

Distributed Data Storage Strategy in Wireless Sensor Networks

<https://doi.org/10.3991/ijoe.v12i11.6238>

Zhang Song-juan ¹, Yang Jian²

¹ Nanyang Institute of Technology, Nanyang, China

²Henan Mechanical and Electrical Vocational College, Henan, China

Abstract—In order to solve the problem of large scalability and low energy efficiency in distributed data storage in wireless sensor networks, the author proposed a temporal-centric storage approach method. By using this method the sensing data are stored in some storage node indexed by a hash function parameterized with detecting time. These nodes act as rendezvous among sink and source nodes. Simulation results show that the proposed approach mitigates the hot-spot problem and can thus improve overall system performance substantially.

Index Terms—distributed data, hash function, storage strategy, wireless sensor networks

I. INTRODUCTION

Wireless Sensor Networks (WSNs) consist of a great number of sensor nodes which are composed of sensing unit, processing unit and communicating unit. Wireless sensor networks have a broad application prospect and great practical significance in human life, because WSNs are good for hiding, quick at network building, widespread, simple and inexpensive, easy to expand, small in size, large in scale, and they have flexible structure. Wireless sensor network is constituted by abundance of cheap, low-power wireless communication nodes in monitoring the region's, it is used to cooperative sensing, collecting and processing information within the network coverage area, and send the information to monitors. Data fusion system can process data fusion effectively, reduce redundant information transmission, which greatly reduce the energy consumption and increase the life time of network[1-2].

Modern wireless sensor networks (WSNs) has been improved a lot due to the development of sensor technique, micro electromechanical system, modern network and wireless communication. People can obtain physical information by dispose wireless sensor networks in certain places that human can't get close to. Therefore, it will play an important role in many areas such as military affairs, environment monitoring, industry control, city management etc. In all these applications, sensor nodes collaborate with each other to perceive, collect and process the data from the objects in the covered network area, then send the results to users for further analysis by multi-hop wireless communication. Therefore, data management is the key technique in WSN since it's a data-centric system. Real time assurance is one of the most important applications of WSN[3]. For instance, the system needs to provide rapid response time in applications such as disaster relief, military attack and defense.

People can obtain information about objective world through wireless sensor networks rather than site investi-

gation, which expands the function of existing networks and improves people's ability of recognizing the world[4]. However, some problems are still needed to research and solve because of the specificities of wireless sensor networks. For the limits of computing ability, storing capacity, communication band, battery capacity and the existence of data redundancy, it is indispensable to design suitable data compression algorithm to adapt above mentioned characters of wireless sensor networks. The data redundancy not only exists in data obtained by the same node in sequent instants but also exists in data obtained by adjacent nodes in the same instant. If nodes directly send data which includes redundancy, the limited communication band and energy won't be used efficiently and the delay of system will be prolonged and the ability of the whole system will be affected seriously.

Wireless sensor network spreads widely, and with the expansion of WSNs there will be a large number of sensors which produce massive sensor data. A two-tier storage architecture based on HBase storing sensor data from different regions and global data management directory can achieve a near real-time storage system. From the researcher's experimental results, it is shown that this system with high scalability, storage and query efficiency can solve the massive sensor data storage. In this paper, the two-tier storage architecture based on HBase storing sensor data from different regions and global data management directory is used as the basic model.

Recent advance of microprocessor and sensor technologies are able to deploy in large-scale wireless sensor networks. In some environment, sensor device are disposable. Since economical on cost to deploy sensor network, sensor nodes are relatively limited in communication and computation capability, small memory, and battery powered. Sensor network enables many applications such as vehicle tracking, battlefield surveillance, ecological and habit monitoring. The basic function of sensor node includes process, store and forward data. By observe from above application, the restriction of sensor node arise many issues of energy efficient and data dissemination. For describe our environment, we called sensor node source if it detects event and we called the user who generate query sink. This paper studies the problem of hot-spot and multidimensional search. By the restriction of limit memory, uniformly distribute data can guarantee data dropped from memory slowly and reserve more power.

