Model Klasifikasi Serangan DoS pada Jaringan Blockchain Menggunakan Algoritma Proximal Policy Optimization

Iffo Elsande Pratama Putra[1], Ricky Eka Putra[2]

[1,2] Program Studi Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya
[1]iffo.21064@mhs.unesa.ac.id
[2]rickyeka@unesa.ac.id

***Abstrak*— Teknologi blockchain menghadirkan pendekatan baru dalam pengelolaan sistem informasi terdesentralisasi yang mampu menjaga keamanan, transparansi, dan integritas data. Namun, karakteristik tersebut menjadikan teknologi blockchain rentan terhadap ancaman siber, terutama serangan *Denial of Service* (DoS) yang berfokus pada gangguan ketersediaan layanan melalui pembanjiran lalu lintas pada node blockchain. Penelitian ini bertujuan untuk merancang dan mengembangkan model klasifikasi serangan DoS pada jaringan blockchain dengan menggunakan algoritma *Proximal Policy Optimization* (PPO). Algoritma PPO merupakan salah satu metode dari *reinforcement learning* yang dikenal memiliki kestabilan tinggi dan efisiensi dalam proses pembaruan kebijakan. Dataset yang di gunakan dalam penelitian ini ada *Blockchain Network Attack Traffic* (BNaT), yang mencakup lalu lintas normal dan serangan DoS pada jaringan Ethereum privat. Proses penelitian meliputi tahap pengumpulan data, pre-pemrosesan (*preprocessing*), pelatihan model, dan evaluasi kinerja menggunakan metrik *accuracy, precision, recall, F1-Score,* dan *Area Under the Curve* (AUC). Hasil pengujian menunjukkan bahwa model PPO berhasil mencapai akurasi 99,65% dan *F1-Score* sebesar 99,65%, dengan nilai AUC mencapai 99,99%. Nilai-nilai tersebut menunjukkan bahwa PPO mampu mengenali pola serangan DoS secara adaptif dan stabil. Oleh karena itu, pendekatan *reinforcement learning* berbasis PPO dapat menjadi alternatif yang menjanjikan untuk pengembangan sistem deteksi ancaman pada jaringan blockchain yang bersifat dinamis dan kompleks.**

***Kata Kunci*— Blockchain, Denial of Service, Proximal Policy Optimization, Reinforcement Learning, Keamanan Siber**

I. PENDAHULUAN

Teknologi blockchain telah berperan sebagai salah satu inovasi dalam transformasi digital di era modern. Sebagai sistem pencatatan yang bersifat terdistribusi, blockchain memiliki karakteristik utama berupa desentralisasi, transaparansi dan perlindungan data berbasis kriptografi tingkat tinggi. Mekanisme konsensus yang diterapkan berperan dalam menjaga integritas data serta membangun kepercayaan antar entitas tanpa bergantung pada otoritas terpusat [1]. Sifat terdistribusi pada blockchain tidak secara inheren menjamin ketahanan penuh terhadap serangan siber, khususnya yang menargetkan aspek ketersediaan layanan (*availability*). Ketergantungan pada node individual menimbulkan potensi kerentanan terhadap serangan berbasis lalu lintas jaringan. Selain itu, kompleksitas komunikasi antar node dalam jaringan menghadirkan tantangan terhadap efisiensi dan stabilitas sistem. Mekanisme konsensus seperti *Proof of Work* (PoW) dan *Proof of Stake* (PoS) berpotensi mengalami degradasi performa ketika beroperasi di bawah kondisi beban jaringan yang tinggi [2].

Serangan *Denial of Service* (DoS) termasuk ancaman serius terhadap keberlangsungan operasional jaringan blockchain. Pola serangan ini dilakukan dengan menghasilkan lalu lintas jaringan dalam jumlah besar ke arah node, sehingga menyebabkan penurunan performa sistem, keterlambatan proses validasi transaksi, dan terganggunya komunikasi antarnode [3]. Beberapa laporan pada jaringan yang mengadopsi *Ethereum Virtual Machine* (EVM) menunjukkan bahwa pelaku dapat memanfaatkan transaksi dan panggilan *Remote Prodecure Call* (RPC) dengan beban komputasi tinggi untuk membuat node kehilangan responsivitas. Kondisi ini mengindikasi bahwa meskipun blockchain dirancang secara terdesentralisasi, gangguan operasional pada tingkat node dapat terjadi akbat eksploitasi sumber daya lokal.

Serangan DoS pada jaringan blockchain umumnya dilakukan melalui pola *single-origin flooding* dengan mengeksploitasi keterbatasan sumber daya node seperti kapasitas pemrosesan, memori, dan manajemen antrean transaksi (*mempool*). Tekanan berlebih pada komponen tersebut mengakibatkan ketidakstabilan dalam proses propagasi blok serta memperlambat penyelesaian transaksi. Pada saat lonjakan lalu lintas jaringan terjadi secara signifikan, node tidak lagi mampu mempertahankan kinerja operasionalnya, sehingga efisiensi jaringan secara keseluruhan menurun. Dalam arsitektur *peer-to-peer* (P2P), latensi propagasi antar-node semakin memperbesar dampak serangan terhadap stabilitas sistem [4]. Oleh karena itu, diperlukan pendekatan adaptif yang mampu mengidentifikasi dan mengklasifikasikan perbedaan antara lalu lintas normal dan serangan secara cepat pada lingkungan blockchain.

Berbagai pendekatan berbasis *machine learning* telah banyak di gunakan untuk menganalisis dan mengidentifikasi serangan pada lalu lnitas jaringan. Algoritma seperti *Random Forest* dan *Deep Neural Network* (DNN) dilaporkan memiliki kinerja yang baik dalam mengenali serangan DoS dengan tingkat akurasi yang tinggi [5]. Namun, metode *supervised learning* sangat bergantung pada dataset berlabel, sehingga sulit beradaptasi terhadap pola serangan baru. Di samping itu, pendekatan berbasis tanda tangan umumnya hanya efektif untuk serangan yang telah terdefinisi sebelumnya, sehingga kurang mampu menangani serangan yang bersifat dinamis dan non-deterministik.

