ISBN 978-952-5726-02-2 (Print), 978-952-5726-03-9 (CD-ROM)
Proceedings of the 2009 International Symposium on Information Processing (ISIP'09)
Huangshan, P. R. China, August 21-23, 2009, pp. 021-024

A Reliable Collaborative Detection Scheme of Event-region in Wireless Sensor networks

Zhang ShuKui * [1,2], Gong ShengRong[2], Cui ZhiMing[1,2] , and Fan JianXi[2]
[1] JiangSu Province Support Software Engineering R&D Center for Modern Information Technology Application in Enterprise, Suzhou, China, 215104
[2]Institute of Computer Science and Technology, Soochow University, Suzhou, China,215006
Email: {zhangsk, szzmcui, shrgong, jxfan}@suda.edu.cn

***Abstract*—Event detection is among the most important applications of wireless sensor networks(WSN). Due to the fact that sensor readings do not always represent the true attribute values, Aiming at the problem of low service quality in WSN, a reliable collaborative distributed detection scheme is presented. This algorithm exploits the spatial correlations among sensor nodes, rather than passing large quantities of raw data through the network, our network instead sends the coefficients of a regression polynomial, aggregation the median information, and judges the final state of the event-region. Simulation results demonstrate the superiority of the proposed scheme in the event-region detection.**

***Index Terms*—wireless sensor network, event-region,data aggregation , distributed detection .**

I. INTRODUCTION

Wireless sensor networks (WSN) are of great significance in resolving many real world problems, and have attracted increasing research interests in recent years. One of the most important applications of WSN is the detection of events,. An event can be defined as an exceptional change in environmental parameters such as temperature, pressure, humidity etc. However, an event may occur in many ways. When a particular sensor depicts a smooth variation over time, then the sensors are said to be spatio-temporally correlated just as the attributes[1],[2],[3],[4]. In accordance with different scenarios, it is necessary to exploit spatio-temporal characteristics of sensors to detect the emergence of event boundary accurately (eliminating faulty readings) and quickly convey this information to the sink node. Reporting the boundary of the event accurately is a challenging task as it may involve faulty readings from some sensors, which may affect the accuracy of the detected area of the event [5],[6],[7],[8],[9] .

Many applications in which the sensor readings have a normal distribution within a bounded range, event recognition can be implemented by using a threshold-based scheme, which involves a marginal computational overhead, rather than using fairly complicated schedules [10],[11],[12],[13] Due to spatial-correlation, at a particular instant, if the sensed area is larger than the coverage of a single sensor, neighboring sensors sense similar data values. Again, a sensor's own reported readings will be similar to the reading it reported in the previous instant due to the property of temporal correlation. Therefore, identification of sudden, irregular readings deviating from its readings at the previous instants or highly different from its neighbors' readings beyond a pre-specified threshold helps detect faulty sensors [14].

In general, it is best that one deployment can satisfy the needs of a variety of applications in WSN. Take the example of a group of sensor networks deployed in the forest, biologists need to study the environment influence to the growth of the zoology and botany according to its returned detection value; environment scientists study the environment quality in this region and the influence of microclimate in the area; and what the forest managers care about is whether this region would have fire or other disaster. Different users have different needs of the detection data returned from sensors, if you were to respond to every request of the user's query, the sensor networks need to return the same value of detection many times. Therefore, it is significant to one-off creditably collect all row data in the detection region (the so-called "credible information" is that the data that the Sink node receives in the user's pre-specified error limits of credibility with 100%), which enables the different user to conduct the inquiry, analyze and process separately so as to obtain information of their respective needs.

A sensor can give faulty readings (readings different from neighboring sensors or its own readings sensed in previous time intervals, beyond a pre-specified threshold) due to several reasons. For example, the reliability of the equipment is not high or the different batches of the same factory and the sensor may give wrong readings, due to the different manufacturing process and other unforeseen reasons. The sensor error can be divided into two categories[15]: one is positive fault, i.e. the sensors report the incident while the environment is in a normal situation; the other is negative fault, i.e., the sensors did not report while there is a specific incident. Therefore, in WSN, to guarantee the credibility of the primary data and eliminate the effects of the error readings is one of the key questions that the event region detection needs to solved.

II. EVENT REGION DETECTION

In case of most applications, we use the function $\sigma(i,t)=F(v_1,v_2,...,v_k)$ to characterize the

* Corresponding author: S.-K Zhang, Tel.:+86 0512 65241247,
E-mail address:zhangsk@suda.edu.cn, Postal address: No 1 Shizhi Street Suzhou China 215006.

readings (e.g., temperature, humidity etc.) by a sensor $i \in N$ at instance t where v_k is a parameter that impacts the sensor reading. In most sensor applications, due to known and unknown factors, it is not clear whether the exact expression of the function can be derived. Except some particular cases, it is not easy to model these factors since they may affect the readings in a time-varying manner and in a linear or non-linear way. Rather than trying to obtain an exact expression, we can formulate the basic properties of the function F, thereby analyzing the sensor readings $(\sigma(i,t))$.

A. General Properties of Sensor Readings

The function F which characterizes sensor readings possesses the following properties:

1) F for a sensor $i \in N$ is independent of other sensors.

2) There exist two constants C_{min} and C_{max} (such that $C_{\min} \le \sigma(i,t) \le C_{\max}$) providing the lower and upper bound respectively of the normal readings from a sensor.

3) Within the range $[C_{min}, C_{max}]$, $\sigma(i,t)$ is statistically continuous and it admits a probability distribution, (i.e. a normal distribution). Thus, a continuous probability density function can be used to express the distribution.

It formulates those applications in which the sensor readings follow the above properties. Property 1 intuitively explains that a sensor independently senses the environmental changes. Property 2 gives the bounded variable space of normal sensor readings. It is to be noted that different applications have different value for the parameter $\varphi(i)$ given by property 3 and variable spaces. In practice, the assumption can be ascertained by applications that can approximately fit as normal distribution such as daily air temperature, wind speed, etc.

B. Single Event Detection

When a leaf node collects readings (i.e. d_i) from its surrounding *NT* nodes, it computes P_{flag} for an event (i.e. event polynomial) or P for a normal phenomenon (i.e. normal polynomial) depending on whether τ_{th} (i.e., $|d_i - E(d_i)| > \tau_{th}$, $E(d_i) = (d_1 + d_2 + \ldots + d_n)/N$, where N is the sample size of historic readings) is exceeded or not. While receiving the reports from its child nodes, the parent node has also collected readings from its own surrounding *NT* nodes and has determined whether an event has occurred or not in its region. If a parent node also lies in the same region as any of its child nodes, it regenerates the child node's reading and computes a new polynomial with its own reported data. This polynomial can again be P_{flag} or P, depending on whether the parent node is inside the event region or not. If the event regions of a parent node and the corresponding child nodes are different, the child node's data packet is sent up unchanged. The process continues and the root finally receives two polynomials and the corresponding ranges. From the received P_{flag} and the corresponding area from a child node, the root node can get an estimation of the event boundary from the event location information. It analyses the corresponding area, x_{Nmin}, y_{Nmin}, x_{Nmax}, y_{Nmax} (where the suffix N represents an event) to get an estimation of the coverage of P_{event}. In Fig. 1, P_{event} has occurred inside R_{event}, at the corner of R. We observe that in the sub-tree with the parent node A as well as its children B and C fall in the event region, R_{event}. In this case, both B and C receive readings from *NT* nodes, with deviation greater than τ_{th} (i.e., $|d_i - E(d_i)| > \tau_{th}$). Again, all the children of B lie inside R_{event}. Therefore, the polynomials received by B from its two children are flagged, Since B itself also receives readings from nearest *NT* nodes with deviation greater than the threshold, it generates a new P_{flag} with its own reported data and regenerated data.

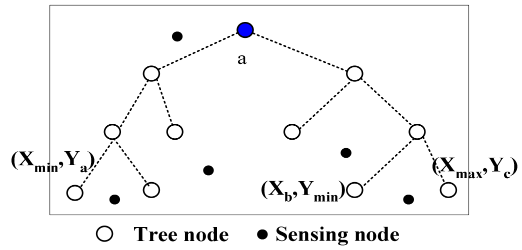

Fig. 1 A node calculate the boundary of the region for data regeneration

C. Multiple Events Detection

In contrast to single-event, the exchange of data packets and computational complexity between tree nodes will increase when multiple events occur. With the increase of the event sources, the effect of each event will reduce. As the packets are sent up to the tree, if one of the child nodes of a sub-tree approximates the data range as that of the parent, then a new polynomial is formed by combining both the data sets. However, if the dataset of none of the child nodes is the same as that of the parent, only their data range and range of area span are sent up to the tree. In this case, their areas are combined to form a larger area if both of two child nodes approximate the same data value.

Suppose in a rectangular area, there are two unusual events and a normal phenomenon. As shown in Fig. 2, two events P_{event1} and P_{event2} occur in region R.B detects event P_{event2}, while its children nodes C and D detect P_{event1} and a normal phenomenon separately. In this case, C and D transmit respective Polynomial and the coordinate scope $\{x_{cmin}, y_{cmin}, x_{cmax}, y_{cmax}\}$ as well as $\{x_{dmin}, y_{dmin}, x_{dmax}, y_{dmax}\}$ to B. B calculates the maximum and minimum data value of the two nodes received, getting the scope of their data. B will transform S_{cmax}, S_{cmin}, C, as well as the coordinates scope of S_{dmin}, S_{dmax}, D to A. If using single event detection, we need six packets, three of which are corresponding to three polynomials and the other three to the range of areas. Furthermore, A's other child G approximates the same phenomenon as D. Therefore, A combines their ranges together and senses the same event P_{event2}. Then, A recreates a new polynomial and sends its polynomial, its range and it's maximum and minimum to its parent H. At the same time, M's estimated value is the same as that sent by A and sends its coefficients to the root after combining N's regenerated data. As both H and I lie in R, the event region of P_{event2} is not modified further by H

and is sent to the root unchanged giving the approximated event region and boundary. In this way, both the events P_{event1} and P_{event2} are detected with limited computation overhead as only one polynomial regression is performed at each tree.

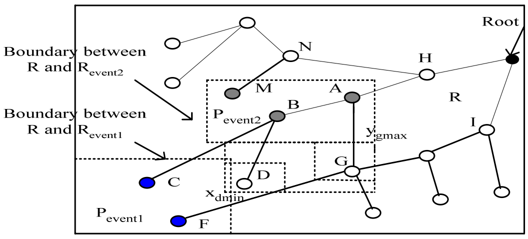

Fig. 2. Occurrence of two new events P_{event1} and P_{event2}.

III. Simulation Results and Discussion

This paper accesses simulation results using discrete event simulation platform NS2,which assumes a collision and contention free MAC protocol with simulation parameters shown in table 1, which focus the data Aggregation algorithm for performance evaluation.

Place randomly about 250 sensor nodes in the square of 800 × 800 units, changes in the temperature 30℃-35℃ for normal and consider the temperature 39℃-49℃ for abnormal as it may be on fire in the region. The depth of the tree is set to 4. From the data set it can be validated that the values of temperature are distributed according to a normal distribution (i.e., $\varphi(i)$) in case of no event. The C_{min},C_{max}, and E(i) are 30, 35, and 32.5 respectively and the variance of $\varphi(i)$ is 1.2. After validation of the assumption, a random location is selected inside R and a single event P_{event1} is generated at the centre with a normal distribution of temperature data in the range 55-65℃ $\tau_1 = \tau_{th}$ is set to 2.5, since E(di)-$C_{\min} = C_{\max}$ -E(di). = 2.5 The event area is called R_{event} as we described. τ_1 is set to 2.5, since $E(\sigma(i,t)) - C_{\min} = C_{\max} - E(\sigma(i,t)) = 2.5$ and $\tau_2 = \tau_{threshold}$ is set to 4.5. The following performance metrics are obtained for multiple rounds of sensing when the event(s) has occurred and the readings have been reported by the *NT* nodes to the nearest tree nodes.

