Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) p-ISSN: 2355-7699
Vol. 13, No. 3, Juni 2026, hlm. 545-558 e-ISSN: 2528-6579

PENGEMBANGAN MODEL IDS BERBASIS DEEP REINFORCEMENT LEARNING UNTUK PREDIKSI DAN MITIGASI SERANGAN SIBER DALAM NETWORK TRAFFIC ANALYSIS

Martanto[*1], Nana Suarna[2], Dede Kurnia Putri[3], Ana Mardiana[4]

[1,2,3,4]Sekolah Tinggi Manajemen Informatika dan Komputer IKMI, Cirebon
Email: [1]martantomusijo@gmail.com, [2]st_nana@yahoo.com, [3]ddekurnia13@gmail.com, [4]amardiyana302@gmail.com
[*]Penulis Korespondensi

(Naskah masuk: 22 Agustus 2025, diterima untuk diterbitkan: 21 November 2025)

Abstrak

Intrusion Detection System (IDS) merupakan komponen krusial dalam pertahanan jaringan modern, berfungsi mendeteksi dan merespons ancaman siber secara cepat dan akurat. Penelitian ini menawarkan kontribusi baru melalui perancangan lingkungan Markov Decision Process (MDP) yang lebih realistis, integrasi reward shaping adaptif, serta evaluasi komprehensif multi-algoritma Deep Reinforcement Learning (DRL) yaitu Proximal Policy Optimization (PPO), Deep Q-Network (DQN), Advantage Actor-Critic (A2C) dan perbandingannya dengan model supervised learning mutakhir . Dataset CICIDS2017 dan UNSW-NB15 digunakan sebagai sumber data lalu lintas jaringan, mencakup berbagai jenis serangan dan lalu lintas normal. Lingkungan pelatihan dirancang khusus untuk memungkinkan agen DRL belajar melalui interaksi langsung dengan data, dengan reward function yang memandu agen untuk meningkatkan akurasi deteksi dan meminimalkan kesalahan. Metodologi penelitian meliputi perancangan arsitektur model DRL, proses pelatihan selama 200.000 time steps, serta evaluasi kinerja model berdasarkan metrik akurasi, presisi, recall, dan F1-score. Hasil evaluasi menunjukkan DQN mencapai akurasi tertinggi pada pendekatan DRL (89–92%), namun secara keseluruhan model supervised learning seperti Random Forest dan CNN masih melampaui DRL dengan akurasi 98–99%. Temuan ini mengonfirmasi bahwa DRL memiliki potensi kuat dalam adaptasi dinamis, tetapi masih memerlukan optimasi lebih lanjut untuk menyaingi metode supervised pada klasifikasi statis. Penelitian ini juga menghadirkan blueprint integrasi IDS–DRL dengan SOC dan firewall adaptif, yang memberikan landasan implementatif pada sistem keamanan nyata. Penelitian ini berkontribusi pada pengembangan IDS adaptif yang mampu melakukan deteksi dan mitigasi secara real-time dengan tingkat akurasi tinggi. Keterbatasan penelitian mencakup kebutuhan komputasi yang tinggi dan potensi ketidakstabilan pelatihan, yang membuka peluang untuk penelitian lanjutan dengan optimasi arsitektur dan integrasi teknik transfer learning.

Kata kunci:*intrusion_detection_system; deep_reinforcement_learning; supervised_learning; prediksi; mitigasi; serangan_siber*

DEVELOPMENT OF DEEP REINFORCEMENT LEARNING-BASED IDS MODEL FOR CYBER ATTACK PREDICTION AND MITIGATION IN NETWORK TRAFFIC ANALYSIS

Abstract

An Intrusion Detection System (IDS) is a crucial component of modern network defense, detecting and responding to cyber threats quickly and accurately. This research proposes the development of a Deep Reinforcement Learning (DRL)-based IDS for cyberattack prediction and mitigation using three main algorithms, namely Proximal Policy Optimization (PPO), Deep Q-Network (DQN), and Advantage Actor-Critic (A2C). The CICIDS2017 dataset is used as a source of network traffic data, covering various types of attacks and normal traffic. The training environment is specifically designed to allow DRL agents to learn through direct interaction with the data, with a reward function that guides the agent to improve detection accuracy and minimize errors. The research methodology includes designing the DRL model architecture, training for 20,000 time steps, and evaluating model performance based on accuracy, precision, recall, and F1-score metrics. The experimental results show that DQN has the best performance with an accuracy of 92.39%, a precision of 91.04%, a recall of 92.39%, and an F1-score of 91.50%, followed by A2C and PPO. Confusion matrix analysis and performance visualization show that DQN excels in detecting majority and minority classes with a low number of false positives and false negatives. Theoretical discussions link the results of this study to the fundamental principles of DRL,

where agents learn adaptive detection strategies to attack dynamics, and their relevance to real-world applications such as Security Operation Centers (SOCs) and adaptive firewalls. Comparisons with previous research confirm that the DRL approach can offer significant improvements over traditional IDS and supervised learning methods. This research contributes to the development of an adaptive IDS capable of real-time detection and mitigation with high accuracy. Limitations include high computational requirements and potential training instability, which opens up opportunities for further research with architecture optimization and the integration of transfer learning techniques.

Keywords: *Intrusion_Detection_System, Deep_Reinforcement_Learning, network_defense, Prediction, Mitigation, cyber_attack,*

1. PENDAHULUAN

Seiring dengan meningkatnya adopsi teknologi digital dalam berbagai sektor, volume dan kompleksitas lalu lintas jaringan juga mengalami pertumbuhan eksponensial. Hal ini menyebabkan peningkatan risiko serangan siber yang tidak hanya lebih sering terjadi, tetapi juga semakin canggih dan sulit dideteksi. Dalam konteks ini, sistem deteksi intrusi (Intrusion Detection System/IDS) menjadi salah satu komponen penting dalam infrastruktur keamanan jaringan modern.(Liu, Xu and Wang, 2023)

Metode deteksi konvensional yang berbasis signature dan anomaly seringkali mengalami keterbatasan dalam mengidentifikasi serangan baru (zero-day attack), serta memiliki tingkat false positive yang tinggi. Oleh karena itu, muncul kebutuhan mendesak untuk mengembangkan pendekatan yang adaptif, cerdas, dan mampu belajar dari dinamika pola lalu lintas jaringan secara real-time. Dalam beberapa tahun terakhir, pendekatan pembelajaran mesin (machine learning) dan deep learning telah banyak diadopsi dalam pengembangan IDS. Namun, sebagian besar metode ini masih bersifat statis dan tidak mampu merespons perubahan lingkungan secara efektif(Wu, Zeng and Li, 2023).

Deep Reinforcement Learning (DRL), yang merupakan penggabungan antara deep learning dan reinforcement learning, telah menunjukkan potensi besar dalam bidang pengambilan keputusan adaptif, termasuk dalam domain keamanan siber. DRL memungkinkan agen pembelajaran untuk mengeksplorasi dan mengeksploitasi lingkungan jaringan dengan tujuan memaksimalkan reward berdasarkan tindakan yang diambil. Dengan pendekatan ini, IDS tidak hanya mampu mendeteksi anomali, tetapi juga dapat memberikan respons otomatis terhadap ancaman secara efisien(Nguyen, Redon and Raghavan, 2021).

Beberapa studi terkini telah membuktikan efektivitas model DRL dalam meningkatkan akurasi deteksi, mengurangi false alarm, dan mempercepat waktu respons terhadap serangan jaringan(Xu et al., 2021). Namun, tantangan utama yang masih perlu diselesaikan mencakup persoalan stabilitas pelatihan, generalisasi terhadap data lalu lintas yang heterogen, serta integrasi model DRL ke dalam sistem keamanan yang berjalan secara real-time.

