Classification of Firewall Log Files with Multiclass Support Vector Machine

Fatih Ertam
Department of Digital Forensics Engineering,
Technology Faculty, Firat University,
Elazig, Turkey
fatih.ertam@firat.edu.tr

Mustafa Kaya
Department of Digital Forensics Engineering,
Technology Faculty, Firat University,
Elazig, Turkey
mkaya@firat.edu.tr

***Abstract*— It is very important to analyze the logs on the Firewall devices and control the internet traffic according to these analysis results. In this study, some logs obtained with the Firewall Device used at Firat University are classified using multiclass support vector machine (SVM) classifier. Linear, polynomial, sigmoid and Radial Basis Function (RBF) functions are used as the activation function for SVM classification. In order to measure the performance of the classifier, the comparison was made by finding the measurement values of sensitivity, recall and their harmonic mean F_1 Score. In this study, 65532 instances have been examined using 11 features. The feature that characterizes any personal data in the selected data has not been used. The Action attribute is selected as the class from these attributes. The "allow", "deny", "drop" and "reset-both" parameters have been implemented for the Action class. Activation functions have been tried and the SVM responses have been evaluated so as to obtain the maximum recall and precision values in the SVM classifier. It was tried to obtain the best activation function for F_1 score value. Receiver Operating Characteristic (ROC) curves were also created for each of the classes. At the end of the study, the activation functions from which the desired SVM responses are obtained are given by comparison.**

Keywords*— Classification, network forensics, log analysis, firewall, network security

I. INTRODUCTION

Positioning according to the activities that the end user has done in a corporate network is very important for managing the network used. According to the size of the network, the amount of data produced by the users in the network can be very large. Firewall devices on a network can either allow or prevent traffic according to the policy used by examining the generated data. The configuration of firewalls is vital for communication networks to work properly and secure. Correct configuration of these systems make a significant contribution to the active and effective use of companies' communication technologies and other existing network devices [1]. Firewalls act as control gates for computer networks. The system administrator sets up firewalls for each organization's needs [2]. Since firewalls are one of the important components of the network, there should be no conflict in the security policies to be selected and should not cause security vulnerability [3]. The designation of rules for firewalls has a critical pressure in terms of no conflicts [4]. Machine learning approaches are often used to discover and reveal relationships in data sets [5]. Analyzes of security firewall records using machine learning approaches are available in the literature. One of these is Golnabi et al. they aimed to reduce active firewall rules by examining approximately 33,000 log entries [6]. In this study, Golnabi et al. they have been working with about 2 times of both the data and feature numbers that they used in their work. Breier and Branišová have attempted to propose a method for anomaly detection in diary files based on data mining techniques to create dynamic rules [7]. In another study using machine learning approaches, Pietraszek and Tanner succeeded in reducing false positive alerts [8].

There are also many studies in the literature related to SVM, which is one of the approaches to machine learning [9], [10]. Although the SVM is used to solve the problem of separating two classes, it is also used to solve problems of two or more classes [14]. In this study, we used the SVM approach in the classification process for a total of 4 classes. The general approach used to do this is to classify multiple classes by trying to group them between binary classes. For the SVM approach, classification can be performed using various activation functions. In order to find the best result for this study, activation functions which are frequently used in the literature were classified by linear, polynomial, RBF and sigmoid activation functions separately. The precision, recall and F measure performance metrics are used to measure the performance of the classifier. Precision; we can calculate how much of the information brought about as a result of the classification is related to the desired information.

Precision is defined as the number of true positives (TP) over the number of true positives plus the number of false positives (FP). Equation 1 shows the precision formula.

$$Precision(P) = \frac{TP}{TP + FP} \tag{1}$$

Recall deals with how much information is retrieved. Recall is the number of true positives over the number of true negatives [11], [12]. Equation 2 shows recall formula.

$$Recall(R) = \frac{TP}{TP + FN} \tag{2}$$

F_1 Score is a harmonic average of the precision and recall values [13]. Also called F-Score or F Measure. Recall and

978-1-5386-3449-3/18/$31.00 ©2018 IEEE

Precision approach the best value of 1 with high values and approach 0 at worst. Equation 3 shows the F_1 Score formula.

$$F_1 Score = 2 \cdot \frac{P \cdot R}{P + R} \quad (3)$$

The ROC curves are used to show the graphical beginnings of the classifiers. ROC curves are calculated by taking the values of TP rate on the y axis and FP rate on the x axis. Approaching the TP rate value to 1 in the graph shows that the classifier is successful. ROC curves belonging to each class have been created separately.

