IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON EMERGING APPROACHES TO CYBER SECURITY

Received June 29, 2020, accepted July 24, 2020, date of publication August 3, 2020, date of current version August 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013840

Malicious Node Identification Strategy With Environmental Parameters

ZHIJUN TENG[1,2], BAOHE PANG[2], CHUNQIU DU[2], AND ZHE LI[2]

[1]Key Laboratory of Modern Power System Simulation and Control and Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132000, China
[2]School of Electrical Engineering, Northeast Electric Power University, Jilin 132000, China

Corresponding author: Baohe Pang (1187571167@qq.com)

This work was supported in part by the National Natural Science Foundation Youth Science Foundation Project under Grant 61501107, and in part by the 13th Five-Year Scientific Research Planning Project of Jilin Province Department of Education under Grant JJKH20180439KJ.

ABSTRACT Wireless sensor network (WSN) works in a complex environment where it is difficult for people to reach or work. The openness of nodes leads to security threats vulnerable to various attacks. The trust and reputation model can be applied in WSN to reduce damage caused by malicious nodes. However, there is a high false-positive rate in trust and reputation models because a node with less reputation due to the communication environment is judged as a malicious one directly. This paper presents a trust & reputation-based malicious node identification strategy with environmental parameters (TRS&EP) to interdict the malicious nodes, such as interrupt attack nodes and selective forwarding attack nodes. Using the linear regression of machine learning and combining the energy of nodes, data volume, number of adjacent nodes, the node sparsity and other deterministic parameters can solve environmental parameters. Then TRS&EP estimates benchmark trust according to the environmental parameters. The Gaussian radial basis function is simplified to calculate the similarity between the benchmark trust sequence and cycle reputation sequence. Furthermore, TRS&EP sets three reputation intervals and an adoptive threshold span to identify the malicious nodes by dynamically considering the work environment and states of nodes. The simulated results show that TRS&EP improves the recognition of malicious nodes above 1% compared to comparison algorithms and reduces the false-positive percentage by more than 1%.

INDEX TERMS Wireless sensor network, interrupt attacks, selective forwarding attacks, reputation model, environmental parameter.

I. INTRODUCTION

WSN is deployable in an extensive assortment of applications, such as the military, industrial, medical, commercial, and other fields [1]. With the continuous development of WSN, users have higher requirements for the network's security performance [2]. Sensor nodes are vulnerable to various attacks because of wireless and distributed nature [3]. The nodes captured by the enemy will be decrypted and tampered with the program of the network. When these tampered nodes return to the WSN, they become malicious nodes to launch almost any attack on the network [4]. Security mechanisms based on cryptography hardly defense malicious nodes entering the network so-called insider attacks bypassing the authentication by their neighbors. They can delete packets, interrupt or selectively forward information, or publish false information, even disguised as normal nodes, to evade the network's intrusion detection system (IDS). In this situation, an effective security model should be established to cope with these attacks in the WSN [5], [6].

In recent years, there is a growing body of the paper that recognizes the importance of the defense of insider attacks in WSN. Previous research has established some schemes to identify and avoid insider intelligence attacks based on more directions, such as intrusion detection, identity identification, and data fusion. For instance, a solution based on vulnerability-aware heterogeneous network devices assignment (VHNDA) for malicious packets attacks was proposed in [7]. Considering the limitation of WSN capability and the openness of the communication environment, a security disjoint routing-based verified message scheme (SDRVM) was designed to resist communication blocking attacks in [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Luis Javier Garcia Villalba.

The trust and reputation can effectively evaluate the historical communication behavior of nodes and recognize the malicious nodes [4]. However, there are high misjudgment rate and low recognition rate in traditional trust and reputation security models. The following reasons can also be found: 1) Some models calculate direct credit values only by evaluating the interaction information. This method is insecure for various types of internal attacks. There are many factors influencing node forwarding behavior. Therefore, the trust assessment should consider more status information of sensor nodes in the real working environment. 2) The communication quality in the real WSN environment is not ideal. Poor-quality links can damage the interaction among nodes and adversely affect the trust value of normal nodes. Therefore, it is necessary to distinguish normal nodes in an inferior environment from sub-attack nodes such as selective forwarding attacks. 3) The process of identifying malicious nodes by setting a single reputation threshold can sort out abnormal nodes with low accuracy roughly. Therefore, an adaptive threshold change with node states to identify the malicious node is essential in the real deployment of WSN.

To address the above concerns and limitations, we propose a malicious node identification strategy based on trust, reputation sequence, and environmental parameters, TRS&EP. We resort to trust, reputation, and node state to single out malicious nodes and reduce the false-positive rate due to poor-states of nodes, respectively while considering the WSN deployment environment. Besides, through a dynamic double threshold interval, TRS&EP can effectively identify sub-attack nodes and cope with the problem of low recognition rate caused by a single threshold in the previous trust and reputation models.

The main contributions of our work are summarized as follow:

- An environmental parameter is calculated with reputation matrix and four-state information of nodes, energy consumption, data volume, number of neighbors, and sparsity of nodes, by the inspiration for machine learning, to predict the trust value in the next period.
- A simplified Gaussian radial basis function is used to calculate the similarity of the actual and predicted trust value matrix.
- Three reputation intervals are designed, and the adaptive threshold of tolerable trust span is set according to the change of dynamic state information of nodes.

The rest of this paper is organized as follows: In Section II, includes related work related to trust and reputation model for WSN. In Section III, the proposed TRS&EP model is introduced. In Section IV, the process of malicious nodes identification is described. In Section V, the performance of the TRS&EP is evaluated. Finally, the conclusion is made in Section VI.

