a) Research Topic

Evaluating Lightweight Adversarial Defenses for Machine-Learning-Based Malware Detection

b) Problem Statement

Machine-learning models are widely used to detect malware, but attackers can make small changes to malware that trick these models into thinking it's safe. While many defenses exist, most are complex, resource-intensive, or tested only under ideal conditions. There is limited understanding of which simple, practical defenses actually improve robustness, both individually and in combination, under realistic attack scenarios. This makes it difficult for organizations to adopt effective, deployable solutions in resource-constrained environments.

c) Aim

To evaluate the effectiveness of lightweight defense strategies in improving the robustness of machine-learning-based malware detection systems against existing adversarial attacks.

d) Research Objectives

1. Build a baseline malware-detection classifier using publicly available malware datasets.
2. Test lightweight defenses against existing, realistic adversarial malware attacks.
3. Evaluate the performance of individual lightweight defenses.
4. Assess whether combining lightweight defenses improves robustness compared to using them individually, while analyzing trade-offs in accuracy and computational cost.

e) Research Questions

1. How effective are individual lightweight defenses at protecting malware-detection models against existing adversarial attacks?
2. Does combining multiple lightweight defenses improve robustness compared to using single defenses?
3. What trade-offs exist between robustness, accuracy, and computational cost for these defenses?
4. Which lightweight defense strategies provide practical, deployable protection in resource-constrained environments?