua untuk mengevaluasi model GAN. Tabel II menunjukkan hasilnya. F1 nyata mewakili skor F1 untuk algoritme ini pada kumpulan data asli, dan F1 palsu mewakili skor F1 7 pada kumpulan data palsu, menunjukkan bahwa LSTM mengungguli algoritme lain dengan skor F1 sebesar 0,99.

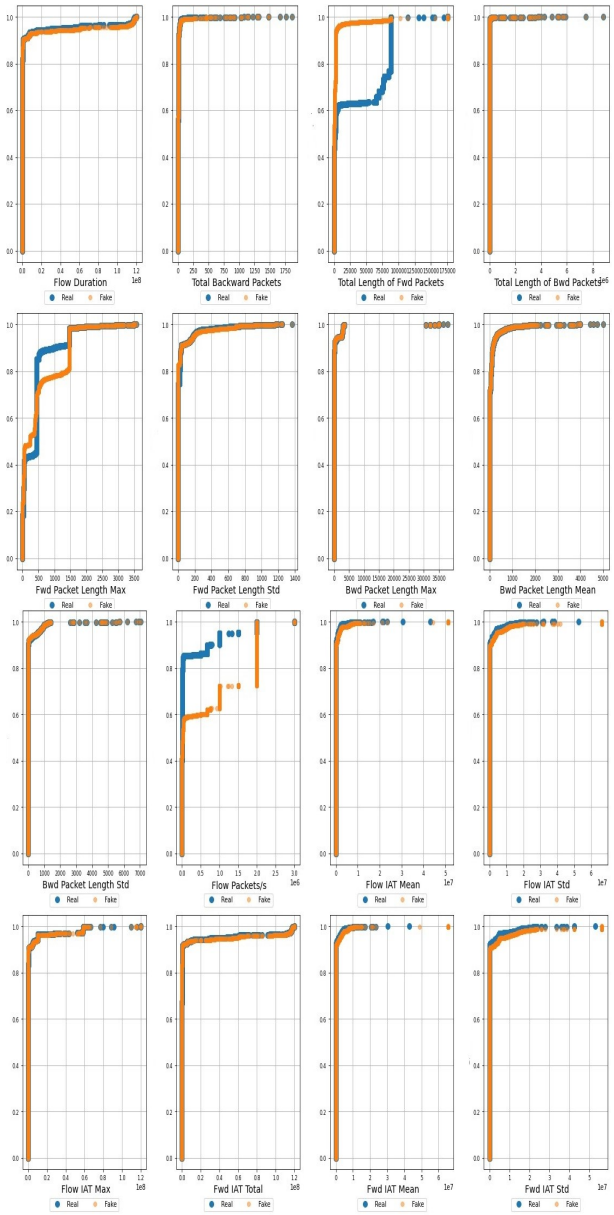
Sebagian besar model yang digunakan dalam percobaan mencapai hasil yang serupa dengan kumpulan data nyata dan data adversarial, yang menunjukkan bahwa sampel adversarial "valid" dan secara akurat mencerminkan sifat statistik dari sampel nyata. Hal ini menunjukkan bahwa model WGAN yang digunakan untuk menghasilkan sampel adversarial berhasil meniru distribusi statistik dari data lalu lintas nyata. Hasilnya, sampel adversarial dapat digunakan untuk mengevaluasi kinerja IDS secara efektif dalam mendeteksi serangan DDoS.

$$F1 = \frac{T.P. + \text{Bahasa Inggris}}{T.P. + \text{Bahasa Inggris} + \text{Bahasa Inggris} + \text{Bahasa Inggris}} \quad (7)$$

IDS berbasis ML kami, sejauh ini, terdiri dari encoder fitur, mesin deteksi (LSTM yang telah dilatih sebelumnya), dan pembuat keputusan (Sistem peringatan) seperti yang diilustrasikan pada Gambar 8.

D. Pembuatan serangan adversarial DDoS yang terganggu

Pada langkah ini, kami menghasilkan instance DDoS menggunakan model WGAN yang sama (Gbr. 2) tetapi kami mempertahankan beberapa fitur fungsional DDoS yang telah diekstraksi menggunakan model SHAP (Gbr. 4) untuk diikuti