II. OVERVIEW

In the data storage fields, we study the main real time feature of data, communication and query respectively and discuss some techniques that guarantee the real time data

management. And the real time requirement will be fulfilled in lots of aspects such as data model, query language, data storage and query processing etc. A data storage approach based on Multi-level mapping Index (MIS) is proposed in Park's paper [5] to support both periodic query and Ad hoc query in large scale wireless sensor network. Considering the constraint of response time for periodic query and network topology, this method divides the whole network into three layers. Each layer uses local storage, data-center storage and external storage respectively. By constructing index for each layer, the low-level data can be mapped into high-level index by multi-level index. The middle level can access the low-level data in time and restrain the transmission of data with little changing. In this way, delay constraints and energy will be satisfied according to different queries. This method will save huge energy compared with local storage, exterior storage and data-centric storage.

Based on MIS approach, a buffer-based real-time query processing method is suggested in Xu's paper [6] to solve the problem that lots of queries will miss the deadline because of long distance transmission in large scale network. A query buffer is built to store data and the query mapping window is adjusted by the deadline in order to fulfill the real time requirement. The buffer-based method abstract the query initiation, query processing and result output into the queue model. The buffer size is selected according to the input speed and service ability and the extent of each network layer can also be regulated dynamically. This approach is much more efficient in energy saving and deadline satisfaction compared with method that partition the network averagely.

A slack factor based real-time query processing algorithm is produced in Mohanasundaram's paper [7] to support real time and energy-saved query processing. The concept of slack factor is proposed to represent the constraint extent, meanwhile the concept of data transfer mode is proposed according to the query selectivity and the router strategy. The efficient real time query plan is established by these factors. Firstly, the location and order of query will be decided by selectivity and the hop threshold will be calculated by the slack factors. Then all the nodes will be divided into different sets by comparing the relation between the hop count and the threshold. After that each set can adopt the same data delivery mode. In this way, the query accuracy will be improved before deadline and the energy will be saved greatly compared with the approach that only makes use of inner-network calculation or centralized calculation.

An efficient event detection technique based on temporal-spatial correlation model is proposed in Sun's paper [8] to assure the precision of detection result. Meanwhile, the constraint of response time can be satisfied by changing the execution time dynamically. The temporal-spatial correlation mode includes two processes: temporal correlation estimation and spatial correlation estimation. The temporal correlation is used to cancel the infrequent temporal mistakes and the spatial correlation is used to cancel the permanent mistakes generated by broken nodes. Taking the feature that WSN has limited energy into account; the group will collaborated by alternate work to enhance the validation time of network.

Dini [9] designs and implements a wireless sensor network data management stimulation system (WSNDM) based on the real-time data management framework and

the exiting research work. Compared with the prototype system, WSNDM system integrates lots of necessary operations such as memory, index, query and schedule in WSN management. Besides implementing and validating all the methods and theories be proposed, the author also accomplishes many current techniques for comparison. It dedicates to study fundamental problems related to data model, query processing, real time event detection in WSNs data management and proposes corresponding technical solutions. Theoretical analysis and stimulation experiments show that such methods could efficiently guarantee real time query processing and event detection. These approaches and techniques could make some referential values to develop high-performance WSN real time data management systems.

WSNs can be used in various applications including: military, environment monitoring, health, space exploration, agriculture, intelligent home, and so on. However, there are many problems and challenges in wireless sensor networks, such as limited energy consumption, limited ability on storage, computing, and communications, low-bandwidth, etc. Therefore, researchers need to consider the various features of wireless sensor networks deeply, and make sure that wireless sensor networks can be efficiently and widely used. It is just the reason that there are a lot of protocols designed for wireless sensor networks, such as routing protocols, MAC protocols, topology control, data processing technology, and so on. However, most of the existing researches on WSNs are carried out after sensor nodes have been deployed, and some studies even completely ignore the structure of WSNs. These researches neglect that WSNs can be configurable. In fact, after sensor nodes are dispersed into the sensing field, Wang's paper [10] can take advantage of the mobility of sensor nodes, or adding some sensor nodes to rebuild the structure of WSNs, so that the performance of wireless sensor networks can enhanced significantly. This is the placement problem in wireless sensor networks that we will focus on in this paper. In fact, placement is the foundation of various applications in WSNs, because only if sensor nodes have been placed in the sensing field with specific number and positions, we could carry out the practical applications, and the design of all kinds of protocols. Placement is actually very important in wireless sensor networks, and in the past few years, there have been many placement algorithms for WSNs. Although, these placement algorithms focus on connectivity, coverage, energy consumption, congestion control, storage capacity and other issues in WSNs, there is still one or more shortcomings in these algorithms, such as: poor application background, too strong assumption, poor scalability, higher degree of complexity on computing and communications and so on. In response to this situation, Gong [11] has conducted a wide range of investigation on placement problem in WSNs. Based on the weaknesses of wireless sensor networks, which are of poor transmission reliability, limited energy, limited storage and computing ability, the lack of placement algorithms for mobile sensor networks, he designed a set of placement algorithms for wireless sensor networks.

