SpecView: Malware Spectrum Visualization Framework with Singular Spectrum Transformation

Project Proposal

Malware Analysis

CY-471

Project Team:

Umar Tariq (2022604)
Ayela Israr (2022130)
M Zeeshan (2022644)
Ahmad Amjad (2022063)

Problem Statement

Traditional malware visualization and classification methods—mainly based on binary image conversion—fail to capture deep structural and textural information of malware samples. These methods often perform poorly against polymorphic, metamorphic, and packed/encrypted malware variants. Additionally, static and dynamic analysis methods have drawbacks:

- **Static analysis** struggles with obfuscation and packing.
- **Dynamic analysis** is time-consuming and vulnerable to anti-VM and sandbox evasion.

Therefore, there is a need for a **universal, efficient, and accurate malware detection and classification framework** capable of handling evolving malware variants and obfuscation techniques.

Objectives

The main goal of this research is to design a **universal malware classification framework** that combines spectrum visualization and machine learning for robust detection.
Specific objectives include:

- Develop **SpecView**, a malware spectrum visualization framework that converts malware binaries into 1D time-series spectrum data.
- Apply **Singular Spectrum Transformation (SST)** to extract meaningful structural and behavioral features from the malware signal.
- Optimize SST parameters using **Particle Swarm Optimization (PSO)** to improve classification performance.
- Use multiple **machine learning classifiers** (e.g., Random Forest, SVM, KNN, Voting Classifier) for malware family detection.
- Evaluate the framework's **cross-platform effectiveness** on both Windows and Android malware datasets.

Methodology

The SpecView framework follows a structured multi-stage pipeline:

1. **Binary to Time-Series Conversion**
   Each malware binary (PE/APK) is read as a sequence of bytes and converted into a **one-dimensional time-series signal**, capturing the raw structural characteristics of the malware.
2. **Signal Resampling**
   The time-series signal is resampled to maintain uniformity across different malware samples, ensuring consistent input length for analysis.

3. **Feature Extraction via Singular Spectrum Transformation (SST)**
   SST analyzes the signal to detect structural changes, generating **Change Point (CP) scores** that represent the internal evolution of malware. These CP scores form the "SST spectrum," which visualizes the malware's unique structure.
4. **Parameter Optimization with Particle Swarm Optimization (PSO)**
   PSO fine-tunes SST parameters (window size, order, and lag) to minimize intra-class variance and maximize classification accuracy.
5. **Classification Phase**
   Extracted SST features are fed into multiple machine learning algorithms for classification, including:
   - K-Nearest Neighbors (KNN)
   - Support Vector Machine (SVM)
   - Gaussian Naive Bayes (GNB)
   - Extra Trees (ET)
   - Random Forest (RF)
   - Voting Classifier (VC) — an ensemble of top-performing models.
6. **Datasets Used**
   - **Windows Malware:** Malimg, Malheur
   - **Android Malware:** Drebin, PRAGuard (Class Encryption)

This methodology allows SpecView to detect malware across multiple platforms and family types with minimal computational cost.

Results

The proposed SpecView framework achieved **remarkable accuracy, robustness, and cross-platform performance** across all test datasets.

- **Malimg (Windows):** 100% accuracy using the Voting Classifier.
- **Malheur (Windows):** 99.68% accuracy with Random Forest, successfully identifying packed and obfuscated variants.
- **Drebin (Android):** 99.08% accuracy using Random Forest, outperforming traditional Android malware detectors such as Drebin, DroidScribe, and AMalNet.
- **PRAGuard Malgenome (Encrypted Android Malware):** 100% accuracy, proving strong resistance to encryption and class obfuscation.

Unlike 2D image-based approaches that require heavy computation and suffer from data distortion, SpecView's 1D SST spectrum method preserves malware structure efficiently and significantly reduces processing time. Additionally, its **cross-platform adaptability** allows consistent detection of both Windows and Android malware families.

Overall, SpecView provides an **accurate, efficient, and scalable** malware detection framework that meets the real-world requirements of speed, precision, and resilience against advanced evasion techniques.

Future Work

Future research directions for SpecView include:

- Exploring **malware homology and evolution analysis** using SST visualization to study variant relationships.
- Integrating SpecView into **real-time malware detection and incident response systems** for adaptive defense.
- Employing **incremental learning techniques** to enhance adaptability for zero-day malware.
- Combining **deep learning** with SST features to further automate feature extraction and improve detection precision.