Research on Data Collection based on Wireless Sensor Networks

Wei Fu

*College of Computer Science and Information Engineering，
Harbin Normal University, Harbin 150076, china,
fullway2005@139.com*

Abstract

The issue of data collection in wireless sensor networks was studied. In this dissertation, a mobile sink node was introduced and a track fixed data collection solution (TDCS) was proposed and realized. TDCS consisted of two phases. Firstly, a number of collection sites were selected according to the sensor distribution. And then a quantum genetic algorithm was performed to calculate the shortest loop across the sites. Simulation tests were carried out. Simulation results showed that TDCS had better performance in network throughput and energy efficiency, and it could collets more data.

1. Introduction

Wireless sensor network is one type which has numerous usages. Data collection is one of its basic applications [1-3]. People deploy plenty of sensor nodes in working environments and that they need to acquire varied data like temperature, moisture, concentration, speed, image, sound, video etc. In reality, many businesses contain huge data volume, e.g. the application relating to image and video. Such practical uses bring about certain challenges to the design of data acquisition solution [4-5]. It involves the following problems:

(1) Funneling effect: nodes locating nearer to the sink node consume more energy and segmentation occurs earlier in the network; it's the same with massive businesses;
(2) Communication overheads: for any kind of routing algorithm, there's definitely control overhead and it's required to consume nodes' energy;
(3) Connectivity limits: when the network is not wholly connected, acquired data of partial nodes cannot be submitted.

The above mentioned shortcomings are partly rooted in the statics of the sink node [6-7]. After mobile nodes like RacemoteZ, quadcopter were introduced, some scholars proposed the solution based on mobile nodes, which partially overcomes the above problems and has advantages in network throughput, transmission reliability. However the method has one problem, which is time delay; besides, it hardly determines the moving path of the sink node. Since most applications in data gathering don't impose higher requirement for instantaneity, the control of movements of the sink node is key [8-9]. So far some achievements were made in the field. The classical strategies refer to random, moving along network boundaries, moving towards nodes which send the biggest data volume; moving towards nodes which have the most residual energy. However, that method doesn't consider the distribution of nodes or overestimates overheads, not quite suitable for wireless sensor network [10-11].

To cope with those difficulties, we put forward a track fixed data collection solution. TDCS. The mobile strategy of TDCS is consisted of two aspects:

(1) Choose one group of collecting points based on the distribution of nodes

(2) Find out the shortest loop traversing those points. The loop is used as the collecting path of the sink node [12-13], along which the node moves and gathers data. According to researches by Akkaya [14], it's an NP problem to find out the best gathering site in the network. To calculate the shortest loop is TSP, which belongs also to NP. For those two points, there's no especially good solution. So for the sake of simplicity and approximation, we choose fewer collecting points for TDCS; to compute the shortest loop, TDCS utilizes quantum genetic algorithm (QGA) [15]. Here we make comparative tests with TDCS in different scenarios. Experimental results indicate that TDCS had better performance especially in network throughput and energy efficiency, and acquired more data as well.

2. Overall Thinking

TDCS protocol makes use of mobile sink node to collect sensory data generated in the network. Mobile paths have great influences on gathering performance. In TDCS protocol, the mobile path of the sink node is determined by the following steps:

(1) Determination of collecting point
(2) Calculation of the shortest path

In the first step above, TDCS protocol needs to select a group of collecting points from the monitored regions. Since its NP problem to locate the best collecting point in the network, no better solution at all, we're in the hope of making as few collecting points as possibly, which will not only reduce the calculated quantities in the second step and also make it probable to obtain the shortest path. In the wireless sensor network, the communication range of the convergent node over each collecting point can be regarded as a circular disc of which the location of the point is circle center. Hence in the first period, the task is to find out a group of coordinate collection which can satisfy these two conditions:

(1) The quantity of coordinates is as few as possible
(2) A group of circular disc can cover the monitored area, whose circle center is those coordinates and radius r.

In the second step, TDCS works out the shortest loop which passes over those collecting points, decreasing the mobile energy consumption of the sink node and time delay of submitted data. For the shortest loop is one NP problem, TDCS uses quantum GA for calculation. That method introduced the idea of quantum calculation on the basis of traditional genetic algorithm. The convergence speed and solving ability of the algorithm were both hoisted. TDCS reaches better results with it.

