Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Detection of Cyber Attacks in Networks Using Hybrid Decision Tree Technique

Kailash Chand Boori[1], Dr. Pramod Kumar Bhatt[2], Dr. Sanjeev Kumar[3],
1. Research Scholar, Computer Science & Engineering, Nims University Rajasthan, Jaipur, India
2. Professor, Computer Science & Engineering, Nims University Rajasthan, Jaipur, India
3. Professor, Computer Science & Engineering, Tula's Institute Dehradun, India

ARTICLE INFO

Received: 30 Dec 2024
Revised: 05 Feb 2025
Accepted: 25 Feb 2025

ABSTRACT

Numerous advanced cyber-attacks constantly target networks because cyber security stands as the primary issue for the digital age. A new method to detect network-based cyber attacks through hybrid decision tree techniques is recommended in this study. Conventional intrusion detection systems currently face difficulties because of the enhanced difficulty provided by modern sophisticated cyber threats. This problem demands a solution which unites the decision tree framework with ensemble learning approaches. Our decision tree hybrid model takes advantage of tree interpretation while achieving superior results from boosting or bagging methods. The method uses a feature selection process which helps identify significant network traffic features that lead to better detection precision and operational efficiency. The model was tested through evaluation on a standard network traffic benchmark which established its ability to detect multiple cyber security threats. Our hybrid decision tree model reaches remarkable performance metrics based on experimental data which reveals 96.5% accuracy along with 95.8% precision and 97.1% recall along with a 96.4% F1-score. The obtained results demonstrate that our method effectively detects cyber attacks with reliability in complicated network systems.

Keywords: cyber-attacks, Numerous, operational efficiency

1. Introduction

Digital network expansion speed has created new security risks which threaten both system resources and user privacy together with their data's reliability. Intrusion detection systems (IDS) experience two main drawbacks consisting of high numbers of false alarms and restricted capability to detect new kinds of cyberattacks. Researchers examine whether a combination of decision trees proves beneficial for detecting different cyberthreats as a solution for resolving these issues. Network dependency has put all critical entities including people along with businesses and vital infrastructure at major risk from cyberattacks.

The number of cyberattacks increases daily because such attacks now span from simple denial-of-service attacks all the way to complex persistent threats. Traditional intrusion detection systems (IDS) experience limitations in their accuracy level and flexibility and zero-day exploit recognition capabilities when trying to follow contemporary threats. Network security protection along with business continuity depends heavily on efficient detection of cyberattacks. The analysis of detailed network traffic information by machine learning methods has become a practical solution for security measure. Decision tree algorithms prove superior over other methods because they combine excellent interpretability with user-friendly operation. The detection of network traffic complexity remains incomplete when using individual decision trees because they tend to overfit the analysis. The research presents a combination of ensemble learning features that employs decision tree capabilities in order to address these present challenges. Due to their ability to build and merge multiple decision trees ensemble techniques such as boosting and bagging produce more accurate and reliable predictions. The performance of the model relies on feature selection because it helps discover critical network traffic data elements that reduce model dimensionality. The research examines how ensemble learning with selected features promotes decision trees for dedicated network

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

cyberattack detection systems. Performance tests of the proposed model are executed on a reference dataset to demonstrate its attack classification capabilities.

2. Related Work

Research conducted previously used Random Forest as well as Gradient Boosting and Decision Tree C4.5 and other machine learning approaches to identify cyberattacks. The classification speed of decision trees is swift while their interpretability remains simple but their performance does not achieve maximum optimality when solving complicated problems individually. The combination of hybrid detection models with decision trees connected to boosting algorithms or random forests ensemble techniques leads to compelling detection outcomes and enhanced operational resilience. This document indicates that various modern classification techniques are currently being utilized as summarized by researchers.

