Information Processing and Data Management Technology in Wireless Sensor Networks

https://doi.org/10.3991/ijoe.v14i09.8270

Yong Liu, Baohua Liang, Jiabao Jiang
Chaohu University, Chaohu, Anhui, China
ahliuyong0328@126.com

Abstract—The wireless sensor network is essentially a data-centric network that processes the Sequential stream of data, which is collected by different sensors. Therefore, the existing data management technologies regard the wireless sensor network, which is named WSN as a distributed database, and it is composed of Sequential data streams from the physical world. Wireless sensor networks are emerging next-generation sensor networks, but their transmission of information is highly dependent. The wireless sensor network processes the sequential stream of data collected by the sensor. Based on the features of wireless sensor networks, this paper presents a topology-dependent model of cluster evolution with low packet loss. Through the limited data management, resources have reasonably configured, while also saving energy. The model has based on the energy-aware routing protocol in its network layer protocols. The key point is the energy routing principle. According to its own local view, the cluster head node builds the inter-cluster topology to achieve low packet loss and energy-saving goals. Simulation results show that the model has low packet loss rate and energy efficiency.

Keywords—Wireless sensor network; Topology evolution model; Energy perception; Routing protocol

1 Introduction

In recent years, advances in microprocessors and wireless communication technologies have resulted in distributed devices capable of local information processing and wireless communication. Such small size, low cost and low energy consumption nodes have called the nodes of wireless communication. Each node cannot only deal with limited information, but many nodes with others help to observe in detail a given physical environment. Many nodes form a wireless sensor network by means of wireless communication technology. The wireless medium can be an infrared device or a radio. Unlike traditional networks, wireless sensor networks rely on intensive dissemination and collaboration to accomplish their tasks. Their purposes are to cooperatively sense, collects and process the network coverage area. Within the monitored object's information, and sent to the observer [1, 2]. It combines embedded computing
technology, sensor technology, and many other related technologies, in military, industrial, medical, and other aspects of tremendous value.

WSN is essentially a data-centric network [3]. The data collected by the sensor is called perceptual data. It has characterized by Sequential data stream with additional operations and approximate fuzzy data, and has Sequential queries. Therefore, the existing research regards the WSN database as a distributed database composed of Sequential data flows from the physical world [4]. Due to the limited computing power, storage capacity and communication capacity of nodes in WSN, and the nodes rely on battery power supply in many occasions; the battery is not replaceable and directly affects the life of the network. The perceptual data in WSN and its own characteristics of sensor nodes bring some new challenges to data management that traditional distributed database technology does not have. Specific performance is in the following areas.

1. Need to study data management techniques for WSN data characteristics. Since the sensor nodes can sequentially collect the data in the monitoring environment, the data in the WSN is often a sequential stream of data, and the statistical characteristics of the data distribution are unknown [5, 6]. The traditional distributed database data is often intermittent, the statistical characteristics of data distribution are well known.

2. There is a need to study data management techniques that balance energy consumption and response time. Energy consumption is an important technical indicator of WSN and directly affects the service life of the network. The response time is another important indicator of WSN, especially for real-time monitoring applications [7]. The energy consumption has reduced as much as possible to meet the response time requirement, which is a subject worth further study.

3. The need to study for flash memory helps to reduce energy consumption as the goal of data storage technology. As flash has some characteristics of high reliability, high density, low power consumption, sensor nodes use flash as a permanent memory. Compared with the hard disk technology, flash has many unique features; the traditional database storage technology is no longer applicable.

Aimed at many requirements and characteristics of wireless sensor network, there are many research contents around this field at present, which can divide into two parts: node level and network level. In wireless sensor networks, routing protocols are not only concerned with the data management of a single node. At the same time, WSNs are data-centric, which is most prominent in routing protocols. It is not necessary for each node to adopt the whole network addressing, and the selection path may not need node-based addressing [8]. More has based on the data of interest to establish a data source to the forwarding path between the aggregation nodes. This paper will elaborate the data management technology of WSN from the aspect of database system architecture, data storage and index, data pattern, and data mining.
2 Material and methods

2.1 Traditional database management model

The system architecture of a typical sensor network includes resource constrained sensor nodes, a resource-rich sink node, the Internet, and a user interface [9]. The distributed database system mapped to the sensor network also adopts two-layer architecture. As shown in FIG. 1, the distributed database system is composed of a local database running on a sensor node and a distributed database management layer running on a sink node and interacting with a local database.

