Development of Automated Malware Analysis Framework

Keiji Takeda
KEIO UNIVERSITY

01/29/2016
Final Report

DISTRIBUTION A: Distribution approved for public release.
**Abstract**

The research project was conducted from June 2014 to June 2015 by the malicious software (malware) research team in Keio University. The outcome of the research includes development of a new method for identification of malware, a new method to monitor behavior of malware binary program and platform to analyze malware using both static analysis approach and dynamic analysis approach.

The goal of the project was to develop automated system to analyze malware with minimum human interaction. The developed technologies through this research project are applied to the platform developed and provided semi-automated functionality. Proposed methods are verified their performance against actual malware on the developed platform. Two research papers were published in academic conferences.

**Subject Terms**

Autonomous adaptation, AOARD
Final Report
"Development of Automated Malware Analysis Framework"
2015/9/13
Keiji Takeda
Keio University
keiji@sfc.keio.ac.jp

1 Overview of the Result

The research project was conducted from June 2014 to June 2015 by the malicious software (malware) research team in Keio University. The outcome of the research includes development of a new method for identification of malware, a new method to monitor behavior of malware binary program and platform to analyze malware using both static analysis approach and dynamic analysis approach.

The goal of the project was to develop automated system to analyze malware with minimum human interaction. The developed technologies through this research project are applied to the platform developed and provided semi-automated functionality. Proposed methods are verified their performance against actual malware on the developed platform. Two research papers were published in academic conferences.

2 Achievements

Through this research activity following outcomes were achieved.

- Design and specification for automated malware analysis system were determined.
- Two algorithms for malware analysis were proposed.
- Developed technologies were verified with actual malware.
- API/interfaces for malware analysis system were defined.

The system developed is a set of servers including honeypot to collect malware and modules to conduct static analysis of malware and benign binary program on windows platform and management functions of connected hardware platform and virtual machines.

The platform is developed as an integrated system that provides required
function for analyst to collect, store, conduct static analysis of binary program, conduct dynamic analysis on binary program then manage obtained data. The system enhance capability of security analyst by providing semi-automated malware analysis functions.

3 Research Process

This research has been conducted through following steps.

a. Specification and design of the prototype components.
b. Procurements for prototype components.c. Implementation of prototype components.d. Preliminary experiments with prototype components.e. Specification and design of prototype systemf. Procurements and implementation of prototype system
g. Completion of experiments with prototype system

4 Research Result

Findings through the research activities were complied as research papers. Those papers are presented at a research symposium and workshop.

The research paper related to the static analysis part is titled as "Proposal for Techniques to Identify Files to Investigate by Executable File." This research proposed an approach to generate a profile model of executable files to individually evaluate and extract executable files present on a computer, and that might cause malicious behavior. This allows investigators to identify files that can potentially cause malicious behavior for detailed investigation in a relatively short time, even if the investigators have limited technical expertise. In this research, we created a prototype system for the method of classifying executable files on a computer by extracting their characteristic attribute information as metadata, and performing a regression analysis of these metadata characteristics against known executable file types. When the created regression model was applied to actual malware, it was able to evaluate the majority of the malware as software with a low level of
trustworthiness. This allows investigators to reduce the number of files that should be subjected to detailed analysis under the circumstances of an actual computer security incident. Profile generation by regression analysis successfully determined 95.6% of the malware in the sample as executable files other than those associated with vendor applications, which are potentially dangerous.

The title of research paper on dynamic analysis is "Stealth malware analysis using taint propagation on virtual machine monitor." In this paper, a method for dynamically interpreting semantics information using taint propagation and automatically analyzing malware that uses code injection or rootkits. Also a prototype system for evaluation is developed and applied to analyze actual malware. Because the proposed method can accurately analyze even sophisticated malware of the type used in targeted attacks, it can contribute to swift comprehension and elimination of malware behavior.

Overhead on some parts of context changes, discontinuity in taint propagation, and insufficient countermeasures for malware that changes its behavior through commands remained as points for future improvement.

However, with regard to the problems that surround semantic gap, the proposed method solved them by monitoring data structures through taint analysis based on pages. Furthermore, the method realized VMI without agent insertion into guest OS, which was a factor in detection by malware.

5 Future Works
This research has presented a concept of automated malware analysis platform so the implementation of the system is still in its early stage. Some more improvements and examination with other approaches are expected as future research.
6 Reference (Publication)