II. RELATED WORK

Trust and reputation model for WSN is regarded as an effective supplement mechanism of cryptography and can further protect against insider attacks launched by malicious nodes. Trust is the anticipation of nodes about their future behavior, while reputation is mainly accumulated by past performance. Therefore, reputation would be one of the parameters to determine trust [9], [10]. There are some classical methodologies using trust and reputation, including the Bayesian trust model, the subjective logic trust model, the entropy trust model, the fuzzy trust model, the game theory trust model, and the biological approach. Among them, Bayesian theory is widely used in WSN's security and IDS recently. Ganeriwal *et al*. [11] proposed a typical reputation framework, RFSN, to identify the misbehaving nodes for various fault scenarios. They adopted a Bayesian formulation, specifically a Beta function, for the algorithm steps of reputation representation and trust update. The Beta reputation-system for sensor networks (BRSN) was proposed based on the Beta distribution and Bayesian formulation. The feasibility of Beta distribution was verified by derivation, and a detailed explanation of calculations for the reputation of the update, aging, and trust was elaborated. BRSN, as a trust model for WSN, has been widely studied and used. However, BRSN sets a single threshold to distinguish between normal and malicious nodes. It is difficult to distinguish between normal nodes and sub attack nodes whose reputation value and behavior are similar to normal nodes. In [12], a reputation model based on BRSN to detect sub attacks was proposed. K-mean is used to aggregate the sub-attack nodes again from the abnormal nodes. Zhang *et al*. [13] proposed a mechanism based on state context and hierarchical trust. In the hierarchical wireless sensor nodes work in an unmanageable environment with variable node state, they evaluated trust from three factors for cluster heads and common nodes respectively and finally detected intrusion according to a comprehensive assessment. Khan *et al*. [14] proposed a trust estimation approach (LTS) for large-scale WSN to detect malicious (faulty or selfish) nodes. In LTS, punishment and trust severity can be tuned according to the application requirement. LTS has excellent decision-making capability. Fang *et al*. [15] proposed BTRES based on monitoring nodes' behavior and beta distribution. BTRES mainly focuses on communication trust, data trust, and energy trust. Sahoo *et al.* [16] introduced the method of error behavior cycle factor and trust value repair to distinguish the selective forwarding attack nodes from temporary fault nodes. In [17], BLTM was proposed based on trust value to beta distribution, which took the effect of link quality on the trust model. Besides, they discussed the weight of communication trust, energy trust, and data trust. In [18]–[20], they improved trust mechanisms to identify bad-mouthing attacks, false-praise attacks, and collusion attacks.

More new ideas and technologies have been introduced into the trust and reputation model to ensure network security. Tariq *et al*. [21] proposed a mobile code-driven trust mechanism (MCTM) for addressing internal attacks by assessing trust value based on nodes' forwarding behaviors. A software-defined network solves the hard problem of data aggregation in WSN. They focused on the message

overhead and energy efficiency in the trust assessment but fail to verify or analyze the detection effects and security performance of MCTM, such as recognition rate, misjudgment rate, or detection efficiency. Under the idea of machine learning, Ren *et al.* [22] established a trust mechanism to evaluate the trust of the data reporter. Fu *et al.* [23] proposed a data clustering algorithm that can detect and isolate malicious cluster heads that have launched selective forwarding attacks by clustering their cumulative forwarding rates.

III. PROPOSED TRS&EP MODEL FOR WSN

In WSN, nodes often carry different sensors deployed in complex environments to sensory and collect data [24]. It is necessary to consider the reputation behavior of each node dynamically for improving the accuracy in the security system. In this section, we give a detailed description of TRS&EP on a clustered WSN, which selects parameters such as node reputation value, node energy consumption, data volume, number of adjacent nodes, and distribution sparsity of nodes to build the environment parameter.

A. NODE REPUTATION MODEL

This work adopts an improved Bayesian reputation evaluation model based on time series information analysis, TS-BRS [25]. In the clustered hierarchical WSN, the model can reduce the trust of interrupting attack nodes quickly and optimize the impact of channel occupation on communication behavior. The calculation formula of its reputation value is as follows:

$$R^N = \frac{\mu\alpha + \Delta\alpha}{\mu(\alpha+\beta) + \Delta\alpha + \Delta\beta} \tag{1}$$

$$\mu = \frac{\theta}{\alpha+\beta} \tag{2}$$

where R^N is the reputation value of the node N. Suppose there are $\alpha+\beta$ times interactions between the node N and neighbors. α and β denote the number of normal communications and abnormal communications in history, respectively. $\Delta\alpha$ and $\Delta\beta$ denote the number of normal communications and abnormal communications during a period Δt, respectively. μ is a reputation maintenance function that ensures the influence of current node behavior on reputation value and reduces the influence of historical behavior. θ is a fixed maintenance parameter used to set the scope of the maintenance function, refer to [25], $\theta = 150$.

B. MATRIX OF REPUTATION

The reputation of node N is periodically updated according to (1) during per period Δt, which is written as $R^N_{i(j)}$. After n periods, we can record a reputation sequence of node N as follows:

$$\{\underbrace{R^N_{i(1)}, R^N_{i(2)}, \cdots, R^N_{i(n)}}_{n}, \underbrace{R^N_{(i+1)(1)}, \cdots, R^N_{(i+1)(n)}}_{n}\} \tag{3}$$

The ith group of reputation time series of node N is defined as $T^N_i = \{R^N_{i(1)}, R^N_{i(2)}, \ldots, R^N_{i(n)}\}$. And next group is $T^N_{i+1} = \{R^N_{(i+1)(1)}, R^N_{(i+1)(2)}, \ldots, R^N_{(i+1)(n)}\}$. So we defined the matrix of reputation T^N_i in the ith period as follows:

$$T^N_i = [R^N_{i(1)} R^N_{i(2)} \cdots R^N_{i(n)}]^T \tag{4}$$

C. MATRIX OF STATE

The nodes have different working states at workplace. In this paper, the states of the nodes are considered from four dimensions, the energy consumption $E^N_{res} = \{e^N_{i(1)}, e^N_{i(2)}, \cdots e^N_{i(n)}\}$, data volume $W^N_{job} = \{w^N_{i(1)}, w^N_{i(2)}, \cdots w^N_{i(n)}\}$, number of neighbor $M^N_{nei} = \{m^N_{i(1)}, m^N_{i(2)}, \cdots m^N_{i(n)}\}$, the sparsity of node $D^N_{den} = \{d^N_{i(1)}, d^N_{i(2)}, \cdots d^N_{i(n)}\}$.

When node N updates reputation value, the current states of the node is recorded. Then the matrix of the state is constructed as follows:

$$S^N_i = \begin{bmatrix} s^N_{i(1)} \\ s^N_{i(2)} \\ \vdots \\ s^N_{i(n)} \end{bmatrix} = \begin{bmatrix} e^N_{i(1)} & w^N_{i(1)} & m^N_{i(1)} & d^N_{i(1)} \\ e^N_{i(2)} & w^N_{i(2)} & m^N_{i(2)} & d^N_{i(2)} \\ \vdots & \vdots & \vdots & \vdots \\ e^N_{i(n)} & w^N_{i(n)} & m^N_{i(n)} & d^N_{i(n)} \end{bmatrix} \tag{5}$$

1) ENERGY CONSUMPTION

The node behaves selfishly, giving up cooperation, or receiving only without forwarding, which is similar to selective attack when its energy resource is insufficient [26]. Therefore, the residual energy of nodes should be considered when evaluating the reputation value of nodes. Residual energy $E^N_{t+\Delta t}$ is the preset energy minus the energy dissipation in the process of sending, receiving, and processing data, which is calculated by LEACH [27]. E^N_t represents the residual energy of node N before period time Δt. The energy consumption of node N during a period time Δt is calculated by the difference between $EN\,t + \Delta t$ and $EN\,t$ as follows:

$$e^N_{i(j)} = E^N_{t+\Delta t} - E^N_t \tag{6}$$

2) DATA VOLUME

Data volume is usually represented in two ways [28]. The first method is to count the size of packets sent and received by the node in period Δt. And the second point is to record the number of packets during a period time Δt. In this paper, node N records data volume $W^N_{job} = \{w^N_{i(1)}, w^N_{i(2)}, \cdots w^N_{i(n)}\}$ in a second way, which can save computing overhead because of the same size of packet set in the simulation experiment.

3) NUMBER OF NEIGHBORS AND SPARSITY OF NODES

$m^N_{i(j)}$ is the number of neighbors within the communication radius at jth Δt of ith group, and M^N_{nei} represents the array of $mN\,i(j)$. The sparsity of node $D^N_{den} = \{d^N_{i(1)}, d^N_{i(2)}, \cdots d^N_{i(n)}\}$ is a degree of distribution in this region. As Fig.1., the density of node distribution varies greatly in the situation with the same communication radius R_c and $m^N_{i(j)}$. The node distribution with different density takes some influence into the transmission.

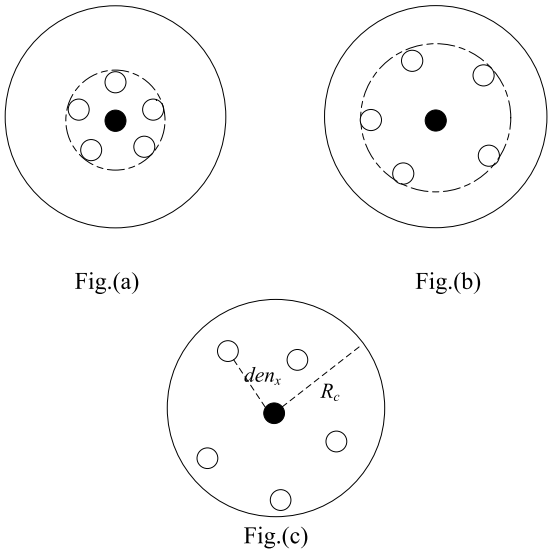


FIGURE 1. Distribution of neighbors.

We calculate nodes' sparsity by the distance between the target node and its neighbors.

$$d_{i(j)}^{N} = \frac{\sum_{x=1}^{m_{i(j)}^{N}} den_x}{m_{i(j)}^{N} R_c} \tag{7}$$

den_x is defined as the distance between each neighbor node and the target node. When the communication radius R_c of nodes is fixed, the denser the node distribution is, the smaller $d_{i(j)}^{N}$ is.

D. MATRIX OF ENVIRONMENTAL PARAMETERS

We define the relationship with a matrix of environmental parameters Q_i^N, matrix of state S_i^N, and matrix of reputation T_i^N. The S_i^N is multiplied by Q_i^N to get T_i^N, as follows:

$$S_i^N Q_i^N = T_i^N \tag{8}$$

The correlation formula of time series of node N in the i group can be expressed as:

$$\begin{bmatrix} e_{i(1)}^N & w_{i(1)}^N & m_{i(1)}^N & d_{i(1)}^N \\ e_{i(2)}^N & w_{i(2)}^N & m_{i(2)}^N & d_{i(2)}^N \\ \vdots & \vdots & \vdots & \vdots \\ e_{i(n)}^N & w_{i(n)}^N & m_{i(n)}^N & d_{i(n)}^N \end{bmatrix} \cdot \begin{bmatrix} q_E^N \\ q_W^N \\ q_M^N \\ q_D^N \end{bmatrix} = \begin{bmatrix} R_{i(1)}^N \\ R_{i(2)}^N \\ \vdots \\ R_{i(n)}^N \end{bmatrix} \tag{9}$$

During an observation period, we can collect and calculate the data in the matrix of node state and reputation, and the emphasis is to apply these actual data to calculate the matrix of environment parameters Q_i^N. Linear regression in machine learning is used to calculate Q_i^N. The mean square error formula MSE of the matrix is as follows:

$$MSE = \frac{1}{n} \sum_{j=1}^{n} (s_{i(j)}^N \cdot Q_i^N - T_i^N)^2 \tag{10}$$

TABLE 1. Main notations adopted in this paper.

Symbol	Description
α	Number of normal communications
$\Delta\alpha$	Number of normal communications during a period Δt
β	Number of abnormal communications
$\Delta\beta$	Number of abnormal communications during a period Δt
μ	Value of the reputation maintenance function
θ	Fixed maintenance parameter
n	Number of Δt in an observation cycle (the length of a time series)
$R_{i(j)}^N$	Reputation value of node N at the jth Δt of the ith period
T_i^N	Matrix of reputation in the ith period
E_{res}^N	Set of energy consumption
$e_{i(j)}^N$	Energy consumption of node N at the jth Δt of the ith period
W_{job}^N	Set of data volume
$w_{i(j)}^N$	number of packets transmitted and received by node N at the jth Δt of the ith period
M_{nei}^N	Set of neighbor nodes
$m_{i(j)}^N$	Number of neighbors within the communication radius at the jth Δt of the ith period
D_{den}^N	Set of node sparsity
$d_{i(j)}^N$	Sparsity of node N at the jth Δt of the ith period
den_x	Distance between each neighbor
R_c	Communication radius of nodes
S_i^N	Matrix of state in the ith period
Q_i^N	Matrix of environmental parameters
$T_{N'i}$	Benchmark matrix of trust
k_i^N	Value of similarity between T_i^N and $T_{N'i}$
p_{N_i}	Change direction of reputation
K_i^N	Similarity between T_i^N and $T_{N'i}$
σ	Adjustment parameter
δ	Threshold for low and medium reputation interval range
ψ	Threshold for medium and high reputation interval range
γ	Threshold of tolerable trust span
η	Parameter for adjusting γ

The first row of S_i^N is multiplied by Q_i^N, subtracted by the first reputation value of T_i^N, and averaged over all the rows. MSE is equivalent to finding the error of the product of matrix S_i^N and matrix Q_i^N with T_i^N matrix. When the error is zero, formula (9) is established. Thus, the optimal solution for Q_i^N is minimizing MSE.

$$Q_i^N = [(S_i^N)^T \cdot S_i^N]^{-1} \cdot (S_i^N)^T \cdot T_i^N \tag{11}$$

$(S_i^N)^{\mathrm{T}}$ is defined as a transpose of matrix S_i^N).

IV. MALICIOUS NODE IDENTIFICATION

To evade the detection of security mechanisms, malicious nodes in the network will show sub-aggressive, selectively discard packets and indirectly forward data, to hide the role of attacker and prolong the life cycle [29]. However, nodes in WSN are distributed randomly, and the working state of each node is different. The poor-quality environment will also lead to abnormal data transmission [30]. Thus, it is an urgent problem to be solved how to distinguish between environmental conditions and packet loss caused by malicious attacks in the WSN security model [31]. The traditional malicious node recognition algorithm sets a single threshold

to judge the nodes in the whole network, which will cause misjudgment in the case of a relatively complex environment. In this paper, the environmental parameter Q_i^N is contacted to set a determination method for each node in the network to adapt to the node state and working environment.

A. BENCHMARK MATRIX OF TRUST

All nodes in the network can be seen as normal nodes after the successful deployment of WSN. The first matrix of reputation Q_i^N is updated, refer to (11). After the next group, S_{i+1}^N is updated, we calculate the benchmark matrix of trust $T_{i+1}^{N'}$ with the current environment parameters Q_i^N.

$$T_{i+1}^{N'} = S_{i+1}^N Q_i^N \tag{12}$$

B. SIMILARITY

Considering the distribution of the WSN node trust value based on the Bayesian evaluation model is approximately subject to be Gaussian. Therefore, in this model, the simplified Gaussian radial basis function is used as a similarity calculation tool.

$$k_i^N = \frac{1}{n}\sum_{j=1}^{n} \exp\left(-\frac{T_{i(j)}^{N'} - T_{i(j)}^{N}}{\sigma^2}\right) \tag{13}$$

$$p_i^N = \begin{cases} 1 & if \quad sumT_{i(j)}^{N'} < sumT_{i(j)}^{N} \\ -1 & if \quad sumT_{i(j)}^{N'} > sumT_{i(j)}^{N} \end{cases} \tag{14}$$

$$K_i^N = p_i^N k_i^N \tag{15}$$

σ is an adjustment parameter and has the radial range of the adjustment function. Its value needs to refer to the span range of node trust value distribution. This paper sets $\sigma = 1$. $k \in [0,1]$. The closer k_i^N is to 1, the more similar the two matrices are. On the contrary, the closer k_i^N is to 0, the less similar the two matrices are. Meanwhile, p_i^N is defined as the change direction of reputation. If the actual reputation is higher than the benchmark trust, p_i^N is equal to 1; if the actual reputation is lower than the benchmark trust, p_i^N is equal to -1. If K_i^N is greater than 0, it is called positive similarity. Otherwise, it is a negative similarity.

C. MALICIOUS NODE RECOGNITION

In the traditional reputation model, only a low threshold value is set. When the reputation value of a node is smaller than this threshold, it is judged to be malicious. We set a double threshold value for the δ and ψ in TRS&EP. Low reputation interval range is 0 to δ, medium reputation interval is δ to ψ, and ψ to 1 for high reputation interval as Fig.2. It is judged as a malicious node when a node is in a low reputation interval. A node in the high reputation interval can be considered as a normal one. There will be sub attack malicious nodes and normal nodes in a relatively bad environment in the credibility interval. In this paper, we aim to identify malicious nodes, especially the sub attack malicious nodes in the medium reputation interval.

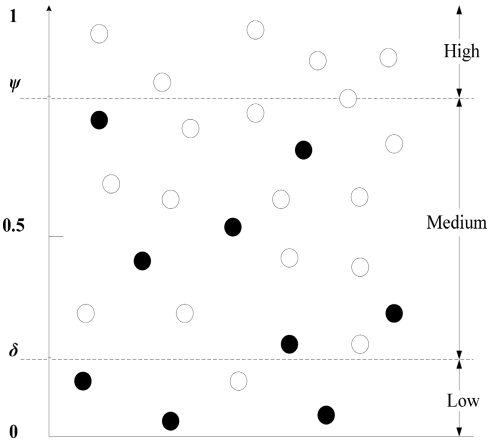


FIGURE 2. Node reputation distribution.

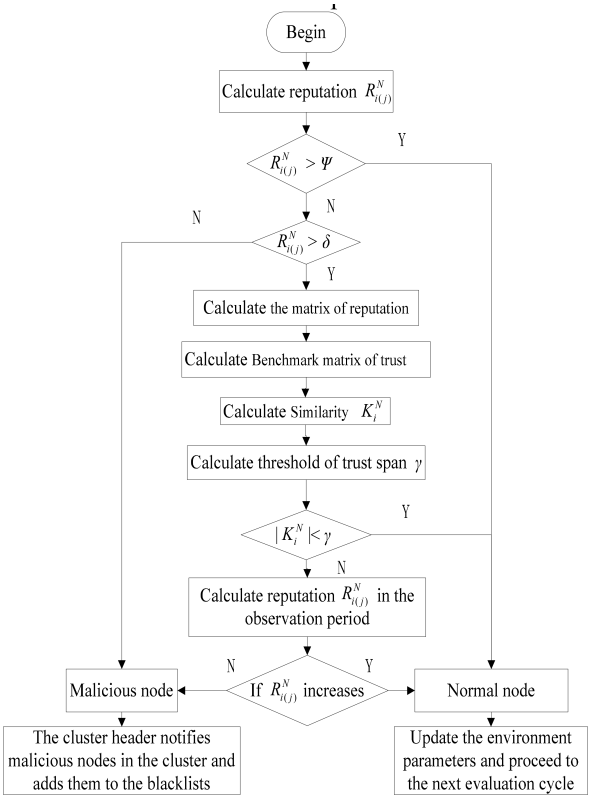


FIGURE 3. Flow chart of the proposed model.