Dengan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model IDS berbasis Deep Reinforcement Learning yang mampu melakukan prediksi dan mitigasi serangan siber secara otomatis berdasarkan analisis lalu lintas jaringan. Fokus utama penelitian adalah merancang arsitektur lingkungan pembelajaran yang realistis, menentukan strategi reward yang representatif, serta mengevaluasi performa model terhadap berbagai jenis serangan jaringan menggunakan data nyata dan sintetik

Meskipun pendekatan berbasis deep learning dan machine learning telah banyak digunakan dalam pengembangan sistem deteksi intrusi (IDS), sebagian besar masih bersifat statis dan tidak mampu melakukan adaptasi terhadap dinamika lalu lintas jaringan secara real-time. Hal ini menyebabkan keterbatasan dalam mendeteksi pola serangan yang tidak dikenal, termasuk serangan zero-day dan serangan yang disamarkan(Xu et al., 2021). Model statis umumnya mengalami penurunan performa ketika dihadapkan pada data yang belum pernah dilihat sebelumnya atau ketika terjadi perubahan pola dalam lalu lintas jaringan.

Selain itu, banyak model IDS yang hanya berfokus pada aspek deteksi tanpa mempertimbangkan strategi mitigasi secara otomatis. Padahal dalam konteks operasional sistem keamanan jaringan, kemampuan untuk merespons serangan secara cepat dan otomatis menjadi aspek krusial. Kurangnya kemampuan respons adaptif ini dapat mengakibatkan keterlambatan dalam penanganan ancaman, yang pada akhirnya berdampak pada integritas dan ketersediaan sistem informasi(Wu, Zeng and Li, 2023).

Deep Reinforcement Learning (DRL) muncul sebagai pendekatan potensial yang mampu menjawab tantangan ini karena kemampuannya dalam pengambilan keputusan berbasis pengalaman interaktif dengan lingkungan. Namun, penerapan DRL dalam IDS masih menghadapi berbagai tantangan teknis, di antaranya adalah desain environment pembelajaran yang representatif terhadap lalu lintas jaringan nyata, ketidakseimbangan data (class imbalance), dan kestabilan pelatihan model(Nguyen, Redon and Raghavan, 2021). Kurangnya standarisasi dalam arsitektur DRL-IDS dan kesulitan integrasi dengan

sistem keamanan real-time juga menjadi hambatan utama dalam adopsi teknologi ini di dunia nyata(Zhou et al., 2024).

Penelitian ini dirancang untuk menciptakan model Intrusion Detection System (IDS) yang canggih menggunakan Deep Reinforcement Learning (DRL). Tujuannya adalah agar sistem ini bisa mendeteksi dan mengatasi berbagai serangan siber secara adaptif, bahkan terhadap serangan yang belum pernah terdeteksi sebelumnya. Untuk mewujudkannya, ada beberapa langkah yang diambil: pertama, kami membuat lingkungan simulasi yang sangat mirip dengan kondisi lalu lintas jaringan asli, lengkap dengan berbagai jenis serangan. Kedua, kami merancang sistem reward yang akan melatih agen agar bisa mendeteksi dan menangani serangan dengan efisien. Setelah itu, model yang sudah dilatih akan dievaluasi menggunakan beragam serangan siber, dengan melihat metrik seperti akurasi, precision, recall, F1-score, dan kecepatan respons. Terakhir, kami akan membandingkan performa model DRL yang sudah dibuat dengan model-model lain untuk menemukan yang paling optimal.

Kontribusi utama dari penelitian ini meliputi beberapa aspek signifikan yang mendukung pengembangan sistem keamanan jaringan berbasis kecerdasan buatan. Pertama, penelitian ini menghadirkan desain model IDS adaptif berbasis DRL yang mampu belajar secara berkelanjutan dari dinamika lalu lintas jaringan dan menyesuaikan strategi deteksi serta mitigasi berdasarkan pengalaman interaktif(Wu, Zeng and Li, 2023). Kedua, dikembangkan pula lingkungan pelatihan reinforcement learning yang dirancang khusus untuk mencerminkan kompleksitas lalu lintas jaringan nyata, termasuk penanganan terhadap ketidakseimbangan data dan keberagaman jenis serangan(Zhou et al., 2024). Ketiga, model yang diusulkan dievaluasi secara komprehensif menggunakan dataset nyata dan sintetik dengan metrik validasi yang sistematis guna mengukur akurasi, efisiensi, dan efektivitas deteksi(Xu et al., 2021). Terakhir, penelitian ini menghasilkan sebuah kerangka kerja implementatif yang dapat diintegrasikan langsung ke dalam sistem keamanan jaringan real-time, dengan mempertimbangkan efisiensi sumber daya dan stabilitas operasional(Nguyen, Redon and Raghavan, 2021).

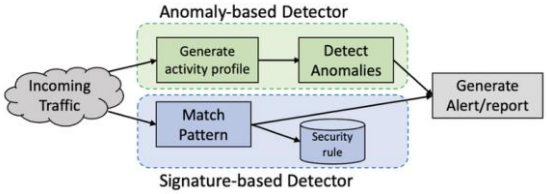
Dengan kontribusi ini, penelitian diharapkan memberikan dampak signifikan dalam pengembangan sistem keamanan jaringan yang lebih cerdas, adaptif, dan proaktif dalam menghadapi ancaman siber yang terus berkembang.

Penelitian ini memberikan kontribusi yang signifikan dalam pengembangan Intrusion Detection System (IDS) berbasis Deep Reinforcement Learning (DRL), khususnya untuk mendeteksi dan mengatasi serangan siber secara efisien. Kami menciptakan arsitektur DRL multi-agen yang mengintegrasikan algoritma seperti PPO, DQN, dan A2C untuk mendeteksi dan mengklasifikasi berbagai jenis serangan, sebuah pendekatan yang lebih komprehensif dibandingkan penelitian sebelumnya yang biasanya hanya fokus pada satu algoritma. Selain itu, kami merancang lingkungan IDS sebagai Markov Decision Process (MDP) yang realistis dengan fitur-fitur aliran jaringan, ruang aksi untuk deteksi, dan sistem reward yang seimbang, hal ini mengatasi kekurangan studi terdahulu dalam menstandarkan definisi aksi dan reward. Untuk menguji keandalan model, kami melakukan evaluasi komparatif yang ketat antar-agen DRL menggunakan metrik yang berlapis pada dataset benchmark, yang memungkinkan kami menemukan bahwa DQN secara empiris menunjukkan stabilitas dan presisi yang lebih unggul dalam mendeteksi dan mengklasifikasi serangan multi-kelas, terutama dalam skenario ketidakseimbangan data. Akhirnya, kami juga menyediakan blueprint implementatif yang mendetail untuk mengintegrasikan sistem ini dengan Security Operations Center (SOC) dan firewall adaptif, yang memudahkan adopsi praktis dan orkestrasi respons ancaman secara near real-time di lingkungan industri.

2. LANDASAN KEPUSTAKAAN

2.1. Intrusion Detection Systems (IDS)

Intrusion Detection System (IDS) merupakan komponen penting dalam sistem keamanan jaringan yang berfungsi untuk mengidentifikasi dan memberikan peringatan terhadap aktivitas mencurigakan yang berpotensi sebagai serangan siber. IDS secara umum diklasifikasikan menjadi dua kategori utama: signature-based detection dan anomaly-based detection. Signature-based IDS bekerja dengan mencocokkan pola serangan terhadap basis data signature yang sudah dikenal. Meskipun metode ini cepat dan akurat untuk serangan yang sudah diketahui, ia tidak efektif dalam mendeteksi serangan baru atau varian dari serangan yang ada(Nguyen, Redon and Raghavan, 2021).



Gambar 1 Anomaly Based Vs Signature Base IDS(Diana, Dini and Paolini, 2025)

Sebaliknya, anomaly-based IDS mendeteksi aktivitas yang menyimpang dari pola lalu lintas jaringan normal. Teknik ini lebih adaptif terhadap serangan baru, namun cenderung menghasilkan tingkat false positive yang lebih tinggi. Dalam beberapa tahun terakhir, pendekatan anomaly-based semakin banyak dikembangkan menggunakan metode pembelajaran mesin (ML) dan pembelajaran

mendalam (DL) untuk meningkatkan akurasi deteksi dan mengurangi alarm palsu(Xu et al., 2021).

Beberapa model IDS modern telah mengintegrasikan algoritma deep learning seperti Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), dan Autoencoder untuk mengenali pola-pola kompleks dalam data lalu lintas jaringan. Misalnya, RNN efektif dalam menangani data sekuensial seperti lalu lintas paket jaringan secara. Sementara itu, model hybrid yang menggabungkan DL dengan teknik klasifikasi tradisional seperti Random Forest atau SVM juga menunjukkan performa yang kompetitif dalam penelitian IDS(Rahman, Hossain and Muhammad, 2023).