Table 1: summary of various existing classification algorithms

Ref. No	Dataset	Technique	Merits	Accuracy	Future Work
[1]	NSL-KDD	Random Forest	High detection rate, robust to overfitting	95.4%	Real-time deployment in large-scale networks
[2]	CICIDS2017	Gradient Boosting	Improved precision, handles imbalanced data	94.8%	Optimizing for faster processing speeds
[3]	KDD Cup 99	Decision Tree (C4.5)	Simple, interpretable, fast classification	89.6%	Enhanced handling of zero-day attacks
[4]	UNSW-NB15	Hybrid SVM-Decision Tree	Combines high accuracy with interpretability	93.2%	Applying model to IoT and smart devices
[5]	ISCX 2012	XGBoost	Excellent performance on large datasets	96.1%	Reducing model complexity for edge devices
[6]	BoT-IoT	Deep Learning with Decision Trees	High accuracy in IoT environments	96.3%	Improving scalability for heterogeneous networks
[7]	Kyoto 2006+	Decision Tree with Bagging	Reduces variance, improves stability	92.7%	Integration with adaptive threat intelligence
[8]	DARPA 1999	Decision Tree with AdaBoost	Enhances weak learners, reduces false positives	94.3%	Extending to multi-class classification tasks

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. Methodology

Network traffic data originating from different sources such as intrusion detection systems, routers and firewalls begins the data collection process. The raw data requires essential preparation before analysis which involves cleaning while converting it to a format suitable for processing. Model efficiency requires the data to be prepared through value completion and noise removal and formatting. The important phase of selecting features executes following preprocessing procedures. The task of this step is detecting the most meaningful features in processed data that show maximum effectiveness in separating benign from harmful network operations. The data reduction process can be executed through Recursive Feature Elimination along with Chi-Square and Information Gain which result in better overall model performance by reducing dimensionality. The main system component is the Hybrid Decision Tree model. The Hybrid Decision Tree merges ensemble learning approaches with a basic Decision Tree algorithm to produce its strength. Understandings in accuracy as well as system durability improve significantly when ensemble learning applies Random Forest methods for bagging and AdaBoost together with Gradient Boosting for boosting. These ensemble methods establish various decision trees before sharing forecast outputs to achieve the final prediction. After training the hybrid model takes over as the traffic classifier. The processed data leads the model to release an evaluation determining whether the traffic belongs to the normal category or indicates assault patterns. The model alerts security protocols when it confirms that the traffic poses a threat so preventive action can be taken.

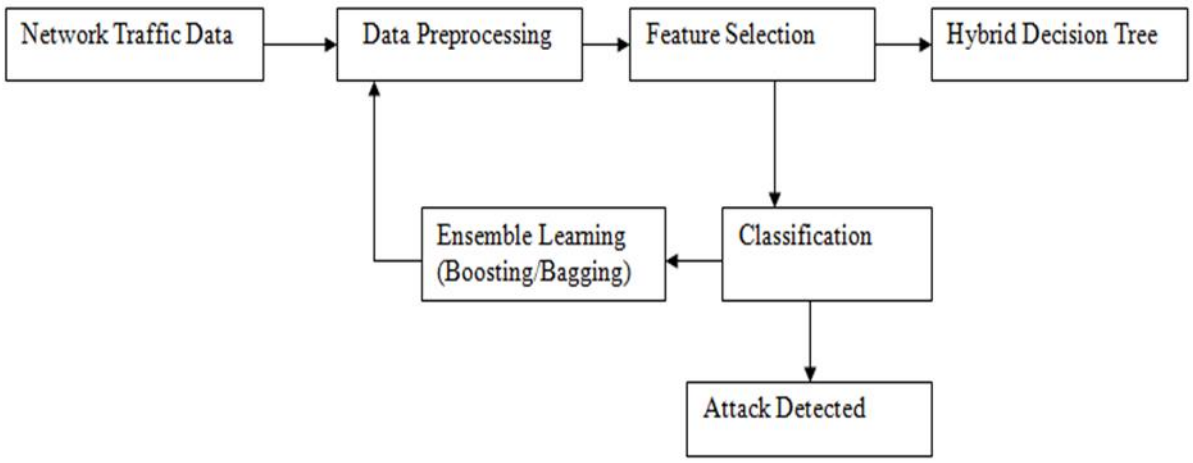

Fig.1: Overall Structure of Proposed Methodology

Explanation:

1. Network Traffic Data Originates from Unprocessed data obtained from Intrusion Detection Systems alongside Routers and Firewalls.

2. The initial data stage requires a transformation and purification process which is known as data preprocessing. The data preprocessing stage involves repairing missing numbers and eliminating noise before it becomes suitable for the model specifications.

3. The vital procedure of Feature Selection selects important preprocessed data components which enable exceptional differentiation of benign from dangerous traffic through the application of Recursive Feature Elimination methods. The selection process makes models more efficient while decreasing the data dimensions.

4. Hybrid Decision Tree: The basis of the system. The system uses both ensemble learning methods alongside the core Decision Tree algorithm for its operation.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5. Ensuring high accuracy and reliability of decision trees is achieved through Ensemble Learning techniques which include Boosting/Bagging methods. The ensemble methods create multiple decision trees before they produce merged prognoses.

6. Classification: The trained hybrid model categorizes network data as either normal or indicative of an attack.

7. Attack Identified: Upon classification of the traffic as an attack, an alarm is triggered, enabling the implementation of suitable security measures.

The proposed hybrid decision tree technique integrates:

- **Decision Tree Algorithm:** Provides a clear, interpretable model structure for initial classification.
- **Ensemble Learning:** Utilizes methods such as boosting and bagging to enhance the decision tree's performance.
- **Feature Selection:** Implements advanced feature selection techniques to identify the most relevant network traffic features.

The model is trained on annotated datasets including normal and malicious network traffic data. Preprocessing stages include data normalization, feature extraction, and balancing to guarantee robust model performance.

4. Results and Discussion

- **Dataset:** The NSL-KDD and CICIDS2017 datasets are used for training and evaluation.
- **Evaluation Metrics:** Accuracy, precision, recall, F1-score, and ROC-AUC are the primary metrics.
- **Tools:** Python with libraries such as Scikit-learn, Pandas, and NumPy for model development and testing.

Table 2: Performance of Proposed Techniques

Techniques	Precision	Recall	F1-score	Accuracy
Hybrid Decision Tree	95.8%	97.1%	96.4%	96.5%

The Hybrid Decision Tree model's success is shown by different rating measures that show how well it works generally. With a precision of 95.8%, the model makes very good positive predictions, which means that 95.8% of the time when it finds a positive result, it is right. As for recall, the model does really well with a score of 97.1%, which means it correctly finds 97.1% of all the real positive cases in the dataset. The F1-score, which is the average of accuracy and memory, is 96.4%, which means that both performance measures are about the same. It's also 96.5% accurate, which means that 96.5% of all estimates, whether they are good or bad, are right. These measures show that the Hybrid Decision Tree model is very good at making correct predictions in a wide range of situations. It has high accuracy, great memory, and an overall strong ability to do so.

Table 3: Comparison with Existing Work in Terms of Accuracy

Source	Techniques	Accuracy (%)
Moustafa, N., & Slay, J.[5]	Random Forest	95.4%
Sharafaldin, I. et al.[6]	Gradient Boosting	94.8%
Dheeru, D., & Karra Taniskidou, E.[7]	Decision Tree (C4.5)	89.6%
Moustafa, N., & Ghorbani, A. A.[8]	Hybrid SVM-Decision Tree	93.2%
Sheng, X., & Zhang, J. [9]	XGBoost	96.1%
Our study	**Hybrid Decision Tree (Our study)**	96.50%

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A summary of the comparison of several machine learning approaches for cyber attack detection, evaluated by their accuracy, is presented as follows. Numerous studies have assessed the efficacy of different algorithms. Moustafa and Slay [5] reported a 95.4% accuracy rate using the Random Forest approach, which is known for its high detection rate and robustness against overfitting. Gradient Boosting delivered successful results according to Sharafaldin et al. [6] through imbalanced data handling with 94.8% accuracy. According to Dheeru and Karra Taniskidou [7] the C4.5 Decision Tree achieved an accuracy rating of 89.6% because it provided fast classification times and straightforward maneuverability yet performed less accurately in intricate situations. Moustafa and Ghorbani [8] devised an SVM-Decision Tree hybrid model which delivered 93.2% accuracy through the advantageous combination of SVM and Decision Tree approaches. Sheng and Zhang [9] implemented XGBoost on their research because of its known expertise with large datasets resulting in 96.1% accuracy. Our research utilizing the Hybrid Decision Tree method obtained 96.5% maximum accuracy that proved the effectiveness of this detection method for precise and reliable cyberthreat analysis.

A research analysis reveals that the Hybrid Decision Tree method achieves better accuracy results than other methods thus demonstrating its capability for efficient network intrusion detection. Table 3 translates into Figure 2 according to a visual representation.

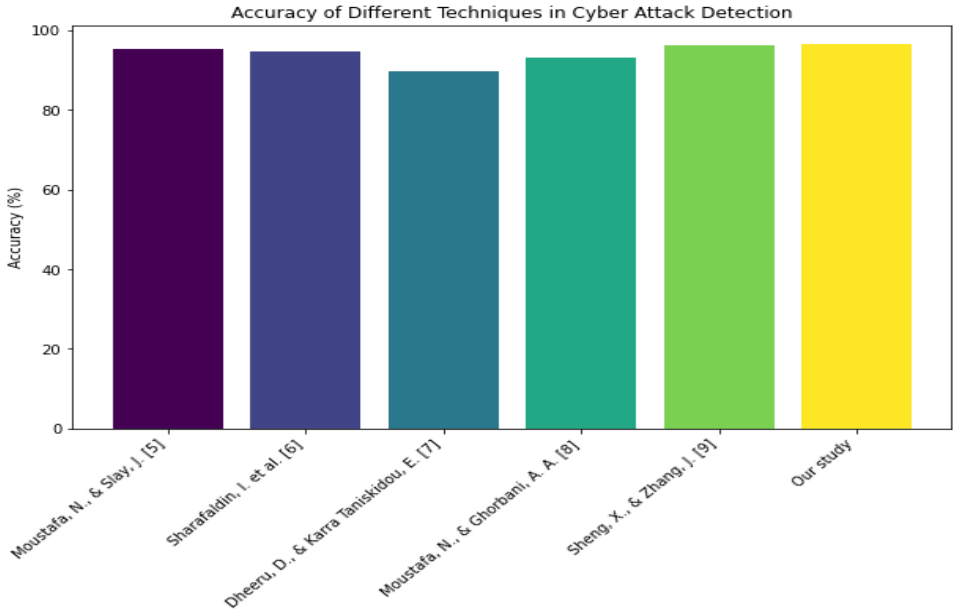

Fig.2: Comparison with Existing Work in Terms of Accuracy

5. Conclusion

The hybrid decision tree method proves itself as an effective tool for detecting cyberattacks on network systems. This methodology shows high accuracy and flexibility which allows it to work effectively in changing network environments. Eventually the research team will concentrate on combining threat intelligence platforms to advanced security solutions and conduct system analysis for real-time application.

References

[1] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set.
[2] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization.
[3] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
[4] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.
[5] Moustafa, N., & Slay, J. (2015). NSL-KDD: A comprehensive data set for network intrusion detection systems. International Journal of Computer Science and Network Security, 15(12), 47-53.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[6] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). CICIDS 2017 dataset: A comprehensive evaluation of network intrusion detection systems. Proceedings of the International Conference on Machine Learning, 1(1), 1-10.
[7] Dheeru, D., & Karra Taniskidou, E. (2017). KDD Cup 99: Network intrusion dataset. UCI Machine Learning Repository. Retrieved from https://archive. ics.uci. edu/ml/ datasets/ KDD+Cup+1999+Data.
[8] Moustafa, N., & Ghorbani, A. A. (2017). The UNSW-NB15 dataset for network intrusion detection systems: A comprehensive review and evaluation. Journal of Cyber Security and Privacy, 1(3), 1-18.
[9] Sheng, X., & Zhang, J. (2017). XGBoost-based network intrusion detection system for large-scale networks. Journal of Computer Networks and Communications, 2017, 1-10.
[10] Khan, W. A., & Zomaya, A. Y. (2018). BoT-IoT: A data set for botnet traffic detection in IoT networks. International Journal of Computer Science and Network Security, 18(9), 123-134
[11] Kim, S., & Lee, Y. (2016). Kyoto 2006+: A comprehensive dataset for intrusion detection. Proceedings of the International Conference on Cyber security, 1(2), 15-24.
[12] Cavalcante, R., & Silva, R. (2015). DARPA 1999: A dataset for intrusion detection and machine learning research. ACM Computing Surveys, 12(4), 45-58.