

Knight Foundation School of Computing and Information Sciences

Spring 2022 Senior Design Project

Ransomware Anomaly Detection

Student: Brendan Wojtczak
Mentor: Mike East, Carl Banzhof, Kaseya, RocketCyber
Instructor/Faculty: Dr. Masoud Sadjadi, Florida International University

PROBLEM

Ransomware is a type of cyber attack which demands payment in exchange for access to a computer's data or functionality. With data security being a top priority of companies and even individual users, the ability to detect such invasive behavior can prove beneficial to combating it as well. The Ransomware Anomaly Detection program was designed to aid the detection of ransomware attacks using machine learning.

CURRENT SYSTEM

The current version is compatible with machines running on Microsoft operating systems, recording event log data from Sysmon archive files, and calculates dissimilarity to find outliers within the system. This project is written in Python and uses various Python libraries to store and calculate data as well as train virtual machines. The current version supports execution via command line prompt with various flags for more detailed functionality.

IMPLEMENTATION

The application utilizes various python libraries in order to calculate and store data. Such libraries include: scikit-learn, and pickle (for serialization). Figure [1] shows the implementation and functionality of our data smoothing portion of the application. Figure [2] shows the main structure of our system, and which files execute with each other.

REQUIREMENTS

- Microsoft Windows Sysmon v13.33
- Python3 IDE
- Sysmon Simulator

VERIFICATION

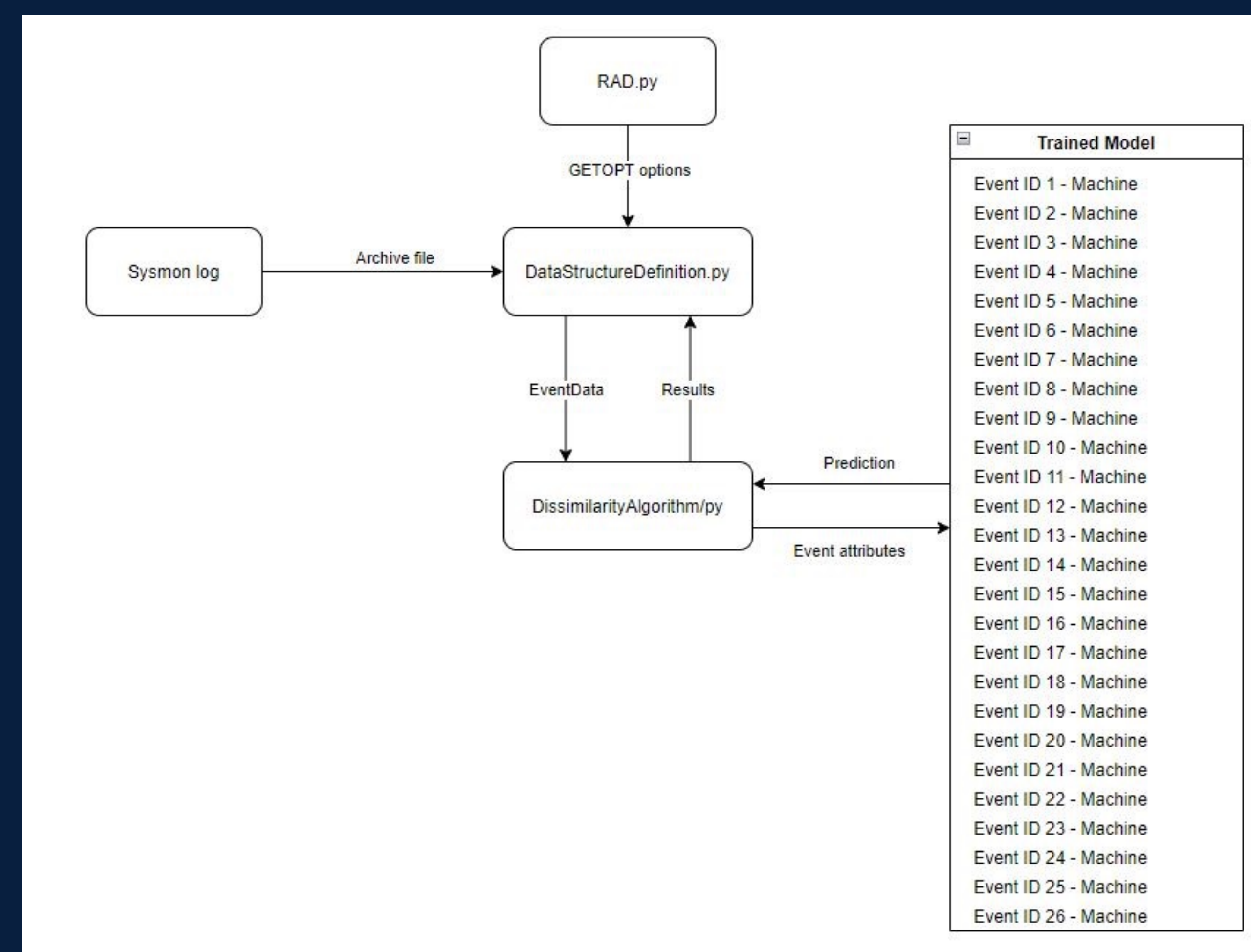
This project was tested using the Sysmon simulator as a source of ransomware data. Training was done with 30 days worth of personal computer logs. Largest recorded accuracy was approximately 75% successful classification.

Key Attribute Implementation

```
1 #Function returning a list of only key attributes. Argument: 'arr' is a list of 1s and 0s
2 def createDataStructure(EventData, EventID):
3     activeValues = []
4     EventID -= 1
5     if len(KeyAttributes[EventID]) > 0:
6         for i, val in enumerate(KeyAttributes[EventID]):
7             if val == 1:
8                 activeValues.append(EventData[i])
9         return activeValues
10    else:
11        print("\nErr: list argument contains no data.\n")
12
13    KeyAttributes = [[ 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1], #1
14                    [ 0, 0, 0, 1, 1, 0, 1], #2
15                    [ 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0], #3
16                    [ 0, 0, 0, 0 ], #4
17                    [ 0, 0, 1, 1, 1], #5
18                    [ 0, 0, 0, 1, 1, 1], #6
19                    [ 0, 0, 0, 1, 1, 0, 1, 1, 1, 1 ], #7
20                    [ 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1 ], #8
21                    [ 0, 0, 0, 1, 1 ], #9
22                    [ 0, 0, 0, 0, 0, 1, 0, 0, 1, 0 ], #10
23                    [ 0, 0, 0, 1, 1, 0], #11
24                    [ 0, 1, 0, 0, 0, 1, 1, 0], #12
25                    [ 0, 0, 0, 0, 1, 1, 1, 0, 0], #13
26                    [ 0, 0, 1, 1, 1, 1, 0], #14
27                    [ 0, 0, 0, 1, 1, 0, 0], #15
28                    [ 0, 0, 1], #16
29                    [ 0, 0, 0, 0, 1, 1, 1], #17
30                    [ 0, 0, 0, 0, 1, 1, 1], #18
31                    [ 0, 0, 1, 1, 1, 1, 1], #19
32                    [ 0, 0, 1, 1, 1, 1, 1], #20
33                    [ 1, 1, 1, 1, 1, 1], #21
34                    [ 0, 0, 0, 0, 1, 1, 0, 1], #22
35                    [ 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], #23
36                    [ 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], #24
37                    [ 1, 1, 1, 1, 1, 1], #25
38                    [ 0, 0, 0, 0, 1, 1, 1, 0 ]] #26
```

[1] Key Attributes

SYSTEM DESIGN



[2] System Structure

SUMMARY

Project Areas I Worked On:

- Serializing and de-serializing calculated data for efficient use.
- Research on key indicators of anomalies.
- Implementation of data smoothing algorithm.
- Research on data mining and outlier detection.

REFERENCES

- Data Mining Concepts and Techniques 3rd edition, Jiawei Han, Micheline Kamber, Jian Pei, August 2000
- <https://docs.microsoft.com/en-us/sysinternals/downloads/sysmon>
- <https://github.com/ScarredMonk/SysmonSimulat> or

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by XXXXX (name of the project sponsor: federal, state or local government, or corporate partners, where applicable). We/I thank XXXX for his/her/their assistance and mentorship that I/we received throughout the senior design project.