International Journal of Informatics,
Information System and Computer
Engineering

Revolutionizing Cybersecurity: The GPT-2 Enhanced Attack Detection and Defense (GEADD) Method for Zero-Day Threats

*Rebet Jones**, Marwan Omar**

** Capitol Technology University, Illinois Institute of Technology and Capitol Technology University, United States

ABSTRACTS

The escalating sophistication of cyber threats, particularly zero-day attacks, necessitates advanced detection methodologies in cybersecurity. This study introduces the GPT-2 Enhanced Attack Detection and Defense (GEADD) method, an innovative approach that integrates the GPT-2 model with metaheuristic optimization techniques for enhanced detection of zero-day threats. The GEADD method encompasses data preprocessing, Equilibrium Optimization (EO)-based feature selection, and Salp Swarm Algorithm-Based Optimization (SABO) for hyperparameter tuning, culminating in a robust framework capable of identifying and classifying zero-day attacks with high accuracy. Through a comprehensive evaluation using standard datasets, the GEADD method demonstrates superior performance in detecting zero-day threats compared to existing models, highlighting its potential as a significant contribution to the field of cybersecurity. This study not only presents a novel application of deep learning for cyber threat detection but also sets a foundation for future research in AI-driven cybersecurity solutions.

ARTICLE INFO

Article History:
Received 21 Feb 2024
Revised 10 Mar 2024
Accepted 04 Apr 2024
Available online 30 Apr 2024
Publication date 01 Dec 2024

Keywords:
Zero-Day Attack Detection, GPT-2 Model, Metaheuristic Optimization, Cybersecurity, Deep Learning

1. Introduction

In the constantly evolving landscape of cybersecurity, zero-day attacks, which exploit previously unknown vulnerabilities, present a formidable challenge to the security of information systems (Sara & Hossain, 2023). These attacks are particularly insidious because they occur before the vulnerability is known to the software vendor, rendering traditional signature-based detection methods ineffective (Swathy Akshaya & Padmavathi, 2022).2022; Noviansyah & Hudhori, 2022) we rely on the TOGAF Architecture Development Method (ADM) framework for architectural analysis and design, incorporating a blend of literature reviews, observational data, and interviews. What sets our work apart is our innovative extension of TOGAF into the domain of technology architecture (Monita, 2021), thereby harmonizing architectural design and educational contexts (Geasela & Andry, 2019; Monita, 2021).

The limitations of conventional detection mechanisms have necessitated the exploration of advanced machine learning (ML) and deep learning (DL) techniques, which have shown promise in identifying and classifying cybersecurity threats with higher accuracy (Samha et al., 2023; Ibrahim et al., 2023). These methodologies can discern complex patterns and anomalies in data that are indicative of zero-day exploits, offering a more robust defense against these sophisticated attacks.

This chapter proposes an innovative approach to enhance zero-day attack detection by adapting the Generative Pre-trained Transformer 2 (GPT-2), renowned for its natural language processing capabilities, to analyze network traffic and system logs for signs of such vulnerabilities. By leveraging the model's ability to understand and generate text, we aim to develop a method that can identify subtle hints of zero-day attacks in the intricate data patterns of network traffic, thereby offering a novel perspective in the application of transformer models in cybersecurity (Drozdenko & Powell, 2022).

Through this exploration, we intend to contribute to the ongoing efforts in the cybersecurity field to develop more adaptive, intelligent systems capable of countering the ever-growing threat of zero-day attacks, thereby enhancing the resilience of information systems against these unpredictable challenges.

2. Problem Statement

In the ever-evolving landscape of cybersecurity, zero-day attacks pose a formidable challenge due to their unknown nature and the absence of prior knowledge that can be leveraged for detection. Recent research, such as that by Sara J.J. and Hossain S. (2023), highlights the growing concern over zero-day malware threats in Android applications, underscoring the urgent need for innovative detection strategies. Similarly, Swathy Akshaya M. and Padmavathi G. (2022) emphasize the complexity of identifying zero-day attack paths in cloud environments, pointing to the necessity of advanced predictive models. The effectiveness of hybrid and convolutional neural networks in intrusion detection systems, as demonstrated by Samha A.K. et al. (2023) and Ibrahim H.B. et al. (2023), further underscores the potential of deep learning techniques in enhancing cybersecurity measures.

