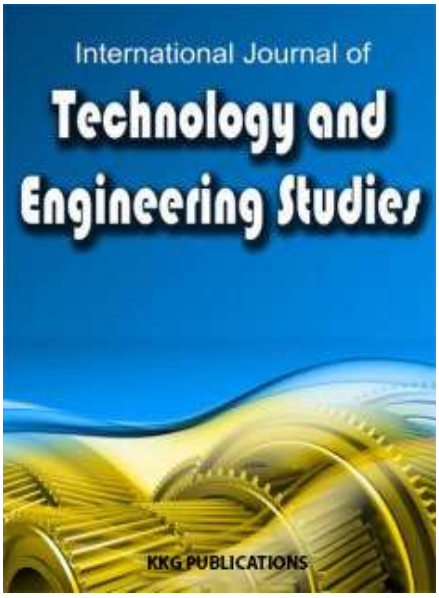


**Key Knowledge Generation**
Publication details, including instructions for author and
Subscription information:
http://kkgpublications.com/technology/

# Detecting TCP Based Attacks Using Data Mining Algorithms

UGTAKHBAYAR N. [1], USUKHBAYAR B. [2], SODBILEG SH. [3], NYAMJAV J. [4]

[1, 2, 3, 4] National University of Mongolia, Mongolia

Published online: 29 February 2016

**To cite this article:** N. Ugtakhbayar, B. Usukhbayar, S. H. Sodbileg, J. Nyamjav, "Detecting TCP based attacks using data mining algorithms," *International Journal of Technology and Engineering Studies*, vol. 2, no. 1, pp. 1-4, 2016.
DOI: https://dx.doi.org/10.20469/ijtes.2.40001-1

**To link to this article:** http://kkgpublications.com/wp-content/uploads/2016/2/Volume2/IJTES-40001-1.pdf

KKG PUBLICATIONS

International Journal of Technology and Engineering Studies

IJTES

Vol. 2, no. 1, pp. 1-4, 2016

# DETECTING TCP BASED ATTACKS USING DATA MINING ALGORITHMS

UGTAKHBAYAR N. [1*], USUKHBAYAR B. [2], SODBILEG SH. [3], NYAMJAV J. [4]

[1, 2, 3, 4] National University of Mongolia, Mongolia

**Keywords:**
Data Mining
Learning Algorithms
Network Attacks
Intrusion Detection
IDS

**Received:** 17 October 2015
**Accepted:** 02 December 2015
**Published:** 29 February 2016

**Abstract.** This research studies the effects of TCP-based attacks on AI algorithms computing time and detection ratio using the KDDCUP dataset and the collected dataset. This study gathers network traffic; normal and abnormal containing attacks are collected by SNORT. It also extracts features in TCP headers of the packets in the collected dataset such as sequence and acknowledges numbers, window size, control flags, and an event which is the time between neighbor segments. First, the feature set is normalized to reduce our input feature space dimensionality and apply Pearson correlation to measure the dependability of the relationship. Finally, the selected subset of the features is given to learning the classifiers: J-48, Naïve Bayes, and ANNs. By adopting machine learning and data mining concepts, we could detect 98% of abnormal traffic containing attacks.


## INTRODUCTION

Network security is still quickly developing in any information technology fields. In the last few years, due to the growing use of computer networks, network traffic is immediately increasing. There are several private as well as business sectors, government organizations that store valuable data over the computer network. Cause, new threats are showing up on quickly, while older often abide relevant. Therefore, more dynamic mechanisms such as Intrusion Detection Systems are should also be utilized.

A Cisco report found the following: "Global IP traffic in 2012 stands at 43.6 Exabyte's per month and will grow threefold by 2017, to reach 120.6 Exabyte's per month, by 2019, there will be 24 billion networked devices and connections globally" [1].

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible attacks [2], [3]. Intrusion detection mechanism is divided in two; anomaly detection and misuse detection. Misuse detection is an approach where each suspected attack is compared to a set of known attack signatures [4].

