RADemics

Transformer Encoder Decoder Frameworks for Intrusion Detection and Cyber Threat Prediction

Mohan B. A, E. G. Satish.
BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT,
NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY.

2. Transformer Encoder Decoder Frameworks for Intrusion Detection and Cyber Threat PredictionAbstract

[1]Mohan B. A., Associate Professor, Department of Computer Science and Engineering, BMS Institute of Technology and Management, Bengaluru, Karnataka, India. ba.mohan@bmsit.in

[2]E. G. Satish, Assistant Professor, Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India. satish.eg@nmit.ac.in

Abstract

This chapter explores the transformative role of Transformer Encoder-Decoder architectures in advancing intrusion detection and cyber threat prediction. Leveraging the power of attention mechanisms, these models capture complex relationships in sequential data, enhancing the identification of malicious activities in network traffic and system logs. The chapter provides a comprehensive overview of Transformer models, including key components such as self-attention, multi-head attention, and the encoder-decoder framework. It highlights how these architectures address the challenges of traditional methods, offering improved scalability and adaptability to evolving cyber threats. The chapter delves into innovative applications of Transformers in cybersecurity, focusing on real-time anomaly detection and predictive threat modeling. By examining recent advancements in Transformer variants, such as sparse transformers, this chapter also addresses the growing need for computational efficiency in large-scale cybersecurity systems. The chapter concludes with an outlook on future directions and research gaps in the application of Transformer models to cybersecurity.

Keywords:

Transformer Architecture, Encoder-Decoder Framework, Intrusion Detection, Cyber Threat Prediction, Anomaly Detection, Computational Efficiency.

Introduction

The increasing complexity of cyber threats and the rapid growth of digital infrastructures have made traditional methods of intrusion detection and cyber threat prediction insufficient [1]. Conventional techniques, such as rule-based systems and signature-based approaches, often struggle to adapt to new, evolving threats [2]. As a result, there was a growing need for more advanced models capable of learning from vast amounts of data and identifying previously unseen attack patterns [3]. Transformer Encoder-Decoder architectures, which have shown remarkable success in natural language processing and machine learning, have emerged as powerful tools to tackle these challenges [4-6]. Their ability to handle long-range dependencies and capture intricate relationships within data makes them particularly effective for intrusion detection and cybersecurity applications [7,8].

The core innovation behind Transformer models lies in their self-attention mechanism, which allows each token in a sequence to attend to every other token [9,10]. This contrasts with earlier models, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs), which process data sequentially [11,12]. Self-attention enables Transformers to capture complex, non-linear relationships in data without being constrained by the limitations of sequential processing [13]. This capability was crucial for analyzing time-series data from network traffic or system logs, where the relationships between events span long distances [14]. The attention mechanism ensures that significant patterns, even those occurring at distant points in the sequence, are not overlooked, making Transformers particularly suited for detecting complex cyber threats [15,16].

One of the most significant advantages of Transformer models was their encoder-decoder framework, which has proven highly effective in sequence-to-sequence tasks [17]. In the context of intrusion detection, the encoder was responsible for processing input data, such as network packets or system logs, and encoding it into a latent representation that captures the underlying patterns [18,19]. The decoder then uses this representation to predict the likelihood of a cyber-attack or identify specific types of anomalies [20]. This separation of encoding and decoding allows for efficient learning of both input features and output relationships, enhancing the model's ability to make accurate predictions in dynamic cybersecurity environments [21,22].

One such challenge was the computational cost associated with training and inference, particularly when dealing with long sequences of data [23]. Transformers traditionally rely on full attention mechanisms, where every token attends to all others, resulting in quadratic complexity with respect to the input size [24]. This can lead to inefficiencies, especially when processing large datasets, such as lengthy network traffic logs or extensive security event histories [25]. To address this, recent advancements have focused on optimizing the attention mechanism to make it more scalable, such as sparse attention models, which limit the number of tokens each token attends to. These innovations aim to strike a balance between maintaining the effectiveness of Transformers while reducing their computational burden.