Project: Graph Visualization and Interaction

Overview

The goal of this project is to develop transformations and visualizations permitting cyber-analysts to explore the complex structure of malicious applications. This project has two phases, visualization and interaction. In the first phase, you will transform the output of Radare2, a reverse-engineering tool, to depictions of the structure of the malware using the Scalable Vector Graphics (SVG) file format. In the second phase, you will use D3 to interact with this depiction, providing contextual information, emphasis, etc.

Background

Radare2 output is a JSON file. It is a list of functions, each function is made of blocks, and each block is made of instructions. We currently process this data structure to resolve the calls and branches providing us with edges of the graphs. These edges represents the flow between instructions, including call and control flow. When we process this data structure using machine learning, it is "projected" on the three different levels of granularity: functions, blocks, and instructions. This project aims at exploring ways to create interactive visualizations of the complete data-structure.

We envisage the following ways to generate the SVG representation of graphs:

- **GraphViz** is good at flow graphs. In addition, it is possible to cluster nodes permitting the representation of blocks and functions.
- **Graph-tool** (in Python) permits one to create compact visualization of large graphs.
- **2D/3D embedding** can be used to position each node in the rendering. A clustering algorithm could be applied to the embedding of all nodes and used for coloring. **See instructor for this part.**

Data

We will provide a dataset of analysis corresponding to 10,000 executable files. It contains the results of multiple analyses for each files:

Considerations

We do not expect it to be perfect, but speed and performance of generating and visualizing the graphs is of utmost importance.

Technology

- Graph visuals should be done in an easily manipulated format, i.e. SVG
- Visualization should be done using an open-source platform, i.e. D3

Related Links

1. D3 - https://d3js.org/