The flow chart of malicious node identification is shown in fig.3. Because of emergencies in small areas in the network environment, we set a node observation cycle parameter as n.

When the node's reputation value is in the medium interval, the similarity and direction between the current matrix of reputation and the benchmark matrix of node trust are judged by combining refer to (15).

The threshold of tolerable trust span of the node is defined as γ. If the node reputation sequence similarity is positive similarity or $|K_i^N| < \gamma$, it is determined to be a normal one and update its Q_{i+1}^N. If the similarity is negative and

TABLE 2. Simulation parameters.

Parameters	Value
Simulation area	100m×100m-200m×200m
Total number of nodes	100-300
Number of cluster heads	4
Number of malicious nodes	10%
Communication radius	20m
Initial total energy	2J
Energy dissipation of send/receiving data	50nJ/bit
Energy consumed by transmitting amplifier circuits in the free space model	10pJ/(bit/m^2)
Energy consumed by transmitting amplifier circuits in the multipath attenuation model	100pJ/(bit/m^2)
Packet size	80bit

$|K_i^N| > \gamma$, the node enters an observation period. If the node's reputation value declines after half an observation cycle(n/2), it is judged to be malicious.

$$\gamma = \eta \exp[\frac{sumT_i^{N'}}{n(\psi - \delta)} - \frac{\delta}{\psi - \delta}] \tag{16}$$

η is used to adjust the size of the trust span threshold γ. $\eta \in (0,1)$. And the smaller the η, the trust threshold span is smaller. The higher the expectation of the benchmark trust sequence is, the higher the span threshold of similarity is.

V. SIMULATION AND ANALYSIS

MATLAB2016a is used to set up the simulation environment. We set the 100m×100m − 200m×200m square area randomly distributed with 100 to 300 nodes, divided into 4 clusters. The effects on nodes in each region are randomly assigned with a probability of 0 to 30%. At the beginning of the simulation, all the normal nodes can fully respond to communication requests. The number of malicious nodes generated after updating the three reputation sequences accounts for 10% of the total nodes, and the packet is discarded with a probability of 50%-100%.

A. NETWORK PERFORMANCE EVALUATION INDEX

In TRS&EP, two indexes are used to determine the performance of the model.

Recognition percentage (*RP*): defined as the number of nodes detected as malicious compared to the total number of malicious nodes.

False-positive percentage (*FPP*): defined as the ratio between the normal nodes incorrectly classified as malicious and the total number of normal nodes.

B. SETTING OF δ AND ψ

To determine the size of the threshold of the δ and ψ of the 100 m by 100 m random distribution 100 nodes, including ten the probability of malicious nodes by 50% to 100% discarded

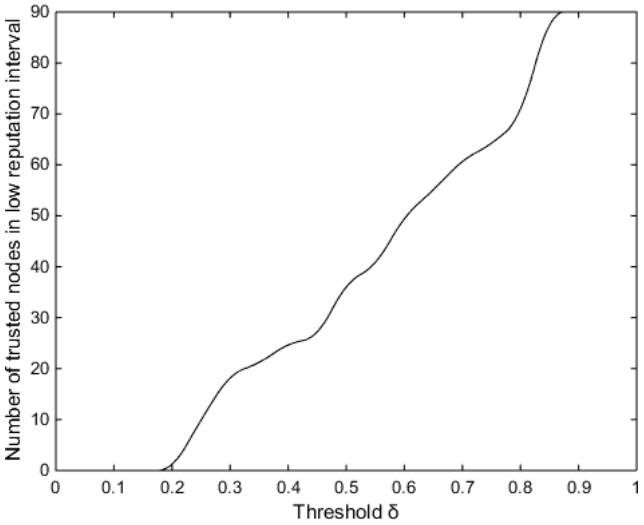


FIGURE 4. Relationship between threshold δ and the number of trusted nodes in the low reputation interval.

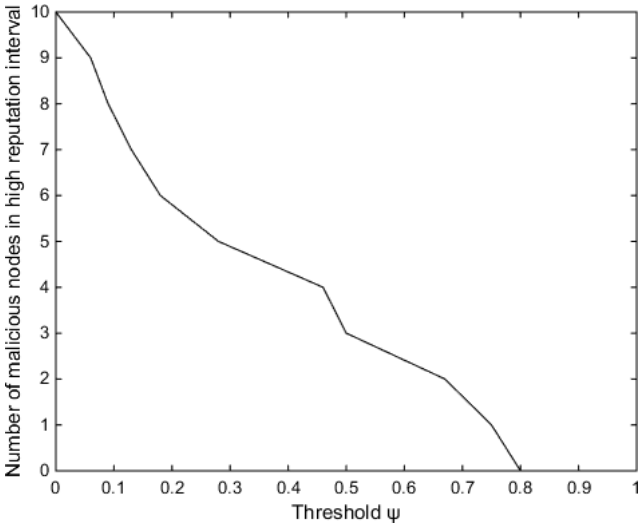


FIGURE 5. Relationship between threshold ψ and the number of malicious nodes in high reputation interval.

packets. First set ψ to 1, identify malicious nodes with TS-BRS reputation model refer to (1) and threshold of the δ changes from 0 to 1, record low prestige area existing trusted node number, the result is shown in Fig.4. Then, under the same simulation environment, to make the threshold of the δ be 0, the threshold ψ of varying from 0 to 1, record the number of malicious nodes appear in the region of the high reputation, shows in Fig.5.

It can be seen from Fig.4. That when the threshold δ moves around 0.2, the first trusted node exists in the low reputation area, which means that the larger the threshold δ is, the more misjudgment will be caused to the trusted node. As Fig.5., when the threshold ψ set as 0.8, there are no malicious nodes in the high reputation interval. Thus, we set threshold δ and ψ set to 0.2 and 0.8, respectively.

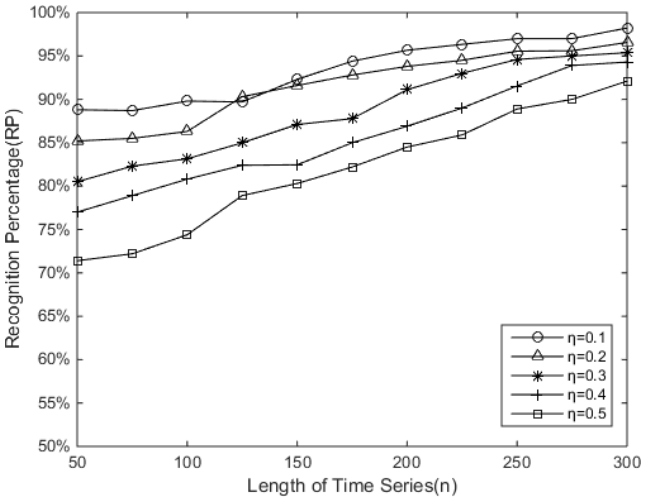


FIGURE 6. Relationship between different η and RP.

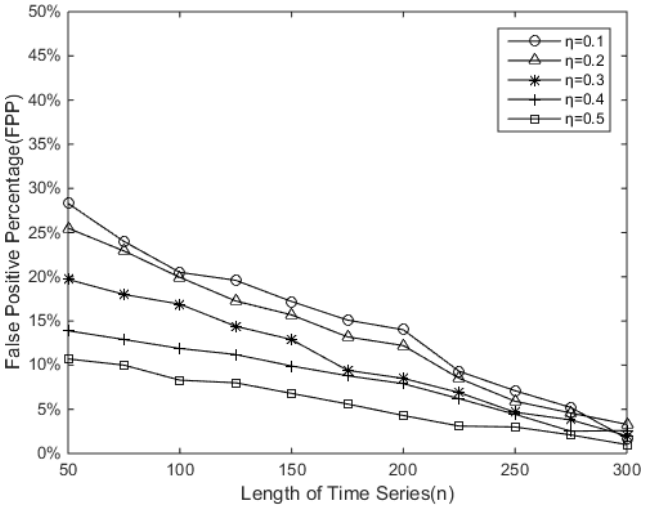


FIGURE 7. Relationship between different η and FPP.

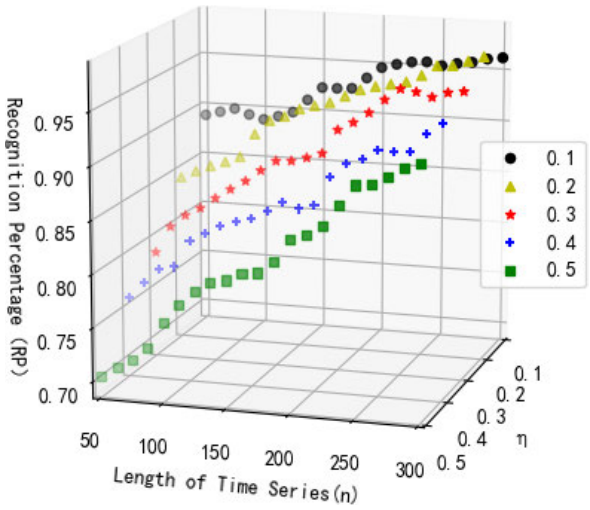


FIGURE 8. Relationship between different η and RP.

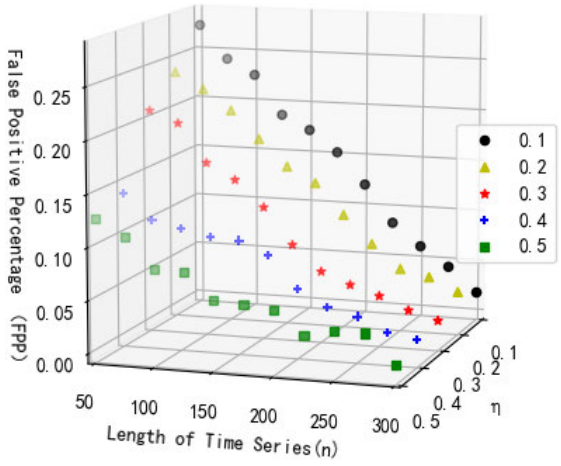


FIGURE 9. Relationship between different η and FPP.

C. SETTING LENGTH OF TIME SERIES n AND PARAMETER η

There are two essential parameters in the TRS&EP scheme: the length n of the time series and the parameter η, which affects the threshold of the reputation span γ. In this paper, the relationship between the two controllable parameters of the simulation experiment and the malicious node recognition rate and the normal node false positive rate of the model is introduced under the unpredictable environment. The length of the time series is from 50 to 300, and the value of the parameter η is from 0.1 to 0.5.

It can be seen from Fig.6. to Fig.9.. The smaller the time series length, the less effective information is extracted for current node's the behavior characteristics. So the recognition and the false-positive rate in the case of using a shorter time series are inaccurate. As the length of the time series increases, the accuracy of the model detection is higher. However, if the time series is too long, the computational overhead will increase, and the node burden will be aggravated when the period becomes longer so that the malicious node has more survival time and has more damage to the network. If a smaller parameter η is chosen, although the ability to identify malicious nodes can be improved, the η decision is so sensitive that small environmental fluctuations can cause misjudgment to normal nodes. The model has insufficient recognition ability with excessive parameters η. After the trade-off, the performance of the system with parameter $n =$ 250, and η =0.3 is better than that of the other parameters.

D. SECURE PERFORMANCE ANALYSIS

When the number of malicious nodes increases in WSN, the pressure on each security model increases. TRS&EP compares the security performance of the model with the LTS and the BTRES model in a complex environment.

Recognition of the three models is reflected under different numbers of malicious nodes in Fig.10. It shows that the recognition percentage of the TRS&EP model is greater than

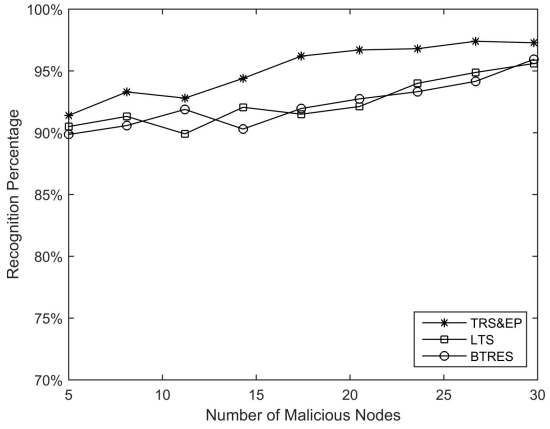


FIGURE 10. Comparison of RP among different models.

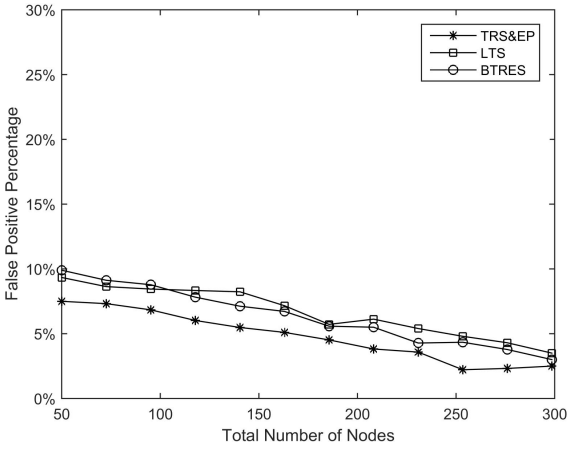


FIGURE 11. Comparison of FPP among different models.

90% and increases steadily, while the recognition rates are lower in the LTS and the BTRES. It is obstructed for the information transmission of the normal node in a relatively complex environment with the existence of sub-attack nodes. The LTS model is more suitable to defend against dishonest suggestion attacks than selective forwarding attacks.

As shown in Fig.11., the false positive rate of the TRS&EP is smaller than LTS's and BTRES's. The security model with a single threshold is affected by environmental factors to produce security performance. Fluctuation leads to an increase in the false-positive rate. The algorithm in this paper set a flexible threshold interval for each node to take into account the impact of the environment. So the overall effect of the model is slightly improved, and the performance is relatively stable.

The results of a random run are shown in Fig.12. The basic parameters are shown in Table 2. In addition, set the threshold δ =0.2, threshold ψ =0.8, the length of the time series n =250, and parameter η =0.3. There are 218 nodes in the whole region, including 20 malicious nodes that have been identified. Recognition percentage is 100% and false-positive percentage is 0%. After 100 simulation experiments,

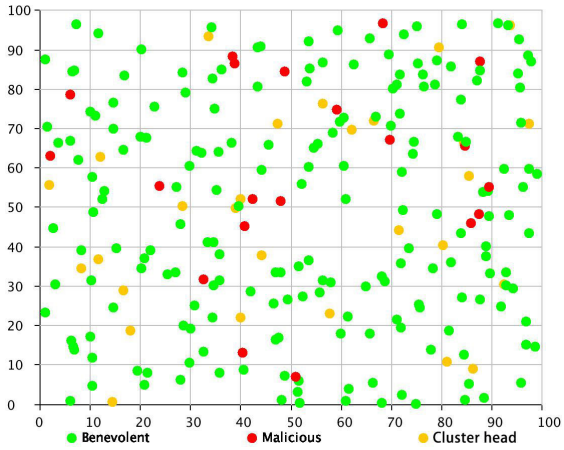


FIGURE 12. Simulation result diagram of malicious node identification.

the average accuracy of identifying malicious nodes in this paper is 97.77%.

VI. CONCLUSION

Trust and reputation model was proved can be used to identify the malicious nodes in WSN. However, nodes deployed in a complex environment with poor-quality links or worse state (less energy or larger workload) will be mistaken for malicious nodes by traditional security mechanisms. Based on the trust and reputation model, this paper constructed a malicious node identification strategy based on environment parameter optimization (TRS&EP). We set the corresponding threshold interval of a reputation for each node in the network by integrating the multidimensional state matrix to calculate the environment parameter. The Gaussian radial basis function is simplified to calculate the similarity between actual reputation and predicted trust. The proposed TRS&EP effectively deals with malicious for sub attacks, such as selective forwarding attacks and interrupt attacks. Simulation results show that TRS&EP can keep the recognition percentage above 90%, and the false-positive percentage is below 8%. Especially, compared to LTS and BTRES, TRS&EP improves the recognition of malicious nodes above 1% and reduces the false-positive percentage by more than 1%. Further development would focus on node energy, forwarding data volume, distribution state, and other factors on information transmission and set the weight of each environmental parameter.

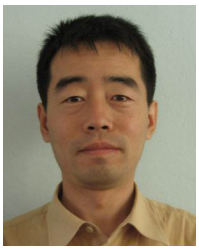
REFERENCES

[1] F. Zawaideh and M. Salamah, "An efficient weighted trust-based malicious node detection scheme for wireless sensor networks," *Int. J. Commun. Syst.*, vol. 32, no. 3, pp. 1–13, Feb. 2019.

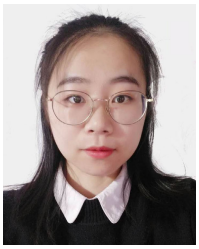
[2] X. Jin, J. Liang, W. Tong, L. Lu, and Z. Li, "Multi-agent trust-based intrusion detection scheme for wireless sensor networks," *Comput. Electr. Eng.*, vol. 59, pp. 262–273, Apr. 2017.

[3] F. Bao, I.-R. Chen, M. Chang, and J.-H. Cho, "Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection," *IEEE Trans. Netw. Service Manage.*, vol. 9, no. 2, pp. 169–183, Jun. 2012.

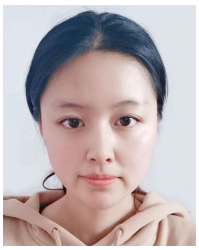
[4] R. Feng, X. Han, Q. Liu, and N. Yu, "A credible Bayesian-based trust management scheme for wireless sensor networks," *Int. J. Distrib. Sens. Netw.*, vol. 11, no. 11, pp. 1–9, Nov. 2015.
[5] O. A. Osanaiye, A. S. Alfa, and G. P. Hancke, "Denial of service defence for resource availability in wireless sensor networks," *IEEE Access*, vol. 6, pp. 6975–7004, Mar. 2018.
[6] B. Jiang, G. Huang, T. Wang, J. Gui, and X. Zhu, "Trust based energy efficient data collection with unmanned aerial vehicle in edge network," *Trans. Emerg. Telecommun. Technol.*, Mar. 2020, doi: 10.1002/ett.3942.
[7] J. Ai, H. Chen, Z. Guo, G. Cheng, and T. Baker, "Mitigating malicious packets attack via vulnerability-aware heterogeneous network devices assignment," *Future Gener. Comput. Syst.*, vol. 111, pp. 841–852, Oct. 2020.
[8] X. Liu, A. Liu, T. Wang, K. Ota, M. Dong, Y. Liu, and Z. Cai, "Adaptive data and verified message disjoint security routing for gathering big data in energy harvesting networks," *J. Parallel Distrib. Comput.*, vol. 135, pp. 140–155, Jan. 2020.
[9] G. E. P. Kumar, I. Titus, and S. I. Thekkekara, "A comprehensive overview on application of trust and reputation in wireless sensor network," *Procedia Eng.*, vol. 38, pp. 2903–2912, Jan. 2012.
[10] D. Qin, S. Yang, S. Jia, Y. Zhang, J. Ma, and Q. Ding, "Research on trust sensing based secure routing mechanism for wireless sensor network," *IEEE Access*, vol. 5, pp. 9599–9609, 2017.
[11] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation based framework for high integrity sensor networks," *ACM Trans. Sensor Netw.*, vol. 4, no. 3, pp. 66–77, May 2008.
[12] L. Zhang, N. Yin, and R. Wang, "Research of malicious nodes identification based on DPAM-DM algorithm for WSN," *J. Commun.*, vol. 36, no. S1, pp. 53–59, Nov. 2015.
[13] Z. Zhang, H. Zhu, S. Luo, Y. Xin, and X. Liu, "Intrusion detection based on state context and hierarchical trust in wireless sensor networks," *IEEE Access*, vol. 5, pp. 12088–12102, Jul. 2017.
[14] T. Khan, K. Singh, L. H. Son, M. Abdel-Basset, H. V. Long, S. P. Singh, and M. Manjul, "A novel and comprehensive trust estimation clustering based approach for large scale wireless sensor networks," *IEEE Access*, vol. 7, pp. 58221–58240, May 2019.
[15] W. Fang, C. Zhang, Z. Shi, Q. Zhao, and L. Shan, "BTRES: Beta-based trust and reputation evaluation system for wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 59, pp. 88–94, Jan. 2016.
[16] R. R. Sahoo, S. Ray, S. Sarkar, and S. K. Bhoi, "Guard against trust management vulnerabilities in wireless sensor network," *Arabian J. Sci. Eng.*, vol. 43, no. 12, pp. 7229–7251, Dec. 2018.
[17] X. Wu, J. Huang, J. Ling, and L. Shu, "BLTM: Beta and LQI based trust model for wireless sensor networks," *IEEE Access*, vol. 7, pp. 43679–43690, Apr. 2019.
[18] T. Suh and Y. Cho, "An enhanced trust mechanism with consensus-based false information filtering algorithm against bad-mouthing attacks and false-praise attacks in WSNs," *Electronics*, vol. 8, no. 11, pp. 1359–1375, Nov. 2019.
[19] J. Yang, W. Li, J. Yan, and L. Hua, "Collision behavior recognizing data forwarding for intermittently connected wireless network," *J. Syst. Eng. Electron*, vol. 39, no. 11, pp. 2571–2579, Nov. 2017.
[20] F. Khedim, N. Labraoui, and A. A. A. Ari, "A cognitive chronometry strategy associated with a revised cloud model to deal with the dishonest recommendations attacks in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 123, pp. 42–56, Dec. 2018.
[21] N. Tariq, M. Asim, Z. Maamar, M. Z. Farooqi, N. Faci, and T. Baker, "A mobile code-driven trust mechanism for detecting internal attacks in sensor node-powered IoT," *J. Parallel Distrib. Comput.*, vol. 134, pp. 198–206, Dec. 2019.
[22] Y. Ren, Z. Zeng, T. Wang, S. Zhang, and G. Zhi, "A trust-based minimum cost and quality aware data collection scheme in P2P network," *Peer-to-Peer Netw. Appl.*, Mar. 2020, doi: 10.1007/s12083-020-00898-2.
[23] H. Fu, Y. Liu, Z. Dong, and Y. Wu, "A data clustering algorithm for detecting selective forwarding attack in cluster-based wireless sensor networks," *Sensors*, vol. 20, no. 1, p. 23, Dec. 2019.
[24] V. R. Prabha and P. Latha, "Fuzzy trust protocol for malicious node detection in wireless sensor networks," *Wireless Pers. Commun.*, vol. 94, no. 4, pp. 2549–2559, Jun. 2017.
[25] Z. Teng, L. Guo, and J. Lv, "WSN Bayes reputation evaluation model based on time series information analysis," *J. Zhengzhou Univ. (Eng. Sci.)*, vol. 40, no. 1, pp. 38–43, Jan. 2019.
[26] Z. Qu, C. Song, Y. Ren, Y. Liu, Q. Niu, and J. Du, "Recommendations based on collaborative filtering by user activity," *J. Northeast Electr. Power Univ.*, vol. 37, no. 5, pp. 4–79, Mar. 2017.
[27] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. HICSS*, Maui, HI, USA, 2000, pp. 1–10.
[28] R. Garg, M. Mittal, and L. H. Son, "Reliability and energy efficient work flow scheduling in cloud environment," *Cluster Comput.*, vol. 22, pp. 1283–1297, Feb. 2019.
[29] R. Kapoor, R. Gupta, L. H. Son, S. Jha, and R. Kumar, "Boosting performance of power quality event identification with KL divergence measure and standard deviation," *Measurement*, vol. 126, pp. 134–142, Oct. 2018.
[30] V. B. Reddy, S. Venkataraman, and A. Negi, "Communication and data trust for wireless sensor networks using D–S theory," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3921–3929, Jun. 2017.



ZHIJUN TENG received the Ph.D. degree in signal and information processing from the Harbin Engineering University of China, in 2013. Since 1994, he has been working with Northeast Electric Power University, as a Teacher, where he has been a Distinguished Professor of information and communication engineering, since 2011. He has published more than 100 articles in international and domestic journals. His research interests include cognitive networks, wireless sensor networks, and map matching.



BAOHE PANG received the B.E. degree in telecommunications engineering from Beihua University and the M.E. degree in information and communication engineering from Northeast Electric Power University, Jilin, China.



CHUNQIU DU received the B.E. degree in smart grid information engineering and the M.E. degree in information and communication engineering from Northeast Electric Power University, Jilin, China.

ZHE LI received the B.E. degree in telecommunications engineering and the M.E. degree in information and communication engineering from Northeast Electric Power University, Jilin, China.