Utilizing machine learning techniques for network traffic anomaly detection

Xiang Li[1], Guojun Shi[2], Yuxin Wu[3,4]

[1]Shanghai Second Polytechnic University, Shanghai, 201209, China
[2]University of Nottingham Ningbo China, Ningbo, 315100, China
[3]University of Illinois Urbana-Champaign, Champaign, 61820-5711, USA

[4]wuaria0@gmail.com

Abstract. The landscape of network traffic anomaly detection has evolved considerably in recent times, driven largely by the advent of pioneering algorithms. This article undertakes an exhaustive comparative exploration of some of the most contemporary algorithms, namely Prophet, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Isolated Forest (IF), and OmniAnomaly. Delving deep into their distinctive features, functionalities, and practical applications, the paper sheds light on the factors that render these algorithms superior to their predecessors. The discourse commences with a prologue underscoring the escalating importance of network traffic anomaly detection, especially in the face of the burgeoning cyber threats of the modern era. This sets the stage for a presentation on the focal algorithms - Prophet, RNN, CNN, IF, and OmniAnomaly. Each algorithm is then dissected to provide readers with a nuanced understanding of its underlying mechanics and methodologies. Furthermore, the discourse amplifies the breakthroughs and innovations underpinning each algorithm, highlighting attributes such as heightened accuracy, lucid interpretability, proficiency in deciphering intricate patterns, and the agility to detect anomalies in real time. Factors like computational agility, resilience, structural intricacy, and versatility across varied operational terrains are assessed in a meticulous comparative framework. Drawing from empirical evidence available in extant literature, the article underscores the stellar performance of these algorithms, benchmarked using quantitative metrics like precision.

Keywords: Network Anomaly Detection, Intrusion Detection, Comparative Analysis.

1. Introduction

As the digital landscape of contemporary society expands, the emphasis on cybersecurity has intensified. Traditional security mechanisms now grapple with the burgeoning intricacies of cyber threats, particularly with the inception of sophisticated incursions like Advanced Persistent Threats (APTs). This changing paradigm underscores the paramountcy of intrusion detection and an astute analysis of threat intelligence in the modern annals of cybersecurity.

Legacy Intrusion Detection Systems (IDS) grapple with the dynamism of novel network threats. Parallelly, the ascendancy of technologies such as machine learning and deep learning heralds a promising era for network security. The recent past has seen the unveiling of innovative algorithms and models tailored to amplify intrusion detection precision, timeliness, and versatility. Noteworthy among these are algorithms like Prophet, Recurrent Neural Networks, Convolutional Neural Networks, Isolation Forest, and OmniAnomaly, each possessing its distinctive strengths and operational niches.

This study seeks to meticulously juxtapose and dissect these avant-garde algorithms, illuminating their edge over conventional methodologies. It ventures deep into the operational ethos, attributes, and real-world deployment scenarios of each algorithm, elucidating the transformative strides they have engendered in the network security landscape. In terms of structure, the article is delineated as follows: Section II offers insights into the various types of anomalies, underpinned by practical applications and exemplars. Section III navigates the myriad machine learning techniques tailored for anomaly detection, elucidating their merits and limitations. Section IV furnishes an overarching discourse capturing the paper's salient features, culminating in a conclusion presented in Section V.

2. Anomaly Detection

2.1. Anomaly types

The classification of anomalies is fundamentally organized into three principal categories. These categories are: 'Point anomalies', encompassing individual data instances differing from the norm; 'Contextual anomalies', which account for irregularities tied to specific contexts, particularly relevant for time-series data; and 'Collective anomalies', involving the collective analysis of data instances to discern deviations from the expected pattern [1].

2.2. Machine Learning Approaches

This section comprehensively examines prominent algorithms in the intrusion detection domain, namely 1) Prophet, 2) Recurrent Neural Networks, 3) Convolutional Neural Networks, 4) Isolation Forest (IF), and 45 OmniAnomaly, elucidating their unique methodologies and strengths [2].

(1) Prophet:

Prophet is a time-series forecasting model that has been adapted for anomaly detection purposes. It leverages an additive model that captures various components of time-series data, including trend, seasonality, and holiday effects [3]. Prophet is particularly well-suited for datasets with seasonal patterns and is characterized by its ability to handle missing data points effectively. The algorithm's strength lies in its interpretability and simplicity.

(2) Recurrent Neural Networks:

RNNs are a class of neural networks designed to capture sequential dependencies within data. In the context of intrusion detection, RNNs excel at modeling temporal relationships within network traffic data. By processing data sequentially and using feedback connections, RNNs can capture patterns that extend over time. This makes them adept at detecting anomalies that evolve gradually. However, RNNs can suffer from the vanishing gradient problem and require careful hyperparameter tuning to prevent issues like overfitting [4].

(3) Convolutional Neural Networks:

CNNs are primarily known for their prowess in image analysis, but they have also found utility in intrusion detection. CNNs can extract hierarchical features from input data, which makes them suitable for identifying complex patterns in network traffic. Their ability to automatically learn features reduces the need for manual feature engineering [5]. However, applying CNNs to network traffic data requires transforming the sequential data into suitable formats, and their performance heavily depends on the architecture design and dataset characteristics.

