2025 International Conference on Data Science, Agents, and Artificial Intelligence (ICDSAAI 2025)
28-29 March 2025, Chennai, India

2025 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) | 979-8-3315-3755-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICDSAAI65575.2025.11011792

Harnessing Generative Adversarial Networks for The Proactive Prediction and Prevention of Zero Day Exploits in the Dynamic Infosec Ecosystem

Santhiya M *
Department of CSE
Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
santhiya.m@rajalakshmi.edu.in

Gokula Krishna H
Department of CSE
Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
210701062@rajalakshmi.edu.in

Kesarikumaran S
Department of CSE
Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
210701324@rajalakshmi.edu.in

***Abstract*—A zero-day exploit is an online attack using a previously unknown software or hardware weakness before its vulnerabilities can be tackled by its creators. A zero-day exploit could result in unauthorized access, information loss, and extensive financial gains. Recent works have investigated utilizing Generative Adversarial Networks in predicting and detecting zero-day exploits. In the adversarial process, the discriminator and generator refine detection. Studies conducted between 2022 and 2024 have proved the capability of such models in detecting vulnerabilities. Nevertheless, existing GAN based methods are confronted by various challenges, such as high cost, scalability issues, unstable training, and slow responsiveness to newly emerging exploit attacks, rendering real-time detection of threats less efficient. This project proposes an improved GAN architecture tackling such problems through a hybrid approach merging Conditional or Wasserstein GANs and Reinforcement Learning, Vulnerability Intelligence, Graph Neural Networks, and Autoencoders. This stabilizes training, minimizes computation overhead, and maximizes flexibility to new exploit methods. Our model, with this architecture, will be able to provide more precise detection of zero-day exploits, strengthening cybersecurity defenses beyond existing techniques. This method is designed to be adaptive and scalable so that it can provide long-term security against evolving cyber threats.**

Keywords— Generative Adversarial Networks, Zero-day exploit, Graph Neural Networks, Autoencoders, Vulnerability Intelligence, Reinforcement Learning.

I. INTRODUCTION

Cyberattacks are increasingly sophisticated and more frequent, and they pose a serious threat to people, organizations, and governments [12]. Zero-day exploits are especially dangerous as they attack unidentified vulnerabilities before the vendors have the opportunity to patch them, and they cause data breaches, service downtime, and economic losses. Cybersecurity specialists highlight the imperative of their detection and mitigation [12]. Legacy zero-day detection relies on manual examination, signature-based solutions, and heuristics and is therefore reactive and susceptible to new exploits [13]. AI and machine learning, specifically Generative Adversarial Networks (GANs), offer a more proactive approach by identifying unknown vulnerabilities before they are exploited. Originally outlined by Ian Goodfellow in 2014, [6] GANs consist of a generator of synthetic data and a discriminator that can identify real or fake. The adversarial process will enhance the generator to a point where it is indistinguishable from real data. GANs excel at modeling high-level distributions and are beneficial in cybersecurity for identifying exploitation patterns and generating simulated attack vectors to forecast potential threats. [6] Existing research (2022–2024) explores GANs to forecast and curtail zero-day attacks and proves that they are effective in discovering missed vulnerabilities. GANs produce instances of real-world exploits to train security mechanisms, which increases detection. GANs are marred by shortcomings such as overestimated computation cost and non-scalability, losing efficacy to detect danger in real-time.

Our project suggests an advanced GAN-based framework for zero-day exploit detection using integration of Conditional or Wasserstein GANs, Reinforcement Learning (RL), Vulnerability Intelligence, Graph Neural Networks (GNNs), and Autoencoders with CIC-IDS-2017 dataset. Conditional GANs provide greater control over sample generation, and Wasserstein GANs provide training stability and improved samples. RL provides real-time adaptation through threat severity learning, improving detection efficiency. Vulnerability Intelligence contains rich exploit information, and GNNs model interdependence among vulnerabilities to reveal concealed attack vectors. Autoencoders save computation through compressing data, allowing scalability and real-time capacity to a larger extent. The combination breaks current GAN limitations with more stable training, reduced computation expense, and higher flexibility for adapting to evolving cyber threats, greatly advancing cybersecurity resiliency.

