Optimization of Network Intrusion Detection Model Based on Big Data Analysis

Jizhou Shan* and Hong Ma

Hainan College of Economics and Business, Haikou, Hainan, 571127, China
E-mail: sjz202405@126.com
**Corresponding Author*

Received 12 July 2024; Accepted 03 September 2024

Abstract

As user usage grows, so do security threats to networks, the Internet, websites, and organizations. Detecting intrusions in such a big data situation is complex. A feature-optimized network intrusion detection model based on extensive data analysis is designed to overcome the limitations of current network intrusion detection models and obtain more ideal results. Firstly, the current modeling status of network intrusion detection is studied, and the influence of features on the results of network intrusion detection is analyzed. Then, the feature optimization mathematical model of network intrusion detection is established. The solution of the feature optimization mathematical model is searched by an adaptive genetic algorithm simulating natural biological evolution. The optimal feature subset of intrusion detection is obtained by back coding the optimal solution. Finally, according to the optimal feature subset, the learning sample of network intrusion detection is modeled, and the optimal network intrusion detection model is designed. Using the standard data set of network intrusion detection for simulation and comparison tests, the average accuracy of this paper's network intrusion

Journal of Cyber Security and Mobility, Vol. 13_6, 1357–1378.
doi: 10.13052/jcsm2245-1439.1366

detection model is about 95%, while other current network intrusion detection models are below 95%. Meanwhile, the time of training and the detection of intrusion detection modeling in this model is significantly reduced, and better efficiency of network intrusion detection can be obtained.

Keywords: Network security, big data analysis, network intrusion detection, learning sample modeling, detection model, feature analysis.

1 Introduction

In recent years, big data, artificial intelligence, cloud computing and 5G technologies have developed rapidly, and the application of network has become more extensive and convenient [1, 2]. With the expansion of Internet users, more network traffic and network attack surface are introduced, which makes it a more challenging problem to protect network information and communication security [3]. Intrusion Detection System (IDS) plays an important role in improving the security level of the system.

Telecom fraud has seriously affected people's sense of acquisition, happiness and security. Our country attaches great importance to the governance of telecom fraud [4, 5]. As a link in the whole chain of telecom fraud, operators also invest a lot of manpower, material resources and technical funds to carry out special governance of telecom fraud. How operators use technical means to fight fraud has always been the focus of research. At first, the characteristics of telecom fraud are not complicated, and good anti-fraud effect can be achieved through some simple behaviour analysis and content detection [6]. The essence of intrusion detection is a classification problem, and machine learning can better complete various classification tasks [7, 8]. Intrusion detection should first extract the important features of computer system and network, then compare these features with normal features and known intrusion features, and find out the intrusion behaviour in advance before it causes negative impact on computer system and network, so as to find out the intrusion behaviour suffered by computer system and network and take corresponding security measures to eliminate intrusion threats [9]. However, traditional intrusion detection technology has been difficult to complete more and more complex intrusion detection tasks. Traditional firewall, user authentication and data encryption technology, to a certain extent, not only lack of intelligence to detect intrusion, but also low detection efficiency [10, 11]. Therefore, we need to apply more intelligent and efficient technology to intrusion detection.

With the rapid development of emerging technologies, the application of information technology has gradually penetrated into all walks of life, bringing convenient and intelligent services to our work and life, but also bringing gradually aggravated network security problems. Among them, intelligent attacks such as zero-day attack, APT attack and ransomware attack are increasingly rampant. This paper puts forward some problems existing in the current network security protection, introduces the necessary factor data set for intrusion detection system training, and studies the application of deep learning model combined with data set in intrusion detection system, so as to realize fraud detection and prevention based on big data.

2 Design of Network Intrusion Detection Model Based on Big Data Analysis Feature Optimization

2.1 Network System Defect Analysis

Intrusion detection and defense systems are important guarantees for data security. The related developers did not fully consider the security of the system when they first designed the software system, which led to some security vulnerabilities in the process of software development [12, 13]. Although there are some testing tools that can find and fix some vulnerabilities, it is impossible to completely fix them. Once there is a serious security risk, this loophole will become a breakthrough for many hackers to attack the network. Hackers modify computer systems without users' permission, resulting in loss or disclosure of user information [14]. The implementation of data security requires the comprehensive use of various methods such as technical means, organizational management, and laws and regulations. Firstly, technological means are the foundation of data security. For example, by using data encryption technology, the security of data during transmission and storage can be guaranteed. Encryption algorithms are used to encrypt data to prevent unauthorized access. Encryption can convert data into a form that is difficult to understand, and data cannot be read unless the decryption key is held. In addition, access control technology is adopted to restrict access to data, ensuring that only authorized personnel can access sensitive data and that only authorized users can access specific data. Manage user access to data through authentication and authorization mechanisms, such as using strong passwords, two factor authentication, and other methods. In addition, establish strong firewalls and intrusion detection systems, use network security devices and software such as firewalls, intrusion detection

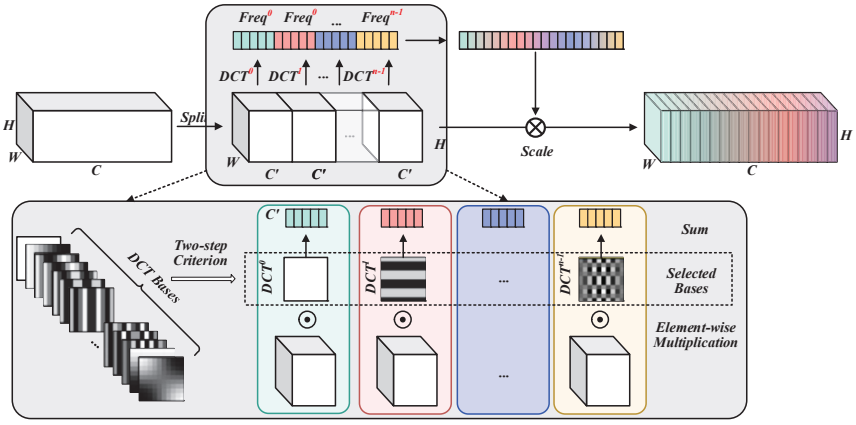

Figure 1 Analysis of network security vulnerability characteristics.

systems (IDS), intrusion defense systems (IPS), etc. to prevent unauthorized access or attacks, and resist network attacks and malicious software intrusion. The comprehensive application of these technological means can effectively protect the security of data.

