# A Fault Detection Technique based on Numerical Taxonomy in WSNs

Song Jia[1, 3*], Wang Bailing[2] and Peng Xiyuan[3]

*[1]School of Information and Electrical Engineering*
*Harbin Institute of Technology at Weihai, Shandong, China*
*[2]School of Computer Science and Technology*
*Harbin Institute of Technology at Weihai, Shandong, China*
*[3]Automatic Test and Control Institute*
*Harbin Institute of Technology, Harbin, China*
**songjia@hitwh.edu.cn*

***Abstract***

*Failures are inevitable in wireless sensor networks due to the fragile features and unattended deployment. To explore an effectual technique for node's fault detection caused by energy consumption, a new algorithm, Fault Detection Technique based on Numerical Taxonomy (FDNT) is proposed in this paper. The algorithm is deployed on the sink node with unconstrained energy consumption and is on the basis of dividing WSNs into clusters according to their geographical distribution. Simulation results indicate that this algorithm can get high detection accuracy and will consume less energy.*

***Keywords:*** *FDNT; Fault detection; WSNs*

## 1. Introduction

Although the concept of a sensor network is proposed in 1980 (DARPA initiated the Distributed Sensor Networks program), there comes an explosive increase of interest in wireless sensor networks (WSNs) recent years. WSNs are always deployed in unattended operation and remote sensing situation, where the harsh environment and the character of organizing networks randomly make fault happen frequently. WSNs' node has constrained energy, storage capacity and computational ability, which makes it easy to become faulty. In practice, faulty sensor node is one of the most important sources of faults in WSNs, which usually associates with a higher risk of the whole network malfunctioning [1-2]. Meanwhile, the faulty nodes will lead to "coverage hole" and a vaster expanse of the network. Therefore, fault management, including fault prevention, fault detection and fault recovery, is as important as other performance metrics such as energy efficiency, latency and accuracy no matter in theoretical or practical conditions [3-5].

Our institute, cooperating with Northern Arctic-alpine Horticultural Genetic Modification and Installation Cultivation Lab of Northeast Agriculture University, structures the WSNs-based greenhouse environment monitoring system (GEMS), aiming for monitoring the growing environment of a certain kind of improved tomato and providing support to the agriculture experts for analyzing the relationship between the development of that kind of tomato and its growing environment [6]. Node Self Detection by History data and Neighbors (NDHN) algorithm [7] and Fault Detection Technique based on Clustering in WSNs (FDTC) [8] are proposed to detect the faulty nodes. In this paper, we propose a Fault Detection Technique based on Numerical Taxonomy (FDNT) to improve the NDHN algorithm. FDNT analyses the measured data transmitted from sensor nodes to the sink node, which has unconstrained energy. By numerical taxonomy method, it is successful to avoid consuming much more energy in detecting personal parameter, calculating and communicating among ordinary sensor nodes in NDHN. And it turns out efficient through the GEMS system simulation.

## 2. Related work

### 2.1. Fault detection in WSNs

Fault detection techniques are used to detect potential faults and search for the source of faults, which are convenient for fault recovery work. There are many research achievements recent years. In PSFQ [9] and GARUDA [10] techniques, packet loss is used as fault characteristic parameter, in which, communication protocol orders the destination node detect the lost pocket in the transmission process of the networks. By the feedback information, it determines if there has a fault. There exist a lot of redundant nodes in the network, who produce plentiful data, so individual packet loss is rarely detected, which is the shortcoming of this type of technique. In the paper [11-12], interruption, delay or lack of regular network traffic are also used as indications of faults. In the paper [13-14], buffer occupancy level and channel loading conditions are used for detection, and there are a great deal of request and reply data, which will consume vast energy and lead to node premature death.

In the paper [15-16], it is proposed to use nodes' residual energy value to indicate the faults, which is used extensively. In the paper [17], nodes' self-detection method is proposed, which arranges within every node to detect the status of itself periodically and transmit to its neighbors. If the update packet does not achieve in a proper time, the node would be regarded as faulty node, and sink node will get an alarm. A modified technique [18] make the ordinary nodes making more decision and transforming less data to sink node, which reduce the energy consumption of leaders. In the paper [19], it is proposed that using neighbor's cooperation can improve the detection efficiency. Cluster leader makes a comparison of node's measurement with the neighbors' and then determines the faulty node, which is also widely used in fault detection in WSNs.

