Abstract

Malicious software and especially computer worms cause significant damage to organizations and individuals alike. The detection of computer worms faces a number of challenges that include incomplete approximations, code morphing (polymorphism and metamorphism), packing, obfuscation, tool detection and even obtaining datasets for training and validation. The challenge of incomplete approximations can partially be solved by feature selection.

Generally, only a small number of attributes of binary or network packet headers show a strong correlation with attributes of computer worms. The goal of feature selection is to identify the subset of differentially expressed fields of network packet headers that are potentially relevant for distinguishing the sample classes and is the subject of this study. The datasets used for the experiments were obtained from the University of San Diego California Center for Applied Data Analysis (USCD CAIDA). Two sets of datasets were requested and obtained from this telescope. The first is the Three days of Conficker Dataset ([2]) containing data for three days between November 2008 and January 2009 during which Conficker worm attack ([4]) was
active. It was found out that is well known dstport, ip len, value, ttl and China were the most instructive features.

References


Index Terms

Computer Science Security
Keywords