III. METHOD

The primary goal in our scheme is solve hot-spot problem by appropriate dispose storage area, second is avoid multidimensional search problem. We will describe multi-

dimensional search problem. Our environment is efficient where large amounts of sensing data are generated, but only a small portion of them are queried. We assume a sensor field where sensor nodes are uniform deployed and time synchronized when initial process. Sensor nodes are stationary and aware of their own location with GPS device or location technology. We using Perimeter Stateless Routing (GPSR) protocol for any packet transmit in the sensor network. GPSR use two algorithms to achieve routing process. First sensor node forward packet progressively to the neighbor node that closest to destination location as greedy algorithm. If greedy forward is not work, we use perimeter forward algorithm [12].

We construct global grid structure in sensor field by odd integer G ($G > 1$). Show as Fig. 1, the global grid structure constructed by G^2 grids and $(G-1)^2$ intersection points, where we called storage area (SA).

After global grid structure constructed, we generate a storage curve in sensor field. As Fig. 2, every intersection point noted by SA_i ($i=1, 2, \dots, (G-1)^2$). Storage curve is known by all sensor nodes in the sensor network.

We define a parameter T_p that split to $(G-1)^2$ section, where each time period (TP) has a storage area (SA) associated with it, as show in Fig. 3.

In our system suppose, event detect can categorize to environment detect and target detect. We define environment detect as phenomenon detect that simply and doesn't need update frequently. For example, temperature detects maybe update with three-hour intervals, frequent temperature update is unnecessary and temperature information is simple value. We define target detect as mobile object that have large information and update frequently. For example, animal detect need frequently update for location tracking, and animal information is large with id, speed, location, direction, etc. In [10] they improve DCS method by index update to decrease traffic complexity. If the ratio of data/index is small as environment detect that we defined, the additional query forward cost will reduce network lifetime. We define environment detect and target detect to decide data update or index update that the source node should process.

At first, some relay nodes are placed to connect the whole wireless sensor network through a greedy scheme to satisfy the basic requirement of WSNs. Next, some redundant relay nodes are placed to maximize the network data storage lifetime. And then, we prove that the network data storage lifetime can achieve maximum value through our algorithms. And then, we study the storage node placement problems in wireless sensor networks, to solve the problem that WSNs have limited storage and computing ability. By placing a certain number of storage nodes to store sensing data coming from back-fence sensor nodes, the data traffic in wireless sensor networks can be reduced. We first formulate the storage node placement problems, with modifying the shortcoming of the definition in the energy model of existing methods. Finally, we study the placement problem in mobile wireless sensor networks, and we focus on the theoretical relationship between the mobile sensor density and sweep coverage.

IV. EXPERIMENT RESULT

In DCS scheme, event type is regarded as attribute and decides which place should to disposal, each attribute regarded as one dimension. The feature of data-centric

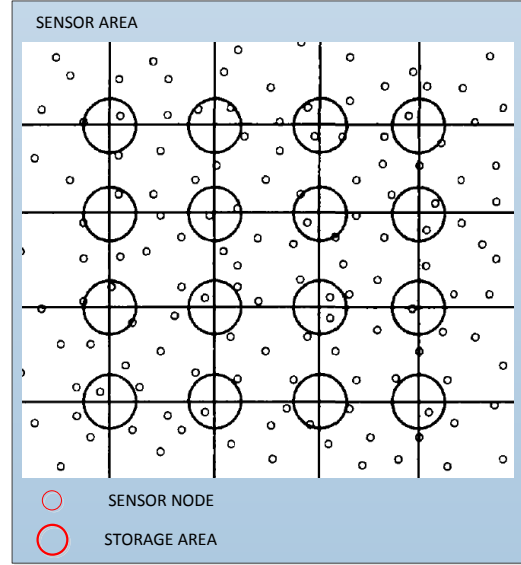


Figure 1. Globe data structure

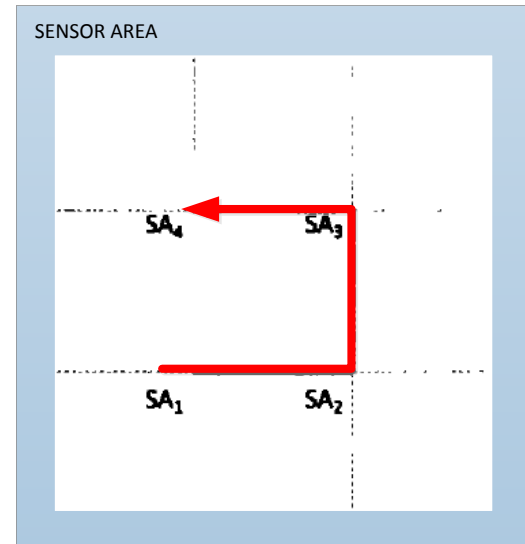


Figure 2. The curve of storage

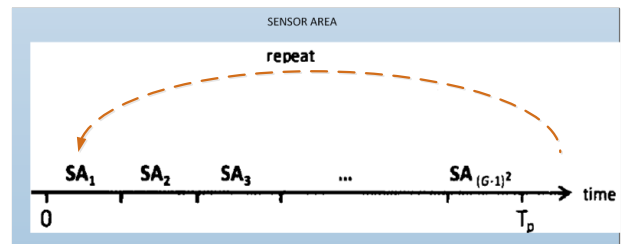


Figure 3. The relations between each time period and storage area

scheme will derive multidimensional query problem. Fig. 4 shows example of multidimensional range query. In DCS scheme, the information of temperature, humidity and snake should deliver to somewhere individual when event detected, that imply multidimensional query must deliver to three places. This feature cause more query cost and also receiving more data than necessary since each storage area can not aware of other attribute, the useful information just few part of receiving data. After multidimensional query (Fig. 4) delivered by sink, it receives $\{s1,s2,s9\}$ from temperature, $\{s1,s6,s9\}$ from humidity and

$\{s1,s4,s2,s3\}$ from snake, actually the desired information is the intersection of all receiving data, finally $\{s1\}$ is the complete result of multidimensional range query. In our method, TCS is temporal-based storage scheme, which imply multidimensional range query only deliver to one corresponding storage node. We assume all the information has already updated to storage node and query duration have not exceeds to next period. Continue use query, we will receive data without any unnecessary information since storage node has entire environment data and index information of target in storage node's period, that imply storage node is able to decide which information is complete conform with query condition. After storage node forward query to corresponding source node, sink receives pure data from source immediately.

```

SELECT *
FROM Snake
WHERE temperature >= 21 AND temperature <= 25 AND
      humidity >= 70 AND humidity <= 80
Duration [now, now+180 min]
    
```

Figure 4. Multidimensional range query

We simulated based on Java program to evaluate and compare the Temporal-Centric Storage (TCS) scheme to Data-Centric Storage (DCS). In this simulation, the communication range of each node is 50 m. 500 sensor nodes are distributed over 500x500 m² field.

We first evaluate the impact of r_{sat} , in TCS. r_{sat} , influence the number of nodes in storage area directly. Fig.5 shows the increased r_{sat} , decrease number of dropped data from memory. Leader receive event update packet and choose one of member to store data, increased number of members means used probability of member is reduced and average residual energy increased. In this simulation the network size is 500x500 m and $G=5$, range r_{sat} , cannot exceed 50m or storage area will cover each other.

Fig. 6 shows impact between query rate and hot-spot traffic. We observed the Fc of hot-spot in DCS is higher 7 Kbyte (63.2%) than TCS when $r_q=1$, and the traffic of hot-spot in DCS is higher 43 Kbyte (61.8%) than TCS when $r_q=15$. The result shows TCS improve hot-spot traffic efficient and DCS causes large hot-spot traffic since the number of storage space restricted by limited event types.

It is just the reason that there are a lot of protocols designed for wireless sensor networks, such as routing protocols, MAC protocols, topology control, data processing technology, and so on. However, most of the existing researches on WSNs are carried out after sensor nodes have been deployed, and some studies even completely ignore the structure of WSNs.

V. DISCUSSION

From the experiment data above, it is proved that the performance ratio of the our placement algorithm is no worse than $(1+[(2\sim 1/2D-2r/2R)])(\ln n - \ln \ln n + (R)(1))$ for the single-tiered wireless sensor network, and the placement algorithm for the two-tiered network is no worse than $(1+[(2\sim 1/2d)/2R])(\ln n - \ln \ln n + (R)(1))$, where D is the size of the sensing field, n is the number of initial sensor nodes, constants $R > r > 0$ are the communication radius of relay node and sensor node.

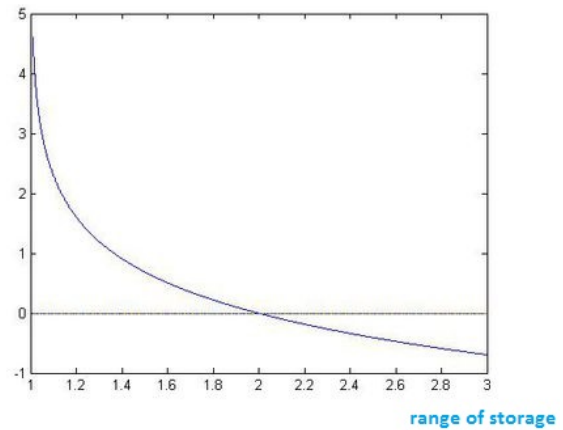


Figure 5. Impact between r_{sat} and total dropped packets in TCS

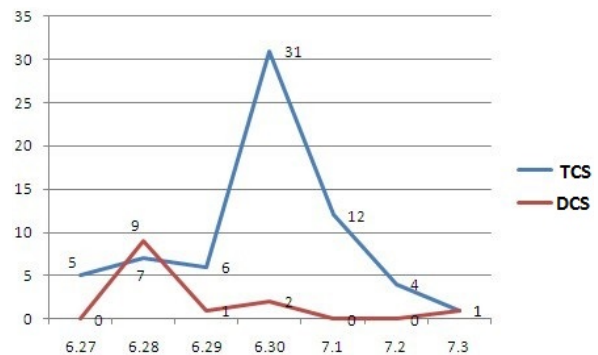


Figure 6. Impact between dimensional and total traffic

Through analyzing the impact of the positions and number of sensor nodes on network data storage lifetime, we design placement algorithm to prolong the lifetime of WSNs. It is proved that the storage node placement problem in wireless sensor networks is NP - hard; it is also proved that the performance ratio under worst-case of our algorithm is $O(\log n)$, where n is the number of sensor nodes.

Our environment is efficient where large amounts of sensing data are generated, but only a small portion of them are queried. We assume a sensor field where sensor nodes are uniform deployed and time synchronized when initial process. Sensor nodes are stationary and aware of their own location with GPS device or location technology. We using Perimeter Stateless Routing (GPSR) protocol for any packet transmit in the sensor network. GPSR use two algorithms to achieve routing process. First sensor node forward packet progressively to the neighbor node that closest to destination location as greedy algorithm. If greedy forward is not work, we use perimeter forward algorithm. So, in mobile wireless sensor networks, there is no need to monitor each point in the sensing field continuously. In fact, periodic monitoring is sufficient to meet the needs of most applications in mobile wireless sensor networks. This new coverage model is called sweep coverage, which can save a large number of sensor nodes. By the analyzing of experiment result, we explored t-sweep coverage problem in mobile wireless sensor networks, and then, we gives a clear definition on it. Next, we calculated the minimum sensor density to satisfy t-sweep coverage. It is assumed that the speed of mobile sensor nodes is v , the sensing radius is R , scanning interval is t , then in mobile wireless sensor networks, if $vt \geq 2(3\sim 1/2)R$, then the den-

sity of mobile nodes should be $O(4/3vtR)$ to satisfy t-sweep coverage, otherwise, it is $O(2/3(3\sim 1/2)R^2)$; in hybrid sensor networks with a certain number of static sensor nodes, if $vt \geq 2(2\sim 1/2)R$, the density of mobile nodes should be $O(4(2\sim 1/2)/evtR)$, otherwise, it is $O(2/eR^2)$.

Traditional external storage-based scheme relies on centralized server, which storing data outside of sensor network. The neighbor nodes of external server are used with high probability and power consume quickly. In local storage-based scheme, sensor node store data itself if it detects event. In this manner, network lifetime decrease quickly because sink deliver query with flood-response pattern. Many data dissemination scheme have recently been proposed to reduce the cost of discovery and query in wireless sensor network.

In this manner, events detected and named by type. The sensing data can be pushed to storage node that calculates by hash function on the event type. Sink deliver query to storage node by same hash function without discovery process. Index-based scheme extend behavior of data dissemination of DCS. In this manner, sensor node deliver index to storage node and store data locally if event detected.

This is efficient when large amounts of sensing data are generated, but only a small portion of them are queried. In TCS, source updates information to storage area without target data when target detected. For alternative, it updates index. TCS causes target data traffic when source deliver data to sink only. In DCS, all the sources that detect target deliver target data to storage area and cause large target data traffic.

We can observe the number of storage node is constrained by number of event type from above scheme. There is also having hot-spot problem when some event type detected frequently. The Two-Tier Data Dissemination (TTDD) method use source based grid structure to achieve data dissemination. In this manner, source constructs grid structure if event detected. Source C flood the advertisement message to entire grid structure. Sink A and B discover with local flooding pattern and deliver query to source C directly. However, TTDD consumes too much energy if targets are also mobile since the grids are generated on the basis of the source locations. In railroad scheme, the virtual infrastructure called rail which is an area where all the metadata of event data are stored. Source delivers metadata to the closest location of rail from itself. Once a query is issued, it circulates around Rail and searches relevant data stored in Rail. When a relevant metadata is found, the source node of the data transmits the corresponding data to the sink node which has issued the query. Some nodes of rail drops data quickly from memory if event detected frequent in certain location.

VI. CONCLUSION

In some environment, sensor device are disposable. Since economical on cost to deploy sensor network, sensor nodes are relatively limited in communication and computation capability, small memory, and battery powered. Sensor network enables many applications such as vehicle tracking, battlefield surveillance, ecological and habit monitoring. The basic function of sensor node includes process, store and forward data. By observe from above application, the restriction of sensor node arise

many issues of energy efficient and data dissemination. The efficient real time query plan is established by these factors. Firstly, the location and order of query will be decided by selectivity and the hop threshold will be calculated by the slack factors. Then all the nodes will be divided into different sets by comparing the relation between the hop count and the threshold. After that each set can adopt the same data delivery mode.

In this paper, we propose a temporal-centric storage method based on a global grid structure. It doesn't restrict to the number of event types as the number of storage area. When some critical event type detected frequency, temporal-centric storage scheme still avoid hot-spot problem since it disseminate data by hash on time. Normally, multidimensional range query request multiple attributes on same time, for example: "List all events whose humidity lies between 70 and 80, temperature lies between 20 and 25 and detected snake target for last 30 minute duration". In order to solve the problem of large scalability and low energy efficiency in distributed data storage in wireless sensor networks, the author proposed a temporal-centric storage approach method. By using this method the sensing data are stored in some storage node indexed by a hash function parameterized with detecting time. For this scenario, our approach stores the sensing data based on time-period and filters data centralized. It is assumed that the speed of mobile sensor nodes is v , the sensing radius is R , scanning interval is t , then in mobile wireless sensor networks, if $vt \geq 2(3\sim 1/2)R$, then the density of mobile nodes should be $O(4/3vtR)$ to satisfy t-sweep coverage, otherwise, it is $O(2/3(3\sim 1/2)R^2)$. The Fc of hot-spot in DCS is higher 7 Kbyte (63.2%) than TCS when $rq=1$, and the traffic of hot-spot in DCS is higher 43 Kbyte (61.8%) than TCS when $rq=15$. Then, the sink can receive accurate result without any unnecessary data. Simulation results show that the proposed approach mitigates the hot-spot problem and can thus improve overall system performance substantially. These nodes act as rendezvous among sink and source nodes. Simulation results show that the proposed approach mitigates the hot-spot problem and can thus improve overall system performance substantially.

