RADemics

AI-Driven SIEM (Security Information and Event Management) Systems Using Long Short-Term Memory (LSTM) for Log-Based Threat Detection

Surbhi Choudhary, S. Kalaiarasi, A Joshua Sundar Raja

PES UNIVERSITY, SAVEETHA SCHOOL OF ENGINEERING, ST. JOSEPH'S COLLEGE (ARTS & SCIENCE)

AI-Driven SIEM (Security Information and Event Management) Systems Using Long Short-Term Memory (LSTM) for Log-Based Threat Detection

[1]Surbhi Choudhary, Assistant Professor, CSE, PES University, EC Campus, Hosur Rd, Konappana Agrahara, Electronic City, Bengaluru, Karnataka 560100, sur.choudhary17@gmail.com

[2]S. Kalaiarasi, Professor, Saveetha School of Engineering, SIMATS. Chennai, kalaiarasis.sse@saveetha.com

[3]A Joshua Sundar Raja, Assistant Professor, English, St. Joseph's College (Arts & Science), Kovur, Chennai - 600128. joshuaresearcher@gmail.com

Abstract

The increasing sophistication of cyber threats necessitates the adoption of advanced techniques for real-time anomaly detection in Security Information and Event Management (SIEM) systems. Traditional rule-based and signature-based approaches are no longer sufficient to address emerging attack vectors and the growing volume of security logs. This chapter explores the integration of Long Short-Term Memory (LSTM) networks into SIEM systems for log-based threat detection, highlighting their capacity to capture temporal dependencies and identify subtle patterns in sequential log data. Despite their effectiveness, challenges such as data privacy concerns, limited access to high-quality labeled datasets, and computational complexity remain. To overcome these obstacles, privacy-preserving data synthesis techniques, such as Generative Adversarial Networks (GANs) and differential privacy, are proposed to generate realistic, high-quality synthetic datasets for model training, ensuring data confidentiality and regulatory compliance. The chapter discusses the potential of LSTM-based SIEM systems in enhancing cybersecurity defenses, as well as ongoing research efforts to address the scalability, accuracy, and interpretability of AI-driven models. Key research gaps and future directions in the application of LSTM to SIEM are also presented. This work provides valuable insights into the development of next-generation AI-driven cybersecurity solutions that can dynamically adapt to the evolving threat landscape.

Keywords: Security Information and Event Management (SIEM), Long Short-Term Memory (LSTM), Log-Based Threat Detection, Privacy-Preserving Data Synthesis, Generative Adversarial Networks (GANs), Cybersecurity.

Introduction

The growing complexity and frequency of cyberattacks present significant challenges to organizations striving to secure their digital infrastructures. Traditional cybersecurity approaches, relying on rule-based and signature-based detection methods, have proven to be inadequate in addressing the dynamic nature of modern cyber threats. With the sheer volume of data generated by network devices, endpoints, and applications, traditional Security Information and Event Management (SIEM) systems often struggle to detect sophisticated and novel attacks. In this context, advanced machine learning (ML) and deep learning (DL) techniques, particularly Long Short-Term Memory (LSTM) networks, have emerged as effective solutions to enhance threat detection and response in SIEM systems. By leveraging LSTM's ability to capture temporal dependencies in log data, these models provide a powerful means to detect anomalies and malicious activities that would otherwise go unnoticed using conventional methods.

SIEM systems aggregate, store, and analyze security-related event data from a variety of sources to detect suspicious activities in real-time. The traditional approach in SIEM solutions, which uses predefined rules and signatures, is no longer sufficient for identifying novel attack vectors such as zero-day exploits or advanced persistent threats (APTs). These limitations have driven the adoption of machine learning techniques to improve the identification of potential threats. LSTM networks, a type of recurrent neural network (RNN), have gained attention for their capacity to handle sequential data, such as security logs, and capture long-term dependencies in these sequences. By automatically learning the patterns of normal behavior and detecting deviations indicative of malicious activity, LSTM-based models have shown significant promise in strengthening SIEM systems and reducing false positives.

The deployment of LSTM-based models in SIEM systems faces several challenges. One of the key hurdles is the availability of high-quality, labeled datasets for model training. Security logs often contain sensitive and confidential information, which limits their accessibility for research and development purposes. collecting large, labeled datasets for training deep learning models can be time-consuming, expensive, and prone to bias. This scarcity of labeled data is a major bottleneck in the development of AI-driven SIEM solutions, as models rely on large, diverse datasets to achieve high accuracy in threat detection. In light of this challenge, there is growing interest in privacy-preserving data synthesis techniques, which can generate realistic synthetic datasets without compromising privacy or violating regulatory requirements.