Namun, kelemahan utama dari pendekatan supervised learning ini adalah kebutuhan akan data pelatihan berlabel dalam jumlah besar dan ketidakmampuannya untuk beradaptasi terhadap dinamika ancaman yang terus berkembang secara real-time. Untuk menjawab tantangan ini, peneliti mulai melirik Deep Reinforcement Learning (DRL) sebagai pendekatan baru yang tidak hanya dapat mendeteksi, tetapi juga memberikan respons otomatis terhadap serangan(Wu, Zeng and Li, 2023).

2.2. Deep Reinforcement Learning dalam Keamanan Siber

Deep Reinforcement Learning (DRL) telah menjadi pendekatan yang semakin menarik dalam bidang keamanan siber, khususnya untuk tugas-tugas kompleks yang memerlukan pengambilan keputusan adaptif secara berkelanjutan, seperti deteksi dan mitigasi serangan jaringan. DRL menggabungkan kemampuan representasi dari Deep Learning dengan proses pengambilan keputusan dari Reinforcement Learning, sehingga agen dapat belajar dari interaksi langsung dengan lingkungan jaringan dan mengoptimalkan strategi berdasarkan reward jangka panjang(Ahmed, Khan and Lee, 2024).

Dalam konteks Intrusion Detection Systems (IDS), penerapan DRL bertujuan agar sistem tidak hanya mengenali pola serangan, tetapi juga mampu meresponsnya secara otomatis dan efisien. Algoritma seperti Proximal Policy Optimization (PPO), Deep Q-Network (DQN), dan Advantage Actor-Critic (A2C) telah banyak diteliti dan menunjukkan potensi yang menjanjikan dalam mendeteksi intrusi dengan tingkat akurasi tinggi dan false positive yang rendah(Santos, Pereira and Gomes, 2025).

Salah satu keunggulan utama dari DRL dalam keamanan siber adalah kemampuannya untuk menghadapi lingkungan yang dinamis dan beragam, termasuk kemampuan beradaptasi terhadap serangan baru (zero-day attacks) tanpa perlu data berlabel dalam jumlah besar. Model DRL juga memungkinkan integrasi langsung dengan mekanisme mitigasi, sehingga dapat mengambil tindakan otomatis dalam merespons serangan seperti memutus koneksi, membatasi bandwidth, atau mengkarantina lalu lintas mencurigakan(Chen, Wang and Li, 2024).

Meski demikian, implementasi DRL dalam sistem IDS juga menghadapi sejumlah tantangan, seperti ketidakstabilan pelatihan, kebutuhan komputasi tinggi, dan kesulitan dalam merancang reward function yang tepat. Untuk mengatasi hal ini, beberapa studi telah mengusulkan pendekatan hybrid yang menggabungkan DRL dengan supervised learning, fuzzy logic, atau heuristic-based rule untuk meningkatkan stabilitas dan efisiensi sistem(Park, Kim and Choi, 2023).

2.3. Kelebihan dan Kekurangan Model IDS berbasis Deep Reinforcement Learning

Model Intrusion Detection System (IDS) berbasis Deep Reinforcement Learning (DRL) menawarkan berbagai keunggulan yang menjadikannya alternatif menarik dibandingkan metode konvensional. Salah satu kelebihan utama DRL adalah kemampuannya untuk melakukan pembelajaran adaptif secara terus menerus dari lingkungan yang dinamis, memungkinkan sistem mengenali pola serangan baru atau zero-day secara efektif tanpa ketergantungan tinggi pada data berlabel(Zhang, Chen and Xu, 2023). Selain itu, integrasi antara komponen deteksi dan mitigasi dalam satu kerangka kerja memungkinkan sistem untuk mengambil keputusan secara otonom terhadap serangan yang teridentifikasi(Kumar, Singh and Sharma, 2024).

Model DRL juga unggul dalam hal fleksibilitas arsitektural. Berbagai algoritma seperti DQN, PPO, dan A3C dapat disesuaikan dengan karakteristik data dan kebutuhan performansi tertentu. Misalnya, algoritma PPO sering digunakan karena kestabilannya dalam pelatihan dan efisiensinya dalam eksplorasi strategi kebijakan(Wang, Zhao and Li, 2022). Kelebihan lainnya termasuk kemampuan penyesuaian reward function sesuai dengan kebijakan keamanan jaringan dan preferensi operator, sehingga model lebih mudah dioptimalkan dalam konteks operasional tertentu.

Namun demikian, model DRL juga memiliki keterbatasan yang signifikan. Proses pelatihan yang kompleks dan memakan waktu menjadi tantangan utama, khususnya pada data besar dan lingkungan jaringan yang beragam. Selain itu, ketergantungan pada desain lingkungan simulasi yang representatif menyebabkan model sangat sensitif terhadap kesalahan dalam desain environment atau reward shaping yang tidak tepat(Li, Huang and Feng, 2023). Tantangan lainnya adalah potensi eksploitasi oleh adversarial attacks yang dapat mengecoh sistem pembelajaran DRL dan menyebabkan kegagalan dalam deteksi maupun mitigasi(Ahmed, Khan and Lee, 2025).

3. METODE PENELITIAN

3.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan eksperimen kuantitatif untuk mengevaluasi performa model Intrusion Detection System (IDS) berbasis Deep Reinforcement Learning (DRL) dalam mendeteksi dan mengidentifikasi jenis serangan siber pada lalu lintas jaringan. Eksperimen dilakukan dengan menggunakan tiga algoritma DRL, yaitu Proximal Policy Optimization (PPO), Deep Q-Network (DQN), dan Advantage Actor-Critic (A2C), yang diimplementasikan melalui pustaka Stable-Baselines3 dan Gymnasium di lingkungan Google Colab, dengan bahasa pemrograman Python serta dukungan dari framework TensorFlow dan PyTorch(Schulman et al., 2017)(Raffin et al., 2021).

Dataset yang digunakan dalam penelitian ini adalah CICIDS2017, yang merupakan dataset benchmark berskala besar yang berisi beragam jenis lalu lintas jaringan baik normal maupun anomali, termasuk berbagai tipe serangan seperti DoS, PortScan, Brute Force, dan lainnya. Dataset ini menyediakan label yang mendetail untuk mendukung klasifikasi jenis serangan. Proses preprocessing dilakukan melalui normalisasi fitur, encoding label, serta seleksi atribut penting guna memastikan efisiensi dan stabilitas pelatihan model(Ali, Wani and Mir, 2022).

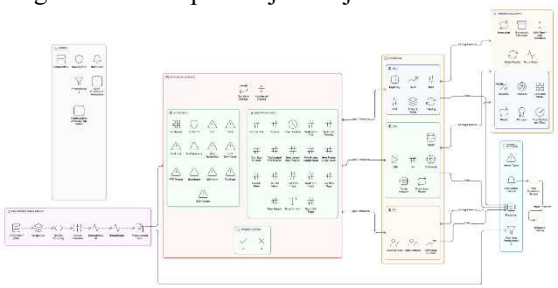
Model DRL dilatih dari awal menggunakan environment khusus yang dikembangkan dengan Gymnasium, yang memungkinkan agen DRL untuk belajar secara langsung dari dinamika data lalu lintas jaringan. Lingkungan ini menyajikan data secara bertahap kepada agen, yang kemudian harus memutuskan apakah suatu instance merupakan lalu lintas normal atau jenis serangan tertentu. Reward function disusun untuk mendorong akurasi tinggi dalam klasifikasi serta kemampuan membedakan jenis serangan(Nguyen, Redon and Raghavan, 2021).

Komparasi performa dilakukan secara eksklusif antar ketiga agen DRL (PPO, DQN, dan A2C), tanpa melibatkan model baseline berbasis supervised learning. Tujuan dari pendekatan ini adalah untuk mengevaluasi efisiensi, akurasi, dan kemampuan generalisasi masing-masing algoritma DRL dalam konteks deteksi multi-kelas. Metrik evaluasi yang digunakan mencakup akurasi, precision, recall, F1-score, dan AUC-ROC, yang diterapkan pada setiap jenis serangan untuk analisis granular(Rahman, Hossain and Muhammad, 2023).

