**Conservation of Feature Sub-Spaces across
Rootkit sub-families**

Prasenjit Das

1 Chitkara University, Baddi, Himachal Pradesh

Prasenjit.das@chitkarauniversity.edu.in

**Abstract.** Modern malware detection systems have largely relied on the definition of signatures to characterize malwares to their corresponding malware families. These signatures that characterize malware families are parts of codes and it is believed that families of malwares share commonalities in their signatures. We hypothesize that changes in these signatures generate newer sub-families of malwares. In the present work we have evaluated the signature conservation across two sub-families of rootkits. We have carried out our experiments to establish the fact that features in the rootkit family of malware are conserved. We report that our feature extraction yielded the accuracy of 84.17% using the Naïve Bayes classification algorithm. The results reported in this work reinforce our belief that there are subsets of independent features that discriminate between sub-families but not exhibiting any trend of conservation. We conclude that certain features (if not all) are preserved and discriminate between sub-families.

**Keywords:** Data mining, Malware, Rootkit, Classification, Clustering, Bi-Clustering

1 Introduction

Information that is available on the standalone as well as network systems is constantly under the threat of being attacked by software trying to malign the working of the system or steal the data. In general, most of the malwares (malicious software’s) are commonly referred to as a virus. Considering the detrimental effects of these malicious software’s, organization largely rely on intermediate intrusion detection systems to detect and block malwares from entering and causing damage. It should be noted that not all ‘virus attacks’ are caused by viruses, requiring intrusion detection systems to be constantly up-to-date with known malware families and their corresponding morphological changes as these malwares constantly evolve, rendering detection ineffective. Malwares are broadly categorized based on the criteria such as: the manner by which these malware enters into the system, the manner by which they execute, and by the associated damage that they can cause.

Reported by Microsoft TM, there exists close to two hundred families of malwares. Some of the major malware families are *Computer Virus, Trojan horse, Worms, Spyware*