3. Determination of Collecting Points

3.1. Determination of the Location of Circular Disc

To avoid gaps among deployed discs, adjacent discs must be overlapping. As shown in Figure 1, disc A and B are mutually close and overlapping; while empty spaces below them are covered with disc C; therefore C is intersected with them. To keep fewer discs, overlapping spaces should be as small as possible. In this case, C's edges should meet right with the crossing point O of A and B; otherwise, there will have triple overlapping among them or empty space.
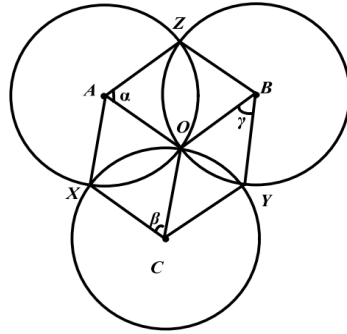
Figure 1. Overlapping Three Discs

3.2. Selection of Collecting Points

After all discs are defined, TDCS will choose one collecting point from each disc. When determining the location, TDCS considers mainly the reason of energy consumption.

TDCS takes advantage of the following way to compute the position of the collecting point.

For any one disc, we assume n nodes in it and (x_s, y_s) is the position of one pending collecting point.

(1) If n=0, no collecting point within the disc;
(2) If n=1, the location of the collecting point is the only one node coordinate;
(3) If n=2, make the central point of connecting line between two nodes as one collecting point;
(4) If n>2, compute the rectangular area of (x_s, y_s); when the area is bigger, calculate the mean value of all node's coordinates and use it as one collecting point; otherwise, use grid partition method to get approximation point.

4. Calculation of Mobile Path

After collecting points are gotten, TDCS uses quantum genetic algorithm (QGA) to search for the shortest path going through those points. QGA evolved on the basis of genetic algorithm. It brought in the concept of quantum computation, characteristic of rapid converging speed and stronger parallel searching capability. It has been widely used to deal with different optimization problems. In QGA, the basic unit forming the population is quantum chromosome. Each quantum chromosome is composed of some quanta. One typical quantum chromosome is shown as formula (1).

The conversion from quantum chromosome to traditional chromosome is realized by quantum observation operation. The update of quantum chromosome is fulfilled by quantum rotation gate. The application of QGA involves with chromosome coding and updating and calculation of the fitness.

$$\begin{pmatrix} \alpha_{11} \mid \alpha_{12} \mid \ldots \mid \alpha_{1k} \mid \alpha_{21} \mid \alpha_{22} \mid \ldots \mid \alpha_{2k} \mid \ldots \ldots \mid \alpha_{n1} \mid \alpha_{n2} \mid \ldots \mid \alpha_{nk} \\ \beta_{11} \mid \beta_{12} \mid \ldots \mid \beta_{1k} \mid \beta_{21} \mid \beta_{22} \mid \ldots \mid \beta_{2k} \mid \ldots \ldots \mid \beta_{n1} \mid \beta_{n2} \mid \ldots \mid \beta_{nk} \end{pmatrix} \quad (1)$$

4.1. Chromosome Coding

TDCS adopts the method which is applicable for encoding the optional links of the collecting point. In this encoding strategy, each gene of the chromosome corresponds to a

collecting point. The value of gene refers to link number chosen by the point. In Figure 2, there're totally seven collecting points from A-G, where connecting line means selectable link (overlong ones were removed). Table 1 shows the number of those links. Figure 3 is an instance of one chromosome, which stands for the loop A-B-C-E-F-G-D-A.
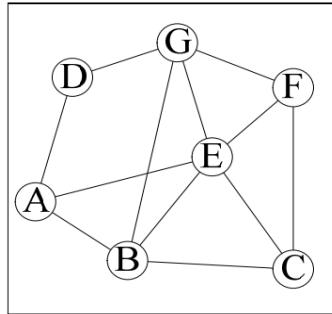

