RADemics

Scalability Challenges and Solutions in Machine Learning Algorithms for High-Throughput Intrusion Detection Systems

Gomathi N, R. Navaneethakrishnan
SENGUNTHAR ENGINEERING COLLEGE, GLOBAL COLLEGE OF ARTS AND SCIENCE

Scalability Challenges and Solutions in Machine Learning Algorithms for High-Throughput Intrusion Detection Systems

[1] Gomathi N, Assistant Professor, Department of Biomedical Engineering, Erode Sengunthar Engineering College, Palakarai, Thuduppathi, Perundurai - 638057, Tamil Nadu, India. gomathiesec@gmail.com

[2] R. Navaneethakrishnan, Head and Assistant Professor, Department of Computer Science, Global College of Arts and Science, Ammaiyappan-613701, Thiruvarur District, Tamil Nadu, India. msgtokrishnan@gmail.com

Abstract

This book chapter explores the scalability challenges and solutions in applying machine learning algorithms to high-throughput Intrusion Detection Systems (IDS). As cyber threats continue to evolve, traditional IDS struggle to keep pace with the enormous volume, velocity, and variety of data in modern network environments. Machine learning offers promising solutions for enhancing detection accuracy and efficiency; however, integrating these models into high-throughput systems presents significant obstacles related to latency, processing time, and resource consumption. This chapter discusses key issues such as the impact of data volume, real-time processing requirements, and computational overhead on scalability. It examines optimization techniques including model compression, pruning, and hybrid approaches that combine machine learning with rule-based systems. The chapter concludes by proposing future directions for scalable IDS, emphasizing the need for innovative algorithms and architectures to ensure robust and timely threat detection in complex environments.

Keywords: Scalability, Intrusion Detection Systems, Machine Learning, High-Throughput, Model Optimization, Real-Time Processing.

Introduction

In today's digital landscape, where cyber threats are becoming increasingly sophisticated, the need for robust Intrusion Detection Systems (IDS) has never been more critical [1]. High-throughput environments, such as large-scale enterprise networks, cloud infrastructures, and Internet of Things (IoT) ecosystems, generate massive volumes of data that must be processed in real time to identify potential security threats [2,3]. Traditional IDS, primarily rule-based systems, often struggle to scale effectively in these environments, facing challenges such as high false positive rates and limited adaptability to new attack patterns [4-6]. Machine learning (ML) offers a promising solution, enabling systems to learn from data and continuously improve detection accuracy [7]. However, the application of ML in high-throughput IDS brings forth several scalability challenges, primarily relating to the computational complexity of algorithms and the need for rapid processing to maintain system performance [8,9].

The scalability of IDS becomes particularly important when considering the data volumes in high-throughput environments [10,11]. Networks today generate enormous amounts of data, often reaching terabytes or even petabytes, that must be analyzed to detect potential intrusions [12-15]. For an IDS to operate effectively in such conditions, it must be capable of processing this data in real-time, which requires efficient algorithms and optimized system architectures [16,17]. Machine learning algorithms, although powerful in detecting complex attack patterns, often struggle with the computational overhead required for processing large-scale data [18]. As the data volume increases, so does the demand for memory, processing power, and storage, which can introduce latency and slow down detection times [19,0]. Real-time processing requirements exacerbate this challenge, as there was a need to balance accuracy and response time [21,22].

Machine learning holds the potential to significantly enhance the efficiency and effectiveness of IDS by automating the detection process and improving the system's ability to identify unknown threats [23]. Unlike traditional signature-based methods, which rely on predefined patterns of known attacks, ML models can learn from historical data, allowing them to recognize new and evolving threats [24]. In addition, ML algorithms can adapt to the changing landscape of cyberattacks, continuously improving their detection capabilities as are exposed to new types of data [25]. However, despite these advantages, the integration of machine learning into IDS introduces a new set of challenges. The complexity of training and deploying these models in high-throughput environments requires addressing issues such as computational overhead, model size, and the time required for data processing. Ensuring that these models can operate efficiently at scale, without compromising detection accuracy, was essential for maintaining system performance.