Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 482191, 13 pages
doi:10.1155/2012/482191

Research Article

Subjective Logic-Based Anomaly Detection Framework in Wireless Sensor Networks

Jinhui Yuan,[1, 2, 3] Hongwei Zhou,[1, 2, 3] and Hong Chen[1, 2]

[1] *Key Laboratory of Data Engineering and Knowledge Engineering, MOE, Beijing 100872, China*
[2] *School of Information, Renmin University of China, Beijing 100872, China*
[3] *Institute of Electronic Technology, Information Engineering University, Zhengzhou 450004, China*

Correspondence should be addressed to Jinhui Yuan, jcyjh@126.com

Received 15 June 2011; Revised 25 September 2011; Accepted 28 September 2011

Academic Editor: Yuhang Yang

In existing anomaly detection approaches, sensor node often turns to neighbors to further determine whether the data is normal while the node itself cannot decide. However, previous works consider neighbors' opinions being just normal and anomalous, and do not consider the uncertainty of neighbors to the data of the node. In this paper, we propose SLAD (subjective logic based anomaly detection) framework. It redefines opinion deriving from subjective logic theory which takes the uncertainty into account. Furthermore, it fuses the opinions of neighbors to get the quantitative anomaly score of the data. Simulation results show that SLAD framework improves the performance of anomaly detection compared with previous works.

1. Introduction

Recently wireless sensor networks (WSNs) have been widely used in military surveillance, traffic monitoring, habitat monitoring and object tracking, and so forth [1, 2]. Such networks deploy lots of sensor nodes with sensing, data processing, and wireless communication capabilities in the monitoring area. Sensor nodes are resource-constrained and susceptible to interference from the environment so that their sensing data are often unreliable. Potential sources of anomalous data in WSNs are classified into three categories: faults (errors), events, and malicious attacks [3, 4]. While sensor nodes fail, their sensing data are faulty data [5]. Once the number of faulty data increases, it will bring great influence on the user query. Thus, they should be eliminated or corrected. When some event happens, the sensing data of the nodes in the area are informational data, which are different from the normal data. They should be reported to user for further deciding. The thirdly potential source of anomalous data is attacks which are beyond the scope of this paper. Anomaly detection is considered as a solution to detect faulty data and informational data.

In existing anomaly detection approaches, sensor nodes turn to neighbors to further determine whether the data is normal while the node itself cannot decide. In this process, existing solutions, including voting algorithms [6, 7] and aggregation frameworks [8–10] which detect anomaly in the process of aggregating data, provide neighbors' opinions being just normal and anomalous. However, no neighbor can always say that the data of the node are absolutely normal or anomalous, and something is neglected by previous works which we call uncertainty. Thus, taking the degree for neighbors' opinions about the data being normal or anomalous into account can more realistically describe the view of neighbors. Consequently, the performance of anomaly detection is able to be improved.

In this paper, we propose SLAD (subjective logic-based anomaly detection) framework, which takes uncertainty into account, to improve the performance of anomaly detection. It includes three phases: preprocessing, self-monitoring, and cooperant detecting. Among them, pre-processing run on sink and self-monitoring execute on each node. After the two phases, sensor nodes send suspicious data to its neighbors to turn to further determine. The third phase is the key of our framework.

The important element of SLAD is ESLB (extended subjective logic-based algorithm), which is the key of the third phase mentioned above. Before plunging into the detail of

ESLB, we first propose SLB (subjective logic-based algorithm) which elementarily describe our work. In SLB, each neighbor gives the quantitative opinion to the suspicious data involving with subjective logic theory. After fusing the opinions of all the neighbors, SLB gets the quantitative anomaly score, which demonstrates the degree of the suspicious data being considered as an anomaly. We extend SLB to ESLB in order to avoid the impact of those neighbors whose data are suspicious, effectively distinguish the faulty data from the informational data, and take the historical spatial correlations of the node and its neighbors into account.

The main contributions of this paper are as follows.

(i) Proposes SLAD framework which takes the uncertainty of neighbors to the data of the node into account. It redefines opinion deriving from subjective logic theory and can more realistically describe the view of neighbors on the data of the node.

(ii) Presents SLB and ESLB algorithms. SLB fuses all the opinions of neighbors for the data of the node to get the quantitative anomaly score of the data. We extend SLB to ESLB to improve the performance further.

(iii) Constructs the experiments to verify the detection performance of the framework we propose. Simulation results show that SLAD framework is effective and gains a lot of performance improvement of anomaly detection compared with the previous approaches.

The rest of the paper is organized as follows. Section 2 summarizes the related work of this paper. Section 3 presents preliminary concepts. Our framework SLAD is introduced in Section 4. Section 5 gives SLB algorithm and its extended algorithm ESLB. Section 6 discusses some problems which are not involve in the above sections. Section 7 describes the experimental setup and evaluates the performance of framework in realistic data set. Finally, Section 8 concludes the paper.

2. Related Work

Lots of efforts have been made in recent years to detect the anomaly in wireless sensor networks. We briefly survey the recent researches relevant to our work as follows.

First category involves voting algorithm and its improved algorithms. Authors in [6] propose majority voting algorithm. If some node v is aware that it's sensing data x maybe anomalous, it sends x to its all one-hop neighbors. Each neighbor v' compares x with its sensing data x'. If the difference is less than the threshold, v' casts a positive vote for v, otherwise casts a negative vote. Node v collects all the votes of its neighbors and gets the determination. If the number of positive votes is more than negative votes, x is thought to be normal, otherwise is anomalous. Based on majority voting algorithm [6, 7] proposes weighted voting algorithm which considers that the neighbors who are closer to the node should have greater weights. Authors in [11] discuss how to detect the faulty (erroneous) data in WSNs. It uses extended Jaccard's coefficient to compute the similarity degree between sensor nodes and set the different levels for the nodes to set up the correlation network. It presents an efficient two-phase voting algorithm called TrustVoting to determine whether the data is faulty. However, the algorithms mentioned above provide neighbors' opinions being just normal and anomalous. In addition, taking the degree for neighbors' opinions about the data being normal or anomalous into account is able to improve detection performance [4].

Second category is to detect anomaly in the process of aggregating data in the network. Authors in [8] propose a robust aggregate framework, which performs the similarity tests among sensor nodes to classify the particular node as anomaly. It returns the aggregate results excluding anomaly, which is also maintained and sent to the users. Furthermore, authors in [9] define minimum support MinSupp, which is the minimum count of sensor nodes to prove the data of the node being normal or anomalous. For some node holds on anomalous data, if it has MinSupp number of nodes whose data are similar to it, it is determined that some events happen, otherwise it is faulty data. On this basis, [10] present the in-network anomaly detection framework based on position sensitive hash function. It achieves the load balance of the network. Using comparison pruning methods, it assures the detection performance and energy efficiency. Authors in [12] introduce PAO framework to reliably and efficiently detect anomaly in WSNs, which is able to operate over multiple window type, and operate in exact or approximate mode suiting for a variety of application requirements. However, the outputs of similarity test for all these frameworks mentioned above are also only yes or no, which depends on the prethreshold, and do not provide quantitative determination, which are similar to the voting algorithms.

The third one regard the sensing data of the nodes as time-series data to some extent. Authors in [13, 14] construct autoregressive (AR) models for sensor nodes. Every sensor node sends the coefficients of the models to sink after establishing AR models, and sink estimates approximate values of the sensor nodes in the following rounds without getting real data from sensor nodes. Thus, it reduces the number of messages sent in the network a lot. Once the data are no longer predicable from AR models, it maybe due to that the models are not suitable to the data or anomalous data arise. If the reason is the former, it needs reconstructing AR models and repeating the process mentioned above. Otherwise, the anomalous data are identified to be eliminated or corrected. Authors in [13] use two thresholds to distinguish them. However, the approach only relies on the predefined thresholds and does not employ the spatial correlations among sensor nodes. If taking spatial correlations into account, it can make full use of neighbors' opinions to achieve better performance of anomaly detection.

According to the above-related works, we can draw the conclusion that providing quantitative opinions is very important for anomaly detection after self-monitoring on each node in WSNs. As we know, in subjective logic theory, the subjects express subjective beliefs about the truth of the objects with degree of uncertainty and indicate subjective belief ownership whenever required [15, 16]. Subjective logic provides the quantitative evaluation for the trust degree of the object. From this perspective, judgment among the adjacent nodes in WSNs is similar to trust evaluation. So we take subjective logic theory into the anomaly detection in WSNs. Subjective logic is involved to offer quantitative neighbors' opinions about the suspicious data of the node.

Besides, authors in [17–19] use machine learning techniques to detect anomaly in WSNs, which are different from our solution. For machine learning techniques are resource intensive that are difficult to be implemented on sensor nodes, the early studies, for example [17], run their algorithms on gateway (or sink). Authors in [17] identify anomalies in critical gas monitoring using offline echostate network in an underground coal mine. The following researches try to do something to make it possible to run the algorithms on sensor nodes. Authors in [18] compares and classifies the input signals in accordance with online learned prototypes on node-level, and then sends the results of classification to a fusion center for further processing. Based on [17], the authors in [19] propose a general anomaly detection framework which unifies fault and event detection. It runs on sensor nodes, distinguishes faults from events, and improves the performance of detection. The focuses of [18, 19] are how to select appropriate machine learning techniques and then decrease the complexity to make the algorithms be suitable to run on nodes. It is different from our solution, the difficulty of which is how to provide the quantitative neighbors' opinions to improve the performance of detection.