Figure 1 is an analysis of the characteristics of network security vulnerabilities, which are weaknesses found in information systems, security programs or internal controls that threat actors can exploit. When threat actors exploit vulnerabilities in the system, they can unauthorized access to the organization's confidential data and steal data. Threatening actors often sell stolen data on dark networks or return it to organizations in exchange for ransom. Cybercriminals may reveal their login credentials when users are subjected to phishing attacks or password hygiene is poor. After revealing a user's login credentials, cybercriminals can gain unauthorized access to an organization's network [15, 16]. Programmers may inadvertently leave flaws and mistakes that can be exploited by cybercriminals. If software does not fix these errors, cybercriminals can have unauthorized access to the organization's hardware, software, data and other resources. When systems are overly complex, organizations can be exploited by cybercriminals by setting up misconfigurations, flaws, and unauthorized access points in their systems [17]. Attack surface refers to all possible entry points that cybercriminals can use to access the system. The more devices and systems connected to an organization's network, the greater the attack surface for cybercriminals to gain unauthorized access. If your organization mismanages user roles, such as giving someone more access than they need or not removing access

from former employees, your network can be vulnerable to both internal and external security vulnerabilities.

2.2 Data Security Protection Based on Feature Analysis

Data encryption is one of the core technologies to ensure data security. By encrypting data, attackers cannot directly obtain plaintext information even if the data is intercepted during transmission or storage. Data encryption can be divided into two types: symmetric encryption and asymmetric encryption. Scalable encryption of static data and transmitted data is very important for implementation across large data pipelines. Scalability is the key point here, and in addition to storage formats such as NoSQL, encrypted data needs to be spanned across the analysis toolset and its output [18, 19].

The large storage and processing demand of big data have driven enterprises towards cloud computing, despite security risks such as exposed API keys and misconfigurations. Automated tools are used to scan public cloud assets for vulnerabilities. In model selection for big data analysis, scalability and efficiency guided our choices, with models like distributed random forests being favoured. Hyperparameter tuning through cross-validation and grid search ensured optimization, with detailed documentation of the tuning process and performance outcomes. This comprehensive approach optimizes models for cloud-based big data workloads while addressing critical security considerations.

Figure 2 shows the analysis of the big data ecosystem. In the big data ecosystem, it is necessary to strengthen access control to ensure that only authorized personnel can access Minggan data. Access control can be achieved through the use of strong passwords, multi factor authentication, permission management, and auditing logs. In addition, for some sensitive data, access control (ACL) or access control matrix (ACM) can be used for fine-grained permission control. Encrypting data transmission and storage, data encryption is one of the important means to protect data security. For sensitive data transmission, the SSL ITLS protocol can be used. Disk encryption technology, database encryption technology, or file encryption technology can be used to protect data from unauthorized access [20, 21]. Regular backup and recovery are crucial to avoid irreparable data caused by data loss or destruction. Backup data should be stored in a secure and reliable location, and regularly tested to verify the recoverability of the backup data. In addition, develop a comprehensive disaster recovery plan to ensure rapid system recovery in the event of a system failure or security

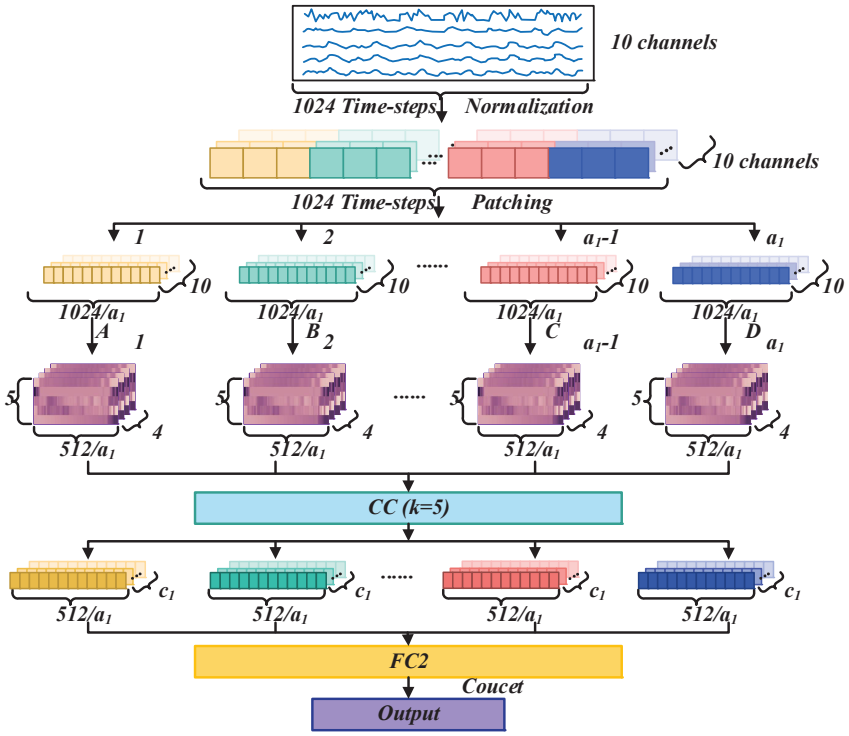

Figure 2 Big data ecosystem analysis.

incident. Continuous monitoring and vulnerability repair, using a security monitoring system to monitor critical systems and networks in real-time, and promptly detect abnormal behaviour and security incidents [22, 23]. Vulnerability repair is an important environment for protecting data security, timely repairing system and application vulnerabilities, and preventing hackers from exploiting vulnerabilities to invade and attack.