Sebagai alternatif terhadap pendekatan *supervised learning, Reinforcement Learning* (RL) menawarkan kemampuan adaptif karena model memperoleh pengetahuan melalui interaksi berulang dengan lingkungan tanpa memerlukan supervisi langsung [6]. Metode ini tidak memerlukan proses

anotasi label secara eksplisit dan memperbarui kebijakan berdasarkan umpan balik dari tindakan. Mekanisme pembelajaran yang bersifat berkelanjutan menjadikan RL sebagai solusi efektif dalam menghadapi dinamika ancaman yang tinggi pada sistem keamanan jaringan. Dalam ekosistem blockchain, RL berpotensi dimanfaatkan untuk mengklasifikasi pola lalu lintas normal dan serangan berdasarkan respons node terhadap variasi beban jaringan.

Algoritma *Proximal Policy Optimization* (PPO) merupakan salah satu algoritma RL berbasis *policy gradient* yang dirancang untuk meningkatkan stabilitas dan efisiensi dalam proses pembaruan kebijakan [7]. Pendekatan ini memanfaatkan mekanisme *clipped surrogate objective* untuk membatasi perubahan kebijakan yang terlalu agresif selama proses pembelajaran [8]. Sejumlah studi melaporkan bahwa PPO memiliki efisiensi yang lebih baik dibandingkan metode *policy gradient* lainnya, khususnya dalam hal stabilitas pembelajaran dan pemanfaatan sampel [9]. Karakteristik tersebut menjadikan PPO sesuai untuk klasifikasi lalu lintas normal dan serangan DoS pada jaringan blockchain dengan dinamika lalu lintas yang tinggi.

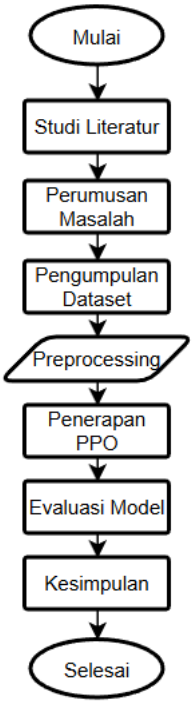
Model *Deep Reinforcement Leraning* telah meningkatkan performa dalam tugas klasifikasi serangan DoS secara dinamis. Pembelajaran mendalam berbasis RL dilaporkan mampu meningkatkan akurasi sekitas 10 % dibandingkan pendekatan berbasis tanda tangan [6]. Selain itu, *Double Deep Q-Network* (DDQN) diketahui mampu mempertahankan stabilitas operasional sistem saat berada pada kondisi serangan DoS yang berat [10]. Temuan tersebut mengindikasikan bahwa pendekatan berbasis RL memiliki sifat adaptif, efisien, dan lebih tangguh terhadap variasi pola serangan yang kompleks

Penerapan PPO dalam hal klasifikasi serangan DoS pada jaringan blockchain masih tergolong terbatas. Mayoritas studi terdahulu berfokus pada sistem deterministrik seperti *Industrial Control System* (ICS) dan *Internet of Things* (IoT) [11]. Dengan demikian, di perlukan analisis terhadap efektivitas PPO pada lalu lintas blockchain untuk menilai kemampuannya dalam mengklasifikasi lalu lintas normal dan serangan pada lingkungan jaringan dengan dinamika dan kompleksitas yang tinggi.

Makalah ini mengusulkan model klasifikasi lalu lintas jaringan blockchain dengan kelas normal dan serangan DoS menggunakan algoritma PPO. Pendekatan ini memanfaatkan RL yang adaptif untuk mengklasifikasikan pola lalu lintas secara dinamis tanpa ketergantungan pada metode berbasis tanda tangan. Evaluasi dilakukan secara komprehensif untuk menilai efektivitas dan stabilitas PPO pada lingkungan jaringan blockchain dengan dinamika dan kompleksitas tinggi.

II. METODE PENELITIAN

A. *Alur Penelitian*



Gbr. 1 Alur Penelitian

Penelitian ini mengembangkan model klasifikasi lalu lintas jaringan berbasis *Proximal Policy Optimization* (PPO) melalui alur metodologi terstruktur sebagaimana ditunjukkan pada Gbr. 1, seluruh tahapan dieksekusi secara berurutan untuk menjamin konvergensi pembelajaran serta keandalan model dalam mengklasifikasi serangan *Denial of Service* (DoS) pada jaringan blokchcain.

B. *Studi Literatur*

Studi literatur dilakukan melalui peninjauan yang sistematis terhadap jurnal ilmiah, prosiding konferensi dan sumber akademik terkait untuk menganalisis perkembangan metode klasifikasi serangan DoS berbasis *machine learning, deep learning,* dan *reinforcement learning* dalam lima tahun terakhir. Kajian ini menunjukkan bahwa berbagai pendekatan telah mencapai kinerja tinggi pada lingkungan jaringan konvensional, namun belum sepenuhnya mengakomodasi kompleksitas lalu lintas jaringan blockchain yang bersifat terdesentralisasi dan non-deterministik. Temuan tersebut menjadi dasar pemilihan algoritma PPO sebagai pendekatan yang diusulkan, sekaligus mengidentifikasi keterbatasan utama penelitian sebelumnya yang menjadi fokus penyempurnaan dalam studi ini.

C. *Perumusan Masalah*

Perumusan masalah didasarkan pada temuan studi literatur yang mengindikasikan bahwa jaringan blockchain rentan terhadap serangan DoS, sementara pendekatan deteksi konvensional masih memiliki keterbatasan dalam menangani pola serangan yang kompleks dan dinamis. Oleh karena itu, permasalahan utama yang dikaji Adalah bagaimana meningkatkan efektivitas klasifikasi serangan DoS melalui penerapan pendekatan RL berbasis PPO seiring dengan

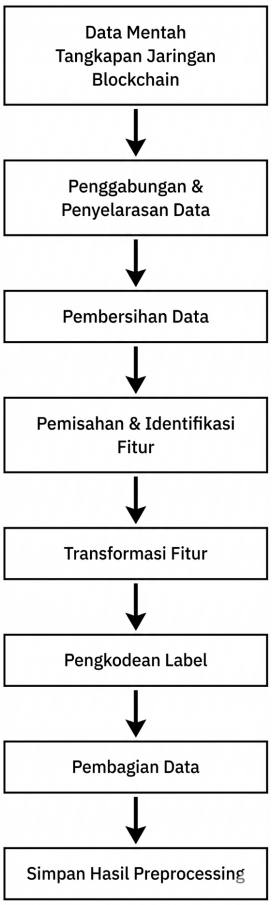
meningkatnya adopsi teknologi blockchain berbagai sektor industri.