### 2.2. NDHN algorithm

The main idea of NDHN is using neighbors' cooperation and node's history records to diagnose the node's fault. Node A invites its neighbors to cooperate by sending them a cooperation request in order to gather their measurements. The neighbor nodes make the decision to cooperate or not based on well defined criteria. A cooperation node responses by sending its measurement to node A. The collected data is concatenated and analyzed in node A to judge whether node A is faulty or not. The judgment will be then sent to the intermediate node B and then to the next intermediate node until reaching the sink node. Every node repeats the same procedure and the sink node will know which node is fault in the cluster. At last, routing will be rescheduled and the faulty nodes will be avoided [7].

It is necessary to detect the node's battery level or other hardware parameters, and also necessary to collect the neighbors' measurements, which will consume much more node's energy. Furthermore, better fault detection accuracy needs higher operation frequency, as a result, there will be more energy consumption, and it will also decrease the lifecycle of the network desperately.

# 3. Fault detection technique based on numerical taxonomy

## 3.1. Principle of FDNT

The NDHN algorithm consumes a great deal of energy in fault detection, especially in the condition of pursuing excellent accuracy. Because of the character of large scale deployment and massive data gathering, we propose using numerical taxonomy method to extract feature of data and assist in network management [20, 21]. Thus, we modify NDHN and FDTC into FDNT for the purpose of reducing the energy consumption while not affecting the detection accuracy. The main idea of FDNT is clustering the measurement data into some taxonomic group in numerical value. Firstly, NDHN algorithm runs in a certain time frame (24 hours in this paper) to ensure the samples of measurements is healthy and correct. Secondly, sink node collects all nodes' measurements and stores them into database. Thirdly, the algorithm makes mathematical model to cluster all the measurements into several groups according to the data similarity, and it is claimed that measurements should have big similarity in taxonomy, but have obvious difference between taxonomies. Finally, the new collected measurements are compared with the taxonomy before activating alarm when the difference value is beyond the threshold. Then NDHN algorithm is reset to detect new fault. In this way, it omits the procedure to detect and transmit nodes' hardware parameters frequently, and does the calculating works in sink node which has unconstraint energy, as a consequence, it realizes the purpose of saving energy.

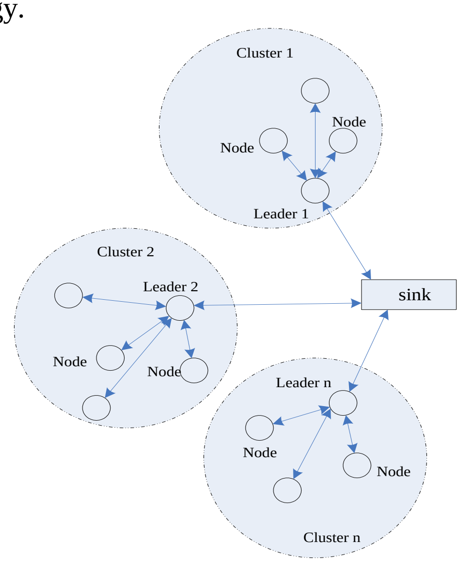



**Figure 1. The Model Schemes of WSNs**

## 3.2. Assumptions

The algorithms in this paper are developed under the following assumptions:

1) WSNs are deployed with clustering model, as showed in Figure 1;

2) All sensor nodes are stationary after deployment;

3) All sensor nodes have the same structure and function, and their status is equal in any way;

4) There is only one sink node, and it is deployed outside the network, which has unrestricted energy;

5) Every node has its location information.

6) We denote $e$ as the initial energy of every node, and the energy consumption model [22] is showed by equation (1) (2).

$$\begin{cases} E_{Tx}(l,d) = lE_{elec} + l\varepsilon_{fs} d^2, if d \le d_0 \\ E_{Tx}(l,d) = lE_{elec} + l\varepsilon_{amp} d^4, if d > d_0 \end{cases} \quad (1)$$

$$E_{Rx}(l) = lE_{elec} \quad (2)$$

Where, we denote $E_{Tx}(l,d)$ as the energy consumption by transforming $l$ bit data crossing $d$ distance, and denote $E_{Rx}(l)$ as the energy consumption by receiving $l$ bit data. $E_{elec}$ is denoted as the emission circuit's loss of energy. The power amplification loss is calculated by free space model if the transmit distance $d$ is less than threshold $d_0$, while calculated by multipath fading model if $d$ is greater than $d_0$. We also denote $\varepsilon_{fs}$ and $\varepsilon_{amp}$ as the power amplifier's energy consumption of the two models mentioned above.

