Abstract

The malicious codes are normally referred as malware. Systems are vulnerable to the traditional attacks, and attackers continue to find new ways around existing protection mechanisms in order to execute their injected code. Malware is a pervasive problem in distributed computer and network systems. These new malicious executables are created at the rate of thousands every year. There are several types of threat to violate these components; for example Viruses, Worms, Trojan horse and Malware. Malware represents a serious threat to confidentiality since it may result in loss of control over private data for computer users. It is typically hidden from the user and difficult to detect since it can create significant unwanted CPU activity, disk usage and network traffic. In existing systems, new
malicious programs can be detected by automatic signature generation called as F-Sign for automatic extraction of unique signatures from malware files. This is primarily intended for high-speed network traffic. The signature extraction process is based on a comparison with a common function repository. The data mining framework employed in this research learns through analyzing the behavior of existing malicious and benign codes in large datasets. We have employed robust classifiers, namely Naïve Bayes (NB) Algorithm, k-Nearest Neighbor (kNN) Algorithm, and J48 decision tree and have evaluated their performance. This involves extracting opcode sequence from the dataset, to construct a classification model and to identify it as malicious or benign. Our approach showed 98.4% detection rate on new programs whose data was not used in the model building process.

References

- Symantec, "Symantec internet security threat report: Volume XII," effectiveness and efficiency of our work is in In their Symantec 2008.
- F-Secure. (2007, 19 August 2009). F-Secure Reports Amount of variants such as the Netsky family of malware using the Malware Grew by 100% during 2007.
- G. Bonfante, M. Kaczmarek, and J. Y. Marion, Morphological Detection


Index Terms

Computer Science

Egovernance And Cloud Computing Services

Keywords

Malware F –sign Naïve Bayes Algorithm J48 Algorithm K-nearest Neighbor Algorithm