Abstract

Internet worms can spread very fast and cause losses both in terms of lost business opportunities as well as human resources required to alleviate the caused damages. There exists two ways of protecting against the worms namely anomaly based and signature based systems. Signature based systems depends on security signatures (patterns) that match particular known attacks while anomaly based systems relies on detecting anomalies with the background idea that abnormal activity is malicious. With the ever increasing internet speeds and growing complexity of data across it, it is necessary to have correspondingly fast ways of analyzing network traffic in order to categorize activities in time. Also the existence of zero-day attacks makes relying of preconfigured signatures unreliable. This study sought to find how to develop an accurate, robust near real time machine driven Internet security signature detection, generation and collection system using big data technologies such as Hadoop Map Reduce programming model and Hadoop Distributed File System. We set up Hadoop Ecosystem at the University of Nairobi Laboratory and gathered and analyzed both malicious and innocuous network traffic and generated documented security signatures for known Internet worms with near real time speeds and also corresponding signatures for synthetic worms to simulate zero-day worms. We realize that adding the number of nodes to the Hadoop cluster not only increases the processing speeds but also eases the resources for the signature generation system. The increased power of the system improves accuracy and the HDFS replication improves system robustness.