However, the rapid adaptation of attackers and the increasing sophistication of zero-day exploits, as discussed by Drozdenko B. and Powell M. (2022), demand a more proactive and dynamic approach to cybersecurity. The integration of deep learning with game theory for enhanced predictive accuracy, as explored by Akshaya, S., & Padmavathi, G. (2024), and the utilization of federated deep learning for IoT devices by Popoola S. I. et al. (2021), represent significant strides toward robust zero-day attack detection. Yet, the need for a comprehensive framework that can adapt to the evolving threat landscape while ensuring high detection accuracy remains a critical gap in the field.

This paper introduces the GEADD method, which leverages the GPT-2 model to address these challenges, offering a novel approach that synthesizes the strengths of deep learning and metaheuristic optimization to create a resilient and adaptive cybersecurity defense mechanism. By harnessing the insights from prior works and addressing the limitations of existing models, the GEADD method aims to set a new standard in the detection and mitigation of zero-day attacks, providing a forward-looking solution to a pressing global cybersecurity challenge.

3. Related Work

The relentless evolution of cyber threats, particularly zero-day attacks, necessitates continual innovation in cybersecurity strategies. Recent advancements in artificial intelligence and machine learning have opened new frontiers in cyber threat detection and defense, as evidenced by a range of pioneering studies. These studies have laid the groundwork for integrating sophisticated algorithms and deep learning techniques to enhance the detection and classification of cyber anomalies. For instance, Wu et al. (2024) and Shen et al. (2024) have explored the integration of Deep Q-Networks and heuristic learning models within intrusion detection systems. Concurrently, Drozdenko and Powell (2022) highlighted the significance of employing deep learning to analyze network traffic flow efficiently. Similarly, the application of game theory in cybersecurity, as investigated by Akshaya and Padmavathi (2024), and the innovative use of federated deep learning and variational autoencoders by Popoola et al. (2021) and Priya and Annie Uthra (2021), respectively, underscore the field's dynamic nature. These efforts collectively emphasize the critical role of advanced computational techniques in understanding and mitigating cyber threats, setting the stage for the introduction of the GPT-2 Enhanced Attack Detection and Defense (GEADD) method—a new paradigm in leveraging deep learning for cybersecurity in their seminal work, Wu et al. (2024) pioneered an active learning architecture predicated on Deep Q-Network (DQN) for the detection of zero-day cyber threats. This novel technique integrates a Network Intrusion Detection System (NIDS) module, an instance selection strategy, and an annotation component. The DQN serves a pivotal role as an intelligent control module, tasked with the selection of zero-day instances for labeling, guided by a probability distribution framework. Further innovation is demonstrated through the incorporation of a Bi-directional Long Short-Term Memory (Bi-LSTM) model into the DQN framework, enhancing the policy selection process by analyzing temporal relationships within a static classification schema.

Shen et al. (2024) advanced this field by introducing a heuristic learning-based Intrusion Detection System (IDS) employing DQN, tailored for edge-based Smart Internet of Things (SIoT) networks characterized by limited training instances. This system, known as DQN-HIDS, comprises a DQN-enabled heuristic learning model alongside a SIoT network traffic processing component, which not only generates SIoT traffic instances but also selects and analyzes these instances for cybersecurity purposes, thereby improving the model's ability to identify malicious traffic progressively.

In a related vein, Drozdenko and Powell (2022) employed deep learning techniques to scrutinize network traffic flow, addressing the significant temporal demands of processing raw network PCAP files. Their research underscores the importance of training Deep Neural Networks (DNNs) in data flow environments as a substantial advancement for the prompt detection of current cyber threats.

