Research on Structure Data Link Detection Method Based on Internet of Things System

Wenwu Yu*, Yue Zhai, Yang Guo, Meijiao Ji
Dalian University of Science and Technology
yww_12@edu.cn

Abstract: In order to improve the fault diagnosis and analysis capabilities of the cyber-physical system in the Internet of Things, it is necessary to detect and identify the vicious data chain of the cyber-physical system in the Internet of Things. Moreover, to realize the detection and identification of the vicious data link in the cyber-physical system in the Internet of Things, a fast extraction algorithm for the vicious data link of the cyber-physical system in the Internet of Things is proposed in the paper to collect vicious data from the cyber-physical system in the Internet of Things. In addition, the malignant data generated during the factor adjustment of the Internet of Things can be divided into periodic structure and non-periodic structure in structure, and the malignant data sensor network model of the cyber-physical system in the Internet of Things is constructed. Therefore, through simulation experiment analysis, the method proposed in the paper has superior performance in the detection and identification of vicious data links in the cyber-physical system of the Internet of Things.

1. Introduction
The Internet of Things is a cyber-physical system composed of physical networks, information equipment, and computing unit modules, which has the advantages of diverse operating modes and high compatibility. However, affected by the use environment of the cyber-physical system in the Internet of things and the technical index parameters of the equipment, malignant data is generated in the cyber-physical system of the Internet of Things [1]. The malicious data of the cyber-physical system in the Internet of Things mainly refers to the user’s abnormal deduction information, malicious attack data, and personal information leakage. What is more, the existence of malicious data not only hinders the operating efficiency of the Internet of Things system, but also causes economic losses to users. In order to solve the above-mentioned problems, it is necessary to add a device for detecting malicious data in the system to realize the multi-level control of the system and accurately identify and send information flow. However, this approach will make the system structure more complex and expensive, leading to uncertainty in the operation of the cyber-physical system in the Internet of Things. Therefore, in order to improve the information output and forwarding control capabilities of the Internet of Things, the research on the detection method of the vicious data link in the cyber-physical system of the Internet of Things is of great significance in the optimization design of the Internet of Things [2,3]. At present, in the field of theoretical analysis of cyber-physical systems in the Internet of Things, some preliminary research results have been obtained at home and abroad.

Considering the periodicity of malignant data, the time series node distance of the data is calculated to obtain its high-dimensional mapping structure, which provides a limited environment for the accuracy of malignant data detection and enhances the detection accuracy. Meanwhile, through the linear fitting of the multiple collinearity feature in the malignant data, the malignant data link is
obtained to calculate the load of the malignant data link in the merged cluster. Then, the fuzzy clustering of the characteristics in the vicious data chain is carried out to merge the feature similar data into one, saving the time and cost of data detection.

2. Vicious Data Chain Collection and Statistical Analysis Model of the Cyber-physical System in the Internet of Things

2.1 High-dimensional Mapping Structure of Malignant Data

First, supposing that the distribution time series of malignant data features in the sensor network model is \{M_n\}, \(n = 1,2,\cdots,N\), each cluster in the distribution time series has a cluster head node (S\(_n\)) and several cluster nodes (V\(_0\)). The distance between the nodes of the malignant data time series is solved by the Euclidean distance formula\([4]\), the calculation formula is as follows:

\[
d(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}
\]

Among them, \(x_i, y_i, x_j, y_j\) represents the abscissa and ordinate of the malignant data time series node \(i\) and node \(j\) in the cyber-physical system of the Internet of Things, and \(d(i,j)\) refers to the Euclidean distance between the two nodes\([5]\).

Based on the node distance of the malignant data time series and combined with the big data fusion method, the similarity fusion of the energy features in the malignant data is performed\([6]\), and the fusion result is:

\[
E(L) = LE_i
\]

Among them, \(E_i\) represents the load of the malignant data node (intermediate node) \(i\); \(L\) is the transmission link set of malignant data characteristics. The load of the transmission load information sent to the i-node in the cyber-physical system of the Internet of Things is:

\[
C(n_j) = E_i l \delta + E(L)
\]