![Wireless Sensor Network Database System Architecture](http://www.i-joe.org)

**Fig. 1.** Wireless Sensor Network Database System Architecture

The local database has three key elements: (1) Query engine. Generate an energy-efficient query plan for executing the query, which is resulting in a query result with a credible value [10]. (2) Data processing methods. Provide multi-granularity data collection for efficient data query; delete some obsolete data to save flash storage space. (3) Energy efficient storage management. It is used to achieve flash memory allocation and management, to improve the query processing speed and establish the relevant indexes. The three component instances depend on the capabilities of the node. The distributed data management layer located on the Sink node, it has also commonly referred to as a proxy database, and the database consists of two key components: a data cache that holds the data aggregates monitored by the lower nodes and the results queried from the sensor nodes [11]. The other for the query-processing engine, decide how to deal with each query command. It inquiries can use the data in the cache for local queries; the query request is sent to the corresponding node, extracting more data from the node. Wireless sensor network nodes and data processing diagram is shown in FIG. 2.


2.2 WSN information processing structure

Wireless ad-hoc network is a query-based routing mechanism. Aggregation nodes send queries through interest messages, and flood messages to all sensor nodes in the entire area or in some areas [12]. The interest message represents the task of querying and express information that the network user is interested in the monitoring area, such as environmental information, temperature, humidity, and light in the monitoring area. During the propagation of the interest message, the protocol establishes a reverse data transfer gradient from the data source to the sink-node on each hop. The sensor node transmits the collected data along the gradient direction.

Directed diffusion routing mechanism can be divided into cyclical interest diffusion, gradient establishment, and path enhancement in three stages. Figure 3 shows the data propagation path and force in these three phases.
2.3 Data Management Model in WSN

Data routing management. Data routing is one of the earliest proposed sensor network routing mechanism, so this paper will focus on analyzing and exploring it. Energy routing selects the data forwarding path based on the available power (PA) of the requirements. The available energy is the current remaining node [13]. The network is shown in Figure 3; capital letters denote nodes, such as node A. The numbers in brackets to the right of the node indicate the available energy. The bidirectional lines in the figure represent the communication links among the nodes, and the numbers on the links represent the energy consumed for sending data on the link. The source node is a general-purpose sensor node that completes the data acquisition [14, 15]. In Figure 3, possible paths from the source node to the sink node are as follows.

Path 1: Source Node B-A-aggregation Node, the sum of all nodes PA on the path is 4, and the sum of energy required for sending packets on this path is 3.
Path 2: Source node-C-B-A-sink node, the sum of all nodes PA on the path is 6, and the sum of the energy required for sending the packet on the path is 6.
Path 3: Source node-D-sink node, the sum of all nodes PA on the path is 3, and the sum of energy required for sending packets on the path is 4.
Path 4: Source node-F-E-sink node, the sum of all nodes PA on the path is 5, and the sum of energy required to send the packet on this path is 6.

Energy routing strategies are the following.

1. Maximum PA route: Select the path with the largest sum of nodes PA from all the paths, which are from the data source. In Figure 3, the sum of PA in Path 2 is the largest, but Path 2 contains Path 1, so it is not efficient to exclude, and Path 4 has selected.
2. Minimum energy consumption routing: Select the path with the least energy consumption among all the paths.
3. Least hop count routing: Select the path from the data source with the fewest hops. Select path 3 in Figure 4.
4. Maximum and minimum PA node routing. There are multiple nodes on each path, which is different from the available energy of the node, and the node with the lowest available energy in each path has selected to represent the available energy of this path [16, 17]. If the available energy of the node in path 4 is a minimum of 1, the available energy of the path is 1. The maximum and minimum PA node routing strategy is to choose the path with the largest available energy.

The energy routing algorithm shown in Figure 4 requires nodes to know the entire information. Due to resource constraints in sensor networks, nodes can only obtain local information. Therefore, the above energy routing method is only an ideal routing strategy.
Multi-path routing. The traditional network routing mechanism often selects the path with the smallest hop count, but in the wireless sensor network. If you frequently use the same-path transmission of data, it will cause nodes on the path; so that the separate parts of the entire network are not connected. Reduce the lifetime of the entire network; Rahul C. Shah et al. have proposed a kind of energy multipath routing mechanism. According to the communication energy consumption of nodes and the remaining energy of nodes, this mechanism gives a certain choice probability for each path, so that the data transmission consumes the energy of the entire network evenly, extend the life of the entire network [18]. Energy multi-path routing protocols include path establishment, data dissemination, and routing maintenance three processes. The path establishment process is the key content of this agreement. Each node needs to know all the next-hop nodes that arrive at the destination node and calculate the probability of selecting the data to transmit. The probability of selection has calculated from the node to the destination node, which has used in the following description $Cost(N_i)$ to represent the communication cost of node $i$. The main process of energy multipath routing can be briefly described as follows.

