Abstract

Malware is basically malicious software or programs which are a major challenge or major threats for the computer and different computer applications in the field of IT and cyber security. Traditional anti-viral packages and their upgrades are typically released only after the malware’s key characteristics have been identified through infection. But by this time it may be too late to protect systems. Multiple sequence analysis is widely used in bioinformatics for helpful the genetic multiplicity of organisms and annotating gene functions through the identification of common genetic regions. This paper adopts a new approach to the problem of malware recognition, which is to use multiple sequence alignment techniques from bioinformatics to align variable length computer viral and worm code so that core, invariant regions of the code occupy fixed positions in the alignment patterns. Data mining (ANNs, symbolic rule extraction) can then be used to learn the critical features that help to determine into which class the aligned patterns fall. Experimental results demonstrate the feasibility of our novel approach for identifying malware code through multiple sequence alignment followed by analysis by ANNs and symbolic rule extraction methods.
Detection of Malware by using Sequence Alignment Strategy and Data Mining Techniques

- Vivek kumar, Dr. Sadhna K Mishra, Prof. Vineet Ricchariya "Detection of malicious software by Using Data Mining Tools and Other Techniques- a Survey", IJCSMR volume (1) issue 4, 2012, pp-746-750.
- Data Mining for Malicious Code Detection and Security Applications
- Data Mining: "Concepts and Techniques Jiawei Han and Micheline Kamber"
- Thomas M. Chen and Gregg W. Tally "Malicious Software. pdf";

Index Terms

Computer Science

Security
Keywords
Malware  sequence alignment  viral signatures  HEX editor  ASCII code  ARFF  Dev C++  
Java eclipse