In the formula, \(E(L)\) is the similarity fusion result, \(\delta\) refers to the load ratio, and \(l\) indicates the number of transmission chains. According to the results of load fusion, the attribute partition characteristics of the malignant data in the Internet of Things cyber-physical system is mined to adopt the multi-mode fusion method, so that the collection of vicious data in the Internet of Things cyber-physical system is realized\([7]\). Moreover, according to the collected malignant data, the spectral offset characteristics of the data feature quantity are extracted, and machine learning algorithms are used to match the spectral features of the malignant data of the cyber-physical system in the Internet of Things to the extracted spectral features. Combined with the big data output time delay, the characteristic space of the vicious data information reorganization of the cyber-physical system in the Internet of Things is obtained. The expression is:

\[
X_n = \{X_n, X_{n-1}, X_{n-2}, \cdots, X_{n-(d-1)r}\}
\]

Mining the average mutual information amount of the vicious data in the cyber-physical system of the Internet of Things in the feature space of information reorganization, the mutual information distribution set is obtained as:

\[
R_d = \{R_1, R_2, \cdots, R_d\}
\]

In the formula, \(R_d\) represents the \(d\) information amount of mutual information distribution. Parallel mining method is used to mine the association rules of the malignant data characteristics in the cyber-physical system of the Internet of Things\([8]\). Unitary decomposition is performed on the mutual information distribution matrix, the vector set of eigen decomposition is:

\[
T_e = \{X_1, X_2, \cdots, X_m\}
\]

In the formula, \(X_m\) represents \(m\) vectors in the feature space of information reorganization. Meanwhile, the distributed fusion method with singular value feature is used to perform high-dimensional mapping on the malignant data of the IoT cyber-physical system, and the data
structure model of the malignant data of the IoT cyber-physical system in the high-dimensional mapping space is obtained. This model can be described by a two-dimensional matrix model:

\[
V = \frac{C(n)}{\sum (R^T R)}
\]  

In the formula, \(T\) represents the malignant data processing time. The high-dimensional mapping structure of the malignant data in the cyber-physical system of the Internet of Things is obtained by formula (8), and statistical analysis of the data can be performed on this structure.

### 2.2 Linear Fit of Malignant Data

Based on the high-dimensional mapping structure of the malignant data in the cyber-physical system of the Internet of Things, adaptive learning and intelligent detection of the parameters such as the scale and delay of the malignant data are carried out to realize the spectral peak search of the malignant data in the cyber-physical system of the Internet of Things [9]. Among them, the scale parameters of the malicious data in the cyber-physical system of the Internet of Things are:

\[
S_i = \frac{L\epsilon_d d^2}{V}
\]  

In the formula, \(\epsilon_d\) represents the characteristic parameter of malignant data; \(d\) refers to the delay parameter. The set of delay parameters for malignant data is:

\[
d_i = \{X_{d_{11}}, X_{d_{12}}, \cdots, X_{d_{1n}}\}
\]

In the formula, \(X_{d_{im}}\) is the set of delay parameters in the m-sequence. The association rule mining method is used to restructure the malignant data of the cyber-physical system in the Internet of Things. The delay parameters of the malignant data satisfy the balance between characteristic parameters and delay of the malignant data.

Therefore, the malignant data sequence analysis model of the cyber-physical system in the Internet of Things is constructed [10]. Meanwhile, the relevant characteristic data mining method is adopted to carry on the inspection statistical analysis, and the cluster processing is performed to the malicious data of the cyber-physical system in the Internet of Things. In the distributed sensor storage medium of the malicious data in the cyber-physical system of the Internet of Things, the output load balance scheduling model is:

\[
\eta^r(\omega) = d_i \{T^r_i | T^r_i > \omega\}, k \in R, w \in W
\]

Among them, \(T^r_i\) represents the test statistics for rapid extraction of malignant data. The test statistic \(T^r_i\) for the rapid extraction of malignant data in the cyber-physical system of the Internet of Things can be expressed as:

\[
T^r_i = E\left[\min_{w} |H_{w, i}^r| w \right] = \frac{1}{\theta} \ln \sum_{w, h} \exp \eta^r_i (\omega), w \in W, h \in H
\]

In the formula, \(\eta^r\) represents the linear fitting value, and \(\theta\) is the test adjustment parameter. The linear fitting of the multicollinear characteristics of the malignant data in the cyber-physical system of the Internet of Things needs to satisfy \(T^r_i < [H_{w, i}^r]\). The spectral peak \(\zeta^r_i(\omega)\) of the malignant data in the cyber-physical system of the Internet of Things can be expressed as:

\[
[H_{w, i}^r] = \min \left\{ \frac{E[T^r_i]}{\gamma_r^r(\omega) \omega + \xi} \right\} = E[T^r_i] + \gamma^r_i(\omega) \quad k \in R, w \in W
\]