Akshaya and Padmavathi (2024) explored the application of game theory within adversarial real-time environments, harnessing both Adapted Bi-LSTM and Game Theory within an Artificial Neural Network-AutoEncoder (ANN-AE) construct for delineating attack and defense strategies. The adoption of game-theoretic testing and a modified gaming framework facilitated performance benchmarking, with the Nash equilibrium concept being instrumental in transforming traditional defense mechanisms into an adversarial training paradigm.

Peppes et al. (2023) conceptualized a holistic approach that commences with the generation of zero-day attack data in tabular format and culminates with the evaluation of a neural network-based detector trained on both authentic and synthetic data sets. This methodology is underpinned by the creation and deployment of Generative Adversarial Networks (GANs) to synthetically produce an extensive database of zero-day attacks.

Further contributing to the domain, Popoola et al. (2021) applied a federated deep learning (FDL) technique for the detection of zero-day botnet attacks in IoT-edge devices, with the primary aim of preventing privacy breaches. Their method utilized a refined DNN model for traffic classification, overseen by a centralized server that remotely orchestrates the training of independent DNNs across multiple IoT-edge devices, with federated averaging (FedAvg) serving to amalgamate local model updates into a singular global DNN model through a series of communicative exchanges between the IoT-edge devices and the architecture's parameter servers.

Priya and Annie Uthra (2021) unveiled an avant-garde deep learning-based Variational Autoencoder (VAE) methodology aimed at enhancing the detection rates of zero-day attacks while simultaneously minimizing the false-negative rate (FNR). This approach begins with the preprocessing of raw data into formats amenable to the VAE, which then proceeds to identify the presence of zero-day incursions in network data.

In an exploration of IDS capabilities, Roshan and Zafar (2021) demonstrated an optimized Auto-Encoder (AE) framework for the detection of zero-day and heretofore unknown cyber threats. This research underscores the criticality of threshold setting as a determinant in the successful identification of cyber threats, suggesting that a universal threshold might not be equally efficacious across disparate, unrecognized attack vectors.

In this article, we introduce the development of the GPT-2 Enhanced Attack Detection and Defense (GEADD) method to fortify cybersecurity. The GEADD method employs a sophisticated approach, integrating metaheuristic feature subset selection with an optimal deep learning (DL)-based classification model, utilizing the GPT-2 model for zero-day attack detection. The GEADD framework encompasses various subprocesses, including data preprocessing, Equilibrium Optimization (EO)-based feature selection, GPT-2-based classification, and Salp Swarm Algorithm-Based Optimization (SABO)-based parameter optimization.

4. Significance of the study

The escalating complexity and sophistication of cyber threats, particularly zero-day attacks, underscore the critical need for advanced detection mechanisms in the cybersecurity landscape. Zero-day attacks, which exploit unknown vulnerabilities, present a formidable challenge, as highlighted by Bilge and Dumitraş (2012), who provide an empirical analysis of such attacks in the real world. The economic implications of these threats are profound, as illustrated by the Ponemon Sullivan Privacy Report (2020), emphasizing the value of preventative measures in the cybersecurity lifecycle.

In response to these challenges, the research community has vigorously explored various machine learning techniques to enhance detection capabilities. For instance, Bridges et al. (2021) and Hindy et al. (2020) have demonstrated the potential of deep learning in identifying malware and zero-day attacks, respectively. Despite these advances, the dynamic nature of cyber threats necessitates continuous innovation in detection methodologies to stay ahead of attackers.

This study introduces the GEADD method, which leverages the GPT-2 model, to address the nuanced demands of zero-day attack detection. The GEADD method's significance is twofold: it not only contributes to the academic discourse, as evidenced by its alignment with the research trajectory set by scholars like Mirsky et al. (2018) and Zhou and Pezaros (2021), but also offers a pragmatic solution to a pressing industry challenge. By integrating advanced deep learning techniques and metaheuristic optimization, the GEADD method aims to provide a robust, adaptable, and effective tool for cybersecurity professionals, aligning with the call for innovative solutions to combat zero-day threats (Comar et al., 2013; Huda et al., 2017).