1. The destination node broadcasts message to the neighboring node to start the path establishment process. The path establishment message contains a cost field, which indicates the energy information on the path. We can set the initial value to 0.
2. When a node sends a message to its neighbor nodes, they need to forward only when they are closer to the source node and farther away from the destination node the news. Otherwise, the message will discard.
3. If the node decides to forward the path establishment message, it needs to calculate the new value to replace the old value. When the path establishment message is sent from the node $N_i$ to the node $N_j$.

$$C_{N_j,N_i} = Cost(N_i) + Metric(N_j, N_i)$$  \(1\)
\( C_{N_j,N_i} \) represents the cost that the node \( N_j \) sends the data to reach the destination node \( N_i \) through the node path, \( \text{Metric}(N_j,N_i) \) represents the communication node energy consumption from node \( N_j \) to node \( N_i \), the relationship is as follows.

\[
\text{Metric}(N_j,N_i) = e_j^\alpha R_i^\beta
\]  

(2)

\( e_j^\alpha \) denotes the energy consumption of node \( N_j \) and \( N_i \) direct communication, \( R_i^\beta \) denotes the remaining energy of node \( N_i \), \( \alpha \) and \( \beta \) is constant.

(4) Node to give up the path of too much cost, node \( j \) will node \( i \) join the local routing table conditions are as follows.

\[
\text{FT}_j = \{i | C_{N_j,N_i} \leq \alpha (\min_k (C_{N_j,N_k}))\}
\]  

(3)

\( \alpha \) is greater than 1.

(5) The node \( N_j \) uses the following formula to calculate the probability of selecting a node \( N_i \).

\[
P_{N_j,N_i} = \frac{1}{\sum_{k \in \text{FT}_j} 1/C_{N_j,N_k}}
\]  

(4)

(6) The node calculates the cost of reaching the destination node based on the entry in the selection probability. It defined as the average of the cost of the nodes \( \text{Cost}(N_j) \), which is arriving at the destination.

\[
\text{Cost}(N_j) = \sum_{k \in \text{FT}_j} P_{N_j,N_i} C_{N_j,N_k}
\]  

(5)

The node \( N_j \) replaces the original cost value in the message with the value \( \text{Cost}(N_j) \), and then broadcasts the route establishment message to the neighboring node. In the data propagation phase, for each data packet received, the node selects one node from among the plurality of next hop nodes.

3 Route simulation with the least energy consumption in data management

3.1 Construction of WSN data management platform

This method is to maintain a routing table on each node, the table marked on each destination address which link should forward. These representations have generated
Throughout the system configuration and remain static for quite some time thereafter. When the network topology is fixed and the traffic is relatively stable, this algorithm is the best. The system constructs the network node. Based on the TI-MAC underlying communication protocol, it independently develops the routing protocol supporting mobile users and constructs a double-layer communication monitoring network [19]. Any one of the research results need to be its performance analysis and evaluation of verification, the current verification methods have divided into two kinds of experimental verification and simulation. The network experiment platform can analyze the large-scale network environment. As a result, researchers have designed and developed many simulation platforms or established wireless sensor network models based on existing platforms. Figure 4 shows the WSN experiment simulation platform.

![Simulation Platform Diagram](image)

Fig. 5. Workflow of self-localization algorithms

3.2 WSN Self-Localization Algorithms

According to whether the location calculation process measures the angle or distance information between nodes, the positioning algorithms have two categories: ranging-based and non-ranging-based. Distance-based or relative azimuth measurements are based on ranging. The TOA ranging algorithm needs to measure the time that the acoustic signal experiences during the transmission of the sound source node, the sound source node and the receiving node can be calculated by multiplying the transmission time by the transmission speed. However, the TOA ranging algorithm strictly requires time synchronization in the network, which is difficult to do in underwater-complicated channels. When the calculated distance error is also relatively high, the core idea of TDOA ranging algorithm is similar to that of TOA except,
which TDOA algorithm calculates the distance. The positioning accuracy of TDOA is high. At present, large parts of positioning algorithms use TDOA to obtain the distance. Rather than distance-based algorithms, positioning errors are greater than those based on ranging, but the hardware requirements of nodes are lower and less affected by environmental factors, which make them more suitable for wireless sensor networks with limited energy.

Fig. 6. Workflow of self-localization algorithms

We use a three-dimensional coordinate system to impose constraints on the bounding box, which is based on the x and y coordinates, so that the distance measurement can be accurately made. In addition, we can also use distance as a beacon, which has also based on x and y coordinates.
Based on the above operation, we can impose constraints on the x and y coordinates, which is more conducive to accurate distance measurement, it is also known as the bounding box. In the same distance from the limits on the x and y coordinates; the same beacons are constantly changing, so the one we choose most is the bounding.

