Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM

ISSN: 2788-7421

Identification Method of Power Internet Attack Information Based on Machine Learning

Yitong Niu[1,*], **Korneev Andrei**[1]

[1]Belarusian-Russian University, Mira Ave 43, Mogilev, 212000, BELARUS

*Corresponding Author: Yitong Niu

DOI: https://doi.org/10.52866/ijcsm.2022.02.01.001
Received October 2021; Accepted December 2021; Available online February 2022

ABSTRACT: To solve the problem of large recognition errors in traditional attack information identification methods, we propose a machine learning (ML)-based identification method for electric power Internet attack information. Based on the Internet attack information, an Internet attack information model is constructed, the identification principle of the power Internet attack information is analysed based on ML, hash fixing is conducted to ensure that the same attack information will be assigned to the same thread and that the deviation generated by noise can be avoided so that the real-time lossless processing of the power Internet attack information can be ensured. The vulnerability adjacency matrix is constructed, and the vulnerability is quantitatively evaluated to complete the design of the optimal identification scheme for power Internet attack information. The experimental results show that the identification accuracy of the method can reach 98%, which can effectively reduce the risk of power Internet network attacks and ensure the safe and stable operation of the network.

Keywords: Machine learning; Power Internet; Attack information; Recognition

1. INTRODUCTION

To meet the needs of users, the power Internet needs to understand various applications in the network, whether it is managing network resources or transforming and upgrading the network [1]. In this context, the identification of attack information of the electric power Internet can effectively solve the aforementioned problems [2].

... [3] proposed a power Internet attack information identification method based on the integrated weight method, which uses the integrated weight theory to divide the power Internet attack graph generation process into multiple sub-regions, each conducting the generation of attack information graphs in two power Internet information sub-networks, and aggregating all the sub-attack graphs based on this information to complete the power Internet attack information identification. Although this method is more scalable, it takes a longer time. ... [4] proposes an attack information identification method based on the TOPSIS algorithm, which uses the algorithm to construct an attack information graph, generate an attack identification optimization scheme, and identify the attack information in the power Internet. Although this method can reduce the time of attack information identification, it encounters poor attack information recognition. To address this problem, we propose a method based on machine learning (ML).

2. PRINCIPLE OF ATTACK INFORMATION IDENTIFICATION

After constructing the power Internet attack information model, the attack behaviour in the power Internet space is modelled and analysed, and the results of the analysis are used to guide the defence against Internet attacks. As the Internet model cannot describe the attack persistence, it cannot reflect the characteristics of the attack scenario. Therefore, the Internet attack information model is extended as shown in Figure 1.

*Corresponding author: itong_niu@163.com

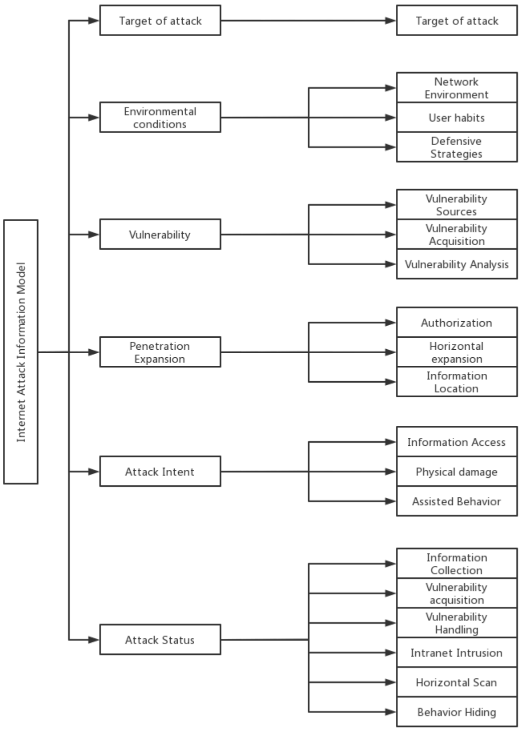

FIGURE 1. Internet Attack Information Model

ML methods are used to simulate human learning behaviour to acquire new knowledge [5]. The environment provides certain information to the learning part of the system and the system uses this information to modify the knowledge base to improve the ability to perform the task of information recognition by attacking it while feeding the acquired information back to the learning part [6]. The information provided by the environment is the main factor that affects ML, and if the information is of high quality, then the ML part of the information is easy to process compared with the general principle. If specific information is provided to the learning system, then the learning system needs to remove the redundant information after acquiring enough information and then summarize and extend it to form the guiding action principles [7]. Through the ML identification method, the correct rules can be left behind and the incorrect rules can be removed from the database.

In the identification process, the following four principles have to be followed: strong expressive ability, easy reasoning, easy modification of the knowledge base, and easy expansion [8]. Based on this, the power of Internet attack information is identified. According to the form of learning, ML is divided into three types, which are supervised learning, unsupervised learning, and semi-supervised learning. ML aims to continuously collect new information from the environment and update the already learned knowledge system, which can continuously improve the recognition efficiency of electric power Internet attack information [9–11].

The main identification process involves the following steps:

1) collecting relevant information from the power Internet attack information and eliminating redundant information,

2) storing the correct and valuable knowledge information in the knowledge base,

3) summarising and processing the valid information collected,

4) solving the problem of power Internet attack based on the laws that already exist in the knowledge base, and

5) collecting valid information from the power Internet and transferring it to the learning module to conduct attack information identification [12].

3. REAL-TIME LOSSLESS PROCESSING OF ATTACK INFORMATION

Efficient thread planning based on supervised ML, combined with hash fixing, guarantees that the same attack messages will be assigned to the same thread for processing to avoid bias [13, 14]. The thread pool scheduling management is used to avoid mutual exclusion amongst a large number of attack messages. The structure of the lossless processing technique is shown in Figure 2.

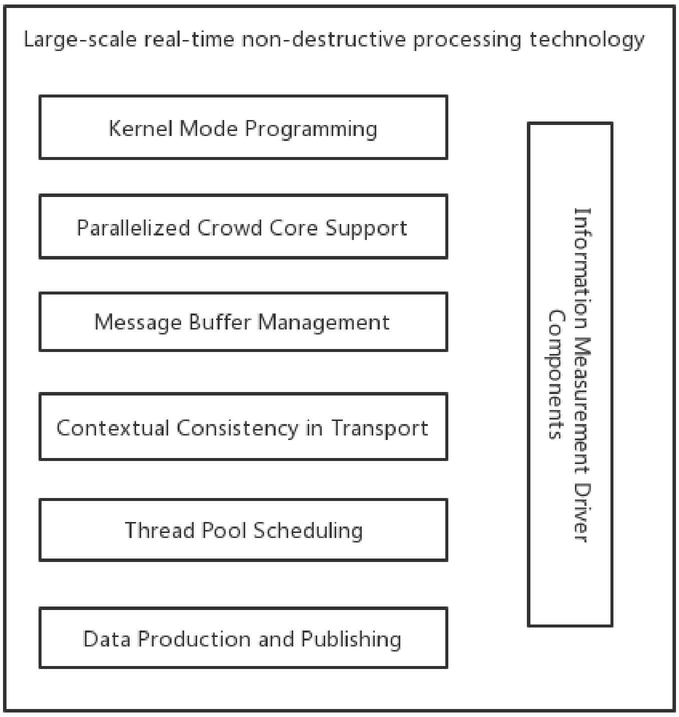
Large-scale real-time non-destructive processing technology
Kernel Mode Programming
Parallelized Crowd Core Support
Message Buffer Management
Contextual Consistency in Transport
Thread Pool Scheduling
Data Production and Publishing
Information Measurement Driver Components

FIGURE 2. **Real-time lossless processing of electrical attack information**

The implementation of lossless processing of power Internet attack information can effectively exploit parallel computing capabilities and use passive thread scheduling to dynamically adjust the number of threads used for analysis and processing, thereby achieving the real-time processing of power Internet attack information [15].