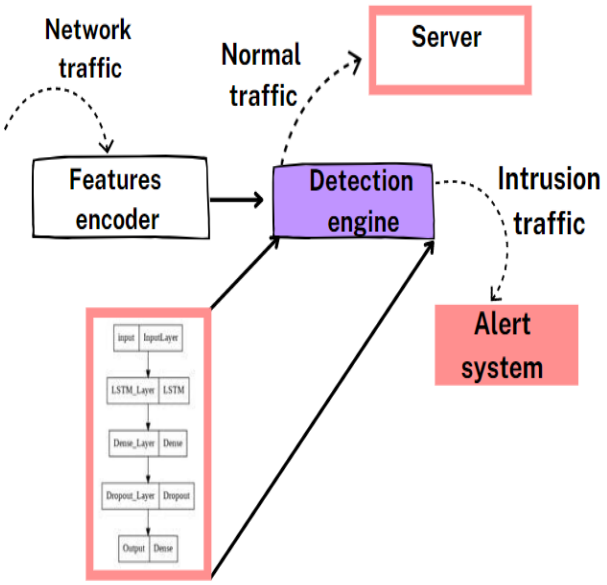


Gbr. 7: Jumlah kumulatif per fitur

Penggolong	f1 nyata	f1 palsu
Klasifikasi pohon keputusan	0.9570	0.9440
Pengklasifikasi MLP	0.9175	0.8470
XGBoost	0.9513	0,9065 tahun
Klasifikasi hutan acak	0,9722 tahun	0,9625 tahun
LSTM	0,99	0,9825 tahun

TABEL II: Hasil pengklasifikasi pembelajaran mesin.

distribusi fitur sampel jinak. Setiap kali kami menerapkan skenario baru, kami mengubah jumlah yang berbeda untuk fitur tersebut (8, 16, 20). Selanjutnya, kami mengevaluasi kemampuan IDS untuk mengklasifikasikannya secara akurat serta apakah akurasinya menurun. Kami melakukan pengujian untuk berbagai skenario ini, dan matriks kebingungan (Gbr. 9) menunjukkan hasilnya. Matriks kebingungan adalah matriks 2 x 2 yang digunakan untuk mengevaluasi kinerja model klasifikasi.



Gambar 8: Arsitektur model IDS

(A)

	Normal	DDoS
Normal	100 %	0 %
DDoS	0%	100 %

(B)

	Normal	DDoS
Normal	99.3 %	0.7 %
DDoS	77.2 %	22.8

(C)

	Normal	DDoS
Normal	99.4 %	0.6 %
DDoS	79.5 %	20.5 %

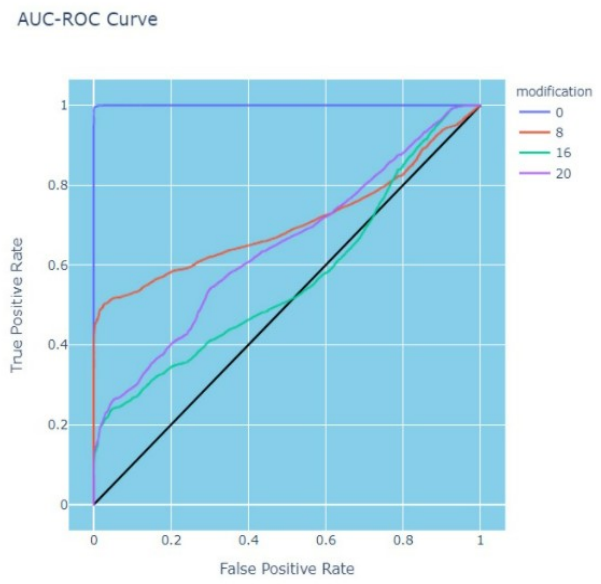
Gambar 9: Matriks kebingungan prediksi lalu lintas jaringan pada IDS dalam berbagai skenario: (a) Pada data asli, (b) Dengan hanya 16 modifikasi pada fitur fungsional, (c) Memodifikasi semua fitur fungsional.

Matriks membandingkan nilai target aktual dengan nilai yang diprediksi oleh model pembelajaran mesin.

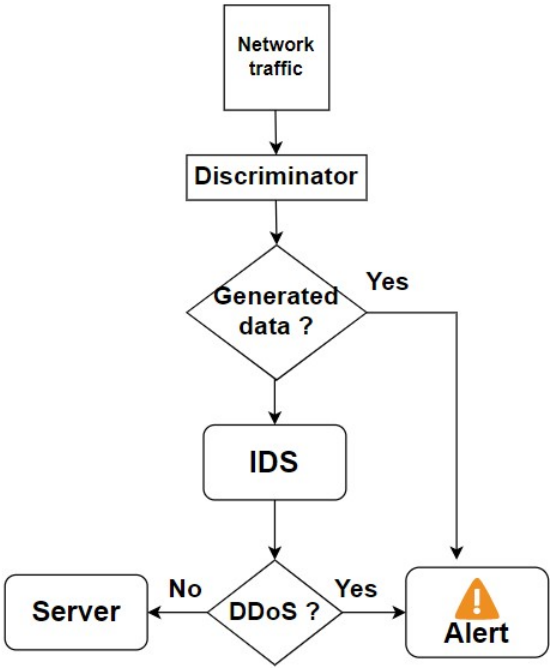
Dengan membandingkan ketiga matriks kebingungan, kami menyimpulkan bahwa IDS mampu mengklasifikasikan semua data secara akurat (Gbr. 9(a)) ketika tidak ada gangguan pada fitur masukan, namun di sisi lain, terdapat penurunan signifikan dalam kinerja ini ketika terdapat gangguan pada fitur masukan (Gbr. 9(b,c)).

