# **Real-time Anomaly Detection using Neural Networks and Big Data in the context of Cybersecurity**

Abstract

Anomaly detection has become increasingly popular in recent years, and there are many different types of algorithms that can be used for this purpose. In this study, we explored the applicability of convolutional neural networks (CNN) for this type of problem.

1. Introduction

Data analytics has evolved dramatically in recent years, developing new techniques and technologies for analyzing large amounts of data. One such approach is the use of neural networks, which shows great promise in the ability to learn from and make predictions based on complex data sets. It has been used in a wide range of applications, from image recognition to natural language processing, with impressive results.

In this paper, we will explore the use of neural networks combined with big data storage and processing technologies to develop new real-time anomaly detection techniques. Anomaly detection is a method of identifying unusual or unexpected patterns in data, which may indicate errors, fraud, or other issues By combining the power of neural networks and the amount of data available through big data technology together, we aim for more accurate and efficient methods to detect anomalies in real-time.

Our approach uses big data technology to collect and store large amounts of real-time data, which can then be analyzed by neural networks. These networks are trained to recognize patterns and relationships in the data, enabling them to detect anomalies that might be difficult to detect with traditional methods that. By using this technique in real time we hope to improve our ability to identify abnormalities as they occur, resulting in faster response times and more effective decision-making.

This paper will provide an overview of our approach, including a discussion of relevant literature and a description of our methodology. We will also present the results of our research, including our findings and implications for further analysis of the field. Finally, we will discuss the limitations of our approach and propose future research directions.

1. Literature Review

Real-time anomaly detection is a key component of cyber security, as it allows for rapid identification and mitigation of potential security threats. In recent years, there has been a growing interest in using neural networks and big data technologies for real-time anomaly detection in cyber security applications.

One of the key advantages of using neural networks for anomaly detection is their ability to learn from and make predictions based on complex data sets. For example, Fu Z. (2016) presented a method to monitor and prevent malicious attacks on computer networks using a recurrent neural network model to process sequential data and do risk analysis. The proposed system was tested through massive experiments, demonstrating its reliability and the effectiveness of using neural networks for real-time anomaly detection.

Sometimes anomaly detection models, especially in the case of intrusion detection systems (IDS), can produce false positives. An example on how deep learning can help in this case can be found in X (20XX). The paper proposes using deep learning models, especially long short-term memory (LSTM), instead of traditional machine learning models, such as support vector machines (SVM), to optimize the anomaly detection performance and reduce the false-positive rate. The paper reports that LSTM achieves 10% lower false-positive rate than SVM, and higher accuracy on unseen data. The paper also claims that LSTM can detect more types of attacks, such as contextual and collective attacks, than traditional models.

For X (2022) a small dataset has been used to train the model, but in the discussion the author mentioned that deep learning models perform better with big data, as seen in Fig. Because of this, in addition to the use of neural networks, there has also been signigicant interest in the use of big data technologies for real-time anomaly detection. These technologies, such as Hadoop and Spark, allow for the efficient storage and processing of large amounts of data, enabling real-time analysis. For example, X (2018) conducted a comprehensive review of existing research on real-time intrusion detection using big data and machine learning techniques, including neural networks. They found that big data technologies can significantly improve the performance of intrusion detection systems by enabling real-time analysis of large amounts of data.

Al Jallad, K., Aljnidi, M. & Desouki (2019) is another good example of how using big data analysis with deep learning in anomaly detection is an excellent combination that may be optimal solution. The system proposed in this paper uses a distributed framework (Apache Spark) to handle large-scale and heterogeneous data sources, such as flow traffic, traffic aggregation, and labels. A recurrent neural network with long short-term memory (LSTM) has then being used to learn the normal pattern of the network traffic and classify each time frame as normal or anomalous. Their model got 10% less false positive than a traditional learning model.

Overall, the existing literature suggests that the combination of neural networks and big data technologies can be highly effective for real-time anomaly detection in cybersecurity applications. The most used neural network is the recurrent type, while in few cases like Lu,Wei, Li & Wang (2018) Convolutional Neural Networks have being used to train a neural network to detect anomalies in system logs. The paper compares the CNN-based model with other approaches using Long Short term memory (LSTM) and Multilayer Perceptron (MLP) on big data logs. The paper claims that the CNN-based model has better accuracy (reaches to 99%) than the other approaches.

Based on the literature mentioned above, convolutional neural networks (CNN) has been used for anomaly detection in network traffic data; this will also provide an additional study on those models were used in this instance.

1. Methodology
2. Results
3. Discussion
4. Conclusion