Automated Malware Infection Signature Generation

Information & Communication Technologies, IT-Security / Advisor: Prof. Dr. Endre Bangerter
Project partner: Non-Disclosed / Expert: Dr. Andreas Greulich

Malicious software, commonly known as malware, has become a serious threat to the security of computer systems. While current anti-virus detection and defense solutions are important and useful, they remain, loosely speaking, not effective enough. Companies and organizations are often confronted with the problem of assessing the extent of infections by new malware. In this thesis, we have built a system for the automated generation of malware infection signatures with the objective to detect infected computers, which enables custom malware detections in organizations.

Problem Statement

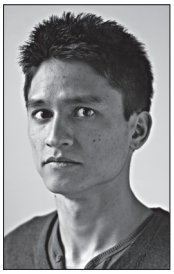
A particular problem for companies and organizations is to assess the extent of malware infections. Thus, how many and which computer systems are infected, once a piece of malware has been identified. Not only is this important for remediating the systems, but as well for preventing the proliferation of the malware more effectively, and to assess what information might has been leaked from infected computers.
An obvious approach to this problem would be to perform an organization wide scan using an anti-virus product. Yet, this only works for malware that is known to the respective anti-virus vendor. This is particularly not the case for targeted malware attacks, where the malware sample is usually unknown. Another approach to solve the problem is that the affected organization could submit the sample to its anti-virus vendor to obtain a detection signature. However, anti-virus vendors in general do not provide such a service for customized signatures. Furthermore, the organization under attack might not want to share the sample with third parties for confidentiality reasons.

Goal

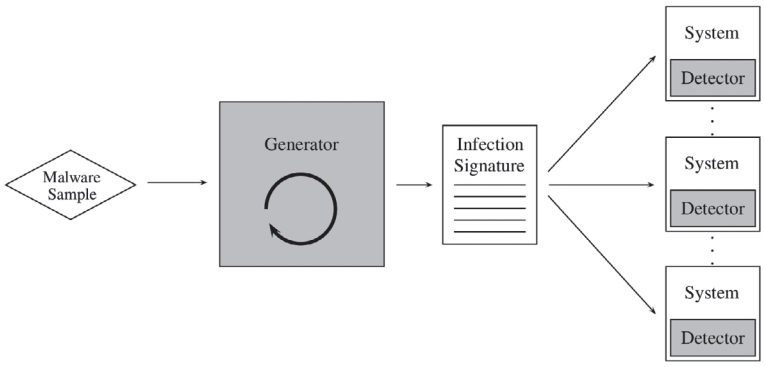
The goal of this thesis was to remedy this situation enabling companies and organizations to perform custom malware detection of infected computer systems. More precisely, the objective was to develop an automated custom infection signature generator that takes an arbitrary sample which is known to be malware as input and outputs a detection signature for the infection. The signature is then fed into a detection component that performs the malware detection on a potentially infected host.

Solution

We have developed a system that generates an infection signature for a given arbitrary malware sample. This is performed by first executing the sample in a virtual machine where it is monitored. The monitoring results are then analyzed by a heuristic algorithm that we have developed in order to identify and extract malicious memory regions. These are regions of the malware that exist in memory due to its infection. Finally, byte string candidates identifying the malware are created out of these regions, and then sound candidates with respect to false positives and negatives are selected. For the selection, we have adopted a machine-learning approach proposed by Hancock (Griffin et al.), which is based on a benign software model. We have as well developed a detector that applies the infection signature in order to detect infected systems.



Jonathan Wey
thesis@trycatch.ch



High-level view of the infection signature generation and detection.

Results

The effectivity of our system has been evaluated in an environment that simulates an organization as realistically as possible. We have conducted a large-scale measurement on a mass malware set. The obtained results suggest that the detection of infected systems with the generated infection signatures is sound in terms of the false positive and negative rate, and as well for infections by variants of the malware.