iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73938

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

A Survey on An Anomaly Detection Improvement in Computer Communication Network using ML Techniques

Nagarathna C[1], Pritiparna Kar[2], Shivani R[3], Sneha G C[4], Vinutaa Reddy[5]
Dept. of CSE Sapthagiri College of Engineering

***Abstract:** In today's digital age, guarding network structure from cyber pitfalls is a major concern. Traditional hand-grounded Intrusion Detection Systems (IDS) often struggle to identify new or unknown attacks. This creates a need for smarter and more flexible results. This project introduces an anomaly-grounded Network Intrusion Detection System (NIDS) that uses ML approaches to spot unusual functioning in network traffic. The system analyzes patterns and differences from typical activity to detect implicit intrusions, including Denial-of-Service (DoS) attacks, probe attacks, User to Root (U2R) exploits, and Remote to Local (R2L) breaches in real-time. The system works with preprocessed network datasets like NSL-KDD or CICIDS. It extracts important features and trains supervised machine learning models such as Random Forest, Support Vector Machine, and K-Nearest Neighbors (KNN) for accurate classification.*

I. INTRODUCTION

In moment's digital age, computer networks serve as the backbone of nearly every field, from personal communication to essential services like healthcare, banking, defense, and e-commerce. As we rely more on interconnected systems, the threat of malicious events and prohibited access to sensitive data has grown significantly. Protecting network resources from these threats is a pressing concern for both associations and individuals. One common approach to tackling this challenge is Network Intrusion Detection (NID). Network intrusion detection involves continuously monitoring, analyzing, and determining network traffic to spot unusual patterns or viciousmovements that could compromise system security. A Network Intrusion Detection System (NIDS) is designed to perform this function by inspecting the data packets flowing across the network and detecting suspicious behavior or policy violations in real-time. By furnishing early warnings about possible intrusions, NIDS play a pivotal role in maintaining the privacy, integrity,durability and obtainability of information resources.

Intrusion attempts in computer networks can be classified into several types based on the attacker's methods and goals. Denial-of-Service (DoS) attacks aim to overwhelm network resources or services by bombarding them with excessive demands, halting legitimate users from acquiring the system. Probe attacks involve reconnaissance activities where an attacker scans the network to gather information about active hosts, open ports, and vulnerabilities, often as a precursor to more serious attacks. Remote-to-Local (R2L) attacks happen when an attacker without authorized access gains local user privileges, typically through password guessing, phishing, or exploiting unpatched software vulnerabilities. Lastly, User-to-Root (U2R) attacks are particularly dangerous, as they involve privilege escalation where a threat actoroperating under a basic user accountgains root or administrative privileges, leading to full system compromise. This variety of attack typesunderscores the need forpowerful recognition methods
that can handle a wide range of threats.

Conventional intrusion detection methods have largely depended on signature-based and rule-based methods. In these systems, predefined patterns of known attacks—called signatures—are stored in a database. An alert is generated whenever the system detects network activity that matches these patterns. While effective against previously identified intrusions, this method is limited because it cannot detect zero-day exploits, new attacks, or polymorphic malware that do not match existing signatures. Additionally, rule-based systems often produce high rates of wrong alarms and undetectedattacks, leading to inefficiencies in real-world applications. Attackers frequently alter or disguise their methods to evade these static rules, causing it challenging for conventionalmeans to keep up with the changing threat landscape.

To address these shortcomings, researchers and practitioners have stressed the significance of intelligent and adaptive intrusion detection methods. By utilizing machine learning (ML), artificial intelligence (AI), and statistical analysis techniques, modern NIDS can go beyond predefined rules to identify anomalies in network behavior.

These systems can be trained on historical datasets of both normal and malicious traffic, allowing them to recognize subtle deviations from typical usage patterns that may signal intrusions. For example, supervised learning algorithms cancategorize network traffic as normal or attacking, while unsupervised learning approaches can detect previously unseen attack patterns by finding outliers. Moreover, deep learning models and hybrid detection frameworks have gained attention lately because of their potential to spontaneously learn complicated features from unprocessed traffic data without relying heavily on manual feature engineering.