4. OPTIMAL IDENTIFICATION OF ATTACK INFORMATION

4.1 ASSEMBLY OF VULNERABILITY ADJACENCY MATRIX

Based on the assumption that the security vulnerabilities in the power Internet are potential attack information, the element $\mathbf{q}_{i,j}$ in the vulnerability adjacency matrix represents the possibility of successfully connecting to node **j** through further attacks from the power Internet attacking node **i**. Then, $\mathbf{q}_{i,j}$ is calculated based on the following formula:

$$q_{i,j} = \begin{cases} \alpha_{acc,j} & i = 1 \\ \beta_{i,j} & i \neq 1 \end{cases} \tag{1}$$

When **i** = 1 or **j** = 1, the place is an attack node.$\alpha_{acc,j}$ indicates the possibility of the selected node appearing as a parallel attack access point, and $\beta_{i,j}$ denotes the possibility of attack information appearing between two nodes [16].

Using ML methods to calculate the N-step vulnerability matrix of the length of the power Internet attack, and then the elements of each matrix corresponding to the position of the summation process, we can calculate the possibility of selecting the node as the attack node, that is, the quantitative value of vulnerability of each network node, expressed by the following formula:

$$W_{i,j} = \sum_{n=1}^{N} q_{i,j} \tag{2}$$

where $\mathbf{W}_{i,j}$ denotes the quantified value of network node vulnerability.

Based on the assumption that the impact of each access point on the power Internet is different after being connected, the probability index of the access point, when any security hole in the power Internet is attacked, is calculated as follows:

$$E_{s,i} = \begin{cases} \dfrac{T_{v,i}T_{a,i}}{\sum_{j\in acc} T_{v,j}T_{a,j}} & i \in S_{acc} \\ 0 & i \notin S_{acc} \end{cases} \tag{3}$$

In equation (3), $\mathbf{T}_{\mathbf{v,i}}$ represents the value of security vulnerabilities, $\mathbf{T}_{a,i}$ represents the difficulty of accessing power Internet security vulnerabilities, $\mathbf{T}_{v,j}$ represents the speed of access to power Internet security vulnerabilities, $\mathbf{T}_{a,j}$ represents the risk of power Internet security vulnerabilities, and $\mathbf{S}_{acc}$ represents a collection of security vulnerabilities that may be selected as attack points in the power Internet [17].

Based on the assumption that $\mathbf{Q}_{v,i}$ represents the impact on the network caused by a successful attack on the power Internet security vulnerability, the following formula is obtained:

$$Q_{v,i} = A_{acc,i} \cdot \theta_i \tag{4}$$

In equation (4), $\mathbf{A}_{acc,i}$ represents the attractiveness of the power equipment resource where the security breach is located, and $\theta_{\mathbf{i}}$ represents the impact factor of the security breach [18].

4.2 DESIGN OF ATTACK INFORMATION OPTIMISATION IDENTIFICATION SCHEME

In the optimal identification of power Internet attack information, the aforementioned vulnerability adjacency matrix is used as the basis to obtain attack information, calculate the probability of a successful attack for each attack path, and obtain the parallel attack metric, which is used as the basis to identify power Internet attack information. The optimised identification process of the power Internet attack information is shown in Figure 3.

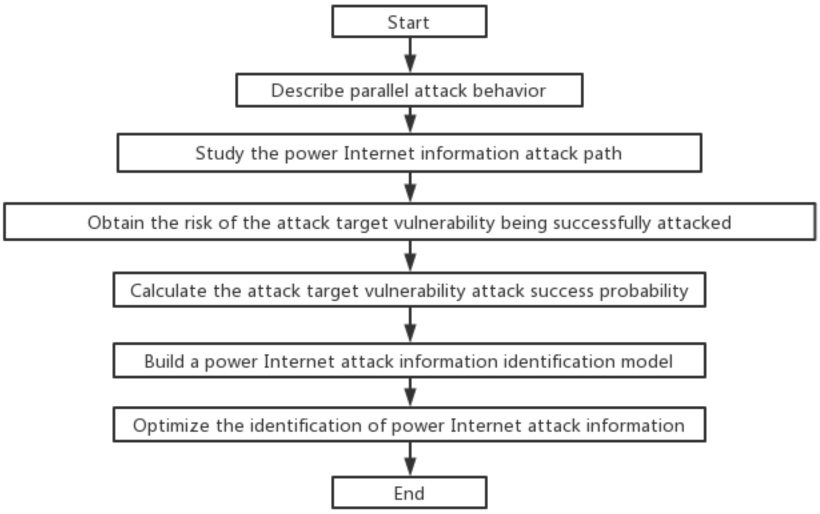
Start
Describe parallel attack behavior
Study the power Internet information attack path
Obtain the risk of the attack target vulnerability being successfully attacked
Calculate the attack target vulnerability attack success probability
Build a power Internet attack information identification model
Optimize the identification of power Internet attack information
End

FIGURE 3. **Process of attack information optimisation identification**

5. EXPERIMENTAL VERIFICATION

To verify the ML-based information of power Internet attack identification method, we conduct experimental validation on the effectiveness of its application in analysing overall network security [19].

