Automatic Extraction of TTPs from a Program's Binary Code

PhD position

Jean-Yves Marion
Jean-Yves.Marion@loria.fr
LORIA, Université de Lorraine

Dylan Marinho
Dylan.Marinho@lip6.fr
LIP6, Sorbonne Université

1. Context

Today, the detection of malicious programs is performed by neural models, supplementing syntactic rules such as Yara. While this approach works well for a fairly wide range of threats, the pressure and offensive capabilities are such that scientific advances are needed to break through the glass ceiling of current malware defenses.

2. Subject

2.1 Problem Statement

This PhD topic falls within the general theme of fighting malware, and more specifically in the field of binary code analysis of obfuscated programs, with a direct but potential application to the detection of suspicious behavior.

Most of the time, only the binary code of malicious programs, such as ransomware, on Windows/ Linux/MacOS is accessible. This code must be analyzed to understand the intentions of the final payload of the attack. This tedious and time-consuming task is carried out by reverse engineering experts. A triage is performed beforehand by trying to associate a malicious program with a known family in order to reduce the number of analyses. The result is a list of tactics, techniques, and procedures (TTPs) that are implemented in the malware, which will subsequently feed Cyber Threat Intelligence (CTI).

2.2 Objective

The objective of this PhD is to extract TTPs from the binary code of a malicious program.

In the context of the DefMal project, this PhD contributes both to work on reverse engineering and on detection. The proposed approach is to start with dynamic analysis to extract the control flow graph and the function call graph, as well as various information, in particular system register modifications, thread creations, and information on communications with the Command & Control C2.

The LORIA part (Carbone team) of the DefMal project has provided a dynamic analysis service that will supply all this information. Then, we will use heuristics for feature identification. This work will require enhancing approaches through various means, notably by dynamic symbolic analysis or generative AI.

2.3 Research Questions

1. Given a program's binary code, how can we identify a pattern corresponding to a procedure?
2. How can we define a technique, then a tactic, from a procedure graph?
3. How can we conclude from the extraction of TTPs that a program's behavior is potentially malicious?

2.4 Expected Results

The results will be published in the best possible conferences. Some parts of the work should be applicable quite quickly, and prototypes will be developed and validated incrementally as scientific advances are made. The tool may also be presented at more technical conferences such as SSTIC, BotConf, or BlackHat. The final prototype is intended to be a component in the DefMal analysis platform, enriching the TTPs associated with malware and the resulting CTI.

3. Organization and Support

A weekly meeting is organized via videoconference with the PhD student, and meetings with the whole team are held every two weeks.

This position will be assigned to a restricted area (ZRR) at Loria in Nancy. The PhD student will benefit from the team's expertise and will be able to interact and collaborate with the team's engineers and post-docs. The PhD student will have access to the High Security Laboratory (LHS). Finally, the laboratory offers a stimulating scientific environment with numerous seminars and a PhD student association.

4. Application

- The PhD position can start anytime from now, and latest in December 2026 - the application deadline is set for September 2026.
- To apply, please contact Jean-Yves Marion (Jean-Yves.Marion@loria.fr) and Dylan Marinho (Dylan.Marinho@lip6.fr).