(4) Isolation Forest:

IF is an unsupervised anomaly detection algorithm that exploits the concept of isolating anomalies by recursively partitioning data points. Unlike traditional clustering methods, IF creates isolation trees to isolate anomalies in a shorter number of splits, making it efficient even in high-dimensional spaces. IF is adept at identifying individual anomalies and performs well when anomalies are sparse and distinct from the majority [6]. However, it might struggle with identifying anomalies that are surrounded by other anomalies.

(5) OmniAnomaly:

OmniAnomaly is a recently proposed algorithm that addresses the limitations of existing anomaly detection methods. It integrates multiple models, including autoencoders and matrix factorization, to capture different facets of anomalies [7]. By leveraging diverse models, OmniAnomaly aims to enhance detection accuracy across various types of anomalies. Its versatility and ability to accommodate complex patterns make it a promising candidate for detecting sophisticated attacks.

2.3. Motivation

The evolving nature of network anomalies poses a significant challenge, prompting the utilization of diverse machine learning (ML) techniques for anomaly detection in network traffic. However, determining the most effective ML techniques tailored to specific datasets for optimal results remains an open question [8]. This knowledge gap underscores the need to gain a comprehensive understanding of supervised, unsupervised, and reinforcement approaches employed in recent anomaly detection applications. Thus, this literature survey primarily aims to enhance comprehension of the array of existing ML techniques in the realm of Intrusion Detection, thereby facilitating future advancements in this field [9].

3. Related Works

3.1. Prophet

The introduction of Prophet, originally designed for time-series forecasting, has provided a fresh perspective on anomaly detection. By explicitly capturing seasonal patterns, trends, and holidays, Prophet has showcased enhanced accuracy in identifying anomalies in network traffic data with inherent temporal dependencies. Its contribution lies in its adaptability to dynamic network environments and its simplicity, making it an accessible solution for both experts and practitioners.

3.2. RNN and CNN

The utilization of Recurrent Neural Networks and Convolutional Neural Networks has revolutionized the analysis of sequential network data and complex patterns. RNN's ability to model sequential dependencies has resulted in improved anomaly detection accuracy for evolving threats that manifest over time. Similarly, CNN's feature extraction capabilities have enabled the identification of intricate anomalies that might be missed by traditional methods. The contributions of RNN and CNN lie in their ability to capture nuanced patterns, leading to enhanced detection accuracy and timely response.

Table 1. Comparision of Anomaly Detection Algorithms, Accuracy Performance.

Article	Year	Model	Algorithm	Dataset	Detection Accuracy
Freeman et al.	2021	AM	Prophet	Personal Dataset	0.990
Ring et al.	2021	WaveNet	CNN	PLAID	0.936
Rajapaksha et al.	2023	LSTM	RNN	Public real data (HCRL CH)	0.980
Liu et al.	2023	SVM	IF	PUMP	0.977
Liu et al.	2023	LSTM	OmniAnomaly	SWaT	0.983

3.3. Isolation Forest

The innovation introduced by Isolation Forest lies in its approach to detecting anomalies by isolating them with a minimal number of partitioning steps. This efficiency allows IF to excel in scenarios where anomalies are distinct and sparse, thus enhancing its applicability in real-world intrusion detection. IF's contribution addresses the challenge of efficiently detecting individual anomalies in high-dimensional data spaces, providing an effective alternative to traditional clustering methods [10].

3.4. OmniAnomaly

The OmniAnomaly algorithm stands out for its holistic approach to anomaly detection. By integrating multiple models and techniques, such as auto-encoders and matrix factorization, OmniAnomaly has demonstrated improved accuracy across diverse types of anomalies. Its contribution lies in addressing the limitations of single-model approaches by leveraging the strengths of various techniques. This innovation has significantly enhanced the algorithm's robustness and detection capabilities.

4. Explanation of Results

The ascent of deep learning techniques has catalyzed significant advancements in intrusion detection. Convolutional Neural Networks and Recurrent Neural Networks, both representing deep learning methods, have excelled in interpreting image and sequential data, autonomously discerning abstract features and potential threat patterns, thus enhancing detection precision and efficiency. Novel algorithmic structures and designs have emerged, addressing prior constraints. As shown in Table 1. For instance, Prophet's incorporation of seasonal pattern modeling benefits time series anomaly detection, including network traffic anomalies. Isolation Forest boosts efficiency in identifying anomalies within high-dimensional data through isolation techniques. This innovation-driven design evolution fosters enhanced intrusion detection performance. The contemporary network landscape's remarkable surge in data volume and diversity provides new challenges and opportunities. The prevalence of expansive, complex network traffic data allows for richer datasets and unveils intricate attack patterns, propelling algorithmic advancements.

Moreover, modern algorithms adeptly process this influx of data, uncover concealed anomaly patterns, and adeptly learn more precise feature representations. Intrusion detection methodologies have embraced multimodal data analysis, amalgamating diverse data types. This multimodal analysis prowess enables algorithms to comprehend and assess data from various angles, markedly heightening detection accuracy. An illustrative example is the OmniAnomaly algorithm's comprehensive performance achieved through the fusion of multiple models and techniques.

