The Analysis and Visualization of Network Security Assessment Data

Li Yuxuan

Xi'an Jiaotong University

Abstract: The proliferation of the Internet and network communication technologies has revolutionized information access and global commerce. However, this has also led to escalating network security challenges that threaten the integrity and privacy of online activities. This study aims to enhance network security by developing an AI-driven intrusion detection system with a focus on network security visualization. We introduce a rule-based detection method for PortScan and DDoS attacks using the CIC-IDS2017 dataset. By analyzing attack patterns and dataset features, we establish classification rules to accurately identify and predict these threats. Additionally, we employ a decision tree algorithm to train a model for flow classification, utilizing random sampling for model prediction. Our system uses the InfluxDB time-series database for data storage and the Grafana tool for visualizing network security events. The model achieved an accuracy of 95% in detecting network anomalies, significantly improving upon traditional methods. The visualization component provides an intuitive understanding of network threats, aiding in swift decision-making for threat mitigation. In conclusion, our research contributes a novel approach to network security by integrating AI algorithms and visualization techniques, offering a robust and efficient solution for network administrators to safeguard against sophisticated attacks.

Key Words:network security; CIC-IDS2017; InfluxDB; Grafana; AI; visualization;

Introduction

With the widespread application of internet technology, cybersecurity issues have become increasingly prominent. The diversification of cyber-attacks poses threats to national and public interests. To address this challenge, cybersecurity visualization technology has emerged. By presenting complex network data in a graphical manner, it assists analysts in quickly identifying abnormal behaviors and potential attack patterns, enhancing the accuracy and efficiency of network anomaly detection, and reducing false positives and false negatives, thereby strengthening cybersecurity defense capabilities. This research aims to explore the relevant theories and technologies of cybersecurity visualization systems, with the expectation of contributing to the improvement of cybersecurity defense levels and the protection of national and public interests.

Related Technologies

1 Common Attack Methods

Common attack methods in the field of cybersecurity include port scanning, denial of service (DoS & DDoS), brute force attacks, browser attacks, botnets, and infiltration attacks. These attacks can lead to anomalies in network traffic, posing a threat to cybersecurity. The CICIDS2017 dataset is used to simulate these attack behaviors, providing an understanding of the mechanisms of network attacks and how to use this knowledge for effective cybersecurity protection.

2 Security Assessment Technologies

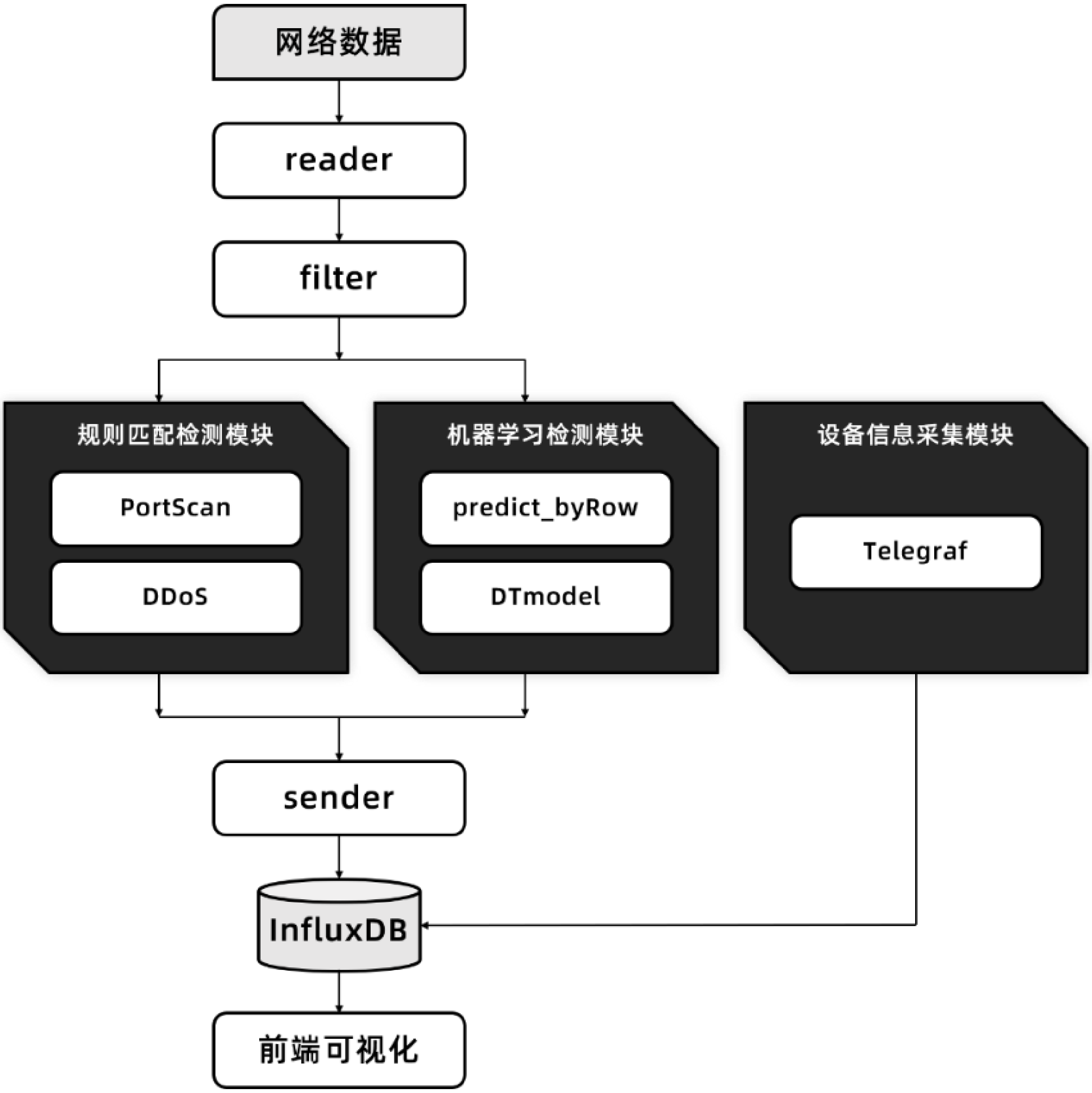
There are four major technologies for cybersecurity assessment: anomaly detection based on statistics, rules, machine learning, and data mining. Statistical methods discover anomalies by generating profiles and comparing data; rule-based methods match data based on predefined attack signatures; machine learning identifies anomalies by learning the characteristics of normal behavior; and data mining is used to uncover potential abnormal patterns. Each method has its strengths and limitations, and they are often combined in practical applications to improve detection accuracy and coverage.

3 Visualization Technologies

Visualization technologies in cybersecurity include Telegraf, InfluxDB, and Grafana. Telegraf is an agent for metric collection and reporting, supporting various input plugins, and can integrate with InfluxDB to simplify data stream management. InfluxDB is a high-performance time-series database, focused on storing time-series data, with the ability to organize data using tags and fields, and supports data retention policies. Grafana is an open-source data visualization and monitoring platform, supporting connections to multiple data sources, providing a rich selection of visualization options and a flexible data query language, with alert and notification features, as well as user access control and authentication mechanisms. These tools together form a powerful cybersecurity data visualization and monitoring system.

Attack Detection and Visualization

1 System Architecture



2 Intrusion Detection Based on Rule Matching

PortScan detection involves analyzing the scanning frequency of target ports and packet characteristics, such as abnormal values of Total Fwd Packets and Fwd Header Length. DDoS detection is based on metrics related to data volume, such as Total Fwd Packets, Total Backward Packets, as well as abnormal patterns in Flow Duration and Destination Port. These methods help identify and defend against network attacks.

3 Intrusion Detection Based on Machine Learning

This study employs a decision tree algorithm to classify traffic data, using 70% of the data for training and 30% for testing. The results show that the decision tree performs best on the CICIDS2017 dataset. The test results indicate that the detection rate of the decision tree algorithm is 91.9%, with a false positive rate of 2.91%. Subsequent tests will continue to use this machine learning algorithm.

4 Visualization Methods

Visualization methods facilitate an intuitive understanding of cybersecurity data. The visualization of PortScan attacks includes the scanning IP, scanned ports, total number of ports, packet length ratio, packet interval time ratio, and connection duration. The visualization of DDoS attacks involves the attacking source IP, attacked ports, total number of attack links, average connection time, proportion of active connection time, and packet information. For other types of attacks in the CICIDS2017 dataset, the visualization displays the attacking IP and port, attacked ports, number of attack links, total attack traffic, and average attack rate per minute, in order to more effectively identify and analyze different types of network attacks.

System Design and Implementation

1 Workflow

The system design is divided into data preprocessing, model training, statistical analysis, and database visualization. Data preprocessing includes merging, cleaning, and encoding the CICIDS2017 dataset. Model training generates and saves the model files. For DDoS and PortScan attacks, statistical analysis methods are used to identify and label them. Finally, the classified data is written into the database and displayed through visualization software.

2 Data Processing

Data preprocessing includes cleaning, label numericalization, feature selection, data normalization, and One-hot encoding. Cleaning removes "Infinity" and "NaN" values and selects files with balanced data volumes. Label numericalization converts string labels to numbers. Feature selection identifies 65 key features through recursive feature elimination and a random forest classifier. Data normalization ensures consistent feature distribution. One-hot encoding converts discrete labels into binary vectors. Model training uses a decision tree with information entropy as the splitting criterion and a maximum depth of 12, achieving an accuracy of 99.77% on the test set.

3 Machine Learning Detection

Machine learning detection involves the generation of data to be predicted and the prediction of model data. It randomly selects 100,000 rows of data from the CICIDS2017 dataset, which only includes the seven attack types that the model has learned. The prediction part first reads the data and the model, processes each row of data, discards features not suitable for machine learning, selects features from the feature list for prediction, converts One-Hot encoding back to original attack labels, and outputs the results as `predicted.csv`.

4 Test Results

The database status shows the four databases in InfluxDB and their data volumes. The result visualization, configured through the Grafana interface, displays the visualization results for attack types such as PortScan and DDoS. Real-time data transmission tests confirm that Grafana can receive and display attack data in real-time. The accuracy evaluation and analysis table shows the detection accuracy rates for different types of attacks, where the overall accuracy rate of machine learning methods is higher, while the accuracy rates of rule-matching methods are lower for DDoS and PortScan detection.

Conclusion

This paper completes the design and implementation of a cybersecurity assessment and visualization system. Firstly, the CICIDS2017 dataset was selected, and the system was developed using Python. During the data processing stage, database connection timeout issues were resolved, and a decision tree model was used for attack type prediction. The system utilizes Telegraf, InfluxDB, and Grafana to build a data visualization platform. Looking forward, it is planned to adopt real-time network traffic prediction and deep learning techniques to improve the system's prediction accuracy for new data and real traffic.