International Journal of Innovative Computing, Information and Control
Volume **5**, Number **11(A)**, November **2009**

 ISSN 1349-4198
pp. 3625–3635

ENHANCED STOCHASTIC LEARNING FOR FEATURE SELECTION IN INTRUSION CLASSIFICATION

Gil-Jong Mun[1], Bong-Nam Noh[2] and Yong-Min Kim[3,*]

[1]INFOSEC Technologies Co., Ltd.
[2]System Security Research Center
[3]Department of Electronic Commerce
Chonnam National University
Yongbong-dong, Buk-gu, Gwangju, 500-757, Korea
alcorjjong@infosec.co.kr; bbong@chonnam.ac.kr
*Corresponding author: ymkim@chonnam.ac.kr

Received July 2008; revised December 2008

Abstract. *Feature selection is very important in various fields due to the enormous data processing requirements. Many researchers are investigating data reduction and feature selection, particularly in network traffic reduction for network intrusion detection systems. In this paper, we suggest a method that selects the useful features for intrusion classification, decreases the storage of rules and computational time, and increases the classification accuracy. A proposed many-to-many Kullback-Leibler (K-L) divergence is applied to the probability distribution that is calculated by the histogram estimator to select the specific features. This method is an improvement on the previous method that only calculated the distance between the normal and each intrusion. To verify the method, we present experimental results of the classification rates and false positive rates for accuracy, the number of rules generated, and the features selected for accuracy and faster detection. The results of the applying method show that the number of selected features is reduced from 41 to 19. Also, the accuracy rate is increased by 0.1588 pecent, whereas the false positive rate is decreased by 0.0033 percent. Therefore, we confirm the classification accuracy of the proposed method and support its usefulness for data reduction.*

Keywords: Feature selection, Data reduction, Intrusion detection, K-L divergence

1. **Introduction.** Data reduction techniques are required in a variety of fields because of the enormous data processing needs. These techniques consist of data filtering, feature selection, and data clustering. Intrusion detection research consists of techniques for intrusion prevention, detection, and data reduction. In this paper, we suggest a method to select useful features for intrusion data reduction and classification accuracy. We repeatedly calculate the Kullback-Leibler (K-L) divergence between each feature of each intrusion datum. This method is an improvement on the previous method (one-to-many learning), which only calculated the distance between the normal and each intrusion datum, and easily determines the characteristics of the intrusion behaviors. In these experiments of features selection, we use forty-one components of statistical information that are in the KDDCUP 99 dataset [1]. We firstly explain the related research on feature reduction, feature selection of intrusion detection, and our previous research. In addition, we present an experimental dataset, the proposed method (many-to-many learning), and the suggested system. Lastly, we use graphs and tables of the experimental results for the selected features to compare them according to the distance, number of rules, and detection performances.