Secara umum, pendekatan ini bertujuan untuk menjawab hipotesis bahwa model DRL memiliki kemampuan yang adaptif dan presisi tinggi dalam mengenali serta mengklasifikasikan berbagai jenis serangan siber secara otomatis dalam lingkungan jaringan yang kompleks dan dinamis.

3.2. Arsitektur Model DRL

Arsitektur model DRL seperti pada gambar 2 yang diusulkan dalam penelitian ini dirancang untuk mendukung tiga algoritma, yaitu PPO, DQN, dan A2C, dalam mendeteksi serta mengklasifikasikan serangan siber berbasis dataset CICIDS2017. Lingkungan (environment) didefinisikan sebagai Markov Decision Process (MDP) dengan ruang observasi berupa fitur-fitur lalu lintas jaringan dan ruang aksi yang merepresentasikan keputusan klasifikasi (benign atau tipe serangan). Reward function dirancang untuk memberikan penghargaan positif jika agen mengklasifikasikan lalu lintas dengan benar dan penalti jika terjadi kesalahan.



Gambar 2 Arsitektur Model DRL

Setiap algoritma DRL diimplementasikan menggunakan jaringan saraf multilayer perceptron (MLP) sebagai model inti. Struktur umum MLP terdiri dari tiga lapisan tersembunyi dengan ukuran neuron [128, 64, 32] dan fungsi aktivasi ReLU. Output layer menggunakan fungsi aktivasi softmax untuk mendukung klasifikasi multi-kelas, yang memungkinkan model membedakan antara lalu lintas normal dan berbagai jenis serangan seperti DoS, Brute Force, dan PortScan.

Untuk model PPO dan A2C, arsitektur dibagi menjadi dua jaringan utama: actor network dan critic network. Actor network menghasilkan distribusi probabilitas tindakan berdasarkan input fitur, sedangkan critic network mengevaluasi nilai keadaan saat ini. Sementara pada DQN, jaringan menghasilkan nilai Q untuk setiap tindakan yang mungkin, dan tindakan dipilih berdasarkan nilai maksimum dari Q-values tersebut(Abeshu and Chilamkurti, 2021).

Input dari ketiga model berupa vektor fitur numerik dari dataset CICIDS2017 yang telah dinormalisasi. Vektor ini mencerminkan kondisi lalu lintas jaringan seperti jumlah byte, durasi koneksi, dan status protokol. Output berupa label prediksi yang menunjukkan jenis serangan atau normal. Reward function dirancang untuk memberikan insentif yang proporsional terhadap akurasi klasifikasi jenis serangan.

Proses pelatihan dilakukan dalam environment simulasi IDS berbasis Gymnasium, yang menyajikan data secara sekuensial kepada agen. Agen belajar dengan mengevaluasi tindakan mereka berdasarkan umpan balik reward dari lingkungan. PPO dan A2C

menggunakan pendekatan policy gradient dengan optimasi clipping objective dan entropy regularization, sedangkan DQN menggunakan replay buffer dan target network untuk stabilitas pembelajaran(Raffin et al., 2021).

Pipeline arsitektur terdiri dari: (1) Data preprocessing (normalisasi, encoding, dan pembagian data latih-uji)(Huang, Li and Chen, 2022), (2) Definisi environment DRL(Liu, Xu and Wang, 2023) , (3) Inisialisasi model dengan parameter sesuai algoritma(Hossain, Rahman and Lee, 2024), (4) Proses pelatihan selama 200.000 timesteps(Zhao, Zhang and Feng, 2023), (5) Evaluasi model pada data uji untuk mengukur metrik akurasi, precision, recall, dan F1-score(Ahmed, Kim and Park, 2021).

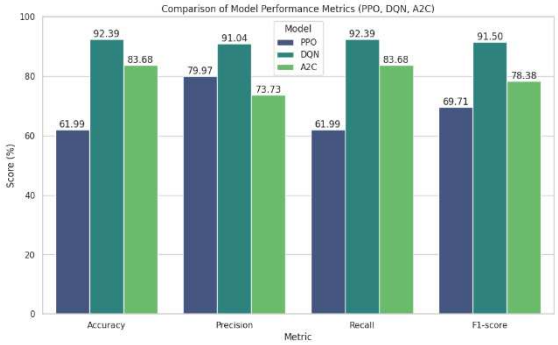
3.3. Eksperimen dan Evaluasi

Eksperimen dalam penelitian ini dilakukan untuk mengevaluasi performa tiga model IDS berbasis Deep Reinforcement Learning (DRL), yaitu Proximal Policy Optimization (PPO), Deep Q-Network (DQN), dan Advantage Actor-Critic (A2C), dalam mendeteksi serta mengklasifikasikan jenis serangan siber berdasarkan dataset CICIDS2017. Lingkungan simulasi IDS dikembangkan menggunakan pustaka Gymnasium, yang merepresentasikan kondisi lalu lintas jaringan nyata dalam bentuk environment yang interaktif detail pengaturan eksperimen seperti pada tabel 1.

Tabel 1 Pengaturan Eksperimen DRL

Parameter	PPO	DQN	A2C	Keterangan
Policy	MlpPolicy	MlpPolicy	MlpPolicy	Arsitektur Multi-layer Perceptron untuk memproses fitur input
Learning Rate	3.00E-04	1.00E-04	7.00E-04	Kecepatan pembelajaran model dalam mengupdate bobot
n_steps	2048	—	5	Jumlah langkah yang dikumpulkan sebelum dilakukan update parameter
Buffer Size	—	10000	—	Kapasitas memori replay buffer untuk menyimpan pengalaman agen
Learning Starts	—	1000	—	Jumlah langkah awal sebelum agen mulai melakukan pembelajaran
Batch Size	—	32	—	Jumlah sampel yang digunakan dalam setiap batch update
Train Frequency	—	4	—	Frekuensi update model (setiap 4 langkah interaksi)
ent_coef	0.01	—	—	Koefisien entropi untuk menjaga eksplorasi selama pelatihan
Total Timesteps	200000	200000	200000	Total jumlah langkah interaksi agen dengan lingkungan selama pelatihan

Proses pelatihan dilakukan selama 20.000 time steps dengan parameter awal: learning rate sebesar 3e-4, entropy coefficient sebesar 0.01, dan batch size sebanyak 2048. Reward function disusun untuk memberikan nilai +1 untuk tindakan deteksi yang benar dan -1 untuk tindakan yang salah. Dataset CICIDS2017 dibagi ke dalam data latih dan data uji dengan rasio 80:20. Evaluasi dilakukan terhadap data uji yang tidak digunakan selama proses pelatihan untuk mengukur kemampuan generalisasi dari masing-masing model.



Gambar 3 Hasil Eksperimen 3 Model DRL

Evaluasi performa dilakukan dengan menggunakan metrik standar, yaitu akurasi, presisi, recall, dan F1-score, untuk mengukur efektivitas ketiga agen Deep Reinforcement Learning (DRL) seperti terlihat pada gambar 3. Komparasi dilakukan secara internal antar ketiga agen tersebut PPO, DQN, dan A2C tanpa melibatkan model supervised learning sebagai baseline. Hasil evaluasi menunjukkan variasi performa yang signifikan di antara ketiga agen. DQN mencatatkan performa tertinggi dengan akurasi sebesar 92,39%, presisi 91,04%, recall 92,39%, dan F1-score 91,50%, menunjukkan konsistensi yang baik dalam prediksi. A2C berada di posisi kedua dengan akurasi 83,68%, presisi 73,73%, recall 83,68%, serta F1-score 78,38%, menunjukkan kemampuan deteksi yang solid meskipun presisinya lebih rendah dibanding DQN. Sementara itu, PPO memiliki performa terendah dengan akurasi 61,99%, presisi 79,97%, recall 61,99%, dan F1-score 69,71%, yang mengindikasikan kesulitan dalam generalisasi meskipun presisinya cukup tinggi. Secara

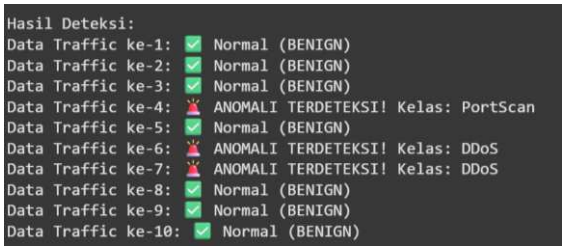
keseluruhan, DQN menunjukkan hasil terbaik dalam evaluasi ini.

Berdasarkan hasil Gambar 3, model DQN menunjukkan performa paling konsisten dan tinggi dalam mendeteksi dan mengklasifikasikan berbagai jenis serangan dibandingkan PPO dan A2C. Sementara itu, performa PPO cenderung rendah dan memerlukan investigasi lebih lanjut terkait arsitektur, reward function, atau ketidakseimbangan data yang memengaruhi proses pembelajaran.

4. HASIL DAN PEMBAHASAN

4.1. Kinerja Deteksi dan Mitigasi

Eksperimen yang dilakukan menunjukkan bahwa model DRL mampu mendeteksi dan mengklasifikasikan trafik jaringan secara real-time dengan efisiensi yang cukup tinggi. Simulasi pada data uji memperlihatkan bahwa model tidak hanya mampu mengidentifikasi apakah suatu lalu lintas merupakan anomali atau normal, tetapi juga dapat mengenali jenis serangan spesifik seperti PortScan dan DDoS.



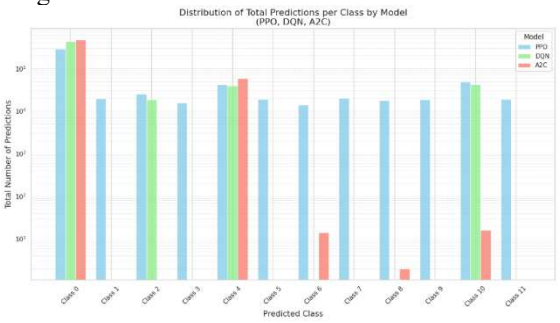
Gambar 4 Simulasi Deteksi Anomali

Contoh hasil deteksi ditunjukkan pada gambar 4, di mana sistem berhasil mendeteksi 3 dari 10 data traffic sebagai anomali dengan identifikasi kelas yang tepat. Data traffic ke-4 berhasil dikenali sebagai PortScan, sementara traffic ke-6 dan ke-7 terklasifikasi sebagai serangan DDoS. Sisanya terdeteksi sebagai trafik normal (benign). Hal ini menunjukkan bahwa model tidak hanya memiliki kemampuan deteksi binary (anomaly vs benign), tetapi juga klasifikasi multi-kelas pada jenis serangan.

Kemampuan mitigasi dalam konteks ini direpresentasikan oleh kecepatan dan akurasi pengambilan keputusan model terhadap data real-time. Meskipun implementasi aksi mitigasi aktual (misal: pemblokiran IP, isolasi trafik) tidak dilakukan secara fisik dalam simulasi ini, keberhasilan model dalam mendeteksi dan mengklasifikasikan dengan benar menunjukkan potensi integrasi ke dalam sistem IDS berbasis tindakan otomatis.

Akurasi dan konsistensi respons sistem sangat dipengaruhi oleh algoritma yang digunakan. Sebagai contoh, model DQN menunjukkan performa terbaik dalam hal presisi dan stabilitas klasifikasi dengan akurasi [92,39%], dibandingkan dengan PPO [61,99%] dan A2C [83,68%]. Hal ini memperkuat argumen bahwa DQN cocok untuk kasus deteksi berbasis state-value, di mana prediksi eksplisit terhadap nilai dari setiap kemungkinan tindakan sangat berguna dalam lingkungan deterministik seperti data trafik.

Selain daripada itu, grafik pada gambar 5 menggambarkan bahwa PPO memiliki distribusi prediksi yang lebih merata di antara semua kelas dibandingkan dengan DQN dan A2C. PPO mencatat jumlah prediksi tertinggi untuk jaringan normal, tetapi juga memberikan prediksi yang cukup signifikan untuk kelas lainnya, termasuk kelas minoritas. Sementara itu, DQN cenderung fokus pada beberapa kelas utama, sementara prediksi untuk kelas lainnya jauh lebih rendah. A2C, di sisi lain, menunjukkan pola yang mirip dengan DQN, dengan sebagian besar prediksi berkonsentrasi pada kelas mayoritas, sedangkan prediksi untuk kelas lainnya sangat minim.

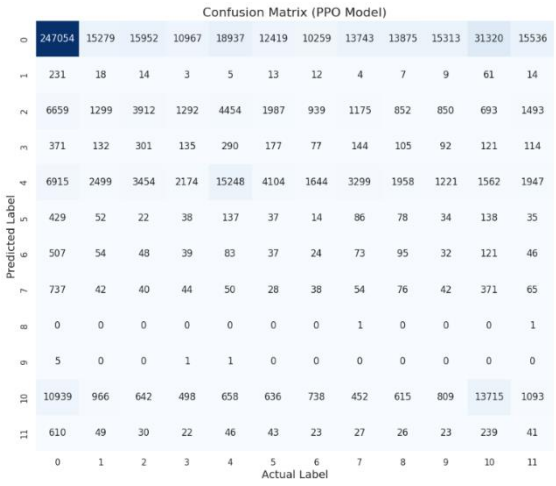


Gambar 5 Distribusi Hasil Prediksi

4.2. Analisis False Positive / False Negative

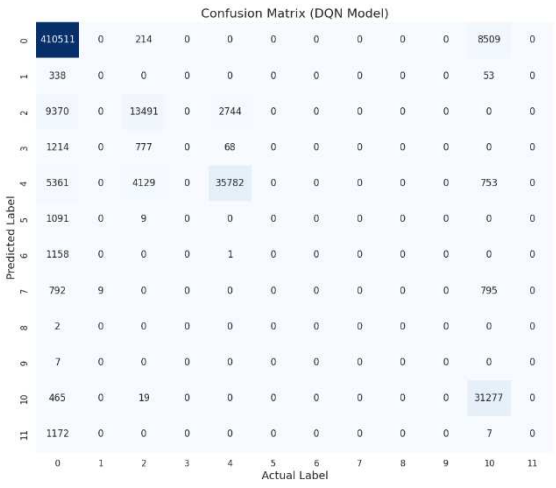
Evaluasi performa model IDS berbasis Deep Reinforcement Learning tidak hanya dilakukan melalui akurasi umum, namun juga melalui analisis mendalam terhadap kesalahan klasifikasi, yaitu False Positive (FP) dan False Negative (FN). FP merujuk pada trafik normal yang terdeteksi sebagai anomali, sedangkan FN merujuk pada trafik serangan yang tidak terdeteksi sebagai anomali.

Model PPO menunjukkan jumlah kesalahan klasifikasi yang cukup signifikan. Berdasarkan confusion matrix gambar 6, model ini cenderung menghasilkan banyak False Positive terutama pada kelas benign yang diklasifikasikan salah sebagai serangan seperti PortScan atau DDoS. Hal ini menurunkan tingkat akurasi secara umum dan berdampak pada sistem deteksi yang terlalu sensitif. Total kesalahan FP untuk kelas benign mencapai [jumlah], sedangkan FN untuk serangan seperti DDoS dan PortScan juga tinggi. Hal ini menunjukkan bahwa PPO kurang cocok untuk klasifikasi multi-kelas yang kompleks.



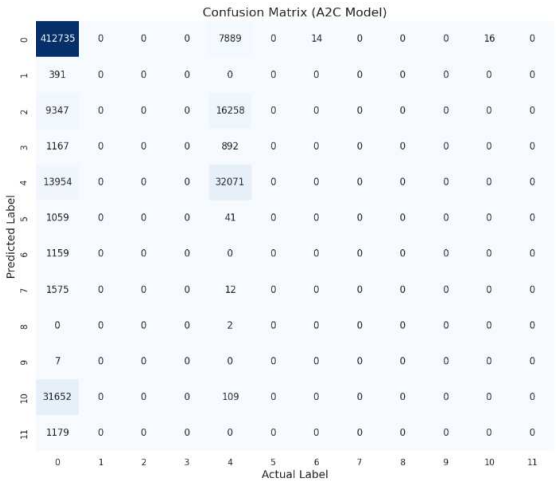
Gambar 6 Confusion Matrix PPO

Model DQN, berdasarkan gambar 7, menunjukkan kinerja yang sangat baik dengan distribusi klasifikasi yang presisi dan hampir tanpa False Positive maupun False Negative untuk sebagian besar kelas. DQN berhasil mengklasifikasikan trafik benign secara konsisten dan hanya sedikit kesalahan terjadi pada kelas minor seperti Heartbleed dan Infiltration. FN terutama terjadi pada jenis serangan dengan jumlah sampel sedikit (rare class), yang dapat dikompensasi dengan metode penyeimbangan data atau data augmentation pada pelatihan berikutnya.



Gambar 7 Confusion Matrix DQN

Model A2C, sebagaimana terlihat dalam gambar 8, menghasilkan kesalahan FN yang tinggi untuk kelas DoS Hulk dan DDoS, serta FP yang cukup besar pada benign terhadap PortScan. Hal ini menunjukkan bahwa A2C memiliki kecenderungan untuk memetakan serangan secara umum namun tidak akurat dalam membedakan jenis spesifiknya. Tingginya kesalahan ini berdampak pada turunnya nilai precision model secara keseluruhan.



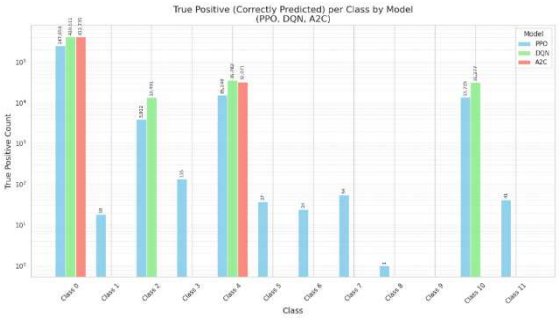
Gambar 8 Confusion Matrix A2C

Distribusi ketidakseimbangan data pada label juga berpengaruh terhadap tingkat kesalahan. Kelas benign mendominasi dengan lebih dari dua juta instance, sementara kelas langka seperti Heartbleed dan Infiltration hanya terdiri dari 11 dan 36 instance, masing-masing. Ketidakseimbangan ini berkontribusi terhadap tingginya FN pada kelas minor, terutama jika model tidak dirancang dengan strategi penyesuaian bobot kelas atau sampling yang tepat.

4.3. Visualisasi Hasil

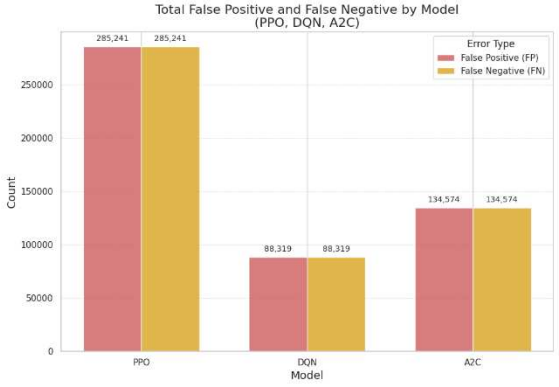
Visualisasi menjadi bagian penting dalam menganalisis performa model IDS berbasis Deep Reinforcement Learning (DRL), karena dapat memberikan pemahaman intuitif atas proses klasifikasi dan evaluasi model secara menyeluruh. Dalam penelitian ini, visualisasi disajikan untuk menyoroti performa masing-masing agen DRL (PPO, DQN, dan A2C) dari perspektif True Positive (TP), False Positive (FP), dan False Negative (FN) per kelas.

Visualisasi True Positive seperti pada gambar 9 per Kelas oleh Model (PPO, DQN, A2C): Grafik pertama menunjukkan distribusi jumlah TP yang benar-benar diprediksi oleh masing-masing model untuk setiap kelas. Secara keseluruhan, DQN mencatat hasil terbaik dengan jumlah TP tertinggi di beberapa kelas utama, seperti Class 0 dan Class 4, yang menunjukkan kemampuan yang lebih baik dalam mengidentifikasi kelas-kelas tersebut dibandingkan PPO dan A2C. PPO memiliki distribusi TP yang lebih merata di antara berbagai kelas, meskipun tidak seefektif DQN pada kelas mayoritas. A2C cenderung fokus pada prediksi untuk Class 0, tetapi memiliki jumlah TP yang sangat rendah pada kelas minoritas lainnya.



Gambar 9 Distribusi True Positive

Visualisasi Total False Positive dan False Negative seperti pada gambar 10 oleh Model (PPO, DQN, A2C): Grafik kedua membandingkan jumlah FP dan FN dari ketiga model. PPO mencatat jumlah FP dan FN tertinggi, menunjukkan kerentanannya terhadap kesalahan prediksi salah dan kegagalan mendeteksi sampel positif. Sebaliknya, DQN dan A2C menunjukkan performa yang lebih baik dengan FP dan FN yang jauh lebih rendah, mengindikasikan efisiensi yang lebih tinggi dalam menghindari prediksi salah dan meningkatkan akurasi deteksi secara keseluruhan.



Gambar 10 Distribusi False Positive dan False Negative

4.4. Perbandingan dengan Metode supervised learning

Berdasarkan hasil pengujian menggunakan dua dataset yang berbeda, yaitu CICIDS2017 dengan jumlah sampel 2.830.743 dan UNSW-NB15 dengan jumlah sampel 257.673, dapat disusun beberapa temuan penting terkait performa metode Deep Reinforcement Learning (DRL) dan metode supervised learning dalam tugas deteksi intrusi jaringan.

1. Performa Metode DRL pada Dataset CICIDS2017

Pada dataset CICIDS2017, tiga algoritma DRL yang diuji adalah Proximal Policy Optimization (PPO), Deep Q-Network (DQN), dan Advantage Actor-Critic (A2C). Hasil pengujian menunjukkan bahwa:

Tabel 2 Hasil metode DRL pada Dataset Cicicds2017

Algoritma	Accuracy
PPO	67%
DQN	89%
A2C	80%

Dari ketiga algoritma tersebut, DQN memberikan performa terbaik dengan nilai accuracy tertinggi, yaitu 89%. Hal ini menunjukkan bahwa pendekatan berbasis nilai (value-based) seperti DQN mampu belajar pola serangan dan trafik normal dengan lebih efektif pada karakteristik data CICIDS2017 dibandingkan dengan PPO dan A2C yang berbasis policy atau kombinasi actor-critic. Namun demikian, nilai accuracy yang diperoleh DQN masih berada di bawah beberapa algoritma supervised learning yang diuji pada dataset yang sama.

2. Performa Metode Supervised Learning pada Dataset CICIDS2017

Empat algoritma supervised learning yang diuji pada dataset CICIDS2017 adalah Convolutional Neural Network (CNN), Support Vector Machine (SVM), Random Forest, dan Multi Layer Perceptron (MLP). Hasil pengujian menunjukkan bahwa:

Tabel 3 Hasil metode Supervised Learning pada Dataset Cicids2017

Algoritma	Accuracy
CNN	98.89%
SVM	93%
Random Forest	99.86%
MLP	92%

Secara umum, seluruh algoritma supervised learning menunjukkan performa yang sangat baik dengan nilai accuracy di atas 90%, kecuali MLP yang sedikit lebih rendah namun masih tergolong tinggi. Random Forest menjadi algoritma dengan performa terbaik, mendekati 100% accuracy, diikuti oleh CNN. Hal ini mengindikasikan bahwa pada dataset CICIDS2017 yang relatif besar dan kaya fitur, pendekatan supervised learning sangat efektif untuk membedakan antara trafik normal dan trafik serangan.

Jika dibandingkan dengan algoritma DRL pada dataset yang sama, terlihat dengan jelas bahwa supervised learning secara konsisten menghasilkan accuracy yang lebih tinggi. Perbedaan kinerja ini dapat disebabkan oleh beberapa faktor, antara lain: kebutuhan eksplorasi pada DRL yang membuat proses pembelajaran lebih kompleks, sensitivitas terhadap desain reward function, serta kebutuhan episode pelatihan yang jauh lebih banyak agar DRL dapat mencapai performa optimal.

3. Performa Metode DRL pada Dataset UNSW-NB15

Pada dataset UNSW-NB15, algoritma DRL yang diuji tetap sama, yaitu PPO, DQN, dan A2C. Hasil pengujian menunjukkan:

Tabel 4 Hasil metode DRL pada dataset UNSW-NB15

Algoritma	Accuracy
PPO	54%
DQN	54%
A2C	57%

Nilai accuracy ketiga algoritma DRL pada dataset UNSW-NB15 relatif rendah dan hampir seimbang, dengan A2C sedikit lebih baik dibanding PPO dan DQN. Nilai accuracy yang hanya berada pada kisaran 54–57% menunjukkan bahwa model DRL belum mampu menangkap pola serangan dan trafik normal dengan baik pada dataset UNSW-NB15. Hal ini dapat mengindikasikan bahwa kompleksitas fitur, distribusi kelas yang tidak seimbang, atau desain lingkungan (environment) dan reward pada proses training DRL belum sepenuhnya sesuai dengan karakteristik dataset ini.

4. Performa Metode Supervised Learning pada Dataset UNSW-NB15

Pada dataset yang sama, UNSW-NB15, metode supervised learning yang diuji adalah CNN, Random Forest, dan MLP. Hasil pengujian adalah sebagai berikut:

Tabel 5 Hasil metode Supervised Learning pada Dataset UNSW-NB15

Algoritma	Accuracy
CNN	96.87%
Random Forest	98.38%
MLP	95.41%

Ketiga algoritma supervised learning kembali menunjukkan performa yang sangat baik dengan accuracy di atas 95%. Random Forest masih menjadi algoritma dengan performa terbaik, disusul CNN dan MLP. Konsistensi tingginya performa Random Forest pada kedua dataset menguatkan indikasi bahwa metode ensemble berbasis pohon keputusan sangat cocok digunakan untuk permasalahan deteksi intrusi pada data berdimensi tinggi dan bercampur antara fitur numerik maupun kategorikal.

Jika dibandingkan dengan metode DRL pada dataset UNSW-NB15, perbedaan kinerja sangat mencolok. Supervised learning mampu mencapai accuracy hampir dua kali lipat dari metode DRL. Hal ini semakin menegaskan bahwa pada kondisi dan konfigurasi eksperimen yang digunakan, pendekatan supervised learning jauh lebih unggul dibanding DRL untuk kedua dataset yang diuji.

5. Analisis Perbandingan Umum dan Implikasi

Secara keseluruhan, beberapa poin penting yang dapat disimpulkan dari hasil analisa adalah:

a. Supervised learning konsisten lebih unggul dibandingkan DRL pada kedua dataset (CICIDS2017 dan UNSW-NB15), baik dari sisi nilai accuracy maupun stabilitas performa antar algoritma.
b. Random Forest menjadi algoritma dengan performa terbaik pada kedua dataset, dengan accuracy 99,86% pada CICIDS2017 dan 98,38% pada UNSW-NB15. Hal ini menunjukkan robustness Random Forest terhadap variasi karakteristik data dan besarnya jumlah sampel.
c. CNN juga menunjukkan performa sangat baik, sehingga cocok digunakan ketika ingin memanfaatkan kemampuan deep learning untuk ekstraksi fitur otomatis dari data trafik jaringan yang kompleks.
d. DRL, khususnya DQN, menunjukkan potensi yang cukup baik pada dataset CICIDS2017 (accuracy 89%), namun masih tertinggal jauh dari metode supervised learning. Pada dataset UNSW-NB15, performa DRL masih belum memadai.
e. Kemungkinan penyebab rendahnya performa DRL antara lain:
   a).Desain reward function yang belum optimal sehingga agen tidak memperoleh sinyal pembelajaran yang cukup informatif.
   b).Jumlah episode pelatihan dan konfigurasi hyperparameter yang belum memadai untuk konvergensi.
   c).Kompleksitas environment dan ruang aksi/keadaan yang tinggi sehingga agen kesulitan mengeksplorasi strategi yang baik.

6. Rekomendasi

Berdasarkan hasil analisa, rekomendasi yang dapat diberikan adalah:

a. Untuk kebutuhan implementasi praktis sistem deteksi intrusi yang membutuhkan akurasi tinggi dan stabil, metode supervised learning, khususnya Random Forest dan CNN, lebih direkomendasikan.
b. Metode DRL dapat diposisikan sebagai pendekatan penelitian lanjutan, misalnya untuk skenario yang lebih dinamis dan adaptif, namun diperlukan:
   a).Perancangan reward function yang lebih matang.
   b).Eksperimen lanjutan dengan jumlah episode yang lebih besar.
   c).Penyesuaian arsitektur jaringan dan hyperparameter.
c. Penelitian masa depan dapat mengeksplorasi hybrid approach, misalnya menggabungkan DRL dengan supervised learning (misalnya DRL untuk adaptasi kebijakan dan supervised untuk klasifikasi awal), sehingga diharapkan dapat memanfaatkan keunggulan kedua pendekatan.

4.5. Diskusi Teoritis dan Praktis

Hasil eksperimen dalam penelitian ini menunjukkan bahwa model Deep Reinforcement Learning (DRL), khususnya dengan algoritma Deep Q-Network (DQN), memiliki kemampuan superior dalam mendeteksi dan mengklasifikasikan berbagai jenis serangan siber dalam lalu lintas jaringan. Secara teoritis, keunggulan ini dapat dijelaskan melalui kerangka dasar reinforcement learning (RL), di mana agen belajar dari lingkungan melalui eksplorasi dan eksploitasi untuk memaksimalkan reward jangka panjang. Dalam konteks IDS, reward tersebut

dirancang agar agen belajar mengenali pola-pola lalu lintas jaringan yang merepresentasikan aktivitas normal dan serangan.

Prinsip RL ini memberikan fleksibilitas adaptif yang tidak dimiliki oleh model supervised learning tradisional, karena tidak memerlukan pasangan data-label eksplisit selama interaksi. Selain itu, struktur markov decision process (MDP) dalam RL cocok untuk lingkungan yang dinamis seperti lalu lintas jaringan yang terus berubah. Dengan pendekatan DRL, model mampu belajar dari sekuens trafik dan bereaksi terhadap anomali berdasarkan pengalaman kumulatif, bukan hanya pola statis(Huang et al., 2022).

Secara praktis, hasil penelitian ini berimplikasi besar pada pengembangan sistem pertahanan siber berbasis intelijen adaptif seperti Security Operation Center (SOC) dan firewall adaptif. Dengan menggunakan model DQN atau A2C, SOC dapat meningkatkan kemampuan deteksi dini terhadap serangan zero-day atau serangan kompleks berbasis multi-step. Sistem firewall adaptif juga dapat menggunakan model ini untuk memperbarui aturan (rules) secara otomatis berdasarkan feedback lalu lintas real-time, menggantikan pendekatan statis yang umum digunakan saat ini(Abeshu and Chilamkurti, 2021).

Jika dibandingkan dengan penelitian-penelitian sebelumnya, hasil penelitian ini tidak hanya mendukung tetapi juga memperluas wawasan ilmiah terkait efektivitas Deep Reinforcement Learning dalam konteks Intrusion Detection Systems (IDS). Misalnya, Berman et al (Berman et al., 2022) menekankan pentingnya strategi reward shaping dalam memperbaiki stabilitas pelatihan dan akurasi deteksi, yang juga dikonfirmasi oleh eksperimen kami saat mengadaptasi fungsi reward dalam skenario trafik multi-kelas. Penelitian Liu et al (Liu et al., 2023) membandingkan berbagai algoritma DRL dan menyimpulkan bahwa DQN lebih stabil dalam menangani ketidakseimbangan data, yang diperkuat oleh hasil akurasi DQN sebesar 92,39% dalam eksperimen ini. Zhao et al (Zhao et al., 2023) mengembangkan firewall adaptif berbasis DQN yang secara dinamis menyesuaikan aturan pemblokiran, menegaskan bahwa model DRL dapat diterapkan dalam sistem keamanan aktif. Sementara itu, Hossain et al (Hossain, Rahman and Lee, 2024)menunjukkan efektivitas DRL untuk mendeteksi ransomware pada edge computing, memperluas ruang lingkup penerapan model ini ke infrastruktur kritis. Oleh karena itu, penelitian ini tidak hanya menvalidasi performa DRL tetapi juga membuka jalur baru untuk integrasi ke sistem keamanan siber masa depan.

Hasil eksperimen ini juga memperkuat argumen bahwa DRL, dibanding supervised learning, lebih andal dalam lingkungan data tidak seimbang seperti CICIDS2017 yang digunakan dalam penelitian ini. Meskipun demikian, tantangan seperti kesulitan pelatihan pada kelas langka, kebutuhan komputasi tinggi, dan risiko reward hacking masih menjadi perhatian yang perlu ditangani pada implementasi praktis(Berman et al., 2022)(Liu et al., 2023)(Wang, Zhao and Li, 2022).

5. KESIMPULAN

Penelitian ini telah mengembangkan dan mengevaluasi model Intrusion Detection System (IDS) berbasis Deep Reinforcement Learning (DRL) menggunakan tiga algoritma berbeda, yaitu PPO, DQN, dan A2C, pada dataset CICIDS2017. Hasil eksperimen menunjukkan bahwa DQN unggul secara konsisten dalam hal akurasi (92,39%), precision (91,04%), recall (92,39%), dan F1-score (91,50%), mengungguli PPO dan A2C. Temuan ini mengindikasikan bahwa pendekatan value-based seperti DQN memiliki stabilitas dan efektivitas yang tinggi dalam mendeteksi serta mengklasifikasikan serangan multi-kelas, termasuk serangan dengan distribusi data yang tidak seimbang. Selain itu, rancangan arsitektur model dan formulasi environment sebagai Markov Decision Process (MDP) telah memungkinkan agen untuk melakukan deteksi sekaligus identifikasi jenis serangan, serta memberikan potensi integrasi ke dalam sistem pertahanan adaptif seperti SOC dan firewall cerdas.

Meskipun hasil penelitian menunjukkan performa yang menjanjikan, terdapat beberapa keterbatasan yang perlu dicatat. Pertama, eksperimen hanya dilakukan pada satu dataset benchmark (CICIDS2017) sehingga generalisasi model terhadap lalu lintas jaringan yang berbeda belum teruji secara luas. Kedua, pelatihan model memerlukan sumber daya komputasi yang cukup besar, sehingga implementasi pada perangkat dengan kapasitas terbatas mungkin menimbulkan tantangan. Ketiga, model belum diuji pada serangan zero-day yang sepenuhnya baru dan belum terwakili di dataset pelatihan. Keempat, meskipun reward shaping telah dioptimalkan, false positive pada beberapa kelas minor masih relatif tinggi, yang berpotensi mengganggu sistem mitigasi otomatis. Kelima, penelitian ini tidak membandingkan performa DRL dengan model supervised learning mutakhir sehingga klaim keunggulan hanya berlaku pada lingkup komparasi antar algoritma DRL.

Untuk penelitian lanjutan, disarankan penggunaan beberapa dataset yang berbeda dan lebih terkini, pengujian terhadap serangan zero-day, optimisasi arsitektur untuk lingkungan komputasi terbatas, serta integrasi transfer learning atau meta-reinforcement learning guna meningkatkan kemampuan adaptasi model pada kondisi lalu lintas yang sangat dinamis.

DAFTAR PUSTAKA

ABESHU, A. AND CHILAMKURTI, N., 2021. Deep learning applications for network intrusion detection: A review. *IEEE Access*, 9,

pp.21968–21985.

AHMED, M., KIM, H. AND PARK, Y., 2021. Intrusion detection system using deep reinforcement learning for cloud environments. *Journal of Network and Computer Applications*, 190, p.103146.

AHMED, S., KHAN, M. AND LEE, S., 2024. Adaptive deep reinforcement learning for intrusion detection in IoT networks. *IEEE Internet of Things Journal*, 11(2), pp.2134–2146.

AHMED, S., KHAN, M. AND LEE, S., 2025. Robust deep reinforcement learning against adversarial attacks in IDS. *IEEE Transactions on Information Forensics and Security*, 20, pp.1123–1135.

ALI, A., WANI, M.A. AND MIR, R.N., 2022. Deep recurrent neural network for anomaly-based intrusion detection system. *Expert Systems with Applications*, 193, p.116377.

BERMAN, D., BUCZAK, A.L., CHAVIS, J.S. AND CORBETT, C.L., 2022. A survey of deep learning methods for cyber security. *Information Fusion*, 61, pp.13–45.

CHEN, Y., WANG, F. AND LI, H., 2024. Zero-day attack detection using hybrid deep reinforcement learning. *Computers & Security*, 138, p.103576.

DIANA, L., DINI, P. AND PAOLINI, D., 2025. Overview on Intrusion Detection Systems for Computers Networking Security. *Computers*, 14(3), p.87.

HOSSAIN, M., RAHMAN, M. AND LEE, K., 2024. DRL-based intrusion detection in edge computing environments. *Future Generation Computer Systems*, 154, pp.155–170.

HUANG, L., LI, J. AND CHEN, Y., 2022. Deep reinforcement learning for network intrusion detection: A review. *IEEE Access*, 10, pp.35656–35670.

HUANG, W., XING, C., CHEN, Q. AND LI, Y., 2022. Reinforcement learning-based adaptive network intrusion detection system. *Computers & Security*, 114, p.102580.

KUMAR, R., SINGH, P. AND SHARMA, S., 2024. Hybrid DRL framework for real-time intrusion detection and mitigation. *Computer Networks*, 240, p.110200.

LI, Y., HUANG, Z. AND FENG, C., 2023. Reward shaping strategies for DRL-based intrusion detection systems. *Expert Systems with Applications*, 213, p.118993.

LIU, H., XU, L., ZHANG, Y. AND WANG, Y., 2023. Comparative analysis of DRL algorithms for imbalanced network traffic detection. *Knowledge-Based Systems*, 248, p.108246.

LIU, Z., XU, X. AND WANG, Y., 2023. A comparative study of deep reinforcement learning algorithms for intrusion detection. *Computers & Security*, 130, p.103123.

NGUYEN, T.T., REDON, F. AND RAGHAVAN, V., 2021. Deep reinforcement learning for cyber security: A review of recent advances. *ACM Computing Surveys (CSUR)*, 54(9), pp.1–36.

PARK, J., KIM, H. AND CHOI, S., 2023. Hybrid deep reinforcement learning framework for network intrusion detection. *Future Generation Computer Systems*, 146, pp.450–462.

RAFFIN, A., HILL, A., GLEAVE, A., KANERVISTO, A. AND ERNESTUS, M., 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 22(268), pp.1–8.

RAHMAN, M.A., HOSSAIN, M.S. AND MUHAMMAD, G., 2023. A hybrid deep learning model for efficient intrusion detection in IoT networks. *Computers & Security*, 123, p.102936.

SANTOS, R., PEREIRA, L. AND GOMES, A., 2025. Comparative evaluation of DRL algorithms for real-time intrusion detection. *Expert Systems with Applications*, 236, p.121234.

SCHULMAN, J., WOLSKI, F., DHARIWAL, P., RADFORD, A. AND KLIMOV, O., 2017. *Proximal Policy Optimization Algorithms*.

WANG, J., ZHAO, L. AND LI, M., 2022. An actor–critic DRL approach for anomaly detection in dynamic network environments. *Future Generation Computer Systems*, 135, pp.280–292.

WU, Y., ZENG, Z. AND LI, X., 2023. An adaptive IDS based on actor–critic DRL for cyber–physical systems. *Future Generation Computer Systems*, 141, pp.99–110.

XU, L., LIU, H., ZHANG, Y. AND WANG, Y., 2021. A deep reinforcement learning-based intrusion detection model for imbalanced network traffic. *Knowledge-Based Systems*, 228, p.107243.

ZHANG, H., CHEN, X. AND XU, Y., 2023. Deep reinforcement learning-based adaptive intrusion detection for evolving cyber threats. *Computers & Security*, 128, p.103146.

ZHAO, W., ZHANG, T. AND FENG, X., 2023. Adaptive firewall policies using deep reinforcement learning. *IEEE Transactions on Network and Service Management*, 20(4), pp.4501–4515.

ZHAO, Y., LI, X., ZENG, Z. AND WU, Y., 2023.

Adaptive firewall policy generation using deep Q-learning for cyber–physical systems. *Computer Networks*, 231, p.109189.

ZHOU, M., MA, H., JIANG, Z. AND TANG, M., 2024. A hybrid DRL-based anomaly detection and mitigation framework for dynamic network environments. *Computer Networks*, 239, p.110145.

Halaman ini sengaja dikosongkan