It is in an exclusive manner the attacks in that database that can be detected, this method does not can for detection of unknown attacks. Unknown attack can be most zero day attacks. The role of anomaly detection is the identification of data points, substance, event and observations or attacks that do not conform to the expected pattern of a give collection [5].

Network traffic speeds and volume are increasing at an exponential rate. The conventional approach of tuning the hardware and software of the NIDS platform to maximize its performance can yield considerable improvements, but falls short in supporting next-generation networks operating at gigabits per second and faster.

This paper worked for neural networks as a data processing technique, it can increasing detection ratio using data mining algorithms and decreasing processing time using correlation technique and feature selection algorithms. We are using DARPA, CAIDA datasets and our university gateway traffic in this research. The DARPA, CAIDA dataset have been used for evaluating the most eminent algorithms available in the literature for feature selection and classification.

In Methodology section, we presented Pearson correlation and J48, Naïve Bayes and ANNs in our collected and DARPA, CAIDA datasets. In finally section we are summarize our paper and give final conclusion. DARPA dataset consists of nearly 5 million training connection records labeled as an intrusion or not an intrusion, and separate testing dataset consists of seen and unseen attacks [6].

## LITERATURE REVIEW

In the last years, network security has been the subject of many researcher. There are many works in the literature that discuss about information security, Intrusion detection system, using artificial intelligence and data mining in intrusion detection system. Intrusion detection and prevention systems used to detect and prevent the known and unknown attacks made by intruders [7], [8]. In this paper, there are presented an IDS that uses IDS for effective intrusion detection. One of the disadvantage of their approach is that it increases the time in training. In the literature

* Corresponding author: Ugtakhbayar.N
E-mail: ugtakhbayar@num.edu.mn

[9] proposed a new method based on Continuous random function for selecting appropriate feature sets to perform network intrusion detection. Also, [10] have used Learning Vector Quantisation neural networks to detect attacks that is supervised version of quantization, which can be used for pattern recognition, multi-class classification and data compression tasks. In paper [11] has written a highly referenced article about intrusion detection using neural networks. In the article, he studies in detail the advantages and disadvantages of neural networks for this application. In the conclusion of this article that neural networks are very suitable for Intrusion detection system. The [12] have used a neural network to detect the number of zombies that have been involved in DDoS attacks. The objective of their work is to identify the relationship between the zombies and in sample entropy. [13] used genetic algorithm to extract optimized information from raw internet packets. [14] applied J48 decision tree algorithm to determine significant features from KDDCUP 1999 dataset for anomaly intrusion detection. And experimental results demonstrate that the Hidden Markov model is able to classify network traffic with approximately 76% to 99%. Most proposed techniques utilize characteristics of network traffics to identify abnormalities absolutely. But, performing the real time network traffic detection with maintaining higher accuracy is restricted due to complex nature of networks. In this paper, we are focus on detection ratio and performance.

## METHOD AND MATERIALS

This research focuses on solving the issues in Intrusion Detection methods that can help the network and system administrators to make pre-processing, classification of network traffic. Most of attacks can be identified only after it happens. Data mining approaches have been implemented by many researchers to solve the abnormal detection problem. In this section, we are explain the proposed methodology for anomaly intrusion detection. We concentrated on data mining such as J48 algorithm, Naïve Bayes and ANN classifiers, because data mining approaches use strong statistical foundations to enhancing the dynamic and accurate learning that gives better accuracy, reduce false alarm rates, performance improvements, ability to detect novelty, protection against zero-day exploits.

The entire framework of proposed methodology shown in figure 1, we are collected our university's internet and intranet traffic using Bro IDS by sensor. In the data collection section,

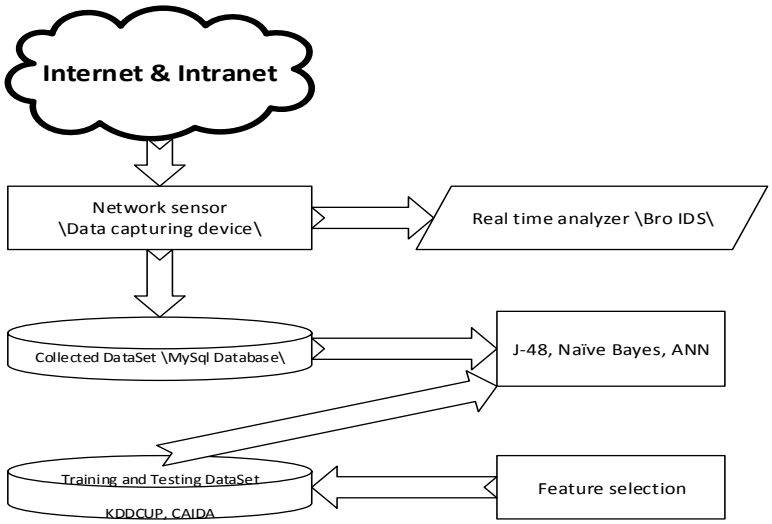


Fig. 1. Anomaly detection proposed method

In our approach, the dataset is divided into training and testing datasets. First, training data sets includes DARPA, CAIDA's labeled datasets. The labeled datasets are applied to J-48, Naïve Bayes, ANN classifier and the model is generated. Then change the datasets by Pearson correlation and generate the latest training datasets for each tuning process. After we are collecting testing datasets from National University of Mongolia's gateway firewall and router. Our sensor system Bro is running with the specifications of 2nd generation Intel Atom Dual core processor, 2GB DDR3 RAM disk, 128GB SSD hard disk. Testing dataset collected by net flow, TCP dump and applied Pearson correlation by static method.

## RESULTS

DARPA dataset has been used in this research work of which 50% is treated as training data and CAIDA is another 50% of training data. The proposed method has been implemented in Weka data mining tool.

First, we used Pearson correlation [shown as fig.2] method to improve performance and results.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

The sampled testing dataset after using Pearson correlation are shown in table 1.

TABLE 1
USED CORRELATION IN TESTING DATA SET

| ip_src | ip_dst | time | tcp_sport | tcp_dport | tcp_seq | tcp_ack | tcp_flags | tcp_win | tcp_urp |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.07121 | 0.549692 | 0.223295 | -0.10574 | 0.002207 | 0.165121 | -0.13162 | -0.13066 | 0.061697 |
| -0.07121 | 1 | 0.54553 | -0.18875 | 0.311146 | 0.217049 | -0.01575 | 0.074959 | -0.04194 | 0.069881 |
| 0.549692 | 0.54553 | 1 | 0.034116 | 0.18847 | 0.261126 | 0.141376 | -0.10384 | -0.08742 | 0.137321 |
| 0.223295 | -0.18875 | 0.034116 | 1 | -0.96924 | -0.463 | 0.442719 | -0.17518 | 0.059534 | -0.09109 |
| -0.10574 | 0.311146 | 0.18847 | -0.96924 | 1 | 0.511442 | -0.4119 | 0.142961 | -0.07498 | 0.106863 |
| 0.002207 | 0.217049 | 0.261126 | -0.463 | 0.511442 | 1 | -0.26352 | -0.04639 | -0.11145 | 0.039886 |
| 0.165121 | -0.01575 | 0.141376 | 0.442719 | -0.4119 | -0.26352 | 1 | 0.083495 | -0.12671 | -0.02341 |
| -0.13162 | 0.074959 | -0.10384 | -0.17518 | 0.142961 | -0.04639 | 0.083495 | 1 | -0.31938 | 0.055782 |
| -0.13066 | -0.04194 | -0.08742 | 0.059534 | -0.07498 | -0.11145 | -0.12671 | -0.31938 | 1 | -0.05166 |
| 0.061697 | 0.069881 | 0.137321 | -0.09109 | 0.106863 | 0.039886 | -0.02341 | 0.055782 | -0.05166 | 1 |

Detection rate for TCP attacks results shown as table 2 before using Pearson correlation. In this table, we are only chosen TCP related features from our features. In table 3, after using Pearson correlation.

