**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Cyber Threat Detection Based on Deep Learning Through AI

***P. Guru Maheshwar Reddy[1], S. Suleman[2], G. Sainath[3], E. V Kalyan Goud[4], S. Muzamil[5]***

[1,2,3,4,5] Students of Dept of CSE, Santhiram Engineering College, Nandyal, Nandyal

**ABSTRACT**

One of the foremost challenges in cyber security is the provision of an automatic effective cyber-threat detection technique. During this paper, we have a tendency to gift an AI technique for cyber-threat detection, primarily based on artificial neural networks. The planned approach transforms the large number of collected security events into individual event profiles and uses deep learning-based detection techniques to improve cyber threat detection. For this task, we have developed an AISIEM system that supports a combination of event identification for knowledge pre-processing and completely different artificial neural network techniques with FCNNs, CNNs, and LSTMs. The system focuses on discriminating between true positive and false positive alerts, therefore serving security analysts to quickly answer cyber threats. All experiments on this examine are completed the usage of the author's benchmark dataset [2] (NSLKDD and CICIDS2017) and datasets accrued with inside the actual world. To judge the performance comparison with existing strategies, we tend conducted experiments using the 5-standard machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study make sure that our projected methods are capable of being used as learning-based models for network intrusion-detection, and show that though it's used within the real world, the performance outperforms the standard machine-learning methods

Keywords: security, intrusion detection, network security, artificial intelligence, deep neural networks.

## I. Introduction

Network intrusion detection refers to the problem of monitoring and differentiating such network flows and activities from the normal expected behaviour of network which can adversely impact the security of information systems. The search of reliable solutions by Governments and organizations to protect their information assets from unauthorized disclosures and illegal accesses has brought intrusion detection and prevention at the forefront of information security landscape. Traditionally, two primary systems have been used to detect cyber-threats and network intrusions. An intrusion prevention system (IPS) is installed in the enterprise network and can primarily use signature-based methods to examine network protocols and flows. Generate appropriate attack alerts called security events and report them to another system. SIEM (Security Information and Event Management) specializes in accumulating and handling IPS alerts. From diverse protection operations solutions. SIEM is not a rare place, but the most reliable answer for reading accumulated protection activity and logs . In addition, security analysts seek to detect malicious behaviour by investigating suspicious alerts based on policies and thresholds, analysing correlations between events, and applying attack knowledge.

## II. Literature Review

This phase discusses preceding studies for deep learning-primarily based totally intrusion detection and real-international protection occasion analysis. Many research in cyber security have centred on AI-primarily based totally intrusion detection in current years, and diverse AI and system learning-primarily based totally strategies were proposed to enhance the capacity of cyber risk detection. Even a thought at research done good size defects the usage of AI and system learning-primarily based totally strategies, they're none theless constrained to unique check datasets inclusive of NSLKDD. Other studies researches, on the alternative hand, have used real-international protection occasions and logs. This research is extra just like ours in phrases of addressing the aforementioned challenges.

**DETECTION OF INTRUSION BASED ON DEEP LEARNING**

Naseer et al. [1] proposed, implemented, and educated intrusion detection fashions using diverse deep neural community architectures including CNNs, Auto encoders, and RNNs. These fashions have been educated at the NSLKDD education dataset and examined at the NSLKDD take a look at datasets. On take a look at datasets, DCNN and LSTM fashions achieved with eighty-five and 89 in step with cent accuracy, respectively.

B. Zhang et al. [2] categorised community intrusion detection techniques into types: direct techniques that use a unmarried set of rules and aggregate techniques that integrate numerous techniques. A new detection version primarily based totally on a directed acyclic graph (DAG) and a notion rule base became proposed via way of means of the author (BRB). The effects confirmed that, while the usage of the KDD ninety-nine dataset, the DAGBRB aggregate version outperformed traditional detection fashions.

