Using Quadratic Discriminant Analysis by Intrusion Detection Systems for Port Scan and Slowloris Attack Classification

Vinícius M. Deolindo[1], Bruno L. Dalmazo[2](✉), Marcus V. B. da Silva[3], Luiz R. B. de Oliveira[1], Allan de B. Silva[1], Lisandro Zambenedetti Granville[3], Luciano P. Gaspary[3], and Jéferson Campos Nobre[3]

[1] University of Vale do Rio dos Sinos, São Leopoldo, Brazil
{vinicius.deolindo,luizbertoldi,fallanbs}@unisinos.br
[2] Federal University of Rio Grande, Rio Grande, Brazil
dalmazo@furg.br
[3] Federal University of Rio Grande do Sul, Porto Alegre, Brazil
{mvbsilva,granville,paschoal,jcnobre}@inf.ufrgs.br

Abstract. Identify and classify attacks through Intrusion Detection Systems is one constant challenge for security professionals. Computer networks are one of the significant IT components that support classification operations. Machine Learning (ML) techniques can aid in this process by providing methods capable of making decisions based on previously known information. In light of this, literature shows that Quadratic Discriminant Analysis (QDA) is barely explored as a classification method for IDS. To fill this gap, this study aims to create a new classifier able to distinguish legitimate network traffic from an attack by adopting ML techniques and QDA algorithms for identifying Port Scan and DoS Slowloris attacks.

Keywords: Intrusion Detection System · Machine Learning · Denial of Service · Quadratic Discriminant Analysis

1 Introduction

Currently, computer networks provide a means of communication between the most diverse devices, providing services to users and businesses. Consequently, networks also become targets of attacks that seek to compromise their security. The rapid detection of a potential attack and the type of attack in question is essential so that the necessary measures can be chosen for each case.

Services such as Intrusion Detection System (IDS) provide to the computer networks a tool capable of identifying events that may endanger the security of the environment, compromising the integrity, availability, and confidentiality of resources and services [10]. An attack identification, coupled with defense strategies and incident response, can provide a basis for network perimeter protection.

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12951, pp. 188–200, 2021.
https://doi.org/10.1007/978-3-030-86970-0_14

However, the IDS must count on complementary detection mechanisms to assist with rapid detection.

Machine Learning is an area of knowledge that aims to develop techniques that allow computer programs to acquire information and knowledge in an automated way. This learning process is built on specific techniques that enable a system to make decisions based on previously evaluated experiences. Making systems able to precisely predicting occurrences is one of the challenges that Data Scientists face in Machine Learning [1,7]. For instance, defining a model capable of assertively predicting information depends on a number of factors and detailed preliminary analysis.

The adoption of Machine Learning techniques for recognizing attack patterns – which can be characterized as an anomaly detection approach – can help to determine the type of attack suffered, aiding decision-making for reaction measures [3]. Automated techniques are ideal for IDS as they allow you to monitor and correlate a large number of patterns, signatures, and anomalies information [8]. As a novelty, this work applies a classification algorithm, namely, Quadratic Discriminant Analysis (QDA) jointly with an IDS in order to perform anomaly detection. QDA is one of the standard approaches to data classification algorithms and performs well when applied to large amounts of data for factoring [15].

Traditional techniques applied by IDS based on signatures can present problems in identifying certain types of attacks. The main reason is an attack may suffer slight variations in its pattern of action. Thus changing its default features and hiding this new behaviour to the system. The combination between IDS, machine learning, along with the classification of network traffic through classifiers using the QDA algorithm, can support in the correct and quick decision making looking for anomalies in the network traffic [5]. To deal with these lacks, this study aims to create a new classifier able to distinguish legitimate network traffic from an attack by adopting ML techniques and QDA algorithms. In this context, the scope of this work is bounded to scenarios for *Port Scan* and *DoS Slowloris* attacks.

The remainder of the paper is organized as follows. Section 2 presents the theoretical basis of the methods and systems adopted in this study. Section 3 describes the steps of a machine learning project and challenges to overcome, whilst Sect. 4 presents the evaluation and discusses the results. Section 5 covers some of the most prominent related work. Section 6 concludes with some final remarks and prospective directions for future research.

2 Background

This section presents the theoretical framework for the technologies and methods used in this study, namely, Machine Learning (ML), Cross Industry Standard Process for Data Mining (CRISP-DM), Intrusion detection System (IDS), the explored attacks and Quadratic Discriminant Analysis (QDA).

2.1 Machine Learning

Machine Learning (ML) belongs to the artificial intelligence field and consists of a set of tools, techniques, and methods for extracting knowledge from a dataset. ML aims to build systems that can automatically learn by analyzing previous making decisions based on acquired information and knowledge. In other words, ML seeks to extract knowledge from one dataset so that this learning can be applied to other similar datasets [7].

2.2 Cross Industry Standard Process for Data Mining

Also referenced by the acronym CRISP-DM, it alludes to a methodology intended for data mining related projects. Developed through a consortium consisting of *DaimlerChryrler, SPSS and NCR*, the framework presents six steps that make up the lifecycle of a Machine Learning project [12].

The expected activities and deliverables in each of the steps of the CRISP-DM framework are as follows.

- *Business Understanding:* The first phase of the project focuses on understanding the goals and requirements as well as business expectations for the delivery. The challenge is to convert doubts and expectations into a clear definition of the problem so that only then can a plan for the project progress can be drawn up.
- *Data Understanding:* Understanding the collected data, assessing the initial data collection, preselecting information sources and identifying the quality of the records are examples of activities that should be performed at this stage of the project. By following this order, each of the information brings more familiarity and better knowledge regarding the whole process, facilitating to perform the first preliminary analyzes of the collected data.
- *Data Preparation:* The third stage of the project is designed for the formulation of the previously evaluated data, in this phase should be built the datasets that will be used in the future, it is also essential to establish the portion that will be used in the training phase and the portion intended to the classification model testing.
- *Modeling:* In this step, the models and techniques available for applying the data prepared previously are evaluated. This step is characterized by using different techniques to identify and assess the benefits and drawbacks of each algorithm.
- *Evaluation:* The fifth step of the project has two goals: to evaluate the models created in the previous stage and define the final model for implementation. Also, if necessary, it is possible to return to the Business Understanding step for data redefinition. This process may occur if the results of model evaluations are not satisfactory.
- *Deployment:* The last step of the cycle, the final model is created and implemented. Creating the data model is not necessarily the end of the project. Here, it is common presenting the project to the stakeholders and sponsors.

All knowledge built throughout the project should be documented and delivered along with the final results so that they can be available as learning in future projects.

2.3 Intrusion Detection System

Intrusion Detection Systems (IDSs) are systems that seek to identify potential threats and attacks that may be occurring on a device or a network. In general, an IDS may detect intrusions using two approaches: through anomalies or an attack signature identification. Anomaly detection seeks to identify standard deviations in the network traffic or device behaviour based on normal patterns to identify outliers. On the other hand, signature-based IDS uses known patterns (part of code or protocols behaviour, for instance) of various types of attacks to try to identify threats.

- Network-based Intrusion Detection System (NIDS): It operates at capturing network packets to protect a system from network-based threats. A NIDS reads all inbound packets and searches for any suspicious patterns.
- Host-based Intrusion Detection System (HIDS): It runs on individual hosts or devices on the network. A HIDS monitors the inbound and outbound packets from the device only and will alert the user or administrator if suspicious activity is detected.
- Hybrid Intrusion Detection Systems: It consists of using network-based and host-based systems to control and monitor the computational security of an environment.

2.4 Attacks

For this study, we consider two well-known attacks frequently triggered in networks. Each of the attacks has different characteristics and objectives, as described below:

- *Port Scan:* This attack performs a port scan, that is a method for discovering communication channels (ports) that can be employed in future attacks. The technique consists of investigating the communication ports of a given target and then evaluating which exploiting methods are the most appropriate. Although some authors do not consider Port Scan as an attack (since there is no intrusion), there is a consensus that it is a first stage in identifying vulnerabilities that could lead to future attacks [11].
- *DoS Slowloris:* Refers to a type of attack performed using the Slowloris tool, which targets HTTP servers. The method consists of requesting multiple connections to a given target server, in order to attempt exhausting the server's ability to respond to requests. The consecutive attacker connection requests seek for maintaining established connections over a long period. Usually, the IDS based on anomaly detection presents some difficulty to detect this attack due to the low volume traffic generated. Also, different from a Denial of Service, the connection is held as much possible to simulating a legitimate user behaviour [4]. Figure 1 illustrates the Slowloris attack based on the dataset [6].

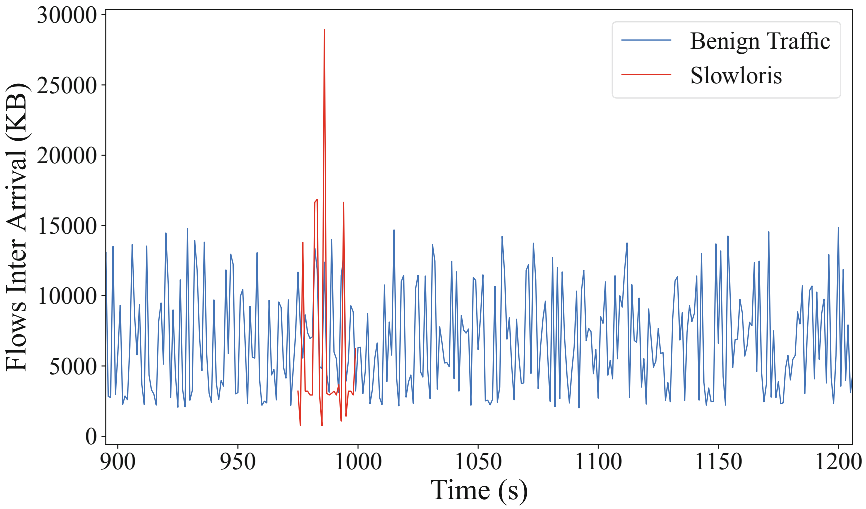

Fig. 1. Attack performed by DoD Slowloris

2.5 Quadratic Discriminant Analysis

Quadratic Discriminant Analysis (QDA) is one of the most used models for machine learning using the supervised learning technique. This method seeks to model the probability of each class based on a Gaussian Distribution, i.e., it is assumed that the measurements from each class are normally distributed. Then, after modeling the classes with the supervised method, the QDA uses a normal distribution to make predictions [15]. Although the Gaussian Model is simple, the QDA presents a limitation: it does not perform well from a small data sample [15].

3 Port Scan and DoS Slowloris Classification

This section presents an alternative for classifying Port Scan and DoS Slowloris attacks by collecting metrics generated by an IDS and machine learning. Intrusion detection plays a vital role in computer network defense processes by enabling the generation of various types of alerts for different malicious behaviours that may occur in the environment [13].

Sharafaldin *et al.* [13] created a complete dataset containing various types of attacks. Then enabling to extract the patterns of attacks adopted for the accomplishment of this work. Mapping of the most relevant information for each type of attack allows to establish patterns that made possible the use of machine learning algorithms for classification and consequent identification of Port San and DoS Slowloris attacks.

CRoss Industry Standard Process for Data Mining (CRISP-DM) was adopted in the context of this work. This framework provides standards for carrying out the activities of a machine learning project [9]. In this section, the data preparation and the proposed model for the classification of the attacks are presented. The environment for modelling this work was based on the RStudio, which is an open-source project that covers several components of R statistical.

3.1 Business Understanding

The business understanding phase can be summarized in the problem that the project seeks to solve. What challenges can be met and how to develop a solution that addresses the issue at hand. In this case, the main problem that this study seeks to solve is the identification of attack patterns drawn for Port Scan and DoS Slowloris jointly with the adoption of the QDA classifier. Therefore, it is necessary to evaluate which data sources are available and which of them may provide attack patterns to perform this study.

At this stage, several IDS datasets and their features were evaluated, so that it was possible to assess which sets would be useful for the study. In this sense, the Sharafaldin's paper [13] brings a complete dataset with valuable information about attacks collected by an IDS. Also, the attacks were performed in a controlled environment using recent techniques. Although using features as the source, destination, and duration of attacks, this dataset present a high quality of the classification labels of each attack, essential information that facilitates the future creation of a QDA model. In the next step was necessary to request the datasets for the Canadian Institute for Cybersecurity. This organization conducted the study around the datasets.

3.2 Data Understanding

Datasets were divided into days, as shown in Table 1, each day contains information from different attacks. As the purpose of this study is to evaluate Port Scan and DoS Slowloris attacks, we used the Friday files, the day in which these attacks were performed.

Several metrics in the dataset archives are available, but due to the diversity of parameters in each set, this study considers just the most relevant metrics, as detailed in [13]. Also, we evaluated the variety of classification labels of each attack contained in the files. Having a consistent diversity of information is a crucial point to create the model using QDA.

Table 1. File organization [13]

Days	Labels
Monday	Benign
Tuesday	BForce, SFTP and SSH
Wednesday	DoS and Hearbleed Attacks Slowloris, Slowhttptest, Hulk and GoldenEye
Thursday	Web and Infiltration Attacks Web BForce, XSS and Sql Injection Infiltration Dropbox Download and Cool disk
Friday	DDoS LOIT, Slowloris ARES, Port Scans (sS, sT, sF, sX, sN, sP, sV, sU, sO, sA, sW, sR, sL and B)

Table 2. Organization of the labels in the dataset

Port Scan	
Var1	*Freq*
BENIGN	127537
PortScan	158930

DoS Slowloris	
Var1	*Freq*
BENIGN	440031
DoS GoldenEye	10293
DoS Hulk	231073
DoS Slowhttptest	5499
DoS slowloris	5796
Heartbleed	11

The data understanding allows to realize some metrics presented in the datasets did not present the type set on. This gap may impact the model construction, depending on the amount of data used. The files presented 85 columns with different features. According to Table 2, we observe that there are 127,537 data classified as *BENIGN* and 158,930 *Port Scan*, in other words, for *Port Scan* more than half of the data is classified as attack. This amount indicates that there are data enough for the model creation. Due to the similarity among the file structure, the parsing procedures for the data from both attack sets could be the same. It is worth noticing that the amounts of *labels* for each attack are the result of an initial assessment, because during the *Data Preparation* stage, the data changes according to the preparation criteria.

3.3 Data Preparation

Data preparation is a key point for creating the model; if it is not done correctly, the results and performance suffer an impact in a negative way. Usually, in case of having a lot of information in the initial dataset, you can build an "auxiliar" dataset. This technique of creating or transforming resources into a new set can assist in more accurate classifiers in later steps [2].

The quality of the assessed data is critical to the success of the project. In the data preparation, all information that will not be used in the future are removed, as most learning algorithms use knowledge extracted from the data without the use of external sources. In order to set up the datasets of each type of attack, we used as guide the study performed by [13]. Table 3 presents the most relevant metrics for each type of attack.

Table 3. Relevant metrics for each type of attack [13]

Label	Feature	Weigh
Port Scan	Init Win F.Bytes	0.0083
	B.Packets/S	0.0032
	PSH Flag Count	0.0009
DoS Slowloris	Flow Duration	0.0431
	F.IAT Min	0.0378
	B.IAT Mean	0.0300
	F.IAT Mean	0.0265

At this stage, we also performed the evaluation of the information whose origin was the internal network itself. In other words, during the creation of the datasets, we introduce all network traffic analyzed by the IDS without removing any information from the internal hosts. This approach could deform the information for analysis, as an attack usually originates from the external perimeter of the network. So, we remove all records that came from the internal network (LAN). However, removing the data affected the amount of benign data of the sets. The updated data are shown in Sect. 4.

After completing the data preparation step, only the columns used to build the models remain. Also, the overall data was reduced at the end of this step, disregarding all traffic from the internal network. In this sense, it was necessary to establish the data types for each column, which helps in the performance of model creation. The data generated during this process is subject to an evaluation to attest to its quality. Depending on the quality of the evaluated data and the results obtained, it could be necessary to return to the previous steps to ensure the data present the required level of quality. This step is essential to avoid future problems in the evaluation of the results.

3.4 Modeling

As previously described, this study seeks to create models that allow the classification and identification of *Port Scan* and *DoS Slowloris* attacks using QDA. In the Modeling phase occurs the tests for the creation of the templates. Firstly, in order to be able to create the model in QDA, it was necessary to revisit some steps of data preparation. In this context, 60% of the data was reserved for the training of each model, the remaining 40% used for evaluating the results.

To create the datasets, we select the records randomly according to the established percentages, not collecting sequence data, which could affect the model performance. For both models (training and testing), the same percentages were adopted, regardless of the label distribution of each dataset. The definition of the percentages for training and evaluation was defined based on similar studies in the literature. The size distribution of each dataset is shown in Table 4.

Table 4. Size distribution of each dataset

	Port Scan	DoS Slowloris
Training	106531	44397
Testing	71022	*29599*

The classification models for each type of attack start to work right after creating the training and testing datasets. The process consists of passing the parameters that the QDA uses to construct each model and which variable should be learned or predicted. In this case, the goal is that the models can predict the label according to each type of attack, distinguishing the regular traffic from the attack. Next, we present the performance evaluation of each case.

4 Evaluation

To evaluate our proposal, we used a dataset produced by the Canadian Institute for Cybersecurity. More specifically, the dataset provides a range of information regarding the main types of attacks detected by IDS. The dataset contains more than 80 metrics available for network traffic evaluation, as well as a description of the configurations applied in the environments and components used in this work, thus allowing the complete identification of the topology used for the instantiation of the scenarios.

During the evaluation of the proposal, several types of attacks presented in the datasets were evaluated. However, not all attacks contained enough information available for the investigation and creation of the classification model. The following metrics were individually evaluated for each of the two models created: Receiver Operating Characteristics (ROC), Accuracy, and Recall. All these metrics are commonly used for evaluating classification algorithms.

- Accuracy: It is one of the most common metrics for evaluating machine learning accuracy. It refers to a statistical measure of how well a binary classification test correctly identifies or excludes a condition. Accuracy is calculated using the total hits made by the classifier (true positive and true negative) on the total number of objects that were classified, as illustrated in Eq. 1.

$$Accuracy = \frac{TruePositive + TrueNegative}{TotalSamples} \tag{1}$$

When the variety of data for the model creation is not large, the Accuracy may not be the best metric. It is crucial to evaluate the available class balance so that the classification standards can be established. Accuracy refers to the number of correct ratings, but it is imperative to evaluate the percentage of correctly rated true positive and true negative. Table 5 presents the results for Accuracy, Sensitivity, and Specificity.

Table 5. Statistics of the classification models

	Accuracy	Sensitivity	Specificity
Port Scan	0.9996	0.9996	0.9996
DoS Slowloris	0.9832	0.9934	0.8616

The values shown in Table 5 range from 0 to 1. Where 1 means the best result possible. *Accuracy* is the overall accuracy, *i.e.*, the percentage that the model is correct. *Sensitivity* is the classifier's ability to identify benign traffic correctly. *Specificity* allows to measure the performance of the classification the attack; it means how often the algorithm hits. Observing results from the evaluation, it is worth noting that all metrics present values higher than 85%, highlighting the scenario for the *Port Scan*, with a hit rate of more than 99%.

- *Receiver Operating Characteristics* (ROC): It refers to a graphical method that aims to evaluate metrics and select them for systems – originally designed for radar signal detection. Currently, it is used for evaluating a wide range of activities, from psychology, medicine, economics, weather, and machine learning. Figure 2 presents the ROC graphs of the models created for *Port Scan* and *DoS Slowloris* respectively.

In Fig. 2, we observe there is no curve in the ROC for *Port Scan*, there is a straight line from (0.0; 0.0) to (1.0; 1.0), unlike the ROC for *DoS Slowloris*. Analyzing the different results, we can notice that for both models created, the positive class is the "Benign", but the amount of "Benign" data in both datasets for each attack is what differentiates the results. Table 6 presents the amount of data for each attack type according to the distribution of the datasets.

Table 6. Characterization of the data according to the datasets

	Benign	Attack
Port Scan	18.623	158.930
DoS Slowloris	68.200	5.796

From the analysis of different distributions for each dataset presented in Table 6, we observe that for *Port Scan* attack the number of records classified as Attack is far superior in comparison to Benign, which means that the model labeled with the Negative class has more data than the Positive

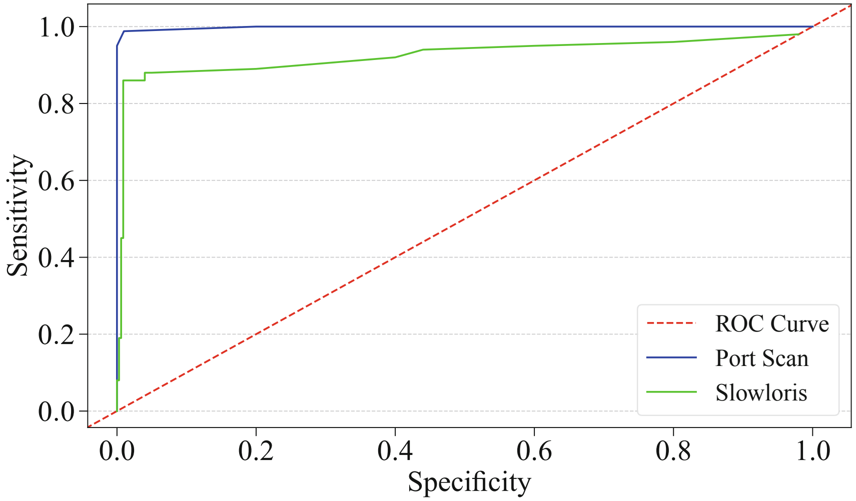

Fig. 2. ROC graph from the models

class. From the point of view of the *Dos Slowloris* dataset, this scenario is reversed. In which case, there is much more data classified as "Benign", which explains the difference in the curves of each of the graphs in Fig. 2. About 89% of data for *Port Scan* is classified as an attack against about 7% for *DoS Slowloris*.

- *Recall:* This metric is also referenced by *True Positive Rate*, which briefly answers the following question: When should the algorithm predict the Positive Class: how often does it hit? Other metrics such as Sensitivity and Specificity can evidence the performance of a classifier [16]. This relationship is presented in Fig. 2.

  The *Recall* results are shown in Table 5 by the column Sensitivity. The lower percentage for Port Scan attack compared to DoS Slowloris can be explained by the low number of "Benign" records. This characteristic confuses the model once there are more attacks than normal network traffic. However, there was a high hit rate for *Recall* in identifying attacks. Recall and accuracy are related metrics, and in some cases, changing parameters to increase one may imbalance the other. The best case is looking for a trade-off where both achieve a suitable result.

5 Related Work

There are several studies related to the use of Machine Learning techniques for identifying attack patterns. On the other hand, usually, generalist methods for identifying and classifying attacks are proposed. Thus, ignoring specific behaviours of each type of attack, which can often make the difference in the result. Sinclair *et al.* [14] present a method based on the Decision Tree algorithm to identify anomalies in network layers. However, the result does not rank the attack or malicious traffic according to the technique used, only suspicious traffic is identified.

Osareh and Shadga [10] conducted a study using a public dataset, KDD, to evaluate the performance of different classification methods. The dataset used for the study is from the year 1998, which makes it challenging to evaluate its efficiency faced with current attacks. Similar to Sinclair *et al.* [14] work, the same methods, and variables were used for traffic classification and identification of attack types. The study explored an algorithm that, on average, could meet the general classification rate. However, this approach does not achieve a classification rate higher than 50% when applied individually for each type of attack. This perspective reinforces the importance and relevance of each specific measurement metric for each type of attack.

Sharafaldin *et al.* [13] developed a complete dataset containing several types of attacks identified by an IDS. This study presents the construction systems, topology overview, and methodology used to construct the available datasets, as well as containing several metrics distributed in more than 80 columns. The variety, amount of data available, as well as all the information used for the construction and execution of the study, allowed us to conduct our research. The diversity of data and the previous identification of the most relevant metrics for a given attack enabled the creation of efficient classification models.

6 Conclusion

Face the relevance of the attack scenarios considered in this study, classification models for assisting in the identification of risks of the network perimeters were proposed. It is worth observing the algorithm proposed is a complement to identification and classification mechanisms already existing in the IDS, and it may support in doubt scenarios or even "double-check" for the definition of defense strategies.

In light of this, the present work proposed a new approach for using QDA for identifying and classifying *Port Scan*, and *DoS Slowloris* attacks, allowing distinguish "benign" traffic from "malicious" traffic. From a test environment and based on detailed datasets, a QDA classification model obtained results with a high percentage of assertiveness in the classification of attacks. By observing the results, it is clear the influence of the correct metrics on the efficiency of each of the models created.

Despite the encouraging results depicted, as future work, we intend to evaluate other types of attacks not considered in this study. Also, to propose a methodology to compare the performance with different classifiers for several types of attacks. It is worth noticing the features used to construct the models were made available in the datasets available in [6]. Finally, an approach for on-the-fly collecting information must be designed in order to evaluate and perform a fair comparison among other classifiers.

References

1. Aksu, D., Aydin, M.A.: Detecting port scan attempts with comparative analysis of deep learning and support vector machine algorithms. In: 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), pp. 77–80. IEEE (2018)
2. Aziz, A.S.A., Sanaa, E., Hassanien, A.E.: Comparison of classification techniques applied for network intrusion detection and classification. J. Appl. Log. **24**, 109–118 (2017)
3. Boutaba, R., et al.: A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. J. Internet Serv. Appl. **9**(1), 16 (2018)
4. Choi, J., Park, J., Heo, S., Park, N., Kim, H.: Slowloris DoS Countermeasure over WebSocket. In: Choi, D., Guilley, S. (eds.) WISA 2016. LNCS, vol. 10144, pp. 42–53. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56549-1_4
5. Dalmazo, B.L., Vilela, J.P., Simoes, P., Curado, M.: Expedite feature extraction for enhanced cloud anomaly detection. In: NOMS 2016–2016 IEEE/IFIP Network Operations and Management Symposium. pp. 1215–1220, April 2016. https://doi.org/10.1109/NOMS.2016.7502990
6. Dalmazo, B.L., Deolindo, V.M., Nobre, J.C.: Public dataset for evaluating Port Scan and Slowloris attacks (2019)
7. Harrington, P.: Machine Learning in Action. Manning Publications Co. (2012)
8. Liao, H.J., Lin, C.H.R., Lin, Y.C., Tung, K.Y.: Intrusion detection system: a comprehensive review. J. Netw. Comput. Appl. **36**(1), 16–24 (2013)
9. Nadali, A., Kakhky, E.N., Nosratabadi, H.E.: Evaluating the success level of data mining projects based on CRISP-DM methodology by a fuzzy expert system. In: 2011 3rd International Conference on Electronics Computer Technology (ICECT), vol. 6, pp. 161–165. IEEE (2011)
10. Osareh, A., Shadgar, B.: Intrusion detection in computer networks based on machine learning algorithms. Int. J. Comput. Sci. Netw. Secur. **8**(11), 15–23 (2008)
11. Sangkatsanee, P., Wattanapongsakorn, N., Charnsripinyo, C.: Practical real-time intrusion detection using machine learning approaches. Comput. Commun. **34**(18), 2227–2235 (2011). https://doi.org/10.1016/j.comcom.2011.07.001
12. Shafique, U., Qaiser, H.: A comparative study of data mining process models (KDD, CRISP-DM AND SEMMA). Int. J. Innov. Sci. Res. **12**(1), 217–222 (2014)
13. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of Fourth International Conference on Information Systems Security and Privacy, ICISSP (2018)
14. Sinclair, C., Pierce, L., Matzner, S.: An application of machine learning to network intrusion detection. In: 15th Annual Computer Security Applications Conference, (ACSAC 1999) Proceedings, pp. 371–377. IEEE (1999)
15. Srivastava, S., Gupta, M.R., Frigyik, B.A.: Bayesian quadratic discriminant analysis. J. Mach. Learn. Res. **8**, 1277–1305 (2007)
16. Su, W., Yuan, Y., Zhu, M.: A relationship between the average precision and the area under the roc curve. In: Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR 2015, pp. 349–352. Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2808194.2809481