RADemics

# Comparative Analysis of Machine Learning Algorithms for Anomaly Detection in Large-Scale Distributed Networks

Rachhapl Singh, Balwinder Kaur
PUNJABI UNIVERSITY CENTRE FOR EMERGING AND INNOVATIVE TECHNOLOGY

# Comparative Analysis of Machine Learning Algorithms for Anomaly Detection in Large-Scale Distributed Networks

[1]Rachhapl Singh, Asssistent Professor (CS), Punjabi University centre for Emerging and Innovative Technology, Mohali, Punjab, Affliated with Punjabi University, Patiala, Punjab, India. rachhpal@pbi.ac.in

[2]Balwinder Kaur, Assistant Professor (CS), Punjabi University, Patiala, Punjab, India. Affiliated with Punjabi University Centre for Emerging and Innovative Technology, Mohali, Punjab, India. balwinder_coem@pbi.ac.in

## Abstract

This chapter presents a comprehensive analysis of machine learning algorithms for anomaly detection in large-scale distributed networks, a critical area in modern network management. As distributed systems, such as cloud computing and IoT, continue to expand, ensuring their security and operational efficiency becomes paramount. The chapter explores various machine learning techniques, evaluating their effectiveness in detecting anomalies such as intrusions, faults, and performance degradation. Key algorithms, including supervised, unsupervised, and reinforcement learning models, are assessed for their ability to identify abnormal patterns in diverse network environments. Emphasis was placed on the challenges of handling large volumes of data, scalability concerns, and real-time processing requirements. The chapter discusses the integration of anomaly detection systems with network monitoring tools to enhance decision-making and response times. Insights provided in this work offer valuable guidance for researchers and practitioners aiming to optimize anomaly detection systems in dynamic, distributed network infrastructures.

**Keywords:** Anomaly Detection, Machine Learning, Distributed Networks, Network Security, IoT, Cloud Computing.

## Introduction

In recent years, the rapid expansion of distributed networks, such as cloud computing, the Internet of Things (IoT), and edge computing, has brought about significant advancements in connectivity, data management, and service delivery [1,2]. However, this growth has also introduced new complexities and challenges in managing and securing these vast, dynamic infrastructures [3,4]. One of the most pressing concerns in large-scale distributed systems was the detection and mitigation of anomalies that could lead to system failures, security breaches, or performance degradation [5]. Anomaly detection plays a crucial role in ensuring the seamless operation of distributed networks, and the integration of machine learning (ML) algorithms has shown significant promise in addressing these challenges [6,7]. By leveraging the power of machine learning, these systems can autonomously detect irregular patterns in network traffic, identify potential threats, and enable timely interventions before escalate into more serious issues [8,9].

The increasing volume of data generated within distributed systems demands efficient and scalable anomaly detection solutions [10]. Traditional methods of detecting anomalies, such as rule-based approaches, are often limited by their inability to scale with the growing complexity and size of modern networks [11,12]. Machine learning offers a more adaptive and robust solution, capable of learning from vast datasets and continuously improving detection accuracy over time [13,14]. With machine learning, systems can detect both known and unknown anomalies by analyzing large volumes of real-time data, identifying patterns thatnot be immediately apparent to human operators [15]. This chapter focuses on exploring various machine learning techniques, including supervised, unsupervised, and reinforcement learning models, to enhance anomaly detection capabilities in large-scale distributed environments [16-19].

One of the major challenges in anomaly detection for distributed networks was the sheer scale and diversity of data involved [20]. Large-scale networks often comprise thousands or even millions of interconnected devices, each generating a unique set of data points [21]. As such, identifying anomalies within this massive volume of information requires algorithms that are not only accurate but also capable of handling large datasets in real-time. Additionally, the complexity of distributed systems means that anomaliesmanifest in different forms, such as abnormal traffic patterns, performance issues, or security breaches, each requiring a different detection approach [22-25]. To address these challenges, machine learning models must be flexible, scalable, and capable of adapting to the evolving nature of the network and its associated data.