5.1 EXPERIMENTAL ENVIRONMENT

The experimental environment includes 5 hosts and 20 network nodes. The attacker can achieve 50% success rate by obtaining user privileges. The experimental environment is shown in Figure 4.

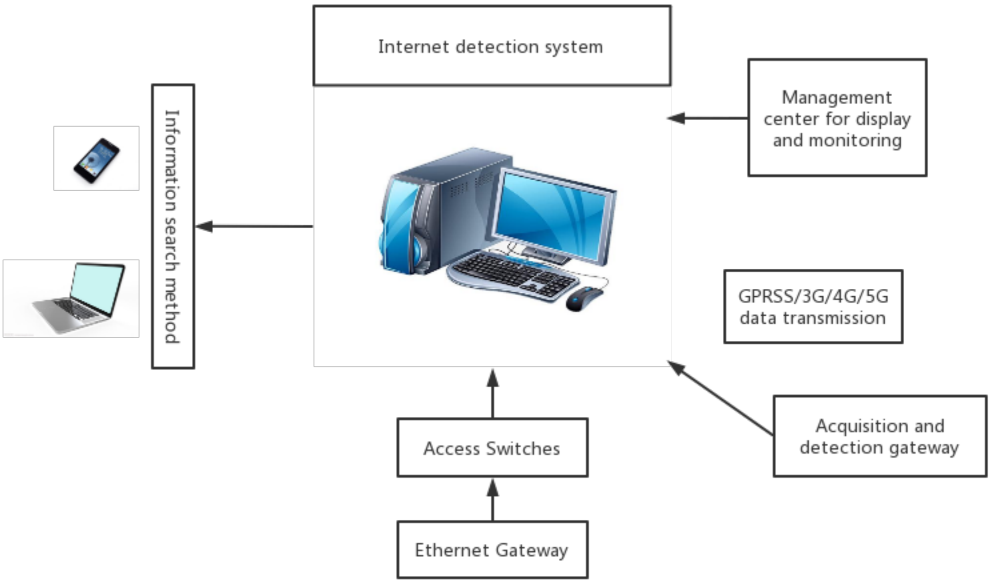
Internet detection system
Management center for display and monitoring
Information search method
GPRSS/3G/4G/5G data transmission
Acquisition and detection gateway
Access Switches
Ethernet Gateway

FIGURE 4. **Experimental environment**

5.2 EXPERIMENTAL PARAMETER SETTINGS

The experimental parameters are set as shown in Table 1.

Table 1. **Experimental parameter settings**

	Parameter	Numerical value
Frequency of information collection	40 kHz	
Network signal window	Window size	30 dimensions
	Split window length	35 ms
	Split frame	250 points
	Frame shift	55 points

The experimental environment includes 5 hosts and 20 network nodes. The attacker can achieve 50% success rate by obtaining user privileges. The experimental environment is shown in Figure 4.

5.3 EXPERIMENTAL RESULTS AND ANALYSIS

Due to the potential risks of vulnerability in the power Internet, the power Internet attack information identification method based on the comprehensive weight method, the attack information identification method based on the TOPSIS algorithm, and the ML identification method are used to identify the attack information. These three methods are compared and the short-time energy comparative analysis is shown in Figure 5.

A comparison of the three methods to analyse the parallel attack fragility of the short-time energy vulnerability is presented in Figure 6.

As shown in Figure 6, the network vulnerability of the three methods is different. The network vulnerability of the identification method based on ML is kept below 10%, that of the identification method based on the integrated weight method is kept between 30% and 40%, and that of the identification method based on TOPSIS algorithm is kept above 40%. The identification method based on ML identifies the attack information by determining the possibility of any security vulnerability in the power Internet being selected as an attack contact point indicator, which reduces the network vulnerability [20].

