Challenge
Malware authors usually obfuscate their original executable to make their code more difficult to detect, decode and neutralize. Malware must be de-obfuscated before it can be detected and subsequently classified.

Focus
This CA Labs research project created a catalog of obfuscation techniques, detection technologies, and a framework of reverse engineering processes that automatically de-obfuscates malware.

Result
A general de-obfuscation system supports a small, efficient multi-purpose detection and reverse engineering engine that identifies malware samples en mass without requiring case-by-case analysis. Malware is detected and classified faster so neutralizing signatures can be published.

Figure 1 Packers have unique byte frequency profiles that are used to automatically identify and reverse engineer malware obfuscation protections.

A wide variety of obfuscation techniques can be used in the formation of malware executables in an effort to hinder detection. Obfuscation techniques employ various packers that can be identified so that malicious code can be unpacked for detection and classification.

Most malware contains multiple common obfuscation techniques that rely on packers such as UPX, FSG, MEW, Upack, PCShrinker, PECompact, Morphine, ASPack, ASProtect, tElock, Armadillo, Themida, VMProtect and more. Untangling the web of obfuscation technologies in malware is the first step toward stopping a malicious attack.

The CA Labs Malware De-Obfuscation research project devised a catalog of known obfuscation techniques and their packers, and invented detection and dissection technologies that work to identify packers based on their unique byte frequency characteristics and reverse engineer the obfuscation to reveal malware so it can be neutralized. The technologies use various algorithms, including one that analyzes byte frequency characteristics from ‘slices’ of a malware instance. The framework of unpacker technologies reverses or removes obfuscations returning malware instances to their pre-obfuscated states.

Anti-virus engineers can benefit from the framework’s automated detection and disassembly of common obfuscation techniques because it helps speed analysis and classification processes.
CA Labs malware de-obfuscation research project algorithms

Three primary algorithms are used in the technologies developed through this research project:

- **Counter measures to VM detection.** When malware runs on a VM, it can change its behavior which makes it harder to detect. Researchers invented an algorithm that makes malware think it is not running on a VM when it is running on a VM. Maintaining the malware in a stable state during analysis allows dissection and unpacking processes to occur.

- **An automated classification system for packers.** A packer generates output with a unique byte frequency distribution when used to obfuscate malware. Researchers invented an algorithm based on Huffman coding that draws from a catalog of packer frequency profiles to determine which packer was used to pack a malware instance.

- **Generic hump-and-dump unpacking algorithm.** Malware must be unpacked into memory before it can be executed. By capturing the memory state directly after the unpacking process, the hump-and-dump algorithm uses this snapshot to reconstruct the original executable for detection and classification. This generic algorithm and technique is applicable to a wide range of packers.

More information on CA Labs Malware De-Obfuscation research project

CA Labs is collaborating with researchers from RMIT University in Australia. The following papers have been published about this research project:


For additional information about this or other CA Labs projects, please contact Steve Versteeg at Steve.Versteeg@ca.com.

About CA Labs and innovation

CA Labs is the research arm for CA Technologies and a hub for the company’s initiatives for innovation. CA Labs collaborates with the world’s foremost researchers in academia, industry and government to perform advanced research to address cloud, software-as-a-service, security, virtualization, automation, mainframe, service assurance, and service and portfolio management challenges. For more information, visit ca.com/calabs.