3. Preliminaries

Suppose that a sensor network is modeled as an undirected connected graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, where $\mathbb{V}$ is the set of all sensor nodes (including n sensor nodes $v_1, \ldots, v_n$ and one sink v_0, denoted as $\mathbb{V} = V_n \cup v_0$) and $\mathbb{E}$ is the set of the edges. An anomaly is defined as a measurement that significantly deviates from the normal pattern of the sensing data [3]. Generally, the anomaly mentioned in this paper includes fault (error) and event, and the anomalous data includes the faulty (erroneous) data and the informational data, respectively.

For the data of sensor nodes can be regarded as time series data [13, 14], we construct AR model on each node. Suppose that the data of node v_i at time t can be denoted by AR(p) as $x_{it} = \sum_{k=1}^{p} \varphi_k x_{i(t-k)} + \varepsilon$, where $x_{i(t-k)}$ is the data of v_i at time $t - k(1 \le k \le p)$, φ_k is the corresponding coefficient of $x_{i(t-k)}$, and ε is the random error and is the normal distribution of the mean being 0 and the variance being σ^2. After that, given $\Phi = [\varphi_1 \cdots \varphi_p]'$and $X_t = [x_{1t} \cdots x_{nt}]'$ we can get $\widehat{\Phi}$ and $\widehat{X}_t$. Among them, $\widehat{\Phi}$ is the linear and the least variance-unbiased estimation of Φ, and $\widehat{X}_t$ is the unbiased estimation of X_t:

$$\widehat{\Phi} = \left[\widehat{\varphi}_1 \cdots \widehat{\varphi}_p\right]' = (Y'Y)^{-1} Y'Z, \\ \widehat{X}_t = \widehat{\varphi}_1 X_{t-1} + \cdots + \widehat{\varphi}_p X_{t-p}, \tag{1}$$

where $Y = [X_j \cdots X_{j-p+1}]_{j=p\cdots M-1}$, $X_j = [x_{1j} \cdots x_{nj}]'$, $Z = [X_{p+1} \cdots X_M]'$. At last, given the confidence level 1-α, the confidence interval of the estimate value $\widehat{X}_t$ is

$$\left(\widehat{X}_t \pm t_{\alpha/2}(M - 2p) \cdot \widehat{\sigma}\sqrt{1 + Y_0 (Y'Y)^{-1} Y_0'}\right). \tag{2}$$

We make the following assumptions about our framework.

(1) The wireless sensor network is static, and the topology does not change in the network lifetime.

(2) All sensor nodes are homogeneous and have the same energy and capabilities, and there is only one sink which holds on infinite energy.

(3) Sensor nodes are deployed densely; that is, if some events happen in the network, adjacent sensor nodes (one-hop neighbors) can monitor them at the same time. Of course, the situation can be extended to not densely deployed, which will be discussed in Section 6.

4. SLAD Framework

SLAD framework consists of three phases: preprocessing, self-monitoring, and cooperant detecting. Among them, preprocessing phase is executed on sink, self-monitoring run on each node, and cooperant detecting is semidistributed algorithm, that is, run on sink and sensor node.

In the first phase, all sensor nodes collect N rounds of data and transmit them to sink. Sink constructs autoregressive models AR(p) and uses the least squares to estimate the coefficients $\varphi_k (1 \le k \le p)$. As for ε, it is estimated by use of the first M rounds of data. Using the least p rounds of data and the coefficients φ_k, we get the estimate value of the nodes. After that, using the last N-M rounds data, we get the confidence interval $(\widehat{X} \pm c_{it})$ under the given confidence level 1-α.

For each node v_i, if its data x_{it} at time t is within the range of its confidence interval $(\widehat{x}_{it} \pm c_{it})$, it is considered as normal, otherwise anomalous. However, this computation run on each node, if it is computed at each round on each node, the computational complexity is so high as to consume too much energy, which significantly leads to increased energy consumption. Consequently, a simple approach is taken to approximate as shown below. Through the use of φ_k, each node predicts the latest N-M rounds of data and compares them with the real data to get the average value of the confidence intervals of those N-M rounds data, which is set as approximate confidence interval $(\widehat{x}_{it} \pm \tau_i)$ at the given confidence level. Then it reduces the computational complexity on each node a lot. Sink sends the messages to each node including p coefficients of its AR model and its respectively approximate confidence interval.

In the second phase, each node uses p coefficients of its AR model and the most recently p rounds of data to predict current round of data. If the difference between the predicative data and the real data is less than the threshold τ, SLAD considers the data as normal. Otherwise, the data is regarded as suspicious which needs to be determined further among adjacent neighbors. It is noted that, if the data is thought to be normal, it does not compute the confidence interval. However, while v considers x_t to be suspicious, it computes $(\hat{x}_t \pm c_t)$ at 1-α. And then, it sends the message to all its one-hop neighbors, which include x_t and $(\hat{x}_t \pm c_t)$.

In the third phase, sensor node whose data is suspicious sends its data to all its neighbors, and each neighbor produces its opinion about the suspicious data. SLAD fuses all the neighbors' opinions and gets the expectation of the consensus opinion. And thus we get the anomaly score of the suspicious data. If the anomaly score is more than the threshold, the suspicious data is anomalous, or else the data is normal. Additionally, to avoid the impact of those neighbors' opinions whose sensing data are suspicious, SLAD removes those opinions from the consensus opinion. In order to take the historical spatial correlations of the node and its neighbor nodes into account, SLAD computes the neighbors' opinions in another way. For the reason of different treatments to faulty data and informational data, SLAD adopt the approach as follows. The suspicious data, if anomalous, is to be marked as faulty data. When the faulty data of sensor nodes at this round are all sent to sink, sink distinguishes faulty data and informational data by employing the spatial correlations of adjacent nodes. The detailed process will be discussed further in Section 5. The third phase is the fundamental step of SLAD framework, which will be discussed in detail in Section 5.

5. Subjective Logic-Based Algorithms

In WSNs, no neighbor can always say that the data of the node are absolutely normal or anomalous, and something is neglected by previous works which we call uncertainty. On the other hand, subjective logic theory is suitable to model the situations with consideration to uncertainty. This drives us to involve subjective logic theory in anomaly detection to improve the detection performance.

Before detailing the subjective logic-based algorithms, it is necessary to address three problems, including expressiveness of neighbors' opinions, value assignment of neighbors' opinions, and consensus of neighbors' opinions. With the solutions of the problems, we propose SLB and ESLB which is the extension of SLB.

5.1. Expressiveness of Neighbors' Opinions

Definition 1. Given sensor network $\mathbb{G} = (\mathbb{V}, \mathbb{E}), v, v_i \in \mathbb{V}, (v, v_i) \in \mathbb{E}$, the opinion of the neighbor v_i about the sensing data of node v is defined as follows:

$$\omega_v^{v_i} = \left(s_v^{v_i}, d_v^{v_i}, u_v^{v_i}, a_v^{v_i}\right), \qquad s_v^{v_i} + d_v^{v_i} + u_v^{v_i} = 1, \tag{3}$$

where $s_v^{v_i}$ is the degree of belief that neighbor v_i considers the data of node v to be normal. $d_v^{v_i}$ is the degree of disbelief that v_i considers the data of node v to be anomalous. $u_v^{v_i}$ is the degree of uncertainty that v_i regards the data of node v as normal or anomalous. $a_v^{v_i}$ is the base rate of that v_i regards the data of node v as normal or anomalous (i.e., a priori probability).

Definition 1 defines neighbor v_i's opinion about the degree of node v's data. $s_v^{v_i}, d_v^{v_i}$ and $u_v^{v_i}$ are combined to express the opinion thoroughly. The following problem is how to determine the opinion $\omega_v^{v_i}$ of neighbor v_i about the data of node v.

5.2. Value Assignment of Neighbors' Opinions. In this section, we discuss how to determine neighbor's opinion $\omega_v^{v_i}$. We compute the similarity degree and difference degree of node v and v_i to denote as $s_v^{v_i}$ and $d_v^{v_i}$, respectively. It is worth mentioning that the sum of $s_v^{v_i}$ and $d_v^{v_i}$ maybe more than one by use of the above method. In the case, we should scale the sum down to no more than one because of the requirement of the subjective logic theory. $u_v^{v_i}$ is equal to subtract the sum of $s_v^{v_i}$ and $d_v^{v_i}$ from one.

To scale them down, we take advantage of the observation that the data of the nodes are changing smoothly most of the time and changing nonsmoothly every some periods for the reason the sampling rates of the nodes are high in WSNs. We have taken into account the data trends while constructing AR model. So we just use the data at the current round to determine neighbors' opinions while the data are changing smoothly. Only while the data are changing non-smoothly, we use several rounds of data to get the neighbors' opinions. As we know, data trends of the nodes can be get according to historical data.

The detailed opinion $\omega_v^{v_i}$ of neighbor v_i about the data of node v is determined as follows.

(1) If the data are changing smoothly,

$$s_v^{v_i} = \begin{cases} \dfrac{x_i}{x}, & x_i \leq x \\[2ex] \dfrac{x}{x_i}, & x < x_i, \end{cases} \qquad d_v^{v_i} = \frac{2 \cdot |x_i - x|}{(x_i + x)}, \tag{4}$$

where x_i and x are the data of node v_i and node v, respectively, at current round. If $s_v^{v_i} + d_v^{v_i} > 1$, the sum is scaled down to no more than one. $u_v^{v_i} = 1 - s_v^{v_i} - d_v^{v_i} \cdot a_v^{v_i}$ is the prior probability of v_i's opinion about v's data, that is, the expectation of the prior opinion. Initially it is set to 0.5; that is, v_i considers the probability of the data of v being normal and anomalous is 0.5.

