World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:8, No:03, 2014

Constructing a Semi-Supervised Model for Network Intrusion Detection

Authors : Tigabu Dagne Akal

Abstract : While advances in computer and communications technology have made the network ubiquitous, they have also rendered networked systems vulnerable to malicious attacks devised from a distance. These attacks or intrusions start with attackers infiltrating a network through a vulnerable host and then launching further attacks on the local network or Intranet. Nowadays, system administrators and network professionals can attempt to prevent such attacks by developing intrusion detection tools and systems using data mining technology. In this study, the experiments were conducted following the Knowledge Discovery in Database Process Model. The Knowledge Discovery in Database Process Model starts from selection of the datasets. The dataset used in this study has been taken from Massachusetts Institute of Technology Lincoln Laboratory. After taking the data, it has been pre-processed. The major pre-processing activities include fill in missed values, remove outliers; resolve inconsistencies, integration of data that contains both labelled and unlabelled datasets, dimensionality reduction, size reduction and data transformation activity like discretization tasks were done for this study. A total of 21,533 intrusion records are used for training the models. For validating the performance of the selected model a separate 3,397 records are used as a testing set. For building a predictive model for intrusion detection J48 decision tree and the Naïve Bayes algorithms have been tested as a classification approach for both with and without feature selection approaches. The model that was created using 10-fold cross validation using the J48 decision tree algorithm with the default parameter values showed the best classification accuracy. The model has a prediction accuracy of 96.11% on the training datasets and 93.2% on the test dataset to classify the new instances as normal, DOS, U2R, R2L and probe classes. The findings of this study have shown that the data mining methods generates interesting rules that are crucial for intrusion detection and prevention in the networking industry. Future research directions are forwarded to come up an applicable system in the area of the study.

Keywords : intrusion detection, data mining, computer science, data mining
Conference Title : ICCSDM 2014 : International Conference on Computer Science and Data Mining
Conference Location : Miami, United States
Conference Dates : March 10-11, 2014