Figure 2. The Distribution of the Collection Points

Figure 3. Chromosome Sample

Table 1. Link Number of Each Collection Point

	0	1	2	3	4
A	(A, B)	(A, D)	(A,E)		
B	(B, A)	(B, C)	(B, E)	(B, G)	
C	(C, B)	(C, E)	(C, F)		
D	(D, A)	(D, G)			
E	(E, A)	(E, B)	(E, C)	(E, F)	(E, G)
F	(F, C)	(F, E)	(F, G)		
G	(G, B)	(G, D)	(G, F)	(G, F)	

4.2. Chromosome Update

The update of quantum chromosomes is executed by quantum rotation gate, as expressed like (2):

$$\begin{pmatrix} \alpha_t' \\ \beta_t' \end{pmatrix} = \begin{pmatrix} \cos(\theta_t) & -\sin(\theta_t) \\ \sin(\theta_t) & \cos(\theta_t) \end{pmatrix} \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} \tag{2}$$

In Table 2, f means fitness value of quantum chromosome; x_i is the ith chromosome; f_b is the fitness value of the best chromosome in one population; b_i is the ith best chromosome; δ is size of rotation, which is normally not fixed. On the initial stage of QGA solution, δ's value is bigger to increase convergence speed of the algorithm; on the later stage, δ's value is smaller to boost the accuracy of the algorithm.

Table 2. Calculation Method of Rotation Angle

$f \geq f_b$?	x_i	b_i	$\Delta\theta_i$	$S(\alpha_i\beta_i)$	
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$
False	0	0	0		
True	0	0	0		
False	0	1	δ	+1	-1
True	0	1	δ	-1	+1
False	1	0	δ	-1	+1
True	1	0	δ	+1	-1
False	1	1	0		
True	1	1	0		

4.3. Calculation of Fitness

The fitness examines two factors:

(1) How many hops of the longest path constituted by links contained by the chromosome;
(2) The length of the longest path. The fitness can be got through expression (3):

$$f = \frac{h^a}{l} \qquad (3)$$

Where f f is calculated from the fitness value, h is the longest path of the chromosome in the formation of the jump number, l is the most length of long path. α is adjustment coefficient, according to the specific network decision.

5. Experiment Design and Discussion

5.1 Testing Scenes and Indicators

TDCS protocol was implemented here in the NS 2.34 environment. To validate its effectiveness, we compared it with static sink node method, random roaming strategy and Peripheral protocol. In the four strategies, TDCS and random roaming adopts passive submission mode: when nodes collect data, they store them locally, which are passed to the sink node after notification. Static sink node and Peripheral take active submission pattern: when nodes acquire any data, they send actively those data to the sink node, necessity for routing protocol. Other testing parameters are set like Table3.

Table 3. Experimental Parameters Setting

Parameter	Values
Scene size	$100*100(m^2)$
Node number	100、200、300、400
Routing protocol	Stationary、Peripheral：AODV DCSR、Random：none
MAC protocol	802.11
Node initial energy	1.8J
Data acquisition speed	40KB/10s
Node communication radius	10m
Sink node moving speed	Stationary：0（m/s） Other protocols: 5、25（m/s）

In the testing, we investigated the following six indicators:

(1) Acquired data volume: the number of effective data packages arriving at the sink node; effective data packages mean those that include the acquired data;

(2) Throughput: the ratio between the total data packages in the network and network working time; all data packages include sent ones, resent ones because of conflicts, transferred ones by nodes, routing query ones, routing control ones etc.;

(3) Efficiency: the ratio between the total effective data packages received by the sink node and the totality of all data packets in the network;

(4) Residual energy: network doesn't work properly, i.e. the remaining energy of all nodes in the network in the case of the sink node unable to collect data;

(5) Energy utilization: the ratio between data amount received by the sink node and total energy of network;

(6) Average time delay: the mean value of time deference of effective data packets arriving at the converging node.