4.1. Rise of Deep Learning Techniques

The rise of deep learning techniques has brought breakthroughs in the field of intrusion detection. Deep learning methods such as Convolutional Neural Networks and Recurrent Neural Networks have excelled in the analysis of image and sequence data. These algorithms can automatically learn abstract features from raw data and recognize potential threat patterns from them, thus improving the accuracy and efficiency of detection.

4.2. Innovations in Algorithm Structure and Design

New algorithms feature innovative structures and designs. For instance, Prophet introduces modeling of seasonal patterns, advantageous for anomaly detection in time series data, including network traffic. Isolation Forest enhances efficiency in anomaly detection within high-dimensional data by isolating isolated points. These innovative designs address the limitations of traditional methods, leading to better performance in intrusion detection.

4.3. An increase in data size and diversity

The current network environment is characterized by an exponential increase in data size and diversity. The availability of large-scale, high-dimensional network traffic data and the emergence of complex attack patterns provide rich datasets and challenges for algorithm development. New algorithms are better able to process this data, identify hidden anomaly patterns, and learn more accurate feature representations.

4.4. Multimodal Data Analysis

The new generation of intrusion detection algorithms are beginning to use multimodal data analysis that combines different types of data information. This multimodal analysis capability allows algorithms to capture and analyze data from multiple perspectives, further improving detection accuracy. For example, the comprehensive performance of the OmniAnomaly algorithm is achieved by integrating multiple models and techniques.

5. Conclusion

The landscape of anomaly detection in network traffic has undergone profound transformation, with a shift towards advanced algorithms that promise heightened efficiency and accuracy. Through an in-depth analysis of these state-of-the-art methods, this paper elucidates the significant enhancements they bring over conventional techniques. One notable change in the realm of intrusion detection has been the assimilation of cutting-edge machine learning strategies. Among the algorithms under scrutiny are Prophet, Recurrent Neural Networks, Convolutional Neural Networks, Isolation Forest, and OmniAnomaly. Each one of these offers unparalleled benefits in managing diverse anomaly patterns. The advent of deep learning strategies, especially as seen in Recurrent Neural Networks and Convolutional Neural Networks, heralds substantial advancements in the timely and precise identification of intricate threat patterns. Certain attributes make these algorithms stand out: the adeptness of Prophet in handling time-series data, and the prowess of Isolation Forest in navigating high-dimensional terrains, are cases in point. Moreover, OmniAnomaly's innovative approach of integrating multiple models points to a bright future, hinting at augmented robustness in detection strategies.

With the realm of network security constantly grappling with increasingly sophisticated cyber threats, these groundbreaking algorithms emerge as the vanguard, defending and fortifying detection mechanisms. Anticipated future explorations might dwell on amalgamating these algorithms or devising bespoke adaptations for specific network scenarios to refine precision and enhance efficiency. This comprehensive examination, thus, paves the way for subsequent breakthroughs in the all-important arena of network traffic anomaly detection.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Ali, W. A., Manasa, K. N., Bendechache, M., Fadhel Aljunaid, M., & Sandhya, P. (2020). A review of current machine learning approaches for anomaly detection in network traffic. Journal of Telecommunications and the Digital Economy, 8 (4), 64-95.

[2] Mokhtari, S., Abbaspour, A., Yen, K. K., & Sargolzaei, A. (2021). A machine learning approach for anomaly detection in industrial control systems based on measurement data. Electronics, 10(4), 407.

[3] NG, B. A., & Selvakumar, S. (2020). Anomaly detection framework for Internet of things traffic using vector convolutional deep learning approach in fog environment. Future Generation Computer Systems, 113, 255-265.

[4] Hwang, R. H., Peng, M. C., Huang, C. W., Lin, P. C., & Nguyen, V. L. (2020). An unsupervised deep learning model for early network traffic anomaly detection. IEEE Access, 8, 30387-30399.

[5] Lin, P., Ye, K., & Xu, C. Z. (2019). Dynamic network anomaly detection system by using deep learning techniques. In Cloud Computing–CLOUD 2019: 12th International Conference, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 12 (pp. 161-176). Springer International Publishing.

[6] Su, T., Sun, H., Zhu, J., Wang, S., & Li, Y. (2020). BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset. IEEE Access, 8, 29575-29585.

[7] Bhatia, R., Benno, S., Esteban, J., Lakshman, T. V., & Grogan, J. (2019, December). Unsupervised machine learning for network-centric anomaly detection in IoT. In Proceedings of the 3rd acm conext workshop on big data, machine learning and artificial intelligence for data communication networks (pp. 42-48).

[8] Xu, H., Sun, Z., Cao, Y., & Bilal, H. (2023). A data-driven approach for intrusion and anomaly detection using automated machine learning for the Internet of Things. Soft Computing, 1-13.

[9] Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEE Access, 9, 103906-103926.

[10] Guezzaz, A., Asimi, Y., Azrour, M., & Asimi, A. (2021). Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection. Big Data Mining and Analytics, 4(1), 18-24.