Malware classification framework for dynamic analysis using Information Theory

ABSTRACT

Objectives: 1. To propose a framework for Malware Classification System (MCS) to analyze malware behavior dynamically using a concept of information theory and a machine learning technique. 2. To extract behavioral patterns from execution reports of malware in terms of its features and generates a data repository. 3. To select the most promising features using information theory based concepts. Methods/Statistical Analysis: Today, malware is a major concern of computer security experts. Variety and in- creasing number of malware affects millions of systems in the form of viruses, worms, Trojans etc. Many techniques have been proposed to analyze the malware to its class accurately. Some of analysis techniques analyzed malware based upon its structure, code flow, etc. without executing it (called static analysis), whereas other techniques (termed as dynamic analysis) focused to monitor the behavior of malware by executing it and comparing it with known malware behavior. Dynamic analysis has proved to be effective in malware detection as behavior is more difficult to mask while executing than its underlying code (static analysis). In this study, we propose a framework for Malware Classification System (MCS) to analyze malware behavior dynamically using a concept of information theory and a machine learning technique. The proposed framework extracts behavioral patterns from execution reports of malware in terms of its features and generates a data repository. Further, it selects the most promising features using information theory based concepts. Findings: The proposed framework detects the family of unknown malware samples after training of a classifier from malware data repository. We validated the applicability of the proposed framework by comparing with the other dynamic malware analysis technique on a real malware dataset from Virus Total. Application: The proposed framework is a Malware Classification System (MCS) to analyze malware behavior dynamically using a concept of information theory and a machine learning technique.

Keyword: Information theory; Malware classification; Mutual information; Neural network