In order to classify the firewall log data, 11 of the attributes in the data set were selected. When selecting data, it is important to select attributes that have more numerical values. The action attribute, which is one of the non-numeric attributes, has been accepted as a class. Table 1 shows the descriptions of the selected attributes.

TABLE I. FEATURES AND DESCRIPTION

Feature	Description
Source Port	Client Source Port
Destination Port	Client Destination Port
NAT Source Port	Network Address Translation Source Port
NAT Destination Port	Network Address Translation Destination Port
Elapsed Time (sec)	Elapsed Time for flow
Bytes	Total Bytes
Bytes Sent	Bytes Sent
Bytes Received	Bytes Received
Packets	Total Packets
pkts_sent	Packets Sent
pkts_received	Packets Received
Action	Class (allow, deny, drop, reset-both)

There are 4 classes in the action attribute used as a class. Descriptions of these classes are shown in Table 2.

TABLE II. SECURITY POLICY ACTIONS

Action	Description
Allow	Allows the internet traffic.
Deny	Blocks traffic and enforces the default Deny Action defined for the application that is being denied.
Drop	Silently drops the traffic; for an application, it overrides the default deny action. A TCP reset is not sent to the host/application.
Reset-Both	Sends a TCP reset to both the client-side and server-side devices.

II. PROPOSED METHOD

The block diagram of the proposed method is shown in Fig.1. Firstly, log records are received via the firewall. The Log records used were taken from the Palo Alto 5020 Firewall device used at Firat University. The receiving log record consists of 65532 records and is obtained as a recording result of approximately 30 seconds. In the receiving log, the attributes are taken with importance to port, byte, packet and time information. The class has the action attribute with "allow", "deny", "drop" and "reset-both" values. SVM was selected as a classifier and linear, polynomial, RBF and sigmoid were selected as activation functions. The precision, recall, F1 score and ROC curves were used to compare the performances.