### 3.3. Procedure of FDNT

1) Sink node collects several samples of measurements, marked $Z_i = \{Z_1, Z_2, \cdots, Z_N\}$;

2) Set a threshold of similarity scale as $\tau$;

3) Choose a sample data $Z_1$ randomly and consider it as the core, $C_1$, of the first taxonomy;

4) Calculate the deviation between $Z_1$ and all measurements $Z_i = \{Z_1, Z_2, \cdots, Z_N\}$, marked $D_{1i}$;

5) If $D_{1i} \le \tau$, it is judged that $Z_i$ belongs to $C_1$, otherwise it belongs to $C_2$, a new taxonomy;

6) Calculate the difference value between $C_2$ and the rest data of $Z_i$ to get $D_{2i}$;

7) If $D_{2i} \le \tau$, it is judged that $Z_i$ belongs to $C_2$, otherwise it belongs to $C_3$, another new taxonomy;

8) Repeat step 4) to 7) until all measurements are processed;

9) When WSNs is activated, calculating the 2-norn between every new measurement $Z_i'$ and the average value of $C_i$;

10) If the 2-norn is greater than threshold $\sigma$, it triggers an alarm and the sink node broadcasts commands to reset NDHN algorithm.

# 4. Simulation and Discussions

## 4.1. Design of simulation experiment

Basing on the voltage data of nodes collected during Aug.10, 2010 to Aug.14, 2010 in the experiments of GEMS, we simulate the algorithm. We did the experiments in the intelligent greenhouse of Northern Arctic-alpine Horticultural Genetic Modification and Installation Cultivation Lab, which is the key provincial laboratory in Northeast Agriculture University[6]. The experiment field is showed in Figure 2.

MATLAB is used to perform the simulations. The GEMS contains 24 nodes in a square region of size 140×100m$^2$, in which active nodes are structured and deployed in advance and 10 redundant nodes are deployed randomly. We use the FDNT technique to detect the voltages of the nodes. When the voltages are lower than 2.2V, nodes in the system will stop transmitting data and be regarded as faulty practically.

**Figure 2. The Sensor Nodes Deployed in the Greenhouse**

In practical, clustering method in WSNs is always based on their geography position. In this paper, we cluster the network into 2 clusters for the reason of the less number of nodes, just like FDTC [8]. We illustrate the proof in detail for the integrity of this paper, as showed in Figure 3.

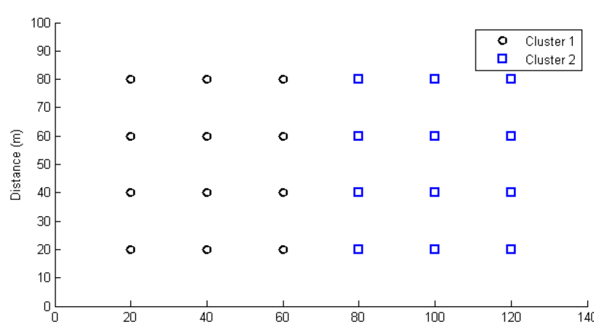



**Figure 3. The Simulation of Clusters in WSNs**

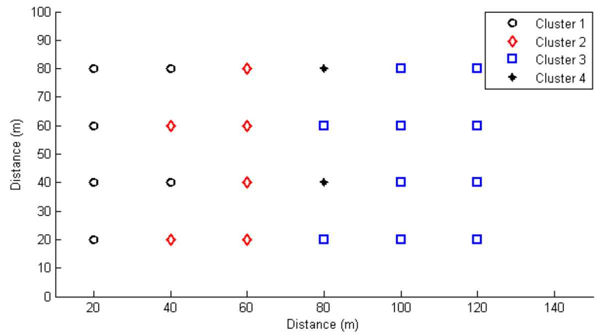



**Figure 4. Clustering Result of FDNT**

We use nodes' air humidity measurements as the samples to cluster the network, and the simulation result without faulty node is showed as Figure 4.