In the formula, \(\gamma^r_i\) represents the maximum peak fitting value, and \(\xi\) refers to the multiple threshold value. Finally, the generalized least squares method is used to linearly fit the malignant data of the cyber-physical system in the Internet of Things to obtain a complete malignant data chain. The fitting process is:

\[
E_{\gamma_r}(L, d) = \begin{cases} 
L\eta^r_i (\omega) + S_i, S > Q \\
S_i, S \leq Q
\end{cases}
\]

In the data fitting process, the characteristic space area of the cyber-physical system data in the
Internet of things is divided into \( S \) and \( Q \), and the solution vector of the endogenous spectrum characteristic quantity \( S \) in the malignant data of the cyber-physical system of the Internet of things is composed of \[ \{s_1, s_2, \ldots, s_n\} \]
then the entropy distribution probability of system malignant data is \( P_i(s_j), \quad i = 1, 2, \ldots, n \), and \( Q \) are composed of the solution vector \( \{q_1, q_2, \ldots, q_n\} \) in the big data fuzz test set. In summary, the fitting of malignant data is completed. By fitting the malignant data, it lays a theoretical foundation for the detection of the malignant data link in the cyber-physical system of the Internet of Things.

3. Malicious Data Link Detection and Identification

3.1 Vicious Data Link Spectral Feature Model

The spectrum feature extraction method is used for the acquired malignant data link to analyze the sample distribution characteristics of the malignant data link in the physical system of the Internet of Things, and the malignant data link is detected according to the distribution characteristics. The detection probability of the corresponding malignant data link in the cyber-physical system of the Internet of Things is \( P_i(q_j), \quad j = 1, 2, \ldots, n \). Under extreme learning training, a model is constructed for reorganizing the characteristics of the vicious data chain in the cyber-physical system of the Internet of Things:

\[
H(s) = -\sum_{i=1}^{n} P_i(s_i) \log_2 P_q
\]  

(14)

In the formula, \( P_i(s_i) \) represents the probability that the concept set \( s_i \) of the malicious data chain feature distribution in the cyber-physical system of the Internet of Things appears in the affine partition area \( S \), and similarly, \( P_q \) refers to the concept set of the malicious data chain feature of the cyber-physical system in the Internet of Things.

Initializing the data link \( A \) from the cluster center to the points within the cluster, the average mutual information of the vicious data link in the cyber-physical system of the Internet of Things can be obtained as follows:

\[
q^* = \sum_{s \in S} H(s)^* \quad W \in W
\]  

(15)

Constructing the vicious data link model of the cyber-physical system in the Internet of Things, the spectrum of the vicious data link in the cyber-physical system of the multi-carrier Internet of Things is decomposed. Moreover, the link random allocation method is used to control the output sensor sequence of the vicious data link in the cyber-physical system of the Internet of Things. The characteristic decomposition sub-sequence of the malicious data chain in the cyber-physical system of the Internet of Things is expressed as:

\[
r_i(n) = \exp\{jq^nT + \theta\}, \quad n = 0, 1, \ldots, (N-3)/2
\]  

(16)

The discrete transformation of point \( (N-1)/2 \) is performed according to \( r_1(n) \) and \( r_2(n) \) to extract the spectral feature quantity of the vicious data chain in the cyber-physical system of the Internet of Things. The spectral feature quantity extraction model is:

\[
U_k = \sum_{i=1}^{n} T_i^* + r_2(n) \quad H(s)
\]  

(17)

In the formula, \( T_i^* \) represents the extraction time of the data link at different output loads. On this basis, the classification and detection of the malignant data link can be carried out based on the extracted spectral features.
3.2 Classification and Detection of Vicious Data Links in Cyber-physical Systems of the Internet of Things

According to the results of the spectral feature system, the vicious data link of the cyber-physical system in the Internet of Things is detected, and the neural network analysis method is used to optimize the detection of the vicious data link in the cyber-physical system of the Internet of Things, so that the load of the vicious data link in the merged cluster can be obtained:

\[ R_i(k) = r_i(n) \exp(-j\omega_T^k / 2), \quad k = 0, 1, \ldots, (N-3)/2 \]

\[ R_i(k) = A_i \exp(j\theta_k), \quad k = 0, 1, \ldots, (N-3)/2 \]

Among them, \( a_i \) is the load prediction error of the vicious data link in the cyber-physical system of the Internet of Things, \( \omega_T \) refers to the time window, and \( A_i \) represents the characteristic offset amplitude of the vicious data link in the cyber-physical system of the Internet of Things. \( \theta_k \) is the output expansion phase. Moreover, linear prediction is performed on the malignant data chain of the cyber-physical system in the Internet of Things to obtain the maximum length of the data block in each merged cluster according to the global optimization result. Then adaptive blind separation processing is carried out on the vicious data chain of the cyber-physical system in the Internet of Things. Besides, given that the solution space of the objective function is from \( R^* \) to \( R \), the outlier cluster \( U \in R^* \) of the vicious data chain in the cyber-physical system of the Internet of Things can be obtained. In other words, according to the spectral feature extraction model established above, a point in the data chain \( A \) of the points in the cluster is found, and the adaptive neural network learning algorithm is used to perform fuzzy clustering of the vicious data chain features in the cyber-physical system of the Internet of Things. When the decision threshold of data clustering meets:

\[ 0 \leq p_{k+1} \leq R_i(k) - p_k \leq 1 \]  \hspace{1cm} (20)

With the help of the fuzzy clustering analysis method, \( N \) data clustering centers are initialized to perform clustering processing of malignant data chains and quickly extract the malignant data chains in the cyber-physical system of the Internet of Things:

\[ \min \text{imize} \quad \frac{1}{2} \| \hat{w} \|^2 + C \sum_{i=1}^{n} \xi_i \]

\[ \text{subject to} \quad y_i - (w^T \Phi(x_i)) \leq \varepsilon - \xi_i \]

\[ \xi_i \geq 0, i = 1, 2, \ldots, n; C > 0 \] \hspace{1cm} (21)

In the formula, \( \Phi(x_i) \) represents the feature space function of malignant data information reorganization, \( C \) refers to the decision threshold, and \( \xi_i \) is the peak of the spectrum set. So far, the vicious data chain of the cyber-physical system in the Internet of Things has been quickly extracted.

4. Simulation Experiment and Result Analysis

In order to test the performance of the method proposed in the paper on realizing the intelligent detection of the vicious data chain in the cyber-physical system of the Internet of Things, a simulation experiment is carried out.

The experiment is based on the Matlab 7 simulation tool, adopting Intel(R)Core(TM)2 Duo CPU 2.94GHz processor. The array of vicious data chain distribution in the cyber-physical system of the Internet of Things is 200*400, and the sampling bandwidth of the vicious data link in the cyber-physical system of the Internet of Things is 12s. In addition, an information data sample with a size of 1024MB is arbitrarily selected in the Oracle system information library. Meanwhile, 27MB of abnormal deduction information, malicious attack data, personal data leakage and other vicious data information are added to it. The period length of the initial sample sequence sampling is T=0.16 s, which is the basis of the vicious data chain in the cyber-physical system of the Internet of Things. The frequency is 100KHz. According to the simulation environment and parameter settings mentioned
above, the algorithm proposed in the paper is applied to the user network quality detection. What is more, the use of the Internet of Things is adopted as an example to analyze the malicious data operation behavior in the monitoring management system to perform the detection and identification of the malicious data link in the Internet of Things cyber-physical system. First, analyze the sample distribution of the data link in the original Internet of Things physical system. The distribution is shown in Figure 1.

![Diagram showing sample distribution of the vicious data chain in the physical system of the Internet of Things](image)

Figure 1 Sample distribution of the vicious data chain in the physical system of the Internet of Things

Figure 1 shows that in the cyber-physical system of the Internet of Things, the vicious data chain is composed of vicious initial data and vicious termination data, which are clustered together. Taking the vicious data link as the research sample, in the cloud computing environment, the vicious data link of the cyber-physical system in the Internet of Things is quickly extracted, and the spectral stripe characteristics of the vicious data link in the cyber-physical system of the Internet of Things are extracted as well. The results are shown in Figure 2.

![Diagram showing spectral stripe characteristics of the vicious data link in the cyber-physical system of the Internet of Things](image)

(a) Vicious data chain extraction in the cyber-physical system of the Internet of Things
Vicious data chain extraction in the cyber-physical system of the Internet of Things

It can be seen from Figure 2 that the method proposed in the paper is used to quickly extract the vicious data chain of the Internet of Things cyber-physical system, and the extracted data chain has a high resolution ability, which improves the accurate detection and identification of the vicious data chain in the Internet of Things cyber-physical system. Meanwhile, the detection accuracy rate of the characteristic data in the malignant data link of the cyber-physical system within the Internet of Things is obtained. Additionally, the calculation formula for the accuracy rate of the detection and recognition of the malignant data link in the cyber-physical system of the Internet of Things is as follows:

$$G(s) = \frac{M_{TTD}}{(F_e^2 + 86.35F_e + 14230)} \quad (22)$$

In the formula, \(s\) represents the rate of change of deviation, and \(F_e\) refers to the membership function.

