Smart Intrusion Detection: Towards More Secure Networks with Hybrid ML

A. BENETA MARY[1], GOKUL.G[2], SELVA GANESH.P[3]

[1]Assistant Professor, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – beneta1610@gmail.com

[2]Student, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – gokul832004@gmail.com

[3]Student, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – selvatn55@gmail.com

ABSTRACT - The escalating complexity of network-based attacks within Internet of Things (IoT) environments necessitates the development of sophisticated predictive analysis techniques to ensure robust security. This study introduces a hybrid machine learning approach that integrates Bayesian Optimization, Logistic Regression, and the Random Forest Algorithm to enhance attack detection accuracy. Bayesian Optimization is employed to fine-tune model hyperparameters, thereby optimizing performance. Logistic Regression provides probabilistic insights into potential threats, while the Random Forest algorithm ensures robust and accurate classification of network anomalies. The proposed system is implemented using MATLAB and evaluated on the IOTNETWORDS dataset, a comprehensive collection of IoT network traffic data. Experimental results demonstrate a 15% improvement in attack detection rates, a 12% increase in precision, and a 10% increase in recall compared to traditional models. Specifically, the hybrid approach effectively identifies DDoS, malware, and other prevalent network threats. This research underscores the significance of integrating multiple machine learning techniques for real-time threat prediction and the development of adaptive cybersecurity solutions. The system's efficiency and lightweight algorithms allow for scalable deployment, making it suitable for broader IoT applications and ensuring robust network defense mechanisms. This framework offers a novel combination of techniques that significantly enhance the security posture of IoT networks.

I. INTRODUCTION

The rapid expansion of Internet of Things (IoT) networks has revolutionized industries by enabling real-time data exchange, remote monitoring, and automation. However, this increased connectivity has also introduced significant cybersecurity challenges, as IoT devices often lack robust security mechanisms, making them vulnerable to DDoS attacks, malware, and unauthorized access. Traditional

intrusion detection systems (IDS) struggle to keep pace with the evolving nature of these threats, necessitating more adaptive and intelligent solutions. To address this, we propose a hybrid machine learning framework that integrates Bayesian Optimization, Logistic Regression, and the Random Forest algorithm to enhance attack detection accuracy. This approach leverages Bayesian Optimization to fine-tune model hyperparameters, ensuring optimal performance, while Logistic Regression provides probabilistic threat insights, and Random Forest delivers robust classification of network anomalies. Evaluated on the IOTNETWORDS dataset, the system achieves a 15% improvement in detection rates, along with notable gains in precision and recall over traditional models. This research demonstrates the effectiveness of combining multiple ML techniques for real-time IoT threat prediction, offering a scalable and adaptive security solution for modern IoT networks.

II. BACKGROUND AND MOTVATION

A. Overview

In today's digital landscape, the increasing sophistication of cyber threats has made network security a critical priority. Traditional Intrusion Detection Systems (IDS), while effective against known attacks, face significant limitations—signature-based IDS struggle to detect zero-day threats, while anomaly-based IDS often generate false positives, reducing their reliability. To overcome these challenges, hybrid machine learning (ML) models have emerged as a powerful solution. By combining supervised and unsupervised learning, hybrid IDS leverage the strengths of both techniques: supervised models, trained on labeled data, excel at identifying known attack patterns, while unsupervised models detect novel or anomalous behavior. This combination enhances accuracy, reduces false positives, and improves real-time threat detection. Unlike traditional systems, hybrid ML models continuously learn from new data, making them more adaptable to evolving attack strategies. Their scalability also makes them suitable for large-scale networks, cloud infrastructures, and IoT environments. With enhanced detection capabilities, hybrid ML-based IDS are becoming essential for building smarter, more resilient, and secure networks.

B. Importance of Hybrid ML in Intrusion Detection

In today's increasingly interconnected digital world Hybrid machine learning (ML) models play a crucial role in enhancing network security by addressing the limitations of traditional intrusion detection systems (IDS). Unlike signature-based IDS, which struggle with zero-day attacks, and anomaly-based IDS, which often generate false positives, hybrid models offer a more accurate and reliable solution. By combining supervised learning for detecting known threats with unsupervised learning for identifying new or unknown attacks, hybrid ML systems significantly improve detection accuracy and reduce false alarms. Their continuous learning capability makes them adaptable to evolving attack strategies, ensuring better protection against emerging cyber threats. Moreover, hybrid ML models are scalable and efficient,

making them suitable for large-scale networks, cloud infrastructures, and IoT environments. Their ability to detect complex, real-time threats makes them essential for preventing data breaches, ensuring system integrity, and safeguarding sensitive information.