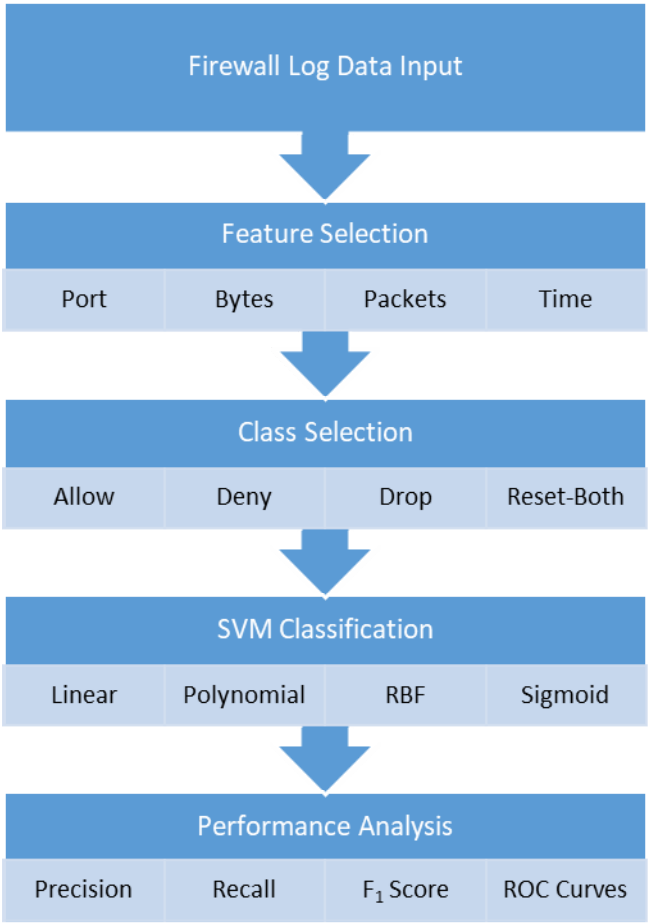

Fig. 1. Proposed Method

III. EXPERIMENTAL STUDIES

The obtained precision, recall and F1 Score values of the classifiers are shown in Table 3. Table 3 shows that the recall value is the best Sigmoid, and the RBF activation function, which is obtained with the worst polynomial activation function, is also successful. Precision value is best obtained by the classifier and the linear activation function is selected by the classifier and the other classifiers are found to be about the same. In the F1 score value calculated as the harmonic mean of the Precision and recall values, it is seen that the best classifier

is obtained with the classifier in which the RBF activation function is selected and the classifiers in which the linear and sigmoid activation functions are selected have good results. For the F1 score, it is seen that the worst result is obtained with the classifier selected as the polynomial activation function as in the precision and recall values.

TABLE III. EVALUATION RESULTS

Method	F_1 Score	Precision	Recall
SVM Linear	75.4	67.5	85.3
SVM Polynomial	53.6	61.8	47.4
SVM RBF	76.4	63.0	97.1
SVM Sigmoid	74.8	60.3	98.5

The graphs of the obtained ROC curves are shown in FIGS. 2 and 5.

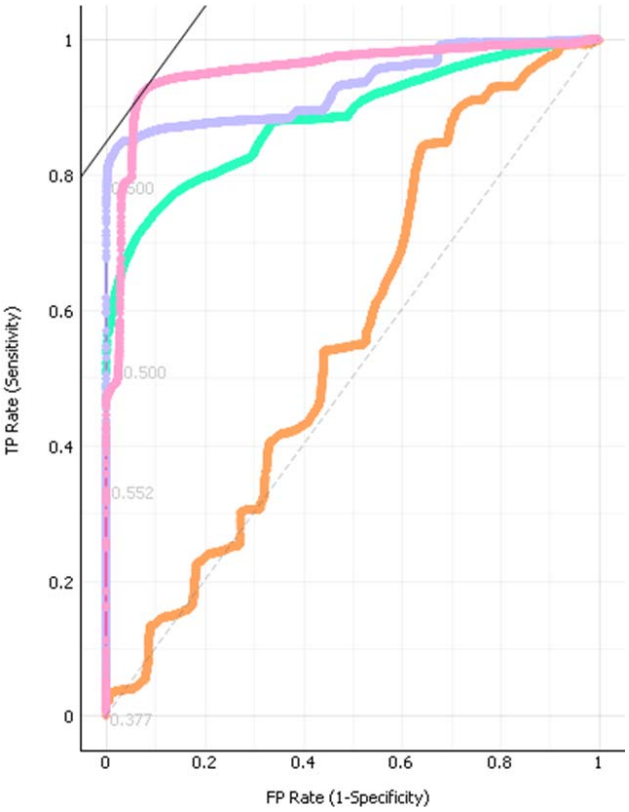

Fig. 2. ROC Curve (Allow Class)