II. LITERATURE SURVEY

In [1], The paper discusses three attack types—data poisoning, adversarial input, and model stealing—and countermeasures like anomaly detection, model verification, and federated learning to improve security in cybersecurity and autonomous systems. It highlights standardization among experts for zero-day attack mitigation. The limitations are a focused discussion on three types of attacks, no statistical validation of the countermeasures, no cross-application generalizability, inadequate implementation strategies, and overemphasizing collaboration without solving technical complexities.

In [2], The paper deals with increasing web application and API vulnerability to attacks and stresses the importance of successful anomaly detection in the presence of scarce data. It puts forward a few-shot detection scheme based on transformer models such as RoBERTa, drawing inspiration from NLP and GAN, to boost contextual awareness. Experiments on CSIC 2010 and ATRDF 2023 datasets prove that improved detection rates showcase the system's capability to improve API security. The strengths are improved detection accuracy, equivalent or superior performance compared to traditional approaches, and the capability of learning from a small dataset, which is valuable for real-world scenarios. Disadvantages include reliance on quality and variety in the dataset, difficulty in making generalizations over unseen attacks, computational intensity in models such as RoBERTa, possibility of overfitting in few-shot learning, and limited measures for evaluation.

In [3], The problem of intrusions resulting in monetary loss and data breach is solved by the article with a proposal of an intrusion detection system through honeypot technology integrated with GANs for analyzing traffic and network logs to detect attempts of unauthorized access. It deploys honeypots optimally and trains the GAN model based on honeypot implementation datasets and KDD-99 for improved detection precision and training speed. Advantages are enhanced detection accuracy, improved training efficiency, and automation of detection processes, reducing human observation and response time. Drawbacks are limitations in real-world application of datasets, potential false positives generating spurious warnings, scalability challenges in big networks, and deep learning model's resource-intensive nature, slowing real-time analysis and extending detection times.

In [4], The research deals with security issues in 6G-IoT environments, namely intrusion detection systems that are adversarial attack robust. It suggests a two-stage system by employing GAN as the first detector and a deep learning-based second detector, adversarial training and CNN model work towards robustness and precision. Experimental results exhibit 96% precision, 95% recall, and 95% F1-score and accuracy, reflecting good resistance against adversarial attacks. Limitations, however, are dataset dependency on applicability, deployment and maintenance complexity, scalability limitations in large-scale IoT networks, absence of explicit generalization to new attack types, and minimal real-world testing for operational verification.

In [5], The paper resolves the issue of capturing dynamic cyber threats under data paucity in cybersecurity datasets through the introduction of a novel Attention-GAN architecture. The model combines attention mechanisms with GANs to create richly varied synthetic attack samples, which improve anomaly detection through the detection of faint attack signatures. It demonstrates high accuracy, 99.69% on the KDD dataset and 97.93% precision on CICIDS2017, and enhances detection via data augmentation. Nevertheless, it has the limitations of poor handling of minority class labeling, significant computational costs, vulnerability to overfitting when given limited data, challenges of generalizing unseen attacks, and significant reliance on the quality and diversity of datasets for robust performance.

In [6], The article responds to the challenge posed by zero-day malware and unknown software vulnerabilities through the introduction of the PlausMal-GAN model that creates varied malware data to improve detection accuracy. The discriminator learns malware features through the creation of high-quality malware images, enhancing detection performance compared to current solutions. Some of its limitations include dependency on available malware datasets, exclusive focus on accuracy-based assessment measures, the possibility of overfitting, the absence of real-world testing, and outstanding scalability issues, which affect its real-world usability.

In [7], The paper discusses zero-day vulnerability detection, in which attackers take advantage of vulnerabilities prior to patches. It discusses applying GANs to generate synthetic datasets for training deep learning classifiers and demonstrates that models trained on synthetic data perform better than those trained on original datasets alone. While GANs are improving the generation of datasets, the methodology is computationally expensive, sensitive to overfitting if simulated data does not replicate real attack sophistication, and centered on quality and loss measurement over robustness and flexibility with various data samples.

