A METHOD OF WINDOWS MALWARE MUTATION USING REINFORCEMENT LEARNING WITH DYNAMIC ANALYSIS-BASED FUNCTIONALITY VALIDATION

Lê Trọng Nhân[1,2]

[1] Trường ĐH Công Nghệ Thông Tin

[2] University of Science HCMC, Vietnam

What ?

This study proposes **a reinforcement learning-based framework for generating evasive Windows malware variants**, integrated with **dynamic analysis to verify functional preservation**. The goal is to produce diverse malware samples that can:

- Bypass machine learning-based detection systems.
- Maintain original malicious functionality without corruption after modification.

Why ?

- **Vulnerability of ML-based detectors:** Current systems fail against adversarial malware variants crafted with subtle modifications.
- **Functional integrity:** Existing mutation techniques often break malware functionality, rendering variants useless. Our RL-driven approach ensures evasion and execution reliability via dynamic analysis.

Overview

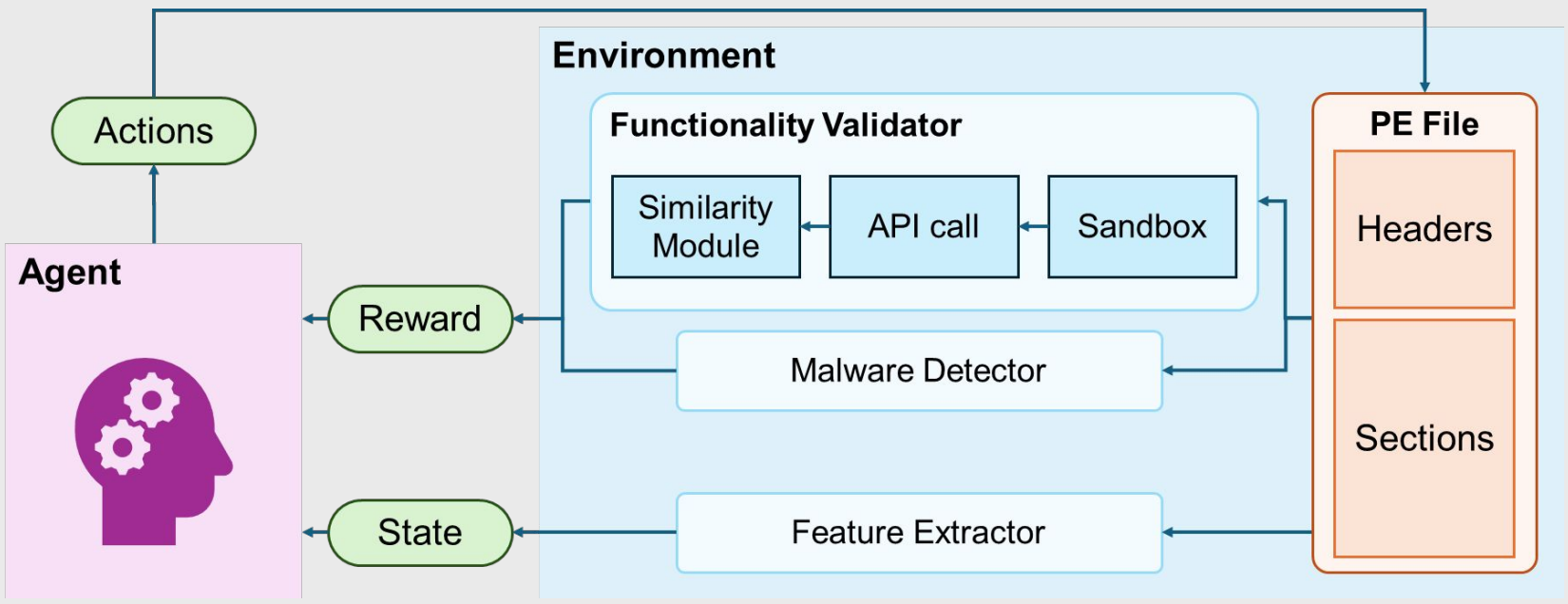

Figure 1. RL-based malware mutator with functionality validator

Description

1. Reinforcement Learning Agent

The agent learns to mutate malware samples by interacting with a controlled environment. It selects actions from a predefined set of PE file transformations (e.g., section renaming, overlay appending) to maximize a custom reward function. The reward considers three factors: evasion success, action efficiency, and most importantly, functionality preservation. Through iterative training, the agent optimizes its policy to produce stealthy yet operational malware variants.

2. Functionality Validator

To verify that the mutated malware retains its behavioral integrity, we introduce a dynamic analysis-based validator. The system runs both the original and modified samples in a sandbox and records their API call sequences. These sequences are processed by:

- Cleaning and de-duplication
- Encoding into symbolic DNA representations
- Aligning using local sequence alignment (e.g., Smith-Waterman algorithm)
- Measuring similarity with metrics such as Jaccard index and gap-based ratios

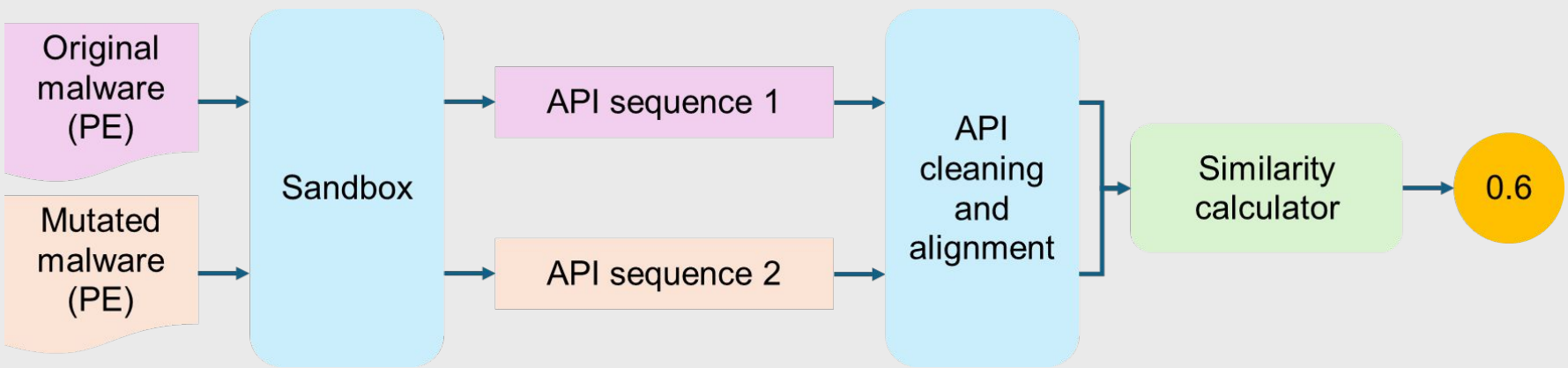

Figure 2. Dynamic analysis-based Functionality Validator.

NII

Lê Trọng Nhân – Trường Đại học Công Nghệ Thông Tin

TEL : 0944002939 Email : nhanlet.19@grad.uit.edu.vn