According to formula (22), the malignant data chain mining is carried out within the information size of 25MB, and different methods are used to compare the accuracy of the malignant data chain mining. The comparison results are shown in Figure 3.

Analyzing the results of Figure 3, it can be known that as the number of iterations increases, the accuracy of fast extraction continues to improve. Moreover, the detection accuracy of malicious data links in the cyber-physical system of the Internet of Things using the method proposed in the paper is 100%, which is 14.6% and 13.4% higher than that of the average accuracy of traditional methods. In addition, test the time cost of fast data extraction under different data scales, and the comparison results are shown in Table 1. Analyzing the results of Table 1, it shows that the method proposed in the paper has a shorter time cost for detecting malicious data links in the physical system of the Internet of Things.
Things, which improves the real-time performance of data monitoring and identification.

Table 1 Comparison of time expenditure (unit/s)

| Data size/Gbit | Method of this article | Spectrum analysis algorithm | PSO algorithm |
|---------------|------------------------|----------------------------|--------------|
| 20            | 0.453                  | 1.234                      | 0.797        |
| 40            | 0.756                  | 1.866                      | 0.986        |
| 60            | 0.865                  | 1.986                      | 1.123        |
| 80            | 0.894                  | 2.346                      | 1.678        |

According to Table 1, in the process of detecting and identifying malicious data links, the method proposed in the paper has less time overhead, which greatly optimizes the efficiency of detecting and identifying malicious data links.

5. Conclusion

Adopting the parallel mining method, the characteristics association rules of the vicious data chain in the cyber-physical system of the Internet of Things are mined to cluster the vicious data chain of the cyber-physical system in the Internet of Things. Meanwhile, the random allocation method of links is used to balance the output sensing sequence control of the vicious data link in the cyber-physical system of the Internet of things, and the vicious data link of the cyber-physical system in the Internet of things can be quickly extracted. Research has shown that the method proposed in the paper has high accuracy in detecting and identifying malignant data links in the cyber-physical system of the Internet of Things. The ability to extract malignant data links is better, while the time overhead is small.

References:

[1] Jia Yimin, Shi Liping, Yan Xin. Improved artificial fish swarm algorithm to optimize wavelet neural network for transformer fault diagnosis[J]. Journal of Henan University of Technology (Natural Science Edition), 2019, 38(02): 103-109.

[2] Li Yong, Chen Yu, Cai Ye,, et al. Reliability analysis of IEC61850 substation automation system based on information physical interface matrix[J]. Electric Power Automation Equipment, 2019, 39(01): 84-90+98.

[3] Cao Wenbin, Wang Xianpei, Tian Meng, et al. Fault diagnosis of SF_6 transformer based on KELM and AT[J]. Transformer, 2018, 55(12): 72-76.

[4] Zhang Xiaoyu, Wang Tianwei, Li Yan, et al. Research on Intelligent Comprehensive Optimization Control of Combustion System of Thermal Power Plant[J]. Proceedings of the Chinese Society of Electrical Engineering, 2019, 39(9): 2544-2552.

[5] He Guiyou, Wang Yonggang. Comprehensive application of high-density resistivity imaging and joint profile method in the detection of hidden karsts in transmission lines[J]. Journal of Engineering Geophysics, 2019, 16(01): 53-60.

[6] Xu Shuwei, Qiu Caiming, Zhang Dongxia, et al. Transmission line fault type identification based on deep learning[J]. Proceedings of the Chinese Society of Electrical Engineering, 2019, 39(01): 65-74+321.

[7] Wang Qian. Power grid industrial control system security monitoring technology based on abnormal network flow detection[J]. Network Security Technology and Application, 2017(10): 129+147.

[8] Zhang Xiaojian. Research on static voltage stability of power system with wind farms[J]. Electrical Engineering Technology, 2018(24):111-114.

[9] Xu Hai, Yang Shangjin, Jin Xuefen, et al. Power system Ward equivalence realization method based on power flow and short circuit calculation[J]. Power System Protection and Control, 2018, 46(24): 104-110.

[10] Zhang Minhao, Cong Haoxi, Shu Xiang, et al. Study on the micro-mechanism of the synergistic degradation of DBDS and hexadecyl mercaptan on transformer insulation paper[J]. Proceedings of the Chinese Society of Electrical Engineering, 2018, 38(24): 7156-7165+7442.