Final Project in NT230 course– Malware’s Modus Operandi

Ransomware Detection and Prevention

Explainable deep learning-based ransomware detection using dynamic analysis

Thành viên nhóm: Hồ Hoàng Diệp (22520249), Nguyễn Đặng Nguyên Khang (22520617), Trần Vỹ Khang (22520628)
Mã nhóm: G05, Mã đề tài: S13, GVHD: Phan Thế Duy

Introduction

Ransomware is an increasingly serious cybersecurity threat that is difficult to detect using traditional methods. While static analysis inspects code without execution and can help identify known patterns, it often fails against obfuscated or encrypted malware. Dynamic analysis, which monitors software behavior during execution, provides diverse information through sequences of API calls, DLLs, and mutexes—key indicators of ransomware activity. Additionally, we have collected and incorporated datasets from various external sources, applying preprocessing and filtering to enrich the data diversity. Our project extracts these ordered sequences from Cuckoo Sandbox reports, concatenates them into a combined vector, and applies a two-layer CNN to classify ransomware accurately. To enhance transparency, we use Explainable AI methods, LIME for local explanations and SHAP for global feature importance, providing insight into the model’s decisions.

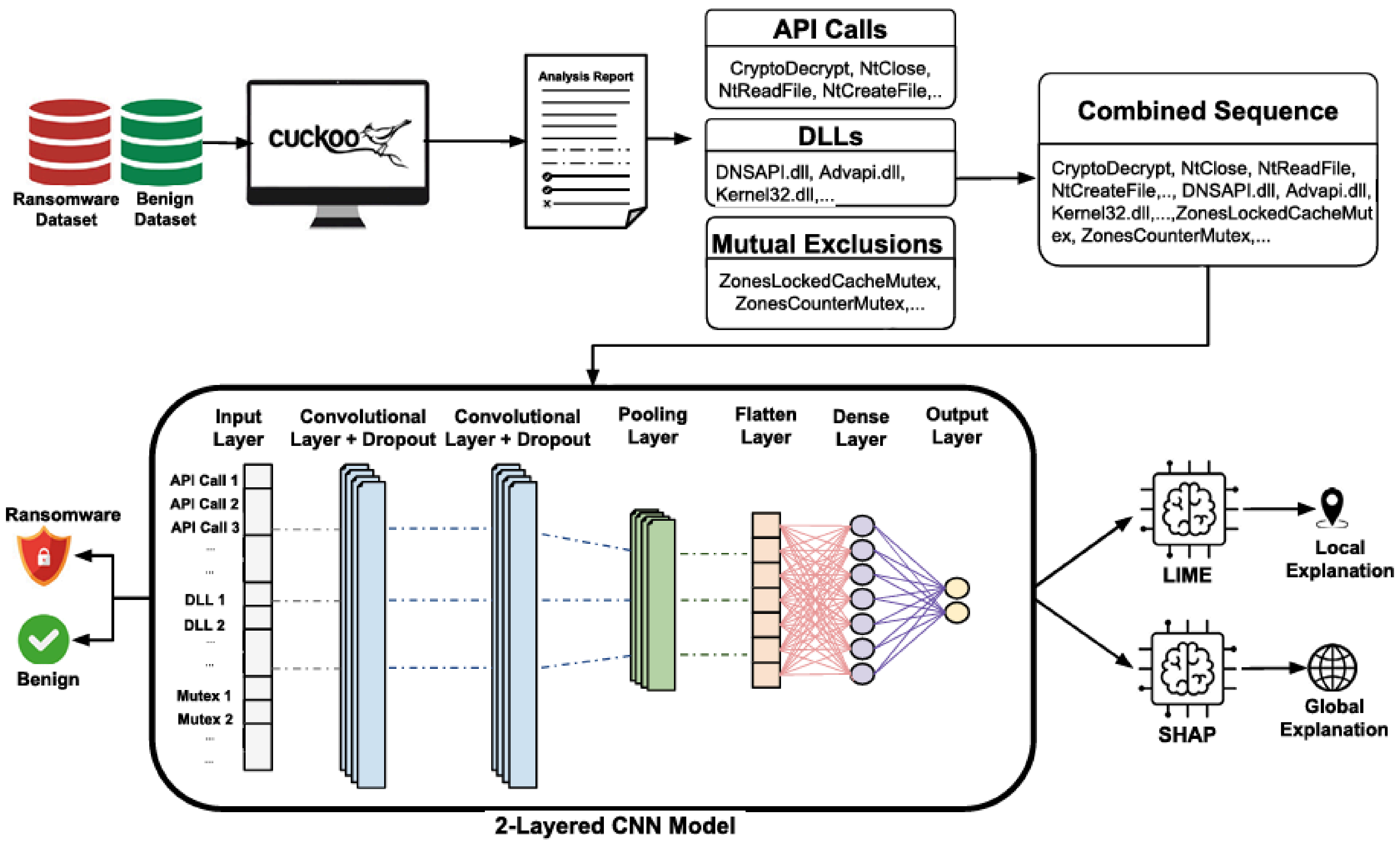
Ransomware Dataset
Benign Dataset
cuckoo
Analysis Report
API Calls
CryptoDecrypt, NtClose, NtReadFile, NtCreateFile,..
DLLs
DNSAPI.dll, Advapi.dll, Kernel32.dll,...
Mutual Exclusions
ZonesLockedCacheMutex, ZonesCounterMutex,...
Combined Sequence
CryptoDecrypt, NtClose, NtReadFile, NtCreateFile,.., DNSAPI.dll, Advapi.dll, Kernel32.dll,...,ZonesLockedCacheMutex, ZonesCounterMutex,...
Input Layer
Convolutional Layer + Dropout
Convolutional Layer + Dropout
Pooling Layer
Flatten Layer
Dense Layer
Output Layer
API Call 1
API Call 2
API Call 3
DLL 1
DLL 2
Mutex 1
Mutex 2
Ransomware
Benign
LIME
Local Explanation
SHAP
Global Explanation
2-Layered CNN Model

***Fig.1** Detailed ransomware/benign classification framework of project*

Methodology

The framework of our proposed design is shown in **Fig 1**.

First, executing ransomware and benign/malware samples in Cuckoo Sandbox to obtain JSON reports of runtime behaviors.

From each report, extracting three ordered sequences–API calls, DLLs and mutexes. Then, concatenating them into a single sequence vector. This vector is embedded and fed into a two-layer 1D-CNN for binary classification.

We train end-to-end with sparse categorical cross-entropy, use a validation split to tune and prevent overfitting, and report accuracy, TPR, FPR, and F1 on the held-out test set. For transparency, we apply LIME for local explanations and SHAP for global feature importance.

Experiments and Results

We conducted our experiments on a machine with **11th Gen Intel(R) Core(TM) i7-11390H,3.4GHz,16GB RAM,** running a Windows 7 guest VM in Cuckoo Sandbox for dynamic behavior collection. Our study focuses on two classification settings: **Ransomware vs. Benign** and **Ransomware vs. Malware.**

As shown in **Table 1** and **Table 2**, our proposed **2L-CNN** outperforms or matches traditional models such as Decision Tree, Random Forest, and standard CNN.

In the ransomware/benign task (Table 1), 2L-CNN achieves **the highest accuracy (99.10%), TPR (98.04%)**, and **zero false positive rate (0.00%)**, along with an **F-Score of 0.9901**.

Finally, we apply **LIME** and **SHAP** to provide both **local** and **global explanations**, are shown in **Fig 2** and **Fig 3.**

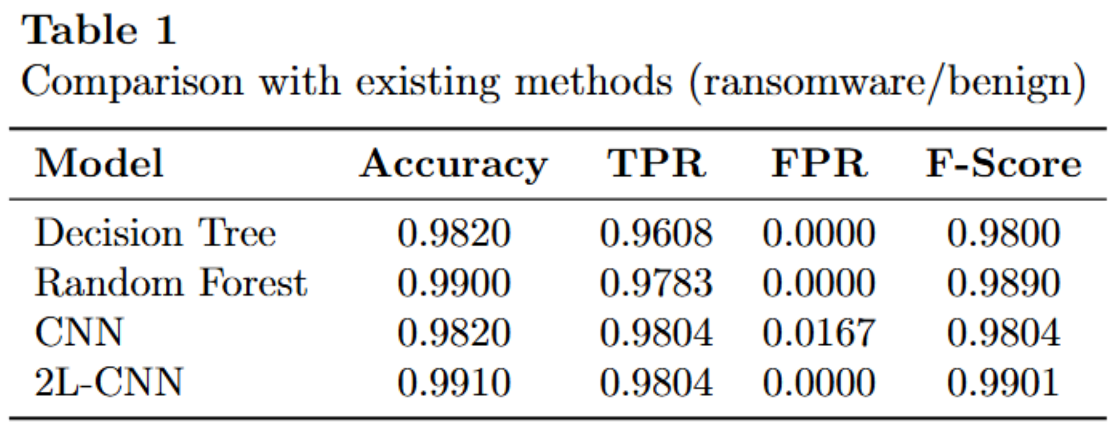

Table 1
Comparison with existing methods (ransomware/benign)

Model	Accuracy	TPR	FPR	F-Score
Decision Tree	0.9820	0.9608	0.0000	0.9800
Random Forest	0.9900	0.9783	0.0000	0.9890
CNN	0.9820	0.9804	0.0167	0.9804
2L-CNN	0.9910	0.9804	0.0000	0.9901

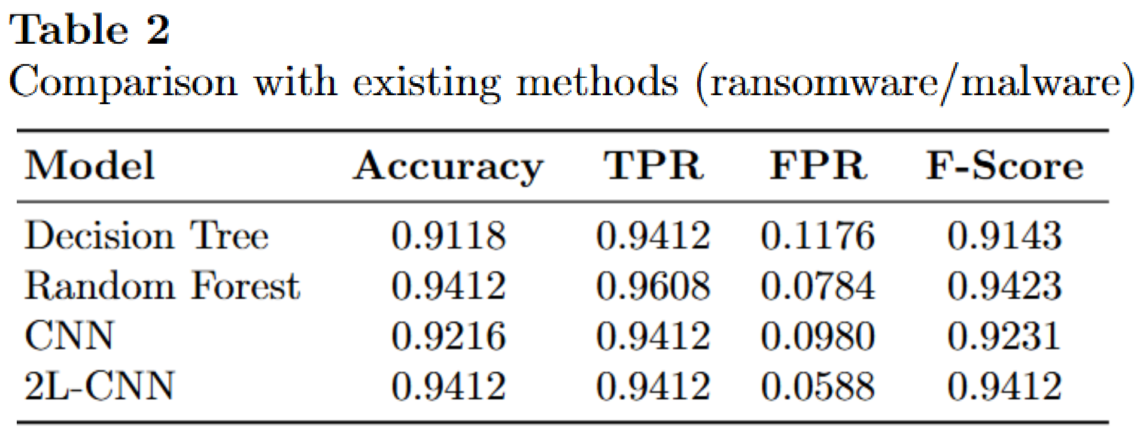

Table 2
Comparison with existing methods (ransomware/malware)

Model	Accuracy	TPR	FPR	F-Score
Decision Tree	0.9118	0.9412	0.1176	0.9143
Random Forest	0.9412	0.9608	0.0784	0.9423
CNN	0.9216	0.9412	0.0980	0.9231
2L-CNN	0.9412	0.9412	0.0588	0.9412

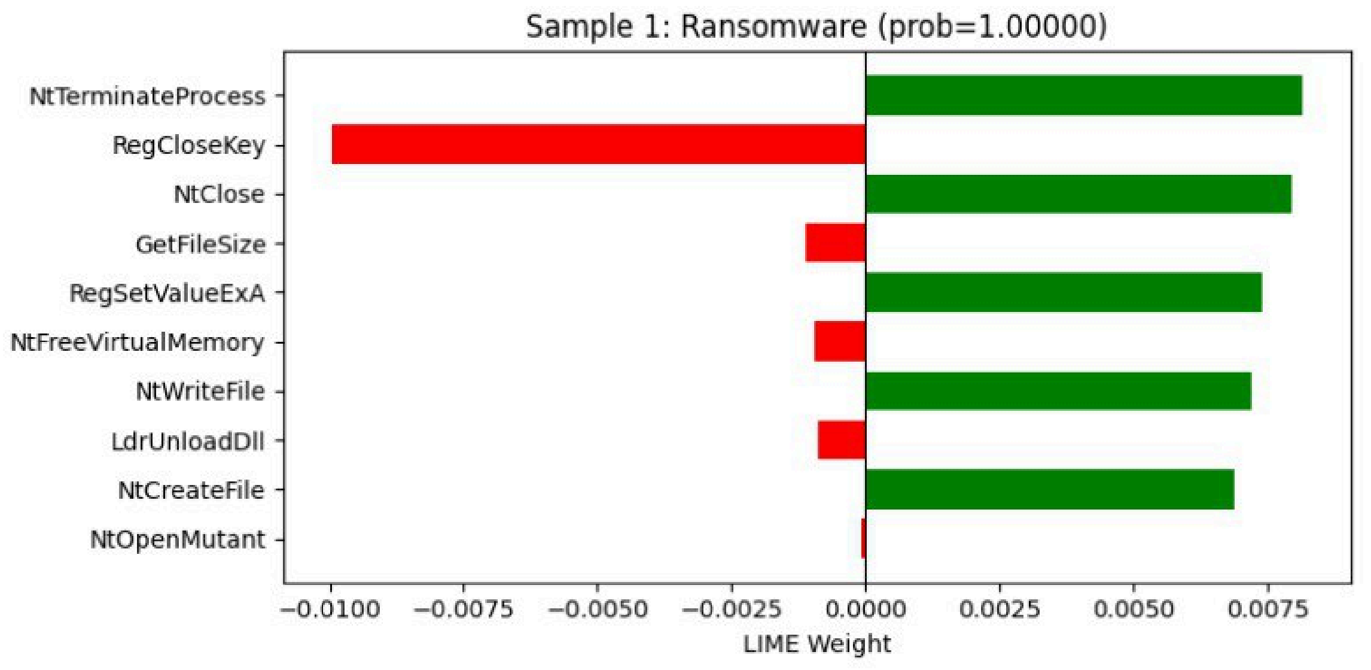

***Fig 2.** LIME feature contributions for Sample 1 (Ransomware, probability = 1.00)*

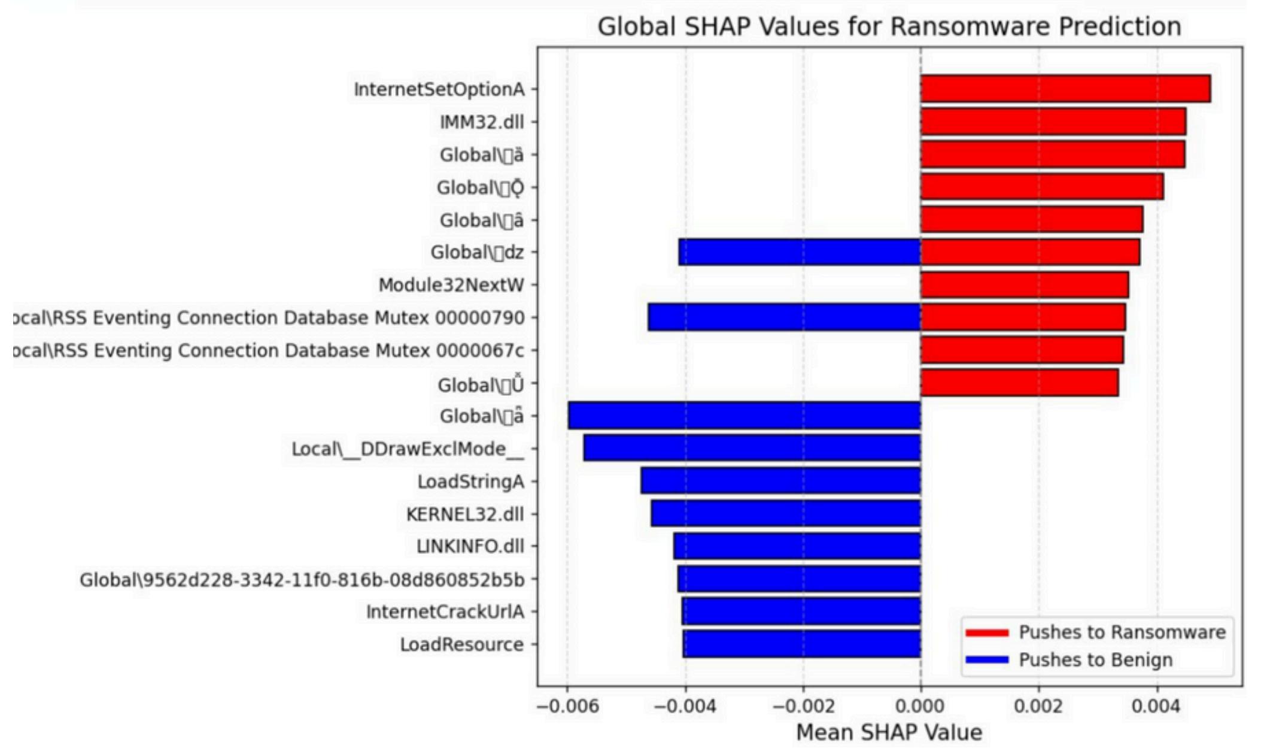

***Fig 3.** Feature-based SHAP feature contributions (ransomware/benign)*

A project at NT230 course– Semester 2 – School Year of 2024-2025, UIT-VNUHCM– Copyright: UIT InSecLab