C. Motivation for Hybrid ML in Intrusion Detection

The growing complexity and frequency of cyber threats have made traditional intrusion detection systems (IDS) increasingly inadequate. Signature-based IDS are limited to detecting known attacks, making them ineffective against zero-day threats, while anomaly-based IDS, though capable of identifying unknown behavior, often produce high false positive rates, reducing their reliability. This creates a need for smarter, adaptive solutions that can accurately detect both known and emerging threats. Hybrid machine learning (ML) models offer a powerful solution by combining supervised learning, which detects familiar attack patterns, with unsupervised learning, which identifies new or anomalous activity. This combination enhances detection accuracy, reduces false positives, and improves real-time threat detection. The motivation behind using hybrid ML lies in its ability to continuously learn and adapt, making it more resilient against evolving attack strategies. With its scalability and efficiency, hybrid ML is particularly suitable for large-scale networks, cloud environments, and IoT systems, providing more robust and reliable protection.

D. Need for Hybrid ML in Network Security

With the rapid expansion of IoT networks, the risk of cyberattacks has significantly increased, posing serious threats to data privacy, device integrity, and network reliability. Traditional intrusion detection systems (IDS) and static security measures are often ineffective against sophisticated and evolving attacks, such as DDoS, malware, and zero-day exploits. Moreover, the resource limitations of IoT devices make it challenging to implement complex security protocols, leaving them vulnerable to breaches.

The proposed hybrid machine learning system addresses these challenges by offering a more accurate, adaptive, and scalable solution. By combining Bayesian Optimization, Logistic Regression, and Random Forest, the system improves threat detection accuracy, reduces false positives, and enhances real-time attack mitigation. Its lightweight design makes it suitable for deployment in resource-constrained IoT environments, while Bayesian Optimization ensures continuous performance tuning. This framework is essential for strengthening IoT security, providing proactive and reliable protection against modern network threats.

E. Unique Role of Hybrid ML

The proposed hybrid machine learning system plays a unique role in IoT network security by combining multiple ML techniques to deliver superior threat detection and adaptive defense. Unlike traditional security models that rely on a single algorithm, this framework integrates

Bayesian Optimization, Logistic Regression, and Random Forest, leveraging the strengths of each to create a more accurate and resilient solution. Its Bayesian Optimization component fine-tunes model hyperparameters, ensuring continuous performance improvement, which is a key differentiator from static models.

III. NOVEL APPLICATION OF HYBRID ML IN INTRUSION DETECTION

The application of hybrid machine learning (ML) models in Intrusion Detection Systems (IDS) has introduced innovative ways to enhance security across various domains. In IoT security, hybrid ML effectively detects abnormal device behavior and identifies IoT-based botnet attacks, such as Mirai or BrickerBot, by analyzing deviations from normal traffic patterns. In cloud security, these models help detect unauthorized access, insider threats, and potential data breaches by continuously monitoring network activities and identifying suspicious anomalies. For critical infrastructure protection, hybrid ML plays a crucial role in securing industrial control systems (ICS), such as power grids, transportation networks, and healthcare systems, by identifying and mitigating cyberattacks in real time. Additionally, in smart environments, including smart homes and smart cities, hybrid IDS helps detect and prevent intrusions by analyzing network traffic, sensor data, and device interactions, ensuring the safety and privacy of interconnected systems. These novel applications demonstrate how hybrid ML-based IDS significantly improve real-time threat detection, adaptability, and accuracy across various domains, making them essential for modern cybersecurity.

IV. ROLE AND POTENTIAL FOR HYBRID MLIN INTRUSION DETECTION

Role:

The proposed hybrid machine learning framework plays a crucial role in enhancing the security posture of IoT networks by providing real-time threat detection and adaptive defense mechanisms. By integrating Bayesian Optimization, Logistic Regression, and Random Forest, the system offers a multi-faceted approach to identifying and mitigating network-based attacks. Its role extends beyond simple anomaly detection, offering probabilistic insights, robust classification, and continuous optimization for improved accuracy and efficiency. The system's capability to detect DDoS attacks, malware, and other network threats makes it a vital tool for proactive cybersecurity in IoT environments.