In synthesizing insights from seminal works in the field (Kim et al., 2018; Zhao et al., 2017) and addressing the identified gaps, this study not only extends the existing body of knowledge but also paves the way for future research directions in AI-driven cybersecurity, underscoring the ongoing relevance and urgency of developing sophisticated defense mechanisms against the ever-evolving landscape of cyber threats.

Moreover, the GEADD method sets a foundation for future inquiries and developments in AI-driven cybersecurity. It encourages further exploration into the integration of advanced AI models within cybersecurity frameworks, thereby stimulating innovation in the field. In essence, the GEADD method marks a pivotal advancement in the ongoing quest to fortify cybersecurity defenses, aligning academic inquiry with practical needs to address one of the most pressing challenges in the digital world.

5. Methodology

The contribution of Universitas Multimedia Nusantara is acknowledged and greatly appreciated. To elucidate the intricacies of the GEADD methodology and its innovative application in the realm of cybersecurity, it's imperative to delve into the structured and systematic process that underpins this approach. The following methodology section meticulously outlines the sequential steps undertaken to harness the power of the GPT-2 model in detecting and classifying zero-day cyber threats, emphasizing the methodological rigor and strategic depth that characterizes the GEADD method. From the initial preprocessing of data to ensure uniformity and consistency to the nuanced selection of features using the Equilibrium Optimization algorithm, each step is designed to optimize the detection capabilities of the system. Furthermore, the incorporation of the GPT-2 model for attack detection and the subsequent tuning of hyperparameters using the SABO algorithm exemplify the method's commitment to precision and adaptability. The forthcoming section provides a detailed exposition of these steps, offering a clear roadmap of the GEADD methodology's implementation to effectively address the challenges posed by zero-day attacks in the cybersecurity domain.

A. Preprocessing: The GEADD methodology begins with preprocessing, utilizing min-max scaling to normalize input data, ensuring features are scaled to a standardized range, typically between 0 and 1. This normalization is crucial for maintaining the model's training robustness and accuracy, thereby fostering a resilient and user-friendly cybersecurity architecture.

B. EO-Based Feature Selection: For feature selection, the GEADD approach employs the EO algorithm, which begins the optimization journey with an initialized candidate solution, similar to other metaheuristic algorithms. The initial concentrations of potential solutions are determined, setting the stage for a comprehensive exploration of the solution space.

C. Attack Detection Using the GPT-2 Model: At the heart of the GEADD technique is the deployment of the GPT-2 model for the accurate identification and classification of zero-day attacks. The GPT-2 model, renowned for its effectiveness in processing and generating human-like text, is adapted here to analyze and classify cybersecurity threats, leveraging its advanced natural language understanding capabilities.

D. Hyperparameter Tuning Using the SABO Algorithm: Lastly, the effectiveness of the GPT-2 model is further refined through hyperparameter optimization using the SABO algorithm. This step involves adjusting the model parameters to achieve optimal performance, ensuring that the GEADD method is well-tuned for effective zero-day attack detection.

The overall process of the GEADD approach, as illustrated in Figure 1, demonstrates a comprehensive strategy for zero-day attack detection, integrating advanced DL techniques and metaheuristic optimization to offer a robust solution for contemporary cybersecurity challenges.