Then we set 4 faulty nodes, and choose $\tau = 10, \sigma = 10$. The simulation result of FDNT is showed in Figure 5.

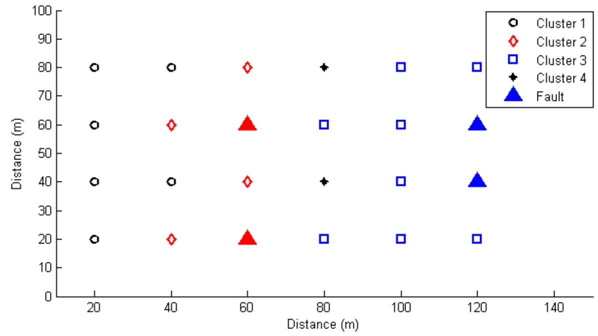



**Figure 5. Fault Detection Result of FDNT**

### 4.2. Discussions

NDHN is also simulated to compare with FDNT, and we choose $\alpha = 0.8$, $\beta = 0.2, \tau = 0.20$. The simulation is showed in Figure 6, and the comparison is showed in Table 1.

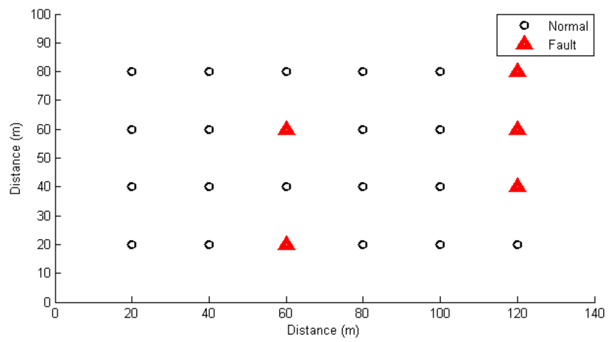



**Figure 6. Fault Detection Result of NDHN**

**Table 1. Details of Comparison**

| Algorithm | Number of nodes | Number of faulty nodes | Number of detected faulty nodes | Detection accuracy |
|---|---|---|---|---|
| NDHN | 24 | 4 | 5 | 75% |
| FDNT | 24 | 4 | 4 | 100% |

It should be specially explained that there are 4 faulty nodes actually, FDNT algorithm detects the faulty nodes correctly, but there appears false alarm in NDHN, where a healthy node is treated as fault. When calculating the detection accuracy, we consider the error rate is $1/4$, so the detection accuracy is 75%. The detection accuracy and false alarm mostly depend on $\tau$ and $\sigma$, which are chosen in experience, and this is the shortcoming of FDNT.

We stabilize $\tau = 10$ and change the value of $\sigma$, then run FDNT to deal with the GEMS, and the detection accuracy and false alarm are showed in Figure 7 and Figure 8 respectively.

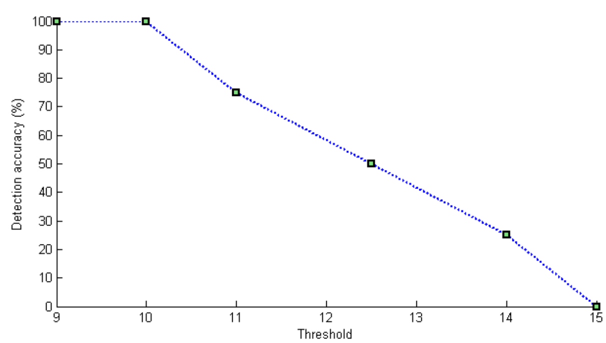



**Figure 7. Detection Accuracy vs. $\sigma$ of 24 Nodes**

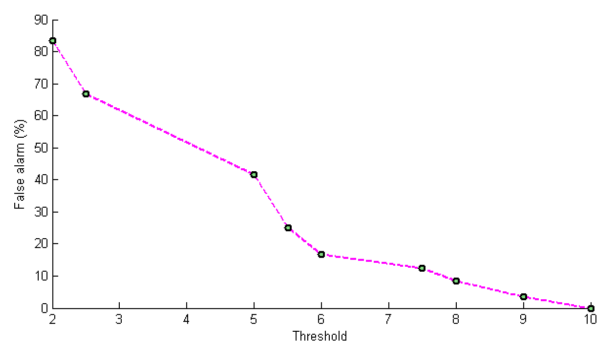



**Figure 8. False Alarm vs. $\sigma$ of 24 Nodes**

Energy resources are very precious in WSNs, so it is the most important content to reduce the energy consumption in fault detection algorithm. The node's energy consumption includes 3 parts: measuring consumption, communicating consumption, and data processing consumption. The communicating consumption is the biggest one, which includes 4 parts: transmitting, receiving, idle and sleeping.

