Damascus University Journal for engineering sciences
Vo 42 number (2) -2026: 58-65

مجلة جامعة دمشق للعلوم الهندسية
المجلد42 العدد (2) – 2026: 58-65

Evaluating the efficiency of Supervised Learning Classification Models in Solving Denial of Service Attacks Problem

Samer mumtaz Sulaiman[*1] Wassim Alsamara[2] Raafah Khazem[3]

*[1]. PhD Student - Computer and networks engineering - Faculty Mechanical and Electrical Engineering- Damascus University
samer.sulaiman@damascusuniversity.edu.sy

[2]. Assistant Professor, Dr, Eng, Damascus University ،Networks and centers of data transformation and distribution .
wasim.alsamara@damascusuniversity.edu.sy

[3]. Lecturer, Dr, Eng, Damascus University ،Remote measurement and driving.
Raafah.Khazem@damascusuniversity.edu.sy

Abstract:

Machine Learning has been widely used in several disciplines nowadays. The rabid development of Denial of service attacks outdated traditional methods for network security regarding the DoS attacks. Several researchers suggest that Machine Learning is a promising technique to fight Denial of Service attacks and they focus on using supervised learning methods to prove their theory. Using Classification to detect Denial of service attacks is expected to succeed especially since it can perform well during the training and testing phases but it wasn't tested in real-life scenarios. In this paper we built seven different classifiers based on the CSE-CIC-IDS2018 and NSL-KDD datasets and tested them in the OMNET++ simulation environment since it is not possible to perform such tests on a real network. We found a large gap between the theoretical accuracy and the resulting one within the simulation which can be caused due to the Covariate shift problem. Tradition Classification might not be suitable to solve this problem. Other models were suggested to be tested in future studies.

Keywords: Denial of Service Attacks, Supervised Learning, Classification, CIC2018, NSL-KDD

Received: 18/7/2023
Accepted: 19/10/2023

http://journal.damascusuniversity.edu.sy/index.php/index/index
ISSN (online): 2789-6854

تقييم فعالية التصنيف باستخدام نماذج التعلم ذي الإشراف في حل مشكلة هجمات الحرمان من الخدمة

سامر ممتاز سليمان*[1] وسيم السمارة [2] رأفة خازم[3]

[1]**. طالب دكتوراه – اختصاص هندسة الحواسيب وشبكاتها – كلية الهندسة الميكانيكية والكهربائية – جامعة دمشق.

samer.sulaiman@damascusuniversity.edu.sy

[2]. أستاذ مساعد، أستاذ، دكتورن مهندس، جامعة دمشق، شبكات ومراكز التحويل وتوزيع المعطيات.

wasim.alsamara@damascusuniversity.edu.sy

[3]. مدرس، جامعة دمشق، القيادة والقياس عن بعد.

Raafah.Khazem@damascusuniversity.edu.sy

الملخص:

شاع استخدام تعلم الآلة في العديد من المجالات في أيامنا هذه. وقد أدى التطور السريع لهجمات الحرمان من الخدمة إلى إبطال وسائل الحماية التقليدية للشبكات فيما يتعلق بهجمات الحرمان من الخدمة. يقترح العديد من الباحثين أنّ تعلّم الآلة تقنية واعدة في مجال مكافحة هجمات الحرمان من الخدمة وتحديدًا بتركيزهم على استخدام طرائق التعلم ذي الإشراف لإثبات نظريتهم. من المتوقع أن ينجح استخدام التصنيف في كشف هجمات الحرمان من الخدمة وخصوصًا أنه كانت نتائجه جيدة في مرحلة التدريب والاختبار إلا أنها لم تكن كذلك في التجارب المنجزة في الحياة العملية. قمنا في هذا البحث ببناء سبعة نماذج مختلفة للتصنيف بالاعتماد على مجموعتي المعطيات CSE-CIC-IDS-2018 و-NSL KDD واختبارها ضمن بيئة المحاكاة OMNET++ نظرًا لعدم إمكانية تنفيذ هذه الاختبارات ضمن شبكة حقيقية. وجدنا فجوة كبيرة بين الدقة النظرية وتلك الناتجة في عملية المحاكاة. يعزى إلى مشكلة الانزياح التبايني. من الممكن أنّ التصنيف التقليدي غير ملائم لحل مثل هذه المشكلة حيث تم اقتراح نماذج أخرى ليتم اختبارها في الدراسات المستقبلية.