To further verify the rationality of the proposed method, we verify the accuracy of the comparative analysis of the three methods. The results are shown in Table 2.

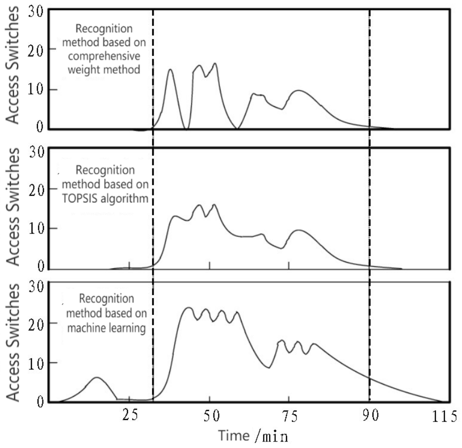

FIGURE 5. **Short-time energy comparison analysis of three methods**

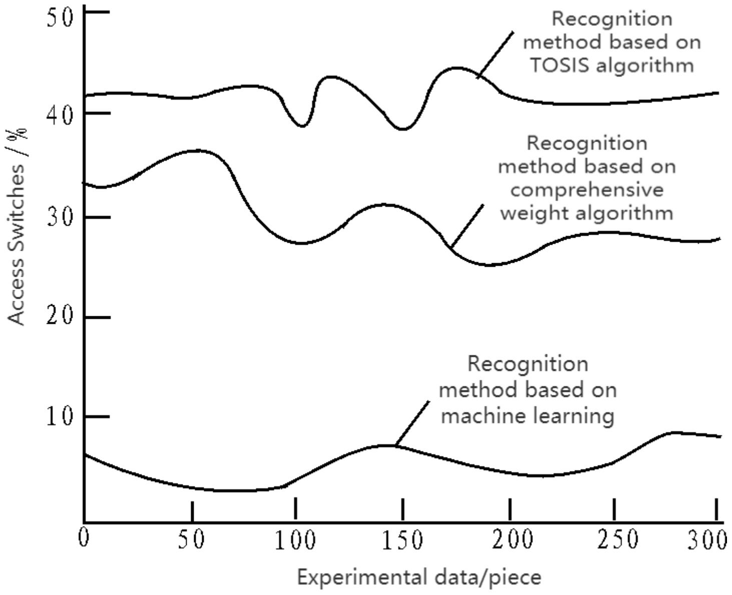

FIGURE 6. **Three methods to attack vulnerability in parallel**

Table 2. **Three methods of recognition accuracy**

Number of experiments/time	1	2	3	4	5
Recognition method based on comprehensive weight	52%	67%	63%	62%	58%
	52%	61%	35%	50%	52%
	50%	54%	52%	50%	49%
	25%	28%	26%	31%	29%
Recognition method based on TOPSIS algorithm	65%	55%	69%	55%	45%
	54%	52%	65%	50%	42%
	52%	48%	42%	42%	38%
	38%	31%	30%	29%	20%
Recognition method based on machine learning	98%	98%	97%	97%	97%
	97%	96%	97%	95%	96%
	93%	92%	92%	93%	91%
	91%	92%	90%	90%	90%

According to the preceding experimental comparison results, amongst the three identification methods, the identification accuracy of the ML-based power Internet attack information identification method is the highest.

6. CONCLUSION

To ensure the efficient identification of electric power Internet attack information, this study proposed an ML-based identification method of electric power Internet attack information. Based on the research on electric power Internet attack information identification and ML, the proposed method can adapt to the characteristics of Internet big data. On this basis, the ML identification method is selected to ensure that the attack information identification error is minimised. The experimental results show that the method can identify the attack information efficiently.

ACKNOWLEDGEMENTS

The first author thanks the reviewers for their useful suggestions that improved the presentation of this paper.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Z. Dong, X. Tang, and J. Cheng, “DDoS attack detection method based on HMM time series prediction and chaos model,” *Computer Engineering and Science*, vol. 40, no. 12, pp. 72–80, 2018.