The relation of node's energy consumption is showed as equation (3).

$$E_T > E_R > E_M > E_P > E_I > E_S \quad (3)$$

Where, $E_T$ means transmitting consumption, $E_R$ means receiving consumption, $E_M$ means measuring consumption, $E_P$ means data processing consumption, $E_I$ means idle consumption, $E_S$ means sleeping consumption.

In NDHN, it will consume more $E_R$ and $E_T$ when collecting neighbors' measurements, and will consume some $E_M$ to measure self hardware parameter, and will consume some $E_P$ to detect the fault state of itself, and will consume much more $E_T$ to transmit the result to the sink node. But in FDNT, this procedure is operated mainly in the sink node, thus the nodes' energy consumption can be cut down a lot. The comparison of energy consumption is showed in Table 2.

**Table 2. Comparison of Nodes' Energy Consumption**

| Algorithm | Communicating consumption | | | | $E_M$ | $E_P$ |
|---|---|---|---|---|---|---|
| | $E_T$ | $E_R$ | $E_I$ | $E_S$ | | |
| NDHN | Yes | Yes | No | No | Yes | Yes |
| FDNT | No | No | No | No | No | No |

## 5. Conclusion

This paper presents a technique FDNT to modify NDHN and FDTC for node fault detection in fault management of WSN. According to the simulation result, the technique proposed in this paper is feasible and effective to reduce energy consumption and false alarm greatly, while still maintains high detection accuracy.

In the future, we will do some research of mathematical method in modeling and the threshold choosing. And we will research the time and space correlation of measurements to perfect FDNT. Furthermore, the detection efficiency of FDNT also needs to be studied when the number of nodes is larger.

## Acknowledgements

The research presented in this paper is supported by National Natural Science Foundation of China （61170262）, National Natural Science Foundation of China (61102038), Program for New Century Excellent Talents in University (NCET-10-0062) and Research Fund for the Doctoral Program of Higher Education (20092302110013). Special thanks to Northeast Agricultural University College of Horticulture and Northern Arctic-alpine Horticultural Genetic Modification and Facility Cultivation Lab for their research support.

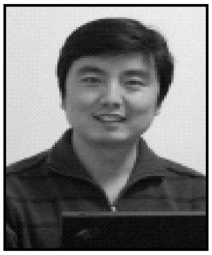
## References

[1] P. Yu, S. Jia and P. Xiyuan, "Survey of Fault Management Framework in Wireless Sensor Networks", Journal of Electronic Measurement and Instrument, no. 23, **(2009)** December, pp. 1-10.
[2] X. Na, X. Pujun, T. Mei and Q. Haowei, "Research on cooperative cluster based data aggregation mechanism