REFERENCES

- [1] H. Jing, "Node deployment algorithm based on perception model of wireless sensor network," *International Journal of Automation Technology*, vol.9, no.3, pp. 210-215, April 2015. <https://doi.org/10.20965/ijat.2015.p0210>
- [2] H. Jing, "Routing optimization algorithm based on nodes density and energy consumption of wireless sensor network," *Journal of Computational Information Systems*, vol. 11, no.14, pp. 5047-5054, July 2015.
- [3] M. Li, and H. Qian, "A Routing Algorithm for Wireless Sensor Networks based on Vector Field Theory," *International Journal of Online Engineering*, vol. 12, no 03, pp. 22-27, March 2016. <https://doi.org/10.3991/ijoe.v12i03.5407>
- [4] J. Yang, S. He, Y. Lin, et al., "Multimedia cloud transmission and storage system based on internet of things," *Multimedia Tools and Applications*, pp. 1-16, October 2015. <https://doi.org/10.1007/s11042-015-2967-9>
- [5] J. Park, D. Seong, H. Kim, et al., "A data-centric storage scheme for high storage utilization in wireless sensor networks," *Cluster Computing*, vol. 18, no. 1, pp. 247-257, 2015. <https://doi.org/10.1007/s10586-014-0357-0>
- [6] J. Xu, S. Guo, B. Xiao, et al., "Energy-efficient big data storage and retrieval for wireless sensor networks with nonuniform node distribution," *Concurrency & Computation Practice & Experience*, vol. 27, no.18:5765-5779, 2015. <https://doi.org/10.1002/cpe.3645>

PAPER
DISTRIBUTED DATA STORAGE STRATEGY IN WIRELESS SENSOR NETWORKS

- [7] R. Mohanasundaram, and P. S. Periasamy, "Clustering Based Optimal Data Storage Strategy Using Hybrid Swarm Intelligence in WSN," *Wireless Personal Communications*, vol.85, no.3, pp.1381-1397, 2015. <https://doi.org/10.1007/s11277-015-2846-8>
- [8] Z. Sun, Y. Zhang, Y. Nie, et al., "CASMOc: a novel complex alliance strategy with multi-objective optimization of coverage in wireless sensor networks," *Wireless Networks*, pp. 1-22, 2016. <https://doi.org/10.1007/s11276-016-1213-3>
- [9] C. Dini, and P. Lorenz, "Prioritized Redundancy of Data Storage in Wireless Sensor Networks," *International Journal on Advances in Systems & Measurements*, vol. 4, pp. 163-185, 2015.
- [10] X. Wang, L. Xu, and S. Zhou, "Restoration strategy based on optimal relay node placement in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 8, pp: 1-13, 2015. <https://doi.org/10.1155/2015/409085>
- [11] B. Gong, et al., "Spatiotemporal Compressive Network Coding for Energy-Efficient Distributed Data Storage in Wireless Sensor Networks," *IEEE Communications Letters*, vol. 19, no.5, pp. 1-15, 2015. <https://doi.org/10.1109/LCOMM.2015.2401008>
- [12] C. Wang, et al., "Practical Spatiotemporal Compressive Network Coding for Energy-Efficient Distributed Data Storage in Wireless Sensor Networks," *Vehicular Technology Conference IEEE*, 2015. <https://doi.org/10.1109/vtcSpring.2015.7146024>

AUTHORS

ZHANG Song-juan is with the College of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang 473004, China (125980657@qq.com)

YANG Jian is with the Henan Mechanical and Electrical Vocational College, Henan, China.

Submitted 09 September 2016. Published as resubmitted by the authors 13 October 2016.