Based on the above, we impose constraints on the x and y coordinates-bounding boxes, by which we improve the accurate distance measurement and the distance to the beacon is treated as the boundary on the x and y coordinates. For the same beacon, we choose the maximum bounding boundary condition and estimate the center of the bounding box as the initial value.
3.3 The results of WSN Self-Localization test

In the C language environment, we run the program; the results have shown as follows. First, we must define the number of the wireless sensor network. Then, we enter the source node number. Finally, we build the network and data in the transmission of energy between the two nodes. Finally, the program automatically calculates the path that consumes the least energy.

Sink nodes in the mobile process, by periodically sending beacons to inform neighboring nodes around their current location. Neighbor nodes of Sink nodes can directly discover Sink nodes and start data transmission. Outer nodes indirectly dis-
cover Sink nodes and transmit data by sensing the data transmission behavior of inner nodes. Throughput reached 13%, packet loss rate of 44%, time delay of 68%, the network life expectancy of 36%. This shows that mobile Sink dynamic load balancing data aggregation algorithm in WSN data processing and data management optimal.

Each system and algorithm has their own characteristics and the scope of application, and no one is the absolute best. Overall, some circular refinement algorithms (such as Cooperative ranging, Two-phase positioning, n-hop multilateration primitive and so on) not only improve the positioning accuracy, but also give users a greater degree of freedom, they’re gradually building up a new category. According communication connectivity between unknown and beacon nodes, and the wireless range, we can calculate the region that the unknown nodes may exist. Through the above experimental tests, we will apply the method to data management in WSN positioning. Figure 11 shows the test results of estimated location.
3.4 Analysis and discussion

For wireless sensor networks, the sensor nodes need to communicate with the central forwarding node, that is, the sensor nodes have connected to the central forwarding node by single hop or multi-hop. Therefore, all nodes in the coverage area of the central forwarding node must control by the MAC protocol to co-channel the capacity of the channel. On the one hand, when the number of nodes increases, not only the signal interference between the nodes increases, but also the link conflicts at the bottom of the link are inevitable. This will result in the single node throughput being approximately $O(1/n^\alpha)$ with respect to the node number $n$. When $\alpha \geq 1$, the throughput of a single node drops sharply. However, the system capacity of the entire network generally does not increase as the number of nodes increases, and even the capacity may have reduced or even collided with all nodes. Since the sensor nodes closer to the center-forwarding node are more heavily loaded with the forwarding traffic, they easily become the "bottleneck" of the system capacity and reliability. Therefore, there exists a conflict area around the center-forwarding node in the wireless sensor network. The nodes included in this area must to send its forwarded and originating traffic competing channels, and the total competitive traffic is approximately proportional to the sum of the originating traffic of all nodes in the entire system. Each node may obtain throughput of only $O(1)$ when each node has the same initial traffic and guarantees fair access.

We must improve the system capacity of wireless sensor networks and reduce the collision-blocking rate, parallel transmission technology can be adopted, and that is, it helps to improve the ability of parallel transmission and spectral multiplexing of sensors at a central forwarding node. In addition, advanced technologies such as adaptive power control, interference cancellation or interference utilization can make it possible for adjacent nodes not to have only one channel available and to make the original
competition easier with multiple channels, which belongs to the category of resource scheduling. For wireless sensor networks, access latency is one of the key performances; multi-node forwarding means that the processing delay increases. The size of the access delay mainly depends on the node's forwarding processing delay and the link-layer competition queuing and retransmission delay. The former has related to the distribution of nodes and the coverage of the network. The latter has related to the average number of competitors per node and the competition. For wireless sensor networks, if the link adaptation strategy has not used to improve the link capacity near the access point. In addition, if the data between nodes is transmitted in parallel, the number of contending nodes in the same conflict area will be greatly reduced, and traffic balancing technology can be used to further reduce the number of link layer conflict, so that the total access delay may be smaller than single-hop access network.

