Filtering events using clustering in heterogeneous security logs

ABSTRACT

Log files are rich sources of information exhibiting the actions performed during the usage of a computer system in our daily work. In this study we concentrate on parsing/isolating logs from different sources and then clustering the logs using data mining tool (Weka) to filter the unwanted entries in the logs which will greatly help in correlating the events from different logs. Unfortunately parsing heterogeneous logs to extract the attribute values becomes tedious, since every type of log is stored in a proprietary format. We propose a framework that has the ability to parse and isolate a variety of logs, followed by clustering the logs to identify and remove unneeded entries. Experiments involving a range of logs, reveals the fact that clustering has the capacity to group log entries with a higher degree of accuracy, thereby assisting to identify correctly the entries to be removed.

Keyword: Event filtering; False; False positive rate; Heterogeneous log parsing; Multi-level clustering