A Quest Called TRIBE: Clustering Malware Families for Enhanced Triage and Analysis

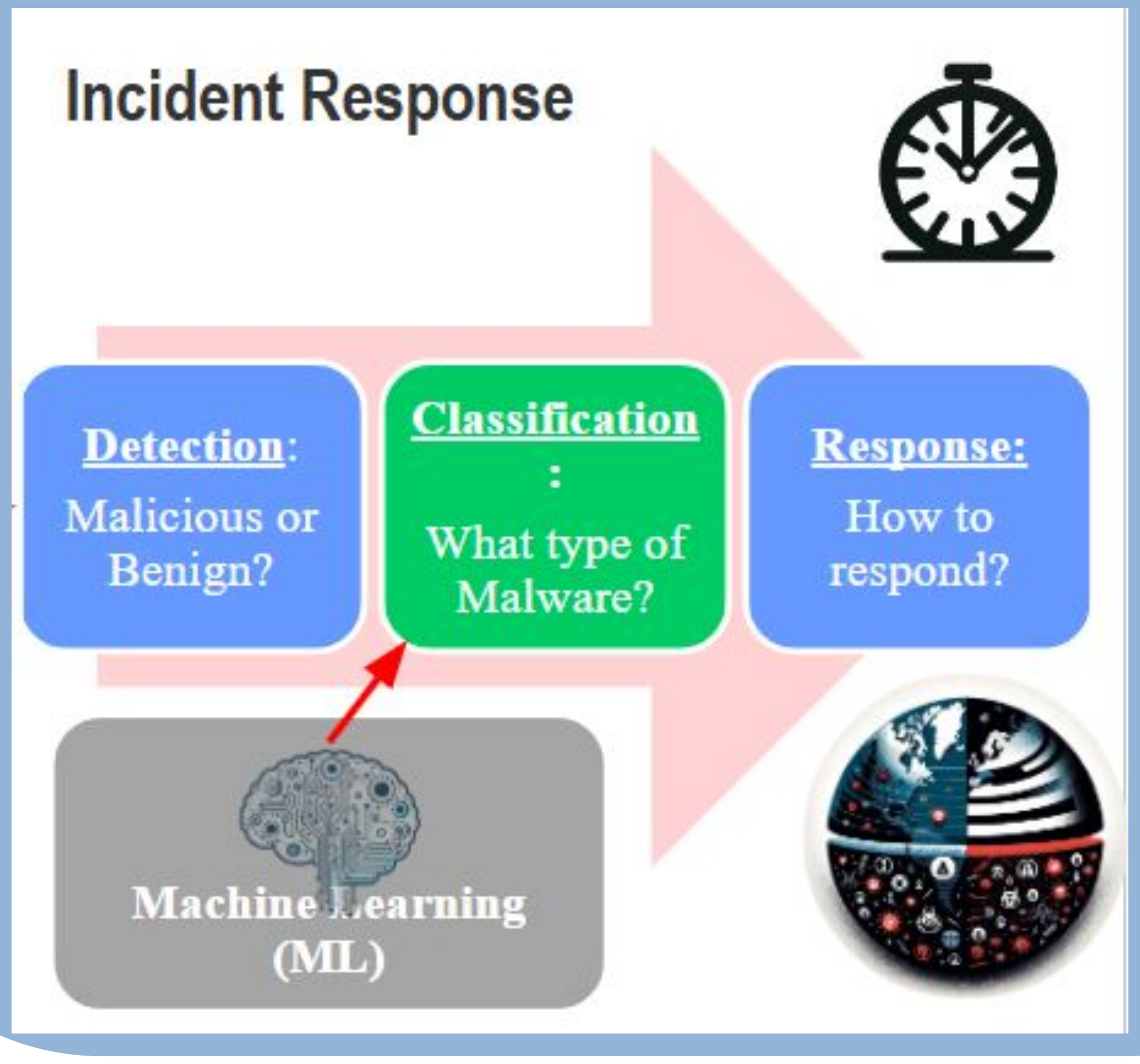
MIDN 1/C: Michael Chen, John Jenness, Chris Kim, Justin Liaw
Professor Dane Brown, Cyber Operations Department
CMDR. Edgar Jatho, Computer Science

