

INTELLIGENT CYBER THREAT PREDICTION USING ML AND GENERATIVE AI APPROACHES

^{#1}Anoosha Kaleru, Assistant Professor, Dept of CSE,

^{#2}Jyothi Macharla, Assistant Professor, Dept of CSE,

^{#3}Dr. B. Anantharam, Assistant Professor, Dept of CSE,

^{#4}N. Manikanta, B. Tech Student, Dept of CSE,

^{#5}M. Jithendrasai, B. Tech Student, Dept of CSE,

^{#1-5}Scient Institute Of Technology(Autonomous), Ibrahimpatnam, R.R. Dist, TG, India.

ABSTRACT: Cyberattacks are a threat to enterprises, financial institutions, and digital infrastructure due to the frequent and sophisticated nature of hackers. This research examines a variety of hacking prediction techniques, including traditional machine learning and contemporary generative AI models. Random forests, support vector machines, and decision trees were implemented to assess historical network data in order to identify malicious activity. While these systems are capable of recognizing established attack signatures, they may fail to recognize emerging threats. Transformer-based models and generative adversarial networks are two recent advancements in generative AI. These technologies allow computers to understand intricate behavioral patterns, simulate plausible attack scenarios, and predict new cyberthreats. Generative AI employs predictive models, adaptive learning, and comprehensive data analysis to identify anomalous activity, attack patterns, and preventative security measures. In terms of assault prediction system accuracy, scalability, and expertise, generative AI outperforms traditional machine learning, according to the study.

Keywords: Cyber Attack Prediction, Machine Learning, Generative Artificial Intelligence, Deep Learning, Cybersecurity, Threat Intelligence, Network Security, Anomaly Detection.

1. INTRODUCTION

As more businesses employ cloud computing, big data networks, and interconnected technology, cybersecurity becomes increasingly important. As internet-based services expand, cyberthreats become increasingly sophisticated. Hackers are always coming up with new ways to exploit weaknesses in systems, rendering traditional security measures useless. The ability to be prepared for hacking has gained popularity. As a result, businesses can promptly identify and address security concerns.

Attack time was first forecasted using outdated machine learning techniques that

looked at past data to find attack patterns. Random forests, logistic regression, decision trees, and support vector machines were employed for problem identification and network activity analysis. To assist security systems in identifying abnormalities, these models examined labeled datasets and attributes. Conventional machine learning was better at identifying dangers than rule-based systems since it relied on manually defined historical patterns and attributes, but it had trouble adjusting to new threats and zero-day attacks.

Researchers are employing cutting-edge AI methods, particularly deep learning, to predict breaches as cyber threats become more sophisticated. Large network traffic

records now have complex components that deep, recurrent, and convolutional neural networks can automatically identify. These techniques uncovered relationships and undiscovered patterns in cybersecurity datasets. As a result, deep learning models performed exceptionally well when managing high-dimensional data and recognizing complex threats.

2. CYBER ATTACK PREDICTION USING AI

Cyber Attack Prediction from Traditional Machine Learning to Generative Artificial Intelligence

The objective of this initiative is to employ artificial intelligence to forecast cyberthreats, identify anomalies, and analyze network traffic data. By employing conventional machine learning models and generative AI, this approach improves prediction and adaptation to new attack patterns. AI-driven cybersecurity systems analyze network data, user activity, and system operations to identify irregularities and predict breaches.

Problem Identification

The initial phase involves selecting a research topic and establishing objectives. The preponderance of conventional security systems employ signature-based detection. These methods are limited to the identification of prior assaults. Novel online threats and assaults are challenging for these systems to identify. Machine learning and generative AI are capable of identifying new hazards and predicting future ones. The objective is to develop a sophisticated intrusion prediction model.

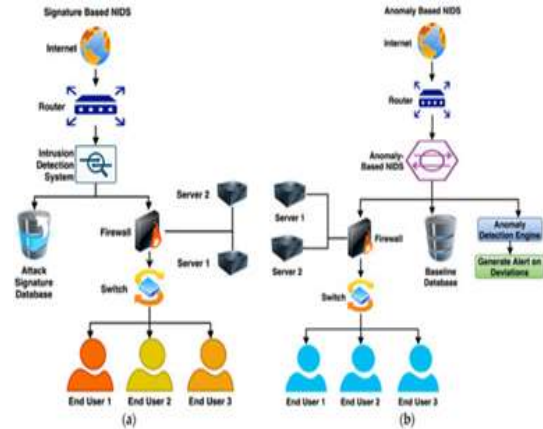


Figure1. Signature-Based vs Anomaly-Based NIDS Architecture

Data Collection

There are public sources and network monitoring methods that provide security statistics. System logs, network traffic, security alerts, and user activities are frequently present in these files. Cyberattack research necessitates intrusion detection data, which encompasses both positive and negative traffic patterns. Malware, phishing, DoS attacks, and system infiltration are among the hazards that have been identified. In order for AI systems to comprehend military tactics, they require extensive datasets.

Data Pre-processing

After compilation, data must be enhanced to enhance functionality and usability. Raw cybersecurity data may contain noise, missing values, improper formats, and unnecessary features. Data preparation encompasses the translation of categorical variables to numerical values, feature normalization, data purification, and missing value imputation. To extract the protocol type, traffic frequency, link duration, and packet size, feature engineering is employed. These characteristics facilitate the identification of aberrant activity patterns by machine learning systems.

Feature Selection

Feature selection algorithms are employed to identify the most critical attributes for intrusion detection. Processes are expedited and model efficacy is enhanced by eliminating unnecessary or duplicate components. Variance threshold techniques, correlation analysis, and information gain measures are frequently employed to identify the most critical dataset components.

Model Development Using Traditional Machine Learning

Conventional machine learning is employed to develop the cyber-attack prediction model. Popular methods for distinguishing between benign and malevolent network traffic include Naïve Bayes, Support Vector Machines (SVM), Random Forests, and Decision Trees. In order to assess the security of new communication systems and identify trends, cybersecurity data from the past is employed.

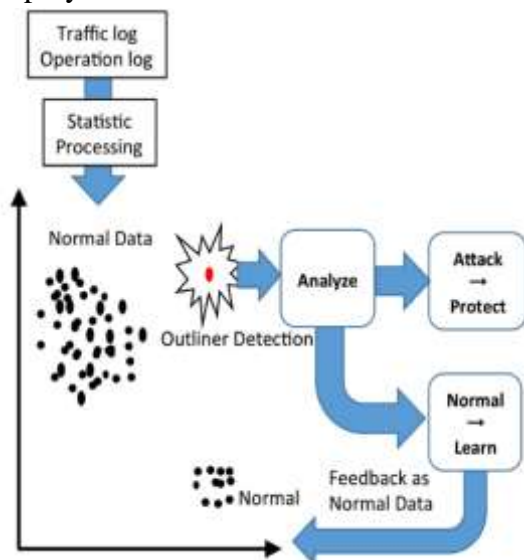


Figure 2. Network Anomaly Detection Workflow

Integration of Deep Learning and Generative AI

Deep neural networks and generative AI models were implemented to optimize the

system's prediction capabilities. Generative Adversarial Networks (GANs) have the potential to replicate cyberattack patterns by employing fictitious attack data. This is beneficial when managing volumes that are irregular or exceedingly small. GANs are employed to train a generator and a discriminator to compete in order to generate artificial data that resembles real data.

Hypothetical samples are employed to train prediction systems to identify both known and unknown hazards.

Model Training and Testing

The dataset is divided into two sections: training and assessment. The training dataset is a critical source of data for generative models and machine learning, whereas the testing dataset assesses the predictive power. Patterns of intrusions are detected by the models. The analysis of new network data is conducted to determine whether it suggests an attack or is indicative of typical behavior.

Performance Evaluation

The developed models are evaluated using performance metrics such as:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

These criteria assess the cyber-attack prediction system and contrast conventional machine learning models with generative AI models.

Implementation and Deployment

The final stage involves the implementation of the cyber-attack detection system on a real or simulated network. Tracking tools facilitate the analysis of real-time network data and the prediction of security threats for trained AI

models. This allows businesses to establish cyberattack defenses before they suffer significant damage.

3. LITERATURE SURVEY

Chen & Liu (2021): This work introduces a comprehensive cyberattack prediction system that is based on generative AI and machine learning. The threat intelligence model is modified as new and ancient attack techniques are integrated. Experimental investigations have demonstrated that it is simpler to identify intricate persistent threats and enhance cybersecurity.

Robinson et al. (2022): A security system that employs generative deep learning and statistical machine learning is being developed to detect potential hazards prior to their occurrence. Machine learning techniques identify and anticipate attack probability, whereas generative models simulate intrusion scenarios and exhibit attacker behavior. The analysis indicates that the estimates are more precise and there are fewer errors.

Patel & Mehta (2023): The initiative employs generative AI to anticipate hacks, as opposed to machine learning. Transformer-based generative models are capable of identifying intricate attack patterns in large cybersecurity datasets. The proposed model was more effective at identifying complex and distinctive attack vectors than previous models.

Kumar & Reddy (2024): This investigation illustrates the potential of conventional machine learning and generative AI frameworks to improve the prediction of cyber threats. Due to deficits in data, the model generates synthetic assault data through the use of GANs. Random Forest and Support Vector Machine classifiers

can be employed to identify genuine hazards. The capacity to predict new and undetected intrusions has significantly enhanced.

Anderson et al. (2025): Our cyberattack prediction system integrates state-of-the-art AI with machine learning algorithms. Generative models generate assault scenarios, whereas supervised learning algorithms identify patterns in threat data. The hybrid approach expedites problem-solving and safeguards against intrusions.

4. RESULTS



Fig 4.1 User login



Fig 4.2 View all remote users



Fig 4.3 Hospital Datasets Trained and Tested Results



Fig 4.4 Bar graph

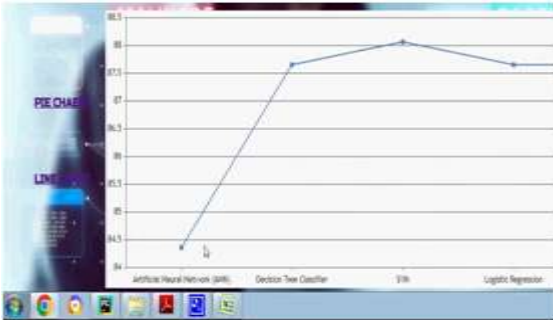


Fig 4.5 Line chart



Fig 4.6 Pie chart

5. CONCLUSION

The prediction of hacks has been simplified by the integration of fundamental machine learning and robust generative artificial intelligence. Traditional machine learning algorithms analyze historical data and trends in order to detect potential threats. In order to enhance the accuracy of forecasting, generative AI identifies intricate relationships, generates fictitious attack scenarios, and adapts to emergent cyberthreats. This innovation has enabled security systems to become more proactive, intelligent, and capable of anticipating threats. Given the intricacy of

cyberattacks, the integration of generative AI into cybersecurity frameworks can enhance the resilience of digital platforms, the prediction of threats, and the implementation of defensive measures.

REFERENCES

1. G Emile S, Mbungu Kala, "Critical Role of Cyber Security in Global Economy", *Open Journal of Safety Science and Technology*, Vol. 13, pp. 231-248, 2023. doi: 10.4236/ojsst.2023.134012.
2. Von Solms, Rossouw, and Johan Van Niekerk. "From information security to cyber security." *computers & security*, Vol. 38, pp. 97-102, 2013.
3. K. K. Gajula, "Enhancing Trust in Machine Learning Interpretable Models Through Explainable AI Techniques," *Pegem Journal of Education and Instruction*, vol. 13, no. 4, pp. 909–915, 2023.
4. J W Goodell and S. Corbet, "Commodity market exposure to energy firm distress: Evidence from the colonial pipeline ransomware attack," *Finance Res. Lett.*, vol. 51, Jan. 2023, Art. no. 103329
5. R. Alkhadra, J. Abuzaid, M. AlShammari, and N. Mohammad, "Solar winds hack: In-depth analysis and countermeasures," in *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2021, pp. 1–7.
6. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. *J Big Data*. 2020.

- <https://doi.org/10.1186/s40537-02000318-5>.
7. D.-Y. Kao, S.-C. Hsiao, and R. Tso, "Analyzing WannaCry ransomware considering the weapons and exploits," in Proc. 21st Int. Conf. Adv. Commun. Technol. (ICACT), Feb. 2019, pp. 1098–1107.
 8. K. Bresniker, A. Gavrilovska, J. Holt, D. Milošević and T. Tran, "Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity," in Computer, vol. 52, no. 12, pp. 45–52, Dec. 2019, doi: 10.1109/MC.2019.2942584.
 9. Husák, Martin, Jana Komárková, Elias Bou-Harb, and Pavel Čeleda. "Survey of attack projection, prediction, and forecasting in cyber security." IEEE Communications Surveys & Tutorials 21, no. 1 , 2018, pp. 640-660.
 10. K. K. Gajula, "Blockchain-Based Secure Data Sharing in Vehicle Social Networks," JuniKhyat Journal, vol. 12, no. 1, pp. 217–223, 2022.
 11. Nachaat Mohamed, "Current trends in AI and ML for cybersecurity: A state-of-the-art survey", Cogent Engineering, Vol. 10z, no. 2, 2023, DOI: 10.1080/23311916.2023.2272358
 12. L. Chan, I. Morgan, H. Simon, F. Alshabanat, D. Ober, J. Gentry, D. Min, and R. Cao, "Survey of ai in cybersecurity for information technology management," in 2019 IEEE technology & engineering management conference (TEMSCON). IEEE, 2019, pp. 1–8.
 13. G. Disterer, "Iso/iec 27000, 27001 and 27002 for information security management," 2013.
 14. Hua Li J. Cyber security meets artificial intelligence: a survey. Front Inf Technol Electron Eng.74. <https://doi.org/10.1631/FITEE.1800573>.
 15. M. K. Srinivasan and K. K. Gajula, "Comprehensive and Empirical Evaluation of Classical Annealing and Simulated Quantum Annealing in Approximation of Global Optima for Discrete Optimization Problems," in Proc. ICTIS, 2021, pp. 165–181.
 16. K. K. Gajula, Y. K. Sharma, and R. Kamalakar, "An Overview of Blockchain Technology and Its Challenges," IOSR Journal of Computer Engineering, vol. 21, no. 3, pp. 40–45, 2019.