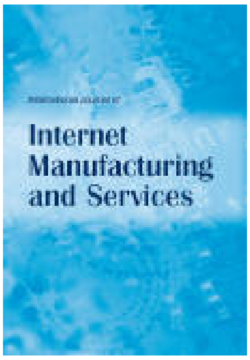


International Journal of Internet Manufacturing and Services

ISSN online: 1751-6056 - ISSN print: 1751-6048
https://www.inderscience.com/ijims

An abnormal node location method of industrial internet of things based on feature fuzzy clustering

Qiaoyun Chen, Chunmeng Lu

DOI: 10.1504/IJIMS.2024.10056157

Article History:

Received:	04 August 2022
Last revised:	18 January 2023
Accepted:	01 March 2023
Published online:	19 February 2024

An abnormal node location method of industrial internet of things based on feature fuzzy clustering

Qiaoyun Chen*

School of Information Engineering,
Jiaozuo University,
Jiaozuo 454000, China
Email: yatia518196@163.com
*Corresponding author

Chunmeng Lu

School of Artificial Intelligence,
Jiaozuo University,
Jiaozuo 454000, China
Email: lchmeng2021@163.com

Abstract: In order to overcome the problems of low positioning accuracy and long time-consuming of traditional anomaly node location methods, a new anomaly node location method of industrial internet of things based on feature fuzzy clustering is proposed in this paper. Firstly, collect the operation data of all nodes in the coverage area of the industrial internet of things. Secondly, according to the node operation data collection results, the standard deviation method is used to standardise the industrial internet of things node data, and extract the characteristics of abnormal nodes. Finally, the fuzzy clustering method is used to cluster the abnormal nodes, and the trilateral localisation method is used to locate the abnormal nodes according to the clustering results. The experimental results show that this method can reduce the positioning time on the premise of improving the positioning accuracy, and the average positioning accuracy reaches 99.11%.

Keywords: feature extraction; fuzzy clustering; industrial internet of things; abnormal node location.

Reference to this paper should be made as follows: Chen, Q. and Lu, C. (2024) ‘An abnormal node location method of industrial internet of things based on feature fuzzy clustering’, *Int. J. Internet Manufacturing and Services*, Vol. 10, No. 1, pp.15–26.

Biographical notes: Qiaoyun Chen received her MS in Communication and Information System from the Dalian University of Technology in 2009. She is currently a Lecturer at the Jiaozuo University. Her research interests include internet of things technology, communication technology and big data.

Chunmeng Lu received his MS in Information, Production and Systems Engineering from the Waseda University in Japan in 2018. He is currently a Teaching Assistant at the Jiaozuo University. His research interests include internet of things and virtual reality.

1 Introduction

With the development of the industrial internet of things, the number of devices in it is gradually increasing, which makes the management of the industrial internet of things more and more difficult (Liu et al., 2021). In particular, the fault alarm mode of the industrial internet of things is mainly threshold alarm. In order to improve the alarm accuracy, it is necessary to set the alarm threshold index more sensitive, so many false alarms will be generated, and the false alarms will greatly interfere with the positioning accuracy of abnormal nodes (Peng et al., 2021; Nayak et al., 2021; Fu et al., 2021). Although the operation and management personnel of the industrial internet of things will manually locate the abnormal nodes, it requires a lot of energy, and the delay of manual positioning is high, which reduces the operation safety and reliability of the industrial internet of things, resulting in the reduction of the service quality of the industrial internet of things (Ma et akl., 2021). Therefore, in order to improve the security of the industrial internet of things, it is necessary to accurately and quickly locate abnormal nodes.

