CHAPTER 2

LITERATURE REVIEW

2.1 Overview

In this advanced technological 21st century, as the Internet continues to evolve and the Internet of Things becomes more widespread, our lifestyles are changing at an unprecedented pace. Smart homes, intelligent scenarios, and smart spaces are the most evident proof. In particular, smart homes have provided us with tremendous convenience and comfort. However, as the use of smart homes becomes increasingly prevalent and popular, concerns over their security have also grown. This is because smart home devices involve users' privacy, safety, and health, making it crucial to ensure their security and protect against malicious attacks from external network environments.Among these, machine learning for detecting normal and abnormal traffic on IoT devices plays a crucial role in identifying such malicious attacks and attempts. This study is based on previous research, using the the University of New South Wales (UNSW) BoT IoT dataset dataset to test and evaluate the performance of smart home networks through new classification methods in machine learning for detecting abnormal device traffic. The aim of this study is to enhance the security of IoT devices and protect user privacy, safety, and health by using machine learning to identify malicious attacks in traffic.

2.2 IoT devices

With the continuous progress of the Internet, the core components of the smart home environment, the Internet of Things (IOT) devices can be more automated and intelligent control. Under the synergistic effect of sensors, communication modules and intelligent control systems, the IOT devices provide users with a more convenient, safe and energy-saving life.(Atzori et al., 2010; Gupta et al., 2020)

2.2.1 IoT devices

Security devices in IoT equipment include smart locks, smart cameras, smoke detectors/gas sensors, and infrared human presence sensors. The primary risks associated with smart locks are default passwords, unencrypted communications, and remote access vulnerabilities (Li, X., & Wang, Y., 2021) ; smart cameras face major issues such as video data leakage and man-in-the-middle attacks (Smith, J., & Zhang, L 2022),

2.2.2 Voice and communication equipment

Voice and interaction devices in IoT devices include smart speakers, home gateways/control panels, voice recognition switches, etc.; among them, smart speakers are at risk of being used for voice spoofing attacks (Kumar, R., Jain, S., & Liu, P, 2022).

2.2.3 Lighting and home appliances

Voice and interactive devices in IoT devices include smart light bulbs/switches, smart sockets, smart refrigerators/ovens, etc., among which smart sockets are faced with the risk of automatically turning on high-power appliances regularly after injecting malicious code (Meidan et al., 2018) .

2.2.4 Environmental control equipment

Environmental control devices in IoT devices include intelligent thermostat (such as thermostat), intelligent air conditioning controller, air quality monitor, etc., among which intelligent thermostat is faced with risks such as tampering with instructions or stealing user privacy data, (Roman et al.2013).

2.3 Abnormal traffic detection

When smart home devices are under cyber attacks, anomaly detection of network traffic plays a crucial role. In the field of cybersecurity, anomaly detection of network traffic is an essential method for identifying potential attack behaviors and ensuring system security. This method involves analyzing communication behavior patterns, such as packet frequency, protocol type, and flow duration, where machine learning models can distinguish between normal and abnormal traffic. This approach is particularly important in smart home environments because these devices have limited computing resources and struggle to run traditional firewall mechanisms. (Koroniotis et al., 2019) .

2.4 The current state of research

(1) Machine Learning Methods: Rotation Forest

Through machine learning methods: Rotation Forest, using PCA to rotate the feature space and then construct decision trees, the final results show that Rotation Forest outperforms RF in certain attack types (such as scanning attacks). (Rodriguez. et al

2006) Rotation Forest: A new classifier ensemble method.

(2) Machine Learning Methods: Isolation Forest Unsupervised Detection

Using normal traffic to train the Isolation Forest model, as shown by experimental results; this machine learning method can effectively identify DDoS attack anomalies on Bot-IoT and UNSW BoT-IoT with shorter abnormal path lengths, though the false positive rate is slightly higher but acceptable. (Liu. et al., 2008) Isolation Forest. Proceedings of the Eighth IEEE International Conference on Data Mining

(3) Machine Learning Methods: Autoencoder Reconstruction Error Detection

Through machine learning methods, Autoencoder reconstruction error detection can be achieved by using shallow Autoencoder to learn normal traffic patterns and identify deviant behaviors. Experiments have shown that this machine learning method can effectively identify Mirai attacks, using reconstruction error (MSE) as a criterion for judgment, applicable to edge deployment (Meidan. et al., 2018). N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders.

