ISSN: 2395-7852

International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 12, Issue 2, March- April 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.028

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

Network Intrusion Detection System using Decision Tree Algorithm

Mrs.R.Karthika M.Sc.[1], K.C.Harshavarthan[2]

Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India [1]

UG Student, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India [2]

ABSTRACT: In today's digital landscape, network security is a critical concern as cyber threats continue to evolve, becoming more sophisticated and damaging. Traditional security measures, such as firewalls and antivirus programs, are often inadequate in detecting and preventing complex cyber-attacks. To address this challenge, Intrusion Detection Systems (IDS) play a pivotal role in identifying and mitigating unauthorized access and malicious activities in a network. This paper presents a Network Intrusion Detection System (NIDS) utilizing the Decision Tree (DT) algorithm for effective and efficient anomaly detection. This paper explores the design and implementation of an effective NIDS capable of detecting intrusions using machine learning-based techniques. The study reviews various intrusion detection approaches, including signature-based and anomaly-based methods, emphasizing their strengths and limitations. Machine learning algorithm, including Decision Tree and Deep Learning models, are evaluated for their effectiveness in detecting malicious network traffic.The proposed NIDS leverages benchmark datasets such as KDD Cup 99 and NSL- KDD, which contain a diverse range of attack types, to train and test various models. Performance metrics such as accuracy, precision, recall, F1-score, and false positive rate are used to assess the detection capabilities of each approach. The study finds that while traditional methods provide reliable results, machine learning-based IDS significantly improve detection rates and adaptability to new attack patterns. Additionally, feature selection and optimization techniques are explored to enhance the system's performance. This research examines real- time implementation challenges, including data preprocessing, model scalability, computational efficiency, and response time. The findings suggest that integrating machine learning with advanced feature engineering and ensemble techniques can create a more robust and efficient NIDS capable of mitigating emerging cybersecurity threats.

KEYWORDS: Network Intrusion Detection System (NIDS), Decision Tree Algorithm, Cybersecurity, Network Security, Anomaly Detection, Feature Selection, Real-time Detection.

I. INTODUCTION

In the modern digital era, cybersecurity threats have grown exponentially, posing significant risks to individuals, enterprises, and governments worldwide. Cyberattacks such as data breaches, ransomware, and distributed denial-of-service (DDoS) attacks disrupt critical systems and cause financial and reputational damage. As a response to the increasing complexity of cyber threats, Network Intrusion Detection Systems (NIDS) have become indispensable tools for monitoring network traffic and identifying malicious activities. These systems play a crucial role in protecting sensitive data, preventing unauthorized access, and ensuring the integrity of network infrastructures.

Traditional signature-based intrusion detection methods, though widely used, have limitations in detecting novel and evolving cyber threats. As attackers continuously develop new techniques to evade security mechanisms, there is a pressing need for more adaptive and intelligent detection approaches. Machine learning-based NIDS has emerged as a powerful alternative, leveraging historical data to recognize attack patterns and classify network traffic dynamically. Among various machine learning techniques, the Decision Tree algorithm is particularly well-suited for intrusion detection due to its interpretability, high detection accuracy, and efficient computational performance.

The Decision Tree algorithm is a supervised learning method that classifies network traffic by recursively partitioning data based on selected features. Its hierarchical structure enables quick decision-making, making it ideal for real-time intrusion detection scenarios. Additionally, Decision Trees can handle both numerical and categorical data, allowing for effective classification of various types of network intrusions.

By applying feature selection techniques, the performance of the model can be further optimized, reducing false positives and improving detection efficiency.

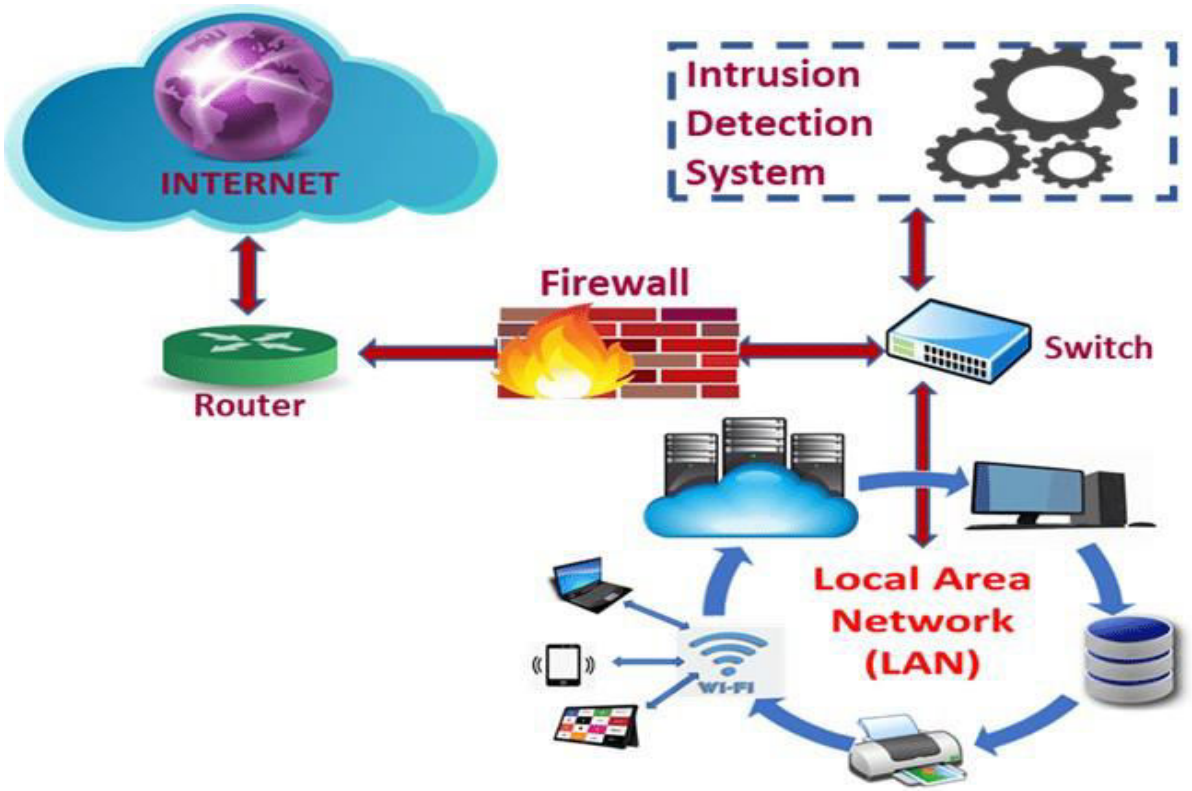

Fig 1: Intrusion detection system

II. RELATED WORK

Overview of Traditional Intrusion Detection Methods

Network Intrusion Detection Systems (NIDS) have traditionally relied on two primary methods**: signature-based detection** and **anomaly-based detection**. Each method has its strengths and weaknesses, and both have been widely used in cybersecurity applications.

Signature-Based Intrusion Detection

Signature-based intrusion detection, also known as misuse detection, operates by comparing network traffic against a database of known attack patterns or signatures. This approach is highly effective at identifying previously encountered attacks with well-defined characteristics. Popular signature-based IDS solutions include Snort and Suricata, which are widely used in enterprise security environments.

Advantages

- High accuracy in detecting known attacks.
- Low false positive rate since detections are based on predefined signatures.
- Fast detection due to efficient pattern matching techniques.

Limitations

- Ineffective against zero-day attacks and novel threats.
- Requires continuous updates to maintain an up-to-date signature database.
- Susceptible to evasion techniques such as polymorphic and encrypted malware.

Anomaly-Based Intrusion Detection

Anomaly-based detection identifies malicious activities by analyzing deviations from normal network behavior. Instead of relying on predefined attack signatures, it builds a model of normal network activity and flags deviations as potential intrusions. This method is more adaptive and can detect zero-day attacks.

Advantages

- Capable of detecting unknown and evolving cyber threats.
- Can adapt to changes in network behavior over time.
- Suitable for real-time monitoring and behavioral analysis.

Limitations

- Higher false positive rates due to variations in legitimate network behavior.

- Requires extensive training data to build an accurate normal behavior model.

III. METHODOLOGY

The methodology for developing a Network Intrusion Detection System (NIDS) using a Decision Tree algorithm consists of structured steps to ensure accurate detection of malicious network activity. This process involves data collection, feature selection and engineering, decision tree and real-time deployment.

A.DATA COLLECTION

The data collection phase is crucial in building a Network Intrusion Detection System (NIDS) as it ensures the availability of high-quality network traffic data for training and testing the model. The collected data should represent both normal and malicious network activity to facilitate effective intrusion detection**.**

Publicly Available Datasets

Several standard datasets are widely used for intrusion detection research:

1. **NSL-KDD Dataset**
   - An improved version of the KDD'99 dataset, which eliminates duplicate and redundant records.
   - Contains labeled traffic, including normal traffic and four attack categories:
     - Denial of Service (DoS)
     - Probing (Probe)
     - User to Root (U2R)
     - Remote to Local (R2L)

2. **CIC-IDS2017 Dataset**
   - Developed by the Canadian Institute for Cybersecurity.
   - Simulates real-world attack scenarios, including:
     - Brute-force attacks
     - DDoS (Distributed Denial of Service)
     - Botnet and malware traffic
   - Contains raw packet captures (PCAPs) and preprocessed CSV files with extracted features.

B.FEATURE SELECTION & ENGINEERING

Feature selection is performed to identify the most relevant attributes that contribute to network intrusion detection. Techniques such as Mutual Information, Recursive Feature Elimination (RFE), and Decision Tree-based feature importance ranking are used to filter the best features. Important features include protocol type, packet size, connection duration, flow rate, and flag-based attributes (e.g., SYN, ACK, FIN flags).

Feature engineering involves transforming and deriving new features to enhance model performance. This includes encoding categorical features (e.g., converting protocol types into numerical values), scaling numerical features, and extracting temporal or behavior-based characteristics of network traffic.

C.DECISION TREE ALGORITHM

The Decision Tree algorithm is a widely used supervised learning method for Network Intrusion Detection Systems (NIDS), capable of classifying network traffic as normal or malicious based on predefined decision rules. The algorithm constructs a tree-like structure where each internal node represents a decision based on a feature, each branch corresponds to an outcome, and each leaf node signifies a classification label such as normal or attack. Popular Decision Tree algorithms include ID3 (Iterative Dichotomiser 3), which utilizes Entropy and Information Gain, C4.5, which extends ID3 by handling continuous values and missing data, and CART (Classification and Regression Trees), which employs the Gini Index for feature selection.

D.REAL-TIME DEPLOYMENT

Real-Time Data Ingestion: The system must continuously capture network traffic data for analysis. This can be done using tools like:

- **Packet Sniffing Tools**: Wireshark, tcpdump, Zeek (Bro) for real-time network packet collection.
- **NetFlow and sFlow**: Collects summary-based network flow data from routers and switches.
- **Intrusion Detection Systems (IDS) Logs**: Uses IDS tools like Snort or Suricata to preprocess network packets and extract relevant features.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

The collected data is preprocessed in real time, which involves converting categorical attributes into numerical values, normalizing numerical features, and structuring the data in a format compatible with the trained Decision Tree model.

Model Integration for Real-Time Classification: The trained Decision Tree model is integrated into the system to classify network traffic as normal or malicious. The classification process follows these steps:

- Receive real-time network packets or flow records.
- Extract relevant features (e.g., protocol type, connection duration, packet size).
- Preprocess incoming data to match the training dataset format.
- Classify traffic using the trained Decision Tree model.
- Trigger alerts if an intrusion is detected.

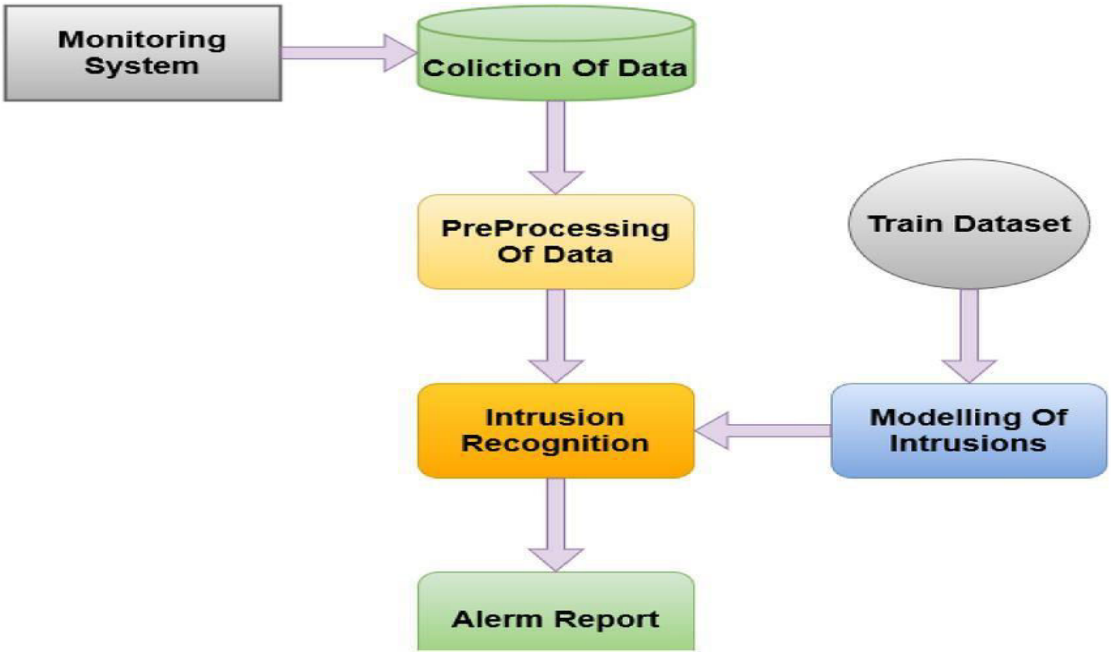

Fig 2: Decision tree algorithm in network intrusion detection

IV. RESULTS AND DISCUSSIONS

Network intrusion detection using the Decision Tree algorithm has shown promising results in identifying and classifying malicious activities in network traffic. The algorithm's ability to create hierarchical, rule-based structures allows for efficient classification of normal and anomalous behavior. Studies indicate that Decision Trees provide high accuracy in detecting known attack patterns while maintaining low false positive rates. However, their performance may degrade when dealing with evolving threats and zero-day attacks, necessitating frequent model updates or integration with ensemble learning techniques. Overall, Decision Tree-based intrusion detection systems are effective in structured datasets, offering a balance between accuracy and computational efficiency in cybersecurity applications.

Network intrusion detection using the Decision Tree algorithm is widely discussed due to its effectiveness in identifying malicious activities within network traffic. The algorithm’s hierarchical structure enables quick classification by breaking down complex decisions into simple, interpretable rules. Researchers highlight its high detection accuracy for known attack patterns, minimal computational overhead, and ease of implementation compared to more complex machine learning models. However, challenges arise when dealing with shisticated cyber threats, such as zero-day attacks and polymorphic malware, which may not follow predefined patterns. Despite these challenges, Decision Tree-based approaches remain a popular choice for network intrusion detection, especially when balancing accuracy, interpretability, and computational efficiency.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

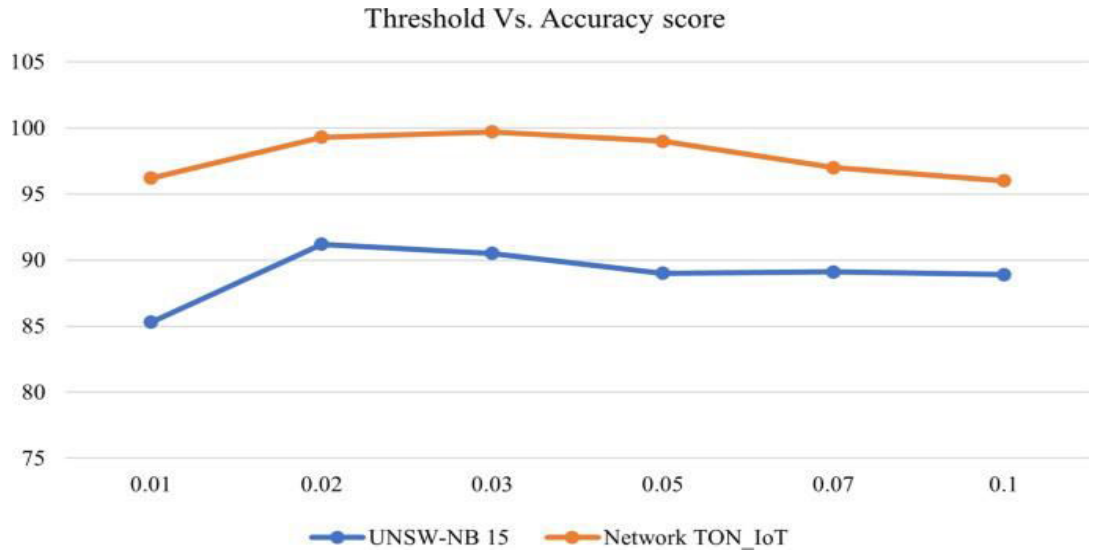

Fig 3: Results on detecting the network intrusion

V. CONCLUSION

Network intrusion detection using the Decision Tree algorithm proves to be an effective approach for identifying malicious activities with high accuracy and efficiency. The algorithm's ability to classify network traffic based on predefined rules enables quick detection of anomalies and cyber threats. Its interpretability and low computational cost make it a suitable choice for real-time intrusion detection systems. However, the model's performance may be influenced by the quality of training data, potential overfitting, and evolving cyber threats. Continuous updates and hybrid approaches integrating other machine learning techniques can further enhance the robustness and adaptability of Decision Tree-based intrusion detection systems.

Additionally, Decision Trees handle categorical and numerical data efficiently, allowing for fast classification of normal and malicious activities. However, despite its advantages, the algorithm faces challenges such as overfitting, especially when dealing with imbalanced datasets or complex attack patterns. Furthermore, as cyber threats evolve, static decision rules may become less effective, necessitating continuous retraining and feature selection to maintain accuracy. Ultimately, while Decision Tree-based intrusion detection systems provide a strong foundation for cybersecurity, their effectiveness depends on regular updates, optimized feature selection, and adaptability to emerging threats in network security.

REFERENCES

[1] Mukherjee, B., Heberlein, L. T., & Levitt, K. N. (1994). Network intrusion detection. IEEE Network, 8(3), 26–41.
[2] Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: applications and issues. Journal of Basic and Applied Sciences, 13(1), 459–465.
[3] Shafiq, M., Yu, X., Laghari, A. A., Yao, L., Karn, N. K., & Abdessamia, F. (2016). Network traffic classification techniques and comparative analysis using machine learning algorithms. In 2016 2nd IEEE International Conference on Computer and Communications (ICCC) (pp. 2451–2455). IEEE.
[4] Sangkatsanee, P., Wattanapongsakorn, N., & Charnsripinyo, C. (2011). Practical real-time intrusion detection using machine learning approaches. Computer Communications, 34(18), 2227–2235.
[5] Farnaaz, N., & Jabbar, M. A. (2016). Random forest modeling for network intrusion detection system. Procedia Computer Science, 89, 213–217.
[6] Ganapathy, S., Yogesh, P., & Kannan, A. (2011). An intelligent intrusion detection system for mobile ad-hoc networks using classification techniques. In Advances in Power Electronics and Instrumentation Engineering: Second International Conference, PEIE 2011, Nagpur, Maharashtra, India, April 21–22, 2011 (pp. 117–122). Springer.
[7] Hussain, J., Lalmuanawma, S., & Chhakchhuak, L. (2016). A two-stage hybrid classification technique for network intrusion detection system. International Journal of Computational Intelligence Systems, 9(5), 863–875.
[8] Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. (2019). A novel hierarchical intrusion detection system based on decision tree and rules-based models. In 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS) (pp. 228–233). IEEE.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

[9] Li, W. (2004). Using genetic algorithm for network intrusion detection. In Proceedings of the United States Department of Energy Cyber Security Group 2004 Training Conference, 1– 8.
[10] Anuar, N. B., Sallehudin, H., Gani, A., & Zakaria, O. (2008). Identifying false alarm for network intrusion detection system using hybrid data mining and decision tree. Malaysian Journal of Computer Science, 21(2), 101–115.
[11] Ramakrishnan, S., & Devaraju, S. (2017). Attack's feature selection-based network intrusion detection system using fuzzy control language. International Journal of Fuzzy Systems, 19(2), 316–328.
[12] Javaid, A. Y., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (pp. 21–26).
[13] Dong, B., & Wang, X. (2016). Comparison deep learning method to traditional methods using for network intrusion detection. In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN) (pp. 581–585). IEEE.
[14] Yash Jadhav, Harsh Deshpande, Ashwini Garole, Komal Sawant, & Aman Patil. (2024). Network Intrusion Detection System Using Decision Tree. Journal Of Network Security, 12(02), 22–33.
[15] Ashwini Garole, Yash Jadhav, Aman Patil, Harsh Deshpande, & Komal Sawant. (2024). Network Intrusion Detection System Using Decision Tree. Journal Of Network Security, 12(02).

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | ijarasem@gmail.com |