(2) If the data are changing nonsmoothly,

$$\begin{aligned} s_v^{v_i} &= \frac{X_i \cdot X}{\|X_i\|^2 + \|X\|^2 - X_i \cdot X}, \\ d_v^{v_i} &= \sum_{j=1}^{l} \frac{2 \cdot |X_i(j) - X(j)|}{l \cdot (X_i(j) + X(j))}, \end{aligned} \tag{5}$$

where $X_i = [x_{1i} \cdots x_{li}], X = [x_1 \cdots x_l]$, supposing the current round is l, X_i and X are the vector data of node v_i and v from 1 round to l rounds, $X_i(j)$ and $X(j)$ are the jth element of X_i and X, l is the length of vector data (X_i and X). If $s_v^{v_i} + d_v^{v_i} > 1$, the sum is scaled down to no more than one. $u_v^{v_i} = 1 - s_v^{v_i} - d_v^{v_i} \cdot a$ is same as above.

5.3. Consensus of Neighbors' Opinions. The opinions of neighbors v_i and v_j about node v's data can be fused to get the consensus which is the new opinion about the proposition on node v's data being anomalous according to Lemma 2.

Lemma 2. *Given* $v, v_i, v_j \in \mathbb{V}, (v, v_i) \in \mathbb{E}, (v, v_j) \in \mathbb{E}, \omega_v^{v_i} = (s_v^{v_i}, d_v^{v_i}, u_v^{v_i}, a_v^{v_i})$ *and* $\omega_v^{v_j} = (s_v^{v_j}, d_v^{v_j}, u_v^{v_j} a_v^{v_j})$ *are the opinions of neighbors* v_i *and* v_j *about the data of node* v, $\omega_v^{v_i,v_j} = (s_v^{v_i,v_j}, d_v^{v_i,v_j}, u_v^{v_i,v_j}, a_v^{v_i,v_j})$ *is the consensus of two neighbors' (*v_i *and* v_j*) opinions about the proposition on node* v*'s node being anomalous, it can be computed as follows. Let* $k = u_v^{v_i} + u_v^{v_j} - u_v^{v_i} u_v^{v_j}$.

If $k \neq 0$,

$$\begin{aligned} s_v^{v_i,v_j} &= \frac{d_v^{v_i} u_v^{v_j} + d_v^{v_j} u_v^{v_i}}{k}, \\ d_v^{v_i,v_j} &= \frac{s_v^{v_i} u_v^{v_j} + s_v^{v_j} u_v^{v_i}}{k}, \\ u_v^{v_i,v_j} &= \frac{u_v^{v_i} u_v^{v_j}}{k}, \\ a_v^{v_i,v_j} &= \frac{(1 - a_v^{v_i})(1 - u_v^{v_i}) u_v^{v_j} + \left(1 - a_v^{v_j}\right)\left(1 - u_v^{v_j}\right) u_v^{v_i}}{k - u_v^{v_i} u_v^{v_j}}. \end{aligned} \tag{6}$$

If $k = 0$,

$$\begin{aligned} s_v^{v_i,v_j} &= \frac{d_v^{v_j} + d_v^{v_i} \gamma}{\gamma + 1} \\ d_v^{v_i,v_j} &= \frac{s_v^{v_j} + s_v^{v_i} \gamma}{\gamma + 1} \\ u_v^{v_i,v_j} &= 0 \\ a_v^{v_i,v_j} &= \frac{\left(1 - a_v^{v_j}\right) + \gamma(1 - a_v^{v_i})}{\gamma + 1}, \end{aligned} \qquad \gamma = \lim\left(\frac{u_v^{v_j}}{u_v^{v_i}}\right) \tag{7}$$

Proof. From [15], we know that posteriori probabilities (ppdf) of binary events can be expressed as

$$\begin{aligned} f(p \mid r, t, a) = {} & \frac{\Gamma(r + t + 2)}{\Gamma(r + 2a)\Gamma(t + 2(1 - a))} p^{r+2a-1} \\ & \times (1 - p)^{t+2(1-a)-1}, \end{aligned} \tag{8}$$

where $0 \leq p \leq 1, r \geq 0, t \geq 0, 0 < a < 1$.

Here r, t, and a represent positive evidence, negative evidence, and relative atomicity (base rate), respectively. The probability expectation value is $E(f(p)) = (r+2a)/(r+t+2)$.

Let $f(p \mid r_v^{v_i}, t_v^{v_i}, a_v^{v_i})$ and $f(p \mid r_v^{v_j}, t_v^{v_j}, a_v^{v_j})$ be two ppdfs, respectively, held by the neighbor nodes v_i and v_j regarding the truth of the suspicious sensing data of the node v. The ppdf $f(p \mid r_v^{v_i,v_j}, t_v^{v_i,v_j}, a_v^{v_i,v_j})$ defined as that [15]:

$$\begin{aligned} r_v^{v_i,v_j} &= r_v^{v_i} + r_v^{v_j}, \\ t_v^{v_i,v_j} &= t_v^{v_i} + t_v^{v_j}, \\ a_v^{v_i,v_j} &= \frac{a_v^{v_i}(r_v^{v_i} + t_v^{v_i}) + a_v^{v_j}\left(r_v^{v_j} + t_v^{v_j}\right)}{r_v^{v_i} + t_v^{v_i} + r_v^{v_j} + t_v^{v_j}}. \end{aligned} \tag{9}$$

Let $\omega = (s, d, u, a)$ be a neighbor node's opinion about the suspicious sensing data, and let $f(p \mid r, t, a)$ be the same neighbor node's probability estimate regarding the same data. For $E(f(p)) = E(\omega)$, that is, $(r+2a)/(r+t+2) = s+au$, and $s + d + u = 1$, it is easy to get $r = 2s/u$, $t = 2d/u$, where $u \neq 0$.

The following is the process to prove that the equations of the lemma are correct. Because we want to get the consensus about the proposition on node v's data is anomalous, we get the equations with exchanging r and t of (9); respectively,

$$r_v^{v_i,v_j} = t_v^{v_i} + t_v^{v_j} = \frac{2d_v^{v_i}}{u_v^{v_i}} + \frac{2d_v^{v_j}}{u_v^{v_j}} = \frac{2d_v^{v_i} u_v^{v_j} + 2d_v^{v_j} u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} = \frac{2d_v^{v_i,v_j}}{u_v^{v_i,v_j}}, \tag{10}$$

$$t_v^{v_i,v_j} = r_v^{v_i} + r_v^{v_j} = \frac{2s_v^{v_i}}{u_v^{v_i}} + \frac{2s_v^{v_j}}{u_v^{v_j}} = \frac{2s_v^{v_i} u_v^{v_j} + 2s_v^{v_j} u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} = \frac{2s_v^{v_i,v_j}}{u_v^{v_i,v_j}}, \tag{11}$$

$$(10) \Longrightarrow \frac{d_v^{v_i} u_v^{v_j} + d_v^{v_j} u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} = \frac{d_v^{v_i,v_j}}{u_v^{v_i,v_j}}, \tag{12}$$

$$(11) \Longrightarrow \frac{s_v^{v_i} u_v^{v_j} + s_v^{v_j} u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} = \frac{s_v^{v_i,v_j}}{u_v^{v_i,v_j}}, \tag{13}$$

$$(12) + (13) \Longrightarrow \frac{(s_v^{v_i} + d_v^{v_i}) u_v^{v_j} + \left(s_v^{v_j} + d_v^{v_j}\right) u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} = \frac{s_v^{v_i,v_j} + d_v^{v_i,v_j}}{u_v^{v_i,v_j}}, \tag{14}$$

$$(14) \Longrightarrow \frac{(1 - u_v^{v_i}) u_v^{v_j} + \left(1 - u_v^{v_j}\right) u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} = \frac{1 - u_v^{v_i,v_j}}{u_v^{v_i,v_j}}, \tag{15}$$

$$(15) \Longrightarrow 1 + \frac{u_v^{v_j} - u_v^{v_i} u_v^{v_j} + u_v^{v_i} - u_v^{v_i} u_v^{v_j}}{u_v^{v_i} u_v^{v_j}} = \frac{1}{u_v^{v_i,v_j}}. \tag{16}$$

Let $k = u_v^{v_i} + u_v^{v_j} - u_v^{v_i} u_v^{v_j}$.

If $k \neq 0$,

$$(16) \Longrightarrow u_v^{v_i,v_j} = \frac{u_v^{v_i} u_v^{v_j}}{k}. \tag{17}$$

Combining (17) onto (12), we get

$$s_v^{v_i,v_j} = \frac{d_v^{v_i} u_v^{v_j} + d_v^{v_j} u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} \times \frac{u_v^{v_i} u_v^{v_j}}{k} = \frac{d_v^{v_i} u_v^{v_j} + d_v^{v_j} u_v^{v_i}}{k}. \tag{18}$$

Combining (17) onto (13), we obtain

$$\begin{aligned} d_v^{v_i,v_j} &= \frac{s_v^{v_i} u_v^{v_j} + s_v^{v_j} u_v^{v_i}}{u_v^{v_i} u_v^{v_j}} \times \frac{u_v^{v_i} u_v^{v_j}}{k} \\ &= \frac{s_v^{v_i} u_v^{v_j} + s_v^{v_j} u_v^{v_i}}{k} \end{aligned} \tag{19}$$

$$\begin{aligned}a_v^{v_i,v_j} &= \frac{(1-a_v^{v_i})(r_v^{v_i}+s_v^{v_i})+\left(1-a_v^{v_j}\right)\left(r_v^{v_j}+s_v^{v_j}\right)}{r_v^{v_i}+s_v^{v_i}+r_v^{v_j}+s_v^{v_j}}\\ &= \frac{(1-a_v^{v_i})(2(s_v^{v_i}+d_v^{v_i})/u_v^{v_i})+\left(1-a_v^{v_j}\right)\left(2\left(s_v^{v_j}+d_v^{v_j}\right)/u_v^{v_j}\right)}{(s_v^{v_i}+d_v^{v_i})/u_v^{v_i}+\left(s_v^{v_j}+d_v^{v_j}\right)/u_v^{v_j}}\\ &= \frac{(1-a_v^{v_i})(1-u_v^{v_i})u_v^{v_j}+\left(1-a_v^{v_j}\right)\left(1-u_v^{v_j}\right)u_v^{v_i}}{k-u_v^{v_i}u_v^{v_j}}.\end{aligned} \tag{20}$$

If $k = 0$, let $\gamma = \lim(u_v^{v_j}/u_v^{v_i})$, it is easy to get the equation (7) which is similar as the above. □

To be simply presented, we denote $\omega_v^{v_i,v_j} = (s_v^{v_i,v_j}, d_v^{v_i,v_j}, u_v^{v_i,v_j}, a_v^{v_i,v_j})$ as $\omega_v^{v_i,v_j} \equiv \omega_v^{v_i}\overline{\oplus}\omega_v^{v_j}$, among which $\overline{\oplus}$ is the new operator which is similar to the consensus operator of subjective logics. The expectation of consensus of neighbors' opinion about the data of node v decides the thorough consideration of neighbors about the data of v. Given consensus of neighbors' opinion $\omega_v^{v_i,v_j}$, the expectation of the opinion is $E(\omega_v^{v_i,v_j}) = s_v^{v_i,v_j} + a_v^{v_i,v_j}u_v^{v_i,v_j}$.

Example 3. Suppose that the opinions of neighbors v_i and v_j about the data of node v are $\omega_1 = (0.7, 0.2, 0.1, 0.5)$ and $\omega_2 = (0.8, 0.1, 0.1, 0.5)$ at some round, respectively, then the consensus of the opinions is $\omega_{1,2} = \omega_1\overline{\oplus}\omega_2 = (s_{1,2}, d_{1,2}, u_{1,2}, a_{1,2}) = (0.158, 0.789, 0.053, 0.50)$, and the expectation is $E(\omega_{1,2}) = s_{1,2} + a_{1,2}u_{1,2} = 0.158 + 0.5 \times 0.053 = 0.184$.

As we all know, each node has many neighbors in WSNs. We need to fuse the opinions of all neighbors into the consensus opinion. Suppose that node v has m neighbors, their opinions about the data of v are $\omega_v^{v_1} = (s_v^{v_1}, d_v^{v_1}, u_v^{v_1}, a_v^{v_1}), \ldots, \omega_v^{v_m} = (s_v^{v_m}, d_v^{v_m}, u_v^{v_m}, a_v^{v_m})$. To get the thorough consideration of neighbors about v's data, we fuse all its neighbors' opinions, which denote as $\omega_v^{v_1,v_2,\ldots,v_m} \equiv \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2}\overline{\oplus}\cdots\overline{\oplus}\omega_v^{v_m}$, that is, $\omega_v = (s_v, d_v, u_v, a_v)$. The consensus process is recursively called by use of Theorem 4.

Theorem 4. *Given i neighbors $v_1, v_2, \ldots, v_i$ of node v, their opinions about the data of v are $\omega_v^{v_1} = (s_v^{v_1}, d_v^{v_1}, u_v^{v_1}, a_v^{v_1}), \ldots, \omega_v^{v_i} = (s_v^{v_i}, d_v^{v_i}, u_v^{v_i}, a_v^{v_i})$, the consensus of their opinions about the proposition on node v's node being anomalous is $\omega_v^{v_1,\ldots,v_i}$, then it can be computed as follows:*

$$\omega_v^{v_1,\ldots,v_i} = \omega_v^{v_1,\ldots,v_{i-1}}\overline{\oplus}\omega_v^{v_i} = \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2,\ldots,v_i} \quad (2 \le i \le m). \tag{21}$$

Proof. We utilize the mathematical induction approach to prove the theorem.

(1) If $i = 2$, $\omega_v^{v_1,v_2} = \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2}$, which illustrates that (21) is true.

(2) Suppose that, if $i = k$, (21) is true; that is,

$$\omega_v^{v_1,\ldots,v_k} = \omega_v^{v_1,\ldots,v_{k-1}}\overline{\oplus}\omega_v^{v_k} = \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2,\ldots,v_k}, \tag{22}$$

we need to prove that (21) is true while $i = k + 1$; that is,

$$\omega_v^{v_1,\ldots,v_{k+1}} = \omega_v^{v_1,\ldots,v_k}\overline{\oplus}\omega_v^{v_{k+1}} = \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2,\ldots,v_{k+1}}. \tag{23}$$

It is equivalent to

$$\begin{aligned}s_v^{v_1,\ldots,v_{k+1}} &= s_v^{v_1,\ldots,v_k}\overline{\oplus}s_v^{v_{k+1}} = s_v^{v_1}\overline{\oplus}s_v^{v_2,\ldots,v_{k+1}},\\ d_v^{v_1,\ldots,v_{k+1}} &= d_v^{v_1,\ldots,v_k}\overline{\oplus}d_v^{v_{k+1}} = d_v^{v_1}\overline{\oplus}d_v^{v_2,\ldots,v_{k+1}},\\ u_v^{v_1,\ldots,v_{k+1}} &= u_v^{v_1,\ldots,v_k}\overline{\oplus}u_v^{v_{k+1}} = u_v^{v_1}\overline{\oplus}u_v^{v_2,\ldots,v_{k+1}},\\ a_v^{v_1,\ldots,v_{k+1}} &= a_v^{v_1,\ldots,v_k}\overline{\oplus}a_v^{v_{k+1}} = a_v^{v_1}\overline{\oplus}a_v^{v_2,\ldots,v_{k+1}}.\end{aligned} \tag{24}$$

(i)

$$s_v^{v_1,\ldots,v_k}\overline{\oplus}s_v^{v_{k+1}} = \frac{d_v^{v_1,\ldots,v_k}u_v^{v_{k+1}} + d_v^{v_{k+1}}u_v^{v_1,\ldots,v_k}}{u_v^{v_1,\ldots,v_k} + u_v^{v_{k+1}} - u_v^{v_1,\ldots,v_k}u_v^{v_{k+1}}} \tag{25}$$

For $\omega_v^{v_1,\ldots,v_k} = \omega_v^{v_1,\ldots,v_{k-1}}\overline{\oplus}\omega_v^{v_k} = \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2,\ldots,v_k}$, we can get the following:

$$(25) \Longrightarrow \frac{d_v^{v_1}u_v^{v_2,\ldots,v_k}u_v^{v_{k+1}} + u_v^{v_1}s_v^{v_2,\ldots,v_k}u_v^{v_{k+1}} + u_v^{v_1}u_v^{v_2,\ldots,v_k}d_v^{v_{k+1}}}{u_v^{v_1}u_v^{v_2,\ldots,v_k} + u_v^{v_1}u_v^{v_{k+1}} + u_v^{v_2,\ldots,v_k}u_v^{v_{k+1}} - 2u_v^{v_1}u_v^{v_2,\ldots,v_k}u_v^{v_{k+1}}}, \tag{26}$$

(ii)

$$s_v^{v_1}\overline{\oplus}s_v^{v_2,\ldots,v_{k+1}} = \frac{d_v^{v_1}u_v^{v_2,\ldots,v_{k+1}} + d_v^{v_2,\ldots,v_{k+1}}u_v^{v_1}}{u_v^{v_1} + u_v^{v_2,\ldots,v_{k+1}} - u_v^{v_1}u_v^{v_2,\ldots,v_{k+1}}}, \tag{27}$$

$$(27) \Longrightarrow \frac{d_v^{v_1}u_v^{v_2,\ldots,v_k}u_v^{v_{k+1}} + u_v^{v_1}d_v^{v_2,\ldots,v_k}u_v^{v_{k+1}} + u_v^{v_1}u_v^{v_2,\ldots,v_k}d_v^{v_{k+1}}}{u_v^{v_1}u_v^{v_2,\ldots,v_k} + u_v^{v_1}u_v^{v_{k+1}} + u_v^{v_2,\ldots,v_k}u_v^{v_{k+1}} - 2u_v^{v_1}u_v^{v_2,\ldots,v_k}u_v^{v_{k+1}}}. \tag{28}$$

Equation (26) = (28); that is, $s_v^{v_1,\ldots,v_{k+1}} = s_v^{v_1,\ldots,v_k} \oplus s_v^{v_{k+1}} = s_v^{v_1} \oplus s_v^{v_2,\ldots,v_{k+1}}$.

It is easy to know that the others (d, u, and a) can be proved as above. So (21) is true while $i = k + 1$.

The above procedure illustrates that (21) is true while i is no less than 2 and no more than m. That is, the theorem is proved to be true as follows:

$$\omega_v^{v_1,\ldots,v_i} = \omega_v^{v_1,\ldots,v_{i-1}}\overline{\oplus}\omega_v^{v_i} = \omega_v^{v_1}\overline{\oplus}\omega_v^{v_2,\ldots,v_i} \quad (2 \le i \le m) \tag{29}$$

□

Given m neighbors $v_1, v_2, \ldots, v_m$ of node v, their opinions about the data of v are $\omega_v^{v_1} = (s_v^{v_1}, d_v^{v_1}, u_v^{v_1}, a_v^{v_1}), \ldots, \omega_v^{v_m} = (s_v^{v_m}, d_v^{v_m}, u_v^{v_m}, a_v^{v_m})$, the consensus of all the neighbors' opinions can be got through the computation of Theorem 4, then the expectation of consensus is $E(\omega_v) = s_v + a_v u_v$, where $s_v = s_v^{v_1,\ldots,v_m}, u_v = u_v^{v_1,\ldots,v_m}, a_v = a_v^{v_1,\ldots,v_m}$. The anomaly score of the node v's data is defined according to the expectation $E(\omega_v)$.

Definition 5. Suppose that the consensus of all the neighbors' opinions about node v's data is ω_v and the expectation of the consensus is $E(\omega_v)$, then the anomaly score of node v is defined as follows:

$$\text{AS}_v = E(\omega_v)\ . \tag{30}$$

Table 1: Notations used in the algorithms.

Notation	Description
m	Number of node v's neighbors
V_{neighbor}	Node v's neighbors set, $\{v_1, \ldots, v_m\}$
x	Suspicious sensing data of node v
X	Suspicious vector data of node v
r	Current round
D_{neighbor}	Sensing data of V_{neighbor} at round r, $\{x_1, \ldots, x_m\}$
VD_{neighbor}	Vector data of V_{neighbor} from $r - l + 1$ to r rounds, $\{X_1, \ldots, X_m\}$
X'	Historical vector data of node v
VD'_{neighbor}	Historical vector data of V_{neighbor}, $\{X'_1, \ldots, X'_m\}$
F_x	Indication of whether x is normal, faulty, or informational data, $F_x = 0 : x$ is normal; $F_x = 1 : x$ is faulty data; $F_x = 2 : x$ is informational data
AS_v	Anomaly score of node v
θ, thre	Predefine thresholds, discussed in Section 6
$\text{Corr}(x, x_i)$	Spatial correlation between x and x_i can be computed using extended Jaccard coefficient or correlation coefficient and so forth

There are some to be said. In the scenario that node v has one neighbor, Lemma 2 is not able to deal with it. To do with that, we suppose an imaginary neighbor who holds the opinion $\omega = (0, 0, 1, 0.5)$ and the neighbor takes part in the consensus with the real neighbor. Thus, we still get the consensus according to Lemma 2.

In the following sections, we present two algorithms to further determine whether suspicious data are normal or anomalous. The notations used to describe the algorithms are shown as in Table 1.

5.4. SLB Algorithm. The process of subjective logic-based algorithm (SLB) is as follows with discussion above. This process is executed among the node and its neighbors. Supposing node v has m neighbors $v_1, v_2, \ldots, v_m$. According to the suspicious data of node v whether it is changing smoothly or nonsmoothly, each neighbor node v_i gives the opinion $\omega_v^{v_i}$ about the data of node v (Line 1–10). Utilizing Theorem 4 to compute, we get the consensus opinion ω_v of all the neighbors of node v (Line 11). The expectation of consensus opinion is obtaind through the equation $E(\omega_v) = s_v + a_v u_v$ (Line 12). And then, the anomaly score AS_v can be get through Definition 5 (Line 13). If the anomaly score is less than the predefined threshold θ, the suspicious data of node v is considered as normal, or it is thought of as anomalous(Line 14–18) (Algorithm 1).

5.5. ESLB Algorithm. SLB algorithm fuses the opinions of all the neighbors about the data of the node to decide whether the data is normal or anomalous. However, it has the following disadvantages. (1) In the process of judgement among the node and its neighbors, the opinions of the neighbors whose data are suspicious are also included so as to affect the performance of anomaly detection. It is more severely affected especially when the proportion of anomalous data is ascending. (2) It does not distinguish the faulty data from the informational data. (3) The base rate a of all the neighbors' opinions is set to 0.5 which is not reasonable. It does not take the historical information of the node and its neighbors into account.

To overcome the disadvantages of SLB, we extend SLB to ESLB. For the first point, ESLB removes the opinions of those neighbors whose data are suspicious. To solve the second point, ESLB employ the correlations of anomalous data. If those data are spatial correlated, they are the informational data or else the faulty data. Thirdly, we define a as follows in considering the historical information.

Suppose that X' and X'_i are the latest l rounds of historical data of node v and neighbor v_i in the pre-processing phase, the historical opinion of neighbor v_i about node v's data is $\omega_v^{v_i\prime} = (s_v^{v_i\prime}, d_v^{v_i\prime}, u_v^{v_i\prime}, a_v^{v_i\prime})$. We set base rate $a_v^{v_i\prime}$ of historical opinion is 0.5; that is, $a_v^{v_i\prime} = 0.5$. Then we have the following definition.

Definition 6. Given the historical opinion of neighbor v_i about node v's data is $\omega_v^{v_i\prime}$, base rate a of current opinion $\omega_v^{v_i}$ of v_i about v's data is defined as follows:

$$a_v^{v_i} = E(\omega_v^{v_i\prime}). \tag{31}$$

Theorem 7. *Suppose that historical opinion of neighbor v_i about node v's data is $\omega_v^{v_i\prime}$, then base rate a of current opinion of v_i about v's data is $a_v^{v_i} = s_v^{v_i\prime} + 0.5u_v^{v_i\prime}$.*

Proof. From the definition of the expectation, we know that

$$E(\omega_v^{v_i\prime}) = s_v^{v_i\prime} + a_v^{v_i\prime} u_v^{v_i\prime}, \tag{32}$$

$$\left.\begin{array}{r} (31) \\ (33) \\ a_v^{v_i\prime} = 0.5 \end{array}\right\} \Longrightarrow a_v^{v_i} = s_v^{v_i\prime} + 0.5u_v^{v_i\prime}. \tag{33}$$

□

We extend SLB to ESLB algorithm as follows. If the data x is suspicious, node v turns to its neighbors set V_{neighbor} to

```
Input: $V_{\text{neighbor}}, x, X, D_{\text{neighbor}}, VD_{\text{neighbor}}$;
Output: $F_x$;
(1) if $r$ is at the time of data changing smoothly
(2)     for $1 \le i \le m$
(3)         compute the opinion $\omega_v^{v_i} = (s_v^{v_i}, d_v^{v_i}, u_v^{v_i}, a_v^{v_i})$ of neighbor $v_i$ about $v$ by use of (4)
(4)     end for
(5) end if
(6) if $r$ is at the time of data changing nonsmoothly
(7)     for $1 \le i \le m$
(8)         compute the opinion $\omega_v^{v_i} = (s_v^{v_i}, d_v^{v_i}, u_v^{v_i}, a_v^{v_i})$ of $v_i$ about $v$ by use of (5)
(9)     end for
(10) end if
(11) get the consensus opinion $\omega_v$ of all the neighbors $v_1, v_2, \ldots, v_m$ about node $v$
(12) compute the expectation $E(\omega_v)$ of the consensus opinion
(13) get the anomaly score $AS_v$ of node $v$
(14) if $AS_v \le \theta$
(15)     $x$ is normal data, $F_x = 0$;
(16) else
(17)     $x$ is anomalous data, $F_x = 1$ //here we do not distinguish faulty data from informational data
(18) end if
(19) return $F_x$;
```

Algorithm 1: Subjective logic-based (SLB) algorithm.

further determine (Line 1–3). If the data of some neighbors are suspicious, they do not provide their opinions about the suspicious data of node v. We exclude the neighbors from the candidate neighbors set V_{neighbor} and get the neighbors set $V_{\text{tneighbor}}$ which provides the opinions about the data of node v (Line 4–8). For each node in $V_{\text{tneighbor}}$, it computes its historical opinion $\omega_v^{v_k\prime}$ of neighbor v_k about v's data by use of X' and X_k', and $a_v^{v_k\prime}$ is set to 0.5 (Line 11). We compute the current opinion $\omega_v^{v_i}$ according to SLB algorithm excluding $a_v^{v_i}$ which is computed through Theorem 7 (Line 12). Then we get the result whether x is normal according to calling SLB algorithm (Line 15). If x is not normal, it sends message M_x to sink, which includes node v, current round l, data x, and flag F_x (Line 21). Sink receives all the messages at round r and further analyzes neighbors who hold on faulty data at this round. If x and x_i are all faulty at the same time and are spatial correlated, they are informational data or else faulty data (Line 24–32) (Algorithm 2).

6. Discussion

There are some problems to be explained further. First, authors in [8, 9] point out that voting algorithms cannot deal with the situation, in which the events are detected by sensor nodes which are not adjacent. However, our framework can do with the situation after minor revision. For example, suppose that node v_i and v_j are not within the radio range of each other and they detect the same event at some time. Suppose that the impact range of events is IR, radio range is $CR, h = \lceil IR/CR \rceil$. Our framework can still detect the event by computing the spatial correlation among h-hop neighbors. For the computation is executed on sink, it does not increase the energy consumption.

Second, in order to reduce the energy consumption, we use the idea proposed by [13] to construct and maintain AR models. (1) It avoids unnecessary data transmission. While the data of nodes are normal, it does not transmit data in the network but estimates the data according to AR models by sink. (2) It reduces the computational complexity of constructing and maintaining AR models. The main computation is executed on sink and not sensor nodes. Please refer to [13] for more detail.

Third, although the thresholds, like θ and thre, are vital to SLAD, we do not pay much attention to them. We focus on how to more realistically quantize the opinion of the neighbors to special sensor node. In this paper, we set them with the historical experience. However, excellent methods are not excluded to improve SLAD further.

7. Simulation Results

7.1. Experimental Setup. We implement our simulation experiments in OMNET++ platform [20]. The topology and the sensing data come from Intel Berkeley research lab data set [21]. 54 sensors are deployed in the Lab of 400 ∗ 700, and the locations of sensor nodes are known in advance. In the experiments of Section 7.2, radio range is set to 150. Section 7.3 shows the impact of different radio ranges on the detection performance. All the experiments suppose that the radio links are reliable and do not fail. The sensing data have four attributes, yet only temperature is selected in our experiments. We use 1000 rounds of data as experimental data, and use the initial 100 rounds of data to construct AR models.

While using AR(p) models to predicate the sensing data in WSNs, AR (3) model can get good estimation and low cost of maintenance [13, 14]. So we use AR(3) as the models constructed on the nodes. If p is set to 3, AR models can express as $X_t = \varphi X_{t-1} + \varphi X_{t-2} + \varphi X_{t-3} + \varepsilon$. In the beginning, we use the first 100(N) rounds as the training data, among

```
Input: $V_{\text{neighbor}}, x, X, D_{\text{neighbor}}, VD_{\text{neighbor}}, X', VD'_{\text{neighbor}}$;
Output: $F_x(, F_{x_i})$;
for each node $v$
    if $x$ is suspicious data
        node $v$ turns to its neighbors $V_{\text{neighbor}}$ to further determine
        for $1 \le i \le m$
            if $x_i$ is suspicious data
                $v_i$ does not provide its opinion to node $v$, $V_{\text{tneighbor}} = V_{\text{neighbor}} - \{v_i\}$
            end if
        end for
        for $1 \le k \le m$
            if $v_k \in V_{\text{tneighbor}}$
                compute historical opinion $\omega_v^{v_k}{}'$ of $v_k$ about $v$ by use of $X$ and $X'_k$, $a_v^{v_k}{}' = 0.5$
                call SLB Algorithm (Line 1–10) to compute current opinion $\omega_v^{v_k}$ excluding $a_v^{v_k}$, and $a_v^{v_k} = s_v^{v_k}{}' + a_v^{v_k}{}' u_v^{v_k}{}'$
            end if
        end for
        call SLB Algorithm (Line 11–18) to get the result whether $x$ is normal
        if $x$ is normal
            $F_x = 0$
        else
            $F_x = 1$
        end if
        send message $M_x$ to sink, $M_x = \{v, r, x, F_x\}$
    end if
end for
sink receives all the messages at round $r$, and analyzes neighbors holding on faulty data at this round
//following executes on sink node
if $x$ and $x_i$ are faulty at the same time
    if $\text{Corr}(x, x_i) > \text{thre}$
        $x$ and $x_i$ are informational data, $F_x = 2, F_{x_i} = 2$
    else
        $x$ and $x_i$ are faulty data, $F_x = 1, F_{x_i} = 1$
    end if
end if
return $F_x, F_{x_i}$;
```

Algorithm 2: Extended subjective logic-based (ESLB) algorithm.

which the first 90 (M) rounds of data are used to estimate the coefficients of AR model and the last 10 (N-M) rounds of data are used to determine the threshold τ.

If the sensing data are changing nonsmoothly, we would use the vector data to compute neighbors' opinions. To compute the base rate of neighbors to the node (historical information), it also needs to utilize the vector data. So, it needs to select the appropriate length of vector data (l). If l is set too small, it cannot express the data trends. Otherwise, it consumes too much energy to exchange sensing data. Figure 1 shows the detection rate of SLAD framework under the condition of different lengths of vector data. While l is not more than 5, the detection rate increases obviously with the increase of l. Once l achieves 5, the detection rate varies not obviously with the increase of l. Consequently, we set the length of vector data (l) to 5.

We randomly change some of normal data as faulty data and define the faulty rate as the proportion of faulty data to the whole data. In the experiments, we compare the performance of different algorithms at various faulty rate, and the results are mean of 20 times of executions.

7.2. Comparison of Detection Performance. In order to compare the anomaly detection performance of different algorithms, we define detection rate, false detection rate, and undetection rate. Among these definitions, the whole experimental data set is denoted as W_D, the real faulty data set is expressed as F_D, and the identified faulty data set which is determined by anomaly detection algorithms is marked as I_D.

Definition 8 (detection rate). It is defined as the faulty data which are determined as faulty in the proportion of the real faulty data:

$$\text{Detection\_rate} = \frac{|F_D \bigcap I_D|}{|F_D|}. \tag{34}$$

Definition 9 (false detection rate). It is defined as those normal data which are determined as faulty in the proportion of the real faulty data:

$$\text{FalseDetection\_rate} = \frac{|(W_D - F_D) \bigcap I_D|}{|F_D|}. \tag{35}$$

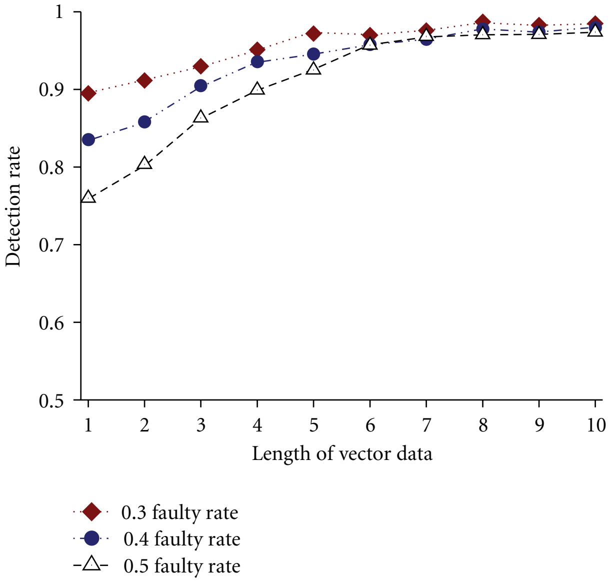


Figure 1: Impact of length on detection rate.

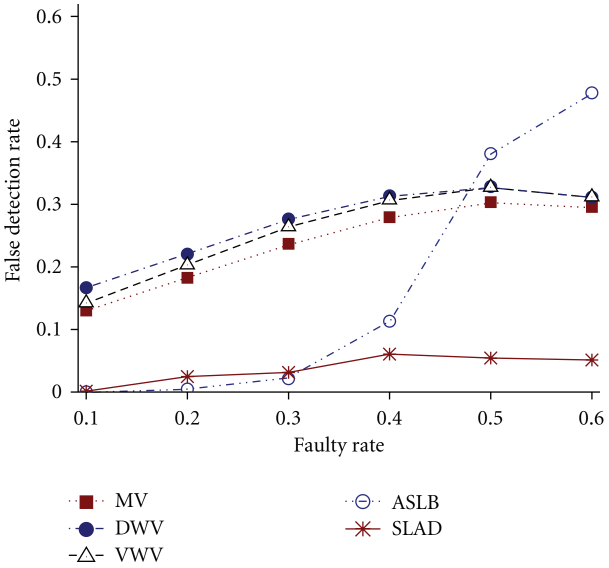


Figure 3: False detection rate of different algorithms.

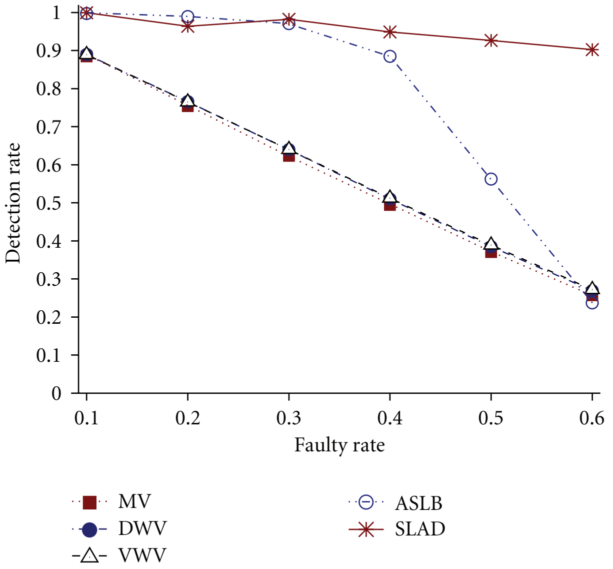


Figure 2: Detection rate of different algorithms.