الكلمات المفتاحية: هجمات الحرمان من الخدمة، التعلم ذي الإشراف، التصنيف، CIC2018، NSL-KDD

تاريخ الايداع :2023/7/18
تاريخ القبول : 2023/10/19

Introduction:

Denial of service attacks remain to be a threat to the whole Internet worldwide. During the past years these attacks were improved dramatically and several countries were affected by them, whether this effect was a minor service interruption or major network failure. (Graewe, 2023) States that bandwidth attacks increased up to 600% in 2019 than its previous value in 2016. Using traditional methods solely such as rule-based filtering to mitigate DoS attacks is not an advisable solution since the implications for such usage are high at cost. The rabid development of Machine learning led to its involvement in nearly all technical disciplines. Machine learning algorithms are roughly divided into three main categories which are Supervised Learning, Unsupervised learning and Reinforcement learning (Kwekha-Rashid, 2023). Numerous researches suggest models using datasets to improve DoS detection and mitigation (Aljuhani, 2021) (Dong Li, 2018) (Gondi Lakshmeeswari, 2020) (Kimmi Kumari, 2022) (Vinícius De Miranda Rios, 2022). A serious issue arises when these models are being evaluated since evaluating them by using the same dataset or even a similar dataset gives high accuracy rates which may be practically overrated when addressing the DoS issue; otherwise the world would have eliminated these threats several years ago, yet major denial of service attacks are bringing down the whole internet in some countries. In this paper we evaluate the efficiency of several supervised learning classification models in detecting DoS attacks, and we study their accuracy by simulating a network using OMNET++.

1. Related Work

Using Machine learning to mitigate denial of service attacks is a common and fertile research area. In (Gondi Lakshmeeswari, 2020) Multiple Linear Regression is used to build a model that detects Distributed Denial of Service attacks. Models are built using both datasets and log files that include traffic packets, authors in this study used the CIC2017 dataset and the highest reported accuracy was 97.86%.The authors in (Kimmi Kumari, 2022) used a Logistic Regression classifier and a Naïve Bayes classifier built using CAIDA 2007 in order to detect denial of service attacks. The achieved accuracy was between 99 and 100% for the logistic regression classifier and between 98 and 99% for the Naïve Bayes one. In (Aljuhani, 2021) an analytical study of the usage of machine learning in fighting DoS attacks was introduced. Authors classify various algorithms, environments and mitigation techniques that can be used to stop these attacks. In (Vinícius De Miranda Rios, 2022) the authors present a survey of the Low-rate denial of service attacks that are emerging threats which further complicate the detection of DoS attacks. The authors also provide the correct means which could be used to detect those attacks. These means are nothing but the features that could be fed to the ML models. In (s. Balaji, 2021) the authors introduce an analytical study of the usage of deep learning in DoS detection, their results show that these methods are highly anticipated to succeed in both detecting and predicting the attacks. In (Dong Li, 2018) the authors introduced an implementation of SVM algorithm to detect DDoS attack and used a multi-vector attack consisting of UDP flood, ICMP flood and SYN flood. The authors compared several algorithms such as KNN, Random Forest (RF), Naïve Bayes and SVM were used but SVM gave the best accuracy among other algorithms. Our main contribution is building multiple supervised learning models and comparing their achieved performance in the testing stage of building the model against their performance in simulated environment.

2. Materials and methods:

Our proposed practical procedure includes two steps. The first one is building machine learning models using seven different classifiers, and the second step is testing each model with a network simulator that accepts machine learning models. There are several datasets that include important data related to the DoS problem, but these datasets typically focus on several types of intrusions and might include additional features that are related to other types of intrusions. In this research, two datasets were used which are the CSE-CIC-IDS2018 dataset and NSL-KDD dataset. Both datasets were used based on the (80-20) % rule this means 80% of the data was used for training and the remaining 20% was used for testing purposes. Each dataset was preprocessed before initiating the training stage. This included the removal of any unnecessary fields which included the remov-

al of timestamps and other fields that are not directly related to the DoS problem, this issues arises in the NSL-KDD dataset since it contains data to detect multiple intrusions along with DoS data. The CIC-IDS-2018 on the other hand has several features that were not useful in the detection such as the IP address and FlowID. In this research, several classifiers including Adaboost, Gradient Boosting, Random Forest, XGB, Ridge, Stochastic Gradient Descent (SGD) and Support Vector Machines (SVM) were chosen to evaluate the usage of traditional supervised learning in detecting DoS, this choice was made based on their common usage in the academia along with their normally reported high accuracy.

Adaboost classifier is considered to be a meta-estimator that fits an initial classifier on a dataset and then fits several other classifiers on the same dataset while adjusting the weights of instances that were wrongly classified to help improve their classification (Schapire, 2013).

Gradient Boosting Classifier is a form of Ensemble learning in which each classifier aims to improve the one that precedes it. The strength point of this method is that the improvement is based on the residual errors of the classifiers and not on the whole data (Natekin, 2013).

