Research on Cloud Computing Network Security Framework Based on Machine Learning

Yan Li, Genjuan Ma
Communication University of China, Nanjing, Nanjing 210000, Jiangsu, China
DOI: 10.32629/jher.v5i5.3062

Abstract: This paper studies and proposes a machine learning-based network security framework for cloud computing, which aims to deal with diverse security threats in cloud environments. By designing data collection, feature extraction, model training and dynamic protection mechanisms, the framework can detect and respond to intrusions, malware and abnormal behaviors in the cloud platform in real time. The results show that machine learning has significant advantages in improving detection accuracy and response speed, and the framework has adaptive learning ability and can dynamically adjust protection strategies.
***Keywords*:** cloud computing network security; machine learning

1. Introduction

With the rapid development of cloud computing, more and more enterprises and individuals migrate their data and services to the cloud, and cloud computing has become an important part of modern information technology infrastructure. However, due to its distributed architecture, multi-tenant sharing and dynamic resource scheduling, cloud computing environment is also facing more complex network security challenges. When faced with attacks and threats in cloud computing environment, traditional security protection mechanisms often appear to be inefficient or lagging behind in response. For example, distributed denial of service attacks (DDoS), virtualization attacks, data leakage and permission abuse in cloud environments may have a serious impact on the availability, confidentiality and integrity of cloud services [1].

At the same time, with the continuous evolution of attack technology, the complexity and concealment of network attacks are also increasing. Traditional rule-based security protection systems are difficult to cope with large-scale and diverse attack patterns. Therefore, how to use more intelligent technical means to improve the network security protection capability in cloud computing environment has become an important research direction. In recent years, machine learning has been gradually introduced into the field of network security due to its excellent performance in big data analysis, pattern recognition and anomaly detection, especially in the cloud computing environment.

2. Network security threats in cloud computing environments

2.1 Cloud computing architecture and security risks

Cloud computing environment adopts advanced technology architectures such as virtualization, multi-tenant sharing and elastic resource scheduling, which makes it have significant advantages in computing efficiency and flexibility. However, these technical features also pose new security risks. First, virtualization technology improves resource utilization by running multiple virtual machines on physical servers, but also increases the potential attack surface between different virtual machines. Furthermore, the multi-tenant architecture in cloud computing environment makes multiple users share resources on the same physical server, a feature that raises security issues of privilege abuse and insufficient data isolation. Although the dynamic resource scheduling mechanism improves computational elasticity, it also enables attackers to carry out attacks through frequent virtual machine migration operations. Ultimately, the centralized storage and management mode of cloud computing also gives attackers the opportunity to obtain massive user data through a single breakthrough.

2.2 Common types of network attacks

2.2.1 DDoS Attack

Distributed denial of service (DDoS) attacks are one of the most common types of attacks in cloud computing environments. By controlling a large number of infected devices, the attacker sends a huge number of requests to the target cloud server, exhausting the computing resources and bandwidth of the server, causing legitimate users to be unable to access the service normally. The high concurrency and complexity of DDoS attacks pose a serious threat to the availability of cloud

computing environment.

2.2.2 Data breach

Data breach is one of the most serious security risks in cloud computing, which is mainly manifested by the data of cloud service providers or users being accessed by unauthorized third parties. Since cloud computing stores a large amount of sensitive information, users privacy and trade secrets may be openly or maliciously used in the event of a data breach. Causes of data breaches can include misconfigured cloud storage, unauthorized access, or insider threats.

2.2.3 Virtualization attacks

Virtualization is one of the core technologies of cloud computing, but virtualization attack is a potential security threat. Attackers may break through virtualization isolation through "virtual machine escape" technology and gain control over the host system or other virtual machines. In addition, frequent migration and resource sharing of virtual machines also provide more attack paths for attackers.

2.2.4 Side channel attacks

Side channel attack is an attack method that uses the physical information leaked during hardware or system operation (such as electromagnetic radiation, processing time, power consumption, etc.) to infer the internal data of the system. In the environment of cloud computing, attackers may use side channels to steal sensitive information, such as encryption keys or user data, by sharing the same physical resources with the target virtual machine.

2.3 Impact of security threats on cloud computing services

Security threats in cloud computing pose significant challenges to the availability, confidentiality, and integrity of services. DDoS attacks can cause cloud services to be unavailable for extended periods of time, causing economic losses to businesses. Data breach incidents can not only damage users trust, but can also lead to legal liability and high compensation costs. Virtualization attacks can break data isolation across multiple tenants, leading to large-scale data leakage or corruption. In addition, the concealment and difficulty of preventing side channel attacks may make attackers obtain sensitive information without being noticed, which brings long-term potential risks to cloud computing systems.

3. Application of machine learning in network security

3.1 Intrusion detection system

Intrusion detection system is an important part of network security protection, which aims to monitor and analyze network traffic and identify potential attack behaviors. Traditional IDS mainly relies on preset rules to detect known attacks, but when faced with increasingly complex and changeable attack patterns, rules cant effectively identify new threats. Machine learning, through its powerful data mining and pattern recognition capabilities, is able to automatically learn and identify abnormal behaviors in network traffic. The IDS system based on supervised learning can be trained with a large number of labeled attack and normal behavior data to improve its detection accuracy and response speed [2]. In addition, unsupervised learning can be used to identify unknown attack patterns, enabling IDS systems to better respond to evolving security threats.

3.2 Anomaly detection and behavior analysis

Machine learning has a wide range of applications in anomaly detection, especially in cloud computing environments. By analyzing traffic patterns, user behavior, and system logs in the network, machine learning models can automatically detect abnormal behavior. For example, a user suddenly starts logging in from multiple geographical locations or an abnormal network request for an application may hint at potential attack behavior. Anomaly detection systems based on unsupervised learning algorithms such as clustering and isolated forests can effectively detect these anomalies and issue timely alarms, thereby reducing the impact of potential threats.

3.3 Malware identification

The identification and prevention of malware is one of the key tasks in network security. Traditional malware detection methods rely on feature libraries, but in the face of constantly mutating and new types of malware, this method gradually becomes ineffective. The malware detection method based on machine learning can determine whether there is malicious activity by extracting the behavioral characteristics of files, such as file access patterns, system calls and network communication behaviors. By training the classifier model, the system can automatically detect and prevent the spread of malware, improving the overall security in the cloud computing environment.

3.4 Network traffic analysis and classification

Network traffic analysis is an important task in network security, which helps to find potential network attacks and

abnormal behaviors. Machine learning algorithms can be used to classify network traffic and identify the difference between normal traffic and malicious traffic. Deep learning-based network traffic analysis models, such as convolutional neural networks and recurrent neural networks, can automatically learn features from complex network traffic data and effectively detect malicious traffic. In addition, machine learning models can also be used to identify encrypted traffic, P2P traffic and DDoS traffic in the network, helping administrators take corresponding defensive measures.

3.5 Application of Reinforcement Learning in Network Security

Reinforcement learning is a machine learning method that continuously learns the optimal strategy through the interaction between agents and the environment. In recent years, its application in network security has attracted more and more attention. In the cloud computing environment, reinforcement learning can be used to dynamically adjust protection strategies to cope with changing attack situations. For example, a reinforcement learning-based firewall system can automatically optimize the rule set to minimize resource consumption while blocking attack traffic to the greatest extent. In addition, reinforcement learning can also be used to automate defense systems to continuously improve the systems defense capabilities by learning against attackers.

4. Cloud computing network security framework based on machine learning

4.1 Overall architecture design of the framework

The cloud computing network security framework based on machine learning aims to provide efficient and intelligent security protection capabilities. The overall architecture of the framework includes multiple modules such as data acquisition, preprocessing, feature extraction and selection, and model training and optimization to ensure that it can dynamically respond to various security threats.

4.1.1 Data acquisition module

The data acquisition module is responsible for obtaining network traffic, user behavior logs and system event information from various data sources in the cloud computing environment. The collected data covers not only real-time network traffic, but also historical records and abnormal behavior logs, ensuring that the model can capture a wide range of security threat characteristics.

4.1.2 Preprocessing module

The data preprocessing module cleans and formats the collected original data, including denoising, normalization, filling missing values and other operations. In order to adapt to the input requirements of machine learning models, the preprocessing stage also needs to convert the data into a form suitable for model learning and reduce the interference of irrelevant features.

4.1.3 Feature extraction and selection

Feature extraction is the core step of the whole framework, which directly affects the accuracy and efficiency of the model. The module extracts key features from the processed data, such as network traffic patterns, file access frequencies, system call sequences, etc. Through the feature selection algorithm, the high-weight features related to the attack are screened out, redundant features are removed, and the complexity of the model is reduced.

4.1.4 Model training and optimization

In the model training stage, the collected historical data set is used for supervised learning or unsupervised learning. By continuously optimizing the hyperparameters and architecture design of the machine learning model, its detection accuracy is improved. The computational efficiency of the model should also be considered during the training process to ensure that it can run in real time in the cloud computing environment.

4.2 Safety protection mechanism

4.2.1 Real-time intrusion detection system

Real-time intrusion detection system (IDS) based on machine learning can continuously monitor network traffic and user behavior, and analyze whether there are anomalies in real time. By using classification algorithms, IDS can identify known and unknown attacks, raise alerts and take appropriate protective measures.

4.2.2 Dynamic threat response system

Relying on the real-time learning ability of machine learning algorithm, the dynamic threat response system can dynamically adjust the protection strategies according to the emerging threats. The system will automatically update security rules according to different attack patterns or trends, timely block attack traffic or limit suspicious behavior.

4.2.3 Monitoring and analysis of user behavior

The user behavior monitoring module uses the machine learning model to analyze the normal behavior patterns and identify abnormal behaviors, such as frequent cross-region login, abnormal resource requests, etc. By analyzing these behaviors, the system can advance warning of possible internal threats.

4.3 Model autoadaptation and continuous learning

4.3.1 online learning

Online learning allows models to continuously learn patterns in new data and adapt to the changing cloud computing environment. By updating the model parameters in real time, online learning avoids the overhead of frequently retraining the model and improves the recognition ability of the new attacks.

4.3.2 Incremental update mechanism

The incremental update mechanism allows the model to update only a fraction of the model weights when detecting new threat patterns instead of retraining the entire model. This way reduces the consumption of computational resources, while ensuring the continuous adaptability and real-time nature of the model.

5. Key technology and algorithm implementation

5.1 Machine learning algorithm selection

5.1.1 Support vector machine

SVM has good performance when dealing with binary classification problems, and is especially suitable for identifying the difference between normal traffic and abnormal traffic. In cloud computing environment, SVM can be used to build high-precision intrusion detection system, especially for small sample learning and high-dimensional feature data.

5.1.2 Decision tree and random forest

Decision tree algorithm divides data by constructing binary tree, which has the advantages of strong interpretability and high computational efficiency. Random forest improves the generalization ability of the model by integrating multiple decision trees. The algorithm can effectively deal with complex and diverse security threats in cloud computing.

5.1.3 Neural network and deep learning

Deep learning models model complex data through multi-layer neural networks and can automatically extract high-dimensional features, especially when dealing with large-scale cloud data. Convolutional neural networks and recurrent neural networks (RNNs) can be used for network traffic analysis and behavior prediction, and the adaptive learning capabilities of deep learning make them an effective tool to deal with complex attacks.

5.2 Feature engineering and dimensionality reduction technology

In order to improve the performance and efficiency of the model, feature engineering is an indispensable step. By using dimensionality reduction techniques such as PCA, the dimension of feature space can be reduced on the basis of retaining key features, and the computational burden can be reduced. Feature engineering also includes using regularization methods to prevent model overfitting, ensuring model robustness.

5.3 Model training and evaluation

5.3.1 Accuracy and recall rate

In the training process, the accuracy and recall rate of the model are important indicators to measure its detection ability. High accuracy means that the model can correctly identify most attack behaviors, while a high recall rate indicates that the model can capture as many potential threats as possible. The balance between the two determines the actual protection effect of the model.

5.3.2 False positive rate and false negative rate

In addition to detection accuracy, the false positive rate and false negative rate of the model are also key evaluation criteria. Too high false alarm rate may lead to waste of system resources, while too high false alarm rate means that threats are not discovered in time.

6. Performance evaluation of cloud computing network security framework

6.1 Experimental environment and data set

In order to verify the effectiveness of cloud computing network security framework, experiments should be carried out in real cloud computing environment. Diversified public data sets should be used in the experiment to simulate different

attack scenarios, and comprehensive tests should be carried out in combination with user behavior data in actual cloud computing environments.

6.2 Detection performance evaluation of the model

Core metrics for performance evaluation include model accuracy, response time, and processing efficiency. Accuracy represents the overall detection power of the model, response time measures the speed that the system goes from detection to response to threat, and processing efficiency reflects the performance of the system under high load. Special attention should be paid to the performance of the model at different attack strengths.

6.3 Comparative analysis of different machine learning algorithms

To select the optimal protection model, a detailed comparison of the detection capabilities of different machine learning algorithms is required. Evaluation metrics include detection accuracy, false alarm rate, time-consuming of model training and inference, and the performance of the algorithm on different attack types. Through horizontal comparison, we can determine the most suitable machine learning algorithm in the cloud computing network security framework.

6.4 The Systems ability to respond to diverse attack scenarios

In addition to common attack types, performance evaluation tests the systems ability to handle a variety of attack scenarios, such as zero-day attacks, multi-tenant attacks, and virtual machine escape attacks. By simulating these complex attack situations, the robust robustness and adaptability of the framework can be evaluated, thus ensuring its reliability in a real cloud computing environment.

7. Advantages and limitations of the machine learning-based network security framework

Machine learning-based cybersecurity frameworks have shown significant advantages in dealing with complex threats. First of all, machine learning algorithms can automatically learn and identify abnormal behaviors, improving the accuracy of intrusion detection and the ability to deal with unknown attacks. In addition, the framework is self-adaptive and can dynamically update the protection strategy according to emerging threats, reducing the need for manual intervention.

However, the framework also has limitations. The model training process may require a lot of computing resources, especially when dealing with large-scale cloud data, and the deployment cost is high. At the same time, the problems of false alarm rate and false negative rate still exist, and the framework relies heavily on data quality and feature engineering. If the data is incomplete or there is too much noise, it may affect the detection effect.

8. Conclusion

This paper proposes and studies a cloud computing network security framework based on machine learning, and discusses the architecture design, key technologies and application scenarios of the framework in detail. By introducing modules such as data acquisition, feature extraction, model training and optimization, this framework can effectively detect and protect various security threats in cloud computing environment. Experiments show that machine learning algorithms can significantly improve the accuracy of intrusion detection, malware identification and abnormal behavior analysis in cloud environment, and have strong adaptability. However, the resource consumption and false alarm rate of model training still need to be further optimized.

References

[1] Su Zibin. Research and Implementation of Cloud Computing Network Security Defense Technology [D]. Beijing University of Posts and Telecommunications, 2016.

[2] Wang Zhe. Development status and prospects of intelligent edge computing [J]. Artificial Intelligence, 2019, (05): 18-25.

[3] Du Junlong, Zhou Jiantao. Research on Network Intrusion Detection System Based on Cloud Computing and Machine Learning [J]. Microcomputer Applications, 2021, Vol. 37(2): 18-20, 59.

[4] Peng Aijun, Ma Jun, Huang Jie, et al. Research on Enterprise Power Load Forecasting Method Based on Cloud Computing and Machine Learning [J]. Science and Innovation, 2024, (18): 127-129.

[5] Liu Jinglin, Hao Jiayu. Intelligent Routing Strategy for Cloud Computing Data Center Networks Based on Reinforcement Learning [J]. Journal of Ningde Normal University (Natural Science Edition), 2023, Vol. 35(4): 374-381.

Author Bio

Yan Li: born in 1986, female,native of Yangzhou, Jiangsu province, Han nationality, undergraduate, research interests: senior engineer, big data, artificial intelligence, cloud computing, Java.

Genjuan Ma: born in 1981, female, Yancheng, Jiangsu Province, master, senior engineer, research direction: Software Engineering, Big Data and artificial intelligence.