Real-Time DDoS Attack Detection Using Apache Flink and AdaBoost Algorithm

T. Jenifer and S. Shriparen
Department of Computer Science, Faculty of Science, University of Jaffna
jeniferfernando@gmail.com, shriparens@univ.jfn.ac.lk

01. Introduction

- DDoS (Distributed Denial of Service) attacks overwhelm networks with malicious traffic, causing disruptions to legitimate services. Identifying and mitigating these attacks quickly is essential to maintain network stability.
- Machine learning techniques, such as the AdaBoost algorithm, can detect patterns in network traffic that indicate potential threats.
- Apache Flink, a stream processing framework, enables real-time data analysis, allowing for swift detection and response to these evolving security challenges.

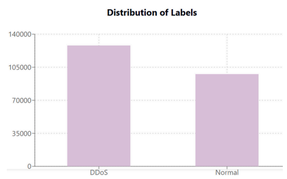
02. Objective

The main objective of this research is to detect DDoS attacks in real-time using the AdaBoost algorithm integrated with Apache Flink, aiming to achieve a faster and more efficient response to potential threats.

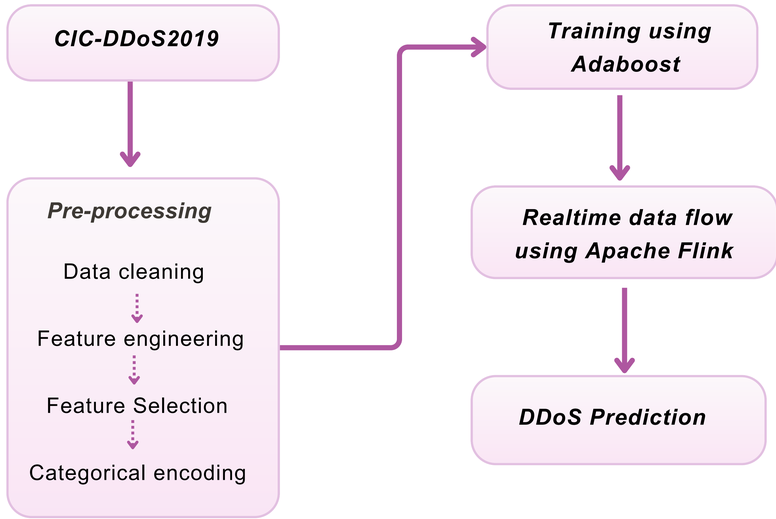
03. DataSet

CIC-DDoS 2019 dataset [05]

- Total number of samples: 225,745
- Number of Features: 85
- Class Distribution:
  DDoS attacks: 128,027 (56.7%)
  Normal traffic: 97,718 (43.3%)



04. Methodology



Initialize Weights $w_i = 1/N$ (No. of samples)

$\alpha = 1/2 \ln\left(\frac{1-e}{e}\right)$

$w_i = w_i \times e^{\alpha}$ if incorrectly classified

$w_i = w_i \times e^{-\alpha}$ if correctly classified

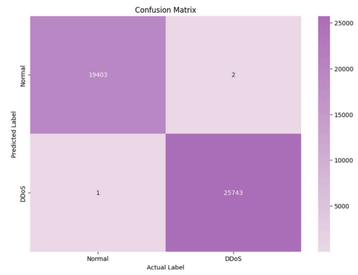
$$e = \frac{\sum_{i=1}^{N} w_i \cdot I(y_i \neq h_t(x_i))}{\sum_{i=1}^{N} w_i}$$

error — weak classifier at iteration t.

Platform	Version
Java	11.0.24
Python	3.9.10
Flink	1.16.3
Flink ML	2.3.0

05. Results/Findings

Studies and Year	Model	Accuracy%	Execution Time (s)	Big Data framework
Zhang, Dai, Li and Zhang, 2018 [01]	Random Forest	97.4%	1.10	Spark, (IDS detection)
Kousar, H. et al. (2021) [02]	Decision Tree Random forest	90.82% 90.86%	-	Spark
Dehkordi, Soltanaghaei and Boroujeni, 2021 [03]	Logistic algorithms	99.62%	1.19	-
Priya, Sivaram, Yuvaraj and Jayanthiladevi , 2020 [04]	Naive Bayes	98.50%	-	-
Our Approach	AdaBoost	99.98%	0.036	Flink



- **Optimized Performance:** Achieved reduced prediction time compared to existing approaches, enabling faster responses to DDoS threats.
- **Minimized Error:** Combined with previous models, the approach significantly reduced overall prediction error.

06. Conclusion

Traditional intrusion detection techniques are effective with slow-speed or small-scale data but often struggle to handle large, high-speed datasets. By leveraging machine learning algorithms in conjunction with Apache Flink, this research enables real-time analysis of big data, significantly enhancing the efficiency and scalability of intrusion detection.

07. References

[01]Zhang, H.; Dai, S.; Li, Y.; Zhang, W. Real-time distributed-random-forest-based network intrusion detection system using Apache spark. In Proceedings of the 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC), Orlando, FL, USA, 17–19 November 2018; pp. 1–7.

[02]Kousar, H. et al. (2021) “DDoS Attack Detection System using Apache Spark,” in 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE.

[03]Dehkordi, A.B.; Soltanaghaei, M.; Boroujeni, F.Z. The DDoS attacks detection through machine learning and statistical methods in SDN. J. Supercomput. 2021, 77, 2383–2415.

[04]Priya, S.S.; Sivaram, M.; Yuvaraj, D.; Jayanthiladevi, A. Machine learning based DDoS detection. In Proceedings of the 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 12– 14 March 2020; pp. 234–237

[05]https://www.kaggle.com/datasets/dhoogla/cicddos2019