Malicious Software (Malware) Detection Using Byte Sequence Analysis

Technology #m01-039

"Lead Inventor: Salvatore J. Stolfo, Ph.D.

Malware and Viral Software Cause Disruption and Data Loss, Requiring Better Screening Strategies Malware and viral software is a serious problem in information technology which costs companies worldwide greater than $10 billion dollars per year. Malicious code embedded in software and/or email correspondence can cause significant disruption and loss of data for corporations and individuals that fall susceptible to an attack. Given the constant evolution and reprogramming of malicious code, techniques which can identify malicious software based upon more general pattern recognition are needed for better screening strategies and identification of culprit software.

Data-Mining Method to Identify Malicious Code with Byte Sequence Analysis This technology demonstrates a novel data-mining method for identification of malicious code using byte sequence analysis. The strategy of this technique is to extract byte sequences from a software executable and compare those sequences to a dataset of known malicious software. The executable can then be classified using a classification rule set derived from a dataset of byte sequence features that are known to be related to malicious or benign software. Using this technique, a probability can then be assigned to the program in question on whether the byte sequences are likely to be malicious or benign. In addition, given the user defining additional executables as malicious, the dataset can be expanded and further used to refine the classification rule.

Applications:
• Virus/Malware Scan
  o Email attachments
  o Software
  o Data

Advantages:
• Novel classification technique for identifying malicious code
• Real-time data mining
• Continuously updated classification rule


Licensing Status: Available for Licensing and Sponsored Research Support

Inventors

Salvatore Stolfo