Abstract

In recent years an increase in polymorphic and mutational malware has been noted by traditional antivirus (AV) software. These traditional AVs are often unable to classify obfuscated malware that should be in the same family leading to a large number of unnecessary malware families. We propose using a transformer-based sequence-to-sequence autoencoder to classify malware. This novel approach will classify malware into Tribal Relation Inferential Binary Encoder (TRIBE) clusters which will reduce the total number malware family classifications.

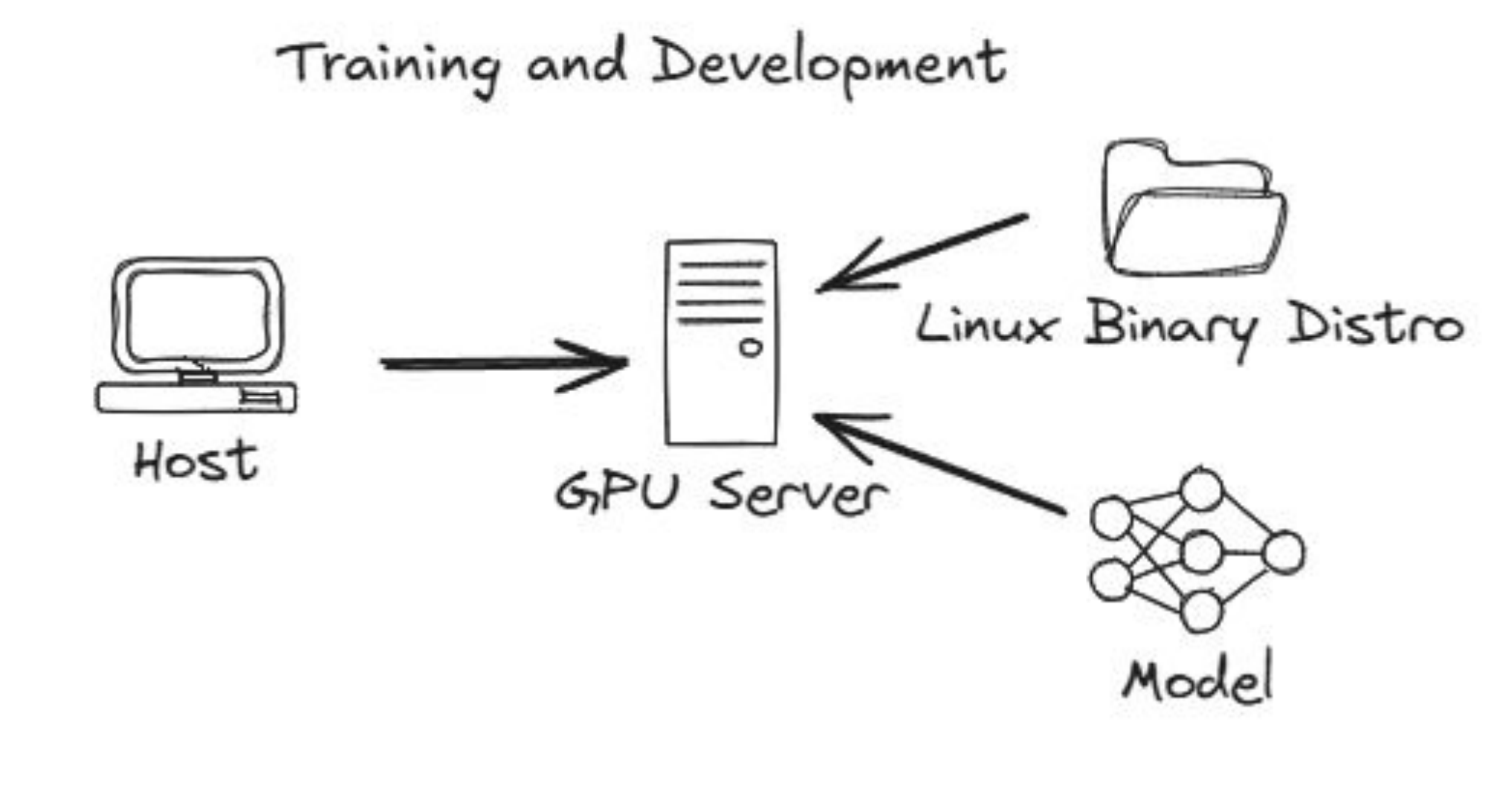
Importance

Our goal is to reduce the triage time it takes for Security Operation Centers (SOCs) to classify malware in order to increase response time.

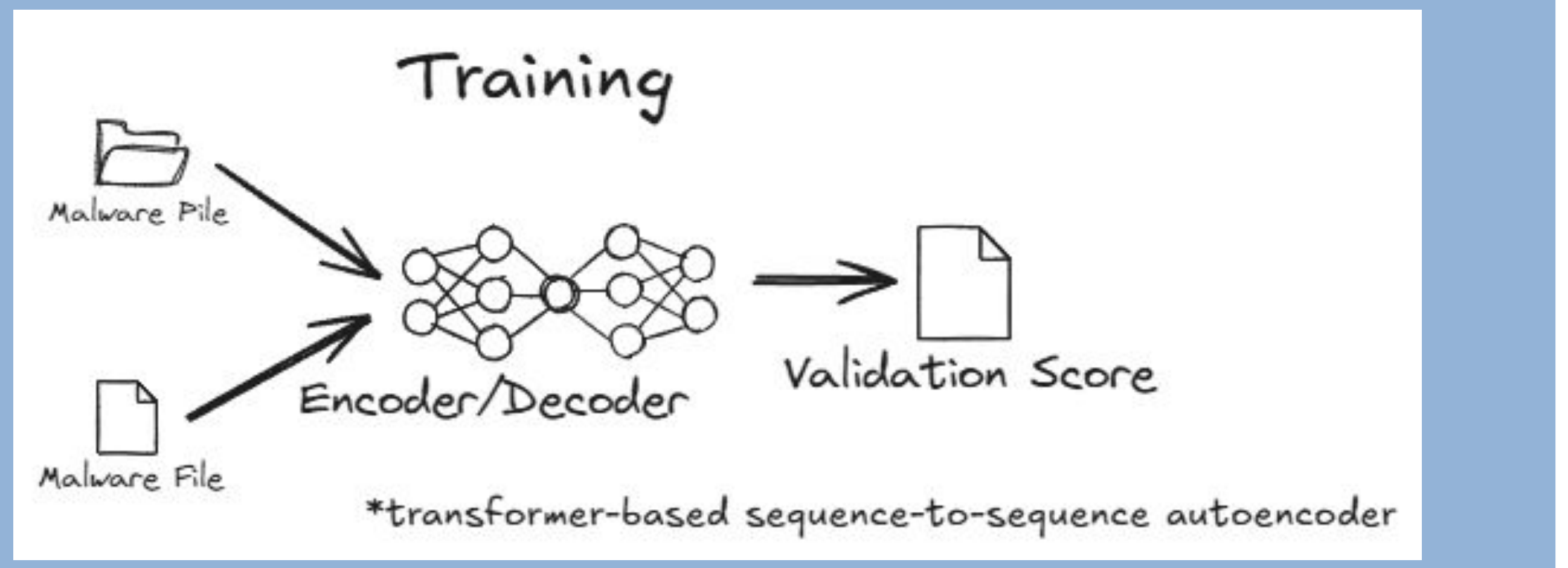

Incident Response
Detection: Malicious or Benign?
Classification: What type of Malware?
Response: How to respond?
Machine Learning (ML)

