Chao GAO, Xiaoya HU, Lejiang GUO, Wei XIONG, Hongliang GAO

Huazhong University of Science and Technology

A Data Dissemination Algorithm using Multi-Replication in Wireless Sensor Networks

***Abstract**. Many data dissemination techniques have been proposed to facilitate data storage and query processing. In this paper, we propose a Multi-Replication Storage (MRS) algorithm in Wireless Sensor Networks (WSNs). In MRS scheme, sensing data is collected and stored at the home nodes, which form an s-hop dominating set of the whole network. Meanwhile, each home node has some replication nodes, when the home node receives a data, it will send data copies to the replica nodes in order to facilitate data query. So the MRS algorithm can provide timely responses to queries. Moreover, proposed data dissemination scheme also discusses load balance. Analysis and simulations are conducted to evaluate the performance of our MRS algorithm. The results show that the MRS algorithm outperforms the external storage (ES) based scheme, local storage (LS) based scheme and the data-centric storage (DCS) based scheme.*

Streszczenie:** W celu gromadzenia i przeszukiwania danych stosuje się wiele technik rozpraszania danych. W prezentowanym opracowaniu proponujemy zastosowanie w bezprzewodowych sieciach czujnikowych algorytmu MRS (gromadzenie przy pomocy wielokrotnej replikacji).W schemacie MRS dane gromadzone i magazynowane są w węzłach wewnętrznych, które tworzą s-przeskokowy układ obowiązujący w całej sieci. Każdy węzeł wewnętrzny ma kilka węzłów replikacji i, w celu ułatwienia przeszukiwania, kopiuje do nich gromadzone przez siebie dane. Tak więc proponowany schemat rozpraszania może zapewnić przeszukiwanie w odpowiednim czasie. W opracowaniu zbadano również zrównoważenie obciążenia. Przeprowadzono analizę i symulację zaproponowanego algorytmu MRS. Wyniki pokazują, że algorytm MRS przewyższa schematy oparte o gromadzenie zewnętrzne (ES), gromadzenie lokalne (LS) i gromadzenie w centrach danych (DCS). **Algorytm rozpraszania danych stosujący wielokrotną replikację w bezprzewodowych sieciach czujnikowych

Keywords: Wireless Sensor Networks, Data Storage, Data Replica, Data Dissemination.
Słowa kluczowe: Bezprzewodowa sieć czujnikowa, Gromadzenie danych, Replikacja danych, Rozpraszanie danych

Introduction

Wireless sensor networks (WSNs) provide a new opportunity for pervasive and context-aware monitoring of physical environments. A WSN is composed of numerous sensor nodes, each being a tiny wireless device that can continuously collect environment information and report to a remote sink through a multi-hop ad hoc network [1]. A WSN is usually deployed in a region of interest to observe particular phenomena or track objects inside the region. Practical applications of WSNs include, for example, habitat monitoring, health care, smart home, and surveillance [2, 3, 4]. As we know, energy and storage capacity are two kinds of the most important resources in wireless sensor networks. Because sensor nodes are typically operated by batteries and recharging is usually infeasible, it is a critical issue to extend the network lifetime by conserving their energy and storage capacity.

Many data dissemination schemes [5, 6] have been proposed for WSNs. However, all of them have some disadvantages. For example, it is very inefficient when the sensing data update very frequently in the external storage-based (ES) data dissemination. Considering a large scale WSN, the network-wide flooding may introduce significant traffic in the local storage-based (LS) scheme. Although queries need not to be flooded, a large amount of sensing data still needs to be transmitted across the whole network to those storage nodes by using data-centric storage-based (DCS) data dissemination scheme, which may introduce a huge communication overhead especially for large scale WSNs.

The main contribution of our work is that we propose a new data dissemination scheme to support scalable handling of large amount of sensing data in large scale WSNs to overcome the drawbacks mentioned above. In our scheme, sensing data is collected and stored at the home nodes, which form an s-hop dominating set of the whole network. Meanwhile, the home node have one replication node each DBC (distance between copies, which is a constant) hops, when the home node receives a data, it will send data copies to the replica nodes in order to facilitate data query. This algorithm also improves the robustness of the data. Our data dissemination framework can minimize the use of limited network and computational resources while providing timely responses to queries and ensure scalability and load balancing of communication as well as adaptivity in the presence of dynamic changes.

Multi-Replication Storage (MRS) algorithm

In this section, we describe the basic idea of the multi-replication data dissemination scheme. In this paper, we discuss static symmetric multi-hop wireless networks. The topology of a wireless sensor network is represented by a graph G = (V, E), where V is the node set and E is the edge set.

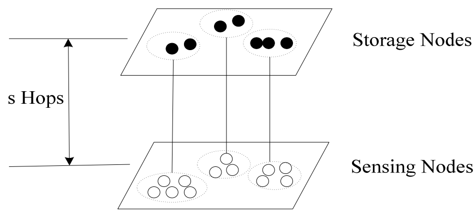

Fig. 1. Hierarchical structure of MRS algorithm

If two nodes are within the transmission range of each other, then there is an edge between them. We assume that all nodes are deployed in a 2-D plane, the area covered by a node is a circle with the radius equals this node's sensing range. There are many targets moving within a vast region. The sensor nodes in the network can detect the status of each target in its sensing range, and periodically generate sensing data. Users may issue a query via sink for data about the current activities of each target. The result report also will be returned to the user via the sink.

In our scheme, the network is logically divided into two layers as Fig.1 shows. The bottom layer contains the sensing nodes that monitor the targets and generate raw sensing data. On the top is storage nodes layer contains the home nodes and replication nodes that are used to store the data. The basic idea of how to choose home nodes is that sensing data are collected and stored at the nodes close to the sensing nodes. Hence, in our work we construct an s-hop dominating set of the whole network and use it as a home node set. Consequently, the maximum distance between the sensing nodes and home nodes is at most s hops. Therefore, the raw data do not need to travel across the whole network. Thus, it can save limited network resource.

The top layer also contains the replication nodes that store the copies information for those home node data. The main design goal of our scheme is to provide timely and efficient response to queries while minimizing the amount of network and computational resources consumed by data dissemination. Hence, we use widespread deployment of replication nodes for the home nodes in the network. According to the MRS scheme, the maximum distance between copies (DBC) is a small variable.

Wherever a query is injected into the network via a sink, it would find the nearest storage node (home node or replication node) which stores the right data information. Thus the access time from the sink to the index node is a constant which is at most DBC hops. That means our scheme can provide timely and efficient response to queries.

Home Nodes Determination. According to the definition of s-hop dominating set, every node in G is at most s hops away from at least one of the nodes in an s-hop dominating set of G. Hence, we construct an s-hop dominating set of the whole network and all the nodes in this set are used as storage nodes. In this way, sensing data are collected and stored at the nodes close to the sensing nodes. Another advantage of using an s-hop dominating set as a storage node set is that storage nodes can combine data from different sources by using functions such as suppression (eliminating duplicates), Min, Max and Average, since sensor nodes might generate a significant amount of redundant data.

Firstly, MRS chooses a root to start a breadth-first search (BFS). The root can be randomly selected or by using other leader election techniques. Then the root initializes a BFS search. After that, every node exchanges information with its s-hop neighbors about the level, degree and ID. After that all nodes should know all the s-hop neighbors' information. A node is called a leaf node if it has no children. We firstly introduce a four tuple variable which is(Level, Energy, Degree, ID), where Level is the hop distance from the root after the BFS search, Energy is the ratio of residual energy and total energy Degree is the number of the neighbors in the network, and ID is the node's unique ID. We assume that each node has a unique ID.

Firstly, a theorem is given as follow:

THEOREM 1. If L1 < L2 or L1 = L2 && E1 >E2 or L1 = L2 && E1 = E2 &&D1>D2 or L1 = L2 && E1 = E2 &&D1=D2 && ID1 < ID2, we say variable (L1, E1, D1, ID1) is larger than (L2, E2, D2, ID2).

At first, the leaf node *u* with the smallest (Level, Energy, Degree, ID) among its neighbors sends a DOMINATING message to its exact s-hop away parent *v* having the largest *(Level, Energy, Degree, ID)* to request *v* to become a dominator (storage node). Whenever *v* receives this DOMINATING message, it becomes black and broadcasts a BLACK message to all of its s-hop neighbors. Upon receiving a BLACK message from its parent, *u* becomes Gray and broadcasts a GRAY message to all of its s-hop neighbors. Then, the next node with the smallest (*Level, Energy, Degree, ID*) among its s-hop neighbors that have not decided their status yet starts this procedure in turn. It becomes Gray and broadcasts a GRAY message to its s-hop neighbors if it has received at least one BLACK message. Otherwise, it sends a DOMINATING message to its exact s-hop away parent with the largest (Level, Energy, Degree, ID) as *u* does.

This procedure stops when the root becomes Gray or Black. After that, all black nodes are used as home nodes that can s-hop dominate the whole graph as well. As shown in Fig.2.

Replication Nodes Determination. We use the Pampa[7] broadcast algorithm to identify replication of Storage nodes. In comparison with other broadcast algorithms, Pampa reduces the number of nodes required to transmit a message by having nodes more distant to the previous forwarder to broadcast the message earlier. Nodes closer to the source (i.e. those whose expected additional coverage would be smaller) do not retransmit. Pampa does not require devices to be aware of their location or of the location of their neighbors. Instead, each node uses the received signal strength indicator (RSSI) of the first retransmission heard to set a hold period. The hold period is set such that nodes with a lower RSSI expire their timers first. During the hold period, nodes count the number of retransmissions heard and, at the end of the hold period, they do not retransmit the message if a predefined threshold was reached (in this paper we use a threshold value of 2). Owing to the store-and-forward nature of the algorithm, Pampa is not used as a black-box. We discuss how Pampa was adapted for our purposes in the following paragraphs.

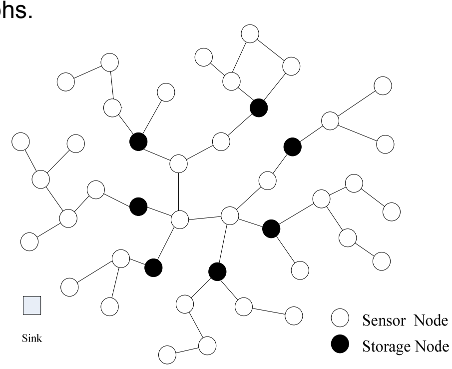

Fig. 2. Home nodes determination where s = 3

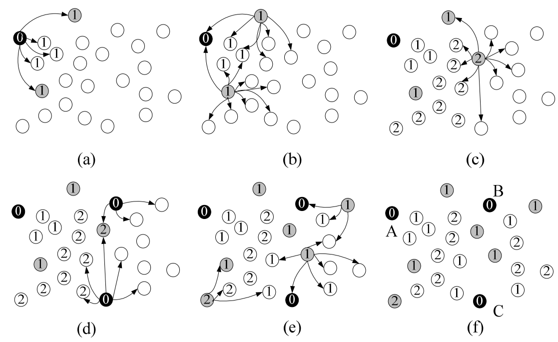

Fig.3. An example of a home node find its replication nodes

An example of the dissemination of an item is depicted in Fig. 3. The dissemination begins with the broadcast of a registration message using Pampa. The item is stored at the producer and included in the message (Fig.3. (a)). The figure depicts in black the nodes that store a replica of the item. Registration messages carry a hop from storage node (HFS) field, which records the distance (in number of hops) from the node sending the message to the closest known copy. The HFS for the message to be forwarded by each node is depicted at the centre of the node.

Fig.3.(b) and (c) show the progress of the dissemination Nodes decide whether to forward the message after the small hold period imposed by Pampa. During the hold period, each node computes the lowest value of all the HFS fields it has received in a variable named minHFS. In the figure, nodes that forward a registration message but did

not store the data item are depicted in grey. When forwarding a registration message, a node sets the TFS field to minHFS + 1, accounting for the additional hop needed to reach the closest copy of the item.

Central to our method is a constant named distance between copies (DBC). The DBC dictates the maximum value of the HFS field and, implicitly, the degree of replication of the items. DBC is expected to be small. In this example, we use DBC =2. Fig.3.(d) shows that a node with minHFS = DBC at the end of the hold period stores a copy of the item and retransmits the message. The HFS of the message is reset to 0 to let other nodes learn about the newly stored copy and update their minHFS variables accordingly (for example, see Fig.3.(e)). Recall that the holding period decreases as the distance from the source of the retransmission increases. This contributes to an improved performance of DSS as it increases the metric distance between the replicas. Because more distant nodes will expire their timers first, they are more likely to store a copy and perform a retransmission, resetting the minHFS of the nodes in the neighborhood. The final state of the system after the dissemination of the item is depicted in Fig.3.(f).Although only a small number of nodes have stored the item, a replica is stored at no more than DBC hops away from any of the nodes.

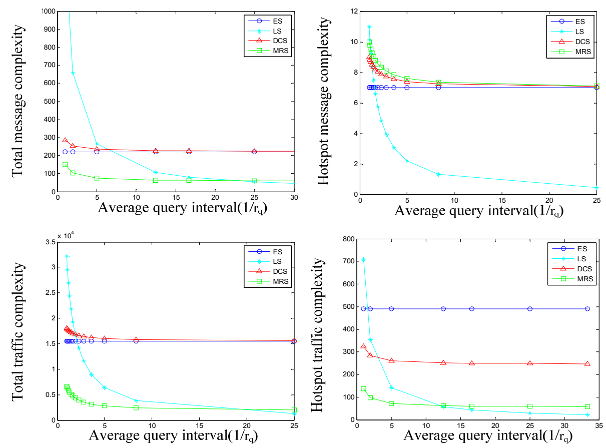

Fig.4. Message complexity and traffic complexity of MRS, ES, LS and DCS

Experimental results

We consider the following metrics when comparing the overhead of the performance of MRS with other data dissemination schemes, ES, LS and DCS:

Total message complexity: the number of messages generated in the whole network.

Hotspot message complexity: the maximum number of messages sent, received and forwarded by one single node in the network.

Total traffic complexity: the amount of data transferred in the whole network.

Hotspot traffic complexity: the maximum amount of data sent, received, and forwarded by one single node in the network.

We choose s = 3 and DBC = 5 and the total number of nodes is 1000 in this simulation. The size of a data message s_d is 70. The size of a query message s_q is 10. The size of an index update message s_i is 10. 10 mobile targets are randomly generated in the network. Each target randomly picks up a direction to move. It returns along the previous path whenever it moves out the boundary of monitored area. A sensor node can detect the target whenever the target is in its sensing range. We also assume that the result of one target is returned for each query. The simulation duration time is 200.

From Fig. 4, we can observe that the total message complexity of MRS is small when r_q is low. Even though the total message complexity of MRS becomes larger as r_q increases, the hotspot message complexity is still as good as DCS. It is obvious that the average network lifetime is decided by the total and hotspot traffic complexities.

We can improve the network lifetime by minimizing the use of limited network and computational resources. Hence, total and hotspot traffic complexities are two important metrics to evaluate the performances of the proposed schemes. From Fig. 4, we also can observe that when r_q is large, both of the total traffic complexity and hotspot traffic complexities of MRS are the smallest ones. For small r_q, the total traffic and hotspot traffic complexities of LS are smaller than others. However, MRS is still better than ES and DCS.

Conclusions

In this paper, we propose a multi-replication based data dissemination and storage method to support scalable handling of large amount of sensing data in large scale wireless sensor networks. Our MRS scheme can provide timely responses to queries while minimizing the use of limited network and computational resources. This data dissemination scheme ensures scalability and load balancing of communication as well as adaptivity in presence of dynamic changes. Analysis and simulation results show that the MRS scheme outperforms the ES, LS and the DCS schemes in overall performance.

Our future work is to address handling updates and removal of data items. Another direction of future work is how to maintain MRS data dissemination scheme in presence of network dynamic changes.

Acknowledgments

This work is supported in part by a grant from the National Science Foundation of China (No.6080202, No.60773190) and the Fundamental Research Founds for the Central University of China (No.2011TS140).

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci. A survey on sensor networks. *IEEE Comm. Magazine*, 40(2002) 102-114.

[2] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen. The gator tech smart house: A programmable pervasive space, *Computer*, 38(2005), 50-60.

[3] C.C.Y. Poon, Y.T. Zhang, and S.D. Bao. A novel biometrics method to secure wireless body area sensor networks for telemedicine and m-health. *IEEE Comm. Magazine*, 44(2006), 73-81.

[4] Y.C. Tseng, Y.C. Wang, K.Y. Cheng, and Y.Y. Hsieh. iMouse: an integrated mobile surveillance and wireless sensor system *Computer.* 40(2007) 60-66.

[5] Intanagonwiwat, C., Govindan, R., Estrin, D., Heidemann, J., Silva, F.. Directed diffusion for wireless sensor networking. *IEEE/ACM Trans. Netw*. 11(2003), 2–16.

[6] Ratnasamy, S., Karp, B., Shenker, S., Estrin, D., Govindan, R., