PESIDIOUS - Create Mutated Evasive Malware Using Artificial Intelligence

Bedang Sen
*Incident Response Consultant, X-Force IRIS, IBM*

Chandni Vaya
*Incident Response Consultant, X-Force IRIS, IBM*
Contents of this Presentation
Contents of this Presentation

Who are we?  
Pesidious  
Implementation  
Project Demo  
Future Work
Who are we?

Bedang Sen
Incident Response Consultant
X-Force IRIS, IBM
https://www.linkedin.com/in/bedangsen/

Chandni Vaya
Incident Response Consultant
X-Force IRIS, IBM
https://www.linkedin.com/in/chandni-vaya-519a05137/
Why Pesidious?
Malware Mutation Using Reinforcement Learning and Generative Adversarial Networks
Malware Mutation Using Reinforcement Learning and Generative Adversarial Networks
Malware Mutation Using Reinforcement Learning and Generative Adversarial Networks
What is Reinforcement Learning

**Agent**

**Goal**

**State**

**Actions**

**Environment**

Q-Value (+ve)

Q-Value (-ve)
**What is |**

Reinforcement Learning

<table>
<thead>
<tr>
<th></th>
<th>↑</th>
<th>↓</th>
<th>→</th>
<th>←</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-0.51</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.84</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.87</td>
<td>0.64</td>
<td>0.05</td>
</tr>
<tr>
<td>34</td>
<td>0.34</td>
<td>0</td>
<td>0.86</td>
<td>0.21</td>
</tr>
<tr>
<td>35</td>
<td>0.55</td>
<td>0</td>
<td>1</td>
<td>0.31</td>
</tr>
</tbody>
</table>

**Q-TABLE**
**What is Deep Reinforcement Learning**

<table>
<thead>
<tr>
<th></th>
<th>Up</th>
<th>Down</th>
<th>Right</th>
<th>Left</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-0.51</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.84</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.87</td>
<td>0.64</td>
<td>0.05</td>
</tr>
<tr>
<td>34</td>
<td>0.34</td>
<td>0</td>
<td>0.86</td>
<td>0.21</td>
</tr>
<tr>
<td>35</td>
<td>0.55</td>
<td>0</td>
<td>1</td>
<td>0.31</td>
</tr>
</tbody>
</table>

**Q-TABLE**

**Neural Network**
What is Generative Adversarial Networks
What is Generative Adversarial Networks

Source Image/Video
Noise Vector
Fake Image/Video

1 0 1 1 0 1 1 1

Generated Output
Generator
Discriminator

Generative Adversarial Network
Malware Mutation Using Reinforcement Learning and Generative Adversarial Networks

If action = add_imports or add_sections or rename_sections

Experience $t$

State $t+1$

Q (State) $t$

Q (State) $t+1$

Reward

Loss

Optimizer

Modifier

Classifier

Memory

Experience Replay

State $t$
Implementation | Extracting Features into Feature Vector Maps

1. Collect the malicious and benign binary dataset.

2. Extract all the features into a single feature vector map

3. Generate feature vectors for each binary data using the feature vector map
Implementation | Extracting Features into Feature Vector Maps

EXAMPLE.EXE

HEADER

SECTIONS

DOS HEADER
PE HEADER
OPTIONAL HEADER
DATA DIR.

SECTION TABLE

CODE

IMPORTS

DATA
<table>
<thead>
<tr>
<th>Feature Vector Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sections</td>
</tr>
<tr>
<td>Number of sections with non zero value or empty name</td>
</tr>
<tr>
<td>No of sections with different characteristics</td>
</tr>
<tr>
<td>DLL</td>
</tr>
<tr>
<td>Functions</td>
</tr>
<tr>
<td>Virtual size</td>
</tr>
<tr>
<td>Debug flag</td>
</tr>
<tr>
<td>Relocation</td>
</tr>
<tr>
<td>Resources</td>
</tr>
<tr>
<td>Signature</td>
</tr>
</tbody>
</table>

**SECTION INFORMATION**

**PROPERTY OF ENTRY POINTS**

**IMPORTS INFORMATION**

**EXPORT INFORMATION**

**MACHINE ARCHITECTURE OS**

**OTHER HEADER INFORMATION**

**BENIGN FILE**
1. Implement the environment for the agent to learn.

2. Design a Deep learning model to select the actions based on the current state of the malware.

3. Use experience replay with prioritized replay buffer.
Implementation | Training a Deep Reinforcement Learning Agent

- Randomly adding import functions and DLLs
- Randomly adding sections and renaming sections
- Appending to existing section
- Appending random bytes
- Removing/adding signature
- Removing debug flag
- UPX pack/unpack
1. Feature vectors are concatenated with noise and fed to the GAN.

2. The GAN generates adversarial feature vectors.

Generating Adversarial Feature Samples with Generative Adversarial Networks
Implementation

Generating Adversarial Feature Samples with Generative Adversarial Networks

BLACK BOX

DEEP LEARNING NETWORKS

- DECISION TREE
- LOGISTIC REGRESSION
- MULTI LAYER PERCEPTRON
- RANDOM FOREST
- SVM
Implementation

Understanding the Generative Adversarial Network

- Benign Feature Vector: 1 0 1 1 0 1 1 1
- Noise Vector
- Malware Feature Vector
- Adversarial Feature Vector
- Discriminator
- BLACK BOX DETECTOR

Generative Adversarial Network
Malware Mutation Using Reinforcement Learning and Generative Adversarial Networks

**Benign Dataset**

**Feature Vector Mapping**

**Benign Feature Vector**

**Malware Feature Vector**

**Noise Vector**

**Generative Adversarial Network**

**Adversarial Feature Vector**

**State**

**Q (State)**

**State**

**Q (State)**

**Q (State)***

**Q (State)**

**Experience Replay**

**OPTIMIZER**

**LOSS**

**CLASSIFIER**

**MODIFIER**

**MEMORY**

If action = add_imports or add_sections or rename_sections

Experience**

**Experience**

**Environment**

**Neural Network**

**Loss**

**Reward**

**Loss**
- **Maintaining the functionality:**
  - Filtering out the PE32 files based on 32 bit
  - Filtering out DLLs and sections
  - Using C++ instead of Python for the malware reconstruction

- **Improving performance:**
  - Using a combination of the machine learning models scores
  - Initially we trained it with backdoors; now we are giving it more diverse malwares
  - For testing we made a comparison between AI and human
Project Demo

Creating Chaos with Mutated Evasive Malware.
Project **Demo**

Run Our Mutated Malware In A Cloud Based **Secure Sandboxed Environment**.

Variant.Ransom.Cerber.171:  
66 detected


Mutated Variant.Ransom.Cerber.171:  
40% more evasive  
100% functionality

Future Work

1. IMPROVE EVASIVENESS

2. MAINTAIN FUNCTIONALITY

3. HELP THE NEXT-GEN ANTI-VIRUS SYSTEMS
Thank You!