3 Optimization of Intrusion Detection Technology

Intrusion Detection Systems (IDS) is a very important software or hardware security tool, which is used to detect possible threats, prevent unauthorized access or abuse, and report attacks to security administrators [24, 25]. According to engine detection mechanism, intrusion detection systems can be classified into IDS based on signature detection and IDS based on abnormal behavior detection.

In our research on constructing a network intrusion fraud detection model based on big data analysis, we've conducted a comparative analysis between our proposed feature selection method and mainstream techniques such as PCA and LDA. Our method, integrating domain-specific insights with statistical significance, was found to outperform PCA and LDA in terms of accuracy and efficiency. Through rigorous experiments on a comprehensive dataset, we demonstrated that our approach's ability to prioritize impactful features while minimizing information loss leads to superior model performance. This comparative analysis substantiates the superiority of our feature selection technique, highlighting its critical role in enhancing the efficacy of intrusion detection models.

3.1 Intrusion Prevention Optimization Based on Signature Detection

According to the known signature detection, this method can effectively identify the existing attacks in the signature library, but cannot identify the unknown attacks and the variants of the known attacks [26]. Intrusion Detection System Q (IDS) is a device or software application used to detect potential threats in networks or systems. It can help us find unknown attacks, such as zero-day attacks, or known attacks, such as DDoS attacks.

NIDS is an IDS that works at the network level. It detects threats by monitoring network traffic. For example, if NIDS detects a known attack pattern, such as a SYN-only flood attack, it can generate a warning, host-based IDS (HIDS): HIDS is an IDS that works at the host level. It detects threats by monitoring system logs, file system changes, or system calls [27, 28]. For example, if HIDS detects that a file has been modified and the file should not be modified, it can generate a warning.

IDS mainly has two working principles. For example, if IDS detects that network traffic contains a known attack pattern, such as SQL injection, it can generate a warning. Anomaly-based detection is to detect threats by comparing network traffic or system behavior with normal patterns. For example, if IDS detects that network traffic contains an abnormal pattern, such as a sudden increase in traffic, it can generate a warning.

When studying the network intrusion and fraud detection model based on big data analysis, it involves many aspects of computing and technology. Data standardization is shown in Equation (1).

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

The information gain of feature selection is shown in Equation (2).

$$IG(T, A) = H(T) - H(T|A) \tag{2}$$

The Gini impurity of decision tree classification is shown in Equation (3).

$$Gini(T) = 1 - \sum_{i=1}^{m} p_i^2 \tag{3}$$

The logistic regression model is shown in Equation (4).

$$P(Y = 1|X) = \frac{1}{1 + e^{-(wx+b)}} \tag{4}$$

The support vector machine classifier is shown in Equation (5).

$$f(x) = sign\left(\sum_{i=1}^{n} a_i y_i K(x_i, x) + b\right) \tag{5}$$

Principal component analysis (PCA) is shown in Equation (6).

$$\mathrm{Cov(X, Y)} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \tag{6}$$

K-means clustering is shown in Equation (7).

$$Distance(x, \mu_k) = \sqrt{\sum_{i=1}^{d}(x_i - \mu_{ki})^2} \tag{7}$$

When designing API interface signatures, the main consideration is to ensure that the request data is correct. When value of a certain field in the request changes, original signature result will change. So, as long as the parameters change, the signature must change, otherwise the request will be invalid. Ensure that the request source is legitimate. Generally, the algorithm that generates the signature will appear in pairs with an appKey and an appSecret, which can identify the caller's identity based on the appKey. A timestamp is the current timestamp corresponding to the client calling the interface, which is used to prevent DoS attacks. When a hacker hijacks the requested URL for DoS attack, the interface will determine the difference between the server's current system time and the timestamp passed

in the interface every time it calls the interface. If this difference exceeds a set time (such as 5 minutes), the request will be intercepted. If it is within the set timeout range, DoS attack cannot be prevented. The timestamp mechanism can only reduce the time of DoS attacks and shorten the attack time. If the hacker modifies the value of the timestamp, it can be processed through the sign signature mechanism. The sign mechanism can prevent parameter tampering, but cannot prevent DOS attacks (third parties using correct parameters constantly request the server, making it unable to provide services normally). Therefore, it is necessary to introduce a timestamp mechanism. The specific operation is: when the client forms the sign value, in addition to using all parameters and tokens, an additional time is added to initiate the request.

3.2 IDS Optimization Based on Abnormal Behavior Detection

By learning the network traffic behavior to classify traffic, unknown attacks can be detected. The first method can identify attacks efficiently and accurately [29]. Through the application of a sophisticated learning mechanism that meticulously analyzes and categorizes network traffic behavior, our approach demonstrates a remarkable capability to detect unknown attacks. To substantiate this claim and showcase its superiority in identifying novel intrusion patterns, we have conducted a series of comprehensive comparative experiments. These experiments meticulously contrast our method against a selection of leading anomaly detection algorithms currently employed in network security.

Our results reveal that the proposed technique not only matches but frequently surpasses the performance of these benchmarks in terms of both efficiency and accuracy when faced with unknown attack types. The method's adeptness in swiftly recognizing and accurately classifying novel threats without extensive prior knowledge or training on such patterns is particularly noteworthy. This capability is attributed to its innovative learning algorithm, which adapts and evolves based on the dynamic nature of network traffic, ensuring a robust defense against emerging security risks.

Figure 3 shows IDS detection based on anomaly statistics, which detects anomalies by modeling and analyzing the statistical characteristics of network traffic and system behavior. Anomaly statistical methods usually use statistics, data mining, machine learning and other technologies to establish a baseline model of normal behavior, and then compare it with the actual observed behavior. The system alerts when the observed behavior differs from the baseline model beyond a certain threshold. IDS based on anomaly

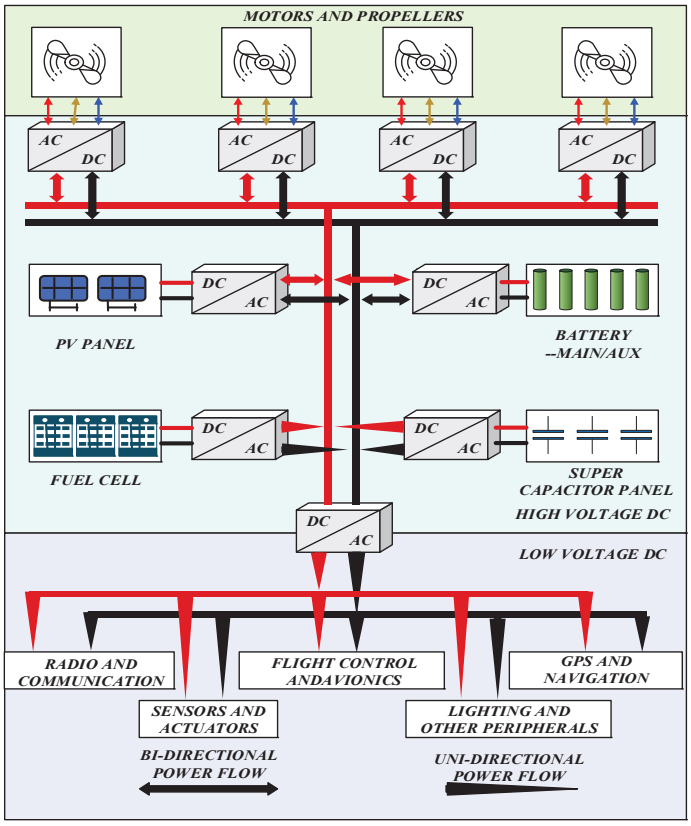

Figure 3 IDS detection based on anomaly statistics.

statistics can detect unknown attacks and has good adaptability, but it may have a high false alarm rate.

IDS based on protocol exception focuses on the abnormal use of network protocols, such as protocol format errors, illegal instructions or behaviors that do not comply with protocol specifications. By checking the protocol packets in network traffic, we can analyze whether there are non-conforming protocol specifications, so as to find potential attacks. IDS based on protocol anomalies can detect attacks against specific protocols, but it has limited ability to detect attacks that do not involve protocol anomalies.

The Naive Bayesian classifier is shown in Equation (8).

$$P(C|x) = \frac{P(x|C)P(C)}{P(x)} \tag{8}$$

Time series analysis-autocorrelation function is shown in Equation (9).

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \mu)(x_{t+k} - \mu)}{\sum_{t=1}^{N}(x_t - \mu)^2} \tag{9}$$

The time series analysis-moving average (MA) model is shown in Equation (10).

$$x_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{10}$$

The time series analysis-autoregressive (AR) model is shown in Equation (11).

$$x_t = \mu + \sum_{i=1}^{p} \varphi_i x_{t-i} + \epsilon_t \tag{11}$$

The time series analysis-autoregressive moving average (ARMA) model is shown in Equation (12).

$$x_i = \mu + \sum_{i=1}^{p} \varphi_i x_{i-1} + \sum_{i=1}^{q} \theta_i \epsilon_{i-1} + \epsilon_i \tag{12}$$

IDS based on traffic anomaly focuses on the abnormal changes of network traffic, such as sudden increase, sudden decrease or abnormal access mode. IDS based on traffic anomaly detects the existence of abnormal traffic patterns by monitoring and analyzing the statistical data of network traffic, thus discovering potential attacks. IDS based on traffic anomaly can detect traffic-related attacks, such as DoS/DDoS attacks, but it has limited ability to detect attacks that do not affect traffic.

Rule-based IDS relies on a predefined set of rules, usually including the characteristics and behavior patterns of known attacks. Rule-based IDS detects potential attacks by matching network traffic or system behavior with rule sets. The rule set needs to be updated constantly to keep the detection ability of the latest attack means.

3.3 Optimization of Host-based Intrusion Detection System

HIDS aims to detect malicious activities and security vulnerabilities on a single computer or server. HIDS detects potential intrusions by monitoring the activity of system logs files and processes on the host and issues alerts to inform administrators to take appropriate action. Intrusions are

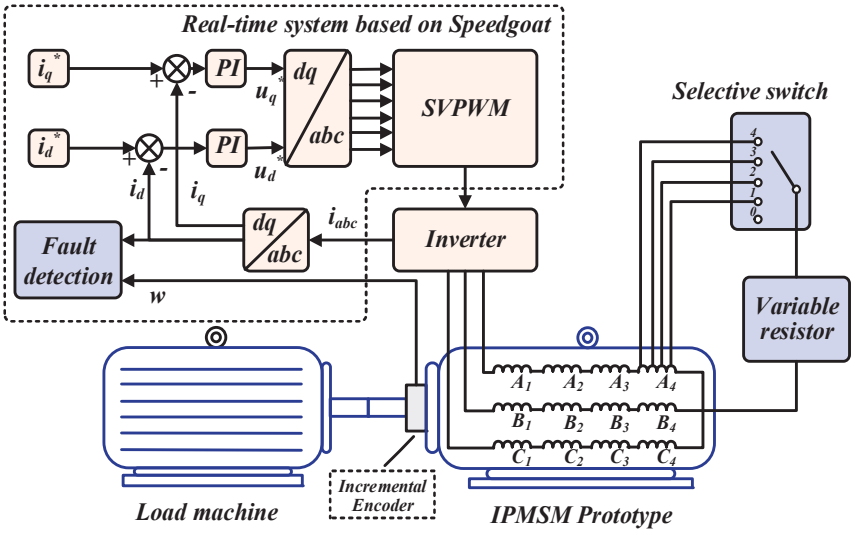

Figure 4 HIDS response architecture for intrusion behavior.

detected by monitoring host system status, event logs and audit records. A key advantage of HIDS lies in its comprehensive monitoring of system log files, processes, and events, which enables it to swiftly detect anomalies and potential intrusions. By closely observing the host's system status, event logs, and audit records, HIDS can accurately identify suspicious activities that might otherwise go unnoticed. This proactive surveillance mechanism ensures timely alerts to administrators, facilitating immediate action and mitigating the impact of security breaches. Moreover, HIDS's focus on a single host allows for a deeper and more detailed analysis of system behavior, making it particularly effective in environments where critical data and resources are concentrated. This targeted approach not only enhances the precision of intrusion detection but also streamlines the response process, making HIDS a critical component in the defense against cyber threats.

Figure 4 shows the HIDS response architecture for intrusion behavior. HIDS can help organizations discover and respond to intrusion behavior in time and reduce the impact of intrusion on a single computer or server. However, HIDS also has some limitations, such as false positives or omissions. Therefore, when selecting and using HIDS, it is necessary to consider the specific security requirements and environment, as well as the feasibility and efficiency of HIDS. In addition, because HIDS only monitors a single

host, it is often used with Network IDS (NIDS) and Intrusion Detection and Prevention System (IDPS) to improve the security of the entire network. The communication data is interpreted after the network stack processes the communication data before the application layer program processes it. The advantages of host-based intrusion detection system are as follows: it can detect the misuse of internal authorized personnel and successfully avoid the traditional system protection methods and penetrate into the intrusion activities inside the network; With the information of operating system and running environment, the detection accuracy is high; After detecting the intrusion, it can cooperate with the operating system to prevent the continuation of the intrusion and respond in time.

Optimizing model processing speed and delay times has been a focal point. We've streamlined feature extraction using dimensionality reduction to accelerate decision-making and reduced latency. A hierarchical processing architecture, complemented by dynamic resource allocation, ensures swift responses and maintains detection accuracy under varying network loads. Real-time anomaly detection algorithms further enhance responsiveness to emerging threats. These enhancements significantly improve the model's efficiency and precision, equipping it to effectively handle big data environments and provide robust security against network intrusion fraud.

4 Experimental Results and Analysis

4.1 Intrusion Detection Experimental Data Set

In evaluating and training anomaly-based Intrusion Detection Systems (IDS), particularly Network IDS, the NSL-KDD dataset, derived from KDDCup99, stands out for its refined data handling. This dataset eliminates redundant records, ensuring unbiased classifier training and more accurate detection rate calculation. It also optimizes the selection of records for different difficulty levels, enhancing the assessment of learning methodologies. Generated using IXIA traffic, NSL-KDD mirrors KDDCup99's 49 features but with a focus on network-oriented attributes. Split into training (175,341 records) and testing (82,332 records) sets, the dataset covers a wide range of attack types, from Fuzzy to Worm attacks. Preprocessing assessment through quantitative analysis of data distribution and feature correlation changes pre and post-processing validates the effectiveness and necessity of preprocessing steps. These steps significantly improve model performance by optimizing data characteristics for better learning outcomes.

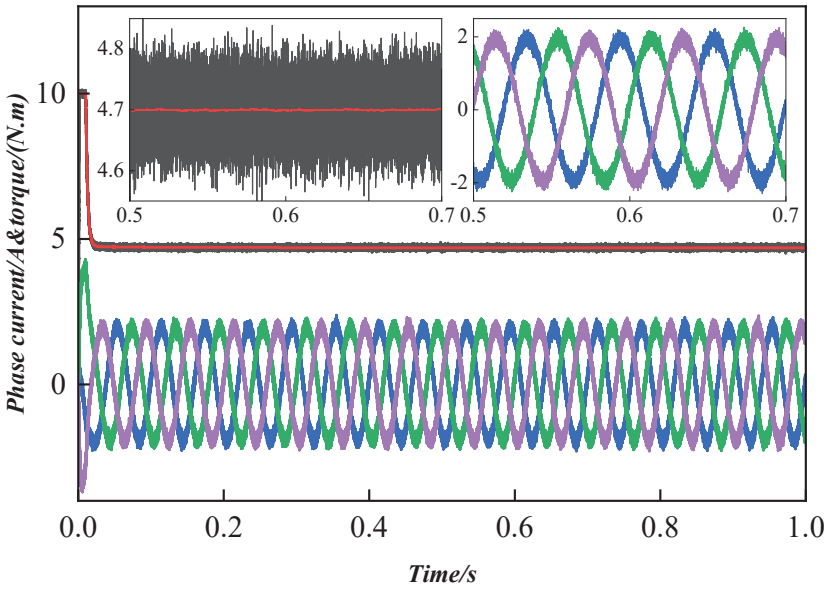

Figure 5 Analysis of intrusion detection results.

4.2 Analysis of Optimal Results of Intrusion Detection System

We propose a deep neural network or DNN algorithm. The input layer of the algorithm includes 41 inputs, 4 hidden layers and 2 output layers.

Figure 5 is the analysis of intrusion detection results. The detection method based on expert system is a common intrusion detection method. By transforming the knowledge about intrusion into IF-THEN structure rules, IF is the condition of intrusion, and THEN is the solution after intrusion is discovered. The detection method based on expert system has the advantage of separating the reasoning control process of the system from the final solution of the problem, that is, the user does not need to understand or interfere with the reasoning process within the expert system, but only needs to regard the expert system as a black box. When applying expert system to intrusion detection, there are the following problems: lack of ability to deal with sequence data, that is, unable to deal with the correlation of data before and after; The performance of detecting attacks depends largely on the knowledge of designers. Only known attack patterns can be detected; Unable to deal with and judge the uncertainty; The rule base is difficult to maintain, so the impact on other rules in the rule base should be considered when changing rules.

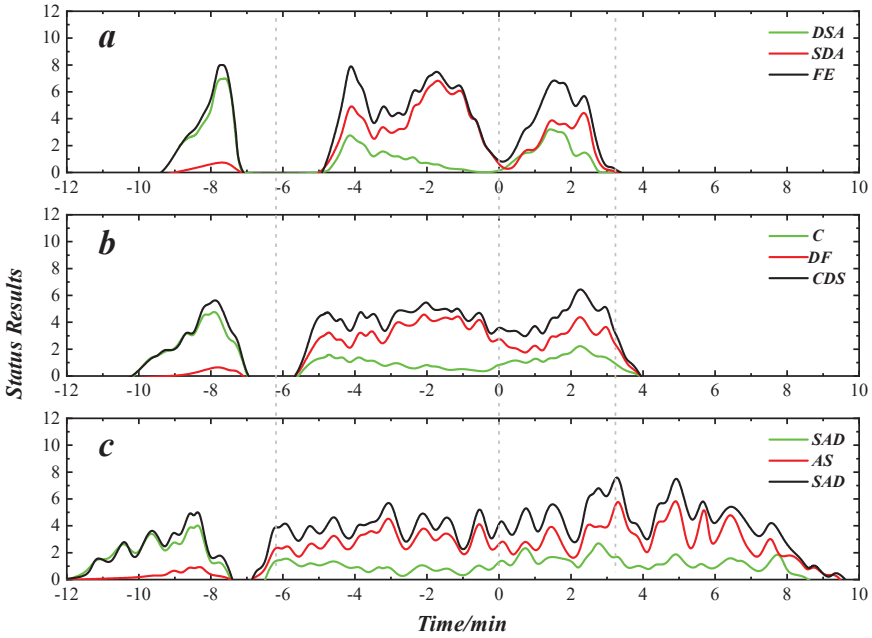

Figure 6 Detection method analysis based on state transition analysis.

The effectiveness of the unbalanced data processing techniques used in this study is investigated and experimentally verified, especially the methods such as SMOTE (Synthetic Minority Over-sampling Technique) and EasyEnsemble. These techniques have shown significant potential to solve the problem of data imbalance, and can effectively improve the detection ability of the model to detect a few categories (such as online fraud), thereby improving the overall classification performance. Experimental results show that the model trained on the SMOTE processed dataset has a significant improvement in the recall rate and F1 score when detecting online fraud, especially in the face of highly unbalanced datasets.

Figure 6 shows the analysis result of detection methods, the known attack patterns are represented and detected by state transition diagram, that is, the known attack patterns are described by system state and state transition expressions, and the intrusion process is represented by finite state machine model. The intrusion state represents the system state after the intrusion. The state transition analysis engine for misuse detection includes a set of state transition diagrams, each representing an intrusion or penetration pattern. Whenever a new behaviour occurs, the state transition analysis engine

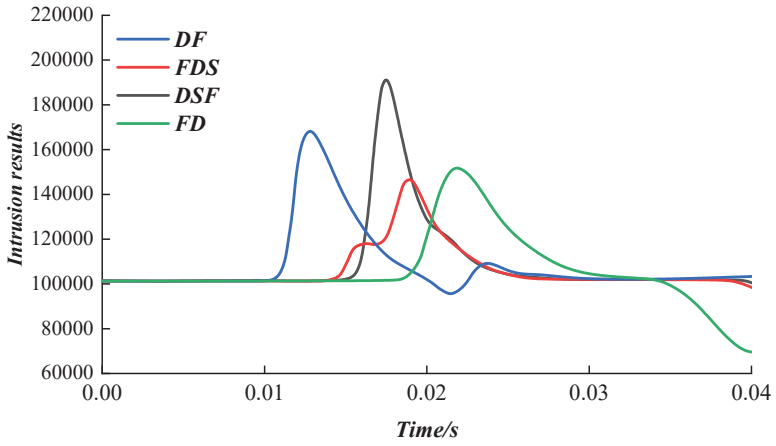

Figure 7 Analysis of misuse detection method results.

examines all state transition diagrams to see if it causes a state transition of the system. If the new behaviour negates the assertion of the current state, the state transition analysis engine traces the state transition diagram back to the state when the assertion is still valid; If the new behaviour transitions the system state to the intrusion state, the state transition information is sent to the decision engine, and the decision engine takes corresponding measures according to the predefined strategy. The intrusion detection process expressed by state transition analysis is only related to the change of system state, but has nothing to do with the intrusion process. The state transition analysis method can detect cooperative attacks and slow attacks. It can detect the attack behaviour when the attack behaviour has not reached the intrusion state, so as to take corresponding measures to prevent the attack behaviour in time. The state transition diagram gives the minimum subset of characteristic behaviours to ensure the success of attack, and can detect different manifestations with the same intrusion pattern. The assertions and characteristic behaviours corresponding to states in state transition analysis method need to be encoded manually, which will have problems when applied to complex intrusion scenarios.