Thus, the necessity for intellectual intrusion detection has become essential in current cybersecurity. While traditional systems provide a foundational level of defense against known threats, intelligent NIDS are the next step toward creating resilient, adaptive, and future-ready security frameworks. They not merelyassure more reliable protection of digital assets but also help strengthen trust in digital communication and safeguard the critical infrastructure that modern society relies on.

II. BACKGROUND AND MOTIVATION

Network protection is a vital concern because of the fast growth of digital infrastructures. Intrusion Detection Systems (IDS) inspect and analyze traffic for harmful activities. Network Intrusion Detection Systems (NIDS) focus specifically on real-time packet-level analysis. Traditional IDS techniques, which are mainly rule-based and signature-based, work well for known attacks but do not cover new or changing threats.

A. Limitations of Existing Systems

Conventional IDS face significant challenges such as an inability to detect zero-day attacks, high false alarm rates, and limited adaptability to dynamic cyber threats. Attackers continuously modify their strategies, making it difficult for static systems to provide comprehensive protection. These shortcomings highlight the demand for more enhances detection mechanisms.

B. Motivation

To tackle these challenges, smart anomaly-based detection methods have turned into more important. By using machine learning, NIDS can spot both known and unknown attacks, adjust to new threats, and lower false positives. This drives the creation of scalable, real-time, and flexible intrusion detection systems that improve cybersecurity resilience in today's network environments.

III. LITERATURE SURVEY

1) *Ahmed, O. (2024). Enhancing Intrusion Detection in Wireless Sensor Networks through Machine Learning Techniques and Context Awareness Integration. International Journal of Mathematics, Statistics, and Computer Science*

METHODOLOGY: Reviewed context-aware machine learning strategies for WSN intrusion detection. Proposed integrating contextual features, such as node energy, location, and time, into models. Validated the approach conceptually against typical WSNscenarios.LIMITATIONS: Lacks large-scale experimental evaluation on real deployments. The context feature extraction overhead and energy cost are not quantified. Generalization to diverse WSN topologies is still unproven.

2) *Shafi, M.; Molisch, A.F.; Smith, P.J.; Haustein, T.; Zhu, P.; De Silva, P.; Tufvesson, F.; Benjebbour, A.; Wunder, G. 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice*

METHODOLOGY:Survey of 5G standards, trials, and deployment strategies. This includes technical requirements, use cases, and architectural choices. It compares different techniques and practical trial results.LIMITATIONS: Focused on 5G era constraints. It does not address post-5G or 6G challenges. The survey's nature limits new algorithm contributions.Rapidgrowth of standards may have prompted some details to become obsolete.

3) *Nagatsuma (2015) — "Terahertz communications: Past, present and future"*

METHODOLOGY:Reviewed physical-layer terahertz technologies and the challenges of system integration. Examined device and circuit-level factors that enable THz communications. Identified application scenarios and created a roadmap.LIMITATIONS: Mostly descriptive. There is limited quantitative validation of system-level performance. Practical deployment barriers, like atmospheric losses and hardware cost, are only partly addressed. Implementation complexity and manufacturing readiness are not fully resolved.

4) Crowe et al. (2017) — "Terahertz RF Electronics and System Integration"

METHODOLOGY: Surveyed RF front-end component designs and system integration strategies for THz electronics. Discussed trade-offs between performance, power, and integration methods. Highlighted prototype demonstrations. LIMITATIONS: Focus on improvements at the component level instead of the whole networking system. Power consumption forecasts look promising, but they need full-stack validation. Thermal and packaging challenges need further experimental research.