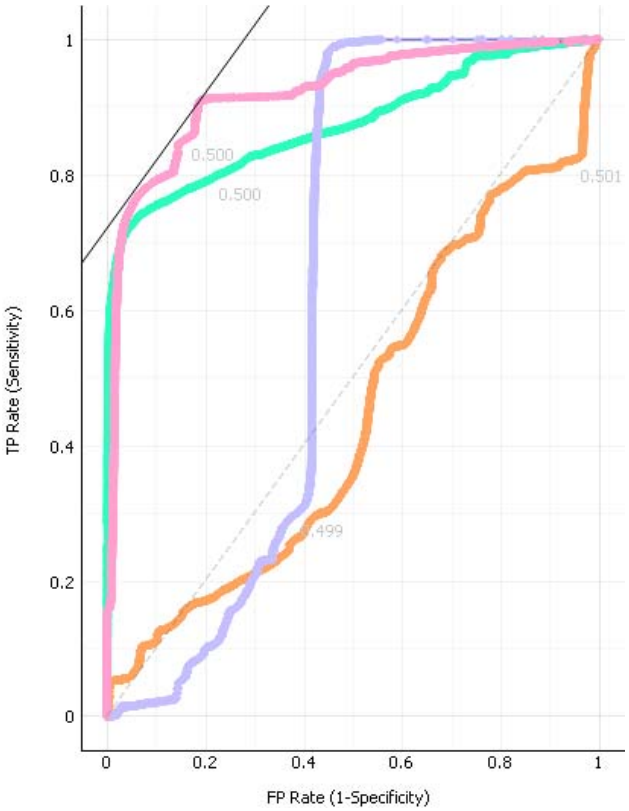

Fig. 3. ROC Curve (Deny Class)

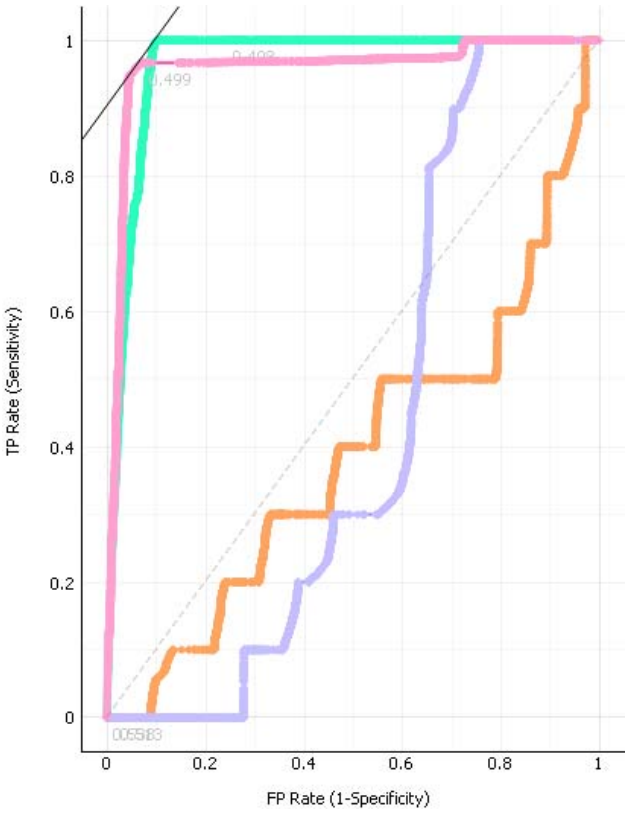

Fig. 4. ROC Curve (Drop Class)

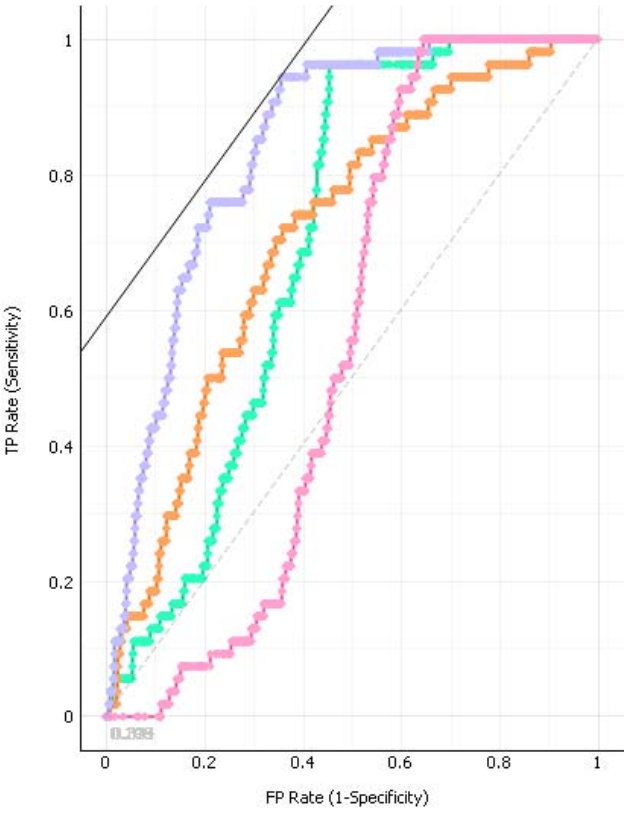

Fig. 5. ROC Curve (Reset-Both Class)