Random Forest classifier generates multiple decision trees using random groups of a selected dataset and then aggregates the outcome of different trees to decide the output (Rigatti, 2017).

XGB classifier is based on gradient boosted decision trees algorithms and it is highly parallelizable. It results an overall good performance in terms of accuracy, training speed and model size (Z. Chen, 2018).

Ridge classifier fits the labeled data into a specific data range [-1, +1] and then solves a regression problem which could be a multi-output regression in the multiclass case. The main advantage of this classifier is its speed when compared to other classifiers (Peng, 2020).

SGD classifier which is a linear classifier combined with SGD training in which the derivative of the loss is calculated based on a single random data point instead of the whole group of data points. Being an optimization method SGD is considered popular with different types of classification and not only linear ones (Pal, 2020).

Support Vector Machine classifier is considered to be robust and can perform linear and non-linear classification (via kernel trick). SVM uses the hinge loss function when data are not linearly separable. In practical implementations several kernel functions could be used, such as linear kernel or Radial Basis Function (RBF) kernel and the choice is governed by the number of features. So when we have a huge number of features it is better to use the linear kernel since data would be more likely linearly separable in high dimensions; otherwise we could use the RBF kernel with a suitable cross-validation which helps eliminate the possibility of over fitting.

After building the seven models and calculating their accuracy by using the 20% of the data, the models were installed within the OMNET++ network simulator which provides a reliable simulation framework to network events. Figure (1) illustrates the chosen network for testing. The network scheme includes five hosts that launches several denial of service attacks based on random times and use random traffic otherwise. OMNET++ provides the robust INET framework which reduces the amount of effort needed to setup the traditional network elements.

The choice of this network scheme was made simply to mimic real-life scenarios. Each host which represents a user that connects to the Internet using a router, there might be several users using the same network which is represented with presence of a switch

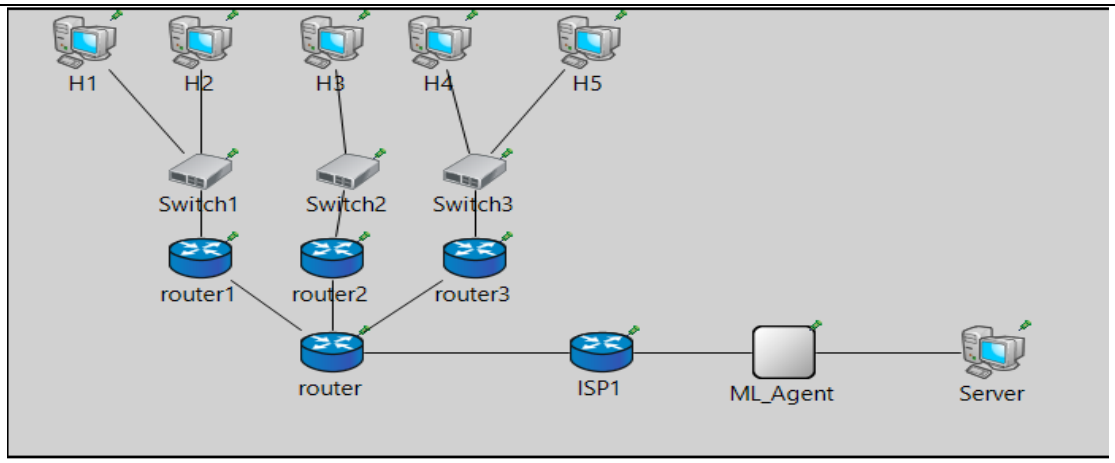

Figure (1) Chosen network diagram for experiments

Each classifier was installed separately on the ML_Agent component and its accuracy to detect the attacks was tested

Table (1) Training and Testing results for the CIC-IDS Dataset

Classifier	Accuracy
Adaboost	0.8642
Gradient Boosting	0.9998
Random Forest	0.99997
XGB	0.99998
Ridge	0.9598
SGD	0.9982
Support Vector Machine	0.99972

Table (2) illustrates the results obtained by using KDD dataset.

Based on launching attacks and working normally we launched 300 attacks randomly and monitored the output of the ML_Agent to specify a more practical accuracy measurement and the results were as depicted in table (3).

