GRAMMATECH

SoURCE CODE Proposer's Day

Malware Origin Analysis

Intro to GrammaTech

- Located in (1) Ithaca, NY, (2) Remote, and (3) On-Site (Government Facilities)
- 30+ years of tackling the hardest software problems
- Highly technical employees
- Work focused on security, resilience, sustainment, automation, and developer productivity
- ➢ 15 current projects with 8 different agencies & DoD services

GrammaTech's Expertise: Program Analysis

- Extensive source code and binary code analysis experience
- Already have experience and technology for binary analysis and feature extraction
  - Could have baseline features in a week
  - Can investigate classification and relevant features almost immediately
  - We have an "infinite" toolkit where we can measure just about everything about a binary – full feature space
    - Some examples: binary parsing; n-grams of opcode mnemonics; advanced disassembly features; malware capabilities; call graph embeddings
  - Technologies work with multiple ISAs and operating systems
- We know what features may be relevant / reliable to identify malware
  - Active project involves identifying features relevant to malware

Relevant Projects: Binary Analysis

- Use binary features to determine if disassembly or binary transform will be successful
- Extract features from binaries for classification of malware/benignware
- Code comparison
  - Determine semantic equivalence based on, shape of control flow graph, extracted from binary
- Program/binary diversity

Relevant Projects:
Source Code Analysis and Refactoring

- Parsing and analyzing various source languages:
  - Code differencing and merging
  - Finding deprecated coding patterns (e.g., unsafe C++ code)
- Source-code transformations:
  - Refactorings, e.g., security-hardening, diversification
  - Code modifications, e.g., adjusting API usage, bug injection
- Cross-language translation:
  - Legacy code into modern languages, e.g., C++ to Rust (semi-automated)

Partnership Information

- Seeking technical partners:
  - Experimental Design and Statistical Analysis
  - Authorship Attribution and Stylometry
  - Personnel with Experience in Threat Intelligence Analytics
  - Cyber-Forensics
- GrammaTech prefers to prime.

Capabilities List

Cyber Security and Cyber-Forensics	GT Expert in Cyber Security
Binary Analysis and Compiler Theory	GT Expert
Programming Language Theory	GT Expert
Software Engineering	GT Expert
Experimental Design and Statistical Analysis	Seeking Partners
Authorship Attribution and Stylometry	Some Expertise
Software Development and Integration	GT Expert
Personnel with Experience in Threat Intelligence Analytics	Seeking Partners
Artificial Intelligence and Machine Learning for Cybersecurity	Some Expertise
Source Code Analysis	GT Expert

Problem Statement

- Extract features from binary code and source code
- Explore full feature spaces in binary code and source code files
- Measure the similarity between files
- Provide information to forensic experts to the likely origins (country, groups, individuals, etc.)
- Enable the automated matching of similar binaries from known samples