Wireless Sensor Data Fusion Algorithm Based on the Sensor Scheduling and Batch Estimate

Kaihong Zhang, Cong Li, and Wenhui Zhang

***Abstract*—As the goal of conserve battery power in very dense sensor networks and ensure the reliability of data, some sensor nodes must be put into sleep mode while other sensor nodes remain active for the sensing and communication tasks. With the feature of large amount of data for wireless sensor networks, high data redundancy and low energy of nodes, we propose the sensor nodes data fusion algorithm based on time-driven network data aggregation with the combination of sensor nodes scheduling and batch estimation. Least sensor nodes are scheduled in cluster to meet the conditions of the acquisition cycle and coverage area to a minimum the time of sensor nodes keep in working state to save network energy, while introducing data fusion mechanism based on batch estimation to enhance the reliability of fusion data. Simulation results show that the algorithm is effective to reduce network energy loss and significantly improve the reliability of the fusion data, and is applicably used in structural health monitoring environment.**

***Index Terms*—Data fusion, scheduling, sensor network, time-driven.**

I. Introduction

With the development of wireless communication technology, smart embedded monitoring technology, MEMS technology, wireless sensor networks and micro-computer system are connected gradually, it brings the unprecedented development and growth of the wireless sensor network technology, which has been widely used in military, agriculture, industry and other fields, including environmental monitoring, intelligent transportation, targeting etc. Among which, the structure of environmental health monitoring is one of the typical application of the time-driven sensor networks, bridges, tunnels, dams and other structures can be applied to performance monitoring. Time-driven based sensor networks have similar characteristics compared with event-driven sensor networks, such as the large amount of data, high data redundancy, and low energy of nodes. However, time-driven based sensor network nodes generally work in the unguarded harsh environment, energy supply mode of sensor nodes is battery-powered, and battery is difficult to maintain the life cycle of monitoring network. How to reduce network energy consumption that held longer network life cycle become a bottleneck in the design of the structure of the harsh environment of structural health monitoring sensor networks. In structural health monitoring process, the sensor data of monitoring area must be transmitted to the base station in real time to and accurately, a huge amount of data and the network latency requirements of distributed data fusion in the network transmission process, in order to reduce data redundancy, and transmission power consumption.

In response to these problems, this paper proposes a combination of sensor network data fusion and sensor scheduling based on time-driven wireless sensor network data fusion algorithm (TSSBEA). Specifically, least sensor nodes are scheduled in cluster to meet the conditions of the acquisition cycle and coverage area to a minimum the time of sensor nodes keep in working state to save network energy, while introducing data fusion mechanism based on batch estimation to enhance the reliability of fusion data.

II. Related Work

At present, the specific wireless network data fusion algorithm is still in developing, common methods are cluster-based data fusion and the tree-based data fusion method. Cluster-based data fusion demonstrates the method in which entire network is divided into several cluster regions and cluster head node is elected in every cluster regions, which is responsible for receiving and data fusion of sensor nodes within a cluster and data sending to sink node, such as LEACH, TEEN, etc. In [1], Haowen Chan et al proposed SIA algorithm, the algorithm used for the maximum, minimum, average, sum functions, to finish data fusion and transmit the results of integration to the base station. In [2], CHEN Hui-fang et al proposed an algorithm of integration of cluster-based and adaptive data fusion, achieve adaptive data fusion on time and space. In order to prevent the low reliability caused by cluster head node failure, in [3], NECCHIL et al proposed EERIA protocol, is the first to use cluster Gossip algorithm to reduce the number of data transfer and elected cluster head in the final stage of fusion progress. Melmet et al proposed data fusion algorithm based on space-time correlation and proposed event detection standards [4]. In tree-based fusion method, the sink node collects data from the distributed sensor nodes hierarchically through the reverse tree; the source node is a leaf node of the tree structure, namely, data fusion tree. KRISHNAMACHARI et al gives energy saving effect with different tree for data fusion: collect near source (CNS), the shortest path tree (SPT) and greedy growth tree (GTI), in the case of certain data, the energy saving effect relationship is: GTI > SPT > CNS [5]. In [6], ISLAM et al proposed the tectonic energy efficient data fusion tree using genetic algorithms (GA). LIAO et al proposed an ant colony algorithm for data fusion to build the best fusion tree by pheromone cumulative method [7]. The sink nodes collect

Manuscript received October 16, 2012; revised November 12, 2012
The authors are with the Chongqing Jiaotong University (e-mail: 916835287@qq.com)

the maximum amount of information is an effective way to improve the quality of the data in the condition of energy and delay satisfied. ORDONEZ et al study in optimal information extraction in energy constrained sensor networks [8]. According to Shannon capacity equation and channel attenuation under additive white Gaussian noise (AWGN), optimal information extraction problem is transformed to nonlinear flow planning issues meet the energy constraints. Limit information amount sent by the node with setting the transmission power and the flow rate, thereby optimizing the amount of information received by the sink node.