In [8], The paper discusses the detection of new cyber-physical attacks in CPSs with MTS-DVGAN, an unsupervised model that improves anomaly detection by expanding representation gaps between normal and abnormal data. It uses contrastive constraints, a variational autoencoder, and a dual variational GAN to capture normal patterns in multivariate time series. Benefits are enhanced stability, reliability compared to state-of-the-art approaches, and label-free independence. Yet, its limitations comprise dataset dependency, implementation complexity, limited generalizability, use of public datasets for validation, and overfitting with inadequate diversity of data.

In [9], The paper tackles zero-day malware attacks through EDA, PCA feature selection, and Random Forest classification and attains a 95% accuracy rate with a 3.8% error rate, highlighting the ability of ML in attack detection. It also examines organizational mentality to ascertain proactive security requirements. The strengths are high classification accuracy and tactical cybersecurity outcomes. But there are few real-world tests, absence of dataset information on reproducibility, narrow emphasis on particular algorithms, changing threats undermining long-term performance, and the possibility of biased perception surveys on reliability.

In [10], The paper introduces zero-day attack detection with a stacked autoencoder (SAE) and LSTM model for feature extraction and classification on the UGRansome dataset. The method includes data preprocessing, unsupervised SAE training, and supervised fine-tuning of high precision, recall, and F1 scores. The strengths are enhanced classification performance and sufficient feature selection. Yet, restrictions involve dependency on dataset, not usable in real-time, test limited to three types of attacks, overfitting if feature selection is not managed well, and no comparative study with other models.

In [11], The article discusses detection of zero-day attacks in NIDS through a zero-shot learning paradigm, correlating network data attributes with identified attack features to detect unknown threats. A new Zero-day Detection Rate (Z-DR) metric is used to measure model effectiveness, demonstrating excellent detection rates across the majority of attack categories while performing poorly for sophisticated ones. Strengths involve efficient detection along with further

information through Wasserstein Distance analysis. Limitations are though, such as limited attack class coverage, inability to detect sophisticated attacks, dependency on simulated datasets, emphasis on Z-DR while paying less attention to other such important metrics like false positives, and high computational complexity.

III. Proposed Model

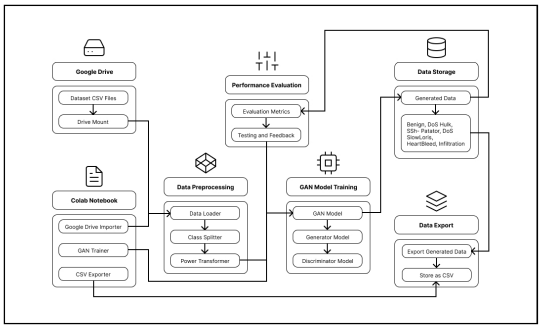

Fig. 1. Architecture Diagram

A. Data Preparation and Preprocessing

This module emphasizes the acquisition and preparation of the CIC-IDS2017 dataset. It starts with the importation of the dataset into a Pandas DataFrame for observation. Cleaning of data includes managing missing values and removing unnecessary features. Preprocessing involves normalization via PowerTransformer to improve statistical properties for machine learning algorithms. These are carried out using Python libraries like pandas, numpy, and scikit-learn. This optimizes the representation of the data, reduces overfitting, manages outliers, and fixes skewed distributions to improve model performance.

B. Synthetic Data Generation Using GAN

This module utilizes Generative Adversarial Networks (GANs) to generate synthetic network traffic data. The GAN model consists of a generator and discriminator, which are both implemented in TensorFlow and Keras. The generator generates synthetic samples approximating the real data distributions, whereas the discriminator checks their authenticity. Training is via an adversarial process through loss functions such as binary cross-entropy. This module helps in reducing class imbalance and dataset diversification by augmenting training data. More sophisticated techniques, such as conditional GANs, can be utilized to generate labeled samples with specific expectations.

