Decoding GAN-Generated Malware using Explainable AI Techniques

Matteo Tiozzo • 4th week 2024/10/08 - 2024/10/15

Overview

Progress

- Extracted mnemonic occurrences from each assembly file
- Generated grayscale images (256px*256px) of assembly and hexadecimal code

Major risk

- Understanding the logic behind extracting mnemonic instructions

Mathematical methods used for matrix generation

Median

- Represents the middle value of a dataset when sorted in ascending order, used to filter values by considering those greater than the median

Mean and Standard Deviation

- A statistical technique that sets thresholds for filtering data based on their distribution. It involves calculating the mean of the data points first, and then using the standard deviation to identify significant deviations from the mean

Percentile with Threshold of 75%

- Determines a value below which 75% of the observations in a dataset fall. This method helps set thresholds less influenced by outliers and skewed distributions

Mathematical method used for matrix normalization

Min-max normalization

- A scaling technique that transforms the values in a dataset to a common scale, ensuring all values are proportionately adjusted to the specified range

Tools used this week

Python

- Used to extract mnemonic instructions and count them from assembly files
- Used to extract and count hexadecimal representations
- Used to create grayscale images for both assembly and hexadecimal code

Questions & doubts

- How am I going to use the generated images? Will the GAN's Discriminator/Generator generate an image of the input malware and compare it with the image dataset?