Wang et al. [3] proposed an intrusion detection gadget (HAST-IDS) primarily based totally on hierarchical spatial and temporal capabilities that routinely examine community site visitors 'capabilities. The important concept is that deep CNNs are used to examine the spatial capabilities of community site visitors, after which LSTM networks are used to examine the temporal capabilities. Datasets DARPA and ISCX finished the experiments.

Vinayakumar et al. [15] created a hybrid intrusion detection gadget which can examine community and host-degree activities. It used a disbursed deep gaining knowledge of version with DNN to procedure and examine very massive quantities of facts in real-time. The DNN version turned into selected after very well evaluating its overall performance to that of classical device gaining knowledge of classifiers on diverse benchmark IDS datasets together with NSLKDD and UNSW-NB15.

Liao et al. [39] added a brand-new method for modelling programme behaviour in intrusion detection-associated device calls primarily based totally at the kNN classifier approach and TF-IDF. The costs of device calls are utilised in [29] to give an explanation for programme interest the use of the k-NN classifier. Text category algorithms including TF-IDF are used to transform every device name fact to a vector and evaluate the similarity of programme device name activities. The TF-IDF-primarily based totally k-NN classifier appears to be nicely suitable to the vicinity of intrusion detection with inside the discipline of malware detection, consistent with the authors.

## III. PROPOSED WORK

The following are the principle contributions of our work: Our proposed device targets to transform a big quantity of safety occasions to character occasion profiles for processing very big-scale records. We created a generalizable safety occasion evaluation approach with the aid of using studying every day and hazard styles from a big quantity of accrued records and thinking of the frequency with which they occur. In this study, we in particular advice a technique for characterising records units the use of base points for the duration of the records pre-processing step. This approach has the capacity to seriously lessen the dimensionality space, that is regularly the principle assignment related to conventional records mining strategies in log evaluation.

Unlike conventional sequence-primarily based totally sample approaches, our occasion profiling technique for making use of synthetic intelligence strategies offers featured enter facts to hire diverse deep-getting to know strategies. As a result, due to the fact our approach allows stepped forward type for proper signals while as compared to standard machine-getting to know methods, it could extensively lessen the range of signals almost supplied to analysts.

To decide applicability, we use actual IPS safety activities from a actual safety operations centre (SOC) to assess our gadget and validate its effectiveness the usage of overall performance metrics inclusive of accuracy, actual wonderful rate (TPR), and false-wonderful rate (FPR), and the F-measure. Furthermore, to examine overall performance with present methods, we ran experiments with 5 traditional machine-mastering methods (SVM, kNN, RF, NB, and DT). In addition, we compare our technique with the aid of using making use of it to 2 benchmark datasets (NSLKDD and CICIDS2017), which can be broadly utilized in community intrusion detection research.

The TF-IDF technique is used on this examine to decompose a massive wide variety of accumulating activities into person occasion prevalence profiles. We additionally generate occasion profiles via way of means of calculating the similarity price among every TF-IDF occasion set and the assigned base points. The generated occasion profiles are fed into the enter layer of the FCNN, CNN, and LSTM fashions walking in AI-SIEM. As a result, we intend to illustrate the applicability of our device for protecting IT structures towards cyber threats via way of means of the use of famous benchmark datasets and actual datasets amassed from running IPS.

## IV. IMPLEMENTATION

We will in brief talk the heritage statistics for our examine on this section. We start via way of means of presenting an outline of the IDS/IPS and SIEM, observed via way of means of an advent to deep studying techniques. Finally, we describe our proposed AI-SIEM system's large information platform.