TABLE 2
DOES NOT USED PEARSON CORRELATION

| Classification algorithms | J-48 | Naïve Bayes | ANN |
|---|---|---|---|
| True positive rate by % | 97.8% | 96% | 97.1% |

TABLE 3
USED PEARSON CORRELATION

| Feature selection | J-48 | Naïve Bayes | ANN |
|---|---|---|---|
| True positive rate by % | 98% | 97.2% | 97.9% |

## CONCLUSION AND RECOMMENDATIONS

In this paper, J-48, Naïve Bayes, ANNs for classification techniques for feature selection and classification techniques used of Intrusion detection has been presented and discussed. In addition, we are collected our University TCP traffics. The scope of this paper includes data mining algorithms, data sets and collected data sets. The Pearson correlation is highlighted in our result. And with this datasets we determined the detection rate of attacked traffic using J-48, Naïve Bayes and ANN algorithms. Further we will do feature selection differently to improve the results. Therefore, we will determine the detection rate after combine the normal and the attacked traffic for the test dataset.

### Declaration of Conflicting Interests

There are no conflicts of interest.

## REFERENCES

[1] I. Cisco, "Cisco visual networking index: Forecast and methodology," 2011–2016. *CISCO White paper, 2016*. (2011).

[2] Scarfone, K. and Mell, P. "Guide to intrusion detection and prevention systems (IDPS)," *NIST Special Publication*, vol. 800, no. 2007, pp. 94, 2007.

[3] L. Hanguang and N. Yu, "Intrusion detection technology research based on apriori algorithm," *Physics Procedia*, vol. 24, pp. 1615-1620, 2012.

[4] A. S. K. Pathan, Ed. *The State of the Art in Intrusion Prevention and Detection*, US: CRC Press, 2014.

[5] M. N. S. Lakshmi and Y. Radhika, "A complete study on intrusion detection using data mining techniques," *International Journal of Computer Engineering and Applications*, vol. 9, no. 6, pp. 130-137, 2015.

[6] M. Stampar and K. Fertalj, "Artificial intelligence in network intrusion detection," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, pp. 1318-1323, May, 2015.

[7] S. Balakrishnan, K. Venkatalakshmi and A. Kannan, "Intrusion detection system using feature selection and classification technique," *International Journal of Computer Science and Application*, vol. 3, no. 4, pp. 145-151, 2014.

[8] M. Moradi and M. Zulkernine, "A neural network based system for intrusion detection and classification of attacks," in *Proceedings of the 2004 IEEE International Conference on Advances in Intelligent Systems-Theory and Applications*, Nov. 2004.

[9] W. Jianping, C. Min and W. Xianwen, "A novel network attack audit system based on multi-agent technology," *Physics Procedia*, vol. 25, pp. 2152-2157, 2012.

[10] J. Li, Y. Liu and L. Gu, "DDoS attack detection based on neural network, in *Aware Computing (ISAC), 2010," 2nd International Symposium on*, pp. 196-199, Nov. 2010.

[11] J. Cannady, "Artificial neural networks for misuse detection," in *National Information Systems Security Conference*, pp. 368-81, Oct. 1998.

[12] B. B. Gupta, R. C. Joshi and M. Misra, "ANN based scheme to predict number of zombies in a DDoS attack," *International Journal Network Security*, vol. 14, no. 2, pp. 61-70, 2012.

[13] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Information Sciences*, vol. 177, no. 18, pp. 3799-3821, 2007.

[14] R., Jain and N. Abouzakhar, "A comparative study of hidden markov model and support vector machine in anomaly intrusion detection," *Journal of Internet Technology and Secured Transactions (JITST)*, vol. 2, no. 1/2, pp. 176-184, 2013.

— This article does not have any appendix. —