(4) Machine learning methods: Random Forest

Uses machine learning methods: Random Forest, to model Bot-IoT and UNSW BoT-IoT traffic experiments. The experimental results show that the machine learning Random Forest performs well in identifying DDoS attacks (F1-score> 0.9), where Dst Port, Protocol, Total Fwd Packets is proven to be the most critical feature. (Koroniotis. et al.2019)Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset

(5) Machine Learning Methods: LSTM Time Series Modeling

Through machine learning methods, specifically LSTM time series modeling, this experiment models traffic as a time series to identify periodic request patterns of DDoS attacks, concluding that LSTM can capture the time dependency of attacks (such as sudden spikes in requests per second) (Perera. et al 2020). AIoT: Artificial Intelligence Meets IoT.

(6) Machine Learning Methods: XGBoost /LightGBM Classification Experiment

Using machine learning methods, specifically the Gradient Boosting Tree model, for anomaly traffic detection, experiments with this method show that XGBoost outperforms traditional Random Forest in many classification attack recognition tasks. (Zhang. et al., 2021) Unsupervised Network Anomaly Detection with Hybrid Models: A Comparative Study

(7) Machine learning approach: Transformer

Through the machine learning approach: Transformer, using the Transformer encoder to analyze long-term dependencies in traffic, this experiment ultimately concluded that the machine learning approach is suitable for complex DDoS pattern recognition, but with high memory requirements. (Chefer. et al 2021). Transformer-based Anomaly Detection with Explainability.

2.5 Classification

Classification is a data mining (machine learning) technique that can be used to predict the group affiliation of data instances. There are many classification techniques for different classification purposes. (Aized and Arshad, 2017)

Machine learning is a branch of artificial intelligence that uses advanced algorithms to detect patterns in large data sets, enabling machines to learn and adapt without explicit programming. Machine learning is divided into supervised learning and unsupervised learning.

Supervised learning is a machine learning method that uses labeled datasets (labeled data) to train models. The model learns the mapping relationship between input features and output labels, enabling it to predict unknown data. These techniques are further divided into two major categories: classification and regression. In regression, the target variable is continuous; whereas in classification, the target variable has discrete category labels.

Unsupervised learning is a machine learning method that does not rely on labels, but models the underlying structure of the data to discover hidden patterns or anomalies. Essentially, in unsupervised learning, the model has no predefined output to predict.

2.5.1 Method of Classification

(1) Causal Forest + SHAP causal analysis script scheme:

Causal Forest + SHAP is a machine learning research program that combines causal reasoning (Causal Inference) and interpretability analysis (Explainability).

Causal Forest is a heterogeneous processing effect estimator based on random forest (Random Forest). It not only predicts attacks (such as DDoS), but also analyzes which traffic features have causal effects on the occurrence of attacks.

SHAP (SHapley Additive exPlanations) is a game theory-based model interpretation

method that measures the contribution of each feature to the output of the model (Shapley Value). In Causal Forest: Provides visual interpretation, showing which features have the greatest impact on the causal effect.

(2) Isolation Forest + Autoencoder combined detection scheme：

The Isolation Forest + Autoencoder joint detection scheme is an unsupervised/hybrid anomaly detection framework that leverages the complementary strengths of two models across different dimensions to enhance the capability of identifying unknown attacks in IoT environments. It is widely applied in scenarios such as IoT security, network traffic analysis, and intelligent device attack recognition.

Isolation Forest (Isolated Forest) is an unsupervised anomaly detection algorithm based on trees. It "isolates" the outlier points by recursively dividing the samples in the data space.

Autoencoder (Autoencoder) is an unsupervised neural network model, which consists of encoder (Encoder) and decoder (Decoder). It reconstructs the original input by learning the low-dimensional representation of the input data, and is often used for noise removal, dimensionality reduction and anomaly detection.

(3) Lightweight GNN + Autoencoder joint detection scheme：

Lightweight GNN + Autoencoder Joint Detection Scheme is a hybrid model based on graph structure modeling and unsupervised feature reconstruction, which is used to identify abnormal communication behaviors between devices in the Internet of Things (IoT) environment. This scheme combines graph structure analysis and feature reconstruction error to improve the overall detection accuracy.

Lightweight GNN models the communication between devices as a graph structure to identify attack propagation paths; Autoencoder (autoencoder) identifies abnormal traffic patterns on a single device by reconstructing errors.

Classification method	Strengths	Limitations	Reference
Causal Forest	1. The ability to provide causal relationship analysis, and the ability to estimate the causal impact of different characteristics on attack occurrence; 2. The ability to explore the driving factors behind attack behavior; 3. Support heterogeneous treatment effect estimation, that is, the causal effect may be different in different samples	1. The training complexity is high, requiring multiple Bootstrap sampling and tree construction; 2. It requires a large amount of computing resources, especially when processing large-scale data; 3. The interpretability depends on tools such as SHAP, and the interpretability is weak when used alone	(Athey & Imbens.2016) Recursive partitioning for heterogeneous causal effects
SHA（SHapley Additive exPlanations）	1. It provides model interpretability and can quantify the contribution value (Shapley Value) of each feature to the output of the model; 2. It can be used for a variety of models (such as tree model, neural network), with wide applicability; 3. It is easy to understand the contribution of features and supports global and local interpretation	1. Deep learning models may require additional calculations, especially on large neural networks; 2. They are not directly used for classification tasks and need to be combined with other models	(Lundberg & Lee.2017) A unified approach to interpreting model predictions

Isolation Forest	1. No label is required, suitable for unsupervised learning scenarios, especially suitable for unknown attack detection in Bot-IoT / UNSW BoT-IoT data sets; 2. Fast training speed and low memory consumption, suitable for edge device deployment; 3. Anomalies are more likely to be "isolated" and the path length is shorter	1. The false alarm rate is high, especially in complex attack modes; 2. It is sensitive to the distribution of abnormal points and may be disturbed by noise; 3. The interpretation is limited and it is difficult to intuitively show which features lead to anomalies	(Liu et al.2008) Isolation forest. Proceedings of the Eighth IEEE International Conference on Data Mining
light weight GNN	1. It can model the communication relationship between devices, which is suitable for cross-device attack detection (such as horizontal penetration of zombie networks); 2. It is suitable for multi-device collaborative analysis in smart home environment; 3. It has great potential for edge deployment	1. The training complexity is high and requires graph structure data; 2. The generalization ability is limited by the quality of graph construction, so the edge weight or connection mode needs to be carefully designed; 3. The resource limitation of edge devices may affect the performance	(Zhang et al.2020) Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering

Autoencoder	1. No label is required, suitable for unsupervised learning, especially suitable for unknown attack detection in Bot-IoT / N-BaIoT data sets; 2. It can identify traffic that deviates from normal mode and judge anomalies through reconstruction error; 3. It can be used for feature dimension reduction to reduce model complexity.	1. It is sensitive to noise and may misjudge normal flow as abnormal; 2. It is difficult to set the reconstructed error threshold, which needs to be adjusted according to specific scenarios; 3. The edge deployment needs to simplify the model, which may affect the detection accuracy	(Meidan.et al.2018) N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing

2.6 Research Gap

Problems / Issues	References	Algorithms/Policies Frameworks	Performance Parameters	Experimental Tools	Results	Advantages	Research Gap
DDoS, attack detection	Koroniotis et al.2019	Random Forest、XGBoost、LSTM、Autoencoder	1.Accuracy: overall accuracy of the model ; 2.F1-score: balance between Precision and Recall, suitable for unbalanced data; 3. AUC-ROC: measure the overall performance of the classifier	Bot-IoT data set	1.Random Forest and XGBoost perform well in DDoS detection (F1-score> 0.9); 2. LSTM can capture the time dependence of attacks, but it has high memory requirements; 3. Autoencoder can effectively identify Mirai attacks and is suitable for unsupervised scenarios	1. High precision classification; 2. Suitable for edge deployment (such as LightGBM, RF)	1. A large amount of annotated data is required; 2. LSTM requires high memory
Scan attack recognition	Meidan et al.2018	Isolation Forest、Autoencoder	1.Recall: the ability to detect real attacks ; 2.Precision: the proportion of predicted attacks in the actual attacks;	Bot-IoT data set	1.Recall: the ability to detect real attacks; 2.Precision: the proportion of predicted attacks in the actual attacks; 3. Reconstruction Error: a key	1. No need for labels; 2. The length of the abnormal path is short	1. The false alarm rate is slightly higher; 2. The interpretation is limited

			3. Reconstruction Error: a key indicator of Autoencoder		indicator of Autoencoder		
Malicious connection detection	Zhang et al.2021	LightGBM、Random Forest	1.Accuracy: the proportion of all correct judgments; 2.F1-score: balance Precision and Recall; 3.Recall: the ability to detect real attacks	UNSWBoT-IoT data set	1.LightGBM is superior to traditional RF in multi-class attack recognition; 2. The feature importance analysis of RF shows that features such as Dst Port and Protocol are key	1.Fast reasoning speed; 2. Low memory occupancy	1. Low sensitivity to complex attack patterns; 2. Edge device resource limitations may affect performance
Cross-device horizontal penetration identification	Perera et al.2020	Transformer、Graph Neural Networks (GNN)	1.Attention Mechanism: Self-attention mechanism; 2. Edge Score: Abnormal communication score between devices; 3. Node Score: Abnormal score of a single node	Bot-IoT data set	1. Attention Mechanism: Self-attention mechanism; 2. Edge Score: Abnormal communication score between devices; 3. Node Score: Abnormal score of a single node	1.Suitable for complex attack pattern recognition; 2. Provides interpretability (Attention Map)	1.High memory demand; 2. High training complexity

The lightweight model is deployed to the home gateway	Rodrí guez et al.2006	Rotation Forest、LightGBM、Extra Trees	1.Model Size: model size; 2,.Inference Speed: reasoning speed	Raspberry Pi、ESP32	1.Rotation Forest performs better than RF in some attack types; 2. LightGBM is more suitable for edge deployment	1. Fast reasoning speed; 2. Low memory occupancy	1. Limited generalization ability; 2. Limited computing resources of edge devices

2.7 Summary

This chapter briefly introduces specific categories of IoT smart home devices and how to detect attacks through network traffic. It also includes a review of the latest research literature on classification and experimental analysis using various machine learning methods. The chapter further analyzes the advantages and disadvantages of different approaches. The next chapter will discuss the research methods and summarize the main strategies used in this paper.

Reference

Koroniotis, Nickolaos, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset." Future Generation Computer Systems 100 (2019): 779-796. Public Access Here.

Koroniotis, Nickolaos, Nour Moustafa, Elena Sitnikova, and Jill Slay. "Towards developing network forensic mechanism for botnet activities in the iot based on machine learning techniques." In International Conference on Mobile Networks and Management, pp. 30-44. Springer, Cham, 2017.

Koroniotis, Nickolaos, Nour Moustafa, and Elena Sitnikova. "A new network forensic framework based on deep learning for Internet of Things networks: A particle deep framework." Future Generation Computer Systems 110 (2020): 91-106.

Koroniotis, Nickolaos, and Nour Moustafa. "Enhancing network forensics with particle swarm and deep learning: The particle deep framework." arXiv preprint arXiv:2005.00722 (2020).

Koroniotis, Nickolaos, Nour Moustafa, Francesco Schiliro, Praveen Gauravaram, and Helge Janicke. "A Holistic Review of Cybersecurity and Reliability Perspectives in Smart Airports." IEEE Access (2020).

Koroniotis, Nickolaos. "Designing an effective network forensic framework for the investigation of botnets in the Internet of Things." PhD diss., The University of New South Wales Australia, 2020.

Li, X., & Wang, Y., 2021） “Privacy Risks in Smart Home IoT Devices: A Case Study of Smart Cameras”

Smith, J., & Zhang, L 2022） “Internet of Things: A Survey on the Security Vulnerabilities and Defense”

Kumar, R., Jain, S., & Liu, P, 2022） “Security and Privacy Implications of Voice-Controlled Smart Home Systems”

Meidan et al., 2018）“N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders”

Roman, et al.2013) “On the features and challenges of security and privacy in distributed internet of things. Computer Networks”

Koroniotis et al., 2019) “Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems”

Athey & Imbens（2016） Recursive partitioning for heterogeneous causal effects

Lundberg & Lee（2017）A unified approach to interpreting model predictions

Liu et al.（2008）Isolation forest. Proceedings of the Eighth IEEE International Conference on Data Mining

Zhang et al.（2020）Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering

Meidan et al.（2018）N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing

Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset . Future Generation Computer Systems , 100 , 745 – 765.

Zhang, Y., Yang, J., & Oski, J. (2021). Unsupervised Network Anomaly Detection with Hybrid Models: A Comparative Study . Future Generation Computer Systems , 115 , 123 – 135.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest . Proceedings of the Eighth IEEE International Conference on Data Mining , 412 – 421.

Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Breitenbacher, D., Zhou, Y., & Elovici, Y. (2018). N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders . IEEE Pervasive Computing , 17 (3), 12 – 22.

Perera, C., Miri, A., & Joshi, R. (2020). AIoT: Artificial Intelligence Meets IoT . IEEE Internet of Things Journal , 7 (11), 10655 – 10667.

Chefer, H., Gur, S., & Wolf, L. (2021). Transformer-based Anomaly Detection with Explainability . arXiv preprint arXiv:2106.09587 .

Rodríguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation Forest: A new classifier ensemble method . IEEE Transactions on Pattern Analysis and Machine Intelligence , 28 (10), 1619 – 1630.

Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019) Towards the

development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems , 100 , 745 – 765.

Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Breitenbacher, D., Zhou, Y., & Elovici, Y. (2018) N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing , 17 (3), 12 – 22.

Zhang, Y., Yang, J., & Oski, J. (2021) Unsupervised Network Anomaly Detection with Hybrid Models: A Comparative Study. Future Generation Computer Systems , 115 , 123 – 135.

Perera, C., Miri, A., & Joshi, R. (2020) AIoT: Artificial Intelligence Meets IoT. IEEE Internet of Things Journal , 7 (11), 10655 – 10667.

Rodríguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006) Rotation Forest: A new classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence , 28 (10), 1619 – 1630.