D. *Pengumpulan Dataset*

Dataset Dataset yang digunakan adalah *Blockchain Network Attack Traffic* (BNaT) yang diperkenalkan dalam studi *"Collaborative Learning for Cyberattack Detection in Blockchain Networks"* [12]. Dataset ini dihasilkan melalui eksperimen laboratorium pada jaringan Ethereum privat dan mencakup lalu lintas normal serta beberapa skenario serangan siber pada blockchain, termasuk *Brute Password* (BP), *Flooding of Transactions* (FoT), *Man-in-the-Middle* (MitM), dan *Denial of Service* (DoS).

Penelitian ini menggunakan subset dua kelas, yaitu Normal dan serangan DoS, untuk memfokuskan analisis pada serangan yang secara langsung memengaruhi aspek *availability* jaringan blockchain. BNaT terdiri atas 21 fitur utama yang mencakup *basic network features* dan *statistical features*, di mana setiap instans merepresentasikan satu sesi komunikasi jaringan berdasarkan *time window*. Data dikumpulkan dari tiga *Ethereum full nodes* pada jaringan eksperimen dan digunakan sebagai sumber utama dalam proses pelatihan serta evaluasi model klasifikasi serangan DoS.

E. *Preprocessing Data*



Gbr. 2 Preprocessing Data

Tahapan *preprocessing* data dilakukan untuk menyiapkan dataset BNaT sebelum digunakan dalam proses pelatihan dan evaluasi model, sebagaimana diilustrasikan pada Gbr. 2. Proses ini mencakup penggabungan berkas lalu lintas jaringan dari tiga *Ethereum full nodes* menjadi satu dataset terpadu, penyelarasan struktur fitur untuk menjaga konsistensi, serta pemilihan dua kelas utama, yaitu Normal dan serangan DoS. Selanjutnya, data duplikat di hapus untuk mengurangi redundansi dan memastikan setiap entri data merepresentasikan satu sesi komunikasi jaringan yang unik. Dataset kemudian dipisahkan menjadi fitur dan label target, dengan klasifikasi fitur kedalam tipe numerik dan kategorikal untuk mendukung proses tranformasi yang sesuai.

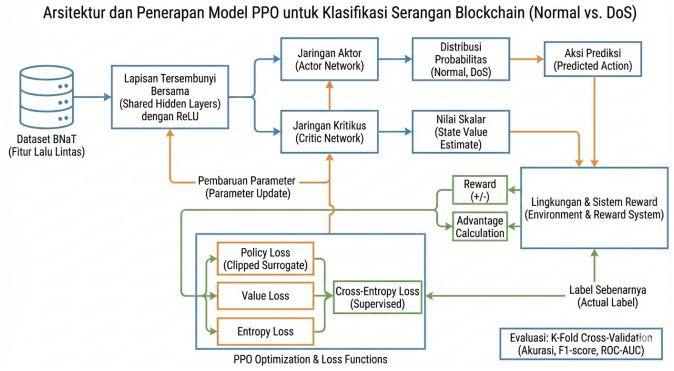
Atribut numerik dinormalisasi menggunakan *Z-score normalitation*, sedangkan atribut kategori dikonversi ke bentuk numerik melalui *Ordinal Encoding*. Seluruh transformasi diintegrasikan ke dalam satu *preprocessing pipeline* untuk menjamin keseragaman antara data pelatihan dan data pengujian. Label target dikodekan secara biner, di mana kelas Normal direpresentasikan sebagai 0 dan kelas DoS sebagai 1. Dataset hasil *preprocessing* selanjutnya dibagi menjadi data latih dan data uji menggunakan metode *stratified sampling* dengan rasio 80:20 untuk menjaga keseimbangan distribusi kelas serta memastikan evaluasi performa model yang objektif. Seluruh tahapan *preprocessing* dilakukan secara berurutan dan konsisten agar menjamin reprodusibilitas hasil penelitian.

F. *Arsitektur dan Penerapan Model*

Algortima PPO diterapkan sebagai pendekatan pembelajaran penguatan berbasis *policy gradient* yang dirancang untuk menjaga stabilitas optimisasi dan efisiensi komputasi selama proses pelatihan. Dalam penelitian ini, PPO di formulasikan untuk tugas klasifikasi biner pada dataset BNaT, dengan dua kelas utama, yaitu Normal dan serangan DoS. Arsitektur PPO mengadopsi kerangka *Actor-Critic*, yang terdiri atas komponen *actor* yang memodelkan kebijakan dalam bentuk distribusi probabilitas tindakan serta komponen *critic* yang mengestimasi nilai keadaan (*state value*). Kedua komponen dioptimalkan secara bersamaan dengan tujuan menyeimbangkan eksplorasi kebijakan dan kestabilan pembaruan parameter.

Jaringan *Actor–Critic* dibangun dengan dua lapisan tersembunyi menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU). Representasi fitur diekstraksi melalui lapisan bersama sebelum diteruskan ke dua cabang jaringan, di mana *actor* menghasilkan probabilitas prediksi kelas dan *critic* menghasilkan estimasi nilai keadaan dalam bentuk skalar. Proses pelatihan memanfaatkan sinyal reward untuk memperkuat prediksi yang benar dan menekan prediksi yang keliru. Pembaruan kebijakan dilakukan menggunakan rasio antara kebijakan baru dan lama yang dibatasi oleh mekanisme *clipping* dengan parameter ε, sehingga perubahan kebijakan tetap berada dalam rentang yang terkendali. Estimasi *advantage function* digunakan sebagai dasar pembaruan parameter actor untuk meningkatkan kualitas kebijakan secara bertahap.

Proses optimisasi melibatkan kombinasi tiga komponen fungsi kehilangan, yaitu *policy loss* berbasis *clipped surrogate objective*, *value loss* untuk menyesuaikan estimasi nilai oleh critic, dan *entropy loss* untuk mempertahankan keberagaman kebijakan selama eksplorasi. Selain itu, PPO diintegrasikan dengan komponen pembelajaran terawasi melalui fungsi *cross-entropy loss* terhadap label aktual, sehingga membentuk skema pembelajaran hibrida. Integrasi ini memungkinkan model memanfaatkan sinyal reward sekaligus informasi berlabel untuk meningkatkan stabilitas pelatihan dan akurasi klasifikasi. Arsitektur dan alur penerapan PPO secara keseluruhan ditunjukkan pada Gbr. 3.



Gbr. 3 Arsitektur dan Penerapan PPO

G. *Evaluasi Model*

Evaluasi dilakukan untuk menilai kinerja model berbasis PPO dalam mengklasifikasikan serangan DoS pada jaringan, dengan focus pada akurasi prediksi, stabilitas pelatihan dan kemampuan generalisasi terhadap data yang tidak digunakan selama proses pelatihan. Kinerja model diukur menggunakan metrik standar, yaitu *accuracy, precision, recall, F1-Score,* dan *Area Under the Receiver Operating Characteristic Curve* (AUC−ROC), serta divalidasi melalui skema *K-Fold Cross Validation* untuk memperoleh estimasi performa yang stabil dan tidak bergantung pada satu pembagian data tertentu. Selain evaluasi kuantitatif, konvergensi proses pembelajaran dianalisis melalui dinamika *reward* dan perubahan nilai fungsi kehilangan untuk memastikan bahwa optimasi PPO mencapai keseimbangan yang konsisten antara eksplorasi dan eksploitasi. Hasil evaluasi ini menjadi dasar empiris dalam menilai efektivitas pendekatan PPO terhadap klasifikasi adaptif serangan DoS pada lalu lintas jaringan blockchain

H. Rancangan Skenario Uji Coba

Skenerio pengujian dirancang untuk menentukan konfigurasi pelatihan optimal pada model PPO melalui eksplorasi sistematis ruang pencarian *hyperparameter* menggunakan pendekatan *grid search*. Parameter yang dievaluasi meliputi *learning rate*, batas pembaruan kebijakan (*clipping parameter* ε), koefisien entropi sebagai pengendali eksplorasi, ukuran *batch*, bobot fungsi nilai (*value loss coefficient*), serta jumlah *fold* pada skema validasi silang. Evaluasi performa dilakukan dengan metode *Stratified K-Fold Cross-Validation* menggunakan dua dan lima *fold* untuk memastikan estimasi kinerja yang stabil terhadap variasi distribusi kelas sekaligus mengurangi potensi *overfitting*. Setiap kombinasi *hyperparameter* diberi identitas konfigurasi unik guna memfasilitasi pelacakan eksperimen dan analisis komparatif secara sistematis. Kinerja model dinilai berdasarkan *F1-score* sebagai metrik utama karena kemampuannya merepresentasikan keseimbangan antara *precision* dan *recall* pada data tidak seimbang, sedangkan *AUC–ROC* dan *Average Precision* digunakan sebagai metrik pendukung. Rincian konfigurasi *hyperparameter* yang diuji ditunjukkan pada Tabel I.

TABEL I
HYPERPARAMETER PENGUJIAN

Parameter	Nilai	Keterangan
Learning Rate	0.0001, 0.0003	Laju pembelajaran untuk mengatur seberapa besar langkah update bobot
Clip Epsilon	0.08, 0.12	Batas perubahan kebijakan pada algoritma PPO
Entropy Coefficient	0.005, 0.01	Koefisien untuk mendorong eksplorasi model
Batch Size	128, 256	Ukuran data per iterasi pelatihan
Value Coefficient	0.5, 0.7	Bobot fungsi nilai dalam total loss PPO
K-Fold	2, 5	Jumlah lipatan dalam validasi silang stratifikasi

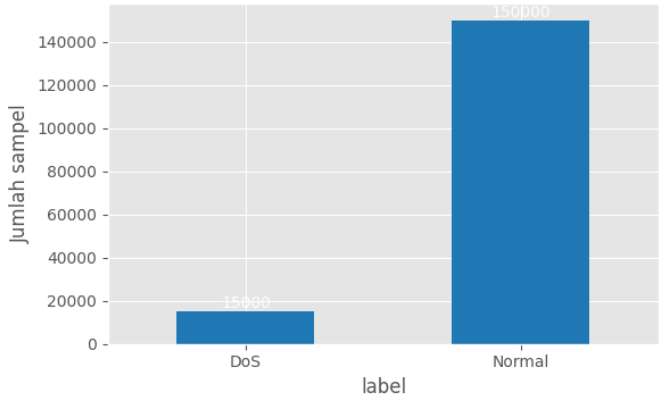
III. HASIL DAN PEMBAHASAN

A. *Hasil Preprocessing Dataset*

Tahapan *preprocessing* dilakukan untuk menyiapkan dataset BNaT agar sesuai dengan kebutuhan pelatihan model PPO. Proses ini di tunjukkan sebagaimana pada Gbr. 2 yang mencakup penggabungan data, penghapusan duplikasi, identifikasi atribut numerik dan kategorikal, transformasi data ke format numerik, pemetaan label, serta pembagian dataset menjadi subset pelatihan dan pengujian. Seluruh tahapan dirancang untuk memastikan integritas data, konsistensi fitur, serta keseimbangan distribusi antar kelas sebelum tahap pelatihan dimulai.

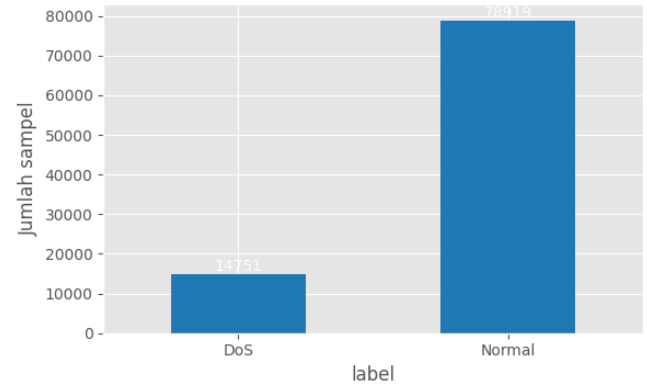
1) Penggabungan dan Penyelarasan Data: Dataset BNaT terdiri atas tiga berkas hasil tangkapan lalu lintas jaringan, masing-masing berisi 70.000 sampel dengan total keseluruhan 210.000 baris dan 22 atribut. Setiap berkas merepresentasikan aktivitas jaringan blockchain pada kondisi normal maupun saat

terjadi serangan siber. Proses penggabungan dilakukan secara otomatis untuk menyatukan seluruh berkas ke dalam satu struktur *DataFrame* terpadu untuk memudahkan proses analisis. Setelah tahap penyaringan kelas, hanya dua kelas utama yang dipertahankan, yaitu Normal dan serangan DoS, sehingga diperoleh 165.000 sampel dengan distribusi 150.000 sampel Normal (90,91%) dan 15.000 (9,09%). Distribusi label awal dataset ditunjukkan pada Gbr. 4.



