

Decision Trees for Signature Recognition and State Classification

Nong Ye and Xiangyang Li
Arizona State University

Purpose

This study uses one of data mining techniques, decision trees, to perform signature recognition for intrusion detection. Decision Tree classifier is used to learn intrusion signatures for intrusion detection, and to classify the activities in a computer and network system into various states for information aggregation and aggregate intrusion analysis.

Method

Computer audit data are used to capture activities in a computer and network system. Computer audit data consist of audit events. There are more than 284 auditable event types in a UNIX-based computer system. The training data for our decision tree classifiers include audit events of intrusive activities and normal activities. A different set of audit events for intrusive activities and normal activities is used as the testing data. Our design of decision tree classifiers is based on the ITI algorithm developed at University of Massachusetts. The ITI algorithm is modified to meet the requirements of intrusion detection.

We design two versions of decision tree classifiers respectively: single-event and moving window. For the single-event version, a decision tree classifier is built by examining the audit events in the training data one at a time and learning signature patterns of both intrusions and normal activities from the training data. This decision tree classifier is then used to examine the audit events in the testing data in the similar way and to classify each audit event by generating a rating of intrusion and a related state ID. This decision tree classifier places 284 types of auditable events into a small number of categories - state IDs - for information aggregation. For the moving-window version, aggregate intrusion analysis is performed to detect coordinated actions of intrusions over time. In training a decision tree classifier is built to examine a sequence of events at a time generated by moving an observation window with fixed size through the event stream and learning signature patterns of both intrusions and normal activities. Then in testing this classifier is used to examine the sequences of events in the testing data and to classify each sequence by giving a rating of intrusion and a state ID for that sequence.

Based on this approach, we make a layered design of decision tree classifiers. At the lower level the streams of audit events in the training data and the testing data are transformed into a stream of state IDs using a single-event classifier. Then this stream of state IDs is used as the input data for a called state-ID classifier at the upper level. This decision tree classifier at the upper level is built in the same as above moving-window classifier except that the sequences of events are replaced by the sequences of state IDs produced by the lower level classifier.

Results

For each of the two versions of decision tree classifiers and the layered design, we compute false alarm rates and detection rates using various decision thresholds, and analyze the Receiver Operating Characteristic (ROC) curves of the intrusion detection performance on the testing data. The moving window classifier shows good performance in terms of false alarm rate and detection rate. From the comparison of ROC curves, the upper level state-ID classifier in the layered design is better than the moving-window classifier depending on the original event stream.

New or Breakthrough Aspects of Work

The automatic learning capability is lacking in the existing signature recognition techniques for intrusion detection. Without the automatic learning capability, it is difficult to manually code all intrusion signatures and update intrusion signatures as computing technology and intrusions change over time. In our research Decision Tree classifier is used to automatically learn intrusion signatures during the training. The various designs of decision classifier show different abilities in finishing this task and provide some insight into this area. Decision tree classifiers for state classification and information aggregation also assist a layered mechanism of intrusion detection for better performance.

Conclusions

The results of this study demonstrate the promising performance of the decision tree technique in intrusion detection and information aggregation.