


Chapter 2

Measuring Throughput and Latency of Machine Learning Techniques for Intrusion Detection

Winfred Yaokumah

 <https://orcid.org/0000-0001-7756-1832>

University of Ghana, Ghana

Charity Y. M. Baidoo

University of Ghana, Ghana

Ebenezer Owusu

University of Ghana, Ghana

ABSTRACT

When evaluating the effectiveness of machine learning algorithms for intrusion detection, it is insufficient to only focus on their performance metrics. One must also focus on the overhead metrics of the models. In this study, the performance accuracy, latency, and throughput of seven supervised machine learning algorithms and a proposed ensemble model were measured. The study performs a series of experiments using two recent datasets, and two filter-based feature selection methods were employed. The results show that, on average, the naive bayes achieved the lowest latency, highest throughput, and lowest accuracy on both datasets. The logistics regression had the maximum throughput. The proposed ensemble method recorded the highest latency for both feature selection methods. Overall, the Spearman feature selection technique increased throughput for almost all the models, whereas the Pearson feature selection approach maximized performance accuracies for both datasets.

DOI: 10.4018/978-1-6684-9999-3.ch002

INTRODUCTION

In recent times network security attacks and data breaches have become rampant while intrusion detection systems (IDS) are being developed with automated response mechanisms to detect and prevent such attacks (Anwar et al., 2017). An intrusion detection system provides active network security protection mechanisms against cybercriminal activities (Liang et al., 2019). These mechanisms monitor network operations and analyze packets for malicious activities (Paul et al., 2018). Packet analysis can take place on a network (network-based) or a computer (host-based). Network-based intrusion detection systems (NIDS) examine network behavior to determine if a node is under attack, whereas host-based intrusion detection systems (HIDS) check the logs kept by a single host for malicious operations (Chan et al., 2016). However, IDS sometimes fails to detect new external attacks and has a low accuracy rate, and a high false alarm rate (Khraisat et al., 2019; Liu & Lang, 2019). The inability of IDS to prevent network intrusion precisely and timely can jeopardize the information security goals of integrity, confidentiality, and system availability (Mishra & Yadav, 2020). In particular, the IDS that are implemented on networks can escalate the amount of delivery time, thereby reducing the reliability of network traffic (Xia et al., 2015). Also, packet processing may decrease network throughput and increase latency (Tsikoudis et al., 2016).

Two major overhead performance metrics, latency and throughput, are used to measure the IDS's timely detection of network intrusion. In the implementation of IDS, high throughput and low latency of an IDS are essential. Latency measures the time it takes for a processor to receive a request for a byte or message from memory (Hasan et al., 2021). In communication networks, latency or end-to-end delay represents the amount of time it takes for data to be received at its endpoint (destination). Intrusion detection systems frequently suffer from severe latency and network overhead, which make them irresponsive to attacks and the identification of malicious activities (Rahman et al., 2020). Likewise, throughput measures the speed at which the data is transmitted effectively (Ingley & Pawar, 2015). It describes the amount of data that is dispatched over a given time. According to Zhang et al. (2018), with high-speed traffic data, conventional intrusion detection systems experience latency and occasionally fail to identify intrusion patterns. However, to identify suspicious traffic momentarily, a real-time network IDS with machine learning (ML) techniques can promptly process vast volumes of network traffic data.

Specifically, machine learning (ML) techniques are enabling the modeling of IDS to provide significantly higher rates of intrusion detection (Srivastava et al., 2019). Some intrusion detection systems use ML classification algorithms to categorize network traffic as normal or irregular (Kaya, 2020; Thaseen & Kumar, 2017). For instance, the IDS implementation that employs ML approaches for classification includes Naive Bayes (NB), Adaptive Boost, PART (Kumar & Doegar, 2018), Active learning Support Vector Machine, Fuzzy C-Means clustering (Kumari & Varma, 2017), Multi-Layer Perceptron, Bayesian Network, Support Vector Machine (SVM), Adaboost, Random Forest (RF), Bootstrap Aggregation, Decision Tree (Halibas et al., 2018), Random Forest (Park et al., 2018); SVM (Kotpalliwar & Wajgi, 2015), Genetic Algorithm, and Support Vector Machine (Gharaee & Hosseinvand, 2017). These ML algorithms may provide varying levels of accuracy, latency, and throughput.

According to Yu et al. (2018), measuring the overhead performance of the learning algorithms accurately is important for determining their effectiveness. However, currently, few studies investigate the efficiency of IDS, considering its training time, testing time, latency, throughput, and detection time (Maseer et al., 2021). Thus, this experimental study focuses on supervised machine learning algorithms to measure latency and throughput using two intrusion detection datasets and two filter-based feature

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/measuring-throughput-and-latency-of-machine-learning-techniques-for-intrusion-detection/334467

Related Content

Architecting IoT based Healthcare Systems Using Machine Learning Algorithms: Cloud-Oriented Healthcare Model, Streaming Data Analytics Architecture, and Case Study

G. S. Karthick and P. B. Pankajavalli (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 198-223).

www.irma-international.org/chapter/architecting-iot-based-healthcare-systems-using-machine-learning-algorithms/307453

New Cloud Computing-Based Strategy for Coordinating Multi-Robot Systems

Claudio Urrea (2023). *Handbook of Research on Machine Learning-Enabled IoT for Smart Applications Across Industries* (pp. 232-257).

www.irma-international.org/chapter/new-cloud-computing-based-strategy-for-coordinating-multi-robot-systems/325999

Power Consumption Prediction of IoT Application Protocols Based on Linear Regression

Sidna Jeddou, Amine Baina, Najid Abdallah and Hassan El Alami (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-16).

www.irma-international.org/article/power-consumption-prediction-of-iot-application-protocols-based-on-linear-regression/287585

Generating an Artificial Nest Building Pufferfish in a Cellular Automaton Through Behavior Decomposition

Thomas E. Portegys (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-12).

www.irma-international.org/article/generating-an-artificial-nest-building-pufferfish-in-a-cellular-automaton-through-behavior-decomposition/233887

A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques

Ramanujam Elangovan and Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 39-56).

www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127