[2] L. Ma and Y. Shu, “DDoS attack detection model based on machine learning algorithm in SDN environment,” *Microelectronics and Computer*, vol. 35, pp. 15–20, 2017.

[3] J. Wen, X. Shi, J. Shao, C. Xin, and L. Ni, “Identification method of vulnerable lines in power system based on comprehensive weight method,” *Electrical and Electrical Engineering*, no. 6, pp. 10–14, 2019.

[4] Z. Geng, L. Cui, and S. Qin, “Identification of key nodes in power communication network based on TOPSIS algorithm,” *Power System Protection and Control*, vol. 46, no. 1, pp. 78–86, 2018.

[5] S. Wang, H. Du, and Y. Meng, “Research on Vehicle Pavement Type Recognition Technology Based on Machine Learning,” *Journal of Military Technology*, vol. 38, no. 8, pp. 189–195, 2017.

[6] J. Peng, P. Chang, and S. Zhang, “Application of fast verification method of telecontrol information in smart substation based on simulation technology,” *Electronic Design Engineering*, vol. 26, pp. 52–56, 2018.

[7] H. Zhao, H. Qi, and X. Wang, “Research on intention perception and control method of man-machine coordination operation based on machine learning,” *Machine Tool and Hydraulics*, vol. 47, pp. 147–150, 2019.

[8] T. Tu and W. Jin, “Radar emitter signal recognition based on automatic machine learning process optimization,” *Computer Applied Research*, vol. 36, no. 1, pp. 197–199, 2019.

[9] X. Lin and J. Sun, “Research on small target detection and tracking algorithm based on machine learning,” *Research on Computer Application*, vol. 35, no. 11, pp. 256–259, 2018.

[10] F. Hu, C. Li, and M. Wang, “SQL injection detection scheme based on machine learning,” *Computer Engineering and Design*, vol. 22, no. 6, pp. 61–65, 2019.

[11] S. Zhao and S. Chen, “Overview and prospect of traffic identification technology based on machine learning,” *Computer Engineering and Science*, vol. 40, no. 10, pp. 34–44, 2018.

[12] Y. Tan, S. Liu, and X. Lu, “Chinese text implication recognition method based on CNN and bidirectional LSTM”,” *Journal of chinese information*, vol. 32, no. 7, pp. 16–24, 2018.

[13] F. Qin, X. Liang, and F. Zhang, “Building target extraction method of array tomography SAR based on machine learning,” *Signal Processing*, vol. 35, no. 2, pp. 22–32, 2019.

[14] Y. Ma, G. He, and B. Zhang, “Simulation study on optimal identification of attack information in power resource network,” *Computer Simulation*, vol. 22, no. 6, pp. 18–19, 2017.

[15] W. Ma, “Research on Security Encryption of Information Collection under Mobile Internet,” *Electronic Design Engineering*, vol. 26, pp. 52–56, 2018.

[16] A. H. Ali, M. Z. Abdullah, S. N. Abdul-Wahab, and &amp; M Alsajri, “A Brief Review of Big Data Analytics Based on Machine Learning”,” *Iraqi Journal For Computer Science and Mathematics*, vol. 1, no. 2, pp. 13–15, 2020.

[17] Y. Shu and X. &amp; Li Cheng-Cheng, *Research on the contribution of regional Energy Internet emission reduction considering time-of-use tariff*, vol. 239. 2022.

[18] S. Sudhakar, K. I. Osamah, S. K. Dilip, H. A. J. A. &amp; A, and Prabhu, “Secured and Privacy-Based IDS for Healthcare Systems on E-Medical Data Using Machine Learning Approach”,” *International Journal of Reliable and Quality E - Healthcare*, vol. 11, no. 3, pp. 1–11, 2022.

[19] A. Sayghe, Y. Hu, I. Zografopoulos, X. Liu, R. G. Dutta, Y. Jin, and &amp; C Konstantinou, “Survey of machine learning methods for detecting false data injection attacks in power systems,” *IET Smart Grid*, vol. 3, no. 5, pp. 581–595, 2022.

[20] H. Karimipour, A. Dehghantanha, R. M. Parizi, K. K. R. Choo, and &amp; H Leung, “A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids,” *IEEE Access*, vol. 7, pp. 80778–80788, 2019.