Figure 7 shows the result analysis of misuse detection method. The advantages of misuse detection are: high accuracy of attack detection; Can identify the type of attack. The disadvantages of misuse detection are: only known attacks can be detected; Lagging behind new attacks, new attacks can only be detected after they are included in the attack feature library; It is difficult to maintain the attack feature library. After new attacks appear, experts

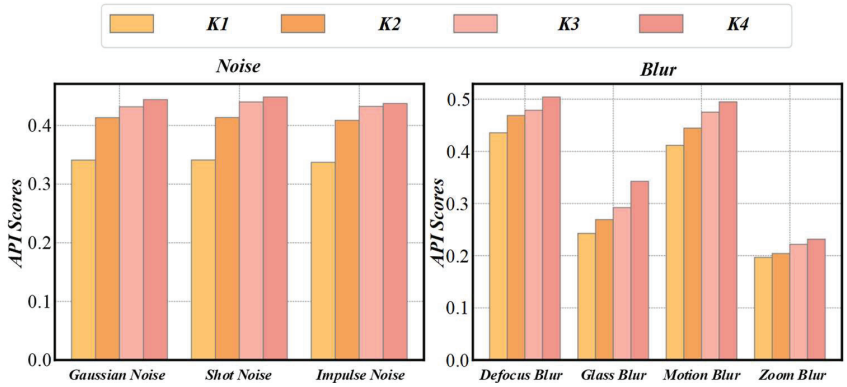

Figure 8 Digital certificate encryption data analysis.

need to extract attack features according to their professional knowledge and update the attack feature library constantly; Attackers can bypass detection by modifying the attack behaviour so that it does not match the features in the attack feature library. Misuse detection and anomaly detection have their own advantages and disadvantages, and they are complementary to each other. Many studies try to combine these two techniques to bring their advantages into play, improve detection efficiency and minimize the error rate. Figure 8 shows the digital certificate encryption data analysis, and the client receiving the certificate can verify the digital signature on the certificate using the digital Certificate Authority's public key. After successful verification, the client understands two things: First, the public key of the authentication server is an accurate and valid digital certificate authority. Second, the public key of the server is trusted. When the client sends a request, the server sends the digital certificate to the client. The client decrypts the encrypted ciphertext (F3) using the public key provided by the CA and decrypts the encrypted ciphertext (F3) in reverse order to obtain F2. Also, use SHA1 to hash the certificate plaintext content (F1) into F2. The certificate is acceptable if the two are equal and the server's public key has not been tampered with. Its core is a certificate signed using the CA authority's private key to verify that the public copper information is trusted.

5 Conclusion

In order to speed up the speed of network intrusion detection, a network intrusion detection model based on feature optimization is designed. Simulation

results of standard data set of network intrusion detection show that the average correct rate of network intrusion detection of the proposed model is better than that of the comparison model, and the modeling training and detection time are relatively less, so it can be widely used in practical network security management. Aiming at the problems of low accuracy, poor real-time performance and low generalization performance of existing network intrusion detection systems, the gain-rate algorithm and network intrusion detection model are designed by taking advantage of the good classification performance and strong generalization ability of big data analysis technology based on feature optimization. The gain rate is used to filter the data features of the data set, which can ensure the accuracy of intrusion detection and shorten the training time of the model. The experimental results show that this model has higher accuracy and stronger generalization ability than other intrusion detection models, and reduces the training time by 77% while ensuring the accuracy. The average accuracy rate of network intrusion detection is 95%, which is significantly higher than other detection models. In future research, emphasis should be placed on novel network application scenarios, integrating big data analysis to enhance real-time adaptability and effectiveness in practical environments, thereby shaping the evolving landscape of network intrusion detection.

Funding

Project supported by the Education Department of Hainan Province, project number: Hnjg2023ZD-67.

References

[1] A. Abirami and S. Palanikumar, "BBBC-DDRL: A hybrid big-bang big-crunch optimization and deliberated deep reinforced learning mechanisms for cyber-attack detection," Computers and Electrical Engineering, vol. 109, pp. 108773, 2023.

[2] Adnan Hasan Bdair Aighuraibawi et al., "Hybridizing flower pollination algorithm with particle swarm optimization for enhancing the performance of IPv6 intrusion detection system," Alexandria Engineering Journal, vol. 104, pp. 504–514, 2024.

[3] Guilherme Nunes Nasseh Barbosa, Martin Andreoni, and Diogo Menezes Ferrazani Mattos, "Optimizing feature selection in intrusion detection systems: Pareto dominance set approaches with mutual

information and linear correlation," Ad Hoc Networks, vol. 159, pp. 103485, 2024.

[4] Arash Bozorgchenani, Charilaos C. Zarakovitis, Su Fong Chien, Tiew On Ting, Qiang Ni, and Wissam Mallouli, "Novel modeling and optimization for joint Cybersecurity-vs-QoS Intrusion Detection Mechanisms in 5G networks," Computer Networks, vol. 237, pp. 110051, 2023.

[5] Ibrahim Hayatu Hassan, Mohammed Abdullahi, Mansur Masama Aliyu, Sahabi Ali Yusuf, and Abdulrazaq Abdulrahim, "An improved binary manta ray foraging optimization algorithm based feature selection and random forest classifier for network intrusion detection," Intelligent Systems with Applications, vol. 16, pp. 200114, 2022.

[6] P. Rajesh Kanna and P. Santhi, "Hybrid Intrusion Detection using MapReduce based Black Widow Optimized Convolutional Long Short-Term Memory Neural Networks," Expert Systems with Applications, vol. 194, pp. 116545, 2022.

[7] Murad Ali Khan, Naeem Iqbal, Imran, Harun Jamil, and Do-Hyeun Kim, "An optimized ensemble prediction model using AutoML based on soft voting classifier for network intrusion detection," Journal of Network and Computer Applications, vol. 212, pp. 103560, 2023.

