A Study on Applying Effective Cluster Analysis Schemes to Intrusion Detection Systems

Lin Dongshen, Cao Weijun

E-mail: 9222744@mail.dyu.edu.tw

Abstract

In recent years, along with the popularity of Internet, the pervasion of the computer usage and the maturity of the network technique, it makes Internet become the media of commercial transaction, and attract the attention from society. On the contrary, it also causes hackers' attacks and a variety of network crimes. The potential damage that caused by the intrusion is not only difficult to estimate, but expose the security problem existing in the system. The host-based intrusion detection system is to defend and detect attack behaviors from hacker via log files in the server. If we can analyze log files precisely and efficiently, it will be useful to establish an intrusion detection system. Therefore, in this thesis we mainly analyze log files by using clustering and classification methods to distinguish normal and abnormal behaviors. Although the clustering method can analyze huge log files, it generally needs to set up a known cluster number in advance before starting clustering; however, it is difficult for us to decide the known number. If we employ a cluster validity index to evaluate automatically the number of clusters, we can acquire better results. When the amount of log files are gradually added, clustering procedure needs to be run again. If we can adopt the advantage of classification, we can handle cumulative log files efficiently. Based on the requirements stated above, we integrate a faster clustering algorithm into the cluster validity index to solve the cluster analysis problems, and further use the classification rule of modified K-nearest neighbor to handle the increasingly cumulative log files. The results derived in this thesis can be used to construct the preceding part of an intrusion detection system. Furthermore, we also develop a practical system. Through this implemented system, we can cluster and classify log files, and validate the feasibility of the proposed methods in this thesis.

Keywords: Intrusion detection system, Clustering, Cluster validity index, Classification.