Methodology

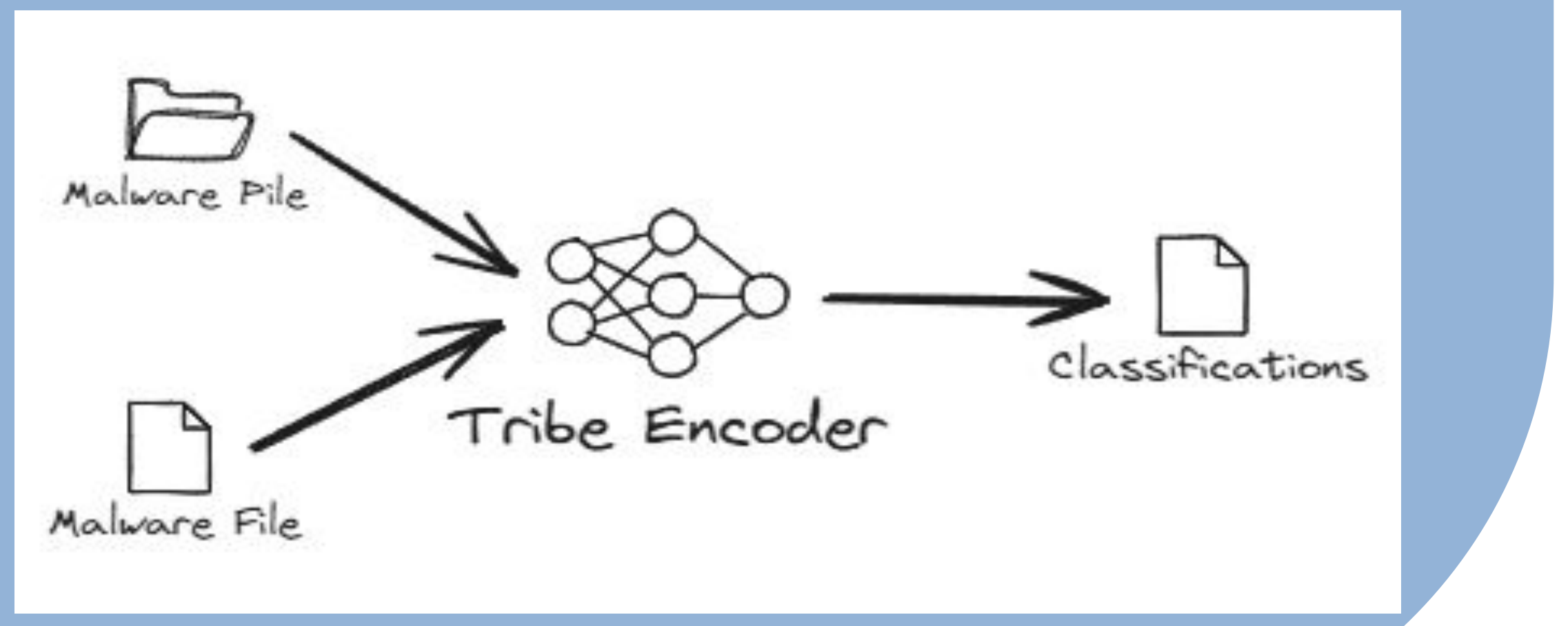
Phase 1: Development using Non-Malicious Dataset Testing



Phase 2: Configuration using known AV Datasets



Phase 3: Reduction of AV families and Productization



Product Details

Our end deliverable will be a Dockerized command line interface tool that can take either a binary file or a folder of binary files and returns the classification of each binary. This tool will arrive pre-trained.

Conclusions (Will Update)

We hope to find that our process can "see" around the obfuscation presented by polymorphism and mutation in order to reduce the total number of families malware can be classified into.

Acknowledgement

Trident Research Program: Provided $8600 GPU for malware.

Future Work

- Integration of recommended course of action based on classification
- User based modification of autoencoder and end-product trainability.

References