5.2 Simulation Results and Analysis

5.2.1 Acquired Data Volume

The data volumes acquired by the four protocols are compared in Figure 4, where, in Figure4(a), the sink node moves at 5m/s; and in Figure4(b) the sink node moves at 25m/s; Stationary's sink nodes are all still, the same below. From that graph, we see with increasing nodes and moving speed, network conflicts grow up and thus acquired data volumes by all solutions decline. In TDCS, since its nodes don't need to transmit data, less energy is consumed and working time is prolonged; as a result, it collects the most data. In Random, due to unreasonable paths, some nodes die earlier and some nodes can't access the network; consequently it collects fewer data. Stationary collects the fewest data for the network interrupts too early.
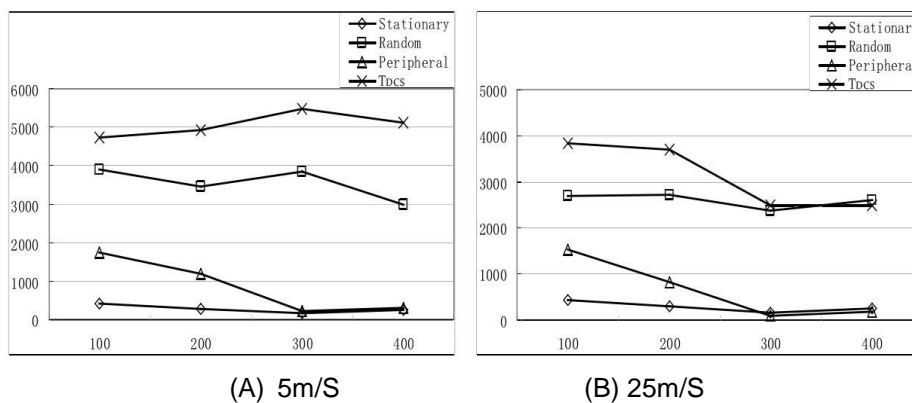

(A) 5m/S   (B) 25m/S

Figure 4. Data Collection

5.2.2 Efficiency

Figure 5 compares their collecting efficiency. TDCS has the highest efficiency thanks to no routing overheads and no needs for nodes to transmit data.
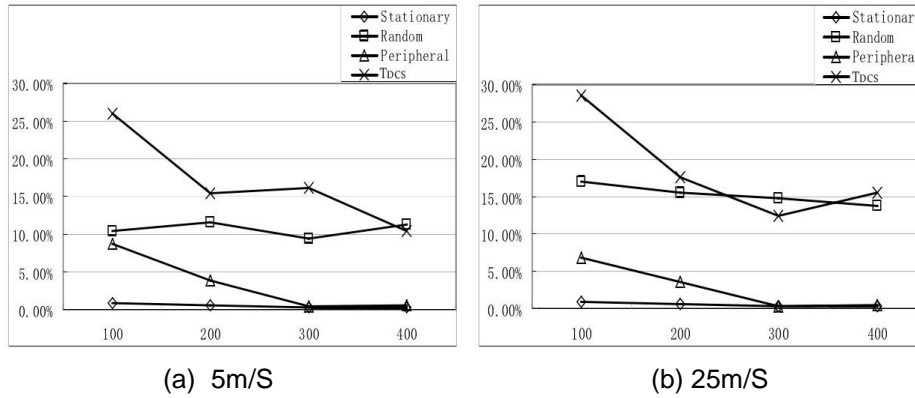
(a) 5m/S   (b) 25m/S

Figure 5. Network Efficiency

5.2.3 Throughput

The comparative results of their network handling capacity are depicted in Figure 6. Network throughput is one of the important indicators for measuring its performance. Clearly, network which takes passive submission way has better throughput than that takes active pattern, because nodes don't need to forward data and that they work ceaselessly rather than early death. For network which takes active submission mode, lots of nodes die prematurely; the sink node of a few nodes is difficultly collected, leading to lower throughput. TDCS has the highest throughput owing to more proper paths than Random and more stable data acquisition.

5.2.4 Residual Energy

The network residual energy of the four protocols is compared in Figure 7. From it, in Stationary, the sink node keeps still, causing nearby nodes to die sooner and network interruption; so it has more left energy, which is actually a waste of energy. Of three strategies which use mobile sink node, TDCS has the most reasonable paths and its nodes are not required to transmit data, a more equalized dissipation of energy. When network finishes working, its energy is almost used up, which is totally for data submission.
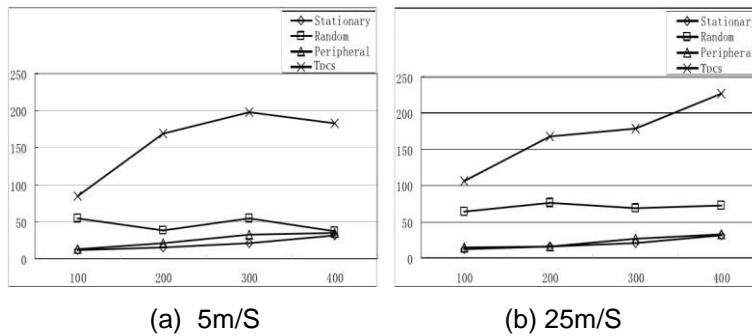

(a) 5m/S   (b) 25m/S

Figure 6. Throughput

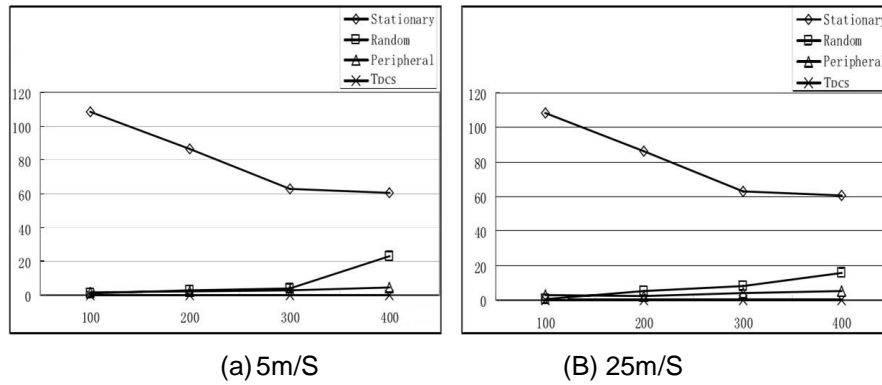(a) 5m/S   (B) 25m/S

Figure 7. Residual Energy

5.2.5 Energy Usage Ratio

Figure 8 gives the comparison of energy usage ratio of the four protocols. It actually means the effective data amount collected by the sink node when every time network consumes 1J energy.
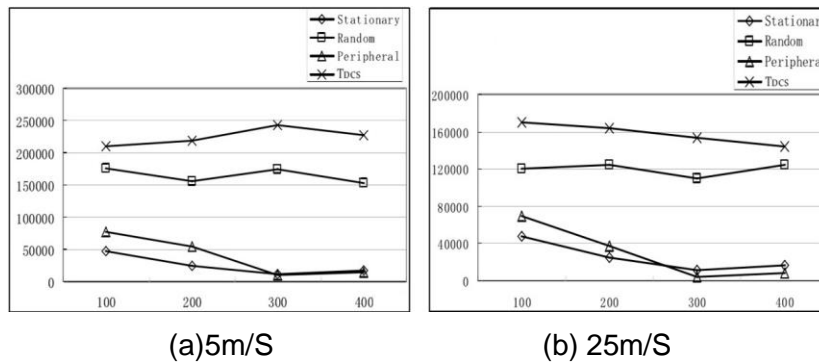

(a)5m/S   (b) 25m/S

Figure 8. Energy Usage Ratio

5.2.6 Average Time Delay

Results of average time delay are compared in Figure 9. From that we find network which takes passive submission mode has longer time lag than that with active mode, for the reason that data transmitting speed is faster than moving speed of the sink node. Random has longer time delay than TDCS, because the latter attempts to find out the shortest path, with more proper paths and less time. It should be noted that most data collection applications, especially non real-time businesses, don't require strict time delay. For them, TDCS is basically suitable.
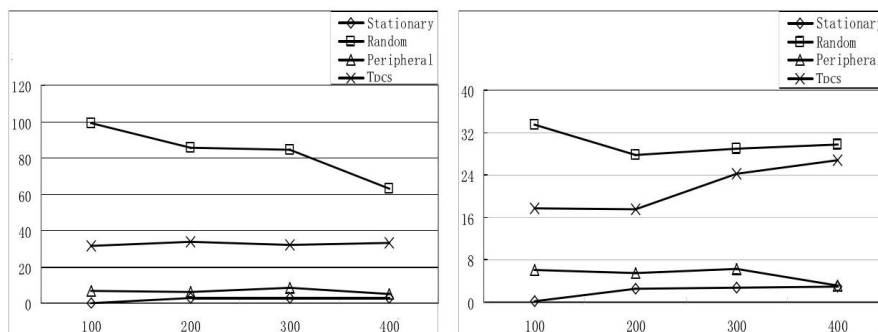

Figure 9. Average Time Delay

6. Conclusion

In data collection applications, many benefits will be brought with the use of mobile sink nodes. They help increase network throughput, equalize nodes' energy consumption and enhance data acquiring efficiency. Even more, when network becomes segmented, it can still work properly.

TDCS protocol we designed here proves well those merits. If being limited by node's mobile speed, TDCS will have certain time delay in collecting data. So for the collection of businesses which have big data volume and not high requirement for time limit, TDCS has good application prospects. Besides apparently, if these technologies like data fusion, node scheduling and parallel transmission are incorporated to sensor nodes, after they're mixed with TDCS, network will realize more favorable performance.

Acknowledgement

References

[1] C. Xia, "The data forwarding method of sensor node and mobile sink node is studied", Telecommunication technology, vol. 1, **(2012)**, pp. 45-48.
[2] C. Yuanyuan, C. Jianxin, Z. Yun and D. Yuelin, "Adaptive data acquisition strategy for mobile sink nodes", Computer science, vol. 11, **(2012)**, pp. 45-50.
[3] D. Jie, "A new type of highly efficient and cooperative mobile wireless sensor network technology research", Beijing University of Posts and Telecommunications, **(2012)**.
[4] L. Wang, "The asymptotic performance of large scale wireless sensor networks and wireless Ad-hoc networks", University of Science &amp, Technology China, **(2012)**.
[5] Y. tiantian, "Key technologies of wireless sensor networks in multi sink node", Nankai University, **(2012)**.
[6] K. Zhengjun, "Study on energy strategies of wireless sensor networks", Jilin University, **(2014)**.
[7] C. Ziyao and L. Yun, "Method for efficient data collection in wireless sensor networks for mobile sink nodes", Journal of Beijing Jiaotong University, vol. 2, **(2011)**, pp. 48-54.
[8] Y. Qianhe and Yan, "A improved design of environmental monitoring system of wireless sensor networks", Southwest University of science and technology, vol. 2, **(2012)**, pp. 54-57.
[9] F. Lei, "Research on mobile data collection mechanism for large-scale wireless sensor networks", Henan University of Science and Technology, **(2014)**.
[10] X. Jun, 'Research on the convergence of node mobility in wireless sensor networks", Beijing University of Posts and Telecommunications, **(2014)**.
[11] W. Yongxing, "Research on the routing protocol of wireless sensor networks and mobile communication networks", Beijing University of Posts and Telecommunications, **(2010)**.
[12] Makui, H. Heqing, S. Jie, D.Y. Yao, L. Haitao and Y.Z. Jie, "Based on hybrid sink wireless sensor network data collection methods", Optics and precision engineering, vol. 9, **(2008)**, pp. 1752-1758.
[13] L. Ping, "Based on mobile sink node of the wireless sensor network location information acquisition technology research. Proceedings of International Conference of China communication and information technology ICCCIT2010", China Institute of communications, youth working committee, **(2010)**, pp. 44-50.
[14] K. Akkaya, M. Younis and M. Bangad, "Sink repositioning for enhanced performance in wireless sensor networks", Computer Networks, vol. 49, no. 4, **(2005)**, pp. 512-534.
[15] J.A. Yang and Z.Q. Zhuang, "Research of Quantum Genetic Algorithm and Its Application in Blind Source Separation", Journal of Electronics, vol. 20, no. 1, **(2003)**, pp. 62-68.

Author
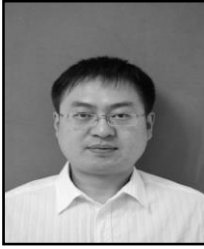
Wei Fu, He received his B.S degree from Heilongjiang University and received his M.S degree from Harbin Institute of Technology. He is a lecturer in Harbin Normal University. Her research interests include Wireless Sensor Network and Software Engineering.