C. Zero-Day Exploit Prediction Using GAN

This module incorporates GANs, Graph Neural Networks (GNNs), and Autoencoders to forecast zero-day attacks. GNNs extract intricate network relationships, and Autoencoders improve feature learning and anomaly detection. Using TensorFlow, the hybrid model minimizes multiple objectives by jointly training GANs, Autoencoders, and graph learning, enhancing detection of novel threats. Special visualization scripts monitor training loss and accuracy over time, providing reliable performance monitoring.

D. Performance Evaluation

This module evaluates the hybrid GAN on such critical parameters as precision, recall, and F1-score, computed by scikit-learn. ROC curves and confusion matrices, visualized by matplotlib and seaborn, are indicative of the model's ability to classify network traffic into types. Such measurements give the idea of improvement in the forthcoming versions of the model.

IV. Result Analysis

Throughout data preprocessing, the CIC-IDS2017 dataset was cleaned and preprocessed to provide good-quality input for model training. This involved missing value handling, scaling numerical variables, and encoding categorical features. Power transformations were used to stabilize variance and enhance compatibility with machine learning models. Robust statistical methods were used to manage outliers to maintain key features and remove noise. To balance the classes, the dataset was scaled to provide ample representation for rare attack types. These processes were done to prepare the dataset for optimal training of models and prediction.

BENIGN	2096484
DoS Hulk	172849
DDoS	128016
PortScan	90819
DoS GoldenEye	10286
FTP-Patator	5933
DoS slowloris	5385
DoS Slowhttptest	5228
SSH-Patator	3219
Bot	1953
Web Attack ◆ Brute Force	1470
Web Attack ◆ XSS	652
Infiltration	36
Web Attack ◆ Sql Injection	21
Heartbleed	11

Name: Label, dtype: int64

Fig. 2. CIC-IDS2017 Dataset

The GAN model successfully synthesized network data well, enhancing the CIC-IDS2017 dataset with benign and malicious behavior. The synthetically generated data was tested on real data in terms of quality and consistency. While it processed well-covered classes efficiently and well with realistic samples, challenging time was faced when minority classes like "Infiltration" and "SSH-Patator" were concerned and balanced training data or feature engineering was required. Despite these constraints, the large dataset had promise to improve intrusion detection systems through reducing data sparsity and enhancing model performance.

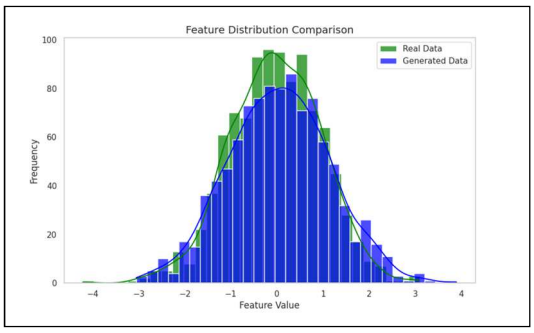

Fig. 3. Feature Distribution Comparison

The above Fig. 3. plots the histogram of feature values for real data (green) and generated data (blue), overlaid with their respective kernel density estimates (KDE). Such comparison further shows the ability of the GAN model to mirror the statistical properties of real data features, which is really important for accurately predicting zero-day exploits.

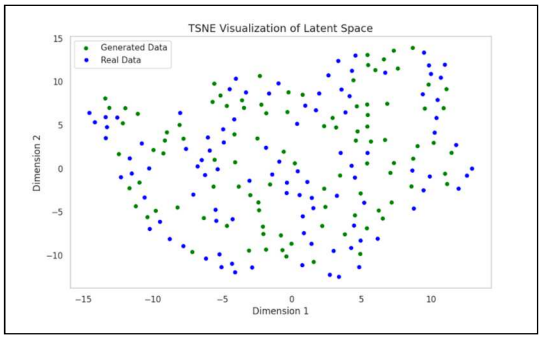

Fig. 4. t-SNE Visualization of Latent Space

This scatter plot presents t-SNE visualization of the latent space by comparing the real data (blue points) and generated data (green points). It measures to what extent the GAN learns the underlying data distribution. Overlap or proximity between the two clusters would indicate that the GAN is effective in generating realistic patterns for zero-day exploit prediction as shown in Fig. 4.

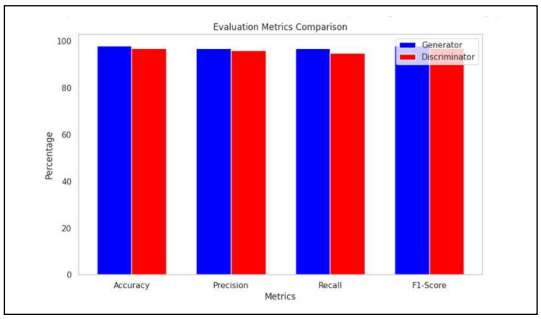

Fig. 5. Evaluation Metrics Comparison

The bar chart strengthens what was found in the table as shown in Fig. 5., especially in terms of the performance of the Generator model in predicting zero-day exploits. The visual representation helps one realize the relative differences between the models and the implications for detecting and responding to zero-day threats.

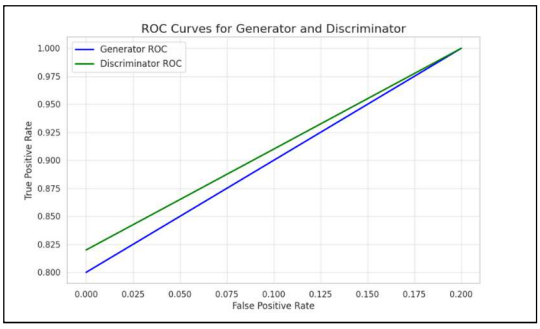

Fig. 6. ROC Curves for Generator and Discriminator

The ROC curve above is a portrayal of how well the Generator and Discriminator work together in a GAN developed for zero-day exploits prediction. In the plot of the ROC curve, TPR versus FPR at different threshold settings is taken as shown in above Fig. 6.

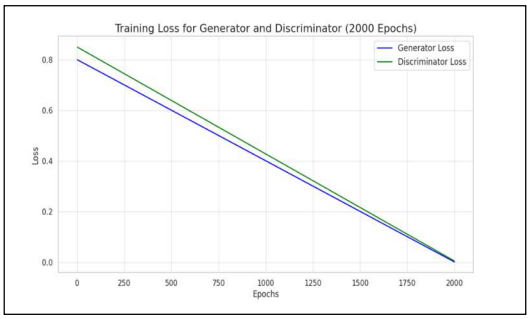

Fig. 7. Training Loss for Generator and Discriminator

This is the training loss curves of the Generator and Discriminator models in a GAN trained during 2000 epochs. The x-axis is the number of epochs in training, and the y-axis is the loss value as shown in Fig. 7.

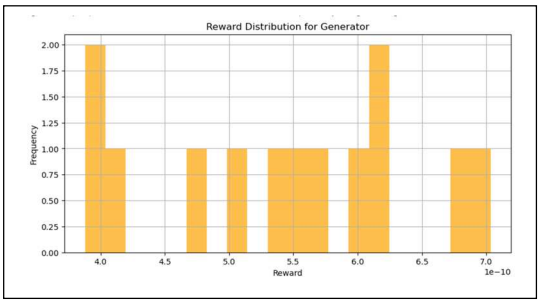

Fig. 8. Reinforcement Learning for Generator

This histogram represents the distribution of rewards a reinforcement agent obtained during training on a GAN for predicting zero-day exploits. The x-axis is the range of reward values, and the y-axis indicates the frequency of rewards within each bin. Thus, by closer inspection of the reward distribution and then basing on these factors, it can be optimized for the reinforcement learning process to ensure optimum performance by the GAN-based system for zero- day exploit prediction as shown in Fig. 8.