Gbr. 4 Distribusi Label Dataset Sebelum Pembersihan

2) Pembersihan Data: pembersihan data dilakukan untuk menghilangkan baris duplikat yang muncul akibat penggabungan data dari beberapa node. Proses deduplikasi mengidentifikasi dan menghapus sebanyak 71.330 entri ganda, sehingga diperoleh 93.670 sampel unik dengan 22 atribut. Dataset hasil pembersihan terdiri atas 78.919 sampel kelas Normal dan 14.751 sampel kelas DoS, dengan proporsi masing-masing sebesar 84% dan 16%. Distribusi kelas setelah proses pembersihan ditunjukkan pada Gbr. 5.



Gbr. 5 Distribusi Label Dataset Setelah Proses Pembersihan

3) Identifikasi Fitur: Dataset yang digunakan terdiri atas 21 atribut prediktor dan satu label target. Seluruh atribut prediktor dikelompokkan berdasarkan tipe datanya untuk mendukung proses transformasi dan pemodelan selanjutnya. Sebanyak 18 fitur diklasifikasikan sebagai numerik, sedangkan 3 fitur lainnya bersifat kategorikal, yaitu *protocol_type, service,* dan *flag.* Rincian klasifikasi atribut disajikan pada Tabel II.

TABEL II
KLASIFIKASI ATRIBUT DALAM DATASET BNAT

Fitur Kategorikal	protocol_type, service, flag
Fitur Numerik	duration, src_bytes, dst_bytes, count, srv_count, serror_rate, same_srv_rate, diff_srv_rate, srv_serror_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_serror_rate, dst_host_srv_diff_host_rate, dan dst_host_srv_serror_rate

4) Transformasi dan Pengkodean Fitur: Untuk menyamakan skala antar atribut dan mencegah dominasi fitur tertentu selama proses optimisasi, seluruh fitur numerik distandarisasi menggunakan *Z-score normalization* melalui penerapan *StandardScaler*. Fitur kategorikal dikonversi ke representasi numerik menggunakan *Ordinal Encoding* agar kompatibel dengan mekanisme pembaruan parameter pada model pembelajaran berbasis gradien. Skema pengkodean yang digunakan untuk setiap atribut kategorikal disajikan pada Tabel III.

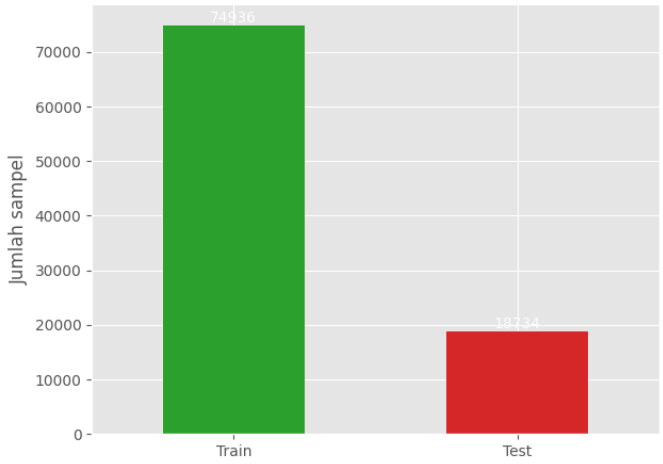
TABEL III
HASIL PEMETAAN UNTUK FITUR KATEGORI

Fitur	**Pemetaan Kategori**
Protocol_type	tcp: 0 udp: 1 icmp: 2
Service	other: 0 http: 1 private: 2 netbios_ssn: 3 oth_i: 4
Flag	OTH: 0 S1: 1 SF: 2 S3: 3

5) Pemetaan Label: Label target dikonversi dari format teks menjadi representasi numerik untuk mendukung proses klasifikasi biner, dengan kelas Normal di representasikan sebagai 0 dan DoS sebagai 1. Proses ini dilakukan bersamaan dengan pemisahan antara fitur prediktor (X) dan label target (y) agar memastikan pemisahan yang tegas antara variabel masukan dan keluaran selama pelatihan model.

6) Pembagian Dataset: Dataset yang telah melalui tahap transformasi dibagi menjadi data latih dan data uji menggunakan skema *stratified sampling* dengan rasio 80:20

untuk mempertahankan konsistensi distribusi kelas pada kedua subset. Proses ini menghasilkan 74.936 sampel pada data latih dan 18.734 sampel pada data uji. Distribusi kelas hasil pembagian ditunjukkan pada Gbr. 6, yang mengkonfirmasi terjaganya keseimbangan proporsional antara kelas Normal dan DoS pada masing-masing subset.



Gbr. 6 Distribusi Sampel Pelatihan dan Pengujian

B. *Implementasi Model*

Model PPO diimplementasikan menggunakan *framework PyTorch* dengan mengadopsi arsitektur *Actor−Critic* sebagaimana dijelaskan pada Bagian II-F. Penerapannya difokuskan pada perancangan pipeline pelatihan adaptif untuk tugas klasifikasi biner pada dataset BNaT yang mencakup dua kelas utama, yaitu Normal dan DoS.

Proses pelatihan memanfaatkan *clipped surrogate objective* untuk membatasi perubahan kebijakan dan menjaga stabilitas gradien selama optimisasi. Selain itu, komponen *value function loss* dan *entropy regularization* diintegrasikan untuk mengontrol keseimbangan antara eksplorasi dan eksploitasi. Pipeline implementasi mencakup pemuatan dataset terstandarisasi, pembentukan memori pengalaman, pembaruan parameter berbasis *policy loss*, *value loss*, dan *entropy loss*, serta validasi berkala menggunakan skema K-Fold Cross-Validation. Seluruh hiperparameter pelatihan dikelola secara modular guna menjamin konsistensi eksperimen dan reprodusibilitas hasil.

C. *Pelatihan Model*

Tahapan pelatihan dirancang untuk mengoptimalkan parameter model PPO dalam melakukan klasifikasi adaptif terhadap lalu lintas jaringan blockchain ke dalam dua kelas, yaitu Normal dan DoS. Fokus utama pelatihan diarahkan pada pencapaian konvergensi yang stabil, pengendalian *trade-off* antara eksplorasi dan eksploitasi kebijakan, serta efisiensi proses pembaruan parameter selama optimasi.

Mengacu pada skenario eksperimen yang di jelaskan pada Bagian II-H, dilakukan 64 percobaan pelatihan menggunakan kombinasi berbagai hiperparameter, meliputi *learning rate*, *clip epsilon*, *entropy coefficient*, *value coefficient*, *batch size*, dan jumlah lipatan validasi silang (*K-Fold*). Eksplorasi ini bertujuan untuk menganalisis pengaruh masing-masing parameter terhadap stabilitas pelatihan dan performa model. Evaluasi setiap konfigurasi dilakukan menggunakan *F1-score* sebagai metrik utama, karena kemampuannya dalam merepresentasikan keseimbangan antara *precision* dan *recall* pada distribusi kelas yang tidak seimbang.

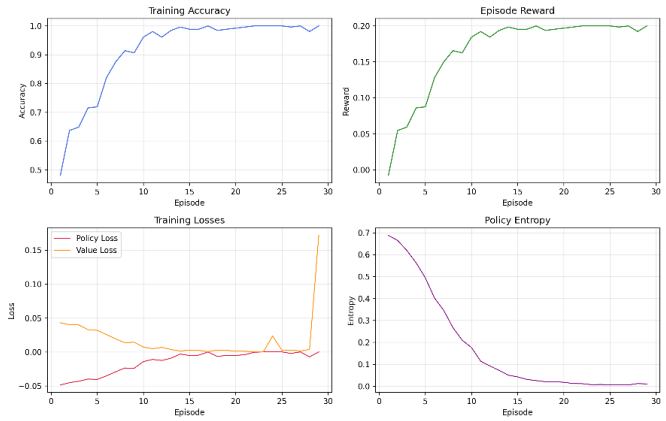
Hasil eksplorasi menunjukkan bahwa konfigurasi dengan *learning rate* sebesar 0.0003, *clip epsilon* 0.12, *entropy coefficient* 0.005, *batch size* 256, *value coefficient* 0.7, serta *K-Fold* sebesar 2 memberikan performa terbaik. Konfigurasi ini menghasilkan F1-score tertinggi dengan pola konvergensi yang stabil, sehingga ditetapkan sebagai konfigurasi utama pada tahap pelatihan akhir model PPO. Rincian konfigurasi hiperparameter optimal disajikan pada Tabel IV.

TABEL IV
KONFIGURASI TERBAIK PADA MODEL PPO

Parameter	Nilai Optimal
Learning Rate	0.0003
Clip Epsilon	0.12
Entropy Coefficient	0.005
Batch Size	256
Value Coefficient	0.7
K-Fold	2

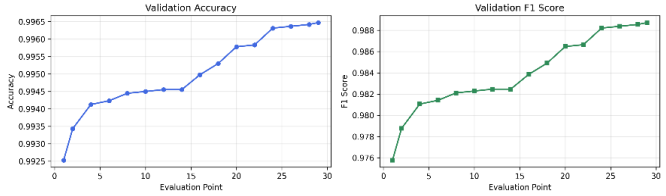
Optimisasi model didasarkan pada tiga komponen fungsi kehilangan utama, yaitu *policy loss*, *value loss*, dan *entropy loss*, yang secara kolektif berfungsi untuk menjaga stabilitas kebijakan, meningkatkan akurasi estimasi nilai keadaan, serta mempertahankan tingkat eksplorasi yang memadai. Pada setiap episode, umpan balik *reward* digunakan untuk memperbarui parameter kebijakan berdasarkan rasio antara kebijakan baru dan kebijakan sebelumnya.

Dinamika pelatihan model PPO ditunjukkan pada Gbr. 7. Sub-gambar (a) memperlihatkan peningkatan *training accuracy* yang cepat hingga mencapai kondisi stabil mendekati 1.0 setelah sekitar 20 episode, mengindikasikan bahwa model mampu mempelajari perbedaan antara lalu lintas normal dan serangan DoS secara efektif. Sub-gambar (b) menampilkan peningkatan *episode reward* yang konsisten hingga mendekati nilai maksimum, mencerminkan efektivitas mekanisme penguatan kebijakan. Pada sub-gambar (c), *policy loss* dan *value loss* cenderung menurun seiring bertambahnya episode, menandakan proses optimisasi parameter yang efisien. Sementara itu, sub-gambar (d) menunjukkan penurunan *policy entropy* yang tajam pada fase awal pelatihan dan kemudian stabil, yang mengindikasikan pergeseran bertahap dari eksplorasi menuju eksploitasi kebijakan.



Gbr. 7 Metrik pelatihan model PPO: (a) training accuracy; (b) episode reward; (c) training losses; (d) policy entropy.

Selain pemantauan metrik pelatihan, evaluasi validasi dilakukan menggunakan subset data yang terpisah untuk menilai kemampuan generalisasi model di luar data pelatihan. Hasil validasi yang disajikan pada Gbr. 8 memperlihatkan peningkatan yang konsisten pada nilai validation accuracy dan F1-score sepanjang proses pelatihan. Validation accuracy mencapai nilai mendekati 0.99 pada tahap akhir, sementara F1-score menunjukkan tren peningkatan yang stabil. Konsistensi antara performa pelatihan dan validasi mengindikasikan bahwa model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik terhadap data uji. Hasil ini menegaskan efektivitas algoritma PPO dalam mempelajari pola lalu lintas jaringan blockchain pada skenario klasifikasi biner.



Gbr. 8 Performa validasi model PPO. (a) validation accuracy; (b) validation F1-score.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma PPO mencapai konvergensi yang stabil dengan peningkatan *reward* yang konsisten, penurunan fungsi kehilangan yang terkontrol, serta dinamika entropi yang sesuai. Temuan ini menegaskan efektivitas mekanisme *clipped objective* dalam menjaga kestabilan pembaruan kebijakan, sekaligus menunjukkan kemampuan PPO dalam mempelajari representasi lalu lintas jaringan blockchain yang relevan untuk membedakan aktivitas normal dan serangan DoS secara adaptif dan efisien.

D. *Evaluasi Model*

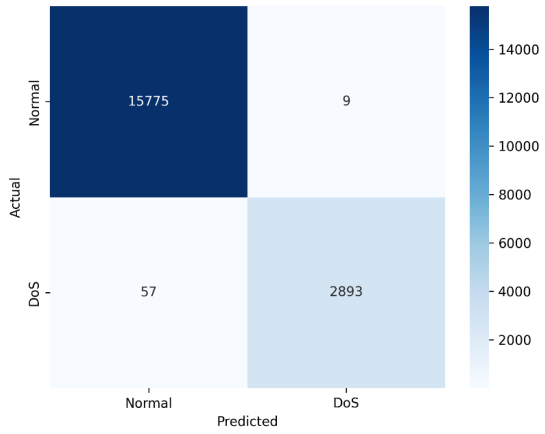
Evaluasi dilakukan untuk menilai kinerja model PPO setelah proses pelatihan pada dataset BnaT, dengan fokus pada kemampuan klasifikasi lalu lintas jaringan blockchain ke dalam dua kelas, yaitu Normal dan DoS. Kinerja model dievaluasi menggunakan metrik klasifikasi standar, meliputi *accuracy, precision, recall, F1-score,* dan *Area Under the Curve* (AUC), yang merepresentasikan ketepatan prediksi serta kemampuan model dalam mengidentifikasi kelas DoS pada distribusi data yang tidak seimbang. Selain itu, *Precision-Recall Curve* (RPC) dan *Receiver Operating Characteristic* (ROC) dianalisis untuk mengevaluasi stabilitas prediksi model pada berbagai ambang keputusan, sehingga memberikan gambaran komprehensif mengenai kemampuan generalisasi model dalam klasifikasi lalu lintas jaringan blockchain.