5) You et al. (2017) — "Pollution Air Monitoring System Based on Space-borne Terahertz Radiometer"

METHODOLOGY: Designing a remote-sensing system with THz radiometry for detecting atmospheric pollutants involved modeling how radiation moves and how the instruments respond. We validated its feasibility through simulation and measurement data. LIMITATIONS: System receptivity under varying weather environments is not fully stated. Spaceborne deployment limits, such as size and power, are only mentioned briefly.

6) Tarish (2024) — "Enhancing 5G communication in business networks with a secured narrowband IoT framework"

METHODOLOGY*:* Proposed a secure narrowband IoT framework designed for enterprise 5G use. Described the changes needed for protocols and architecture, as well as the security methods. Evaluated this framework through conceptual analysis and small-scale tests.LIMITATIONS: Lacks large-scale deployment and testing in different enterprise settings. Performance and security trade-offs, like latency and overhead, need careful measurement.

7) Rey, S.; Eckhardt, J.M.; Peng, B.; Guan, K.; Kürner, T. Channel Sounding Techniques for Applications in THz Communications.

METHODOLOGY: Developed and described channel sounding techniques which are apt for ultra-wideband THz measurements. Built a correlation-based sounder and showed measurement procedures. Analyzed channel impulse responses.LIMITATIONS: Measurements are mainly in controlled or static situations. Studies on limited mobility could also be a factor. The complexity of the equipment and the need for synchronization may restrict its use in the field.

8) MacCartney & Rappaport (2017) — Flexible mmWave channel sounder

METHODOLOGY: Presented a flexible millimeter-wave channel sounder design with precise timing. I showed measurement campaigns and described channel statistics. I used the findings to improve mmWave models.LIMITATIONS:Focused on mmWave, not the full THz band. The complexity of the sounder hardware may limit replication.

9) Pirkl & Durgin (2008) — Optimal sliding correlator channel sounder design

METHODOLOGY:.Analysis of sliding-correlator based channel sounders and their optimal design parameters. It includes mathematical models and design recommendations. These were validated through simulations and experiments.LIMITATIONS: Targeted classical sliding-correlator designs, newer digital alternatives are not covered. Practical implementation constraints in ultra-high-frequency bands are not fully addressed.

10) Raheema & Tarish (2023) — Secure User Authentication Protocol for WSNs

METHODOLOGY: Designed a user verification protocol for WSNs. It includes cryptographic steps and message flows. It formally examined the security properties.LIMITATIONS: Energy and latency overheads in constrained WSN devices are not fully measured. Resistance to advanced attacks, such as side-channel and physical tampering, has not been experimentally validated.

11) Sengupta, Nagatsuma, Mittleman (2018) — Terahertz integrated electronic & hybrid systems

METHODOLOGY: Reviewed electronic and photonic methods for THz systems. Compared hybrid designs and component technologies. Summarized plan for integration.LIMITATIONS: High-level roadmap, but no full system performance demonstrations are included. Economic and production feasibility will be addressed in future work.

12) Tarish & Raheema (2021) — Central Multipath Routing using Neural Networks
METHODOLOGY: Proposed a routing method by means of neural networks to reduce TCP/IP congestion. Simulated multipath scenarios and compared routing results. Evaluated congestion metrics.LIMITATIONS: Simulations might not reflect all real-world network behavior. The overhead of neural models and the training costs in routers have not been measured.

13) Barros et al. (Integrated THz with reflectors for small-cells)
METHODOLOGY: Investigated how reflectors can extend THz small-cell coverage. Modeled propagation using reflectors and simulated improvements in link performance. Proposed guidelines for deployment. LIMITATIONS:Channel variability and blockage still limit reliability. The cost-benefit analysis of alternative solutions is not fully provided.

14) Shin et al. (Feasibility of wireless datacenters)
METHODOLOGY: Explored the possibility of fully wireless data centers using high-bandwidth wireless links. Modeled traffic and evaluated link budgets. Examined interference patterns and discussed practical challenges.LIMITATIONS: Thermal and power trade-offs for wireless transceivers are not quantified. Migration complexity from wired to wireless data centers is not resolved.