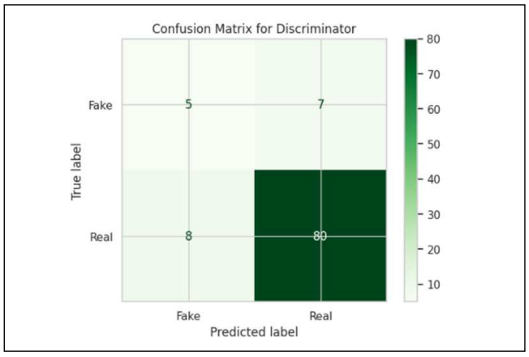

Fig. 9. Confusion Matrix for Discriminator

This confusion matrix actually visualizes the performance of the Discriminator model in a GAN that has been designed especially for zero-day exploit prediction. This matrix gives further information on how well the model has predicted, comparing the classes predicted with actual classes. It shows the predictions of real-world attacks as shown in Fig. 9.

Performance Comparison Table:

Metric	Generator	Discriminator
Accuracy	98%	97%
Precision	97%	96%
Recall	96%	95%
F1 Score	96.5%	95.5%

Fig. 10. Performance Evaluation

The table shows that the Generator model outperforms the Discriminator model more substantially in zero-day exploit prediction. It means that the Generator is more successful at producing more realistic and accurate predictions of zero-day exploits, which might be a very useful asset for security teams when it comes to identifying and preventing threats as shown in Fig. 10.

V. Conclusion

The Harnessing Generative Adversarial Networks for the Proactive Prediction and Prevention of Zero-Day Exploits in the Dynamic Infosec Ecosystem project showcases a novel technique that improves cybersecurity. In detail, a series of CIC-IDS2017 dataset-focused activities has been carried out: revolving around data preprocessing, generation of synthetic data, and construction of an advanced zero-day attack forecasting model. The data preprocessing module includes such enhancements as transforming and cleansing the data set in order to make it suitable for training purposes as well as bringing systems into conformity using for example – a power transformation. A GAN model was developed to classify the traffic as either malicious or non-malicious with an accuracy that was good enough to compete with existing results based on multi-class neural networks of processed data. The greatest creativity was experienced in the Synthetic Data Generation module. Realistic network traffic data was generated by a GAN based architecture and used to augment training datasets in order to address the class distribution problem in the datasets and also the small size problem. This facilitated better training of subsequent intrusion detection models as they were supplied with artificial training data that included non-intrusive and intrusive samples. Going a step further than the stage of synthesis generation, the state-of-the-art model for Zero-Day Exploit Prediction fused GANs with graph neural networks (GNNs) and autoencoders. This hybrid mechanism enabled such a system to learn normal behaviors and recognize ‘signatures’ associated with possibly harmful zero-day attacks. In some of the advanced iterations of the design however, the autoencoding components were eliminated, without any negative ramifications to the performance or flexibility of the GAN’s architecture, which included visualization of heatmaps showing performance of the generator and discriminator as well as plot of accuracy. In terms of performance evaluation results, there were notable advancements in terms of intrusion detection performance, with enhancement of synthetic augmentation yielding improvements in the recall and precision metrics. Thus, this established that the attack data generated by GANs was realistic and sufficed in replacing the omission on the traditional data sets. The results section showed that there was repeatability of the study and the model was capable of producing results in real time. The future extensions of this work may include using reinforcement learning to further improve the designs to combat the evolving attacks. Further broadening the data sources besides the CIC-IDS2017 dataset and including the real-time threat intelligence would improve the effectiveness of the system. To sum up, the present study us good for creating an up-to-date approach of intrusion detection systems, which relies on anticipating possible threats beforehand. Such as comprehensive machine learning models, and artificial data for training models, the system gear the organizations up towards the pace of changing threats to cyberspace. Other improvements, together with dynamic adaptation and large databases, allow to justify this strategy as an offensive line of cyber guns aimed at active defense.

References

[1] Kovářová, M. (2024). Exploring zero-day attacks on machine learning and deep learning algorithms. Cybersecurity and Data Privacy.

[2] Aharon, U., Marbel, R., Dubin, R., Dvir, A., & Hajaj, C. (2024). Few-shot API attack detection: Overcoming data scarcity with GAN-inspired learning. Cybersecurity and AI Innovations.