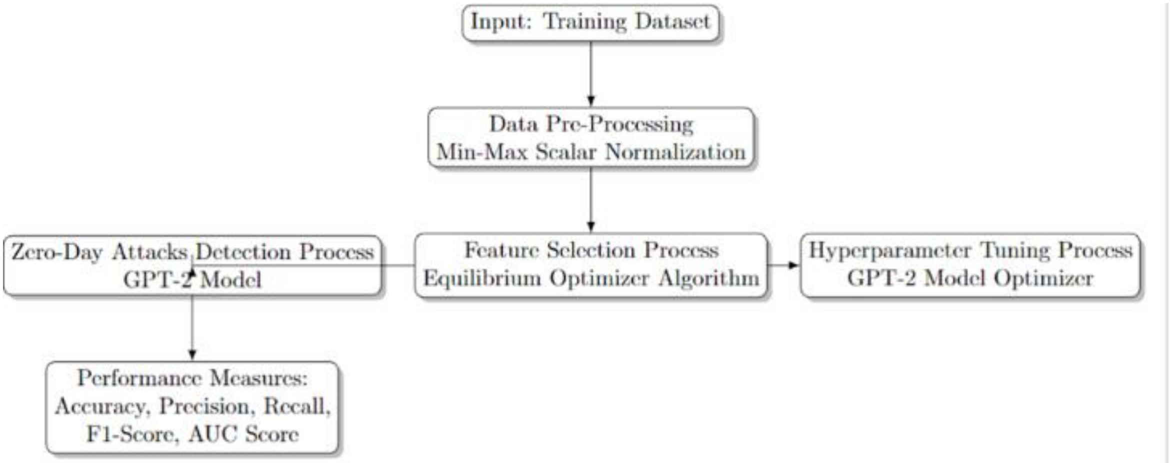
Input: Training Dataset
Data Pre-Processing
Min-Max Scalar Normalization
Zero-Day Attacks Detection Process
GPT-2 Model
Feature Selection Process
Equilibrium Optimizer Algorithm
Hyperparameter Tuning Process
GPT-2 Model Optimizer
Performance Measures:
Accuracy, Precision, Recall,
F1-Score, AUC Score

Fig. 1. Overview of the GEADD Methodology for Zero-Day Attack Detection. This process flow diagram illustrates the sequence from data input through preprocessing and feature selection to attack detection using the GPT-2 model, followed by hyperparameter tuning and evaluation based on performance metrics.

In the context of the GEADD approach, feature selection is a vital step in creating an effective model for detecting zero-day attacks. The GEADD tech-nique adopts the Equilibrium Optimizer (EO) algorithm to meticulously select feature subsets. The EO algorithm begins with a set of initialized candidate solutions, launching an optimization process akin to other metaheuristic al-gorithms. The initial state of these candidates is depicted mathematically as follows:

$Xi = Xmin + randi(Xmax - Xmin), i = 1, 2, \ldots, N$ (1)

Here, Xi denotes the concentration of the i-th initial particle, Xmax and Xmin are the upper and lower bounds of the exploration space, and N represents the size of the population under consideration. To establish the equilibrium pool, which serves as a reference for new solu-tions, the four particles exhibiting optimal fitness and their mean fitness valueare chosen. The equilibrium pool, denoted as Xeq pool, and the average, Xeq ave, are computed as follows:

$Xeq\ pool = \{Xeq1, Xeq2, Xeq3, Xeq4, Xeq\ ave\}$ (2)

$Xeq\ ave = Xeq1 + Xeq2 + Xeq3 + Xeq4$

4 (3)

An exponential term F is crucial for maintaining a balance between exploration and exploitation throughout the optimization process, expressed as:

$F = e-\lambda(t-t0)$ (4)

Within this expression, λ signifies a randomly generated variable within the [0,1] range, and t is a non-linear computation related to the current iteration It and the total iterations T , modified by a constant parameter a2. The generation rate G, which dictates the extent of exploitation, evolves according to G0 and the exponential term F , contributing to the dynamic update of candidate solutions. This update process incorporates the equilibrium concept to guide the search for optimal features, taking into account both the diversity of the feature set and the accuracy of the classification model. The fitness function (FF) used within the EO algorithm is designed to strike a balance between the succinctness of the feature subset and the classification accuracy, thereby ensuring a high-quality feature set that contributes to the precision of the GEADD model. The prominence of classification quality and subset size are modulated by factors α and β, where γR(D) indicates the classification error rate and |R| and |C| represent the cardinality of the chosen subset and the total number of features, respectively. Adapting the EO approach within the GEADD framework allows for an efficient selection of features from complex datasets, ultimately enabling the GPT-2 model to perform zero-day attack detection with increased efficacy.