4 Conclusions

Wireless sensor network is a new information acquisition and processing technology. In special fields, it has the incomparable advantages of traditional technology, and it will also open up a lot of new and valuable commercial applications. Because wireless sensor network is an emerging technology, there are still few researches done in this special field in our country. In this paper, the concept, architecture, characteristics, and research progress of wireless sensor networks have briefly introduced. Low-power wireless transceivers can effectively collect useful information from different environments to distant base stations for processing. Due to the extremely limited battery energy of the sensor nodes, communication between system nodes should utilize energy efficiently to extend the life cycle of the network. In view of this, because of digesting and assimilating the research results of others, the paper deeply analyzes the principle of energy routing in wireless sensor networks and gives a clear explanation based on the legend. The current use of the more common routing protocol made a summary and comparison. In this paper, we use a brief program simulation to illustrate the process of selecting the minimum energy path in the network. This work made a comprehensive study of wireless sensor network energy routing. However, for each specific routing procedure, the specific application environment has played a guiding role. In the future, we should do the model and quantization positioning algorithm performance evaluation, WSN needs to establish standard simulation technology and simulation system. We can also achieve low-cost, high-precision positioning system in a ultra-large-scale network and self-adjusting positioning algorithm in a mobile network environment.

5 Acknowledgment

This work is financially supported by the natural science research project of Chaohu University(No:XLY-201610), The natural science research key project of education department of Anhui Province(No:KJ2017A452).The key teaching research project of Anhui Province(No. PX-1216536).
6 References

[1] Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. IEEE Transactions on industrial informatics, 10(4), 2233-2243. https://doi.org/10.1109/TII.2014.2300753

[2] Jin, J., Gubbi, J., Marusic, S., & Palaniswami, M. (2014). An information framework for creating a smart city through internet of things. IEEE Internet of Things Journal, 1(2), 112-121. https://doi.org/10.1109/JIOT.2013.2296516

[3] Hossain, E., & Hasan, M. (2015). 5G cellular: key enabling technologies and research challenges. IEEE Instrumentation & Measurement Magazine, 18(3), 11-21. https://doi.org/10.1109/MIM.2015.7108393

[4] Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314-347. https://doi.org/10.1016/j.ins.2014.01.015

[5] Abbasi, A. Z., Islam, N., & Shaikh, Z. A. (2014). A review of wireless sensors and networks' applications in agriculture. Computer Standards & Interfaces, 36(2), 263-270. https://doi.org/10.1016/j.csi.2011.03.004

[6] Ortiz, A. M., Hussein, D., Park, S., Han, S. N., & Crespi, N. (2014). The cluster between internet of things and social networks: Review and research challenges. IEEE Internet of Things Journal, 1(3), 206-215. https://doi.org/10.1109/JIOT.2014.2318835

[7] Fadel, E., Gungor, V. C., Nassef, L., Akkari, N., Malik, M. A., Almasri, S., & Akyildiz, I. F. (2015). A survey on wireless sensor networks for smart grid. Computer Communications, 71, 22-33. https://doi.org/10.1016/j.comcom.2015.09.006

[8] Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2017). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. IEEE Systems Journal, 11(1), 88-95. https://doi.org/10.1109/JSYST.2015.2460747

[9] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. Information Systems, 47, 98-115. https://doi.org/10.1016/j.is.2014.07.006

[10] Han, Q., Liang, S., & Zhang, H. (2015). Mobile cloud sensing, big data, and 5g networks make an intelligent and smart world. IEEE Network, 29(2), 40-45. https://doi.org/10.1109/MNET.2015.7064801

[11] Aazam, M., & Huh, E. N. (2014, August). Fog computing and smart gateway based communication for cloud of things. In Future Internet of Things and Cloud (FiCloud), 2014 International Conference on (pp. 464-470). IEEE.

[12] Rawat, P., Singh, K. D., Chauoshi, H., & Bonnin, J. M. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. The Journal of supercomputing, 68(1), 1-48. https://doi.org/10.1007/s11227-013-1021-9

[13] Zhong, N., Ma, J., Huang, R., Liu, J., Yao, Y., Zhang, Y., & Chen, J. (2016). Research challenges and perspectives on Wisdom Web of Things (W2T). In Wisdom Web of Things (pp. 3-26). Springer International Publishing. https://doi.org/10.1007/978-3-319-44198-6_1

[14] Tran, Tuyen X., Mohammad-Parsa Hosseini, and Dario Pompili. "Mobile edge computing: Recent efforts and five key research directions." IEEE COMSOC MMTC Commun.-Frontiers (2017).

[15] Mahmood, M. A., Seah, W. K., & Welch, I. (2015). Reliability in wireless sensor networks: A survey and challenges ahead. Computer Networks, 79, 166-187. https://doi.org/10.1016/j.comnet.2014.12.016
7 Authors

Yong Liu, Baohua Liang, and Jiabao Jiang are with Chaohu University, Chaohu, Anhui, China.

Article submitted 19 January 2018. Resubmitted 15 and 17 April 2018. Final acceptance 01 August 2018. Final version published as submitted by the authors.