IV. Conclusions

It is very important to analyze the log records on the Firewall devices and control the internet traffic according to the results of these analyzes. In this study, a part of log records obtained through Firewall Device used at Firat University was classified by using SVM classifier. Linear, polynomial, sigmoid and RBF activation functions were used for classification with SVM. In order to measure the performance of the classifier, the precision, recall and F measure values were compared. The study was conducted using 65532 instances and 11 features. The action, which is one of the attributes used, is selected as the class. The values of this class are "allow", "deny", "drop" and "reset-both". It was observed that the highest recall value was obtained in the SVM classifier with 98.5% of the Sigmoid activation function selected and the highest precision value was obtained in the SVM classifier with the linear activation function of 67.5%. As the F_1 score value, it was observed that the best result was achieved with the classifier in which RBF activation was used with 76.4%.

Both the precision and the recall values have been observed to be at a low level in the classification made by selecting the polynomial activation function. When the mean values are taken into consideration, it is seen that the best classifier is the classifier made with the RBF activation function. Receiver Operating Characteristic (ROC) curves were also created for each of the classes. Thus, true positive and false positive rates are compared graphically.

In future work, it is planned to prepare the design of an intelligent system that can decide on its own to handle a lot more data with information to be extracted from the firewall logs. This work has created an infrastructure for the ultimate goal to be achieved. With the high accuracy of the prepared system, it is seen that successful results can be obtained in future studies.

References

[1] E. Ucar and E. Ozhan, "The Analysis of Firewall Policy Through Machine Learning and Data Mining," Wirel. Pers. Commun., vol. 96, no. 2, pp. 2891–2909, Sep. 2017.

[2] B. Khan, M. Mahmud, M. K. Khan, and K. S. Alghathbar, "Security analysis of firewall rule sets in computer networks," in Proceedings - 4th International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2010, 2010, pp. 51–56.

[3] R. Hunt, "Internet/Intranet firewall security - Policy, architecture and transaction services," Comput. Commun., vol. 21, no. 13, pp. 1107–1123, 1998.

[4] M. G. Gouda and A. X. Liu, "Structured firewall design," Comput. Networks, vol. 51, no. 4, pp. 1106–1120, 2007.

[5] E. Alpaydın, Introduction to machine learning, vol. 1107. 2014.

[6] K. Golnabi, R. K. Min, L. Khan, and E. Al-Shaer, "Analysis of Firewall Policy Rules Using Data Mining Techniques," 2006 IEEE/IFIP Netw. Oper. Manag. Symp. NOMS 2006, pp. 305–315, 2006.

[7] J. Breier and J. Branišová, "A Dynamic Rule Creation Based Anomaly Detection Method for Identifying Security Breaches in Log Records," Wirel. Pers. Commun., vol. 94, no. 3, pp. 497–511, 2017.

[8] T. Pietraszek and A. Tanner, "Data mining and machine learning—Towards reducing false positives in intrusion detection," Inf. Secur. Tech. Rep., vol. 10, no. 3, pp. 169–183, 2005.

[9] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," Mech. Syst. Signal Process., vol. 21, no. 6, pp. 2560–2574, 2007.

[10] C. Cortes and V. Vapnik, "Support vector machine," Mach. Learn., pp. 1303–1308, 1995.

[11] K. Ting, "Precision and Recall," in Encyclopedia of Machine Learning, 2011, p. 1031.

[12] P. Flach and M. Kull, "Precision-Recall-Gain Curves: PR Analysis Done Right," Adv. Neural Inf. Process. Syst. 28, vol. 1, pp. 838–846, 2015.

[13] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, vol. 8725 LNAI, no. PART 2, pp. 225–239.

[14] Z. Wang and X. Xue, "Multi-Class Support Vector Machine," in Support Vector Machines Applications, 2014, pp. 23–48.