Table 1 Values of simulation parameters used

Parameter	Variation
Area A	800×800
Communicate radius R	40m
Nodes total D	1630
Node density A/D	0.0025
Event area A_s	400×400
Father node choose probability P "	0.33
Tree depth p	4
Average of reports n_s	12

(1) Percentage error: Percentage Error is defined as the absolute deviation of the approximated reading (i.e., $\overline{z}$) from the true sensor reading (say, z) taken over all the sensors present on the boundary of the event:

$$E = \left(\sum_{k=1}^{n_b} \left(\frac{\lfloor z_k - \overline{z}_k}{z_k} \times 100\right) / n_b\right) \le \varepsilon_{th} \tag{1}$$

where ε_{th} = 10% is the error threshold given n_b (number of sensors present on the boundary).(2) Event recognition delay: This is defined as the period between the time of occurrence of an event and the final event.

A. Error Rate

Without changing the network size and the number of nodes, we employ STERD to detect a single event and the Extended STERD to detect two events. Fig.3 depicts the actual percentage error in event detection with variation of communication range. The observed percentage errors for different communication ranges are within 10% for both cases of STERD as shown in Fig.3.However, the percentage error of using Multi-STERD (error range from 6.8-8.9) to detect event is 6% higher than using single STERD(error range from 6.4-7.6). For single event detection with STERD, the range is slightly smaller because the approximate boundary can be more accurately pinpointed by regenerating data values from the final P_{flag} which is not available with the Multi-STERD scheme. As shown in Fig. 3 that the error level for the event detection does not significantly vary with communication range, unlike[3]where error increases substantially with the communication range. With increase in the communication range, more sensors with readings from region R_{event} may report to a tree. In STERD, since the tree nodes generate different polynomials corresponding to readings indicating different events, the overall error rate is independent of the communication ranges of each sensor. It can be observed from the result in Fig.6 that with an increase in the density of *NT* nodes, the error percentage increases slightly. With increase in the node density, the probability of a tree node lying at the border of the two events increases, and therefore, the fraction of *NT* nodes lying at the boundary between events in the network increases. Again, with an increase in the node density, accuracy of the approximated polynomial increases, as more readings are considered by tree nodes from reported sensors, relatively larger area is covered and a better approximation is provided of the sensed parameter over the region .The latter positive effect on accuracy controls a massive increase in the percentage error due to increase in the node density.

B. Delay Incurred in Event Detection

Fig. 4 shows that the delay in the event detection remains almost constant with increase in the node density. This is because the size of the tree is fixed, irrespective of the number of nodes in the network. Event detection delay mainly results from three parts: computational delay for event recognition (called event recognition delay), computational delay for polynomial (called computation delay), and the delay for event report by packets (called transmission delay). It can be explained from Fig. 4 that the event recognition and the computational delay for the polynomial are much smaller than the communication delay. It suggests that communication delay is almost 77% of the total delay. Once the aggregation tree is fixed, the communication delay remains almost constant and is independent of the density of the *NT* nodes. When the node density increases, the number of *NT* nodes around each tree node also increases. However, the result also shows that the computational overhead for event recognition and the construction of the polynomial remain almost constant when the number of nodes (n_{nt}) increases. It is due to the increase of node numbers (n) in the entire network does not cause any significant increase of n_s.STERD reduces the complexity of event recognition and event report because of the following reasons. Firstly, event recognition and polynomial-based data aggregation do not involve any complicated computation in STERD. Secondly, the tree-based network architecture makes the event recognition localized. Thirdly, when the network size (n) increases, the number of tree nodes increases accordingly so that STERD is made scalable.

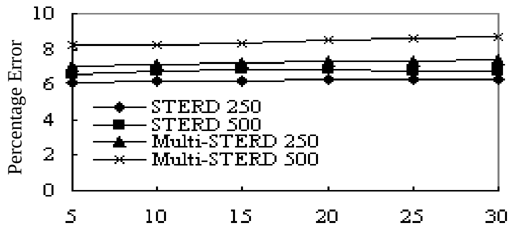

Fig. 3. Percentage error vs. communication range and node density.

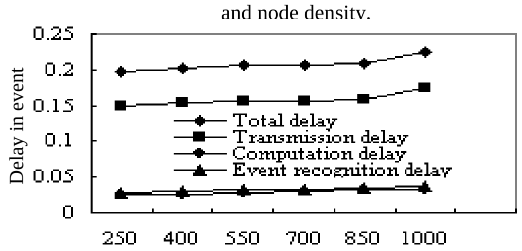

Fig. 4. Delay distribution in event detection vs. node density.

