Georgia Institute of Technology

mabrams8@gatech.edu

Introduction to Firewalls   
and   
University of Firat Internet Firewall Data Analysis

Open Data Statistical Analysis

Abstract

Ever wondered how schools or workplaces monitored or blocked access to certain social media sites and suspicious internet activities? The following document presents the importance and role of firewalls in the protection of private networks. Network security is essential for protection of private data on academia, industry, and residential networks; firewalls are software and hardware devices that act as the first boundaries between the data traffic from outgoing and ingoing messages. This document will evaluate the functions of the firewall, analyze firewall data logs from The University of Firat, and an overview of the importance of network security and the role of firewalls in securing personal, administrative, and industry data.

Introduction to Firewalls   
and   
University of Firat Internet Firewall Data Analysis

Mia Abrams   
*School of Electrical and Computer Engineering*, *Georgia Institute of Technology*

Atlanta, GA   
[mabrams8@gatech.edu](mailto:mabrams8@gatech.edu)

*Abstract*— Ever wondered how schools or workplaces monitored or blocked access to certain social media sites and suspicious internet activities? The following document presents the importance and role of firewalls in the protection of private networks. Network security is essential for protection of private data on academia, industry, and residential networks; firewalls are software and hardware devices that act as the first boundaries between the data traffic from outgoing and ingoing messages. This document will evaluate the functions of the firewall, analyze firewall data logs from The University of Firat, and an overview of the importance of network security and the role of firewalls in securing personal, administrative, and industry data.

Keywords—firewall, network security, data analysis, and the University of Firat

# Introduction

In recent years, concerns surrounding internet security are open avenues for research focused on better protection and safeguards towards private networks for individuals, universities, companies, and so on. Firewalls are one of the first lines of protection for private networks, and with security threats ever advancing, optimizing firewalls are a vital component of network maintenance. Firewalls are software and/or hardware devices designed to protect private intranet networks from suspicious/unauthorized access connections over the internet. Incoming and outgoing messages are transmitted through a firewall which is programmed to access what action protocol to conduct based on specified security criteria by the network administrator. [2] This study will detail the functionality of the firewalls and assess how this first line of defense performs and can be optimized to safeguard networks. The research data has obtained through an open data platform called the UCI Machine Learning Repository; this data however was contributed by the Firat University: Department of Digital Forensics Engineering. Other resources included Indiana University: Knowledge Base and the University of Nevada supplying firewall background information and data logging analytical techniques. The goal of this document is to introduce firewalls to an general audience and provide a detailed data analysis for machine learning beginners. The following subtopics will be outlined as follows: Abstract, Introduction, Related Works, Proposed Methods, Results, Conclusions, Future Work and References.

# Related Works

## Classification of Firewall Log Files with Multiclass Support Vector Machine

The research paper associated with the data set analyzed for this document, which explores the classification of firewall logs based on machine learning algorithms. Each of the data log entries are classified into four categories: allow, deny, reset-both, and drop based on the security criteria and destination ports of the packages.

## Knowledge Base: About Firewalls

The site gives a detailed high-level overview of firewalls including different types and their functionalities. The resource goes into further detail introducing the varies types of firewalls and their respective purposes in internet security. Firewalls act as control gates for computer networks.

## A Firewall Data Log Analysis of Unauthorized and Suspicious Traffic

In this research paper, the authors describe how to accurately analyze the firewall system data logs and determine potential threats to a computer network system. The researchers, in addition, convey the safety risks, number of attacks, and where the attacks originated; to alert the overall effectiveness of the firewall system.

# Proposed Methods

The primary data models used on the Firat University Firewall Data included binary decision tree and linear regression analysis.

The following lines of code were used to execute the data model in R (packages and libraries not included):

Train <- createDataPartition (y=FirewallData$Bytes, p=0.8, list=F)

training <- FirewallData [Train,]

testing <- FirewallData[-Train,]

decisionTreeBinary <- rpart (Action ~ . , data = log2, cp=0.1)

fancyRpartPlot(decisionTreeBinary)

The decision tree assesses the pathway of the packages and first the percentage of which were allowed. In the second layer, the rejected packages at a common destination port are split into either deny or drop categories.