in WSNs", Journal of Electronic Measurement and Instrument, vol. 4, **(2010)**, pp. 307-313.
[3] J. Wang, X. Yang, T. Ma, M. Wu and J. -U. Kim, "An energy-efficient competitive clustering algorithm for wireless sensor networks using mobile sink", International Journal of Grid and Distributed Computing, vol. 5, no. 4, **(2012)**, pp. 79-92.
[4] B. Tang, J. Wang, X. Geng, Y. Zheng and J. -U. Kim, "A Novel Data Retrieving Mechanism in Wireless Sensor Networks with Path-Limited Mobile Sink", International Journal of Grid and Distributed Computing, vol. 5, no. 3, **(2012)**, pp. 133-140.
[5] A. Nikdel, M. S. Kohshoori and S. M. Jamei, "An Intelligent and Energy Efficient Area Coverage Protocol for Wireless Sensor Networks", International Journal of Grid and Distributed Computing, vol. 5, no. 1, **(2012)**, pp. 41-56.
[6] P. Yu, X. Yong and P. Xiyuan, "GEMS: A WSN-based Greenhouse Environment Monitoring System", C. 2011 IEEE Instrumentation and Measurement Technology Conference, **(2011)**, pp. 1000-1005.
[7] P. Yu, S. Jia and P. Xiyuan, "A Self Detection Technique in Fault Management in WSN", C. 2011 International instrumentation and measurement technology conference, **(2011)**, pp. 1755-1758.
[8] S. Jia, P. Yu and P. Xiyuan, "Fault Detection Technique based on Clustering in WSN", Chinese Journal of Scientific Instrument, **(2012)** October, pp. 2214-2220.
[9] C. Y. Wan, A. T. Campbell and L. Krishnamurth, "Pump slowly fetch quickly (psfq): A reliable transport protocol for wireless sensor networks", IEEE JSAC, vol. 23, no. 4, **(2005)**, pp. 862–872.
[10] S. J. Park, R. Vedantham, R. Sivakumar and I. F. Akyildiz, "A scalable approach for reliable downstream data delivery in wireless sensor networks", ACM MobiHoc Conference, **(2004)** May, pp. 78-89.
[11] J. Staddon, D. Balfanz and G. Durfee, "Efficient tracing of failed nodes in sensor networks", Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, **(2002)** September, pp. 122–130.
[12] N. Ramanathan, K. Chang, L. Girod, R. Kapur, E. Kohler and D. Estrin, "Sympathy for the sensor network debugger", Proceedings of SenSys, **(2005)** November, pp. 255-267.
[13] Y. Sankarasubramaniam, O. B. Akan and I. F. Akyildiz, "Esrt: Event-to-sink reliable transport in wireless sensor networks", Proceedings of ACM MobiHoc, **(2003)** June, pp. 177-188.
[14] C. Y. Wan, S. B. Eisenman and A. T. Campbell, "Coda: Congestion detection and avoidance in sensor networks", Proceedings of SenSys, **(2003)** November, pp. 266-279.
[15] Y. Zhao, R. Govindan and D. Estrin, "Residual energy scan for monitoring sensor networks", Proceedings of IEEE WCNC, vol. 1, **(2002)** March, pp. 356-362.
[16] R. Mini, A. Loureiro and B. Nath, "The distinctive design characteristic of a wireless sensor network: the energy map", J. Elsevier Computer Communications, vol. 27, no. 10, **(2004)** June, pp. 935-945.
[17] M. M. Asim and H. Merabti, "A Fault Management Architecture for Wireless Sensor Networ", Wireless Communications and Mobile Computing Conference, **(2008)**, pp. 779-785.
[18] G. Yoo, J. Jung and E. Lee, "Fault Management for Self-Healing in Ubiquitous Sensor Network", Future Generation Communication and Networking Symposia, **(2008)**, pp. 21-25.
[19] M. Ding, D. Chen, K. Xing and X. Cheng, "Localized fault-tolerant event boundary detection in sensor networks", 24th Annual Joint Conference of the IEEE Computer and Communications Societies, **(2005)**, pp. 902-913.
[20] P. Yu, Luo Qinghua and Peng Xiyuan, "Analysis of Uncertain Data Processing Methods in Networking Test Framework", Chinese Journal of Scientific Instrument, vol. 31, no. 1, **(2010)** January, pp. 229-244.
[21] P. Yu, Luo Qinghua and Peng Xiyuan, "UIDK-mean: A Multi-dimensional Uncertian Measurement Data Clustering Algorithm", Chinese Journal of Scientific Instrument, vol. 32, **(2011)**, pp. 1201-1207.
[22] W. B. Heinzelman, A. P. Chandrakasan and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks", Wireless Communication, vol. 1, no. 4, **(2002)**, pp. 660-670.

## Authors

**Song Jia** received the B.S. and M.S. degree from Harbin Institute of Technology, in 2000 and 2007 respectively. Now he is studying for the doctor degree at Harbin Institute of Technology, his research interest is Wireless Sensor Networks.



**Wang Bailing** is working for Harbin Institute of Technology (abstract as HIT) as an associate professor. He got the Ph.D. degree from HIT in 2006. His research is mainly on information security, network security, parallel computing.