Selain itu, untuk lebih memahami hasilnya, kami menunjukkan kurva AUC-ROC. Kurva AUC-ROC adalah pengukuran kinerja untuk masalah klasifikasi pada berbagai pengaturan ambang batas. ROC adalah kurva probabilitas dan AUC menunjukkan derajat atau ukuran keterpisahan. Kurva ini menunjukkan seberapa besar model mampu membedakan antara kelas. Semakin tinggi AUC, semakin baik model tersebut dalam memprediksi kelas jinak sebagai jinak dan kelas DDoS sebagai DDoS.

Secara analogi, semakin tinggi skor AUC, semakin baik model tersebut dalam membedakan antara lalu lintas jaringan: apakah itu DDoS atau jinak. Kurva ROC diplot dengan True Positive Rate (TPR) terhadap False Positive Rate (FPR) di mana TPR berada pada sumbu y dan FPR berada pada sumbu x. Berdasarkan hasil ini, dengan membandingkan hasil kurva AUC-ROC (Gbr. 10), kami menemukan bahwa IDS tidak dapat



Gambar 10: Kurva ROC-AUC



Gambar 11: Arsitektur model IDS

untuk mengkategorikan serangan semacam ini, dan kinerja IDS menurun ketika jumlah fitur yang terganggu meningkat.

V.E. Bahasa IndonesiaDITINGKATKANSKEAMANANAKEUNTUNGANPTERGANGGUDDHAIS ATACK

Setelah mengevaluasi model serangan, dan menemukan bahwa model IDS tidak dapat mengkategorikan jenis serangan ini, tujuan kami berikutnya adalah untuk meningkatkan kinerja IDS dalam mendeteksi serangan DDoS semacam ini, terlepas dari apakah ada penggantian atau gangguan pada fitur serangan, atau tidak. Untuk mendeteksinya, kami berhipotesis bahwa GAN dapat digunakan untuk meningkatkan kinerja IDS dengan melatih model baru berdasarkan sampel adversarial yang dihasilkan. Untuk mengatasi kebutuhan ini, kami memperbarui arsitektur IDS sebelumnya (Gbr. 8) sehingga alih-alih menggunakan satu model untuk menentukan apakah lalu lintas input adalah DDoS atau bukan, ia menggunakan dua model yang berbeda. Model pertama mencegah lalu lintas jaringan yang terganggu (dihasilkan) dari mengelabui model IDS, dan lalu lintas jaringan yang sebenarnya kemudian diteruskan ke model kedua. Model kedua memblokir lalu lintas jaringan abnormal (DDoS), dan lalu lintas jinak diteruskan ke sistem atau mesin target seperti yang ditunjukkan Gbr. 11.
Bagian selanjutnya menjelaskan rincian model pertama.

A. Model Diskriminator

Untuk mendeteksi data yang dihasilkan dan meningkatkan keamanan terhadap jenis DDoS, kami membangun model pertama, yaitu diskriminator. Saat membangun diskriminator, kami mempertahankan arsitektur model ini agar sama dengan model IDS sebelumnya. Satu-satunya perbedaan adalah bahwa model ini akan dilatih untuk mendeteksi lalu lintas yang terganggu, dan untuk menghindari kerentanan DL terhadap serangan adversarial, kami melatih model ini berdasarkan kombinasi kumpulan data antara kumpulan data asli dan kumpulan data palsu yang dihasilkan, termasuk semua fitur kecuali yang berkontribusi pada fungsi DDoS. Kumpulan data yang dihasilkan yang digunakan di sini adalah hasil dari model GAN yang sama yang disebutkan sebelumnya.

Setelah membangun model dan di akhir fase pelatihan, kami mengevaluasi model dengan data terganggu yang dihasilkan dari model GAN yang sama. Gambar 12 menunjukkan hasil pengujian kami. Dari 12000 sampel lalu lintas jaringan, model diskriminator mampu mendeteksi 95,4% data dengan benar sebagai palsu, dan 88,2% data dengan benar seperti data sebenarnya.

	Real	Fake
Real	88.2 %	11.8 %
Fake	4.6 %	95.4 %

Gambar 12: Matriks kebingungan lalu lintas jaringan yang diprediksi pada diskriminator (Asli/Palsu)

B. Model deteksi IDS

Setelah membangun dan menguji diskriminator yang mencapai akurasi tinggi sebesar 91,75%, kami memperbarui model kedua yang hanya menangani data nyata yang diteruskan menggunakan diskriminator. Kami melatih LSTM menggunakan kumpulan data asli, yang hanya mencakup fitur fungsional DDoS.