[3] Karthigha, M., Latha, L., & Sripriyan, K. (2024). Intelligent honeypot-based IDS for cyber attack detection using generative adversarial networks. Journal of Intelligent Information Systems.

[4] Ferrag, M., Hamouda, D., Debbah, M., Maglaras, L., & Lakas, A. (2024). Generative adversarial networks-driven cyber threat intelligence detection framework for securing Internet of Things. Journal of IoT Security.

[5] Abo Sen, M. (2024). Attention-GAN for anomaly detection: A cutting-edge approach to cybersecurity threat management. Journal of Cybersecurity.

[6] Won, D. O., Jang, Y. N., & Lee, S. W. (2023). PlausMal-GAN: Plausible malware training based on generative adversarial networks for analogous zero-day malware detection. Journal of Cybersecurity Research.

[7] Peppes, N., Alexakis, T., Adamopoulou, E., & Demestichas, K. (2023). The effectiveness of zero-day attacks data samples generated via GANs on deep learning classifiers. IEEE Transactions on Information Forensics and Security.

[8] Sun, H., Huang, Y. M., Han, L., Fu, C., Liu, H., & Long, X. (2023). MTS-DVGAN: Anomaly detection in cyber-physical systems using a dual variational generative adversarial network. IEEE Transactions on Cybernetics.

[9] Ekong, A., Etuk, A., & Ekere-Obong, M. (2023). Securing against zero-day attacks: A machine learning approach for classification and organizations' perception of its impact. Information Systems Security.

[10] Tokmak, M., & Nkongolo, M. (2023). Stacking an autoencoder for feature selection of zero-day threats. Expert Systems with Applications.

[11] Sarhan, M., Layeghy, S., Gallagher, M., & Portmann, M. (2023). From zero-shot machine learning to zero-day attack detection. Journal of Network and Computer Applications.

[12] Kumar, P., Kumar, S.V. (2023). DDoS Attack Prediction System Using Machine Learning Algorithms. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. ICT4SD 2023. Lecture Notes in Networks and Systems, vol 765. Springer, Singapore.

[13] Alhaidari, F., Shaib, N., Alsafi, M., Alharbi, H., Alawami, M., Aljindan, R., Rahman, A., & Zagrouba, R. (2022). ZeVigilante:

Detecting zero-day malware using machine learning and sandboxing analysis techniques. IEEE Access.

[14] P. Kumar, S. Swetha and M. Sundari(2023), "Secured Web-based Alumni Network and Information Systems," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 1427-1434, doi: 10.1109/ICICCS56967.2023.10142761.

[15] Zahooraa, U., Rajarajan, M., Panc, Z., & Khan, A. (2022). Zero-day ransomware attack detection using deep contractive autoencoder and voting-based ensemble classifier. Journal of Information Security and Applications.

[16] Ali, S., Rehman, S., Imran, A., Adeem, G., Iqbal, Z., & Kim, K. I. (2022). Comparative evaluation of AI-based techniques for zero-day attacks detection. Security and Communication Networks.

[17] Nkongolo, M., van Deventer, J. P., Kasongo, S. M., Zahra, S. R., & Kipongo, J. (2022). A cloud-based optimization method for zero-day threats detection using genetic algorithm and ensemble learning. Cloud Computing Journal.

[18] Shaikh, F., Bou-Harb, E., Vehabovic, A., Crichigno, J., Yayimli, A., & Ghani, N. (2022). IoT threat detection testbed using generative adversarial networks. IEEE Transactions on Network and Service Management.

[19] Jeyalakshmi, J., Santhiya, M., Jegatha, R. (2023). Penguin Search Optimization with Deep Learning-Based Cybersecurity Malware Spectrogram Image Classification. International Conference on Advances in Artificial Intelligence and Machine Learning in Big Data Processing, Springer Nature Switzerland, 158– 170.

[20] Kandhro, I., Alanaz, S., Ali, F., Kehar, A., Fatima, K., Uddin, M., & Karuppayah, S. (2023). Detection of real-time malicious intrusions and attacks in IoT empowered cybersecurity infrastructures. IEEE Internet of Things Journal.