15) Yang, Shutler, Grischkowsky (2011) — Atmospheric transmission measurement 0.2–2 THz
METHODOLOGY: Measured atmospheric transmission across the 0.2 to 2 THz range using THz-TDS. I analyzed absorption features and the impacts of water vapor. I derived implications for link design.LIMITATIONS: Measurements may depend on the site and weather. We need wider sampling over space and time. High atmospheric attenuation restricts many outdoor uses without mitigation strategies.

16) Gordon et al. (HITRAN2016)
METHODOLOGY: Compiled and standardized molecular spectroscopic data from HITRAN for radiative transfer modeling. Provided a database and validated it against experimental spectra. This work improved atmospheric modeling. LIMITATIONS: Database completeness depends on the available laboratory measurements. Some bands are sparse. High-precision modeling needs expertise and computing resources.

17) Ma et al. (Channel performance for indoor/outdoor THz links)
METHODOLOGY: Analysis of channel performance metrics for indoor and outdoor THz links. Experiments and simulations were used to measure path loss, multipath, and achievable rates. Modeling parameters were proposed... LIMITATIONS: Limited dataset diversity across building types and outdoor environments. Mobility scenarios and Doppler effects require more measurement.

18) Goldsmith (Wireless Communications book)
METHODOLOGY: A detailed discussion of the basics of wireless communication. It includes important models for fading, capacity, and resource allocation. Mathematical proofs and examples were used throughout.LIMITATIONS : Application to THz/6G specifics needs extra material. Implementation-level engineering constraints, like manufacturing and packaging, are not the focus.

19) Kato et al. (2020) — "Ten challenges in advancing ML technologies toward 6G"
METHODOLOGY: Identified and discussed ten research challenges for using ML in 6G systems. Synthesized insights from the community and suggested research directions. Framed open problems. LIMITATIONS: High-level roadmap. It lacks empirical validation for the proposed approaches. The prioritization of challenges may differ by application and region.

20) Zhang, Patras, Haddadi (2019) — "Deep learning in mobile and wireless networking: A survey"
METHODOLOGY: Surveyed deep learning applications in mobile and wireless networking. Categorized methods, datasets, and tasks. Reviewed strengths and weaknesses, then suggested research paths. LIMITATIONS: Survey breadth limits depth on individual method reproducibility. Newer DL advances after 2019, such as transformers and federated updates, were only partially covered.

IV. METHODOLOGY

The proposed project involves collecting and preprocessing data using benchmark datasets like NSL-KDD and CICIDS. We clean, encode, normalize, and balance the raw data to prepare it for training. Next, we apply feature selection to keep the most relevant attributes. This step helps optimizeexactness while curtailing computational load. The prepared parameters is then utilized to train and develop machine learning models, with Random Forest, SVM, and KNN. We fragment the dataset into training and testing segments. Every single model is analyzed with the help of performance parametersnamely accuracy, precision, recall, and F1-score to find the most efficient classifier. Finally, we deploy the best model in dynamic. It persistentlyobserves network traffic, spots anomalies, and initiates alerts for administrators.

A. Data Acquisition and Preprocessing

The first step is to gather high-quality datasets that reflect real-world network environments. For this project, we use well-known datasets like NSL-KDD and CICIDS2017/2018. These datasets include labeled examples of both normal and abnormal network traffic, covering various types of attacks, including DoS (Denial of Service), R2L (Remote to Local), U2R (User to Root), and probing attacks. After acquiring the data, we need to preprocess it to ensure it is clean and suitable for machine learning. This involves: removing missing or irrelevant values to prevent errors during model training, encoding categorical variables like protocol type or service (for example, converting 'TCP' and 'UDP' into numbers), normalizing numeric features such as duration or packet size to ensure all data points areconfined to the same scale, and balancing the dataset if there is a class imbalance, such as having more normal traffic than attacks. This helps prevent bias in the model.

B. Feature selection

Feature selection is an important step that affects the model's accuracy and efficiency. We identify and keep only the most relevant attributes from the dataset. For example, features such as "number of failed logins," "duration of connection," or "source bytes" often show strong signs of suspicious activity. By reducing noise and dimensionality, we help the model concentrate on the most useful data. This improves detection performance and lowers the computational cost.

C. Model selection and training

Three supervised learning algorithms are incorporated and compared. Random Forest (RF) is an ensemble model that develops numerous decision trees combined with their output to upgradereliability and robustness. Support Vector Machine finds the best boundary that separates classes (normal vs. anomaly) in the feature space. K-Nearest Neighbors (KNN) classifies latest data by correlating it to the most similar known examples. Each model is trained using the preprocessed and feature-engineered data. We split the dataset into a training set, typically 80%, and a testing set, 20%, to evaluate the model's ability to generalize.

D. Model Evaluation

After the models are considered as trained, they are tested on novel data using key performance metrics: Accuracy: the overall rightness of the model. Precision: Fraction of true positives withinestimated positives (how many flagged attacks are real). Recall: Percentage of true positives detected (how many actual attacks were detected). F1-Score: A proportionwith respect to precision and recall. Confusion Matrix: A visual representation of true positives,, true negatives, false negativesand false positives. These metrics help us compare model performance and choose the best algorithm for real-time use.

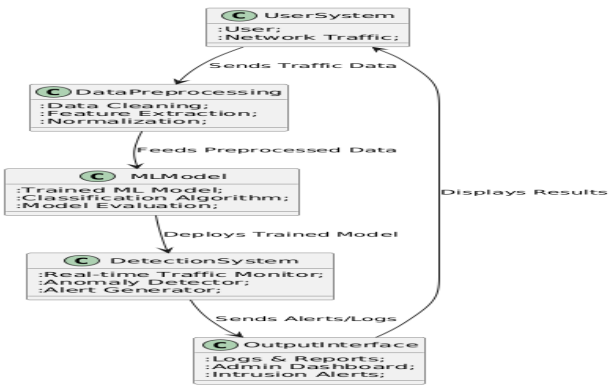


Fig. Methodology

E. Real time detection pipeline

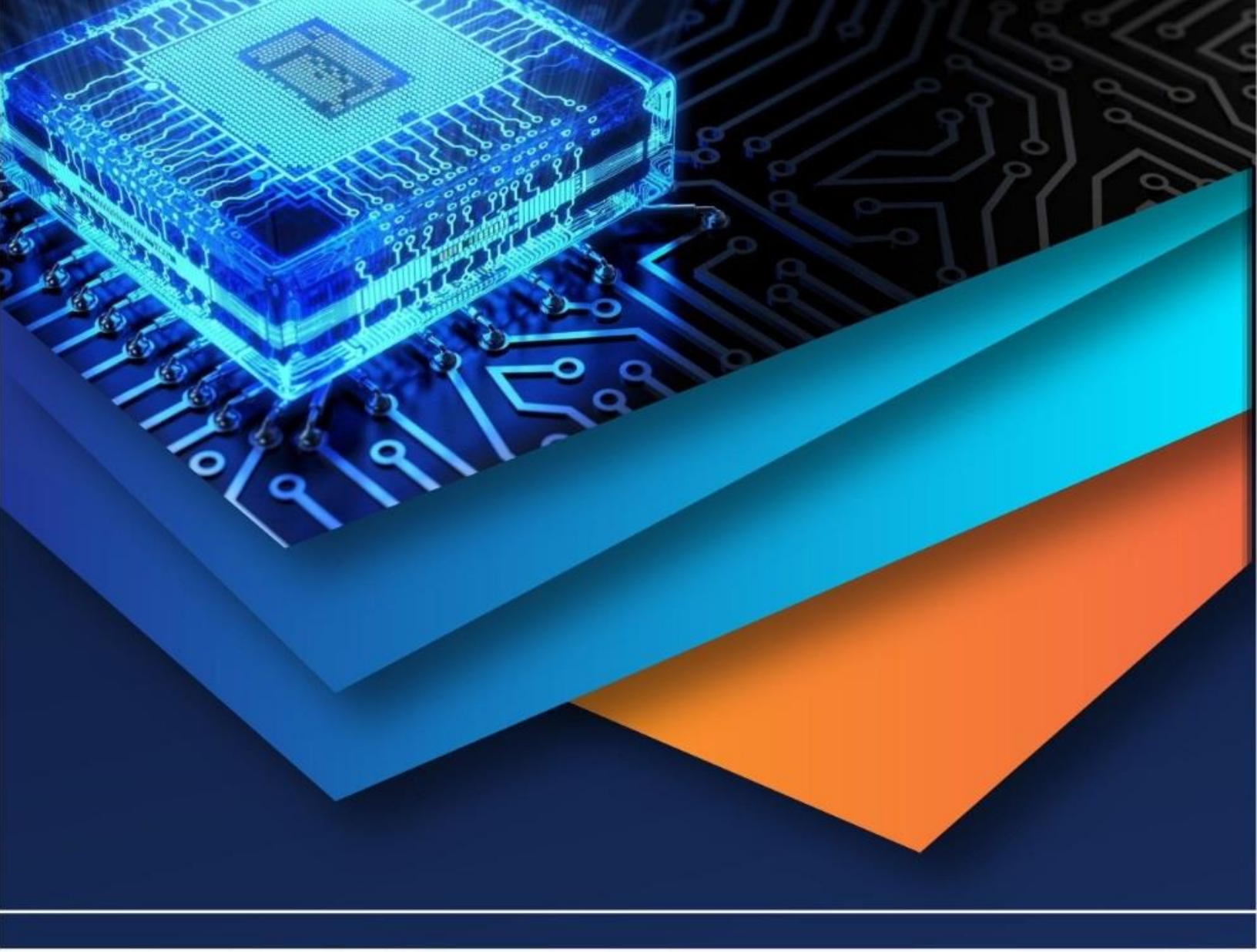
After selecting the best model, we put it into a real-time detection system. Live network traffic or logs flow to the model nonstop. The system pulls features from each connection. The trained model examines the traffic and classifies it as normal or unusual. If it finds something unusual, the system creates an alert, records the event, and updates the interface. The detection system runs all the time, allowing for constant monitoring of network activity and quick threat identification.

V. CONCLUSION

The project designs a machine learning modelwhichflags anomalies and malicious activities in real time. It improves accuracy and reduces false alarms, and it works better than traditional intrusion detection systems. Overall, it boosts cybersecurity through automation, scalability, and smart data analysis.

REFERENCES

[1] Ahmed, O. (2024). Enhancing Intrusion Detection in Wireless Sensor Networks through Machine Learning Techniques and Context Awareness Integration. International Journal of Mathematics, Statistics, and Computer Science,
[2] Shafi, M.; Molisch, A.F.; Smith, P.J.; Haustein, T.; Zhu, P.; De Silva, P.; Tufvesson, F.; Benjebbour, A.; Wunder, G. 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice. IEEE J. Sel. Areas Commun. 2017
[3] Yu, H.; Lee, H.; Jeon, H. What is 5G? Emerging 5G Mobile Services and Network Requirements. Sustainability 2017
[4] Nagatsuma, T. Terahertz communications: Past, present and future. In Proceedings of the 2015 40th International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz), Hong Kong, China, 23–28 August 2015; pp. 1–2.
[5] Crowe, T.W.; Deal, W.R.; Schröter, M.; Tzuang, C.K.C.; Wu, K. Terahertz RF Electronics and System Integration. Proc. IEEE 2017, 105, 985–989
[6] Tarish, H.A., "Enhancing 5G communication in business networks with an innovative secured narrowband IoT framework", Journal of Intelligent Systems
[7] Raheema, A.Q., Tarish, H.A., "Analyze and Design of Secure User Authentication Protocol for Wireless Sensor Networks", AIP Conference Proceedings
[8] Shafi, M.; Molisch, A.F.; Smith, P.J.; Haustein, T.; Zhu, P.; De Silva, P.; Tufvesson, F.; Benjebbour, A.; Wunder, G. 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice. IEEE J. Sel. Areas Commun. 2017, 35, 1201–1221
[9] Segan, S. What Is 5G?
[10] Sengupta, K.; Nagatsuma, T.; Mittleman, D.M. Terahertz integrated electronic and hybrid electronic photonic systems. Nat. Electron. 2018, 1, 622–635
[11] Tarish, H.A., Raheema, A.Q., "Central Multipath Routing to Minimize Congestion in Tcp/Ip Networks Using Neural Networks", Lecture Notes in Networks and Systems2021, 243, pp. 499–507
[12] Gordon, I.E.; Rothman, L.S.; Hill, C.; Kochanov, R.V.; Tan, Y.; Bernath, P.F.; Birk, M.; Boudon, V.; Campargue, A.; Chance, K.; et al. The HITRAN2016 molecular spectroscopic database. J. Quant. Spectrosc. Radiat. Transf. 2017, 203, 3–69
[13] Yang, Y.; Mandehgar, M.; Grischkowsky, D.R. Understanding THz Pulse Propagation in the Atmosphere. IEEE Trans. Terahertz Sci. Technol. 2012
[14] Goldsmith, A. Wireless Communications; Cambridge University Press: New York, NY, USA, 2005.
[15] Z Ahmad, A Shahid Khan, K Nisar, I Haider, R Hassan, MR Haque, “Anomaly detection using deep neural network for IoT architecture”, Applied Sciences 11(15), 7050, 2021
[16] C. Zhang, P. Patras, and H. Haddadi, “Deep learning in mobile and wireless networking: A survey,” IEEE Communications surveys & tutorials, vol. 21, no. 3, pp. 2224–2287, Mar. 2019.
[17] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, “Ten challenges in advancing machine learning technologies toward 6G,” IEEE Wireless Communications, vol. 27, no. 3, pp. 96–103, Apr. 2020.
[18] [Barros, M.T.; Mullins, R.; Balasubramaniam, S. Integrated Terahertz Communication with Reflectors for 5G Small-Cell Networks. IEEE Trans. Veh. Technol. 2017, 66, 5647–5657
[19] Hilbert, J.L. Tunable RF Components and Circuits, 1st ed.; Applications in Mobile Handsets, CRC Press: Boca Raton, FL, USA, 2018
[20] Shin, J.Y.; Sirer, E.G.;Weatherspoon, H.; Kirovski, D. On the feasibility of completely wireless datacenters. In Proceedings of the 2012 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS), Austin, TX, USA, 29–30 October 2019
[21] You, R.; Lu, Z.; Hou, Q.; Jiang, T. Study of Pollution Air Monitoring System Based on Space-borne Terahertz Radiometer. In Proceedings of the 10th UK-Europe Workshop on Millimetre Waves andTerahertz Technologies, Liverpool, UK, 11–13 September 2017
[22] Ericsson AI and Automation, “Employing AI techniques to enhance returns on 5G network investments,” Ericsson, Tech. Rep., May 2019
[23] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, “Federated learning in mobile edge networks: A comprehensive survey,” IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 2031–2063, Apr. 2020
[24] Mendis, R.; Nagai, M.; Wang, Y.; Karl, N.; Mittleman, D.M. Terahertz Artificial Dielectric Lens. Sci. Rep. 2016
[25] Federici, J.F.; Su, K.; Moeller, L.; Barat, R.B. Experimental comparison of terahertz and infrared data signal attenuation in dust clouds. JOSA A 2012, 29, 2360–2366.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY