An Improved Distributed Fault Diagnosis Algorithm in Wireless Sensor Networks

Lin Chang, Zhi-Qing Huang

Faculty of Information Technology, Beijing University of Technology

Beijing Engineering Research Center for IoT Software and Systems, Beijing, China

E-mail: clyby@emails.bjut.edu.cn

Yan-Xin Zhang

College of Electronic Information Engineering, Beijing Jiao tong University, Beijing, China

E-mail: yxzhang@bjtu.edu.cn

Abstract-In wireless sensor networks (WSNs), nodes fault diagnosis is the important measure for the continuous monitoring service. Efficient and accurate methods for diagnose nodes fault is a hot research topic in the field of WSNs in the recent years. This paper propose an improved distributed fault diagnosis method namely Cross-Space Distributed Fault Diagnosis (CS-DFD) which follows the DFD algorithm, building high-dimensional vector space model through the collected information from each node, establishing the cross type of sliding window through its historical data and the data from neighbor nodes, setting the cross direction customizable weights of fault. Ultimately, achieving the goal of fault diagnosis by detecting abnormal vector and threshold. The experimental results show that it has reduced the amount of calculation; simplified harsh conditions of fault judgments; reduced the consumption of electricity, etc. which in DFD algorithm. Comparing with the traditional DFD algorithm, the fault diagnosis accuracy improved by 5.14%, the fault of false alarm ratio (FAR) decreased by 2.01% and the time of diagnosis has shortened obviously.

Keywords-wireless sensor networks (wsns); vector space model; cross-sliding window; node fault diagnosis

I. INTRODUCTION

With the development of sensor hardware technology and wireless sensor networks (WSNs), particularly in Micro-Electro-Mechanical System has facilitated the development of smart sensors (e.g., Mica motes from Crossbow, Tmote Sky from Moteiv , the MKII nodes from UCLA, etc.). Smart sensors nodes are low power devices subject to tighten the communication, storage and computation constraints. WSNs have the potential to enable a substantial class of applications like battle field surveillance, environmental monitoring, intruder detection systems, intelligent infrastructure monitoring and scientific data collection [1].

Sensor nodes are expected to operate autonomously in unattended and hostile environments for applications with short mission time and applications that last for months to years. For example, the forest surveillance project Green Orbs with more than 330 sensor nodes in the wild [2, 3]. Large-scale deployment of low-cost sensor nodes in uncontrolled, harsh or hostile environments is the inherent property of WSNs. It is common for the sensor nodes to become faulty and unreliable. Faults may range from simple crash faults where a node becomes completely inactive to faults where the node behaves arbitrarily or maliciously [4].

The erroneous outputs from faulty sensors might result in wrong interpretation or undesirable alarms. Therefore, node faults diagnosis in time, is the key to successful applications of the wireless sensor networks [5].

The major contributions of this work are follows:

- We study the problem of nodes fault diagnosis and find some characteristics in the nodes faults in WSNs.
- We put forward the distributed fault diagnosis method which based on Vector Space Model; sliding window; the judgment through history data and neighbor nodes follows the DFD algorithm.
- Based on the idea of sliding window, providing a strategy which can calculate the sum of the weight of arbitrary directions in similarity through the cross sliding window.
- Through comparing the method in this paper with the DFD algorithm, which are both belong to the distributed approaches in three aspects of the diagnosis of false alarm ratio, the accuracy of fault diagnosis and the time of diagnosis.

II. RELATED WORK

Usually, the methods of node fault diagnosis in WSNs would be divided into two categories: centralized approaches and distributed approaches.

In distributed approaches is like many other approaches the neighbor coordination approaches, cluster-based-approaches, soft computing approaches. Every sensor node decides independently the state of the network. This allows the diagnostic framework to scale easily to much larger and denser sensor networks.

Kung et al. [6] exploit the PMC model to identify faulty sensor nodes in WSNs. They suggest that it is much more feasible to manage and update the topology of some substructure of a WSNs rather than the whole network. Weber et al. [8] propose a strategy of mutual tests among the sensor nodes in a region where t numbers of faulty sensor nodes are present such that the system graph representing the region of the WSNs is t-diagnosable. Weber et al. presented a diagnosis approach namely energy-efficient test assignment without reciprocal tests [7] which are based on their previous work [8].Chen et al. [9] propose

and evaluate a localized fault detection algorithm named Distributed Fault Detection to identify the faulty sensors. Mahapatro and Khilar [10] propose an On-line distributed fault diagnosis method called cluster-based distributed fault diagnosis algorithm. Miao et al. [11] present an online lightweight failure detection method namely Agnostic Diagnosis. This approach is motivated by the fact that the system metrics of sensors usually exhibit certain correlation patterns. Mahapatro and Khilar [12] propose a hybrid approach which exploits both the sensor node coordination and sensor node self-detection. In [13], providing Local-Diagnosis (LD2), which conducts the diagnosis process in a local area and achieves diagnosis decision through distributed evidence fusion operations. Xi-Liang et al. [14] present an approach for detecting and identifying sensor node failure based on Principal Component Analysis, and wavelet decomposition. Salton et al. [15, 16] put forward Vector Space Model in the 1970 s, and successfully applied to the famous SMART text retrieval system.

In centralized approaches, the central node periodically sends diagnostic queries into the network to obtain the state of the individual sensor nodes in WSNs. After analyzing the diagnostic response messages it takes a decision about failed or suspicious sensor nodes. Yunhao et al. [17] propose PAD which is a passive diagnosis approach. In this approach, a probabilistic inference model is suggested. Ruiz et al. [18] propose a failure detection method using management architecture for WSNs, called MANNA. Ramanathan et al. [19] propose Sympathy, a centralized fault detection method, using a message-flooding approach to pool general system metrics such as sensor nodes' next-hops and neighbors.

Different practical applications may require the fault diagnosis which are conducted with low latency, the amount of calculation, high throughput. Distributed approaches address these issues and limitations. In techniques like centralized approaches have not good at this issues.

We have absorbed the experience over the papers and discovered some large amount of computing, large consumption of energy, the conditions of fault judgment which are harsh, this paper puts forward the distributed fault diagnosis method closely follows the DFD algorithm.

III. IMPROVED DISTRIBUTED FAULT DIAGNOSIS

In this part, mainly includes three stages: data preprocessing; establishing the vector space model and the cross sliding window in the diagnosis of stage.

A. Data Preprocessing

In the application of WSNs, usually by a combination of multiple sensor nodes perception for monitoring the environment , because of the configuration with the device itself and the influence of the environment, the electromagnetic wave, will produce the noise signal, as the following:

$$S(x) = \begin{cases} s(i) = f(i) + g(i) + e(i) \\ s(i+1) = f(i+1) + g(i+1) + e(i+1) \\ \quad\quad\quad \vdots \\ s(j-1) = f(j-1) + g(j-1) + e(j-1) \\ s(j) = f(j) + g(j) + e(j) \end{cases} \quad (1)$$

$S(x)$ is the signal with noise, $f(x)$ is the real signal, $g(x)$ is the signal with failure and $e(x)$ is the noise.

The wavelet transform de-noising effect is better than others, can focus on any minute details of the signal and remove noise from the signal , it is defined as:

$$W_{x(a,b)} = \int_{-\infty}^{+\infty} \psi_{a,b}(t)dt \quad (2)$$

$\psi_{a,b}(t)$ is the continuous wavelet basis function, a is the scaling factor and b is the parallel factor.

Encountered in the process of removing noise, dramatic change data, excessive decomposition may peak data changes as the signal noise elimination, so need to comprehensive consideration to remove the noise of the decomposition process.

$$S = A_N + \sum_{i=1}^{N} D_i \quad (3)$$

A_N as original signal, S as the signal after processing and N is the times of decomposition, D_i is the signal with noise.

B. Vector Space Model

In the general environment can use Euclidean distance or cosine similarity determines two individuals are similar, but the process of fault diagnosis:

Euclidean distance has disadvantages: if the dimensional vector in one dimensional numerical change can affect the location of the space mapping, indirect influence the distance, individual differences is not obvious.

Cosine similarity has disadvantages: if the dimensional vector in one dimensional numerical changes, the changes will affect the direction of the vector.

Therefore, introducing the fault contribution as the decision basis of fault diagnosis.

Contribution of faults: two inverse cosine similarity value of the individual and the value of a percentage of the product of Euclidean distance. The greater of λ, the greater chance of failure.

$$\lambda = \left| \frac{\|S_i\| \times \|S_j\|}{\overline{S_i} \, \overline{S_j}} \right| \times \sqrt{\sum_{i,j=1}^{n} \left(S_i - S_j \right)^2} \times 100\% \quad (4)$$

Both S_i and S_j represent different perception of the nodes to construct vector space model respectively, i and j is positive integer.

In WSNs, the node would be perceived by different sensing module, and for a separate node, can be abstracted as a three-dimensional space vector model:

$$S = S(s_1, s_2, s_3, \ldots, s_n) \tag{5}$$

Equation (5), n represent different kinds of module. Illustrate the vector space model, as shown in Fig. 1:
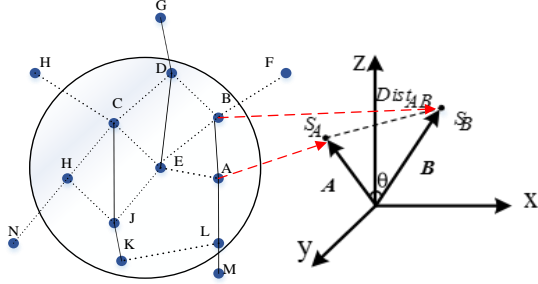

Figure 1.   Vector space model.

As shown in Fig 1, the perception of nodes deployed in monitoring are S_A and S_B , through the establishment of vector space model, its sensory information mapped to high-dimensional space of two points, vector A and vector B are map respectively corresponding to space vector.

Through the establishment of vector space model, has overcome the traditional fault diagnosis methods of large amount of calculation, the reason is that through the establishment of the vector space model, the node's information mapped to a certain point in the space, different perception of the node in space will be mapped to different points, the fault decision fault contribution by calculation.

C.   Cross Sliding Window

Sliding window is a kind of flow control technology in [27].The cross window has based on sliding window, increase the lateral sliding window which is a good way to overcome shortages like relying on neighbor node's data and their own historical data overly, so we introduce the cross sliding window for nodes fault diagnosis.

Learning from the information in monitoring area in WSNs, generally the information from sensor nodes have the dimensional and temporal correlation. If use the longitudinal sliding window on a particular node, fault judgments according to the historical data, if the data is a fixed value, easy to be mistaken for normal nodes; if the lateral sliding window to similarity matching of nodes and neighbor nodes, determine fault or not, its drawbacks are that when a node and its neighbor nodes is completely similar, can't accurately determine whether the node is normal.

Therefore, it has formed a model through combining horizontal and vertical sliding window, called a cross sliding window model, can overcome the following disadvantages: traditional diagnosis methods have relied on neighbor node data overly, its own historical data and the large data sets.

This article uses history data itself as a vertical cross sliding window and the data from neighbor nodes as the horizontal window of transverse sliding window, establish the cross sliding window diagram is shown in Fig. 2:
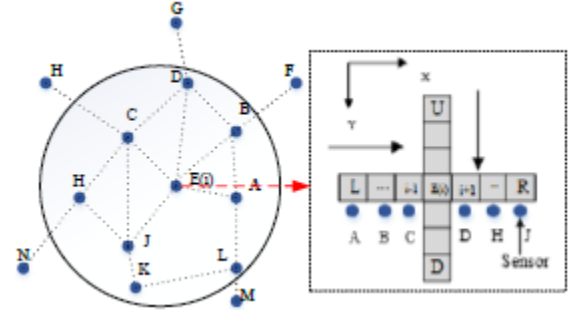

Figure 2.   Sliding window model.

By figure 2, point with the blue color presents sensor node, node i as the center, respectively, the left and the right of the neighbor node sequence is L and R, the lateral size of the window to neighbor nodes is U and D, and the longitudinal size of the window to nodes i in the number of historical data from moment t_1 to moment t_2.

The cross sliding window moving order: taking a perceptual node i as the center, in the form of queue, scan from left to right horizontally, from top to bottom longitudinally, delete the original vector from the original queue and insert a new vector. After the operation of the team and completing queue to update the vector updates at the same time.

On the four directions of the cross sliding window, the custom fault weight respectively is 25%, as is shown in the equation of (6):

$$Weight_{fault} = \begin{cases} 25\%; Dir_L \\ 25\%; Dir_R \\ 25\%; Dir_U \\ 25\%; Dir_D \end{cases} \tag{6}$$

According the fault weights on the four directions, accumulate the contribution ratio of fault in all directions values, if more than threshold, shows that the center vector of cross sliding window is unusual.

A test result that $Res_{i,j} \in \{1,0\}$ is generated by sensor S_i based on its neighbor S_i 's measurement using predefined threshold values θ_1 and θ_2 .

For any nodes $S_j \in N(S_i)$, the node S_i first set $Res_{i,j} \in \{1,0\}$ to 0. This algorithm next calculates λ and $Weight_{fault}$. The comparison test result $Res_{i,j}$ is set to 1 if $\lambda > \theta_1$ and $Weight_{fault} > \theta_2$.

If $Res_{i,j}$ is 0, most likely either both S_i and S_j are good or both are faulty. Otherwise, if $Res_{i,j}$ is 1, S_i and S_j are most likely in different status.

For any sensor node S_i its test results with each sensor node in the neighbor set $N(S_i)$ is obtained. If there are more than $N(S_i)/2$ of sensor nodes whose comparison test results

are 1 in $N(S_i)$, then initial detection status of sensor node S_i is possibly faulty (*AN*), otherwise, it may be possibly good(*CN*), i.e.,

$$Current_{status} = \begin{cases} AN, otherwise \\ CN, if \left\lceil \dfrac{N(S_i)}{2} \right\rceil \leq \sum_{S_i \in N(S_i)} Res_{i,j} \end{cases} \quad (7)$$

Where $N(S_i)$ represents the number of one-hop neighbors of S_i . The number of Abnormal nodes whose test with is 1 is subtracted from the number of Abnormal nodes whose test result with S_i is 0.

That is $\forall$ $S_j \in N(S_i)$ and $Current_{status} = AN$, $\sum(1\text{-}2 \ Res_{i,j})$ must be greater or equal to $[N(S_i)/2]$ to detect S_i as fault-free. This can be defined as

$$Fin_{status} = \begin{cases} FN, otherwise \\ FY, if \sum_{S_i \in N(S_i) and Cur_{status} = CN} (1 - 2Res_{i,j}) \geq \left\lceil \dfrac{N(S_i)}{2} \right\rceil \end{cases} \quad (8)$$

A sensor node S_i that has failed to pass the threshold test of the equation (8) is marked as *FY*.

IV. EXPERIMENTS DESIGN

A. *The experimental data and environment*

In experiment, using the perception of source data set from the Intel research Berkeley lab 54 Mica2Dot sensor nodes deployed in the laboratory will different rooms and different position, acquisition of sensory data [21]. Laboratory deployment nodes distribution is shown in Fig. 3:
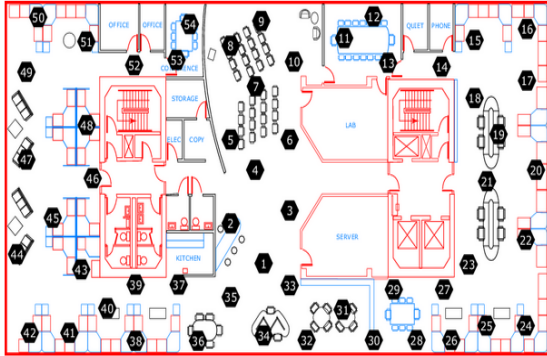

Figure 3. Laboratory profile node deployment.

The node's perception of the module includes: the temperature sensor, humidity sensor, light sensor. Every 31 seconds humidity, temperature, light and update changes in voltage value, running time is 37 days.

B. *Measure*

Get rid of the noise in the data preprocessing effect evaluation indexes are divided into signal-to-noise ratio (SNR) and smoothness (SR) in [14].

Signal Noise Ratio (SNR): the original signal in measured signal and de-noising signal since the square root of the variance.

$$SNR = 10 \times \log \left[\frac{\sum_{i=1}^{N} S_i'^2}{\sum_{i=1}^{N} \left(S_i' - S_i \right)^2} \right] \quad (9)$$

Smoothness (SR): refers to the de-noising signal after poor scores of root of variance and the original signal poor score variance ratio of the root.

$$SR = \frac{\sum_{i=1}^{N-1} \left(S_{i+1} - S_i \right)^2}{\sum_{i=1}^{N-1} \left(S_{i+1}' - S_i' \right)^2} \quad (10)$$

In the equation (10), S_i is the original signal, S_i' is the signal with de-noising.

Fault diagnosis accuracy definition: it is defined as the ratio of the number of faulty sensor nodes detected to the actual number of faulty sensor nodes in the network.

$$f_{FDA} = \frac{N_{fault}}{N_{totalfault}} \times 100\% \quad (11)$$

Fault diagnosis ratio of false alarm definition: refers to the ratio of the number of fault-free sensor nodes diagnosed as faulty to the actual number of fault-free sensor nodes.

$$f_{FAR} = \frac{N_{normal\_fault}}{N_{normal}} \times 100\% \quad (12)$$

V. RESULTS ANALYSIS

Under the different failure probability of the sensor nodes, this paper puts forward the CS-DFD algorithm and the traditional DFD algorithm are both make a contrast experiment, the experimental comparison results as shown in Fig. 4:
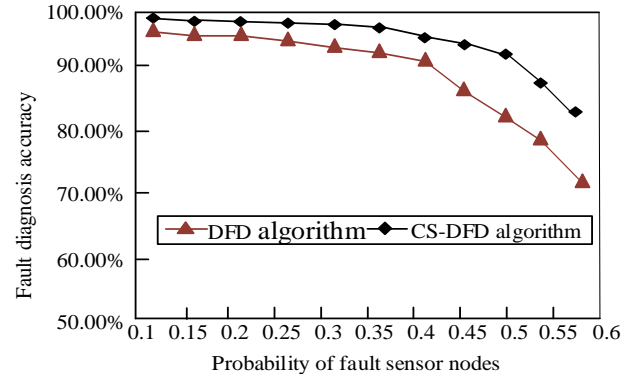

Figure 4. Fault diagnosis accuracy comparison chart.

By Fig. 4, the diagnostic accuracy big decline in DFD algorithm when the failure probability is about 30%, the reason may be overly dependent on the state of the neighbor nodes and judgment is quite harsh. The CS-DFD algorithm changes in the diagnostic accuracy of smooth, the reasons are that the cross sliding window has compared the data from neighbor nodes with historical data involved in the node itself at the same time.

Under the different failure probability of the sensor nodes, the CS-DFD algorithm has companied with the classic DFD algorithm, the experimental comparison results as shown in Fig. 5:
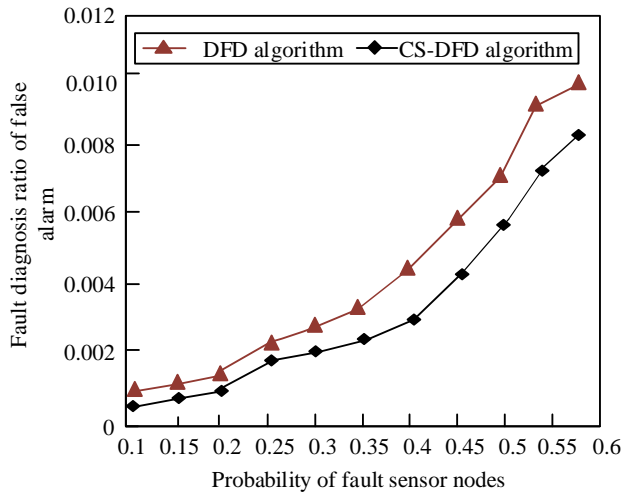

Figure 5.   False alarm ratio comparison diagram

From the Fig. 5, the false alarm ratio of the CS-DFD algorithm is on the rise, when the failure probability is about 35%, DFD algorithm corresponding to the change of the false alarm ratio range is larger, the reason may be ignored the node itself in the process of diagnosis history status. But the CS-DFD algorithm corresponding fault than the former is characterized by low false alarm ratio trend, the reason is that diagnosis by historical data and the data from neighbor nodes comprehensive judgment.

From the time point of view, the consuming time of fault diagnosis results are in table I:

TABLE I.   THE CONSUMING TIME OF DIAGNOSIS TABLE

Fault Probability of sensor nodes	DFD	CS-DFD
10%	2.133ms	1.154ms
15%	2.328ms	1.215ms
20%	2.531ms	1.233ms
25%	3.012ms	1.242ms
30%	3.418ms	1.302ms
35%	3.854ms	1.317ms
40%	3.022ms	1.351ms

From the table I, the CS-DFD algorithm diagnostic time is less than the traditional DFD algorithm relatively, the specific reason is that the amount of calculation has reduced, then shorten the time of diagnosis.

VI.   CONCLUSION

In the research field of wireless sensor networks, effective and accurate method for diagnosing nodes faults is a hotspot of research in recent years, aiming at the limitation of data itself too reliant on node, expert experience dependence are strong, such as large amount of calculation, this paper proposes a   method which named CS-DFD closely the DFD algorithm based on distributed approaches, through the establishment of a high dimensional space vector, introducing historical data and the data from neighbor nodes and constitute the cross sliding window, detecting the abnormal vector through similarity calculation, achieve the goal of sensor nodes fault diagnosis. This article also compared with the traditional DFD algorithm, the results show that the method has improved the accuracy of fault diagnosis and fault diagnosis ratio of false alarm control at low range. Mainly aiming at node fault diagnosis for the network fault diagnosis needs to be further research.

ACKNOWLEDGMENT

REFERENCE

[1]   Huang P, Xiao L, Soltani S, et al. The Evolution of MAC Protocols in Wireless Sensor Networks: A Survey[J]. IEEE Communications Surveys & Tutorials, 2013, 15(1):101-120.

[2]   Mao X, Miao X, He Y, et al. CitySee: Urban CO 2 monitoring with sensors[C]//INFOCOM, 2012 Proceedings IEEE. IEEE, 2012: 1611-1619.

[3]   Zhu T, Cao Z, Gong W, et al. Illuminations and the revelations: lessons learned from GreenOrbs project development [J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2013, 17(4): 42-46.

[4]   Roux M, Chauve D, Keil-Simon G. Fault management in wireless sensor networks[J]. IEEE Wireless Communications, 2007, 14(6):13-19.

[5]   Ren FY, Huang HN, Lin C. Wireless sensor networks. Journal of Software, 2003,14 (7):1282~1291.

[6]   Kung T L, Chen H C, Tan J J M. On the faulty sensor identification algorithm of wireless sensor networks under the PMC diagnosis model[C]// International Conference on Networked Computing and Advanced Information Management2010:657-661.

[7]   Weber A, Kutzke A R, Chessa S. Diagnosability evaluation for a system-level diagnosis algorithm for Wireless Sensor Networks[J]. 2010:241-244.

[8]   Weber A, Kutzke A R, Chessa S. Energy-aware test connection assignment for the self-diagnosis of a wireless sensor network[J]. Journal of the Brazilian Computer Society, 2011, 18(1):19-27.

[9]   Chen J, Kher S, Somani A. Distributed fault detection of wireless sensor networks[C].Proceedings of the 2006 workshop on Dependability issues in wireless ad hoc net-works and sensor networks, New York, 2006: 65-72.

[10]   A. Mahapatro and P. Khilar,"Online distributed fault diagnosis in wireless sensor networks," Wireless Personal Communications,pp.1–30,2012,doi:10.1007/s11277-012-0916-8.

[11] X. Miao, K. Liu, Y. He, Y. Liu, and D. Papadias, "Agnostic diagnosis: Discovering silent failures in wireless sensor networks," in INFOCOM,2011 Proc. IEEE, April 2011, pp. 1548 –1556.

[12] Mahapatro A, Khilar P M. Energy-efficient distributed approach for clustering-based fault detection and diagnosis in image sensor networks[J]. Iet Wireless Sensor Systems, 2013, 3(1):26-36.

[13] Ma Q, Liu K, Miao X, et al. Sherlock Is Around: Detecting Network Failures with Local Evidence Fusion [J]. Parallel and Distributed Systems, IEEE Transactions on, 2015, 26(5): 1430-1440.

[14] Zhang X L, Zhang F, Yuan J, et al. Sensor fault diagnosis and location for small and medium-scale wireless sensor networks[C]// International Conference on Natural Computation. 2010:3628-3632.

[15] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

[16] Turney P D, Pantel P. From frequency to meaning: Vector space models of semantics [J]. Journal of artificial intelligence research, 2010, 37(1): 141-188.

[17] Liu Y, Liu K, Li M. Passive diagnosis for wireless sensor networks[J]. Networking IEEE/ACM Transactions on, 2010, 18(4):1132-1144.

[18] Ruiz L B, Siqueira I G, Oliveira L B E, et al. Fault management in event-driven wireless sensor networks[C]// International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, MSWIM 2004, Venice, Italy, October. 2004:149-156.

[19] Ramanathan N, Chang K, Kapur R, et al. Sympathy for the sensor network debugger[C]// International Conference on Embedded Networked Sensor Systems, SENSYS 2005, San Diego, California, Usa, November. 2005:255--267.

[20] Chang JL, Cao F, Zhou AY. Clustering evolving data streams over sliding windows. Journal of Software,2007,18(4):905−918.

[21] "TheIntelLab,Berkeleydataset"[OL]http://berkeley.intelresearch.net/labdata/.