Tang (2021) proposes an abnormal node location method based on background perception, which analyses the process and state of node motion, and calculates the anomaly degree of the node through the background perception method. According to the calculation results of anomaly degree, the lost state is inserted into the node state, and the motion trajectory of the node is re planned. Finally, the strategy of iterative calculation is used to detect abnormal nodes according to the sequence of node states. However, there is a certain gap between the abnormal node location results of this method and the actual abnormal nodes. Lin et al. (2020) proposes an abnormal node location method based on random matrix theory, which constructs the node operation matrix according to the characteristics of the original node data, and reduces the dimension of the node data. According to the results of dimensionality reduction, the anomaly of nodes is judged by the average spectral radius of nodes. Singular value method is used to locate abnormal nodes. However, the calculation process of this method is complex, resulting in a long time for the final positioning. Lu et al. (2020) proposes an abnormal node location method based on graph signal processing. In this method, sensors are used to build the nearest neighbour graph signal model of nodes, and the smoothness of nodes is counted according to the model. According to the statistical results, the decision threshold is set to determine whether the node is abnormal and complete the positioning of abnormal nodes. However, this method has the problem of insufficient accuracy in locating abnormal nodes.

In order to solve the problems of low positioning accuracy and long positioning time in the traditional positioning methods mentioned above, a feature fuzzy clustering-based anomaly node location method for industrial internet of things is proposed. The overall research scheme of this method is as follows:

1 Analyse the coverage area of the industrial internet of things and collect the operation data of all nodes in the internet of things.

2 Based on the collected industrial IoT node data, the standard deviation method is used to standardise the industrial IoT node data and extract the characteristics of abnormal nodes. According to the extracted abnormal node characteristics, the fuzzy clustering method is used to cluster the abnormal nodes, and the trilateral positioning method is used to locate the abnormal nodes according to the clustering results.

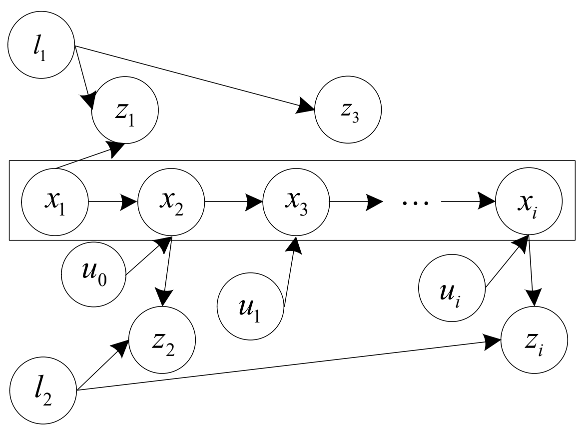
3 Experimental verification: Take the time consuming and positioning accuracy of abnormal nodes in the industrial internet of things as the experimental comparison indicators to carry out comparative verification of various methods.

2 Abnormal node location of industrial internet of things based on feature fuzzy clustering

2.1 Industrial IoT node operation data collection

Suppose the side length of the node distribution area in the industrial internet of things is M, the nodes in the area will form a cluster head $Sch(u)$ with N_f frames, and the time window between node i and link j is $t(n) = [(n - 1)T_w, nT_w]$. The node distribution in the industrial internet of things is shown in Figure 1.

Figure 1 Node distribution of industrial internet of things



In order to locate abnormal nodes in the industrial internet of things, it is necessary to accurately collect a large number of operation data of internet of things nodes, and extract node characteristics through data analysis, so as to determine the location of abnormal nodes. Therefore, first of all, we need to collect the operation data of industrial IoT nodes.

According to the initial position of the industrial IoT node, the optimal networking processing is carried out for the industrial IoT, so the node position control function as shown in formula (1) is constructed:

$$V_{res}(r) = \frac{\sum_{i=1}^{N} \left\{ E'_{res}\left(n_i - E_a^r\right) \right\}^2}{N} \tag{1}$$

In the formula, E'_{res} represents the survival probability of industrial IoT nodes, n_i represents nodes, r represents the transmission link layer of industrial IoT, and E_a^r represents the residual transmission energy of industrial IoT nodes (Lu et al., 2020).

With the support of node position control function, the distributed energy characteristics of industrial IoT nodes are calculated:

$$E_R = \sum_{r=1}^{L_i} \sum_{n_j \in S_i^r} E_{Rx}(l) \tag{2}$$

In the formula, L_i represents the node correlation of the industrial internet of things, S_i^r represents the area covered by the node, and $E_{Rx}(l)$ represents the energy function of the node.

Through the optimisation of the node location of the industrial internet of things, the node energy consumption in a single cluster is calculated:

$$E_T = \sum_{r=1}^{L_i} \sum_{n_g \in N_i^r} E_{Tx}\left(l, d_{(n_i, n_g)}\right) \tag{3}$$

In the formula, N_i^r represents the sparsity coefficient of node data (Zhang, 2021).

According to the calculation results of node energy consumption, the industrial internet of things nodes is thinned. The transformation basis of thinning is $\psi(N \times N)$, and the thinned nodes can be expressed as $x = s\psi$, and the optimal solution of node energy scheduling can be calculated:

$$E_F = \sum_{r=1}^{L_i} l_r E_{DF} \tag{4}$$

In the formula, E_{DF} represents the energy consumed by node scheduling (Venkatesan and Prabhavathy, 2019).

Combined with the above calculation results, build the energy cost of node operation data collection:

$$E_{int} = E_R + E_T + E_F \tag{5}$$

The coverage area of industrial IoT nodes should meet the following conditions:

$$E_0 + \cdots + E_N - L_1 p_1 - \cdots - L_{N+1} p_{N+1} = 0 \tag{6}$$

In the formula, E_0 represents the energy of the initial node in the industrial internet of things. According to the coverage area of the node, build the operation data collection formula of the industrial internet of things node as shown in formula (7):

$$F = f\left(p_{N+1}\right) + \cdots + f\left(p_1\right) - m_1 \log\left(E_0 - L_1 p_1\right) - \cdots m_N \log\left(E_{N-1} - L_N p_N\right) = 0 \tag{7}$$

Through the above calculation, the operation data of all nodes in the industrial internet of things can be collected.

2.2 Abnormal node feature extraction of industrial internet of things

For the active industrial internet of things, to locate abnormal nodes, it is necessary to extract the characteristics of abnormal nodes, so as to avoid positioning errors.

For m nodes in the industrial internet of things, the dimension of the characteristic vector of the node is n_i, and the characteristic space of the node in the industrial internet of things space Ω is A_i. Therefore, the characteristic vector of the abnormal node of the industrial internet of things can be expressed as:

$$a_i = \delta \in \Omega\left[a_{i,1}, a_{i,2}, \ldots, a_{u,n_i}\right] \in A_i \tag{8}$$

The standard deviation method is used to standardise the processing of industrial IoT node data (Venkatesan and Prabhavathy, 2019; George et al., 2019; Meng et al., 2021). Because the positioning of abnormal nodes is based on the abnormal node operation data, the standard deviation method can more accurately describe the change characteristics of node operation data, so as to improve the accuracy of node feature extraction.

The combination standardisation processing formula of the eigenvector between the i^{th} node and the j^{th} node is:

$$a_{i,j}^{*} = \frac{a_{i.j} - \overline{a}_{i}^{j}}{\sqrt{a_{i,j} - \left(a_{i}^{j}\right)}} \tag{9}$$

After the standardisation of node feature data, extract the one-dimensional feature vector of abnormal nodes:

$$T1 = \left[a_1, a_2, \ldots, a_m\right] \times \delta \in \Omega + \left[a_{1,1}, \ldots, a_{1n1}, a_{2,1}, \ldots, a_{2,1}, \ldots, a_{2,n2}, \ldots, a_{m,1}\right] \tag{10}$$

In the formula, T represents the combined characteristic two-dimensional matrix of nodes (Mohammed et al., 2020).

According to the different operating states of industrial IoT nodes, the two-dimensional eigenvector matrix of abnormal nodes is constructed:

$$T2 = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,n} \\ a_{m,1} & a_{m,2} & a_{m,n} \end{bmatrix} \tag{11}$$

Through the above calculation, the abnormal node characteristics of the industrial internet of things are accurately extracted to provide effective support for the location of abnormal nodes.

2.3 Abnormal node location of industrial internet of things based on fuzzy clustering

According to the above extracted abnormal node characteristics of the industrial internet of things, the fuzzy clustering method are used to locate the abnormal nodes. Fuzzy clustering analysis is an analysis method of clustering objective things by establishing fuzzy similarity relationship according to the characteristics, closeness and similarity between objective things. The result of fuzzy clustering makes each data may eventually belong to multiple clusters. Each data assigns a membership degree to each cluster, and the result of clustering can be expressed as a fuzzy matrix. Therefore, we can get more intuitive clustering results and improve the accuracy of abnormal node location in the industrial internet of things.

The distance of the abnormal node is set as $\{d_1, d_2, \ldots, d_n\}$, and the distance data is divided according to a group of three distance information, and then the trilateral positioning algorithm is used to obtain the node positioning results (Yu et al., 2021). Firstly, the initial sample nodes are clustered by fuzzy clustering algorithm to screen the distance information of nodes.

The sample data of the industrial internet of things node is $X = \{(x_i, y_i) \mid i = 1, 2, \dots, M\}$, and the membership matrix of the sample data is $U^{(0)} = (u_{ik}^{(0)})$, then the fuzzy clustering centre of abnormal nodes is constructed:

$$v_i^{(l)} = \frac{\sum_{k=1}^{n} u_{ik}^{(l-1)^m} x_k}{\sum_{k=1}^{n} u_{ik}^{(l-1)^m}}, i = 1, 2, \dots, c \tag{12}$$

In the formula, l represents the number of iterations (Pan et al., 2019).

Correct the membership of the sample data:

$$u_{ik}^{(l)} = \frac{1}{\sum_{j}^{c} = \left(\frac{d_{ik}^{(l)}}{d_{jk}^{(l)}} \right)^{\frac{2}{m-1}}}, i = 1, 2, \dots, n \tag{13}$$

According to the modified membership degree, the objective function of clustering is constructed:

$$J^{(l)}\left(U^{(l)}, V^{(l)}\right) = \sum_{k=1}^{m} \sum_{i=1}^{c} \left(u_{ik}^{(l)}\right)^m \left(d_{ik}^{(l)}\right)^2 \tag{14}$$

In the formula, $d_{ik}^{(l)} = \left\| x_k - v_i^{(l)} \right\|$.

The termination tolerance of membership degree is defined as $\varepsilon_k > 0$, and the maximum number of iterations is $L_{\max}$. If $\max\{| u_{ik}^{(l)} - u_{ik}^{(l-1)} |\} < \varepsilon_k$ exists, the clustering iteration is stopped, otherwise the iterative calculation needs to be repeated until the above conditions are met.

After the above iterative clustering, the membership matrix and clustering centre of abnormal nodes in the internet of things can be obtained, and when the objective function reaches the minimum value, the attribution of different abnormal nodes can be determined. When there is $u_{jk} = \max_{l \le i \le c}\{u_{ik}\}$, sample x_k can be classified as class c.

According to the clustering results of nodes, the abnormal nodes of the industrial internet of things are located by using the trilateral location algorithm. If the beacon node is set as (x_i, y_i), $i = 1, 2, \dots, M$ and the coordinates of the unknown abnormal node are (x, y), the following equations can be obtained:

$$(x - x_i)^2 (y - y_i)^2 = d_i^2, i = 1, 2, \dots, M \tag{15}$$

$$\frac{1}{M} \sum x - x_i = -\overline{x}_i, \frac{1}{M} \sum y - y_i = \overline{y}_i \tag{16}$$

$$\frac{1}{M} \sum x^2 - x_i^2 = \tilde{x}_i^2, \frac{1}{M} \sum y^2 - y_i^2 = \tilde{y}_i^2, d_i^2 - \frac{1}{M} \sum d_i^2 = \tilde{d}_i^2 \tag{17}$$

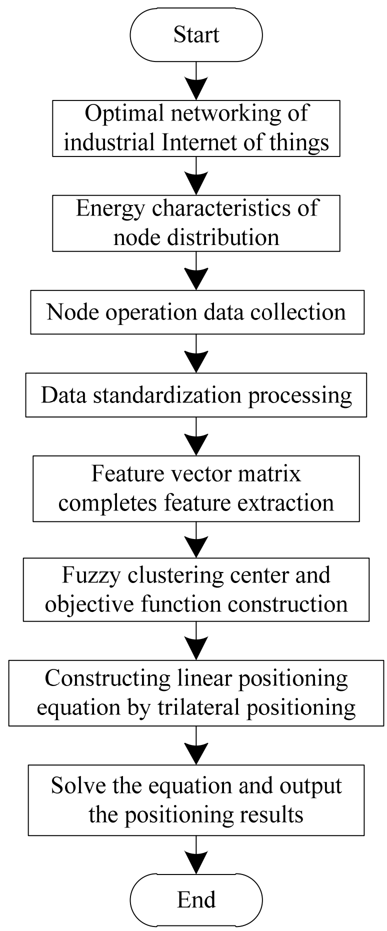
According to the trilateral positioning principle, when the number of beacon nodes is greater than 3, the above equations can be converted into linear equations as follows:

$$AX = b,\ A = \begin{bmatrix} \bar{x}_1 & \bar{y}_1 \\ \bar{x}_M & \bar{y}_M \end{bmatrix},\ X = \begin{bmatrix} x \\ y \end{bmatrix},\ b = \begin{bmatrix} \frac{1}{2}\left(\tilde{d}_i^2 + \tilde{x}_i^2 + \tilde{y}_i^2\right) \\ \frac{1}{2}\left(\tilde{d}_M^2 + \tilde{x}_M^2 + \tilde{y}_M^2\right) \end{bmatrix} \tag{18}$$

By solving the linear equation shown in formula (18), the location results of abnormal nodes in the industrial internet of things can be obtained, and the location of abnormal nodes can be completed.

The abnormal data location process of industrial internet of things based on feature fuzzy clustering is shown in Figure 2.

Figure 2 Abnormal node location process



3 Experimental verification

The above process completes the research on the abnormal node location of the industrial internet of things from the theoretical part. In order to verify the practical application performance of the proposed location method, comparative test experiments are carried out.

3.1 Experimental data

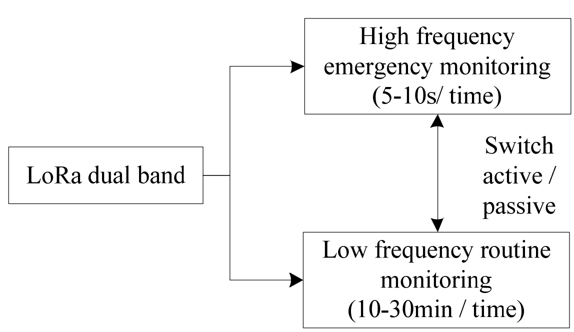
The data of this experiment is collected from an industrial internet of things. LoRa dual frequency data acquisition system is used to collect the operation data of nodes in the industrial internet of things. Industrial IoT parameters are shown in Table 1.

Table 1 Industrial IoT parameters

Parameter	*Parameter value*
Internet of things area	1,000 $*$ 1,000 m
Sink node coordinates	0.5 $*$ 100, 0.5 $*$ 100
Number of nodes	100
Number of sink nodes	1
Number of malicious nodes	1
Probability of cluster head node selection	0.1
Maximum number of rounds	100
Node communication radius	20 m

LoRa dual frequency data acquisition system includes low-speed acquisition frequency band and high-speed acquisition frequency band, which are respectively applicable to data acquisition of different frequency bands, and can meet the needs of data acquisition of this node operation. Collect the operation data of 1,000 nodes and take it as sample data. The schematic diagram of operation data collection of IoT nodes is shown in Figure 3.

Figure 3 Industrial internet of things node data collection



3.2 Experimental scheme and index

In order to fully verify the abnormal node location performance of this method, taking the abnormal node location accuracy and abnormal node location time as experimental comparison indicators, this method is compared with Tang (2021) and Lin et al. (2020) methods:

- Positioning accuracy of abnormal nodes: The positioning accuracy of abnormal nodes refers to the consistency between the positioning results of abnormal nodes of different methods and the actual abnormal nodes. The higher the positioning accuracy of abnormal nodes, the stronger the positioning performance of the method.
- Abnormal node location time: Abnormal node location time refers to the time spent by different methods to locate the abnormal node. The shorter the location time is, the higher the location efficiency of the method is.

3.3 Analysis of experimental results

3.3.1 Positioning accuracy of abnormal nodes

As the fundamental performance of the positioning method, the positioning accuracy of abnormal nodes can directly reflect the positioning performance of different methods. Therefore, the positioning accuracy of abnormal nodes is selected as the method performance test index. In order to improve the reliability of the experimental results, Tang (2021) and Lin et al. (2020) methods are selected for comparison and verification. The comparison results of abnormal node positioning accuracy of the three methods are shown in Table 2.

Table 2 Positioning accuracy of abnormal nodes

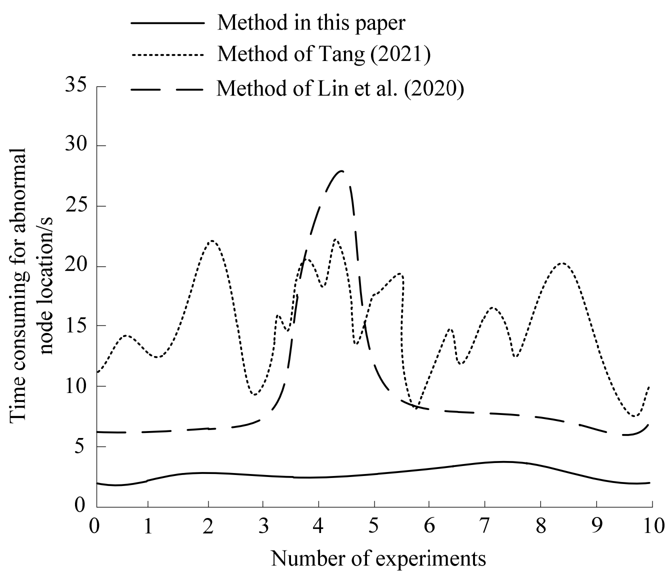
Number of experiments	*Positioning accuracy of abnormal nodes / %*		
	Method in this paper	*Method of Tang (2021)*	*Method of Lin et al. (2020)*
1	99.9	79.8	78.5
2	99.6	79.4	85.5
3	99.0	78.8	88.6
4	99.8	79.7	81.9
5	99.7	79.5	77.8
6	99.5	79.3	70.6
7	98.7	86.7	76.7
8	96.7	81.4	81.8
9	98.6	72.3	74.8
10	99.6	82.6	79.3
Mean value	99.11	79.95	79.55

From the comparison results of abnormal node positioning accuracy shown in Table 2, it can be seen that during many tests, the positioning accuracy of Tang (2021) and Lin et al. (2020) methods is always lower than that of this method. The average value of abnormal node positioning of this method is 99.11%, the average value of abnormal node positioning of Tang (2021) method is 79.95%, and the average value of abnormal node positioning accuracy of Lin et al. (2020) method is 79.55%. Therefore, this method can effectively improve the effectiveness of abnormal node location in the industrial internet of things.

3.3.2 Time consuming for abnormal node location

As the industrial internet of things involves the transmission of a number of industrial data, once there is an abnormal node in the industrial internet of things for a long time, it will seriously affect the security of data transmission in the industrial internet of things. Therefore, higher requirements are put forward for the time-consuming method of abnormal node location, so it is necessary to verify the time-consuming method of abnormal node location in the industrial internet of things. The time-consuming results of abnormal node location of different methods are shown in Figure 4.

Figure 4 Time consuming for abnormal node location



Observing the time-consuming comparison of abnormal node location shown in Figure 4, it can be seen that among the three location methods, the time-consuming of abnormal node location in this method is the shortest, no more than 5 s; the time-consuming of locating abnormal nodes in Tang (2021) method shows a fluctuating trend, with the highest value of about 23 s; the time-consuming of locating abnormal nodes in the method of Lin et al. (2020) shows a sudden rise and then decline, with the highest value of about 28 s. Compared with the two literature comparison methods, this method can effectively reduce the time-consuming of abnormal node location and ensure the security of industrial IoT data.

4 Conclusions

In order to improve the security of the operation of industrial internet of things, an abnormal node location method of industrial internet of things based on feature fuzzy clustering is proposed, and the performance of the method is verified from both theoretical and experimental aspects. This method has higher accuracy and shorter time-consuming when locating abnormal nodes in the industrial internet of things. Specifically, compared with the abnormal node location method based on background perception, the abnormal node location accuracy of this method is significantly improved, and the average location accuracy reaches 99.11%; compared with the abnormal node location method based on random matrix theory, the abnormal node location time of this method is significantly shortened, and the maximum time is no more than 5 s. Therefore, it fully shows that the proposed location method based on feature fuzzy clustering can better meet the requirements of abnormal node location in the industrial internet of things.

References

Fu, L., Zhang, W., Tan, X. and Zhu, H. (2021) 'An algorithm for detection of traffic attribute exceptions based on cluster algorithm in industrial internet of things', *IEEE Access*, Vol. 9, No. 2, pp.53370–53378.

George, M., Kokare, P., Jose, B. and Mathew, J. (2019) 'Autoencoder-based abnormal activity detection using parallelepiped spatio-temporal region', *IET Computer Vision*, Vol. 13, No. 1, pp.23–30.

Lin, C., Zheng, L., Zhang, W. and Deng, X. (2020) 'Normal node location algorithm for WSN based on random matrix theory', *Computer Engineering*, Vol. 46, No. 1, pp.157–163.

Liu, Y., Zhao, R., Kang, J., Yassine, A. and Peng, J. (2021) 'Towards communication-efficient and attack-resistant federated edge learning for industrial internet of things', *ACM Transactions on Internet Technology*, Vol. 22, No. 3, pp.1–22.

Lu, G., Zhou, L., Lv, S., Shi, C. and Su, K. (2020) 'Outlier node detection algorithm in wireless sensor networks based on graph signal processing', *Journal of Computer Applications*, Vol. 40, No. 3, pp.783–787.

Ma, L., Wang, X., Wang, X., Wang, L. and Huang, M. (2021) 'TCDA: truthful combinatorial double auctions for mobile edge computing in industrial internet of things', *IEEE Transactions on Mobile Computing*, Vol. 41, No. 19, pp.141–149.

Meng, L., Wei, Y., Pan, R., Zhou, S., Zhang, J. and Chen, W. (2021) 'VADAF: visualization for abnormal client detection and analysis in federated learning', *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 11, No. 3, pp.1–23.

Mohammed, A.H., Ibrahim, A.A. and Aysa, M.H. (2020) 'IoT Ddos attack detection using machine learning', *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*.

Nayak, S., Ahmed, N. and Misra, S. (2021) 'Deep-learning-based reliable routing attack detection mechanism for industrial internet of things', *Ad Hoc Networks*, Vol. 123, No. 45, pp.156–161.

Pan, Z., Zhao, M., Luo, Z. and Yang, H. (2019) 'Detection and simulation of WSN wormhole attack node based on RSSI localization', *Computer Simulation*, Vol. 36, No. 1, pp.387–391.

Peng, H., Qian, Z., Kan, Z., Zhao, D., Yu, J. and Han, J. (2021) 'Cascading failure dynamics against intentional attack for interdependent industrial internet of things', *Complexity*, Vol. 33, No. 14, pp.1–15.

Tang, W. (2021) 'Location detection algorithm for abnormal nodes based on background perception', *Control Engineering of China*, Vol. 28, No. 8, pp.1621–1627.

Venkatesan, M. and Prabhavathy, P. (2019) 'Graph based unsupervised learning methods for edge and node anomaly detection in social network', *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, IEEE.

Yu, J., Lee, Y., Yow, K.C., Jeon, M. and Pedrycz, W. (2021) 'Abnormal event detection and localization via adversarial event prediction', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 67, No. 9, pp.1–15.

Zhang, J. (2021) 'Adaptive and lightweight abnormal node detection via biological immune game in mobile multimedia networks', *Algorithms*, Vol. 14, No. 12, pp.368–375.