### *A. Intrusion Detection Systems (IDS/IPS) and SIEM*

#### *1) IDS / IPS*

An intrusion detection device (IDS) video display units' community pastime and reviews on any safety violations observed [6]. An intrusion prevention gadget (IPS), not like an IDS, can block a detected community2) SIEM

SIEM is a key thing of organization networks and protection infrastructure that specializes in organization facts technology (IT) protection and gives a entire image of protection management. In general, SIEM collects applicable records generated in an employer from diverse sources, permitting cyber threats to be detected via way of means of matching patterns [17], [18], [19SIEM structures allow the mixing and complete evaluation of protection indicators and logs amassed from community protection structures (inclusive of firewalls and IDS / IPS).Analysts use pre-described protection regulations and thresholds to locate cyber-assaults while analysing IDS/IPS indicators (protection activities) in SIEM. Furthermore, to locate consolidated malicious behaviour, they examine correlations among protection activities and applicable conditions primarily based totally on formerly diagnosed cyber hazard patterns.

Security occasions are constantly generated with the aid of using numerous kinds of community protection connection with the aid of using final the port or losing the packets. Because of the ever-increasing nature of records and the internet, intrusion detection systems (IPS) have emerged as an important gadget for maximum sorts of corporations or industries. Nonetheless, shrewdcommunityassaultslive on in modern-day networks, and an IPS gadget's

capacity to stumble on and reply to community intrusions is limited [15]. This is due to the fact they more often than not hire much less successful signature-primarily based totally detection strategies as opposed to anomaly detection strategies.

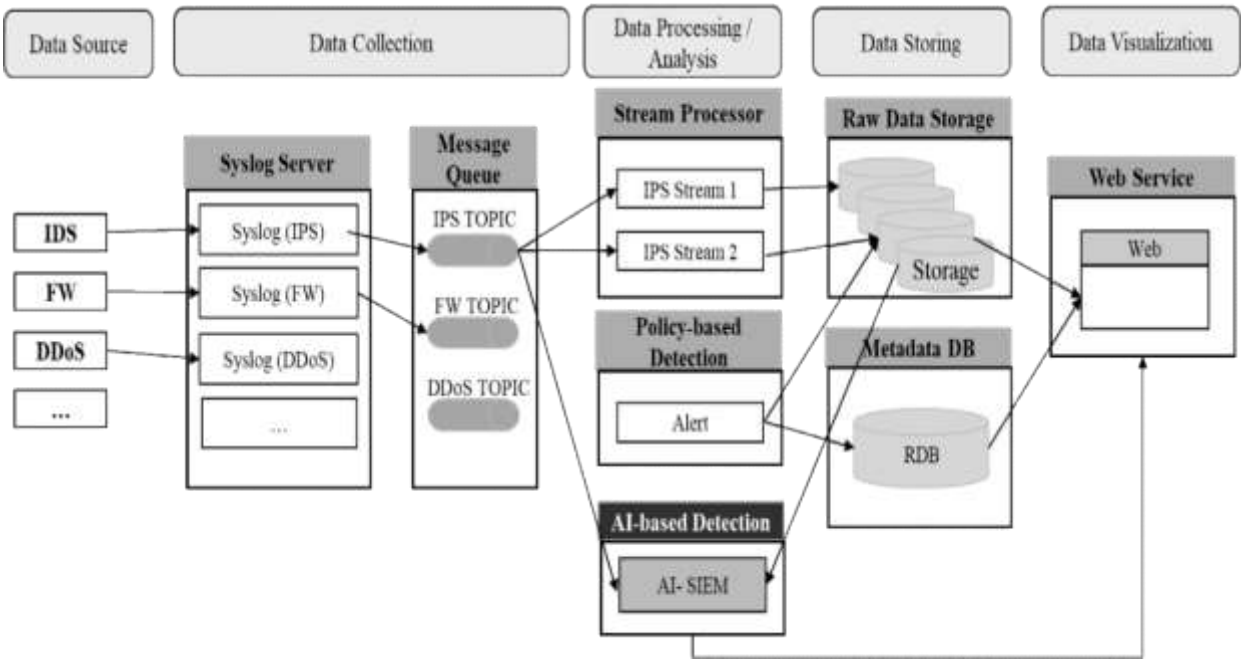


FIGURE 1. The architecture of our big data platform for AI-based SIEM

**BIG DATA PLATFORM**

A large statistics platform is usually used to accumulate statistics on protection occasions from IPS and to maintain protection logs for lengthy intervals. The large statistics platform also can be tailor-made to examine statistics and stumble on cyber threats quickly [35], [36]. This is due to the fact historic statistics accumulated at the platform over lengthy intervals can useful resource with inside the research and reaction to cyber threats. We created a scalable large statistics platform primarily based totally on disbursed computing technology for collecting, processing, storing, correlating, and analysing protection occasion logs.

## VI. RESULTS

This section reports the results of experiments performed on two benchmark datasets and two actual datasets collected. First, I will explain the test environment using the test bed. The following are the indicators of the experiment. We will continue to present SVD and traditional machine learning methods for various comparisons to assess performance. Subsection E describes the experimental results and finally shows the system implemented in the proposed way**.**

We have installed a test bed specially designed for performance evaluation. The big data platform and AISIEM system make up this testbed. In addition, SOC collected actual IPS data over several months.

| | | Accuracy | | | |
|---|---|---|---|---|---|
| | | NSLKDD | CICIDS2017 | ESX-1 | ESX-2 |
| Conventional Machine Learning | SVM | 0.897 | 0.968 | 0.901 | 0.867 |
| | k-NN | 0.909 | 0.978 | 0.905 | 0.858 |
| | Random Forest | 0.930 | 0.979 | 0.900 | 0.858 |
| | Naive Bayes | 0.698 | 0.621 | 0.692 | 0.616 |
| | Decision Tree | 0.919 | 0.979 | 0.900 | 0.858 |
| Our Proposed Method | EP-FCNN | 0.958 | 0.995 | 0.933 | 0.947 |
| | EP-CNN | 0.952 | 0.988 | 0.952 | 0.936 |
| | EP-LSTM | 0.950 | 0.986 | 0.923 | 0.926 |

TABLE 1 Results of accuracy tests for our suggested and several traditional machine-learning techniques.

## VII. CONCLUSION

In this project, we've got proposed the AI-SIEM machine the use of occasion profiles and synthetic neural networks. The novelty of our paintings lies in condensing very massive-scale facts into occasion profiles and the use of the deep gaining knowledge of-primarily based totally detection strategies for greater cyber-danger detection ability. The AI-SIEM machine permits the safety analysts to cope with massive safety indicators directly and correctly with the aid of using evaluating long time safety facts. By decreasing fake wonderful indicators, it may additionally assist the safety analysts to hastily reply to cyber threats dispersed throughout a massive range of safety events. For the assessment of overall performance, we done a overall performance evaluation the use of benchmark datasets (NSLKDD, CICIDS2017) and datasets accrued within side the actual world. First, primarily based totally at the evaluation test with different strategies, the use of widely recognized benchmark datasets, we confirmed that our mechanisms may be implemented as one of the gaining knowledge areas of-primarily based totally fashions for community intrusion detection. Second, thru the assessment the use of actual datasets, we supplied promising consequences that our generation additionally outperformed traditional device gaining knowledge of strategies in phrases of correct classifications.

**References**

[1]. S. Naseer, Y.Saleem, S. Khalid, M. K. Basher, J. Han, M. M. Iqbal,K. Han, "Enhanced Network Anomaly Detection Based on Deep Neural Networks," IEEE Access, vol. 6, pp. 48231-48246, 2018.

[2]. B. Zhang, G. Hu, Z. Zhou, Y. Zhang, P. Qiao, L. Chang, "Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base", ETRI Journal, vol. 39, no. 4, pp. 592-604, Aug. 2017

[3]. W. Wang, Y. Sheng and J. Wang, "HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection," IEEE Access, vol. 6, no. 99, pp. 1792-1806, 2018.

[4]. M. K. Hussein, N. Bin Zainal and A. N. Jaber, "Data security analysis for DDoS Defense of cloud based networks," 2015 IEEE Student Conference on Research and Development (Scored), KualaLumpur, 2015, pp. 305-310.