Definition 10 (undetection rate). It is defined as those faulty data which are determined as normal in the proportion of the real faulty data:

$$\text{UnDetection\_rate} = \frac{|F_D - (F_D \bigcap I_D)|}{|F_D|}. \quad (36)$$

In this section, we compare the performance of different algorithms. These algorithms are listed as follows. (1) MV (majority voting algorithm) [6]. (2) DWV (distance weight voting algorithm) [7]: it use the Euclidean distance of sensor nodes as the weight, and the weight is smaller with the distance being farther. Please refer to Section 2 about the details of MV and DWV algorithms. (3) VWV(value weight voting algorithm): it is different from DWV, and it uses the distance of the data of node and its neighbors as the weight, that is, the difference of the data. It considers that the neighbors whose data are closer to that of the node should have greater weights. (4) ASLB(autoregressive model and SLB): it combines autoregressive models with subjective logic-based algorithm (SLB algorithm). (5) SLAD (subjective logic-based anomaly detection framework): it integrates autoregressive model and extended subjective logic-based algorithm(ESLB algorithm).

Figure 2 shows the detection rate of five algorithms at different faulty rate. It indicates that detection rates of all the algorithms are greater than 0.8 when faulty rate is low. The performances of ASLB and SLAD are better than MV, DWV, and VWV. The detection rates of MV, DWV, and VWV decrease sharply as faulty rate increases. ASLB keeps the high detection rate when faulty rate is less than 0.4, which decreases sharply once faulty rate reaches 0.4 and holds this trend with the increase of faulty rate. However, the detection rate of SLAD is still greater than 0.9 even though faulty rate increases, which shows the best performance compared with the other algorithms.

Figure 3 presents the detailed comparison results of these algorithms at different faulty rate. The false detection rate of all the algorithms increases as faulty rate becomes larger. The false detection rate of MV, DWV, and VWV keeps in some specified scope as faulty rate increases, and ASLB increases suddenly once faulty rate achieves 0.4. SLAD holds the false detection rate within limits which is no greater than 0.1. The false detection rate of SLAD is much less than the others.

We then study the impact of different faulty rate on undetection rate of these algorithms. The undetection rate of MV, DWV, and VWV decreases as faulty rate increases. The undetection rate of ASLB increases abruptly while faulty rate achieves 0.4, and it keeps the rising trend with the increase of faulty rate. SLAD preserves very low undetection rate which

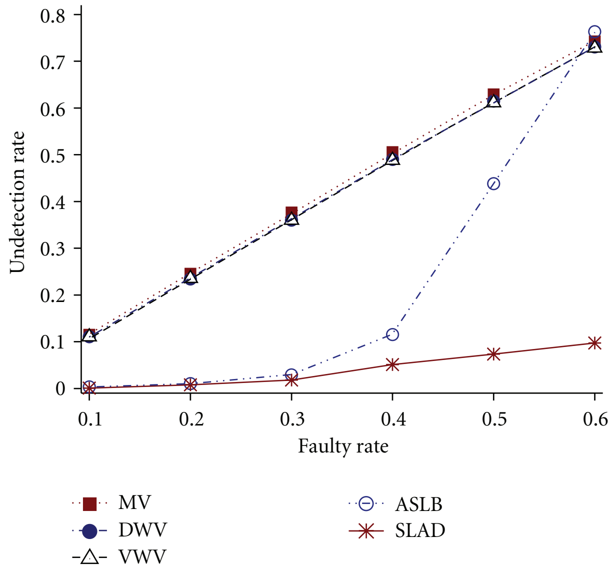


Figure 4: Undetection rate of different algorithms.

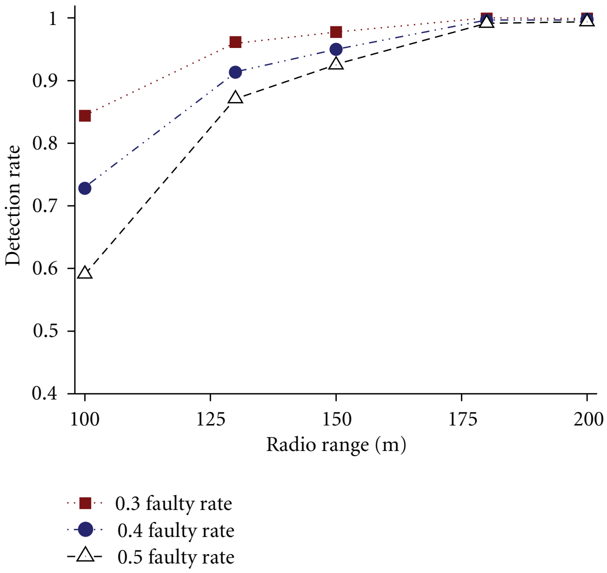


Figure 5: Impact of radio range on detection rate.

does not exceed 0.1 even though faulty rate is high. The undetection rate of SLAD is much less than other algorithms though it increases as faulty rate increases.

From the above figures, we note that ASLB suddenly changes its trends of detection performance when faulty rate is 0.4. The reason is presented as follows. When faulty rate is 0.4, the number of neighbors whose sensing data are right is more than those data being anomalous on average. It results that the detection performance does not decline too much. However, once faulty rate is more than 0.4, the number of neighbors whose data are faulty is no less than that whose data are normal. It is hard to decide whether the suspicious data is normal for ASLB, and it results to the poor detection performance significantly.

We also draw the following conclusion according to Figures 2, 3, and 4. The overall performance of SLAD is much better than the other algorithms, and the performance of ASLB is better than MV, DWV, and VWV when faulty rate is low. The cause is the combination of subjective logic. Using subjective logic, ASLB and SLAD fuses the quantitative opinions of neighbors which avoid the problems other algorithms are facing. Because MV, DVW, VWV, and ASLB use the opinions of all the neighbors, the number of faulty data of neighbors may be rising along with faulty rate increasing, which shows the bad impact on the detection rate, false detection rate, and undetection rate. However, SLAD has removed the opinions of the neighbors whose data are suspicious before providing their opinions and takes historical spatial correlations of the nodes and their neighbors into account. So, SLAD holds significantly superior performance than other algorithms, especially when faulty rate is high.

The above experiments discuss the cases that the network are only involving the faulty data, and not including the informational data. In the monitoring area, some events randomly arise. The anomalous data of sensor nodes detecting the events are spatial correlations (i.e., informational data). The faulty rate, which is defined as the number of informational data in proportion of the whole data, is set to 0.2. The experiment shows that detection rate of SLAD framework for informational data reaches more than 0.9, and MV, DWV, VWV are only about 0.7. The reason is that SLAD framework utilizes subjective logic to fuse the quantitative opinions of neighbors so as to improve the detection performance obviously.

7.3. Impact of Radio Range on Detection Performance. In this section, we analyze the impact of radio range on detection rate, false detection rate, and undetection rate at different faulty rate. The number of neighbors affects the detection performance of the algorithm. Different radio range of the nodes leads to different number of neighbors. Thereby, we discuss the detection performance of SLAD framework under the condition of different radio ranges.

We conduct the experiments to compare the detection performance of SLAD framework under different radio ranges. We set faulty rate to 0.3, 0.4, and 0.5 in the experiments. Figures 5, 6, and 7 show the detection rate, false detection rate, and undetection rate of SLAD, respectively. These figures indicate that detection rate decreases; false detection rate and undetection rate increase with the increase of faulty rate. They also show that detection rate increases; false detection rate and undetection rate decrease as the radio range becomes larger. The reason is that there are more neighbors providing the opinions with the radio range increasing.

8. Conclusions

In this paper, we present SLAD framework which considers the uncertainty of neighbors to the data of the node. It includes three phases: pre-processing, self-monitoring, and cooperant detecting. In the first phase, sink constructs AR

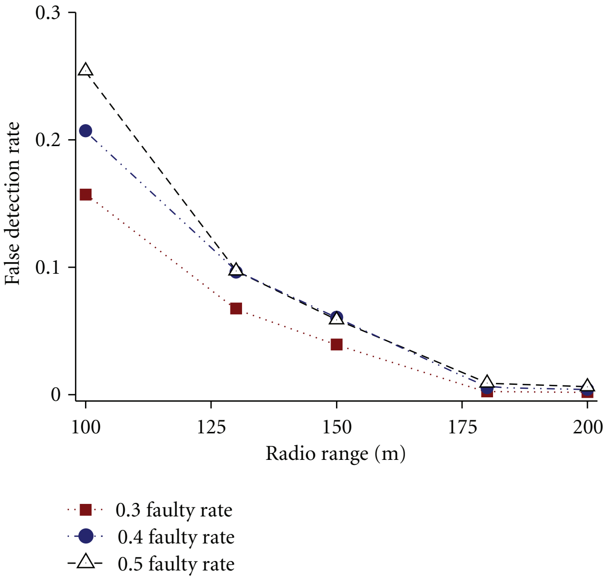


Figure 6: Impact of radio range on false detection rate.

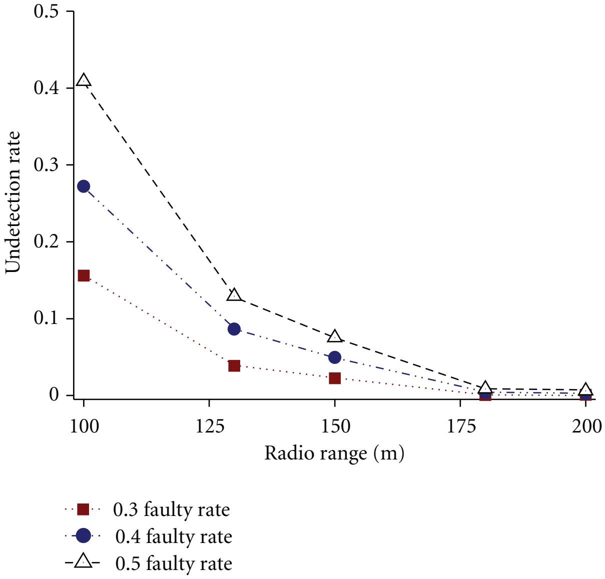


Figure 7: Impact of radio range on undetection rate.

model for each node. In the second phase, it uses AR models to check whether the sensing data are suspicious. In the third phase, it presents two novel algorithms SLB and ESLB. The third phase is the key of our framework. In SLB, each neighbor gives the quantitative opinion to the suspicious data involving with subjective logic theory. After fusing the opinions of all the neighbors, SLB gets the expectation of the consensus opinion and anomaly score, which demonstrates the degree of the suspicious data being considered as an anomaly. We extend SLB to ESLB in order to avoid the impact of those neighbors whose data are suspicious, effectively distinguish the faulty data from the informational data, and take the historical spatial correlations of the node and its neighbors into account. Simulation results show that SLAD framework improves the performance of anomaly detection effectively compared with previous works.

However, we find there is something to do for further improving SLAD. We believe that the opinion of the neighbor, who holds the higher historical spatial correlation with the node, should be paid more attention to. An example is given to demonstrate that. Suppose node *A* and node *B* are the neighbors of node *C* and node *A* and node *C* are located in the room while node *B* is out of the room. Generally, the historical spatial correlation between node *A* and node *C* is higher than that between node *B* and node *C*. Thus, the opinion of node *A* to node *C* should be given more attention. Unfortunately, the subjective logic, which works as the foundation of SLAD, treats the opinions equally and has no capability to deal with it. As the preparatory work, we proposed an operator for subjective logic which is capable of making the consensus on several neighbors' opinions with their weights in a fair way [22]. With the support of the new operator, we can map the historical spatial correlation to the weight of the opinion to improve SLAD. In theory, we believe it will improve the performance of anomaly detection for SLAD. It is our future work.

Acknowledgments

This work is supported by the National Science Foundation (61070056, 61033010), the National 863 High-tech Plan (2008AA01Z120), Program for New Century Excellent Talents in University, and the Research Funds of the Renmin University of China (10XNI018).

References

[1] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next century challenges: mobile networking for "smart dust"," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, pp. 271–278, Seattle, Wash, USA, August 1999.

[2] D. Culler, D. Estrin, and M. Sivastava, "Overview of sensor networks," *Computer*, vol. 37, pp. 41–49, 2004.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[4] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.

[5] S. Jeffery, G. Alonso, M. J. Franklin, W. Hong, and J. Widom, "Declarative support for sensor data cleaning," in *Proceedings of the 4th International Conference on Pervasive Computing*, pp. 83–100, Dublin, Ireland, May 2006.

[6] B. Krishnamachari and S. Iyengar, "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks," *IEEE Transactions on Computers*, vol. 53, no. 3, pp. 241–250, 2004.

[7] M. Krasniewski, P. Varadharajan, B. Rabeler, S. Bagchi, and Y. C. Hu, "TIBFIT: trust index based fault tolerance for arbitrary data faults in sensor networks," in *Proceedings of the International Conference on Dependable Systems and Networks*, pp. 672–681, Yokohama, Japan, July 2005.

[8] Y. Kotidis, A. Deligiannakis, and V. Stoumpos, "Robust management of outliers in sensor network aggregate queries," in

Proceedings of 6th International ACM Workshop on Data Engineering for Wireless and Mobile Access, pp. 17–24, Beijing, China, June 2007.

[9] A. Deligiannakis, Y. Kotidis, V. Vassalos, V. Stoumpos, and A. Delis, "Another outlier bites the dust: computing meaningful aggregates in sensor networks," in *Proceedings of the 25th IEEE International Conference on Data Engineering (ICDE '09)*, pp. 988–999, Shanghai, China, April 2009.

[10] N. Giatrakos, Y. Kotidis, A. Deligiannakis, V. Vassalos, and Y. Theodoridis, "TACO: tunable approximate computation of outliers in wireless sensor networks," in *Proceedings of the International Conference on Management of Data (SIGMOD '10)*, pp. 279–290, Indianapolis, Ind, USA, June 2010.

[11] X. Y. Xiao, W. C. Peng, C. C. Hung, and W. C. Lee, "Using sensor ranks for in-network detection of faulty readings in wireless sensor networks," in *Proceedings of the 6th International ACMWorkshop on Data Engineering for Wireless and Mobile Access*, pp. 1–8, Beijing, China, June 2007.

[12] N. Giatrakos, Y. Kotidis, and A. Deligiannakis, "PAO: power-efficient attibution of outliers in wireless sensor networks," in *Proceedings of the 7th International Workshop on Data Management for Sensor Networks*, pp. 33–38, Singapore, September 2010.

[13] D. Tulone and S. Madden, "PAQ: time series forecasting for approximate query answering in sensor networks," in *Proceedings of the European Conference on Wireless Sensor Networks*, pp. 21–37, Zurich, Switzerland, February 2006.

[14] D. Tulone, "A resource—efficient time estimation for wireless sensor networks," in *Proceedings of the Joint Workshop on Foundations of Mobile Computing (DIALM-POMC '04)*, pp. 52–59, Philadelphia, Pa, USA, October 2004.

[15] A. Jøsang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279–311, 2001.

[16] A. Josang, "Fission of opinions in subjective logic," in *Proceedings of the 12th International Conference on Information Fusion*, pp. 1911–1918, Seattle, Wash, USA, July 2009.

[17] O. Obst, X. R. Wang, and M. Prokopenko, "Using echo state networks for anomaly detection in underground coal mines," in *Proceedings of the ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN '08)*, pp. 219–229, St. Louis, Mo, USA, April 2008.

[18] M. Wälchli, "Efficient signal processing and anomaly detection in wireless sensor networks," in *Proceedings of the EvoWorkshops on Applications of Evolutionary Computing: EvoCOMNET, EvoENVIRONMENT, EvoFIN, EvoGAMES, EvoHOT, EvoIASP, EvoINTERACTION, EvOmUSART, EvoNUM, EvoSTOC, EvoTRANSLOG*, pp. 81–86, Tübingen, Germany, April 2009.

[19] M. Chang, A. Terzis, and P. Bonnet, "Mote-based online anomaly detection using echo state networks," in *Proceedings of the 5th IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 72–86, Marina Del Rey, Calif, USA, June 2009.

[20] A. Varga, "The OMNET++ discrete event simulation system," in *Proceedings of the European Simulation Multiconference*, pp. 319–324, Prague, Czech, June 2001.

[21] Intel Berkeley Research Lab, http://berkeley.intel-research.net/labdata/.

[22] H. Zhou, W. Shi, Z. Liang, and B. Liang, "Using new fusion operations to improve trust expressiveness of subjective logic," *Wuhan University Journal of Natural Sciences*, vol. 16, no. 5, pp. 376–382, 2011.

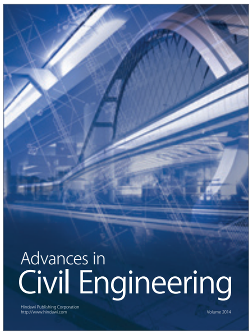
Journal of
Engineering

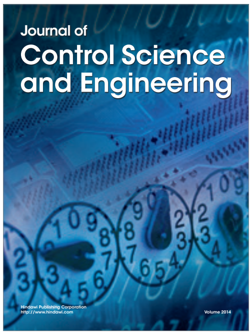
The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

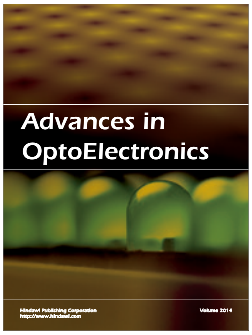
International Journal of
Distributed
Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science
and Engineering

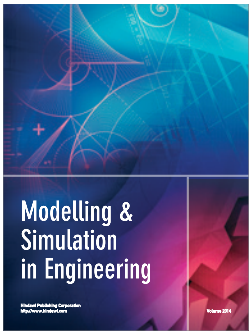
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

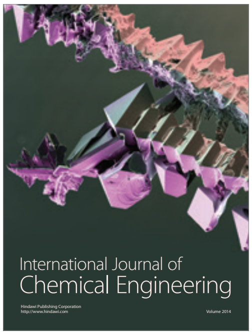

Advances in
OptoElectronics

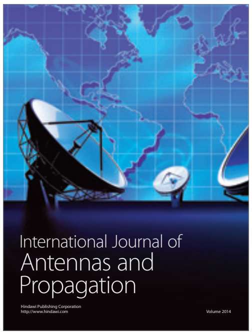
VLSI Design

International Journal of
Navigation and
Observation

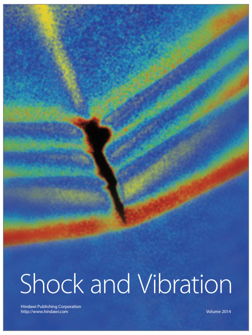

Modelling &
Simulation
in Engineering

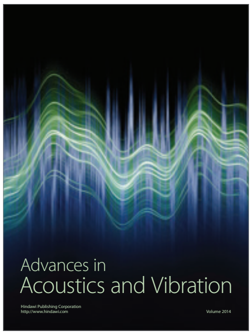
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration