

AI-Powered Security Information and Event Management: A Review of Deep Learning Approaches for Modern Cyber Security

Hithendra B M¹, Khushi Kashinath², Rahul T R³, Rishab Urs A⁴ and Pooja M R⁵

¹⁻⁵Vidyavardhaka College of Engineering, Mysuru, India

Email: hithendra.bm09@gmail.com, {khushikashinath03, rahul.agumbeth, reach.rishaburs}@gmail.com, pooja.mr@vvce.ac.in

Abstract— Cyber threats are increasing with rapidly emerging digital infrastructure. In that scenario, even a reasonable pace cannot be maintained by traditional SIEM systems. Deep learning technologies such as Variational Autoencoders and Graph Neural Networks have to be embedded into SIEM based systems so that anomaly activity might be detected and zero-day threats also identified with accurate and timely generation of the alerts. We critically compare recent advances in deep learning-based SIEM with their more practical applications in detecting complex cyberattacks and focus on how AI improves the effectiveness of the efficacy of reports from SIEMs and technical challenges associated with those methodologies.

Index Terms— Deep Learning, Variational Autoencoder (VAE), Graph Neural Network (GNN), Anomaly Detection, Cybersecurity and Security Information and Event Management (SIEM).

I. INTRODUCTION

This has led to exponential rising figures as well as complexity of cyber threats in recent years by the speed at which businesses across the world are going digital. The growing number of sophisticated threats and complex, dynamic threats have now become too difficult for the traditional SIEM systems, which rely significantly on preset rules and signature-based techniques to detect.

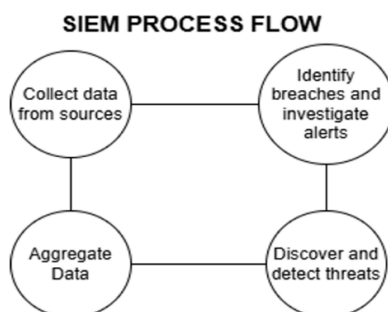


Figure 1. SIEM Process Flow

Improving the SIEM system based on AI development, especially deep learning models, may help them find more resourceful ways of reducing the impact of security threats. It explores a few specific deep learning models and their applicability in the context of the SIEM framework: Graph Neural Networks (GNNs) and Variational Autoencoders (VAEs). These would take its effectiveness to the next level in threat detection and sometimes even in anomaly detection.

II. BACKGROUND

A. Contemporary Significance of SIEM in Cyber Security

Advanced Persistent Threats, insider threats, and zero-day vulnerabilities represent some of the attacks that an enterprise has to worry about as part of today's threat landscape. Too new and strong for old perimeter-based protections, SIEM becomes almost essential for real-time monitoring and security in a multitude of network scenarios.

SIEM enhances the overall security posture of an organization by gathering all log data from various sources, such as firewalls, IDS, or network devices. Thus, it offers central visibility and also helps the analyst respond speedily to incidents.

B. Challenges: Limitations of Traditional SIEM Solutions

Besides that, traditional SIEM systems have several good numbers of disadvantages:

Data Volume

The quantity of log data in the organization's new, modern network may be too huge for the SIEM to handle, making it potentially poor real-time processing and storage.

False Positives

Conventional rule-based SIEM systems produce hundreds of thousands of false positive alarms that drown the security analyst, offering a window of opportunity whereby a real threat may slip past this observer.

Lacked Flexibility

The signature-based method has trouble detecting emerging threats or complicated attack patterns that don't follow set rules, for example APTs and zero-day exploits.

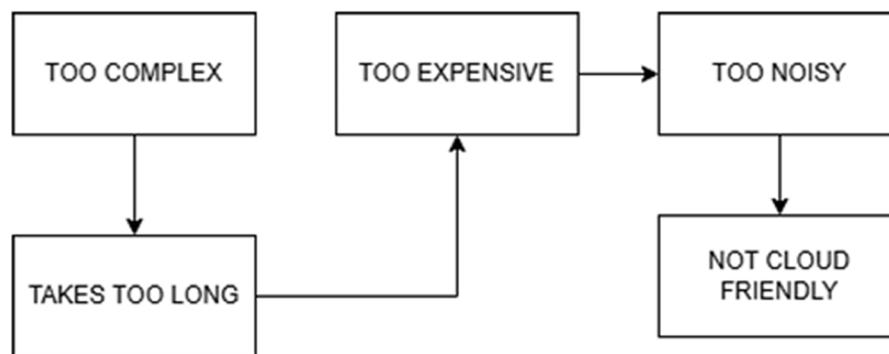


Figure 2. Limitations of Traditional SIEM

C. Motivation: AI and Its Role in Improving SIEM Capabilities

Due to such artificial intelligence potential, the overall SIEM may undergo the revolution since AI has allowed for sophisticated threat detection pattern recognition, anomaly detection and predictive analysis. It goes without saying that big dataset analyzers through machine learning could identify anomalies in real-time without using predefined rules while learning patterns of normal behavior. Graph Neural Networks along with Variational Autoencoders are deep learning models who have peculiar benefits.

- VAEs detect well unknown threats, so-called "zero day" attacks, since they can find patterns that are different from normal behavior.
- GNNs can reproduce the linkages between network entities, including devices, users and events, in order to monitor complex attack patterns, such as lateral network movement.

SIEM improves adaptability and accuracy through the use of AI, while saving the security analyst from unnecessary workload in threat identification and prioritization of critical alerts.

D. Purpose and Scope of This Review

Purpose The review will be on the insertion of the AI methods in the SIEM system, namely VAE and GNN. Essentially, the study aims to enlighten readers on whether such models are fit for current needs in cybersecurity based on how it addresses the limitations of the SIEM as it is standing nowadays. We would take into consideration real world applications, contrast approaches, and address issues and future research directions.

III. LITERATURE REVIEW

Ban Tao. [1] develops in-depth analysis of the methods, approaches, and issues which occurred in the fight against alert fatigue of the SIEM systems. IDS is an important detection tool of unauthorized activities; it can either be HIDS or NIDS. Most cases tend to have it as an integrated view of network security by running them together, but the nature of it suffers due to a significant challenge because of an overwhelming degree of false positives that lead to alert fatigue. Samuel Ndichu. [1] Mention a few of the limitations of the traditional SIEM solutions, for example, their failure to handle huge volumes of alerts that are typically produced by security appliances. They lay down the foundation of the requirement for executing advanced filtering and correlation practices to extend the capabilities of incident response in SOCs. Numerous filtering techniques are proposed for managing alert fatigue. Daisuke Inoue [1] proposed correlation-based techniques, which group correlated alerts to eliminate redundancy, though the techniques are different for various data streams. Isolation forests and SVMs based on ML are known to work effectively in filtering out uninteresting events, as shown by Daisuke Inoue. [1]. These can potentially rank alerts based on their severity so that SOC analysts would only work on critical events. Takeshi Takahashi [1] points out prioritization on the basis of alert risk models; these decide what impact the alerts generated by the IDS would have and prioritize those based on the same. Feature engineering has primarily been centered around standardizing the log formats from any security appliances. It reveals the problem of dealing with diffused alert logs and provides approaches to normalize approaches in order to identify and analyze threats more sensitively. Accepted progressions, such as SOAR (Security Orchestration, Automation, and Response), can be progressed with acceptability in minimizing alert fatigue. Investigating SOAR/SIEM integration is proposed to make the work of SOC's efficient for an automated response to incident. Data visualization tools, like the Augmented Tile Graph that would be proposed in this research, will now provide critical insights about alert patterns and relationships. Incidents can thus be analyzed faster and with greater accuracy. Visualization techniques allow security analysts to develop trends and identify anomalies with the intent of avoiding the overshooting of time reviewing logs manually. Indeed, significant challenges still exist in correlating alerts between sources due to differences in detection mechanisms and formatting. The paper highlights AI, machine learning, and advanced visualization for the design of next-generation SIEM systems that can combat alert fatigue. Overall, the result of this work opens avenues for future work toward enhancing operations in the SOC through the development of methodologies for automated incident response approaches as well as enhanced threat detection techniques.

Pulyala, S. R. [2] analyzed the future of security information and event management (SIEM) in the context of machine learning-driven cybersecurity landscapes. This paper introduces traditional challenges of SIEM which face the awareness of delayed threat detection and high false positives, and presents machine learning as an approach toward more accurate threat detection, faster response, and more resilient systems. The methodology applied is that of supervised ML for known threats and unsupervised learning on unknown threats to be able to identify anomalies. Such advanced techniques that are applied include NLP and behavioral analytics, to which log data is analyzed with in order to find patterns.

It is, therefore, an ML-based SIEM system that will allow proactive risk management since predictive analytics have threats predicted before happening. **Results** The results confirm the idea that ML-based SIEM improves real-time threat detection and reduces false positives while enhancing general risk management. Although the technology is advanced, the proper remedies to issues like data privacy, algorithm bias, and lack of skill in machine learning would be given by proper data governance, constant training, and collaboration of industries, as the author mentions.

Pulyala, S.R. From Detection to Prediction: AI-Powered SIEM for Proactive Threat Hunting and Risk Mitigation, by [3], is a comprehensive review on the evolution of the Security Information and Event Management (SIEM) system through the infusion of AI.

The paper discusses the flaws in traditional SIEMs based mainly on detection and back-looking findings. Most of the SIEMs were reactive in nature, thereby leading to alert fatigue. He further adds that these AI-based SIEM solutions help in better threat detection accuracy and predictive capabilities where the proactive threat hunting and risk management can be conducted. According to the report, several AI methods were applied in SIEM,

such as machine learning algorithms on anomaly and pattern analysis and for the automated synthesis of alerts from disparate sources. With these methodologies in place, the system is optimally enabled to differentiate between false positives and actual threats and therefore reduces some of the burden to SOC. For example, data enrichment and context-dependent analysis bring forth deeper insight into security incidents and therefore enhance decision-making processes involved. Putting everything in one place, Pulyala concludes that AI has a transformative capability for upgrading management of cybersecurity through SIEM systems toward more efficient and proactive management.

Pulyala, S.R., Jangampet, V.D., and Desetty, A.G. [4] carry out an exhaustive review on combining ML with SIEM systems in order to enhance the detection and prioritization of threats. Traditional SIEM's incapability to handle large volumes and high false positive rates in risks. The problem can be solved by using ML by allowing real-time threat detection, minimizing false positives, and optimization of response times. Their approach uses a combination of supervised and unsupervised algorithms, which they believe improves threat detection and response, and they are using deep learning and NLP to handle big data and intricate patterns of threats. It makes use of User and Entity Behavior Analytics (UEBA) to monitor anomalies and for the detection of possible security threats.

Research outputs also explain benefits of Machine Learning (ML) augmented SIEM, such as enhanced anomaly detection, predictive analytics for predictive defense, and enhanced risk assessment and prioritization that form a much more robust foundation for cybersecurity. While doing so, authors identify the challenges that include biasness, issues of privacy toward data and the lack of ML-related expertise.

The authors believe that such challenges can be overcome through inducing robust data governance and encouraging responsible AI culture.

Gattani, T.S., and Kumar, S.A. [5] presented a detailed discourse of AI in the context of cybersecurity referred alongside as critical limitations, techniques, and benefits in applications.

The main limits of the applications of AI identified are feature extraction challenges, dependency on domain knowledge, and the inaccuracies of classification that deteriorate model performance and robustness.

Their method encompasses critical stages of data collection, feature design, model training, and classification with supporting modules in the form of dynamic connectors and network emulators to enhance testing and data flow. Results of the study showed AI's capabilities to transform cybersecurity through threat detection improvement, efficiency in terms of incident response, and networking, besides providing predictive insights for proactive defense.

Explainable AI is used to build in transparency and trust. In summary, the study focuses on the importance of AI in creating a strong cybersecurity framework and underscores the need for interdisciplinarity to solve the ethical and pragmatic issues.

IV. PROBLEM STATEMENT

Traditional SIEM systems fail to identify and heal sophisticated cyber attacks, such as zero-day exploits and multi-stage attacks. Signature-based detection techniques are their greatest weakness, severely limiting flexibility and the ability to identify new or complex attack patterns effectively. An AI-based SIEM approach that incorporates improved pattern recognition and anomaly detection will prove better than traditional approaches in terms of identifying complex threats.

V. OBJECTIVE

As such, this research focuses on investigating and evaluating the usage of deep learning models, particularly VAEs and GNNs for an SIEM system in order to enhance pattern recognition and anomaly detection and even zero day threat identification. Ultimately, the paper should strive to propose a framework that exploits AI capabilities in proposing a more intelligent and responsive SIEM solution.

VI. METHODOLOGY

A. Preparation of Data Set

It is a well-known intrusion dataset of network, representing various cyber risks-the UNSW-NB15. To attract attention to these anomalies and typical attack types, it pre-processes the UNSW-NB15 data set. The raw network packets of the UNSW-NB 15 dataset was created by the IXIA PerfectStorm tool in the Cyber Range Lab of UNSW Canberra for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviours.

The tcpdump tool was utilised to capture 100 GB of the raw traffic (e.g., Pcap files). This dataset has nine types of attacks, namely: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. The Argus, Bro IDS tools are used and twelve algorithms are developed to generate totally 49 features with the class label. These features are described in the UNSW-NB15_features.csv file.

Source

- <https://research.unsw.edu.au/projects/unsw-nb15-dataset>
- <https://www.kaggle.com/datasets/dhoogla/unsw-nb15-data>

Train Set

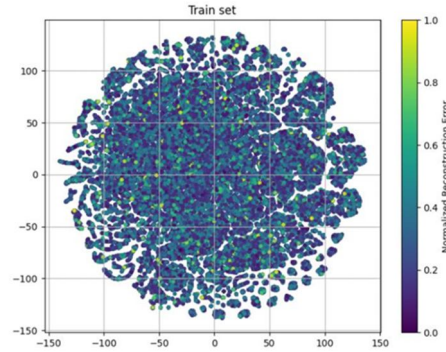


Figure 3. Visualization of Train Set

Test Set

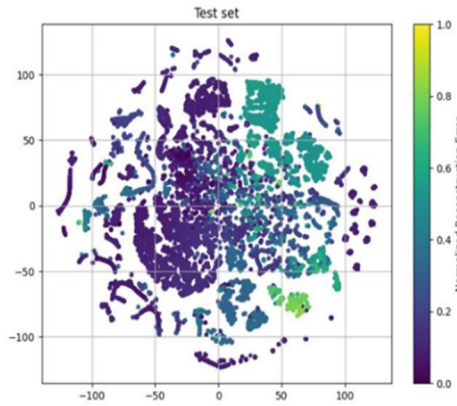


Figure 4. Visualization of Test Set

B. Design of the Model

- Anomaly detection using VAE VAEs will find possible zero-day threats by discovering anomalies in normal network activity patterns.

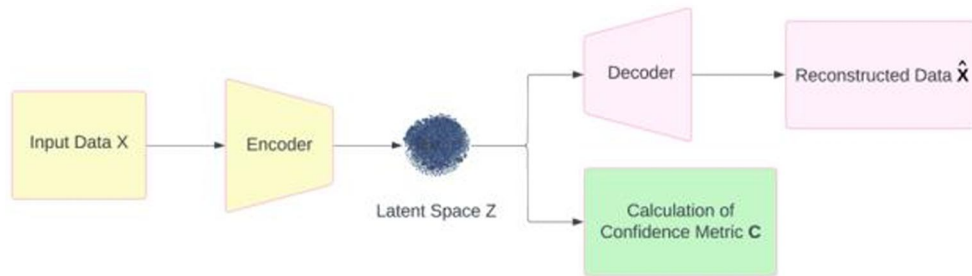


Figure 5. The main components of our VAE architecture

Optimal Threshold and Anomaly Predictions

In Fig 6. Observed to show the number of predicted output to the actual output.

- Precision: 0.9093
- Recall: 0.6444
- F1-score: 0.7543
- Accuracy: 0.7610.

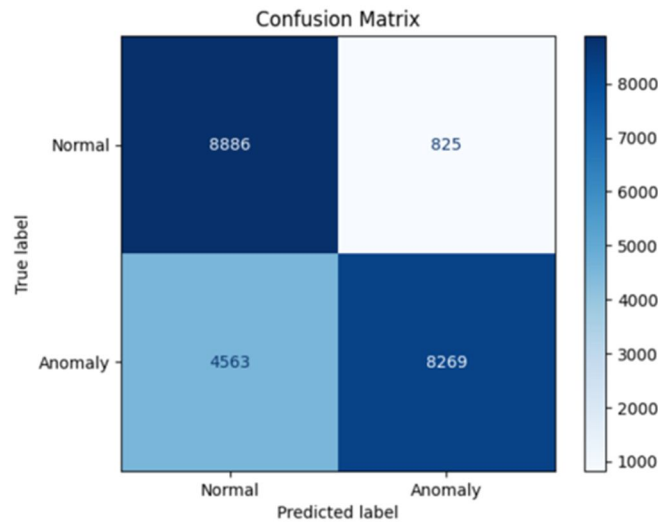


Figure 6. Confusion Matrix

- Relationship Mapping via GNN
GNNs can identify multi-step attack chains by modelling network elements (people, devices and IPs) along with their relations.
- In Fig 7. we can see the real-time processing and visualization Tools like Apache Kafka data stream-ing can be used, combined with ELK Stack (Elasticsearch, Logstash, and Kibana) for visualization of anomalies that are detected to include real-time monitoring in the system.

