Q-ID: A Reinforcement Learning Framework for Adaptive Intrusion Detection

Maisha Maliha, Mohammed Atiquzzaman
School of Computer Science
University of Oklahoma

Motivation

- Cyber threats growing in complexity & frequency
- Traditional IDS struggle with novel attacks
- Need: adaptive & intelligent intrusion detection

Problem

- Supervised models depend on large labeled datasets
- Assume static distributions
- Fail against new attack types
- Goal: adaptive, robust, and generalizable IDS

Our Contribution

- Explicit RL formulation: state, action, reward
- Hybrid training strategy: supervised + RL signals
- Extensive evaluation on CICIDS2017 dataset

Dataset (CICIDS2017)

- 2.8M records (83% benign, 17% attacks)
- Attack types: DoS, PortScan, DDoS, Web Attacks, Bot, etc.
- Class imbalance challenge
- Feature selection: Bwd Packet Length Std, Flow Bytes/s

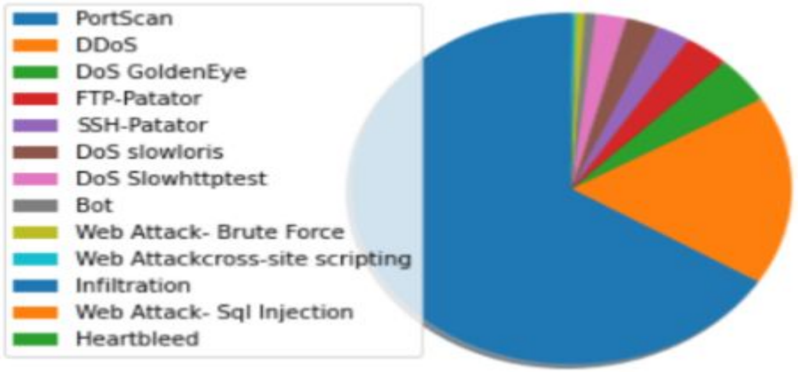

Fig. 1. Distribution of classes in the CICIDS2017 dataset.

Q-ID Method

- State = flow feature vector
- Action = classify as benign or attack type
- Reward = +1 correct, –1 wrong
- Hybrid Objective = Cross-entropy (supervised) + TD loss (RL)

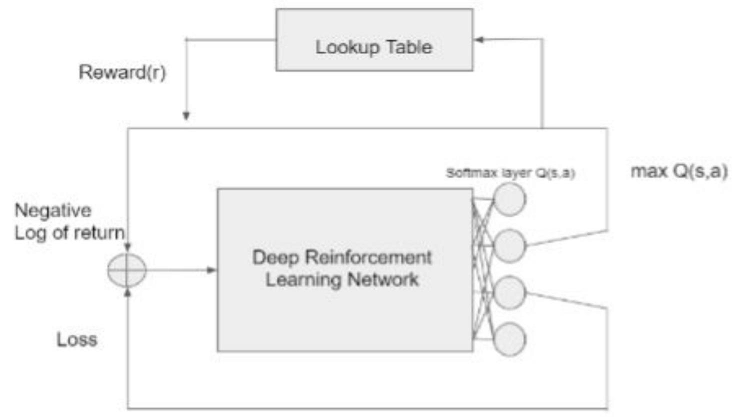

Fig. 2. End-to-end training and evaluation pipeline for the hybrid supervised+RL IDS.

Architecture

- Input ⟶ Fully connected layers (128 units)
- Gating + residual pathway for feature emphasis
- Output = Q-values (actions)
- Softmax only for supervised loss

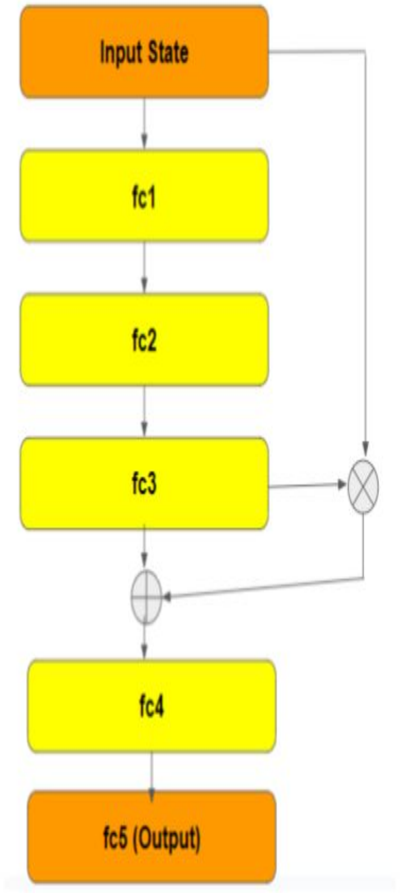

Fig. 3. Architecture of the proposed Q-network used by the RL module.

Results

- Accuracy = 99.3%
- Macro F1 = 0.982, Recall = 0.994
- Outperforms FT-Transformer, TabNet, CatBoost, XGBoost, LightGBM
- Low latency (0.07 ms/sample) → real-time feasible

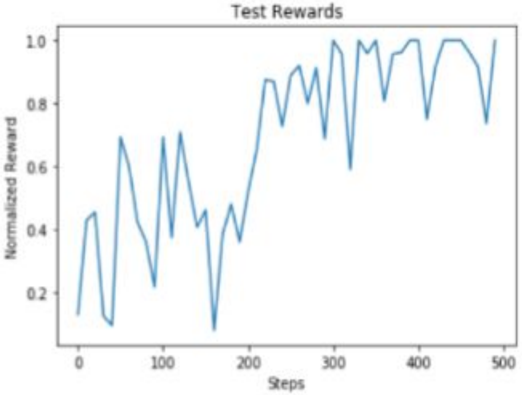

Fig. 4. Normalized reward versus training episodes/steps. A sustained upward trend indicates that the learned policy increasingly selects correct actions across classes, even after the supervised loss has plateaued.

Ablation Study

- Removing TD loss $\longrightarrow$ biggest drop in performance
- Class weighting & exploration critical for rare attacks
- Gating-residual helps stability
- Each component contributes to robustness

Ablation Study

TABLE I

COMPARISON WITH MODERN BASELINES ON THE CICIDS2017 EVALUATION SPLIT. BEST RESULTS PER COLUMN ARE IN **BOLD**. "LATENCY" IS SINGLE-SAMPLE INFERENCE TIME (MEDIAN)—DEEP MODELS ON A T4-CLASS GPU; TREE ENSEMBLES ON CPU (LOWER IS BETTER).

Model	Accuracy (%)	Macro F1	Macro Recall	Macro Precision	Macro AUROC	Macro PR-AUC	Latency (ms)
DRL (ours)	**99.3**	**0.982**	**0.994**	**0.991**	**0.999**	**0.997**	**0.07**
FT-Transformer	99.0	0.976	0.986	0.975	0.998	0.993	0.35
TabNet	98.8	0.972	0.983	0.971	0.997	0.991	0.60
CatBoost	98.7	0.971	0.978	0.972	0.998	0.990	0.12
XGBoost	98.5	0.968	0.975	0.970	0.997	0.988	0.18
LightGBM	98.6	0.969	0.974	0.971	0.997	0.989	0.08
ResMLP (5×128)	98.3	0.965	0.972	0.966	0.996	0.986	0.28
Random Forest	96.1	0.967	0.969	0.961	0.990	0.972	0.15
SVM (RBF)	85.0	0.830	0.852	0.851	0.910	0.740	1.20
KNN (k=5)	98.4	0.960	0.964	0.958	0.992	0.979	0.90

Conclusion

- DRL framework: adaptive IDS with high accuracy
- Handles imbalance & unseen attacks better than baselines
- Suitable for real-time network defense
- Future: model compression, explainability, continual learning