IV. CONCLUSION

This paper has studied high energy efficiency fault-tolerant detection of event region. Firstly, we proposed a novel data aggregation algorithm through the construction of the splay tree and this algorithm can also be able to detect the event attribute value lacking of the sensor node position, using the spatio-temporal correlation of the detected event and the error rate in the range of acceptable. In addition, on this basis, an event region detection algorithm based on the splay tree was proposed, the algorithm can detect a number of events and identify event, keeping the error ratio (due to faulty sensors) of the overall aggregated data reported to the *BS* under control, quickly conveying of this information to the *BS*, thereby reducing the energy consumption and the delay in data transmission.

ACKNOWLEDGMENT

Supported by the the National Natural Science oundation of China (60673092,60873116,60873047); Natural Science Foundation of Jiangsu Province of China(BK2008161, BK2008154);Funded by Preresearch Project of Soochow University; The Opening Project of JiangSu Province Support Software Engineering R&D Center for Modern Information Technology Application in Enterprise (SX200903);The Higher Education Graduate Research Innovation Program of Jiangsu Province in 2009.

REFERENCES

[1] F. Aurenhammer, R. Klein, Handbook of Computational Geometry.Amsterdam, Netherlands: North-Holland, 2000.201-290.

[2] M. Berthold, Intelligent Data Analysis, Springer Press,2nd edition, Feb 2007

[3] J. Chen, S. Kher, A. Somani, Distributed fault detection of wireless sensor networks, In Proc of the 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS 2006).2006.65-72.

[4] Q. Chen, K.Y. Lam, P. Fan , distributed Bayesian algorithms for fault tolerant event region detection in wireless sensor networks, IEEE Transactions on Computers ,2005,54(9):1182-1183

[5] K.K. Chintalapudi, R. Govindan, Localized edge detection in sensor fields[C],In Proc. IEEE Workshop on Sensor Network Protocols and Applications, Los Angeles May 2003.59-70

[6] M. Ding, D. Chen, et al.Localized fault tolerant event boundary detection in sensor networks, In Proceedings of the Annual IEEE Conference on Computer Communications (INFOCOM),Miami,2005,2:902-913

[7] C. Guestrin, P. Bodi, R. Thibau et al,Distributed Regression: an Efficient Framework for Modeling Sensor Network Data, IPSN. New York, ACM Press,2004

[8] F, Koushanfar. M, Potkonjak, et al, On-Line fault detection of sensor measurements, In Proc of the IEEE Sensors. 2003,2:974-979.

[9] F. Koushanfar, M. Potkonjak, et al, Fault-Tolerance in sensor networks. In: Handbook of Sensor Networks. CRC Press, 2004.

[10] F. Koushanfar, M. Potkonjak, et al. Error models for light sensors by statistical analysis of raw sensor measurements, In Proc of the IEEE Sensors, 2004,3,1472-1475.

[11] B. Krishnamachari, S.S. Iyengar, Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks, IEEE Trans. on Computers, 2004,53(3):241-250.

[12] T. Le, N. Ahmed, N. Parameswaran, S.Jha, Fault repair framework for mobile sensor networks, in Proc of IEEE International Conference on Communication System Software and Middleware, COMSWARE, 2006,1-8

[13] X. Luo, M. Dong, Y, Huang, On distributed fault-tolerant detection in wireless sensor networks, IEEE Trans. on Computers, 2006, 55(1):58-70.

[14] J.Z. Li, J.B. Li, S.F. Shi, Concepts, issues and advance of sensor networks and data management of sensor networks. Journal of Software, 2003,14(10):1717-1727

[15] R. Nowak, U. Mitra, Boundary estimation in sensor networks: Theory and methods, In Proc. 2nd International Workshop on IPSN, April 2003,2634,86-95

[16] F.Y. Ren, H.N. Huang, C. Lin, Wireless sensor networks. Journal of Software, 2003,14(7):1282-1291

[17] K. Ren, K. Zeng, W. Lou, Secure and Fault-Tolerant Event Boundary Detection in Wireless Sensor Networks, IEEE transactions on wireless communications, 2008, 7(1): 354-363