A. Network Model

Sensor networks based on time-driven often vary according to environmental complexity, assume that wireless sensor N nodes randomly distributed in the environment of a bridge structure M, and the network has the following properties:

1) The structure environmental region is changeless, sensor nodes can not move any longer when installed because nodes are static layout.
2) Base stations are fixed and unique. The base station may be arranged in the sensor topology center position or a fixed location outside bridge structure environment, in accordance with the actual situation.
3) Network is left unattended does not require human maintenance after installation layout.
4) Network is congenetic and heterogeneous, all nodes have similar processing, communication capabilities, but the type of the transmitted data, the structure is probably different.
5) The node contains same amount of energy and can not be replenished, energy consumption in each round is different, and the base station calculates the residual energy of each node in the end of each round.
6) The sensor nodes obtain the location information from the base station.

B. Energy Consumption Model

In recent years, domestic and foreign researchers did a lot of research work in low-power wireless communication. In this paper, we introduce wireless communication protocols such as LEACH, PEGASIS mode where 50 nJ/bit used for driving the transmitter and receiver circuit, 100 pJ/bit/m^2 for driving the transmission amplifier. Consume energy of data fusion operations often can not be ignored, usually assumed fusion consumption of each bit is 5 *nJ* or 5 *nJ*/bit/message. In addition, we assume that the wireless channel is symmetric, i.e. the energy consumption of message m transmitted from node *i* to node *j* is equal to energy consumption of message m transmitted from node *j* to node *i*. Related operation energy consumption is shown in Table I.

TABLE I: SENSOR NODE OPERATING ENERGY CONSUMPTION

Operations	Energy consumption
Send ($E_{TX/bit}$)	50 nJ/bit
Receive ($E_{RX/bit}$)	50 nJ/bit
Signal amplification (ε_{amp})	100 pJ/bit/m^2
Data fusion ($E_{Data\text{-}fu}$)	5 nJ/bit/message

According to the LEACH protocol wireless communication model, given a distance threshold value d_{ref}, the threshold value is decided by the specific structure of the environmental conditions. Under this condition, the relationship between the node transmitting energy consumption and the distance between sending node and the receiving node shows as following:

$$\begin{cases} E_{TX} \propto d^2, d \le d_{ref} \\ E_{TX} \propto d^4, d > d_{ref} \end{cases} \quad (1)$$

The energy consumption of node receiving operation is $E_{TX/bit}$, $E_{bit} = E_{RX/bit} = E_{TX/bit}$ can be set due to the energy consumed in the transmitting and receiving each bit data are both 50 nJ. Obtained energy consumption equation when sending and receiving n bit data and the transmission distance is d is calculated as follows:

$$E_{RX}(n) = n \cdot E_{bit} \quad (2)$$

$$E_{TX}(n,d) = \begin{cases} n \cdot E_{bit} + \varepsilon_{amp} \cdot n \cdot d^2, d \le d_{ref} \\ n \cdot E_{bit} + \varepsilon_{amp} \cdot n \cdot d^4, d > d_{ref} \end{cases} \quad (3)$$

III. ALGORITHM DESCRIPTION

Be considered from the point of view of network energy savings, based on a time-driven sensor network nodes feedback data according to a fixed threshold frequency after arranged in a structure in the area of the environment. In sensor node scheduling algorithm, sensor nodes are scheduled with minimal time in working state by dynamically setting the acquisition frequency threshold and turns sleep; using a distributed network data fusion algorithm to minimize the amount of data redundancy.

