International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

A Malicious Bot Traffic Detection Method in IoT Network Using Machine Learning Techniques

Ms. Surabhi K S[1], Thivagaran P[2]

[1]Assistant professor, Department of Computer Applications, Nehru College of Management, Coimbatore, TamilNadu, India
[2]Student of II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, TamilNadu, India

ABSTRACT:

In the context of growing Internet of Things (IoT) technology, this project utilizes the Bot-IoT dataset for detecting malicious traffic. A feature selection technique is proposed to improve the performance of machine learning algorithms for detecting attacks in IoT network traffic. The system implements machine learning algorithms such as Support Vector Machine (SVM) and Naive Bayes (NB). The experimental results show that the proposed method is highly efficient, achieving over 98% accuracy on average for each algorithm

INTRODUCTION

Effective feature selection and accurate identification of Bot-IoT attacks in IoT network environments, a newly developed dataset is used. This dataset includes both normal IoT traffic and various cyber-attacks, specifically botnet attacks.

As IoT technology rapidly grows, numerous devices are being connected every minute. Botnets, which are networks of multiple bots designed to perform malicious activities on a target network, are controlled by a single entity known as the botmaster. Detecting such malicious activities in IoT networks is critical for network security

RELATED WORKS

Various studies in IoT security and access control, particularly in detecting and mitigating cyber threats in IoT environments. Some notable works cited are:

- **J. Qiu et al. (2020)**: A survey on access control in the age of the Internet of Things (IoT).
- **Y. N. Soe et al. (2019)**: Implementation of lightweight IoT-IDS (Intrusion Detection System) using correlation-based feature selection and its performance evaluation on Raspberry Pi.
- **J. Qiu et al. (2019)**: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart cities.
- **J. P. Anderson (1980)**: Computer security threat monitoring and surveillance.
- **L. Wu et al. (2018)**: An out-of-band authentication scheme for IoT using blockchain technology.
- **Z. Tian et al. (2020)**: Vcash, a reputation framework for identifying denial of traffic services in Internet-connected vehicles.

These works discuss different techniques and approaches to enhance IoT security, feature selection, and machine learning applications for detecting malicious traffic

METHODOLOGY

Detecting malicious traffic in IoT networks using machine learning techniques. It includes the following processes:

1. **Data Selection**:
   - The Bot-IoT dataset is used, which includes normal IoT traffic and various botnet attack traffic flows.
   - The dataset's features are labelled based on attack categories and subcategories to enhance detection.
2. **Pre-processing**:
   - Unwanted data such as missing values are removed or replaced with appropriate values (e.g., NaN values are replaced with 0).
   - Categorical data is encoded into a format suitable for machine learning.
3. **Feature Selection**:
   - Two feature selection methods are employed: Correlation and Chi-Square.
   - These methods help reduce the number of input variables, reducing computational overhead and improving the predictive model's efficiency.
4. **Data Splitting**:
   - The dataset is split into training (70%) and testing (30%) sets.
   - This division allows for model training and subsequent evaluation on unseen data.
5. **Classification**:
   - Two machine learning algorithms are applied:
     - **Naive Bayes (NB)**: A supervised learning algorithm based on Bayes' Theorem.
     - **Support Vector Machine (SVM)**: Uses a kernel trick to transform data and find an optimal decision boundary.
6. **Performance Analysis**:
   - The performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score to measure the effectiveness of the proposed method

ALGORITHMS

In this project, two machine learning algorithms, **Naive Bayes (NB)** and **Support Vector Machine (SVM)**, are used for classifying and detecting malicious traffic in IoT networks. Here's how they are applied:

1. **Naive Bayes (NB)**:
   - **Use**: Naive Bayes is a probabilistic classifier that applies Bayes' Theorem, assuming that the features are conditionally independent. In this project, it is used to classify network traffic into malicious or benign categories.
   - **Why NB?**: The algorithm is fast, works well with large datasets, and is particularly effective when the assumption of feature independence holds. It helps in situations where computational efficiency is essential, such as in IoT networks with large volumes of traffic data.
2. **Support Vector Machine (SVM)**:
   - **Use**: SVM is used to find the optimal hyperplane that separates different classes of traffic data (malicious vs. benign) by maximizing the margin between classes. It's particularly useful for handling high-dimensional datasets and cases where the classification boundary is not linear.

- **Why SVM?**: SVM is known for its accuracy in classification tasks and its ability to handle both linear and non-linear relationships in data. In the context of detecting Bot-IoT attacks, SVM helps provide a more precise classification, especially when the traffic patterns are complex and overlapping.

Both algorithms are chosen to complement each other in their ability to handle the dataset's diversity and ensure efficient, accurate detection of malicious traffic, achieving over 98% accuracy in the project

DATA SET

The **dataset** used in this project is the **Bot-IoT dataset**, which is specifically designed for detecting malicious traffic in Internet of Things (IoT) networks. Key characteristics of the dataset include:

- **Content**:

  The dataset consists of both normal IoT traffic and several types of malicious traffic, particularly **botnet attack flows**. It includes various types of cyber-attacks, making it ideal for training machine learning models to detect anomalies.

- **Labeled Data**:

  The features in the dataset are labeled according to attack categories and subcategories, enabling precise identification of different types of attacks.

- **Size**:

  As the IoT network grows rapidly, the dataset is designed to handle a large number of devices and their traffic flows, making it scalable for real-world applications.

- **Variety of Attacks**:

  The Bot-IoT dataset includes a wide variety of attacks such as Denial of Service (DoS), Distributed Denial of Service (DDoS), data theft, and reconnaissance, providing a comprehensive foundation for training ML models to detect multiple forms of malicious traffic.

In this project, the dataset is split into training (70%) and testing (30%) sets, enabling the evaluation of model performance on unseen data

Timestamp	Source IP	Destination IP	Protocol	Source Port	Destination Port	Packet Size	Label
2023-09-25 10:00:00	192.168.1.10	192.168.1.20	TCP	49152	80	1500	Normal
2023-09-25 10:00:01	192.168.1.10	192.168.1.25	UDP	49153	53	100	Malicious
2023-09-25 10:00:02	192.168.1.15	192.168.1.30	TCP	49154	443	1200	Malicious
2023-09-25 10:00:03	192.168.1.20	192.168.1.10	ICMP	N/A	N/A	64	Normal

RESULT

The effectiveness of the proposed methodology for detecting malicious traffic using the Bot-IoT dataset and the implemented machine learning algorithms. Here are the key findings:

- **Accuracy**: The proposed method achieved over **98% accuracy** on average for both machine learning algorithms (Naive Bayes and Support Vector Machine).
- **Performance Metrics**: Various performance metrics were evaluated, including:
  - **Precision**: Indicates the proportion of true positive results in the predicted positive instances.
  - **Recall**: Measures the ability of the model to identify all relevant instances (true positives) in the dataset.

- **F1-Score**: A harmonic mean of precision and recall, providing a balance between the two.

- **Efficiency**: The feature selection techniques employed (correlation and chi-square) contributed to improved performance by reducing the number of input variables, leading to faster training times and better model generalization.
- **Comparison with Existing Methods**:

  The experimental results indicate that the proposed approach outperformed existing systems in terms of detection accuracy and efficiency, demonstrating its potential for practical applications in IoT security

Metric	Naive Bayes	Support Vector Machine
Accuracy	98.2%	98.5%
Precision	97.5%	98.7%
Recall	96.8%	97.9%

DISCUSSION POINTS

1. **Effectiveness of Feature Selection**:
   - The use of feature selection techniques, such as correlation and chi-square, significantly improved the model's performance. By reducing the number of input variables, the models were able to focus on the most relevant features, enhancing both accuracy and efficiency.
2. **Comparative Performance**:
   - The results indicate that both Naive Bayes and Support Vector Machine algorithms performed exceptionally well, achieving accuracies above 98%. This suggests that either method can be effectively employed in real-world scenarios for malicious traffic detection.
   - While SVM showed slightly better results, Naive Bayes provided a good balance between performance and computational efficiency, making it suitable for resource-constrained IoT environments.
3. **Implications for IoT Security**:
   - The high detection accuracy of the proposed models underscores the importance of machine learning techniques in securing IoT networks against evolving threats. As botnet attacks become more sophisticated, the ability to detect them effectively is critical for maintaining network integrity.
4. **Challenges and Limitations**:
   - Despite the promising results, there are challenges in scaling these models to accommodate the vast amounts of data generated by IoT devices. Future work may focus on optimizing algorithms for even larger datasets.
   - Additionally, the performance might vary with different types of attacks or in different network conditions, suggesting the need for continuous model training and updating.
5. **Future Work Directions**:
   - Further research could explore the integration of other machine learning algorithms or ensemble methods to enhance detection capabilities.
   - Investigating real-time detection systems could also be beneficial, allowing for proactive measures against malicious activities.

CONCLUSION OF DISCUSSION

Overall, the findings from this study validate the effectiveness of the proposed methodology in detecting malicious IoT traffic and highlight the critical role of machine learning in enhancing IoT security

REFERENCES

1. **J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su, and B. Fang**: "A survey on access control in the age of internet of things," IEEE Internet of Things Journal, 2020.
2. **Y. N. Soe, Y. Feng, P. I. Santosa, R. Hartanto, and K. Sakurai**: "Implementing lightweight IoT-IDS on Raspberry Pi using correlation-based feature selection and its performance evaluation," in International Conference on Advanced Information Networking and Applications. Springer, 2019, pp. 458–469.
3. **J. Qiu, L. Du, D. Zhang, S. Su, and Z. Tian**: "Nei-tte: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city," IEEE Transactions on Industrial Informatics, 2019.
4. **J. P. Anderson**: "Computer security threat monitoring and surveillance," 1980.
5. **L. Wu, X. Du, W. Wang, and B. Lin**: "An out-of-band authentication scheme for internet of things using blockchain technology," in 2018 International Conference on Computing Networking and Communications (ICNC). IEEE, 2018, pp. 769–773.
6. **Z. Tian, X. Gao, S. Su, and J. Qiu**: "Vcash: A novel reputation framework for identifying denial of traffic service in internet of connected vehicles," IEEE Internet of Things Journal, vol. 7, no. 5, pp. 3901–3909, May 2020.
7. **R. Xue, L. Wang, and J. Chen**: "Using the IoT to construct a ubiquitous learning environment," in 2011 Second International Conference on Mechanic Automation and Control Engineering. IEEE, 2011, pp. 7878–7880.