International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 1173-1187
https://doi.org/10.70917/ijcisim-2026-4511
Article

Fedagentshield: An Adaptive Federated Reinforcement Learning Framework for Privacy-Preserving Autonomous Cyber Defence

Supriya P Belkar [1], Pilli Mounika[2], Nagalakshmi Panchakatla [3], G. Kalpana[4], Sanjeevini Harwalkar [5], T G Ramnadh Babu[6]

[1]Department of CSE, KLE Technological University, Udyambagh, Karnataka, India- 580031
[2]Department of CSE, R.V.R.&J.C College of Engineering, Guntur, Andhra Pradesh, India-522307
[3]Department of CSE, Malla Reddy Engineering College for Women (Autonomous), Hyderabad, Telangana, India-500100, lakshmip29@gmail.com
[4]Department of CSE, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad, Telangana, India-500043
[5]Department of CSD, SHARNBASVA UNIVERSITY, Kalaburagi, Karnataka, India-585103, sanjeevini706@gmail.com
[6]Department of CSE, R.V.R.&J.C College of Engineering, Guntur, Andhra Pradesh, India-522307.

Abstract: The growing sophistication and frequency of cyber threats, including Advanced Persistent Threats, ransomware, zero-day exploits, and AI-assisted attacks, highlight the limitations of traditional centralised cybersecurity systems in scalability, adaptability, and data privacy. Federated Learning (FL) and Reinforcement Learning (RL) have shown significant promise for distributed learning and intelligent decision-making across a variety of applications. However, their synergistic combination for privacy-preserving autonomous cyber defence is still an open research issue. The paper proposes FedAgentShield, an adaptive Federated Reinforcement Learning (FRL) framework that enables multiple distributed organisations to collaboratively learn effective cyber defence policies without sharing cybersecurity data. The proposed scheme features a PPO-based autonomous defence agent and a federated learning architecture for decentralised policy learning in distributed cyber environments. The framework can enhance the quality of global policy updates by introducing an Adaptive Federated Policy Aggregation (AFPA) mechanism. The AFPA mechanism forms aggregation weights by leveraging multiple client performance indicators, including detection performance, reinforcement learning reward, response efficiency, and privacy confidence, rather than conventional data-size-based aggregation. We evaluate the proposed framework in a custom cyber defence environment with the NF-UNSW-NB15 benchmark dataset. Findings reveal that FedAgentShield enables collaborative policy learning while protecting user privacy and enhancing overall policy quality through adaptive weighting. The proposed structure is a scalable, adaptive, and privacy-preserving approach for next-generation autonomous cyber defence in distributed network environments.

Keywords: Federated Reinforcement Learning; Federated Learning; Autonomous Cyber Defence; Proximal Policy Optimisation; Privacy-Preserving Learning; Adaptive Federated Policy Aggregation; Network Intrusion Detection.

1. Introduction

A. Background

Due to the rapid digital transformation of modern organisations, the threat landscape has become more challenging than ever. Cyber threats are increasing in scale, frequency, and sophistication, targeting critical infrastructure, cloud computing platforms, IoT devices, and enterprise networks. Recent cyberattacks, such as advanced persistent threats (APTs), ransomware, zero-day attacks, and AI-based attacks, are becoming more sophisticated. That allows attackers to bypass detection by classical signature-based attacks and exploit vulnerabilities in real time. Cybersecurity systems must therefore evolve into intelligent, adaptive, and autonomous defence systems capable of continuously responding to a dynamic attack scenario, rather than merely passive monitoring solutions. In recent years, we have witnessed significant advancements in Artificial Intelligence (AI), enabling organisations to

Copyright: © 2026 by the authors

develop intelligent cybersecurity solutions that enhance intrusion and anomaly detection, malware analysis, and automated threat response.[1] Among AI techniques, Reinforcement Learning (RL) has emerged as a promising approach for autonomous cyber defence, enabling the intelligent agent to learn optimal cyber defence actions through interaction with the environment. Compared with reinforcement learning (RL), supervised learning uses static datasets and predefined labels, whereas RL algorithms continuously improve their policies based on feedback from the environment. Thus, it is a good candidate for any dynamic cybersecurity environment. PPO is one of the most widely used RL algorithms due to its stable learning performance, efficient policy optimisation, and strong convergence properties, all of which make it well-suited to sequential cyber defence decision-making.[2] [3], [4]

While cybersecurity developments have occurred, organisations remain hesitant to share cybersecurity data due to privacy regulations, confidentiality requirements, and organisational security policies. Standard centralised machine learning procedures require sensitive network traffic and security logs to be collected in a single location for model training, leading to data leaks, limited cooperation among disparate organisations, and increased reliance on centralised architectures.[5] Federated Learning (FL) solves this problem by enabling multiple participants to jointly train a global model without sharing raw data. This type of learning preserves organisations' data privacy. It enhances regulatory compliance.[6] [7]Similarly, it fosters collaborative intelligence without disclosing sensitive information. The combination of Federated Learning and Reinforcement Learning appears to be a promising direction for developing an intelligent, privacy-preserving autonomous cyber defence system. Using Federated Reinforcement Learning, organisations can collaborate to enhance policies without compromising any organisation's data privacy. Nonetheless, traditional federated aggregation methods, especially Federated Averaging (FedAvg), assign fault contributions based purely on local dataset size. These methods do not account for the quality of the learned defence policies or the client's performance. As a result, clients displaying better threat detection capability, reinforcement learning performance, or operational efficiency are not necessarily contributing to the global policy and may therefore impair collaborative learning.[4], [8], [9]

To address these limitations, we propose FedAgentShield, an adaptive Federated Reinforcement Learning framework for autonomous cyber defence that preserves privacy. This work proposes a PPO-based framework for autonomous cyber defence, with a federated learning architecture that jointly learns effective cyber defence policies across multiple distributed organisations without exchanging sensitive network traffic. In addition, the framework features an Adaptive Federated Policy Aggregation (AFPA) mechanism. This mechanism computes real-time aggregation weights by drawing on several key performance indicators (KPIs) the client uses. The proposed framework enhances the efficacy of global policy learning by focusing on policy quality rather than data-size-based aggregation, while simultaneously protecting the organisation's data privacy. The framework is evaluated using the NF-UNSW-NB15 benchmark dataset in a custom cyber defence environment.[10], [11]

B. Problem Statement

Conventional cybersecurity engines heavily depend on a centralised machine-learning architecture in which organisations share their intrusion network traffic data to train the model. Although centralised learning can yield good detection performance, it poses significant privacy risks and increases the risk of data exposure and violations of organisations' or regulators' policies. Federated Learning Solutions address these difficulties by enabling joint model training without data transfer. Nonetheless, existing Federated Reinforcement Learning work often employs Federated Averaging (FedAvg), which aggregates local models primarily based on dataset size rather than the quality of the learned RL policies. Organisations may learn effective cyber defence strategies, but their influence on global policy is limited. That undermines the effectiveness of collaborative learning and the quality of the aggregated model.[2], [12]

C. Research Gap

Existing studies have conducted extensive research on reinforcement learning for autonomous cyber defence and on federated learning for privacy-preserving intrusion detection. Nonetheless, a limited number of works have integrated these technologies into a Federated Reinforcement Learning framework for autonomous cyber defence.[2] Moreover, traditional federated aggregation approaches usually depend on static aggregation strategies. FedAvg is an example of this, as it determines client contributions mostly based on dataset size while ignoring the quality of the reinforcement learning policy and operational performance. Performance indicators such as detection effectiveness, cumulative reinforcement-learning reward, response efficiency, and privacy confidence are ignored during policy aggregation. As a result, there remains a need for an adaptive Federated Reinforcement Learning framework that can

adjust local policies based on clients' actual performance while ensuring the privacy of organisations' data.[4], [13], [14]

D. Proposed Solution

To address the research gap identified earlier, this paper presents FedAgentShield, an adaptive Federated Reinforcement Learning framework for privacy-preserving autonomous cyber defence. The proposed framework utilises a local reinforcement learning algorithm, Proximal Policy Optimisation (PPO), to help autonomous cyber defence agents learn effective defence strategies in a bespoke cyber defence environment. Various organisations train independent local PPO policies using their private cybersecurity data, while only the model parameters are exchanged via a federated learning server. The framework also includes the AFPA mechanism, which dynamically estimates aggregation weights for detection performance, reinforcement learning reward, response efficiency, and privacy confidence, rather than relying mainly on data-size-based aggregation. This dynamic aggregation method produces higher-quality global policy updates while preserving the privacy of organisational data, enabling scalable collaborative cyber defence.

E. Contributions

The essential contributions of this study are summarised as follows.

1. A novel FedAgentShield framework is proposed to allow the integration of Federated Reinforcement Learning with autonomous cyber defence. It enables different organisations to co-operatively learn cyber defence policies without sharing cybersecurity data.
2. The FedAgentShield framework proposed in this paper employs an Adaptive Federated Policy Aggregation (AFPA) mechanism that dynamically computes aggregation weights based on detection performance, reinforcement learning reward, response efficiency, and privacy confidence. AFPA thus improves global policy aggregation more effectively than FedAvg, which relies on data-size-based weighting.
3. Utilising the NF-UNSW-NB15 Benchmark Dataset to Model Autonomous Cyber Defence as a Reinforcement Learning Problem for Federated Policy Learning: A custom cyber defence environment is created.
4. The proposed FedAgentShield framework is extensively evaluated on multiple fronts, including reinforcement learning performance, cybersecurity detection metrics, aggregation behaviour, and privacy-preserving federated learning, all developed in the cyber defence environment.

2. Related Work

A. Federated Learning for Cybersecurity

Federated Learning (FL) enables a distributed machine learning paradigm that allows for the shared training of a model while keeping training data decentralised. In contrast to centralised learning methods that require organisations to share raw cybersecurity data, FL allows participating clients to train local models independently and transmit only model parameters to the central server.[15] The Strategy considerably mitigates privacy risks and enables collaborative intrusion detection by geographically distributed organisations. Recent research has shown that FL can be efficient for network intrusion detection, malware classification, IoT security, and distributed anomaly detection. Moreover, FL enhances the model's generalisation by leveraging heterogeneous data distributions. Most existing FL-based cyber defence frameworks, however, utilise the conventional Federated Averaging (FedAvg) algorithm, which aggregates local models mainly based on dataset size. Strategies for aggregation often overlook differences in local model quality, threat-detection capabilities, and homegrown attack distributions, thereby limiting their real-world relevance. [16], [17]

B. Reinforcement Learning for Autonomous Cyber Defence

Reinforcement Learning (RL) has gained popularity in autonomous cyber defence for enabling intelligent agents to learn optimal defence strategies through continuous interaction in dynamic network environments. In contrast to supervised learning approaches, which require labelled datasets for training, RL agents learn defence policies by receiving rewards for taking security-related actions. Thus, they are a suitable option for intrusion response and threat mitigation. A few studies have employed DQN, DDPG, A2C, and PPO for intrusion detection, malware defence, moving-target defence, and network resource protection. Among these algorithms, PPO has been shown to

converge more stably, improve policies, and perform well across a range of sequential decision-making tasks. Nonetheless, most existing RL-based cyber defence systems are isolated, do not support cooperation and collaborative learning across organisations, and thus limit scalability and knowledge sharing.[3], [18]

C. Federated Reinforcement Learning and Research Gap

Federated Reinforcement Learning (FRL) integrates the collaborative learning of Federated Learning with the adaptive decision-making of Reinforcement Learning to improve defence policies through distributed agents without exposing local data. Recent investigations in the Flexible Radio Link (FRL) have investigated distributed policy optimisation, decentralised intelligence at the edge, autonomous network management, and privacy-preserving decision-making.[19] Even though the approaches show that collaborative reinforcement learning is indeed possible, most of them use standard aggregation methods, e.g., FedAvg, which are designed for other paradigms that assume equal importance among clients. They do not account for the quality of the policies learned at each client. As a result, organisations that perform better at threat detection or deliver more effective responses contribute just as much as weaker actors do, thereby reducing the overall effectiveness of global policy.[20] Moreover, research on adaptive policy aggregation for autonomous cyber defence using reinforcement learning is scarce. There remains a need for an adaptive federated reinforcement learning framework that dynamically aggregates local defence policies, ensures privacy-preserving interactions, and improves collaborative learning across diverse cyber environments.[12], [21]

Table 1. Summary of Related Work

Approach	Technique	Strengths	Limitations
Federated Intrusion Detection[22]	Federated Learning	Privacy-preserving collaborative learning	Uses FedAvg; ignores policy quality
RL-based Cyber Defense[3]	PPO / DQN	Adaptive autonomous defense	Centralised learning only
Federated Reinforcement Learning [24]	FRL	Distributed policy learning	Conventional aggregation strategy
Privacy-Preserving IDS[22]	Federated Learning	Data privacy protection	Limited adaptability in heterogeneous environments
Proposed FedAgentShield	PPO + Federated Learning + AFPA	Adaptive policy aggregation, privacy preservation, autonomous cyber defence	Addresses limitations of existing approaches

D. Research Gap Summary

The existing literature demonstrates that Federated Learning (FL) effectively enables collaborative model training while preserving data privacy, whereas Reinforcement Learning (RL) provides adaptive decision-making capabilities for autonomous cyber defence. The integration of these paradigms via Federated Reinforcement Learning (FRL) offers a promising approach to privacy-preserving distributed cyber defence. However, existing FRL approaches predominantly employ conventional aggregation strategies, such as Federated Averaging (FedAvg), that base client contributions primarily on local dataset size and do not account for the quality of locally learned reinforcement learning policies.[23] Consequently, important client performance indicators, including threat-detection performance, cumulative reinforcement-learning reward, response efficiency, and privacy confidence, are not incorporated into the global policy aggregation process. To address these limitations, this paper proposes FedAgentShield. This adaptive Federated Reinforcement Learning framework incorporates an Adaptive Federated Policy Aggregation (AFPA) mechanism that dynamically computes aggregation weights based on client performance metrics. By integrating adaptive policy aggregation with PPO-based federated reinforcement learning, the proposed framework supports privacy-preserving collaborative policy learning for autonomous cyber defence.[12], [24]

3. Proposed FedAgentShield Framework

A. Overall Framework Architecture

This paper proposes FedAgentShield, an adaptive Federated Reinforcement Learning (FRL) framework for privacy-preserving autonomous cyber defence. The framework enables multiple distributed clients to collaboratively learn autonomous cyber defence policies without exchanging raw cybersecurity data. Instead, each client independently trains a local reinforcement learning policy using its own network traffic data. At the same time, only the learned model parameters are transmitted to a federated learning server for global policy aggregation. The proposed framework integrates Proximal Policy Optimisation (PPO) with Federated Learning (FL) to support decentralised reinforcement learning for cybersecurity applications. Each participating client trains a PPO-based cyber defence agent in a custom Gymnasium environment built using the NF-UNSW-NB15 benchmark dataset. After local training, the learned policy parameters are transmitted to the federated server, where the proposed Adaptive Federated Policy Aggregation (AFPA) mechanism computes aggregation weights based on client performance measurements. The aggregated parameters are used to update the global PPO policy, which is redistributed to all participating clients for subsequent federated learning rounds.[10]

Unlike conventional federated learning approaches that aggregate client updates primarily based on local dataset size, the proposed framework incorporates multiple client performance indicators into policy aggregation. Specifically, AFPA evaluates each client's detection performance, normalised reinforcement learning reward, response efficiency, and privacy confidence to determine its contribution to the global policy. This adaptive aggregation mechanism improves the quality of global policy updates while preserving data privacy by ensuring that raw cybersecurity data remain exclusively at the client side throughout the federated learning process. The proposed FedAgentShield framework consists of four major components:

1. Distributed PPO-based cyber defence agents,
2. A custom Gymnasium cyber defence environment,
3. Federated learning server, and
4. Proposed Adaptive Federated Policy Aggregation (AFPA) mechanism.

B. FedAgentShield Workflow

The working steps of the proposed FedAgentShield framework are presented in the following sequence.

Step 1: The NF-UNSW-NB15 dataset is processed by removing unnecessary attributes, selecting numerical attributes, encoding attack labels for detection, and splitting it into training and test sets.

Step 2: The training dataset is divided into several client partitions, technically, which indicates that participants are distributed. Throughout the training process, each client retains its local training data.

Step 3: Once ready, each client initialises an independent PPO-based cyber defence agent that interacts with the customised Gymnasium cyber defence environment. Throughout the training phase, the agent can observe the network traffic state, select defence actions, and receive a reward based on the effectiveness of those actions.

Step 4: Once local PPO training is performed for each client, each client sends only its learned policy parameters to the server. Raw cybersecurity data is kept on the client's local system and never shared.

Step 5: The federated server uses the Adaptive Federated Policy Aggregation (AFPA) system to evaluate the performance of different clients based on estimation performance, normalised reinforcement learning reward, response efficiency, and privacy confidence. The weights of the aggregators are computed, and the weighted local policy parameters are aggregated to produce a global PPO policy.

Step 6: The new worldwide policy is returned to participating customers for evaluation in a future federated learning run.

The proposed framework's full workflow is shown in the given figure.

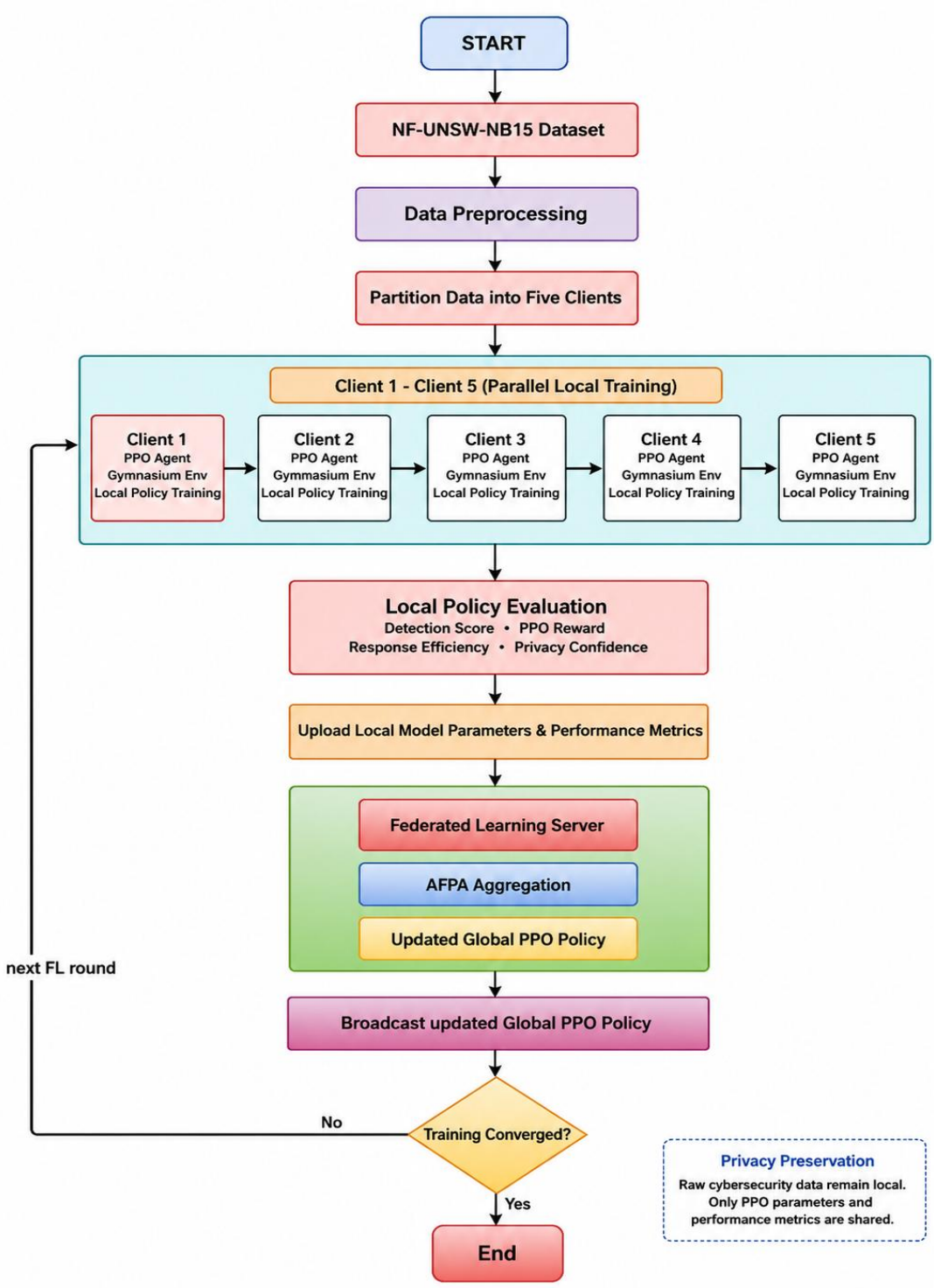

Figure 1: Overall architecture of the proposed FedAgentShield framework.

C. Cyber Defence Environment

We developed a custom-built Gymnasium environment using the NF-UNSW-NB15 benchmark dataset to evaluate autonomous cyber defence policies. Each observation corresponds to a network traffic flow, represented by 41 numerical features describing communication characteristics, traffic statistics, and protocol behaviour. In every interaction, the PPO agent observes the network's current state and selects one of the four available cyber defence

actions. The environment assesses the executed action by examining the corresponding label in the underlying network traffic and returns a reward. Positive rewards are given for correct defensive actions, and penalties are assigned for inappropriate ones. The PPO agent should therefore be encouraged to maximise cumulative rewards over the long run while learning effective autonomous cyber defence strategies.[10]

Table II. Action Space

Action ID	Defense Action	Description
0	Monitor	Continue monitoring network traffic without intervention
1	Alert	Generate a security alert for suspicious traffic
2	Block	Block malicious network traffic
3	Quarantine	Isolate suspicious network traffic or affected hosts

D. PPO-Based Local Policy Learning

Every client that participates in the system is enhanced by enabling each local cyber defence agent to train using Proximal Policy Optimisation (PPO). PPO is a policy-gradient reinforcement learning technique that ensures stability by regulating the extent of policy updates. Through continuous interaction with the cyber defence environment, the PPO agent learns an effective mapping between observed network traffic states and defensive actions. At each decision step, the agent observes the current environment state. , selects an action according to the policy , and receives an immediate reward . The objective of PPO is to maximise the cumulative discounted reward while maintaining stable policy updates. statπθat|strt

The cumulative discounted reward is defined as:

$$\boldsymbol{R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} - (1)}$$

where:

- R_t is the discounted reward accumulated over time.;
- r_{t+k} $denotes\ the\ reward\ obtained\ at\ the\ future\ time\ step\ t+k$;
- $\gamma$is the discount factor, where $0 < \gamma < 1$.

The objective function of PPO is given as

$$\boldsymbol{L^{PPO}(\theta) = \mathbb{E}\left[\min\left(r_{t'}(\theta)\widehat{A}_t \text{clip}(r_{t'}(\theta)', 1-\epsilon 1+\epsilon)\widehat{A}_t\right)\right] - (2))}$$

where:

- $The\ probability\ ratio\ between\ the\ new\ and\ old\ policies\ is\ denoted\ by\ r_t(\theta)$
- $\hat{A}_t$ represents the estimated advantage function.
- $\epsilon\ The\ clipping\ parameter\ to\ control\ policy\ updates$

Upon completing local PPO training, each client forwards just the learned policy parameters to the federated learning server for global aggregation.

E. Adaptive Federated Policy Aggregation (AFPA)

The FedAgentShield framework proposes Adaptive Federated Policy Aggregation (AFPA) to enhance global policy aggregation. AFPA, unlike the traditional method, Federated Averaging, which aggregates in proportion to the size of local datasets, uses clients' performance measurements to calculate aggregation weights. After training the local PPO, four evaluation metrics are used to assess the client.

- Detection Performance (D_i)

- Normalised Reinforcement Learning Reward (R_i)
- Response Efficiency (E_i)
- Privacy Confidence (P_i)

Response efficiency is computed as

$$\boldsymbol{E_i = \frac{1}{1 + RT_i}} - (\mathbf{3})$$

where RT_i denotes the client's measured response time. Under the current scheme, all clients involved retain their raw cybersecurity data locally throughout the federated learning process. Consequently, all clients receive the same privacy confidence score.

The overall client performance score is computed as

$$\boldsymbol{S_i = \alpha D_i + \beta R_i + \gamma E_i + \delta P_i}\text{-(4)}$$

subject to

$$\boldsymbol{\alpha + \beta + \gamma + \delta = 1}\text{-(5)}$$

where:

- α represents the detection performance weight;
- β represents the reinforcement learning reward weight;
- γrepresents the response efficiency weight;
- δrepresents the privacy confidence weight.

The aggregation weight assigned to each client is calculated as

$$\boldsymbol{w_i = \frac{S_i}{\sum_{k=1}^{N} S_k}} - (\mathbf{6})$$

ensuring that

$$\sum_{i=1}^{N} \boldsymbol{w_i = 1} - (\mathbf{7})$$

The global PPO policy is then updated using weighted parameter aggregation:

$$\boldsymbol{\theta^{t+1} = \sum_{i=1}^{N} w_i \, \theta_i^t} - (\mathbf{8})$$

where θ_i^trepresents the locally trained PPO policy parameters of client i, and θ^{t+1}denotes the updated global policy distributed to all participating clients.

F. Computational Complexity

Assume that the federated learning system consists of Nparticipating clients, each performing PPO training over Etraining episodes with Minteraction steps per episode.

The computational complexity of local PPO training is

$$\boldsymbol{O(NEM)}\text{-(9)}$$

while the AFPA aggregation process requires

$$\boldsymbol{O(N)}\text{-}\mathbf{(10)}$$

operations to compute client performance scores and aggregation weights. Since local reinforcement learning dominates the computational cost, the overall computational complexity of the proposed FedAgentShield framework is

$$\boldsymbol{O(NEM)} - \mathbf{(11)}$$

4. Experimental Setup

A. Experimental Environment

The Google Colab platform is used to implement the FedAgentShield framework in Python 3.11. Stable-Baselines3 with PPO was used for the reinforcement learning implementation, while a custom federated learning implementation handled collaborative policy learning across clients. We evaluated the framework using the NF-UNSW-NB15 benchmark dataset. Before building the model, all numerical features were standardised using StandardScaler, and an 80/20 train-test split was performed on the dataset. The datasets were subsequently divided into five federated clients, each of which independently trained a local PPO agent in a custom Gymnasium cyber defence environment. The agent perceives the current network state at every interaction step. Then it executes one system-level defensive action from the four given actions: Monitor, Alert, Block, and Quarantine. The agent receives a reward for the selected action, depending on whether the action is correct. Rewards that signal positive outcomes guide a good defence decision against cyberattacks, and wrong actions are penalised, enabling the PPO agent to learn a good policy through continuous interaction with the environment. Table III summarises the setup details, dataset configuration, and cyber defence action space.

Table III. Experimental Setup Configuration

Category	Parameter	Configuration
Implementation	Programming Language	Python 3.11
	Development Platform	Google Colab
	Deep Learning Framework	PyTorch
	Reinforcement Learning	Stable-Baselines3 (PPO)
	Federated Learning	Custom Federated Learning Implementation
	Environment	Gymnasium
	Data Processing	Pandas, NumPy, Scikit-learn
	Visualisation	Matplotlib
Dataset	Dataset	NF-UNSW-NB15
	Features	41 Numerical Features
	Label	Binary (0 = Benign, 1 = Attack)
	Attack Column	Multi-class Attack Categories
	Feature Scaling	StandardScaler
	Train-Test Split	80% / 20%
	Federated Clients	5
Action Space	Action 0	Monitor
	Action 1	Alert
	Action 2	Block
	Action 3	Quarantine

D. PPO Training Configuration

Using their private dataset, each client independently trains a local PPO policy in the custom Gymnasium cyber defence environment. After local training, only the parameters of the learned PPO policy are transmitted to the federated server, while the raw cybersecurity data remain on the client end. The federated server utilises the proposed Adaptive Federated Policy Aggregation (AFPA) mechanism to compute the adaptive aggregation weights based on

detection performance, normalised PPO reward, response efficiency and privacy confidence. Upon receiving the weights of the local PPO policies, the global policy is updated by aggregating them. Furthermore, the updated policy is sent to the clients for the next round. Table IV summarises the PPO training configuration, the federated learning settings, and the evaluation metrics.

Table IV: PPO Training, Federated Learning, and Evaluation Configuration

Category	Parameter	Configuration / Description
PPO Training	Reinforcement Learning Algorithm	Proximal Policy Optimisation (PPO)
	Policy Network	Multi-Layer Perceptron (MLP)
	Learning Framework	Stable-Baselines3
	Training Environment	Custom Gymnasium Cyber Defence Environment
	Local Training	Performed Independently at Each Client
Federated Learning	Number of Clients	5
	Local Learning Algorithm	PPO
	Global Aggregation Method	Adaptive Federated Policy Aggregation (AFPA)
	Shared Information	PPO Policy Parameters
	Raw Data Sharing	Not Permitted
	Privacy Mechanism	Federated Learning
Evaluation Metrics	Detection Performance	Measures the capability of the PPO agent to identify network traffic correctly
	Normalised PPO Reward	Evaluates the quality of the learned reinforcement learning policy
	Response Efficiency	Measures client responsiveness based on execution time
	Privacy Confidence	Represents privacy preservation during federated learning
	AFPA Policy Quality Score	Combined client performance score used for adaptive aggregation
	AFPA Aggregation Weight	Contribution of each client to the global PPO policy

5. Results and Discussion

A. Overall Performance of the Proposed FedAgentShield Framework

The proposed FedAgentShield framework was assessed using a five-client setup, with clients trained independently using PPO in the custom Gymnasium cyber defence environment. Each client was assessed using detection performance, reinforcement learning reward, response efficiency, privacy confidence, and its related Adaptive Federated Policy Aggregation (AFPA) weight after local training. The experimental results are summarised in Table V.

Table V: Local PPO Performance and AFPA Evaluation

Client	PPO Reward	Detection Score	Normalised Reward	Response Efficiency	Privacy Confidence	AFPA Weight
Client 1	1997	0.993	0.9906	0.5124	1.0000	0.1953
Client 2	2011	0.987	0.9975	0.5052	1.0000	0.1949
Client 3	2004	0.984	0.9940	0.6919	1.0000	0.2026
Client 4	2014	0.992	0.9990	0.7013	1.0000	0.2040
Client 5	2016	0.983	1.0000	0.6978	1.0000	0.2031

According to the results, the five clients learned to execute autonomous cyber defence policies. The cumulative PPO rewards from 1997 to 2016 and the normalised rewards above 0.99 for each client indicate a steady reinforcement learning function in a distributed environment. The detection scores ranged from 0.983 to 0.993, confirming the consistent threat-detection ability of locally trained PPO agents. The response efficiency was not uniform across clients due to differences in policy execution time. Moreover, clients 3-5 had higher efficiency values than clients 1 and 2.

Also, because all clients stored their raw CYBERSECURITY data locally throughout the federated learning process, all clients were assigned the same privacy confidence score of 1.0 in the current implementation. Using the measured performance indicators, the proposed AFPA mechanism computed adaptive aggregation weights ranging from 0.1949 to 0.2040. These weights enabled higher-performing clients to contribute more to the global PPO policy.

B. Confusion Matrix Analysis

To further assess the cyber defence capability of the trained PPO policy, a confusion matrix was generated by mapping the agent's actions to binary predictions. Based on the confusion matrix in Figure 2, the proposed framework correctly classified 952 benign network traffic instances and 6 attack instances. One benign sample was misclassified as malicious, and 41 attack samples were classified as benign. Findings show strong capability to classify normal traffic, allowing future work to focus on further enhancements to attack detection.

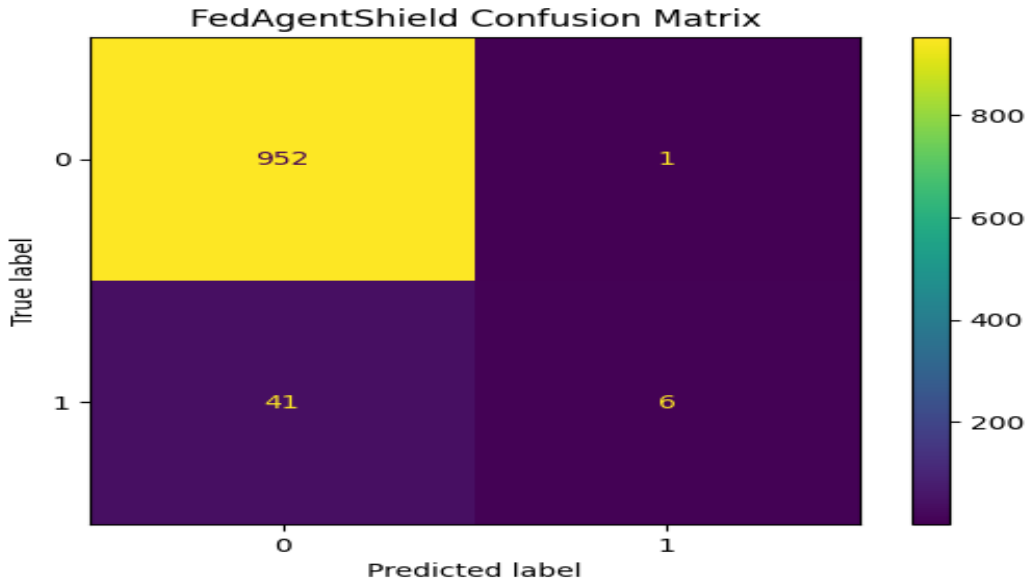

Figure 2. Confusion Matrix of the Proposed FedAgentShield Framework

C. Client Reward Analysis

The total PPO rewards received by the five clients who joined the game are shown in Figure 3. The cumulative reward values are consistently high, indicating that the PPO agents learned effective cyber defence policies for all five clients. Client 5 received the highest overall reward (2016), while clients 4, 2, 3, and 1 received rewards of 2014, 2011, 2004, and 1997, respectively. The small variation in reward values indicates that the PPO algorithm learns defence policies that are almost equally effective across all clients, demonstrating the stability of reinforcement learning in the proposed federated cyber defence settings.

Figure 3. Local PPO Rewards Across Federated Clients

D. Detection Performance

As depicted in Figure 4, the detection performance of each participating client is shown. The detection scores of all participating clients ranged from 0.983 to 0.993, indicating consistently high performance under the implemented evaluation metric. Client 1 recorded the highest detection score, 0.993, followed closely by Client 4 at 0.992. The remaining clients display only small differences in performance. According to the results, the locally trained PPO agents learned almost identical cyber defence policies and maintained stable performance.

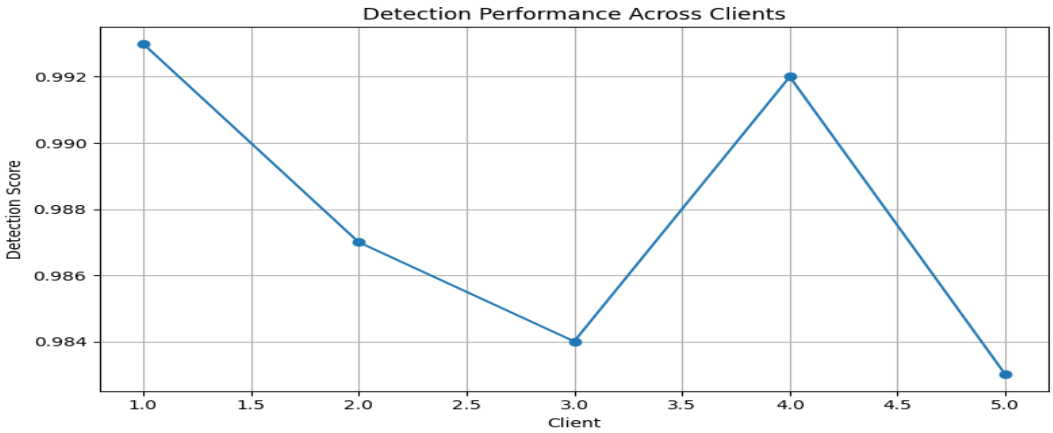

Figure 4. Detection Performance Across Federated Clients

E. Response Efficiency Analysis

The response efficiency values computed for each client are presented in Figure 5. The execution time for the policy evaluation for clients 1 and 2 was notably longer than for clients 3, 4, and 5. Client 4 has the highest response efficiency at 0.7013, followed by Client 5 at 0.6978 and Client 3 at 0.6919. Conversely, the efficiency values for Clients 1 and 2 are lower (0.5124 and 0.5052). The findings indicate that the AFPA mechanism can differentiate client performance based on client response efficiency. Moreover, faster, more efficient local policies contribute more to the universal policy when the mechanism is effective at handling downweighs.

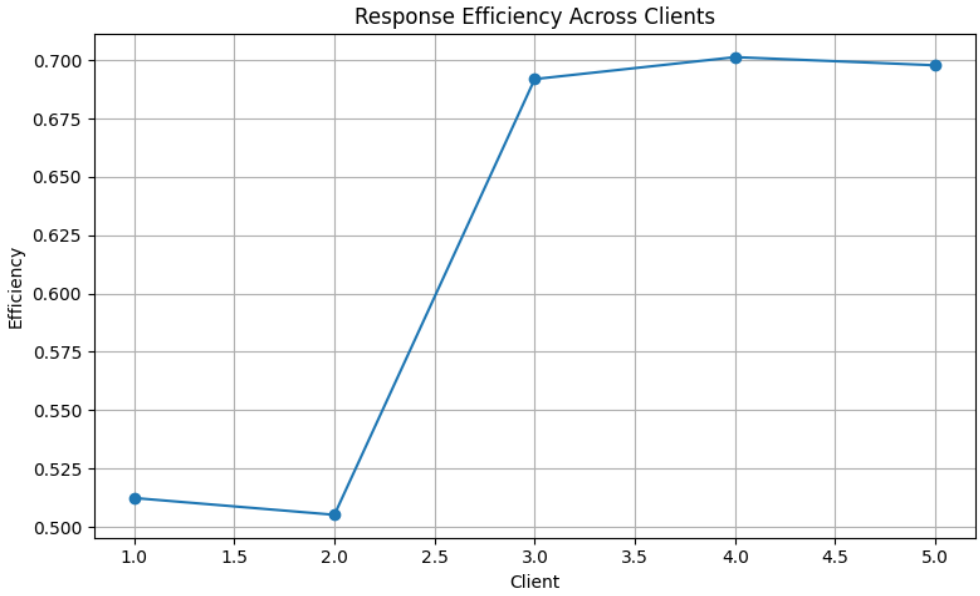

Figure 5. Response Efficiency Across Federated Clients

F. AFPA Aggregation Analysis

The graph in Figure 6 shows the adaptive aggregation weights computed using the proposed AFPA mechanism. The aggregation weights range from 0.1949 to 0.2040, indicating varying measured performance among the clients. Client 4 was assigned the highest aggregation weight owing to its superior overall performance on all evaluation indicators, while Client 2 was assigned the lowest weight. In contrast to classic FedAvg, which depends solely on dataset size, AFPA allocates client contributions based on measured policy quality, allowing higher-performing local PPO policies to exert greater influence over the global model.

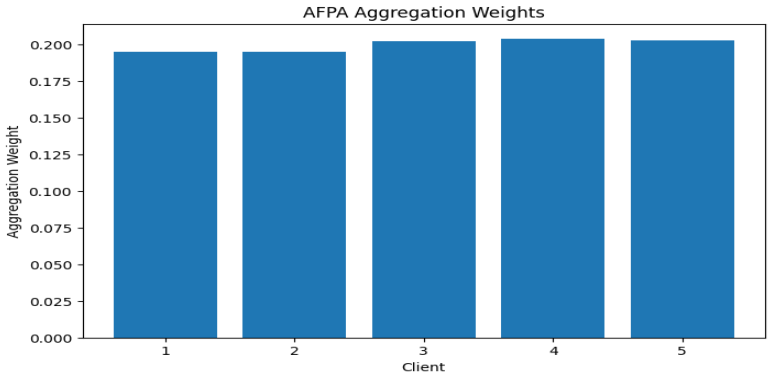

Figure 6. Adaptive Federated Policy Aggregation Weights

G. Summary of Experimental Results

Table VI summarises the average values from the five participating clients, providing an overall assessment of the framework. The overall results show that the proposed FedAgentShield framework successfully integrates federated learning with reinforcement learning to enable privacy-preserving autonomous cyber defence. The PPO rewards and detection scores are consistent, indicating that local policy learning is effective. The AFPA mechanism provides balanced, performance-aware aggregation, enabling clients to obtain adaptive weights based on measured policy performance.

Table VI Overall Experimental Summary

Metric	Value

Average PPO Reward	2008.40
Average Detection Score	0.9878
Average Response Efficiency	0.6217
Average AFPA Weight	0.2000

6. Conclusion

The proposed FedAgentShield is an adaptive Federated Reinforcement Learning (FRL) framework that ensures privacy in autonomous cyber defence. The framework combines Proximal Policy Optimisation (PPO) with Federated Learning (FL), enabling many distributed clients to learn cyber defence policies using PPO without sharing raw cybersecurity data. The Adaptive Federated Policy Aggregation (AFPA) mechanism organises global policy aggregation so that the aggregation weights are adaptive to measured client performance indicators, such as detection performance, RL reward (centred), response efficiency, and privacy guarantee. Using the Gymnasium cyber defence environment, the proposed framework was implemented and evaluated on the benchmark dataset NF-UNSW-NB15. The locally trained PPO agents consistently achieved high cumulative rewards and detection performance across all clients, as observed in the experimental results. In addition, the AFPA method can assign adaptive weights based on clients' performance. Consequently, it allows better-performing local policies to contribute more to the global policy than aggregation strategies based on dataset size and related factors. The experimental results indicate that the proposed FedAgentShield framework can be an effective solution for privacy-preserving collaborative reinforcement learning for autonomous cyber defence. The structure enables secure collaboration while protecting local data privacy and improving the quality of global policy updates by combining decentralised learning with adaptive policy aggregation. The current implementation tests the framework in a simulated federated learning setting with five clients. Future work will extend the framework to larger federated deployments, explore other Reinforcement Learning algorithms, assess the framework on different cybersecurity datasets, and investigate more sophisticated adaptive aggregation techniques for distributed autonomous cyber defence.

References

1. N. Kshetri, "Transforming cybersecurity with agentic AI to combat emerging cyber threats," Telecommun. Policy, vol. 49, no. 6, p. 102976, Jul. 2025, doi: 10.1016/j.telpol.2025.102976.
2. A. Molina-Markham, C. Miniter, B. Powell, and A. Ridley, "Network Environment Design for Autonomous Cyberdefense," Mar. 13, 2021, arXiv: arXiv:2103.07583. doi: 10.48550/arXiv.2103.07583.
3. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimisation Algorithms," 2017, arXiv. doi: 10.48550/ARXIV.1707.06347.
5. M. Towers et al., "Gymnasium: A Standard Interface for Reinforcement Learning Environments," 2024, arXiv. doi: 10.48550/ARXIV.2407.17032.
6. B. Ghimire and D. B. Rawat, "Recent Advances on Federated Learning for Cybersecurity and Cybersecurity for Federated Learning for Internet of Things," IEEE Internet Things J., vol. 9, no. 11, pp. 8229–8249, Jun. 2022, doi: 10.1109/JIOT.2022.3150363.
7. F. Balassone et al., "Cybersecurity AI: Evaluating Agentic Cybersecurity in Attack/Defense CTFs," 2025, arXiv. doi: 10.48550/ARXIV.2510.17521.
8. J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated Optimization: Distributed Machine Learning for On-Device Intelligence," Oct. 08, 2016, arXiv: arXiv:1610.02527. doi: 10.48550/arXiv.1610.02527.
9. R. Özbay, M. Çelebı, and U. Yavanoğlu, "The AI-Cybersecurity Nexus: How Large Language Models are Reshaping Threat Intelligence and Digital Defense," IEEE Access, vol. 14, pp. 15558–15587, 2026, doi: 10.1109/ACCESS.2026.3658308.
10. Y. Liu, X. Liu, H. Yu, B. Guo, and X. Liu, "A Federated Intrusion Detection System for Edge Environments Using Multi-Index Hashing and Attention-Based KNN," Symmetry, vol. 17, no. 9, p. 1580, Sep. 2025, doi: 10.3390/sym17091580.
11. N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia: IEEE, Nov. 2015, pp. 1–6. doi: 10.1109/MilCIS.2015.7348942.
12. Z. Ning and L. Xie, "A survey on multi-agent reinforcement learning and its application," J. Autom. Intell., vol. 3, no. 2, pp. 73–91, Jun. 2024, doi: 10.1016/j.jai.2024.02.003.
13. D. J. Beutel et al., "Flower: A Friendly Federated Learning Research Framework," 2020, arXiv. doi: 10.48550/ARXIV.2007.14390.

14. Q. Zhu, "Game Theory Meets LLM and Agentic AI: Reimagining Cybersecurity for the Age of Intelligent Threats," 2025, arXiv. doi: 10.48550/ARXIV.2507.10621.
15. I. Adabara, B. Olaniyi Sadiq, A. Nuhu Shuaibu, Y. Ibarahim Danjuma, and M. Venkateswarlu, "A Review of Agentic AI in Cybersecurity: Cognitive Autonomy, Ethical Governance, and Quantum-Resilient Defense," F1000Research, vol. 14, p. 843, Sep. 2025, doi: 10.12688/f1000research.169337.1.
16. M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke, and L. Shu, "Federated Deep Learning for Cyber Security in the Internet of Things: Concepts, Applications, and Experimental Analysis," IEEE Access, vol. 9, pp. 138509–138542, 2021, doi: 10.1109/ACCESS.2021.3118642.
17. B. Sarada, C Gazala Akhtar, Ramesh Wadawadagi, N. Shaleen Saroj, K Sandhya Rani, and Lakkadi Sphoorthi, "Dynamic Subgroup Adaptive Secure Multi-Modal Heterogeneous Federated Learning (DSAS-MHFL)," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 18, no. 6s, pp. 298–309, Jul. 2026, doi: 10.70917/ijcisim-2026-2888.
18. B. Sarada and S. Vaishnavi, "Federated Averaging with Multi-Label CNN for Robust Lung Disease Detection in Chest," in 2025 Eighth International Conference on Image Information Processing (ICIIP), Solan (Near Shimla) Himachal Pradesh, India: IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/ICIIP68302.2025.11346185.
19. A. K. Pati, "Agentic AI: A Comprehensive Survey of Technologies, Applications, and Societal Implications," IEEE Access, vol. 13, pp. 151824–151837, 2025, doi: 10.1109/ACCESS.2025.3585609.
20. E. Khramtsova, C. Hammerschmidt, S. Lagraa, and R. State, "Federated Learning For Cyber Security: SOC Collaboration For Malicious URL Detection," in 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS), Singapore, Singapore: IEEE, Nov. 2020, pp. 1316–1321. doi: 10.1109/ICDCS47774.2020.00171.
21. M. Alazab, S. P. Rm, P. M, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, "Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions," IEEE Trans. Ind. Inform., vol. 18, no. 5, pp. 3501–3509, May 2022, doi: 10.1109/TII.2021.3119038.
22. H. H. Zhuo, W. Feng, Y. Lin, Q. Xu, and Q. Yang, "Federated Deep Reinforcement Learning," Feb. 09, 2020, arXiv: arXiv:1901.08277. doi: 10.48550/arXiv.1901.08277.
23. E. M. Timofte et al., "Federated Learning for Cybersecurity: A Privacy-Preserving Approach," Appl. Sci., vol. 15, no. 12, p. 6878, Jun. 2025, doi: 10.3390/app15126878.
24. M. Rahmati and A. Pagano, "Federated Learning-Driven Cybersecurity Framework for IoT Networks with Privacy Preserving and Real-Time Threat Detection Capabilities," Informatics, vol. 12, no. 3, p. 62, Jul. 2025, doi: 10.3390/informatics12030062.
25. J. Qi, Q. Zhou, L. Lei, and K. Zheng, "Federated Reinforcement Learning: Techniques, Applications, and Open Challenges," Intell. Robot., 2021, doi: 10.20517/ir.2021.02.