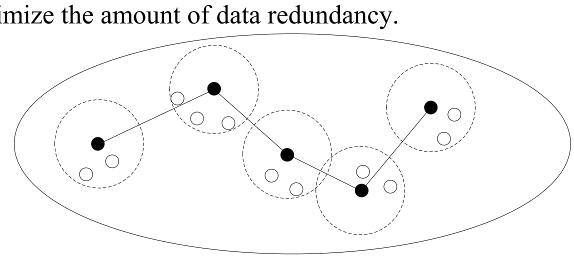

Fig. 1. Sensor nodes scheduled at the time k

A. Sensor Nodes Scheduling Algorithm

In densely distributed sensor networks, the data between adjacent nodes are related, and there may be some nodes which send the same type of monitoring data to the base station. Under the threshold condition of sensor scheduling based on time-driven, sensor nodes may be scheduled in the way of making sensor nodes inoperative in groups by timesharing basis. Specifically, scheduling model of sensor nodes in region A at the time k shown in Fig. 1, the scheduling model at the time $(k+\Delta t)$ in Fig. 2, the periodical node schedule working process in groups shown in Fig. 3. In Fig. 1 and Fig. 2, the connected nodes respectively show the sensor nodes scheduled in different time; the dotted circle represents area of the sensor nodes can sense. The principle of packet scheduling is to schedule the least sensor nodes and get full coverage and make sure sensor nodes scheduled in different time meet the conditions of the structure of all monitoring data.

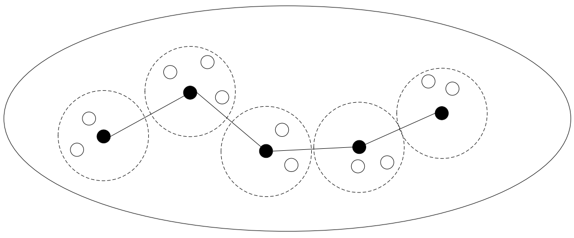

Fig. 2. Sensor nodes scheduled at the time $(k+\Delta t)$

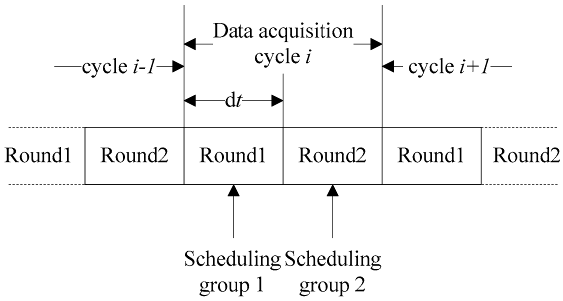

Fig. 3. The cycle scheduling work process in group

In the scheduling, the scheduled sensor node and nodes in its sense area form a cluster; the cluster-head is the node was scheduled. In some cases, node monitoring data of same cluster is related; therefore, in order to save network energy, the node in the same cluster can keep work state by turns. Clusters in the network is $N(i)$, $(i=1,2,\cdots,k)$, different clusters may have a different number of sensor nodes, the sensor node in the cluster is $S(j)$, $(j=1,2,\cdots,n)$, assume the cluster collection frequency threshold value is f_{th-i}, the acquisition cycle $T_i(T_i=1/f_{th-i})$ is divided into j time slots, each sensor node of the cluster only keep in working state in one time slot, and the rest of the nodes in a sleep state. Average acquisition cycle in time slots then the active time of sensor node is T_i/j. In process of scheduling, sink node send scheduling information to sensor nodes in random order, the node working process shown in Fig. 4.

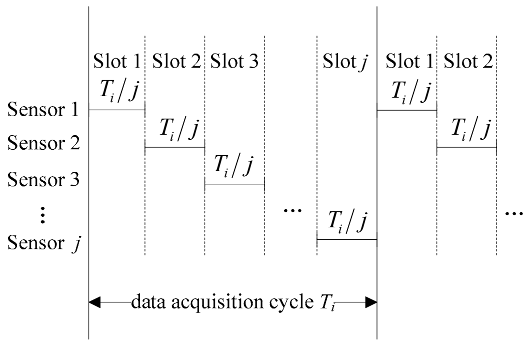

Fig . 4. Working process in nodes scheduling

Suppose that in period $T_i(t_1,t_2,\cdots,t_k)$, there are $A_i(i=1,2,\cdots,n)$ active nodes, the node coverage is C_i, the monitoring area is M, and then the sensor network Q of the scheduling algorithm is as follows:

$$Q:\ (\bigcup_{i=1}^{n} C_i \cap M \geq M) \cap 1\{S_i, S_j \notin \{t_i\}\} \tag{4}$$

$$\begin{aligned}(\sum x_i(\alpha(\delta P_i+\phi/(M_iN_i))+\\ \beta(U_i-w_i)))\cap(\sum x_i w_i \leq E)\end{aligned} \tag{5}$$

where (4) suggest that the nodes coverage get close to real monitoring area in widest degree, and the adjacent nodes in same layer can not select time slot t_i in the period T_i, it means two adjacent nodes can not sleep at the same time to ensure minimize blind spots. (5) describes the node monitoring information distribution and network node energy balancing problem, α factor term is probabilistic information P_i of monitoring point in the monitoring area, node location information M_i and the adjacent node information N_i; β factor term is residual energy, w_i is an energy consumption factor, x_i is scheduling information for the node. The algorithm detailed description as follows:

We introduce backoff algorithm to solve the issue of same level redundancy adjacent nodes select the same time zone, sink node randomly selected nodes in the monitoring area, and assigned equal scheduling period.

The first step of algorithm:

Condition A: nodes in same layer select same time slot

Condition B: nodes in different layer select different time slot

While (A && B)
Do
Set reselect time interval Δt
Set reselect threshold time value m
Set parameter
$k=Min[reselet-times,m], k\leq m$
R = random $[0,1,\cdots,(2^k-1)]$
Rescheduling time slot = $r\cdot\Delta t$
If (reselect times == m && scheduling finish == false)
Send failure information to sink node
Else
Maximum energy priority scheduling adjustments
Stop

Let the probability density function of the network monitoring point of the sensor node S_k is $P_{S_k}(x)$, then the probability $P\{S_k|E\in C(i,j)\}$ of node S_k in its neighborhood classification monitoring region $C(i,j)$ is:

$$P\{S_k|E\in C(i,j)\}=\int_{x\in C(i,j)} P_{S_k}(x)dx \tag{6}$$

By introducing Gaussian function approximated accurate monitoring performance of network coverage node, the continuous cumulative probability of node S_k is:

$$P\{S_k,\tau,\alpha_{S_k}|E\in R_{S_k,\tau}\}=e^{\frac{\tau^2}{2\alpha_{S_k}^2}}, \tau\in(0,d_{S_k}) \tag{7}$$

where in, τ is the approximate distance between the event E and the sensor node S_k, α_{S_k} is the monitoring accuracy of the sensor nodes. Binding equation (7), the average monitoring probability of overall information of all nodes in the neighborhood of the node S_k in a scheduling period is:

$$P\{S|E\in R\}=\sum_{=0}^{l-1}\sum_{=0}^{h-1}P\{S|E\in C(i,j)\}\times P\{E\in C(i,j)\} \quad (8)$$

Assume position appeared in the provided network monitoring area of the monitoring events obeys uniform distribution, and then the event probability in the monitoring area is:

$$P\{E\in C(i,j)\}=\frac{1}{l\times h} \quad (9)$$

Node average energy consumption of the entire node scheduling process:

$$E_s=\int_0^{x'}E_{active}(x)\frac{x}{R^2}dx+E_s^{'}\frac{R^2-(x')^2}{R^2} \quad (10)$$

Where in, E_{active} is energy consumption of active nodes and x' is equal distance parameter of equal energy consumption network nodes.

Network node monitoring probability and network energy information is given by the formula (3) (4),

$$T=uP+v(E_0-E) \quad (11)$$

where in, u,v is weight value factor, considering the energy maximum priority scheduling principles, set $v>u$ and E_0 is initial energy of node, P,E respectively represent the node monitoring the probability information and energy consumption. If $T>T_0$ (T_0 is threshold value set according to the structure), then the scheduling node select a relatively large time period, otherwise maintaining the previous scheduling configuration.

B. Network Data Fusion Algorithm

Nodes redundant working degree of network has been reduced after sensor node scheduling, but the data must undergo further fusion for sending to the sink node. This paper take into account the frequency characteristics of the special environment of the bridge structure and time-driven, based on Kalman filtering batch estimation fusion algorithm, multiple acquisitions results of a single sensor can be fused into a value well, and this method can also be used for multi-sensor data fusion. As below, we take the strain sensor group of a bridge expansion joints for example, give analysis of the principle of the algorithm.

Discrete Kalman optimal filtering observation equation is $Z(k)=C(k)X(k)+V(k)$, wherein, $Z(k)$ is the system observation vector, $C(k)$ is the system matrix, $X(k)$ is the state vector of the system, and $V(k)$ is a zero-mean white noise vector. Batches estimated recurrence formula is:

$$\begin{aligned}&\hat{X}(k/k)=A(k/k-1)\hat{X}(k-1/k-1)+\\&K_k(Z(k)-C(k)A(k,k-1)\hat{X}(k-1/k-1))\\&\hat{X}(k/k)=(I-K_kC(k))\hat{X}(k/k-1)+K_kZ(k)\\&=\{[P(k/k-1)+\\&C(k)^TR_k^{-1}C(k)]^{-1}P^{-1}(k/k-1)\}\hat{X}(k/k-1)+\\&P(k/k)C(k)^TR_k^{-1}Z(k)\end{aligned} \quad (12)$$

Suppose there are n strain sensors arranged in a key area of expansion joints on demand of which observations were divided into two groups, group j can be express as $S_{j1},S_{j2},\cdots,S_{jn},n_j\geq 2,j=1,2$, and $\sum_{j=1}^{2}n_j=n$.

Then the average of the two sets of values is:

$$\overline{S}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}S_{1i}\ ,\overline{S}_2=\frac{1}{n_2}\sum_{i=1}^{n_2}S_{2i} \quad (13)$$

Corresponding variance:

$$\sigma_{\overline{S}_1}^2=\frac{s_1^2}{n_1},\sigma_{\overline{S}_2}^2=\frac{s_2^2}{n_2} \quad (14)$$

where in,

$$s_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(S_{1i}-\overline{S}_1\right)^2>0$$

$$s_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_1}\left(S_{2i}-\overline{S}_2\right)^2>0$$

Set strain true value S_s , the strain data measurement equation $S=HS_s+V$ is obtained under optimal Kalman filter observation equation, wherein, S is the strain measured values, $H=\begin{bmatrix}1\\1\end{bmatrix}$ is the matrix of the system, V is the measurement noise of the Gaussian distribution. Namely:

V_1,V_2 is respectively measurement noise of S_1,S_2 and its mean is 0, and V_1,V_2 are mutually independent.

Measurement noise covariance:

$$\begin{aligned}&R=E\left[VV^T\right]=\\&\begin{bmatrix}E\left[V_1^2\right]&E\left[V_1V_2\right]\\E\left[V_2V_1\right]&E\left[V_2^2\right]\end{bmatrix}=\begin{bmatrix}\sigma_{S_1}^2&0\\0&\sigma_{S_2}^2\end{bmatrix}\end{aligned} \quad (15)$$

Since no strain data on the environment of the structure prior to measurement, $P(1/0)=\frac{I}{\varepsilon}\rightarrow\infty$ is known by the initial conditions priori knowledge, thus the variance of the previous measurement result is $P^-=\infty,(P^-)^{-1}=0$, and the variance of the strain data fusion value can be introduced according to the error covariance matrix $P(k/k)=\left[P^{-1}(k/k-1)+C(k)^TR_k^{-1}C(k)\right]^{-1}$:

$$P^{+} = \left[(P^{-})^{-1} + H^{T} R^{-1} H\right]^{-1} = \left\{ \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sigma_{S_1}^2} & 0 \\ 0 & \frac{1}{\sigma_{S_2}^2} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}^{-1} = \frac{\sigma_{S_1}^2 \sigma_{S_2}^2}{\sigma_{S_1}^2 + \sigma_{S_2}^2} \tag{16}$$

Formula (1) export batches estimated strain data fusion value:

$$\hat{S}^{+} = \frac{\sigma_{S_1}^2 \sigma_{S_2}^2}{\sigma_{S_1}^2 + \sigma_{S_2}^2} \cdot \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sigma_{S_1}^2} & 0 \\ 0 & \frac{1}{\sigma_{S_2}^2} \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \frac{\sigma_{S_1}^2 \overline{S}_2 + \sigma_{S_2}^2 \overline{S}_1}{\sigma_{S_1}^2 + \sigma_{S_2}^2} \tag{17}$$