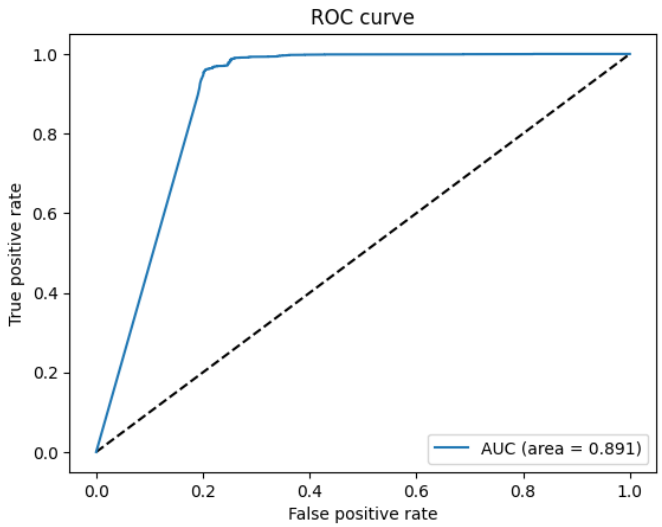
Hasilnya, model ini mampu mendeteksi semua sampel dengan tepat antara yang jinak atau DDoS seperti yang ditunjukkan matriks kebingungan (Gbr. 13).

	Normal	DDoS
Normal	100 %	0 %
DDoS	0%	100 %

Gbr. 13: Matriks kebingungan lalu lintas jaringan yang diprediksi pada IDS (DDoS/Normal)

Akhirnya, dengan menggabungkan kedua model, diskriminator, dan mesin deteksi IDS, kami mengevaluasi kinerja IDS terhadap serangan adversarial yang terganggu. Berdasarkan hasil kurva ROC-AUC untuk kinerja IDS dalam mendeteksi serangan DDoS terlepas dari ada atau tidaknya gangguan, IDS kami yang telah diperbarui mampu mendeteksi, dan memblokir serangan DDoS dengan ketahanan yang lebih baik terhadap sampel DDoS yang terganggu.

(Gbr. 14) menunjukkan antisipasi terhadap penyerang yang dapat menyusun paket dan mengganggu aliran berbahaya untuk meniru karakteristik aliran jinak.

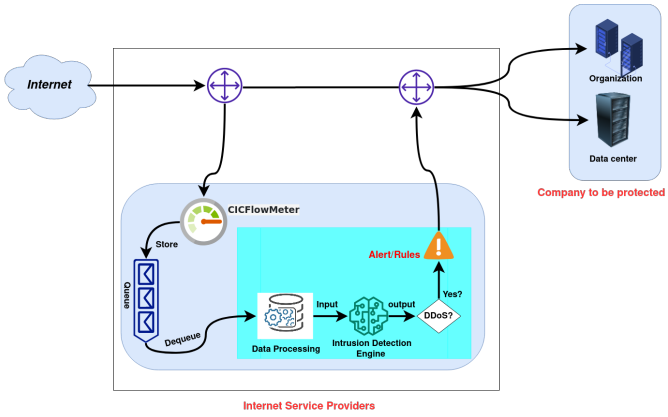


Gbr. 14: Kurva ROC-AUC untuk IDS yang ditingkatkan

KETIGAPEKERJAAN DISAYAINTERNETSLAYANAN PPENYEDIA

Sebagai solusi terhadap serangan DDoS, Penyedia Layanan Internet (ISP) dan perusahaan menggunakan banyak solusi dan Pusat Scrubbing seperti Radware DefensePro, Radware Cloud DDoS Protection Service, Cloudflare DDoS Mitigation Services, Akamai Edge DNS, Arbor Cloud, Oracle Dyn DDoS Protection, Azure DDoS Protection, dll. Semua solusi ini menawarkan fitur-fitur seperti perlindungan berlapis-lapis, deteksi ancaman waktu nyata, pelaporan, dan analitik. Namun, serangan DDoS yang sebenarnya lebih cepat dengan tingkat yang belum pernah terjadi sebelumnya, dan lebih canggih. Botnet lebih terdesentralisasi dan sangat aman. Menggunakan pusat scrubbing cerdas adalah arah baru untuk meningkatkan otomatisasi dan presisi pusat. Oleh karena itu, pusat scrubbing berbasis pembelajaran mesin dianggap sebagai Pusat Scrubbing Generasi Berikutnya (NGSC), menyebarkan IDS berbasis pembelajaran mesin pada ISP memerlukan perencanaan dan pelaksanaan penyebaran IDS berbasis pembelajaran mesin secara hati-hati pada jaringan ISP, untuk memastikan bahwa ia menyediakan keamanan yang efektif dan tidak berdampak negatif pada kinerja jaringan atau mengganggu pengguna. Dalam kasus kami, IDS yang diusulkan mampu mendeteksi lalu lintas langsung secara real-time, beserta pengukur aliran untuk mengekstraksi fitur-fitur yang diperlukan untuk pengambilan keputusan. Dalam pengujian kami, kami menemukan bahwa dengan menggunakan "GPU Tesla V100-PCIE-16GB", kami mampu menganalisis satu paket dalam waktu sekitar 40 ms, termasuk prediksi kedua model dan tanpa fase ekstraksi fitur. Hal ini memungkinkan deteksi ancaman keamanan potensial secara real-time yang efisien dan efektif.
Untuk meniru skenario ISP dalam konteks daya yang dibutuhkan, sistem multi-threading sederhana seharusnya cukup untuk menangani pemrosesan beberapa sampel secara paralel. Ini akan membantu memastikan bahwa IDS dapat memproses dan menganalisis lalu lintas secara real time tanpa penundaan yang signifikan. Seperti yang ditunjukkan pada Gambar 15, IDS yang kami usulkan dapat digunakan dengan cara berikut:

1) Router pertama meneruskan semua paket jaringan ke ISP.
2) CICFlowmeter mengekstrak fitur yang dibutuhkan oleh IDS dan menyimpannya dalam antrean memori.
3) IDS mengeluarkan paket dari antrian dan memprosesnya setiap 40 ms.



Gbr. 15: Penerapan IDS pada ISP

4) Berdasarkan keputusan IDS, sistem peringatan dan aturan terkait diterapkan.

Strategi penerapan ini memungkinkan pemantauan lalu lintas jaringan secara real-time yang efisien dan efektif untuk ancaman keamanan."

VII. BAB VKESIMPULAN DAN PEKERJAAN MASA DEPAN

Deteksi DDoS masih menjadi masalah yang menantang dalam keamanan siber. Baru-baru ini, kami telah menyaksikan peningkatan minat dalam deteksi DDoS menggunakan algoritma pembelajaran mesin (ML) dan pembelajaran mendalam (DL). Ironisnya, meskipun ML/DL dapat meningkatkan akurasi deteksi, keduanya masih dapat dihindari dengan menggunakan teknik ML/DL untuk membuat lalu lintas serangan. Makalah ini membahas aspek-aspek di atas dari teknik deteksi dan antideteksi DDoS berbasis ML. Pertama, kami membangun metode deteksi DDoS berdasarkan model Long Short-Term Memory (LSTM). Skema deteksi menunjukkan tingkat akurasi yang tinggi dalam mendeteksi serangan DDoS dengan akurasi 100%. Kedua, kami menguji teknik yang sama terhadap serangan DDoS adversarial yang dihasilkan menggunakan GAN. Berdasarkan perbandingan kurva ROC dan akurasi berbagai skenario, hasil yang diperoleh menunjukkan penurunan kinerja IDS. Akhirnya, kami menunjukkan cara meningkatkan skema ini untuk mendeteksi serangan DDoS adversarial dengan membangun dua model yang berbeda. Yang pertama digunakan untuk mendeteksi apakah lalu lintas masukan jaringan palsu untuk memblokirnya, jika tidak, meneruskannya ke IDS yang bertanggung jawab untuk mendeteksi apakah itu DDoS atau lalu lintas normal. Hasil eksperimen kami menunjukkan bahwa model deteksi kami efisien dan akurat dalam mengidentifikasi lalu lintas DDoS adversarial yang dihasilkan GAN dengan rasio deteksi berkisar antara 91,75% dan 100%. Sebagai bagian dari pekerjaan kami di masa mendatang, perlu untuk mengevaluasi kinerja IDS kami pada data yang dihasilkan oleh model lain seperti auto-encoder. Selain itu, pekerjaan lebih lanjut diperlukan untuk mempelajari dan menyelidiki penggunaan algoritma pembelajaran daring, yang memungkinkan IDS memperbarui modelnya secara real-time saat memproses data baru. Dengan menggabungkan kemampuan pembaruan inkremental, IDS dapat mempertahankan efektivitasnya bahkan dalam menghadapi metode serangan yang terus berkembang.

VIII. SebuahLAMPIRAN

Simbol-simbol yang digunakan dalam persamaan dijelaskan pada Tabel III

RReferensi

[1] "Ddos statistik laporan untuk 1hari "Tahun 2022." https://blog.nexusguard.com/threat-report/ddos-statisticalreport-for-1hy-2022, 2022.
[2] A. Shield, "Laporan lanskap ancaman – q1 2020." https://awsshield-tlr.s3.amazonaws.com/2020-Q1-AWS-Shield-TLR.pdf, 2020.

TABEL III: Notasi dan definisi:

Simbol	Definisi
X	Data nyata
dari	Vektor laten
$D()$	Evaluasi diskriminator terhadap data nyata atau palsu
$C()$	Evaluasi kritikus terhadap data asli atau palsu
$G()$	Evaluasi generator terhadap data asli atau palsu
$P_{data}(X)$	Distribusi data pada sampel asli x
$P_{dari}(dari)$	Distribusi data atas sampel palsu z
$_{Bahasa\ Inggris x}$	Nilai yang diharapkan atas data asli
$_{Bahasa\ Inggris dari}$	Nilai yang diharapkan dari input data acak ke Generator
$_{Sepakbola}$	Lapisan yang terhubung sepenuhnya
H	Lapisan tersembunyi

[3] Spamhaus, "Rumah Spam" jaringan bot ancaman ke atas-tanggal." https://www.spamhaus.com/custom-konten/unggahan/2022/07/2022-Q2-Pembaruan-Ancaman-Botnet.pdf, 2022.

[4] biji.

[5] CyberSecurity dan ISA (CISA), "Tips keamanan (st04-015). memahami serangan penolakan layanan." Online, November 2019.

[6] MS Elsayed, N.-A. Le-Khac, S. Dev, dan AD Jurcut, "Ddosnet: Model pembelajaran mendalam untuk mendeteksi serangan jaringan," *Simposium Internasional IEEE ke-21 tahun 2020 dengan tema "Dunia Jaringan Nirkabel, Seluler, dan Multimedia" (WoWMoM)*, hlm. 391–396, 2020.

[7] L. Yong dan Z. Bo, "Model deteksi intrusi berdasarkan multi-skala cnn," dalam *Konferensi Teknologi Informasi, Jaringan, Elektronik, dan Kontrol Otomasi (ITNEC) IEEE ke-3 tahun 2019*, hlm. 214–218, 2019.

[8] Y. Jia, F. Zhong, A. Alrawais, B. Gong, dan X. Cheng, "Flowguard: Mekanisme pertahanan tepi cerdas terhadap serangan iot ddos," *Jurnal IEEE tentang Internet of Things*, vol. 7, no. 10, hlm. 9552–9562, 2020.

[9] Y.-Y. Zhu, J.-F. Wu, dan Z. Ming, "Penelitian tentang deteksi intrusi berdasarkan peristiwa jaringan dan analisis protokol mendalam," *Jurnal Institut Komunikasi Tiongkok*, vol. 32, no. 8, hal. 171–178, 2011.

[10] KO Adefemi Alimi, K.Ouahada, AM Abu-Mahfouz, S. Rimer, dan OA Alimi, "Menyempurnakan deteksi intrusi berbasis lstm untuk serangan penolakan layanan di internet of things," *Jurnal Jaringan Sensor dan Aktuator*, jilid. 11, tidak. 3, hal. 32 Agustus 2022.

[11] D. Akgun, S. Hizal, dan U. Cavusoglu, "Model deteksi intrusi serangan ddos baru berdasarkan pembelajaran mendalam untuk keamanan siber," *Komputer & Keamanan*, vol. 118, hal. 102748, 2022.

[12] X. Yuan, C. Li, dan X. Li, "Deepdefense: mengidentifikasi serangan ddos melalui pembelajaran mendalam," dalam *Konferensi internasional IEEE 2017 tentang komputasi cerdas (SMARTCOMP)*, hlm. 1–8, IEEE, 2017.

[13] S. Sambangi dan L. Gondi, "Pendekatan pembelajaran mesin untuk deteksi serangan ddos (penolakan layanan terdistribusi) menggunakan regresi linier berganda," *Prosiding Institut Penerbitan Digital Multidisiplin*, jilid. 63, tidak. 1, hal. 51, 2020.

[14] G. Usha, M. Narang, dan A. Kumar, "Deteksi dan klasifikasi serangan dos terdistribusi menggunakan pembelajaran mesin," dalam *Jaringan Komputer dan Teknologi Komunikasi Inovatif*, hal. 985–1000, Springer, 2021.

[15] C. Yinka-Banjo dan O.-A. Ugot, "Tinjauan jaringan adversarial generatif dan penerapannya dalam keamanan siber," *Ulasan Kecerdasan Buatan*, vol. 53, no. 3, hlm. 1721–1736, 2020.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, dan Y. Bengio, "Jaringan adversarial generatif," *Kemajuan dalam sistem pemrosesan informasi saraf*, jilid 27, 2014.

[17] K. Liu, H. Yang, Y. Ma, B. Tan, B. Yu, EFY Young, R. Karri, dan S. Garg, "Serangan gangguan adversarial pada cad berbasis ml: Studi kasus pada deteksi hotspot litografi berbasis cnn," *ACM Trans. Des. Sistem Elektron Otomatis.*, vol. 25, Agustus 2020.

[18] O. Depren, M. Topallar, E. Anarim, dan MK Ciliz, "Sistem deteksi intrusi (ids) cerdas untuk deteksi anomali dan penyalahgunaan dalam jaringan komputer," *Sistem Pakar dengan Aplikasi*, vol. 29, no. 4, hal. 713–722, 2005.

[19] KS Sahoo, A. Iqbal, P. Maiti, dan B. Sahoo, "Pendekatan pembelajaran mesin untuk memprediksi lalu lintas ddos dalam jaringan yang ditentukan perangkat lunak," dalam *Konferensi Internasional tentang Teknologi Informasi (ICIT) 2018*, hlm. 199–203, IEEE, 2018.

[20] J. Mirkovic dan P. Reiher, "Taksonomi serangan ddos dan mekanisme pertahanan ddos," *Tinjauan Komunikasi Komputer ACM SIGCOMM*, vol. 34, no. 2, hal. 39–53, 2004.

[21] I. Sharafaldin, AH Lashkari, dan AA Ghorbani, "Menuju pembuatan dataset deteksi intrusi baru dan karakterisasi lalu lintas intrusi.," *ICISS*, vol. 1, hal. 108–116, 2018.

[22] I. Sharafaldin, AH Lashkari, S. Hakak, dan AA Ghorbani, "Mengembangkan kumpulan data dan taksonomi serangan penolakan layanan terdistribusi (ddos) yang realistis," dalam *Konferensi Internasional Carnahan tentang Teknologi Keamanan (ICCST) 2019*, hlm. 1–8, IEEE, 2019.

[23] M. Tavallaee, E. Bagheri, W. Lu, dan AA Ghorbani, "Analisis terperinci dari kumpulan data kdd cup 99," di *Simposium IEEE 2009 tentang kecerdasan komputasional untuk aplikasi keamanan dan pertahanan*, hal. 1–6, Ieee, 2009.

[24] Q. Yan, M. Wang, W. Huang, X. Luo, dan FR Yu, "Secara otomatis mensintesis jejak serangan dos menggunakan jaringan adversarial generatif," *Jurnal Internasional Pembelajaran Mesin dan Sibernetika*, vol. 10, no. 12, hlm. 3387–3396, 2019.

[25] J. Aiken dan S. Scott-Hayward, "Menyelidiki serangan adversarial terhadap sistem deteksi intrusi jaringan di SDNS," di dalam *Konferensi IEEE 2019 tentang Virtualisasi Fungsi Jaringan dan Jaringan Terdefinisi Perangkat Lunak (NFV-SDN)*, hlm. 1–7, IEEE, 2019.

[26] M. Shahpasand, L. Hamey, D. Vatsalan, dan M. Xue, "Serangan adversarial pada deteksi malware seluler," dalam *Lokakarya Internasional IEEE ke-1 tentang Kecerdasan Buatan untuk Seluler (AI4Mobile) tahun 2019*, hlm. 17–20, IEEE, 2019.

[27] J. Zhang, Q. Yan, dan M. Wang, "Serangan penghindaran berdasarkan jaringan adversarial generatif wasserstein," dalam *Komputasi, Komunikasi, dan Aplikasi IoT (ComComAp) 2019*, hlm. 454–459, IEEE, 2019.

[28] M. Tavallaee, E. Bagheri, W. Lu, dan AA Ghorbani, "Analisis terperinci dari kumpulan data kdd cup 99," di *Simposium IEEE 2009 tentang kecerdasan komputasional untuk aplikasi keamanan dan pertahanan*, hal. 1–6, Ieee, 2009.

[29] M. Arjovsky, S. Chintala, dan L. Bottou, "Jaringan adversarial generatif Wasserstein," dalam *Konferensi internasional tentang pembelajaran mesin*, hlm. 214–223, PMLR, 2017.

[30] D. Arp, M. Spreitzenbarth, M. Hubner, H. Gascon, K. Rieck, dan C. Siemens, "Drebin: Deteksi malware android yang efektif dan dapat dijelaskan di saku Anda.," dalam *Tidak ada komentar*, vol. 14, hal. 23–26, 2014.

[31] S. Chenette, "Arsip dokumen berbahaya untuk pengujian tanda tangan dan pembuangan malware penelitian-penularan," 2011.

[32] C. Zhang, X. Costa-Pérez, dan P. Patras, "Tiki-taka: Menyerang dan mempertahankan sistem deteksi intrusi berbasis pembelajaran mendalam," dalam *Prosiding Lokakarya Keamanan Komputasi Awan ACM SIGSAC 2020*, hlm. 27–39, 2020.

[33] I. Sharafaldin, AH Lashkari, dan AA Ghorbani, "Menuju

menghasilkan kumpulan data deteksi intrusi baru dan karakterisasi lalu lintas intrusi,"*ICISS*, vol. 1, hal. 108–116, 2018.
[34] M. Abdelaty, S. Scott-Hayward, R. Doriguzzi-Corin, dan D. Siracusa, "Gadot: Pelatihan adversarial berbasis Gan untuk deteksi serangan ddos yang kuat," dalam*Konferensi IEEE tentang Komunikasi dan Keamanan Jaringan (CNS) 2021*, hlm. 119–127, 2021.
[35] SM Lundberg dan S.-I. Lee, "Pendekatan terpadu untuk menafsirkan prediksi model,"*Kemajuan dalam sistem pemrosesan informasi saraf*, jilid 30, 2017.
[36] C. Moreira, Y.-L. Chou, M. Velmurugan, C. Ouyang, R. Sindhgatta, dan P. Bruza, "Linda-bn: Pendekatan probabilistik yang dapat ditafsirkan untuk demistifikasi model prediktif kotak hitam,"*Sistem Pendukung Keputusan*, vol. 150, hal. 113561, 2021.
[37] L. Merrick dan A. Taly, "Permainan penjelasan: Menjelaskan model pembelajaran mesin menggunakan nilai shapley," dalam*Konferensi Lintas Domain Internasional untuk Pembelajaran Mesin dan Ekstraksi Pengetahuan*, hal.17–38, Springer, 2020.
[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, dan Y. Bengio, "Jaringan adversarial generatif,"*Kemajuan dalam sistem pemrosesan informasi saraf*, jilid 27, 2014.
[39] Z. Zhang dan M. Li, "Jun yu. tentang konvergensi dan keruntuhan mode gan,"*Ringkasan Teknis SIGGRAPH Asia 2018*, hal. 21, 2018.
[40] H. Xie, K. Lv, dan C. Hu, "Metode efektif untuk menghasilkan data serangan simulasi berdasarkan jaringan adversarial generatif," dalam *Konferensi Internasional IEEE ke-17 tentang Kepercayaan, Keamanan, dan Privasi dalam Komputasi dan Komunikasi/Konferensi Internasional IEEE ke-12 tentang Sains dan Rekayasa Big Data (TrustCom/BigDataSE) tahun 2018*, hlm. 1777–1784, IEEE, 2018.
[41] Z. Zhou, J. Liang, Y. Song, L. Yu, H. Wang, W. Zhang, Y. Yu, dan Z. Zhang, "Jaringan adversarial generatif Lipschitz," dalam *Konferensi Internasional tentang Pembelajaran Mesin*, hlm. 7584–7593, PMLR, 2019.
[42] AH Lashkari, A. Seo, GD Gil, dan A. Ghorbani, "Cicab: Pemblokir iklan online untuk browser," dalam*Konferensi Internasional Carnahan tentang Teknologi Keamanan (ICCST) 2017*, hlm. 1– 7, IEEE, 2017.
[43] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, dan AC Courville, "Peningkatan pelatihan wasserstein gans,"*Kemajuan dalam sistem pemrosesan informasi saraf*, jilid 30, 2017.
[44] X. Zhang, Y. Zou, dan W. Shi, "Jaringan saraf konvolusi dilatasi dengan leakyrelu untuk klasifikasi suara lingkungan," di dalam*Konferensi internasional ke-22 tentang pemrosesan sinyal digital (DSP) tahun 2017*, hlm. 1–5, IEEE, 2017.
[45] DP Kingma dan J. Ba, "Adam: Sebuah metode untuk optimasi stokastik," *Koroner*Jurnal Ilmu Kebidanan, vol.abs/1412.6980, 2015.
[46] PH Swain dan H. Hauska, "Pengklasifikasi pohon keputusan: Desain dan potensi,"*Transaksi IEEE pada Elektronika Geosains*, vol. 15, no. 3, hal. 142–147, 1977.
[47] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K.Chen,*dan lain-lain.*, "Xgboost: peningkatan gradien ekstrem,"*Paket R versi 0.4-2*, vol. 1, no. 4, hal. 1–4, 2015.
[48] T. Windeatt, "Akurasi/keanekaragaman dan desain pengklasifikasi mlp ensemble,"*Transaksi IEEE pada Jaringan Saraf*, vol. 17, no. 5, hal. 1194–1211, 2006.
[49] M. Pal, "Pengklasifikasi hutan acak untuk klasifikasi penginderaan jauh,"*Jurnal internasional penginderaan jauh*, vol. 26, no. 1, hal. 217–222, 2005.
[50] S. Hochreiter dan J. Schmidhuber, "Memori jangka pendek panjang," *Komputasi saraf*, vol. 9, no. 8, hal. 1735–1780, 1997.