The following code was used to assess the linear regression of Bytes and Elapsed Time:

FirewallDataModel1<-lm(FirewallData$Bytes~FirewallData$`ElapsedTime (sec)`)

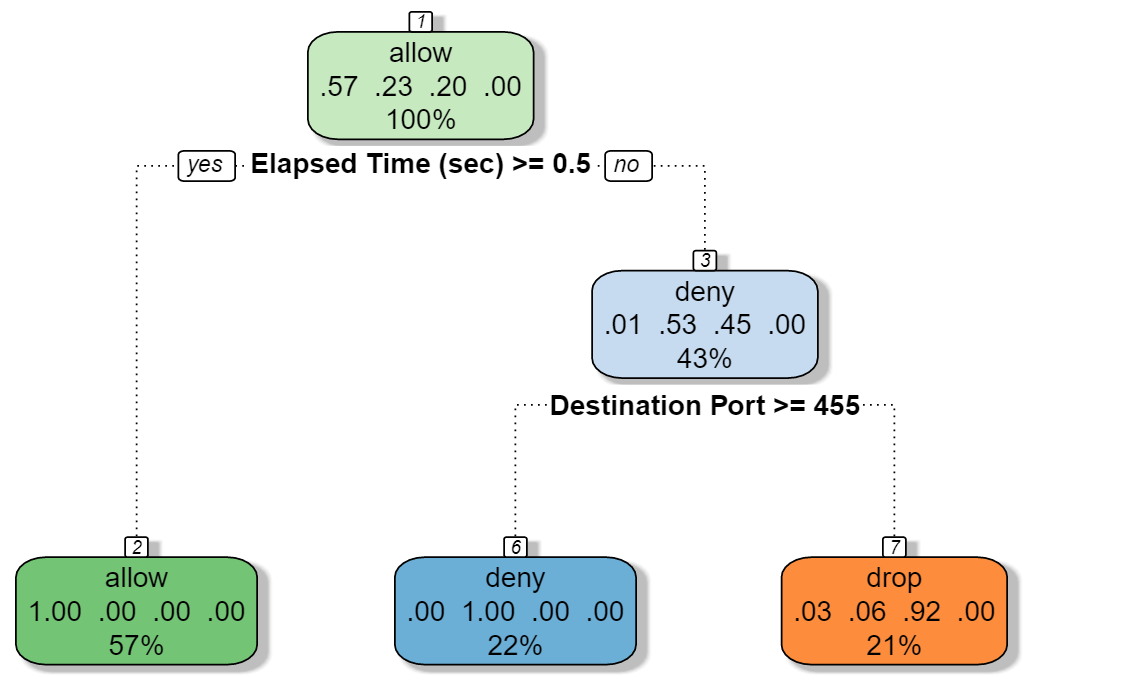
summary(FirewallDataModel1)

The null hypothesis is more bytes would equate to longer elapsed time for package transfer. The null hypothesis would be rejected in this case with R-squared of 0.02215.

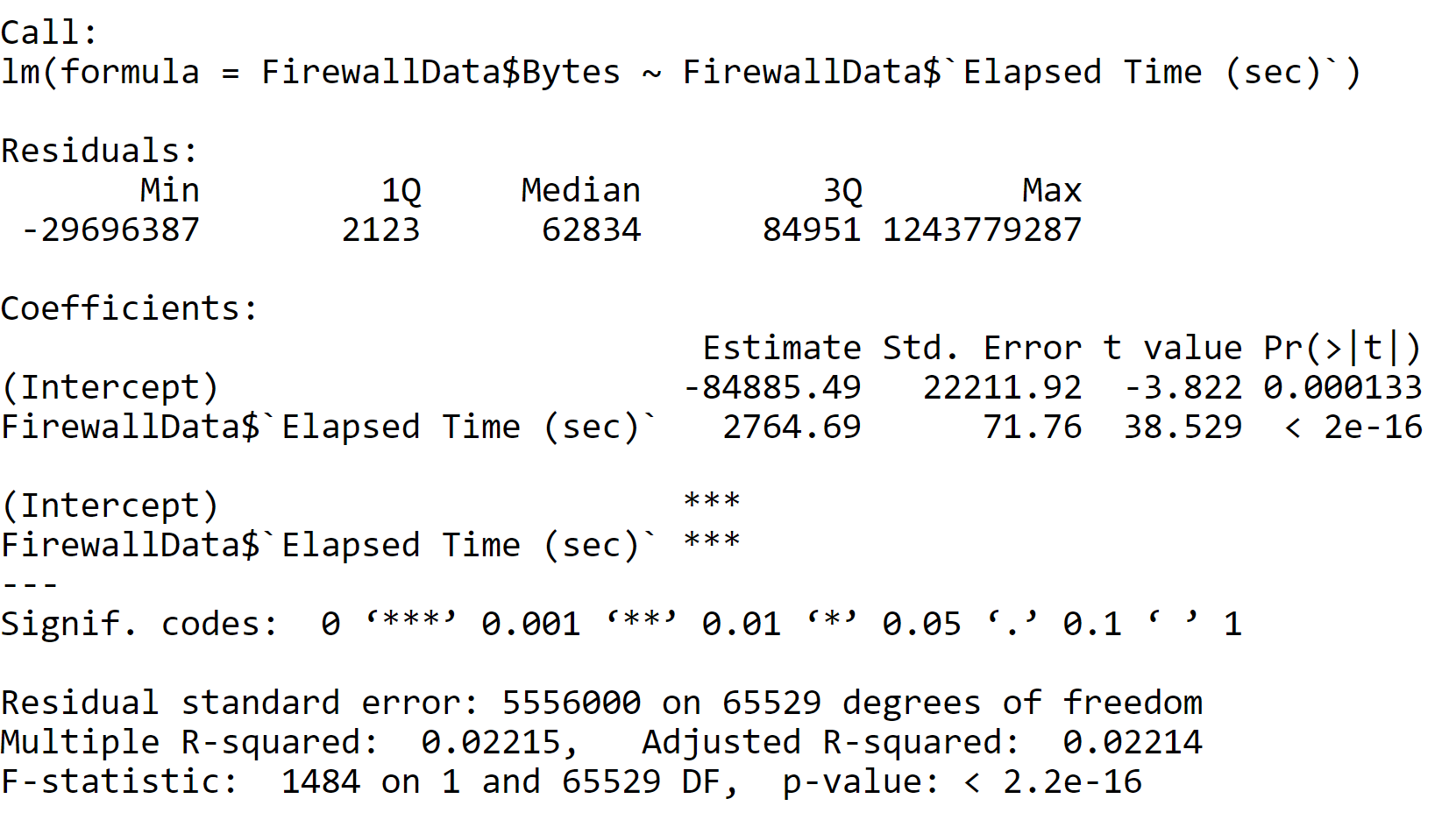
# Results

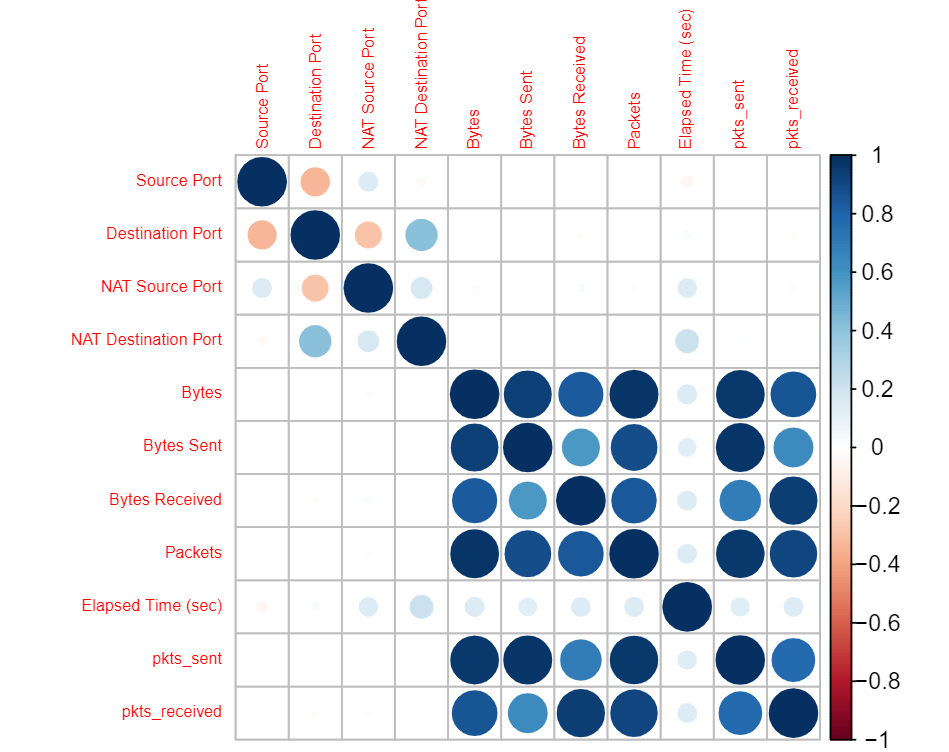
SVM Decision Tree

The figure below is the decision tree described in the proposed methods.



Linear Regression





## Denied Traffic

Approximately 23% of the incoming and outcoming traffic were attacks on the university’s intranet and as a result were denied access to the requested destination port.

# Conclusion

Firewalls are software and/or hardware devices designed to protect private intranet networks from suspicious/unauthorized access connections over the internet. Incoming and outgoing messages are transmitted through a firewall which is programmed to access what action protocol to conduct based on specified security criteria by the network administrator. [2] The research data was contributed by the Firat University: Department of Digital Forensics Engineering. Each of the data log entries are classified into four categories: allow, deny, reset-both, and drop based on the security criteria and destination ports of the packages. The null hypothesis of more bytes equating to longer elapsed time for package transfer was rejected with a R-squared of 0.02215. In addition, approximately 23% of the incoming and outcoming traffic were attacks on the university’s intranet and as a result were denied access to the requested destination port.

# Future Work

## Future Analysis Authors and Affiliations

If more time was allotted for research and data analysis, an overview of the specified destination ports which corresponded with the deny incoming messages would be assessed. In addition, the clustering of the source and destination port related to the chosen action: drop, deny, reset-both, and allow would be interesting and useful to understand.

##### References

1. Ertam, F.; Kaya, M. (2018). *Classification of Firewall Log Files with Multiclass Support Vector Machine.* Department of Digital Forensics Engineering, Firat University.

In this conference paper based on the research conducted by the Department of Digital Forensics Engineering of Firat University, Ertam and Kaya study logs produced by the university’s firewall devices. The data was collected via a multiclass support vector machine (SVM) classifier, where 655532 cases were obtained and evaluated for investigation. The firewalls are used as control gates to secure the university’s computer network, thus analyzing the firewall records was used to access the performance of these devices. The Action class attribute used in the study had the following parameters: allow, deny, drop, reset-both; which assess where data messages could be received by the university’s computer network safety. The research data is clearly defined and effectively assess the sensitivity of the firewall devices through machine learning; thus the vulnerability and effectiveness of the firewall system.

1. Knowledge Base. (2020). *About Firewalls.* Indiana University: Knowledge Base

On this basis knowledge information hub website, the Indiana University: Knowledge Base department delivers a detailed description of firewalls and their functionalities. The resource goes into further detail introducing the varies types of firewalls and their respective purposes in internet security. Firewalls act as control gates for computer networks. The system administrator sets up firewalls for each organization's needs. The writer addresses the audience in the introduction of the reference page by clearly defining firewalls, as to create a basis level of understanding of the topic before further description. The language is plain and easily comprehensible for the audience to digest. The packet-filtering explanation describes how firewall systems effectively allow or deny access to computer networks.

1. Week, J.; Ivanova, P.; Week, S.; McLeod, A. (2009) *A Firewall Data Log Analysis of Unauthorized and Suspicious Traffic.* University of Nevada.

In this research paper, the authors describe how to accurately analyze the firewall system data logs and determine potential threats to a computer network system. The researchers, in addition, convey the safety risks, number of attacks, and where the attacks originated; to alert the overall effectiveness of the firewall system. The firewall logs contain the information detailing the source and destination points, Action items, and assess of the number of data attacks as well as a description of the attack. The researchers extensively describe the variables of the data set, eloquently present the research data, and concisely/effective convey their intended message to the audience.