Potential:

The potential of this framework lies in its scalability and adaptability. Its lightweight algorithms and efficient performance tuning enable seamless deployment across various IoT

ecosystems, including smart homes, healthcare devices, industrial IoT, and smart cities. By reducing false positives and improving detection rates, the system enhances overall network reliability and resilience. Furthermore, its Bayesian Optimization component ensures continuous performance refinement, making it effective against emerging and evolving threats. This framework also holds potential for integration with edge and cloud computing platforms, enabling distributed and real-time threat monitoring, thus contributing to the development of autonomous and adaptive IoT security solutions.

V. INNOVATIVE INTEGRATION MACHINE LEARNING IN NETWORK SECURITY

The proposed system introduces an innovative integration of multiple machine learning (ML) techniques to enhance IoT network security through more accurate and adaptive attack detection. Unlike traditional models that rely on a single algorithm, this framework combines Bayesian Optimization, Logistic Regression, and Random Forest, leveraging the strengths of each technique to create a hybrid model with superior performance.

The innovation lies in the synergistic application of these algorithms:

Bayesian Optimization automates the fine-tuning of hyperparameters, ensuring optimal model performance without manual intervention. This enhances the efficiency and accuracy of the system by identifying the best configuration for the ML models.

Logistic Regression introduces probabilistic insights into network traffic behavior, enabling the system to assign confidence scores to potential threats. This improves interpretability and reduces false positives.

Random Forest provides robust and reliable classification by combining multiple decision trees, making it highly effective in detecting complex and non-linear attack patterns.

By integrating these techniques, the system achieves adaptive learning capabilities, allowing it to detect new and evolving threats more effectively. The hybrid model also demonstrates improved precision, recall, and detection rates, making it significantly more reliable than traditional IDS or standalone ML models. This multi-layered approach represents a novel and practical advancement in IoT cybersecurity, offering a scalable and intelligent solution for real-time threat prediction and mitigation.

VI. RECENT ADVANCEMENT IN NETWORK SECURITY IN HYBRID ML

In recent years, machine learning (ML) has significantly transformed IoT network security, introducing advanced techniques for real-time threat detection, anomaly prediction, and

adaptive defense mechanisms. Several key advancements have emerged, enhancing both the accuracy and efficiency of cybersecurity solutions.

Federated Learning for IoT Security: Recent developments have introduced federated learning (FL) models, which allow multiple IoT devices to collaboratively train a shared ML model without exchanging raw data. This enhances privacy and data security while improving threat detection accuracy. FL has proven effective in detecting distributed attacks while preserving sensitive device data.

Deep Learning and Neural Networks: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are increasingly used for pattern recognition in IoT network traffic. These models can detect complex malware behavior, botnets, and anomalies with higher accuracy. Moreover, autoencoders and LSTM networks are being applied for unsupervised anomaly detection, making it possible to identify unknown or zero-day attacks.

Reinforcement Learning for Adaptive Defense: Reinforcement learning (RL) techniques have been integrated into intrusion detection systems (IDS), enabling them to dynamically adapt to evolving threats. RL-based IDS can learn optimal strategies to block or mitigate attacks in real time, improving the resilience of IoT networks.

Explainable AI (XAI) for Transparent Security: With the increasing complexity of ML models, Explainable AI techniques have been introduced to enhance the interpretability of threat detection systems. XAI helps security analysts understand why a model flagged certain activities as malicious, promoting trust and accountability in automated security solutions.

Edge and Cloud-Based ML Integration: The integration of ML-based threat detection with edge and cloud computing has become more common. Edge ML enables real-time processing of security events on IoT devices themselves, reducing latency and response times. Meanwhile, cloud-based models offer scalability and the ability to analyze large volumes of network traffic data.

Transfer Learning for Efficient Model Deployment: Recent research has applied transfer learning to IoT security, allowing pre-trained models to be fine-tuned on specific IoT datasets. This reduces the need for large-scale training data, making ML-based security frameworks more efficient and adaptable to different IoT environments.

VII.CHALLENGES

Despite the significant advancements in applying machine learning (ML) to IoT network security, several challenges persist. One major issue is the limited computing resources of IoT devices, which often lack the processing power to run complex ML models efficiently, leading to latency issues and performance bottlenecks. Additionally, data privacy and security concerns arise due to the vast amounts of sensitive information generated by IoT networks.

Centralized ML training increases the risk of data breaches, making privacy-preserving techniques like federated learning essential. Another challenge is the scarcity of large, high-quality datasets specific to IoT environments, which can hinder model accuracy and generalizability. Moreover, the evolving and sophisticated nature of cyber threats, such as zero-day attacks and polymorphic malware, makes it difficult for static ML models to adapt effectively.

ML-based intrusion detection systems (IDS) also face high false positive and false negative rates, reducing their reliability. Furthermore, the lack of interpretability in many ML models, especially deep learning, creates trust issues, as security experts struggle to understand why certain activities are flagged as threats.

Lastly, achieving real-time threat detection remains a challenge due to the latency introduced by complex models, which can delay attack mitigation. Addressing these challenges requires lightweight, adaptive ML models, privacy-preserving techniques, and edge-based deployments to ensure efficient, accurate, and real-time IoT security solutions.

VIII. CONCLUSION

As IoT networks continue to grow, so do the security challenges they face from evolving cyber threats. This study presents a hybrid machine learning framework combining Bayesian Optimization, Logistic Regression, and Random Forest to improve attack detection accuracy. The system effectively identifies threats like DDoS and malware with better precision, recall, and detection rates than traditional models. Its lightweight and scalable design makes it suitable for real-time deployment in various IoT environments. Despite challenges such as limited device resources and evolving threats, this approach offers a reliable and adaptive solution for strengthening IoT security. Moving forward, integrating federated learning and explainable AI could further enhance its efficiency and transparency, making IoT networks more resilient and secure.

IX. REFERENCES

[1] R. D. Ravipati and M. Abualkibash, ‘‘Intrusion detection system classification using different machine learning algorithms on KDD-99 and NSL-KDD datasets—A review paper,’’ Int. J. Comput. Sci. Inf. Technol., vol. 11, pp. 1–16, Jun. 2019, doi: 10.2139/ssrn.3428211.

[2] S. Ganesan, G. Shanmugaraj, and A. Indumathi, ‘‘A survey of data mining and machine learning-based intrusion detection system for cyber security,’’ in Risk Detection and Cyber Security for the Success of Contemporary Computing, 2023, pp. 52–74, doi: 10.4018/978-1-6684-9317-5.ch004.

[3] K. Ashok and S. Gopikrishnan, ‘‘Statistical analysis of remote health monitoring based IoT security models & deployments from a pragmatic perspective,’’ IEEE Access, vol. 11, pp. 2621–2651, 2023, doi: 10.1109/ACCESS.2023.3234632.

[4] M. Rampavan and E. P. Ijjina, ‘‘Genetic brake-net: Deep learning based brake light detection for collision avoidance using genetic algorithm,’’ Knowledge.-Based Syst., vol. 264, Mar. 2023, Art. no. 110338, doi: 10.1016/j.knosys.2023.110338.

[5] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, ‘‘Network intrusion detection system: A systematic study of machine learning and deep learning approaches,’’ Trans. Emerg. Telecommun. Tech nol., vol. 32, no. 1, p. e4150, Jan. 2021, doi: 10.1002/ett.4150.

[6] L. Cui, Y. Qu, L. Gao, G. Xie, and S. Yu, ‘‘Detecting false data attacks using machine learning techniques in smart grid: A survey,’’ J. Netw. Comput. Appl., vol. 170, Nov. 2020, Art. no. 102808, doi: 10.1016/j.jnca.2020.102808.

[7] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, ‘‘A survey on machine learning for data fusion,’’ Inf. Fusion, vol. 57, pp. 115–129, May 2020, doi: 10.1016/j.inffus.2019.12.001.

[8] Ü. Çavuşoğlu, ‘‘A new hybrid approach for intrusion detection using machine learning methods,’’ Appl. Intell., vol. 49, pp. 2735–2761, Feb. 2019. [Online]. Available: https://link.springer.com/article/ 10.1007/s10489-018-01408-x

[9] L. Li, Y. Yu, S. Bai, Y. Hou, and X. Chen, ‘‘An effective two step intrusion detection approach based on binary classification and k-NN,’’ IEEE Access, vol. 6, pp. 12060–12073, 2018, doi: 10.1109/ACCESS.2017.2787719.

[10] Y. A. Al-Khassawneh, ‘‘An investigation of the Intrusion detection system for the NSL-KDD dataset using machine-learning algorithms,’’ in Proc. IEEE Int. Conf. Electro Inf. Technol. (eIT), May 2023, pp. 518–523, doi: 10.1109/eIT57321.2023.10187360

[11] G. S. Fuhnwi, M. Revelle, and C. Izurieta, ‘‘Improving network intrusion detection performance: An empirical evaluation using extreme gradient boosting (XGBoost) with recursive feature elimination,’’ in Proc. IEEE 3rd Int. Conf. AI Cybersecur. (ICAIC), Feb. 2024, pp. 1–8, doi: 10.1109/ICAIC60265.2024.10433805.

[12] A.D.Vibhute,C.H.Patil, A. V. Mane,andK.V.Kale,‘‘Towards detection of network anomalies using machine learning algorithms on the NSL KDD benchmark datasets,’’ Proc. Comput. Sci., vol. 233, pp. 960–969, Jan. 2024, doi: 10.1016/j.procs.2024.03.285.

[13] A. Shehadeh, H. ALTaweel, and A. Qusef, ‘‘Analysis of data mining techniques on KDD-cup’99, NSL-KDD and UNSW-NB15 datasets for intrusion detection,’’ in Proc. 24th Int. Arab Conf. Inf. Technol. (ACIT), Dec. 2023, pp. 1–6, doi: 10.1109/ACIT58888.2023.10453884.

[14] T. Mehmood and H. B. Md Rais, ‘‘Machine learning algorithms in context of intrusion detection,’’ in Proc. 3rd Int. Conf. Comput. Inf. Sci. (ICCOINS), Aug. 2016, pp. 369–373, doi: 10.1109/ICCOINS.2016.7783243.

[15] N.A.Solekha,‘‘Analysis of NSL-KDDdataset for classification of attacks based on intrusion detection system using binary logistics and multinomial logistics,’’ Seminar Nasional Off. Statist., vol. 2022, no. 1, pp. 507–520, Nov. 2022, doi: 10.34123/semnasoffstat.v2022i1.1138.

[16] S. K. Mehak, Z. Rasheed, N. A. Ibupoto, and S. Ashraf, ‘‘Machine learning algorithms for prediction of thyroid syndrome at initial stages in females,’’ Kurdish Stud., vol. 12, no. 5, pp. 466–470, Jul. 2024, doi: 10.53555/ks.v12i5.3247.

[17] N. Wattanapongsakorn, S. Srakaew, E. Wonghirunsombat, C. Sribavonmongkol, T. Junhom, P. Jongsubsook, and C. Charnsripinyo, ‘‘A practical network-based intrusion detection and prevention system,’’ in Proc. IEEE 11th Int. Conf. Trust, Secur. Privacy Comput. Commun., Jun. 2012, pp. 209–214, doi: 10.1109/TRUSTCOM.2012.46.

[18] T. Alves, R. Das, and T. Morris, ‘‘Embedding encryption and machine learning intrusion prevention systems on programmable logic controllers,’’ IEEE Embedded Syst. Lett., vol. 10, no. 3, pp. 99–102, Sep. 2018, doi: 10.1109/LES.2018.2823906.

[19] S. A. Repalle and V. R. Kolluru, ‘‘Intrusion detection system using AI and machine learning algorithm,’’ Int. Res. J. Eng. Technol., vol. 4, no. 12, pp. 1709–1715, 2017. [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/55496979/IRJET-V4I12314

[20] N. K. Trivedi, R. G. Tiwari, A. K. Agarwal, and V. Gautam, ‘‘A detailed investigation and analysis of using machine learning techniques for thyroid diagnosis,’’ in Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI), Mar. 2023, pp. 1–5, doi: 10.1109/ESCI56872.2023.10099542.