[8] S. Lakshmi Narayanan, M. Kasiselvanathan, K. B. Gurumoorthy, and V. Kiruthika, "Particle swarm optimization based artificial neural network (PSO-ANN) model for effective k-barrier count intrusion detection system in WSN," Measurement: Sensors, vol. 29, pp. 100875, 2023.

[9] Shahid Latif, Wadii Boulila, Anis Koubaa, Zhuo Zou, and Jawad Ahmad, "DTL-IDS: An optimized Intrusion Detection Framework using Deep Transfer Learning and Genetic Algorithm," Journal of Network and Computer Applications, vol. 221, pp. 103784, 2024.

[10] K. G. Maheswari, C. Siva, and G. Nalinipriya, "Optimal cluster based feature selection for intrusion detection system in web and cloud computing environment using hybrid teacher learning optimization enables deep recurrent neural network," Computer Communications, vol. 202, pp. 145–153, 2023.

[11] Nadir Omer, Ahmed H. Samak, Ahmed I. Taloba, and Rasha M. Abd El-Aziz, "A novel optimized probabilistic neural network approach for intrusion detection and categorization," Alexandria Engineering Journal, vol. 72, pp. 351–361, 2023.

[12] Mariya Princy Antony Saviour and Dhandapani Samiappan, "IPFS based storage Authentication and access control model with optimization

enabled deep learning for intrusion detection," Advances in Engineering Software, vol. 176, pp. 103369, 2023.

[13] D. Suja Mary, L. Jaya Singh Dhas, A. R. Deepa, Mousmi Ajay Chaurasia, and C. Jaspin Jeba Sheela, "Network intrusion detection: An optimized deep learning approach using big data analytics," Expert Systems with Applications, vol. 251, pp. 123919, 2024.

[14] Man Wang, "Optimization of Network Security in University Laboratories Based on Anomaly Intrusion Detection in Public Cloud Networks," Computers and Electrical Engineering, vol. 111, pp. 108968, 2023.

[15] Lan Xia and Xuefei Xia, "Network Security Intrusion Detection Methods Combining Optimization Algorithms and Neural Networks," Procedia Computer Science, vol. 228, pp. 582–592, 2023.

[16] Samed Al and Murat Dener, "STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment," Computers & Security, vol. 110, pp. 102435, 2021.

[17] Feilu Hang, Linjiang Xie, Zhenhong Zhang, Wei Guo, and Hanruo Li, "Research on the application of network security defence in database security services based on deep learning integrated with big data analytics," International Journal of Intelligent Networks, vol. 5, pp. 101–109, 2024.

[18] Fuhua Huo, "Computer network big data detection based on internet of things technology," Measurement: Sensors, vol. 33, pp. 101222, 2024.

[19] S. H. Mousavi, M. Khansari, and R. Rahmani, "A fully scalable big data framework for Botnet detection based on network traffic analysis," Information Sciences, vol. 512, pp. 629–640, 2020.

[20] Dibin Shan, Xuehui Du, Wenjuan Wang, Na Wang, and Aodi Liu, "KPI-HGNN: Key provenance identification based on a heterogeneous graph neural network for big data access control," Information Sciences, vol. 659, pp. 120059, 2024.

[21] Yan Wang et al., "An evolutionary computation-based machine learning for network attack detection in big data traffic," Applied Soft Computing, vol. 138, pp. 110184, 2023.

[22] Fang Xu, Qiang Chen, Qi Liu, and Ning Li, "Intelligent Analysis Algorithm for Hidden Danger Identification of Intelligent Network Monitoring System from the Perspective of Big Data," Procedia Computer Science, vol. 228, pp. 57–63, 2023.

[23] Ijaz Ahmad, Zhong Wan, and Ashfaq Ahmad, "A big data analytics for DDOS attack detection using optimized ensemble framework in Internet of Things," Internet of Things, vol. 23, pp. 100825, 2023.

[24] Hui Gao, "Design of Network Data Information Security Monitoring System Based on Big Data Technology," Procedia Computer Science, vol. 228, pp. 348–355, 2023.
[25] Ramkumar M.P., P. V. Bhaskar Reddy, J. T. Thirukrishna, and Ch Vidyadhari, "Intrusion detection in big data using hybrid feature fusion and optimization enabled deep learning based on spark architecture," Computers & Security, vol. 116, pp. 102668, 2022.
[26] A. Ponmalar and V. Dhanakoti, "An intrusion detection approach using ensemble Support Vector Machine based Chaos Game Optimization algorithm in big data platform," Applied Soft Computing, vol. 116, pp. 108295, 2022.
[27] A. Satish Kumar and S. Revathy, "A hybrid soft computing with big data analytics based protection and recovery strategy for security enhancement in large scale real world online social networks," Theoretical Computer Science, vol. 927, pp. 15–30, 2022.
[28] Tianyue Zhang, Wei Chen, Yuxiao Liu, and Lifa Wu, "An intrusion detection method based on stacked sparse autoencoder and improved gaussian mixture model," Computers & Security, vol. 128, pp. 103144, 2023.
[29] Liu Zhiqiang, Ghulam Mohiuddin, Zheng Jiangbin, Muhammad Asim, and Wang Sifei, "Intrusion detection in wireless sensor network using enhanced empirical based component analysis," Future Generation Computer Systems, vol. 135, pp. 181–193, 2022.

Biographies

Jizhou Shan is an associate professor working on Hainan College of Economics and Business. He graduated from Northeastern University and has dedicated his research to the field of Computer Networks and Simulation, publishing over 30 papers.

Hong Ma is a professor at Hainan College of Economics and Business. She graduated from Northeastern University and specializes in Technology for Computer Applications. Her research interests include communication security and computer technology and applications.