Table (3) Model Accuracy in Simulation

Classifier	CIC-IDS	NSL-KDD
Adaboost	0.1	0.1
Gradient Boosting	0.2	0.1
Random Forest	0.2	0.3
XGB	0.3	0.3
Ridge	0.3	0.3
SGD	0.4	0,1
Support Vector Machine	0.2	0.4

3. Results and Discussion :

By using the CIC2018 dataset we obtained the results shown in the table (1):

Table (2) Training and Testing results for the NSL-KDD Dataset

Classifier	Accuracy
Adaboost	0.7854
Gradient Boosting	0.9832
Random Forest	0.9992
XGB	0.9899
Ridge	0.9678
SGD	0.9913
Support Vector Machine	0.9998

Comparing these results to the accuracy in table (1) and table (2) It can be noticed that there is a huge gap between both accuracy values for each classifier. Despite having high accuracy when building and testing a model over a dataset, DoS classification problem suffers mostly from the covariate shift which leads to the fact that traditional classification algorithms are not well-suited for this problem. Features extracted by each classifier were suitable to detect the attack within the scope of the attack but no correct generalization rule can be extracted based on these features Building an ML model with different features with respect to a time scope could further enhance detection and have better results.

4. Conclusions:

http://journal.damascusuniversity.edu.sy/index.php/index/index
ISSN (online): 2789-6854

In this Paper we used seven different common classifiers to help detect the denial of service attacks. These models were later evaluated within a reliable network simulator which resulted that traditional classification algorithms are not quite suitable without proper improvement. Using Adaptive machine learning could reduce the severity of the covariate shift that causes this poor performance of such algorithms. The concept of adaptive machine learning includes the addition of the time dimension to classification process which can be suitable to the DoS problem but further tests should be considered and applied in that matter. The DoS Problem can be analyzed from the practical point of view where the expert can not make a decision based on low-level information such as the number of packets between the client and server, instead he uses high-level information such as resource consumption and number of requests per user within a period of time. The expert can after examining this information to make a decision that moves the state of the service from one state to another. Such description is very close to the Reinforcement learning methods which we expect to be a very promising solution to the DoS problem.

Funding: This Research is funded by Damascus University-Funder no (501100020595)

References:

1-Aljuhani, A. (2021, March 01). Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments. IEEE Access, 9, pp. 42236 - 42264. doi:10.1109/ACCESS.2021.3062909

2-Dong Li, C. Y. (2018). Using SVM to Detect DDoS Attack in SDN Network. The 2nd annual International Conference on Cloud Technology and Communication Engineering, 466. Nanjing. doi:10.1088/1757-899X/466/1/012003

3-Gondi Lakshmeeswari, S. S. (2020). A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression. 14th International Conference INTER-ENG 2020 Interdisciplinarity in Engineering. Mures: MDPI.

4-Graewe, K. (2023, May 5). Artificial Intelligence (AI) for DDoS Mitigation. Retrieved from link11: https://www.link11.com/en/glossar/artificial-intelligence-ai-for-ddos-mitigation

5-Kimmi Kumari, M. M. (2022, April 28). Detecting Denial of Service attacks using machine learning algorithms. Journal of Big Data. Retrieved from https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00616-0

6-Kwekha-Rashid, A. S. (2023). Coronavirus disease (COVID-19(cases analysis using machine0learning applications. Applied Nanoscience, 2013-2025. Retrieved from https://link.springer.com/article/10.1007/s13204-021-01868-7

7-Natekin, A. &. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics. Retrieved from https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021/full

8-Pal, K. &. (2020). Emotion Classification with Reduced Feature Set SGDClassifier, Random Forest and Performance Tuning. International Conference on Computing Science, Communication and Security (pp. 95-108). Singapore: Springer. Retrieved from https://link.springer.com/chapter/10.1007/978-981-15-6648-6_8

9-Peng, C. &. (2020). Discriminative Ridge Machine: A Classifier for High-Dimensional Data or Imbalanced Data. IEEE transactions on neural networks and learning systems, 2595-2609. Retrieved from https://ieeexplore.ieee.org/abstract/document/
Rigatti, S. J . (2017, 1 1)

10 . Random Forest. Journal of Insurance Medecine. Retrieved from https://meridian.allenpress.com/jim/article-abstract/47/1/31/131479/Random-Forest

11-s. Balaji, G. (2021). Impact of Machine Learning and Deep learning techniques for Denial of service attack detection. High Technology Letters, 27(8), 113-123. Retrieved from https://www.researchgate.net/publication/362208061

12_Impact_of_Machine_Learning_and_Deep_learning_techniques_for_Denial_of_service_attack_detection
Schapire, R. E. (2013). Empirical Inference. Berlin: Springer . Retrieved from http://rob.schapire.net/papers/explaining-adaboost.pdf

13)Vinícius De Miranda Rios, P. R. (2022, July 15). Detection and Mitigation of Low-Rate Denial-of-Service Attacks: A Survey. IEEE Access, 10, 76648 - 76668. doi:10.1109/ACCESS.2022.3191430

14Z. Chen, F. J. (2018). XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-Based Cloud. IEEE International Conference on

Big Data and Smart Computing (BigComp) (pp.
251-256). Shanghai: IEEE.