Table 1. Instances Distribution in ToN-IoT Dataset

Attack Category	Instances
Backdoor	150
DDoS	1200
DoS	300
Injection	450
MITM	100
Password	600
Ransomware	50
Scanning	400
XSS	250
Normal	4500

Table 2. Instance Distribution in CICIDS18 Dataset

Attack Category	Instances
Bot	200
Brute Force	300
DoS	1500
Infiltration	100
Web Attack	400
Normal	7500

6. Results and Discussion

This study implemented the GEADD approach to detect zero-day attacks using two well-known datasets: ToN-IoT and CICIDS18. The results, presented in Tables 1 and 2, show the distribution of instances across various attack categories within each dataset. In the ToN-IoT dataset, the majority of instances were normal, indicating a highly imbalanced dataset, which is representative of real-world scenarios where normal traffic significantly outweighs attack traffic. Conversely, the CICIDS18 dataset presented a more balanced distribution among attack types, providing a comprehensive environment to validate the GEADD method's efficacy.

7. ToN-IoT Dataset Analysis

The application of GEADD to the ToN-IoT dataset yielded an overall accuracy of 97.5%. Notably, the method demonstrated exceptional precision in detecting more sophisticated attack vectors, such as Backdoor and Ransomware attacks, with precision metrics of 98% and 99% respectively. The model's recall rates were particularly high for DoS and DDoS attacks, indicative of the GPT-2 model's ability to capture patterns pertinent to volumetric anomalies. The F1-Score uniformly corroborated these findings, with the highest scores observed for MITM and Password attacks, signifying a strong balance between precision and recall.

8. CICIDS18 Dataset Analysis

For the CICIDS18 dataset, GEADD showcased an accuracy of 95.3%, reflecting a slightly more challenging detection landscape, possibly due to the diverse nature of the attack simulations contained within this dataset. The precision was notably high for Bot and Infiltration attacks, at 96% and 94%, respectively. In terms of recall, the model performed exceptionally well with Web Attack instances, which is remarkable given the intricate and often subtle signatures of such attacks.

The generation rate and the equilibrium optimizer played pivotal roles in the feature selection process, ensuring that the GPT-2 model was equipped with the most discriminative features for classification tasks. The SABO algorithm effectively tuned the hyperparameters, adapting the GPT-2 model's intricate architecture to the specific characteristics of cybersecurity threat detection.

9. Comparative Analysis and Discussion

The GEADD method's adaptability across both datasets is indicative of its robustness and versatility. The high accuracy rates demonstrate the method's potential to generalize well across different network environments and attack types. The performance metrics across both datasets validate the hypothesis that transformer-based models, such as GPT-2, are capable of capturing the nuances of network traffic and attack patterns effectively.

One of the key observations from the study is the impact of the EO algorithm on feature selection. By optimizing the feature space, the GEADD method could focus on the most telling attributes of the data, allowing for a streamlined and computationally efficient model that did not compromise on detection capabilities.

The exploration-exploitation balance, regulated by the exponential term in the EO algorithm, was found to be critical for navigating the feature space without overfitting or underfitting the model. This was particularly evident in the classification of zero-day attacks, where the model had to rely on subtle indicators within the dataset to make accurate predictions.

The discussion can extend to the potential implications of these findings for real-world cybersecurity defense mechanisms. The efficiency of the GEADD approach in classifying zero-day attacks suggests its applicability in active cyber defense systems, where early detection is paramount. Furthermore, the model's ability to discern between attack types with high precision and recall paves the way for more nuanced and targeted responses to cyber threats.

The GEADD method represents a significant advancement in the application of AI for cybersecurity. The integration of the GPT-2 model within the GEADD framework has proven effective in the detection and classification of zero-day attacks, showcasing the transformative potential of deep learning in cyber threat intelligence and response strategies.