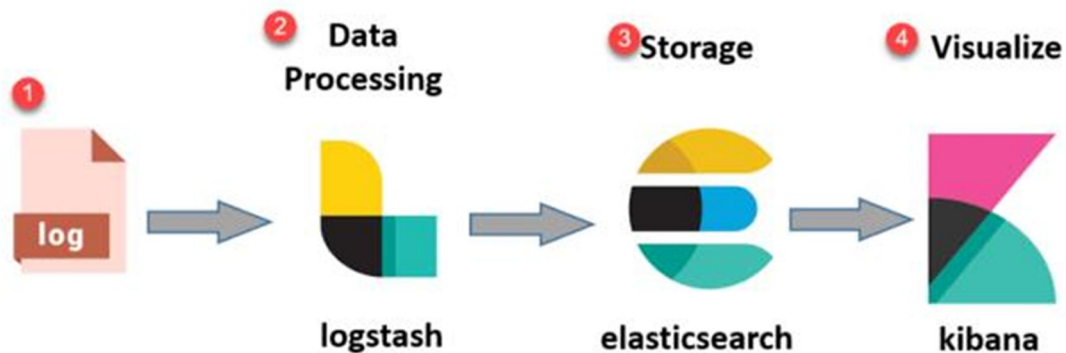


Figure 7. ELK Stack Architecture

VII. CONCLUSIONS

Indeed, the integration of AI, VAEs, and GNNs in SIEM represents a significant evolution in cybersecurity. Such models would therefore most probably allow organizations to keep up with the emerging cyber-threats due to such improved mapping connection strength and anomaly identification, among other features. Some of the important takeaways that might be inferred from such review include the advantages that come with the use of AI in SIEM; the enhancement in proactive cybersecurity posture.

Future work should start with efforts to improve the effectiveness of and extend the models for potential application in any other domain or field of study.

REFERENCES

- [1] Tao Ban, Takeshi Takahashi, Samuel Ndichu and Daisuke Inoue " Breaking Alert Fatigue: AI-Assisted SIEM Framework for Effective Incident Response." Appl. Sci. 2023, 13, 6610. <https://doi.org/10.3390/app13116610>.
- [2] Srinivas Reddy Pulyala , " The Future of SIEM in a Machine Learning - Driven Cybersecurity Landscape.
- [3] Srinivas Reddy Pulyala , " From Detection to Prediction : AI Powered SIEM for Proactive Threat Hunting and Risk Mitigation.
- [4] Srinivas Reddy Pulyala, Vinay Dutt Jangampet, Avinash Gupta Desetty " REVOLUTIONIZING SIEM WITH ML-DRIVEN RISK ASSESSMENT AND PRIORITIZATION".
- [5] Gattani Tanuj Subhash, Dr. S Anupama Kumar, " Artificial Intelligence Approaches to Uncover Cyber Security".
- [6] Chen, Y., Wang, Z., Li, J., "A Survey of Artificial Intelligence in Cybersecurity," IEEE Access, Vol. 11, pp. 3153-3170, 2023.
- [7] Zhang, Y., Liu, J., Chen, H., "Artificial Intelligence for Cybersecurity: A Review of Approaches, Challenges, and Open Research Problems," Frontiers in Cybersecurity, Vol. 3, Article 668686, 2022.
- [8] Hsiao, K., Yang, C., "Using Machine Learning for Intrusion Detection: A Review," Journal of Information Security and Applications, Vol. 70, Article 102428, 2022.
- [9] "Artificial Intelligence in Cybersecurity: Challenges, Advances, and Future Perspectives" by C. Liang, R. Zhang, and Y. Liu - Computers & Security, Vol. 118, pp. 102426, 2022.
- [10] "Machine Learning for Cyber Threat Intelligence: A Survey" by M. Ahmadi, M. Dehghantanha, K. Choo, and S. Singh - Journal of Network and Computer Applications, Vol. 188, pp. 103049, 2021.