IV. Simulation and Results Analysis

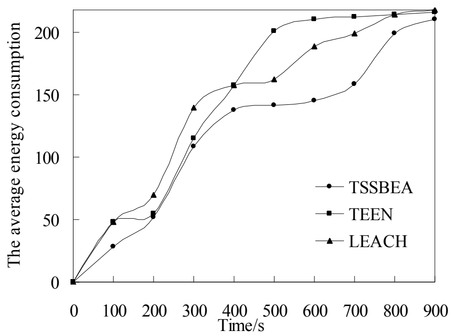

Fig. 5. Node average energy consumption

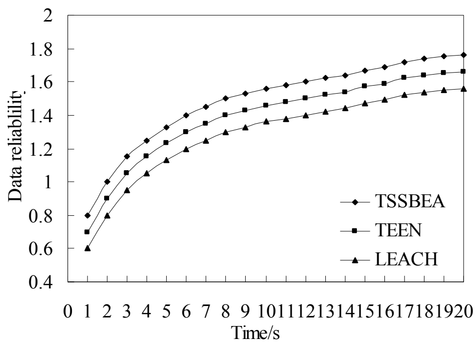

Fig. 6. Data reliability analysis

The main indicators of network simulation are node average energy consumption and data reliability, the experimental simulation platform used in this paper is NS-2.27. Fig. 5 shows the node of the average energy consumption of TSSBEA, LEACH, and TEEN. From the figure it can be seen with the network working time becomes longer, the average energy consumption using the algorithm presented in this paper less than LEACH and TEEN in a relatively long activation time. The main reason is that the monitoring probability scheduling algorithm using the principle of maximum energy priority scheduling. Fig. 6 shows a comparison of the three algorithms in the node data reliability. Data reliability of TSSBEA algorithm is higher than TEEN and LEACH, since fully take into account the emergence of network blind spots, enhanced the coverage and network connectivity while scheduling, and therefore to some extent, improve the network the reliability of the data.

V. Conclusion

Bridge health monitoring sensor network is a typical application of wireless sensor networks, wireless sensor networks taking into account the network energy consumption and data reliability requirements, this paper binding structure of environmental monitoring and sensor time-driven features, proposed TSSBEA algorithm, combine the sensor work time scheduling and batch estimate based on Kalman filtering data fusion, the reliability of the network data and network energy consumption of TSSBES are better than LEACH and TEEN. In addition, this paper does not refer network latency and node survival analysis; researches will be doing in further works.

References

[1] H. Chan, A. Perring, and B. Przydatek *et al*., "Secure Information Aggregation in Sensor Networks," in *Proc of the first ACM Conference on Embeded Networked System*, 2003, pp. 255-265.

[2] H. F. Chen, H. Mineno, and T. Mizuno, "Adaptive data agrregation scheme in clustered wireless sensor networks," *Computer Communications*, 2008, vol. 31, no. 25, pp. 3579-3585.

[3] A. N. Bonivento and L. Lavagno *et al*., "Eerina:an energy efficient and reliable in-network aggregation for clustered wireless sensor networks," in *Proc of Wireless Communication and Networking Conference*, 2007, pp. 3364-3369.

[4] M. C Vurean, O. B Akan, and I. F Akyildiz, "Spatiotemporal Crrelation: Thery and Applications for Wireless Sensor Networks," *Computer Networks*, vol. 45, pp. 245-259, 2004.

[5] B. Krishnamachari, D. Estrin, and S. Wicker, "Modeling data-centric routing in wireless sensor networks," in *Proc of IEEE Infocom*, New York, IEEE Computer society, 2002, pp. 2-14.

[6] O. Islam, S. Hussain, and H. Zhang, "Genetic algorithm for data agrregation trees in wireless sensor networks," in *Proc of the 3rd IEEE International Conference on Intelligent Environments*, 2007, pp. 312-316.

[7] W. H. Liao, Y. C. Kao, and C. M. Fan, "Data aggreation in wireless sensor networks using ant colnoy algorithm," *Journal of Network and Computer Aplications*, vol. 4, pp. 387-401.

[8] F. Ordonez and B. Krishnamachari, "Optimal information extraction in energy-limited Wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1121-1129, 2004.

Kaihong Zhang was born in Sichuan Province, China, 1973. He received the BSc.Degree in Electronic and Information Engineering from Southwestern Normal University, Chongqing, China in 1996. He received the MSc.Degree in Computer Application Technology from Southwestern Normal University, Chongqing, China in 2005. Currently, he is an associate professor and master instructor at School of Information Science and Engineering, Chongqing Jiaotong University. He is a doctoral student in Control Theory and Control Engineering of Chongqing University. His research interests include bridge health monitoring, sensor and sensing technology. Mr. Zhang is a director of EDA/SOPC Research Association in Southwest China. He is a member of Chongqing Artificial Intelligence Society and Chongqing Electronic Society. He is an evaluation expert of Chongqing Municipal Government procurement.