Hasil Hasil pengujian model terhadap test set disajikan pada Table V. Model PPO menunjukkan performa klasifikasi yang sangat tinggi dengan *accuracy* keseluruhan sebesar 0.9965, *precision* rata-rata tertimbang sebesar 0.9965, *recall* sebesar 0.9965, serta *F1-score* sebesar 0.9965. Untuk kelas Normal, model mencapai *F1-score* sebesar 0.9979, sementara kelas DoS memperoleh nilai 0.9887, yang menunjukkan kemampuan model dalam membedakan lalu lintas normal dan serangan DoS dengan tingkat kesalahan yang sangat rendah. Konsistensi kinerja di seluruh metrik mengindikasikan efektivitas arsitektur PPO dalam menjaga stabilitas pembaruan parameter selama proses pembelajaran kebijakan.

TABEL V
CLASSIFICATION REPORT MODEL PPO PADA SET DATA UJI

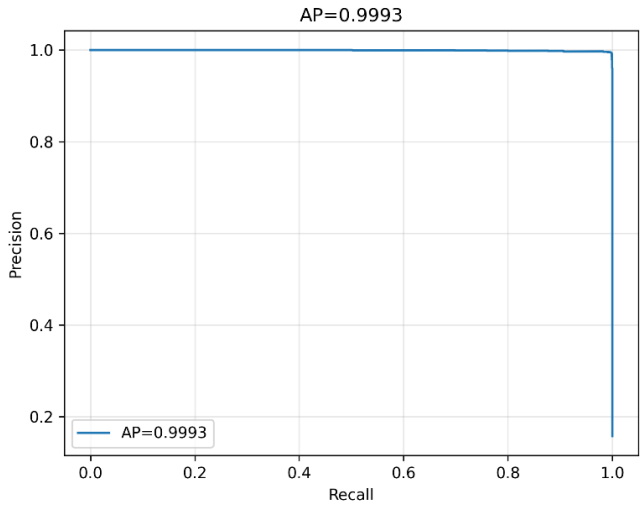
Label	Precision	Recall	F1-score	Support
Normal	0,9964	0,9994	0,9979	15.785
DoS	0,9969	0,9807	0,9887	2950
Accuracy			0,9965	18.734
Macro Avg	0,9966	0,9901	0,9933	18.734
Weighted Avg	0,9965	0,9965	0,9965	18.734

Distribusi prediksi hasil klasifikasi divisualisasikan pada Gbr. 9 melalui *confusion matrix* pada data uji. Dari total 18,734 sampel, model berhasil mengklasifikasikan 15,775 sampel Normal secara benar dengan hanya sembilan kesalahan prediksi, serta 2,893 sampel DoS dengan tingkat ketepatan yang tinggi. Hasil ini menunjukkan kemampuan model dalam mempertahankan *true positive rate* yang tinggi pada kelas DoS sekaligus menjaga *false positive rate* pada tingkat yang sangat rendah. Keseimbangan ini menegaskan efektivitas PPO dalam tugas klasifikasi biner lalu lintas jaringan blockchain.



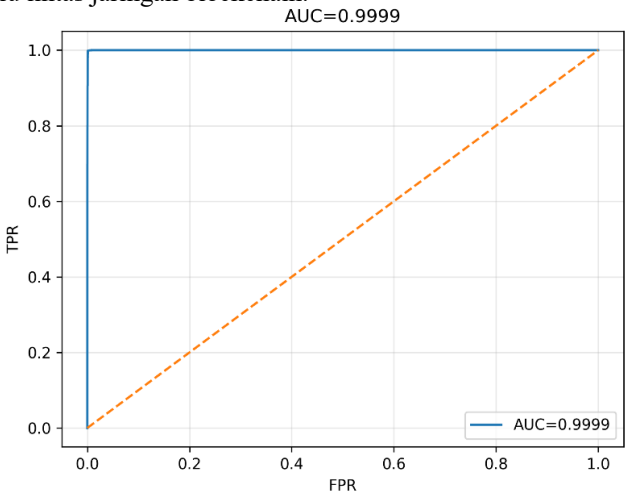
Gbr. 9 Confusion Matrix Hasil Pengujian Model PPO

Kinerja model dalam menjaga keseimbangan antara *precision* dan *recall* dievaluasi menggunakan PRC yang ditunjukkan pada Gbr. 10. Model PPO mencapai nilai *Average Precision* (AP) sebesar 0.9993, yang mencerminkan konsistensi kinerja klasifikasi pada berbagai ambang batas keputusan. Kurva PRC yang berada dekat dengan area maksimum menunjukkan bahwa model mampu mempertahankan tingkat *precision* yang tinggi tanpa mengorbankan *recall,* bahkan ketika ambang keputusan divariasikan. Hasil ini mengindikasikan bahwa PPO menunjukkan ketahanan yang baik terhadap ketidakseimbangan distribusi kelas serta mampu mempertahankan performa klasifikasi yang stabil pada kelas DoS tanpa penurunan sensitivitas yang signifikan.



Gbr. 10 Kurva Precision–Recall Model PPO

Kemampuan model dalam membedakan dua kelas target dievaluasi menggunakan ROC yang ditunjukkan pada Gbr. 11. Model PPO memperoleh nilai AUC sebesar 0.9999, yang menunjukkan tingkat separabilitas yang sangat tinggi antara lalu lintas Normal dan DoS. Kurva ROC yang mendekati sudut kiri atas mengindikasikan kombinasi *True Positive Rate* (TPR) yang tinggi dengan *False Positive Rate* (FPR) yang rendah pada berbagai ambang keputusan. Hasil ini menunjukkan bahwa model PPO mampu mempertahankan kinerja klasifikasi yang konsisten dan stabil, serta memperkuat efektivitas pendekatan pembelajaran penguatan dalam tugas klasifikasi lalu lintas jaringan blockchain.



Gbr. 11 Kurva ROC dan AUC Model PPO

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model PPO mencapai kinerja yang tinggi dan stabil dalam klasifikasi lalu lintas jaringan blockchain. Nilai *F1-score,* AP, dan AUC yang tinggi menunjukkan kemampuan model dalam membedakan lalu lintas normal dan serangan DoS secara konsisten. Kesesuaian antara hasil pelatihan dan evaluasi mengindikasikan bahwa model tidak mengalami overfitting dan mampu mempertahankan kemampuan generalisasi yang baik terhadap data yang tidak terlihat selama pelatihan. Dengan konvergensi yang stabil dan performa yang konsisten pada berbagai metrik evaluasi, pendekatan PPO menunjukkan potensi yang kuat sebagai metode pembelajaran penguatan untuk tugas klasifikasi lalu lintas jaringan blockchain.

IV. KESIMPULAN

Model PPO telah dirancang dan diimplementasikan untuk melakukan klasifikasi serangan DoS pada jaringan blockchain menggunakan dataset BNaT. Dengan memformulasikan tugas klasifikasi dalam kerangka pembelajaran penguatan berbasis arsitektur *Actor–Critic*, model mencapai proses pelatihan yang stabil dan konvergen. Mekanisme pembaruan kebijakan berbasis *clipped surrogate objective* memungkinkan PPO menjaga keseimbangan yang efektif antara eksplorasi dan eksploitasi, sehingga mampu membedakan lalu lintas jaringan normal dan DoS secara konsisten.

Hasil evaluasi menunjukkan bahwa model PPO mencapai kinerja klasifikasi yang sangat tinggi dengan *accuracy dan F1-score* sebesar 0,9965, serta nilai AP dan AUC masing-masing mencapai 0,9993 dan 0,9999. Konsistensi nilai pada seluruh metrik tersebut mencerminkan kemampuan diskriminatif yang kuat serta stabilitas pembelajaran yang baik, tanpa indikasi overfitting. Temuan ini menegaskan bahwa algoritma PPO merupakan pendekatan pembelajaran penguatan yang efektif dan andal untuk klasifikasi lalu lintas jaringan blockchain,

khususnya dalam membedakan aktivitas normal dan serangan Denial of Service (DoS).

V. SARAN

Penelitian selanjutnya dapat mengeksplorasi penerapan PPO pada skenario multikelas dengan cakupan jenis serangan yang lebih beragam serta integrasi dengan arsitektur pembelajaran mendalam lanjutan, seperti *Graph Neural Network* dan *Transformer-based Network*, untuk meningkatkan representasi pola komunikasi blockchain. Selain itu, optimisasi *hyperparameter* otomatis dan evaluasi pada lalu lintas blockchain *real-time* perlu dilakukan gunak menilai skalabilitas, efisiensi dan kelayakan pendekatan ini dalam lingkungan operasional berskala.

REFERENSI

[1] K. K. Vaigandla, M. Siluveru, M. Kesoju, and R. Karne, "Review on Blockchain Technology: Architecture, Characteristics, Benefits, Algorithms, Challenges and Applications," *Mesopotamian J. CyberSecurity*, vol. 2023, pp. 73–84, 2023, doi: 10.58496/MJCS/2023/012.

[2] J. Ahn, E. Yi, and M. Kim, "Blockchain Consensus Mechanisms : A Bibliometric Analysis (2014 – 2024) Using VOSviewer and R Bibliometrix Blockchain Consensus Mechanisms : A Bibliometric Analysis," *Information*, vol. 15, no. 10, p. 644, 2024, [Online]. Available: https://www.mdpi.com/2078-2489/15/10/644

[3] M. Raikwar and D. Gligoroski, "DoS Attacks on Blockchain Ecosystem," *Lect. Notes Comput. Sci.*, vol. 13098, 2022, [Online]. Available: https://doi.org/10.1007/978-3-031-06156-1_19

[4] D. Grandjean, L. Heimbach, and R. Wattenhofer, "Ethereum Proof-of-Stake Consensus Layer: Participation and Decentralization," *Blockchain – ICBC 2023 Lect. Notes Comput. Sci.*, vol. 14222, no. September 2022, pp. 253–280, 2023.

[5] M. M. Rasheed, A. K. Faieq, and A. A. Hashim, "Development of a new system to detect denial of service attack using machine learning classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 1068–1072, 2021, doi: 10.11591/ijeecs.v23.i2.pp1068-1072.

[6] S. Veluchamy and R. S. Kathavarayan, "Deep reinforcement learning for building honeypots against runtime DoS attack," *Int. J. Intell. Syst.*, vol. 37, no. 7, pp. 3981–4007, 2022, doi: 10.1002/int.22708.

[7] X. Wang and L. Liu, "Risk-Sensitive Deep Reinforcement Learning for Portfolio Optimization," *Risk Financ. Manag.*, vol. 18, no. 7, pp. 1–22, 2025, [Online]. Available: https://doi.org/10.3390/jrfm18070347

[8] W. Meng, Q. Zheng, G. Pan, and Y. Yin, "Off-Policy Proximal Policy Optimization," *Proc. 37th AAAI Conf. Artif. Intell. AAAI 2023*, vol. 37, pp. 9162–9170, 2023, doi: 10.1609/aaai.v37i8.26099.

[9] J. Queeney, I. C. Paschalidis, and C. G. Cassandras, "Generalized Proximal Policy Optimization with Sample Reuse," *Adv. Neural Inf. Process. Syst.*, vol. 15, no. NeurIPS, pp. 11909–11919, 2021.

[10] W. Deng, C. Huang, and Q. Shuai, "Double reinforcement learning for cluster synchronization of Boolean control networks under denial of service attacks," *PLoS One*, vol. 20, no. 7 July, pp. 1–21, 2025, doi: 10.1371/journal.pone.0327252.

[11] J. F. C. Garcia and G. E. T. Blandon, "A Deep Learning-Based Intrusion Detection and Preventation System for Detecting and Preventing Denial-of-Service Attacks," *IEEE Access*, vol. 10, no. August, pp. 83043–83060, 2022, doi: 10.1109/ACCESS.2022.3196642.

[12] T. V. Khoa *et al.*, "Collaborative Learning for Cyberattack Detection in Blockchain Networks," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 54, no. 7, pp. 3920–3933, 2024, doi: 10.1109/TSMC.2024.3374280.