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

The given reference paper is about analyzing the firewall log files using machine learning techniques, specifically a multiclass support vector machine (SVM) algorithm together with activation functions, to analyze the log records on the Firewall devices and control the internet traffic**. The input data** is the log records received via the firewall consisting of 65532 instances. **The proposed methodology** consists of

1) **Feature Selection**: select 11 important features related to port, bytes, packets, and time information including Client Source Port, Client Destination Port, Network Address Translation Source Port, Network Address Translation Destination Port, Elapsed Time for flow, Total Bytes, Bytes Sent, Bytes Received, Total Packets, Packages Sent and Packets Received

2) **Class Selection**: using Action attribute consists of 4 classes (allow, deny, drop and reset-Both) as class

| 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. |

3) **Model Training**: using SVM classification algorithm and different activation functions including linear, polynomial, RBF, and sigmoid functions to train the model

4) **Model Evaluation**: measure the performance metrics including precision, recall, F-1 score and ROC curves to compare the performance of the model related with different feature selection and activation functions.

**The study result** suggests that the highest recall value (98.5%) is obtained from using SVM classifier with Sigmoid function. The highest precision value (67.5%) is from using SVM classifier with linear function. And the highest F-1 score (76.4%) is from using SVM classifier with RBF function.

| 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 |

However, using specific dataset and specific machine learning algorithm in this study can lead to **possible limitations** including the results may not be generalized for different dataset or log files, the SVM approach might not be the best approach for other classification problems and the classification performance can be depended on other unselected features. The authors also planned to prepare a system that can handle more data and information extracted from the firewall logs for future studies.