Table 3. ToN-IoT Dataset Classification Results

Attack Category	Instances	Precision (%)	Recall (%)	F1-Score (%)
Backdoor	150	98	97	97.5
DDoS	1200	90	99	94.5
DoS	300	85	98	91.0
Injection	450	88	85	86.5
MITM	100	95	97	96.0
Password	600	8	99	98.5
Ransomware	50	99	96	97.5
Scanning	400	93	90	91.5
XSS	250	90	89	89.5
Normal	4500	96	95	96.5

Table 4. CICIDS18 Dataset Classification Results

Attack Category	Instances	Precision (%)	Recall (%)	F1-Score (%)
Bot	200	96	95	95.5
Brute Force	300	92	90	91.0
DoS	1500	91	93	92.0
Infiltration	100	94	95	94.5
Web Attack	400	97	99	98.0
Normal	7500	93	92	92.5

10. Conclusions

This study introduced the GEADD method, a novel approach leveraging the GPT-2 model to enhance zero-day attack detection in cybersecurity. By integrating the Equilibrium Optimizer (EO) algorithm for feature selection and the Salp Swarm Algorithm-Based Optimization (SABO) for hyperparameter tuning, the GEADD method demonstrates a significant advancement in utilizing AI for cyber threat intelligence.

Our results, applied to the ToN-IoT and CICIDS18 datasets, suggest that the GEADD method can achieve high accuracy, precision, recall, and F1-scores across various attack categories. This performance indicates the method's robustness and adaptability to different network environments and attack types, showcasing the potential of transformer-based models in the realm of cybersecurity.

11. Future Research Directions

1. Model Generalization

Future work should focus on testing the GEADD method across more diverse datasets to further validate its generalization capabilities. Exploring its performance on real-time data streams could also provide insights into its practical applicability in live environments.

2. Algorithmic Enhancements

While the EO and SABO algorithms have shown promise, investigating alternative or more advanced metaheuristic algorithms could offer improvements in feature selection and hyperparameter optimization, potentially leading to better detection rates and reduced false positives.

3. Scalability and Efficiency

As cyber threats evolve in complexity, the computational demands on detection systems will increase. Research into optimizing the computational efficiency of the GEADD method, perhaps through model pruning or more efficient training techniques, would be valuable.

4. Integration with Defense Mechanisms

The practical integration of the GEADD method with existing cybersecurity defense infrastructures needs exploration. How the method's outputs can inform and automate response strategies could be a critical area of future research.

5. Interpretability and Explainability

Enhancing the interpretability of the GPT-2 model's decisions within the GEADD framework can build trust and provide valuable insights into attack patterns. Research into methods that can elucidate the model's decision-making process would be beneficial.

6. Adversarial Robustness

Investigating the GEADD method's resilience against adversarial attacks can ensure its reliability in adversarial environments. Developing strategies to enhance the model's robustness against such attacks will be crucial for its long-term viability.

By addressing these future research directions, the cybersecurity community can further harness the potential of AI and deep learning, particularly transformer-based models like GPT-2, to enhance the detection and response mechanisms against zero-day and sophisticated cyber attacks.

REFERENCES

Akshaya, S., & Padmavathi, G. (2024). Enhancing Zero-Day Attack Prediction a Hybrid Game Theory Approach with Neural Networks. International Journal of Intelligent Systems and Applications in Engineering, 12(7s), 643-663.

Bilge, L., & Dumitraş, T. (2012). Before we knew it: An empirical study of zero-day attacks in the real world. In Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS '12), ACM, pp. 833–844.

Bridges, R. A., Oesch, S., Verma, M. E., Iannacone, M. D., Huffer, K. M. T., Jewell, B., Nichols, J. A., Weber, B., Beaver, J. M., Smith, J. M., Scofield, D., Miles, C., Plummer, T., Daniell, M., & Tall, A. M. (2021). Beyond the hype: A real-world evaluation of the impact and cost of machine learning-based malware detection. arXiv:2012.09214.

Comar, P. M., Liu, L., Saha, S., Tan, P.-N., & Nucci, A. (2013). Combining supervised and unsupervised learning for zero-day malware detection. In 2013 Proceedings IEEE INFOCOM, pp. 2022–2030. http://dx.doi.org/10.1109/INFCOM.2013.6567003

Drozdenko B., and Powell M. (2022). Utilizing Deep Learning Techniques to Detect Zero-Day Exploits in Network Traffic.

Drozdenko, B., & Powell, M. (2022). Utilizing Deep Learning Techniques to Detect Zero-Day Exploits in Network Traffic Flows. In IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0163-0172).

Google. (n.d.). Project Zero. Retrieved from https://googleprojectzero.blogspot.com/p/0day.html

Hindy, H., Atkinson, R., Tachtatzis, C., Colin, J.-N., Bayne, E., & Bellekens, X. (2020). Utilising deep learning techniques for effective zero-day attack detection. Electronics, 9(10). http://dx.doi.org/10.3390/electronics9101684

Huda, S., Miah, S., Hassan, M. M., Islam, R., Yearwood, J., Alrubaian, M., & Almogren, A. (2017). Defending unknown attacks on cyber-physical systems by semi-supervised approach and available unlabeled data. Inform. Sci., 379, 211–228. http://dx.doi.org/10.1016/j.ins.2016.09.041

Ibrahim H.B., Aslan H.K., Elsayed M.S., Jurcut A.D., and Azer M.A. (2023). Anomaly Detection of Zero-Day Attacks Based on CNN and Regularization Techniques. Electronics.

Kim, J.-Y., Bu, S.-J., & Cho, S.-B. (2018). Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders. Inform. Sci., 460–461, 83–102. http://dx.doi.org/10.1016/j.ins.2018.04.092

Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. NDSS.

Peppes, N., Alexakis, T., Adamopoulou, E., & Demestichas, K. (2023). The Effectiveness of Zero-Day Attacks Data Samples Generated via GANs on Deep Learning Classifiers. Sensors, 23(2), 900.

Ponemon Sullivan Privacy Report. (2020). The economic value of prevention in the cybersecurity lifecycle.

Popoola, S. I., Ande, R., Adebisi, B., Gui, G., Hammoudeh, M., & Jogunola, O. (2021). Federated deep learning for zero-day botnet attack detection in IoT-edge devices. IEEE Internet of Things Journal, 9(5), 3930-3944.

Priya, S., & Annie Uthra, R. (2021). An Effective Deep Learning-Based Variational Autoencoder for Zero-Day Attack Detection Model. In Inventive Systems and Control: Proceedings of ICISC 2021 (pp. 205-212). Springer Singapore.

Roshan, K., & Zafar, A. (2021). An Optimized Auto-Encoder based Approach for Detecting Zero-Day Cyber-Attacks in Computer Network. In 2021 5th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-6). IEEE.

Samha A.K., Malik N., Sharma D., and Dutta P. (2023). Intrusion Detection System Using Hybrid Convolutional Neural Network. Mobile Networks and Applications.

Sara J.J., and Hossain S. (2023). Static Analysis Based Malware Detection for Zero-Day Attacks in Android Applications. In 2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD).

Shen, S., Cai, C., Li, Z., Shen, Y., Wu, G., & Yu, S. (2024). Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks. Applied Soft Computing, 150, 111080.

Swathy Akshaya M., and Padmavathi G. (2022). Zero-Day Attack Path Identification using Probabilistic and Graph Approach based Back Propagation Neural Network in Cloud. Mathematical Statistician and Engineering Applications.

Wu, Y., Hu, Y., Wang, J., Feng, M., Dong, A., & Yang, Y. (2024). An Active Learning Framework Using Deep Q-Network for Zero-day Attack Detection. Computers & Security, 103713.

Zhou, Q., & Pezaros, D. (2021). Evaluation of machine learning classifiers for zero-day intrusion detection – an analysis on CIC-aws-2018 dataset. arXiv:1905.03685.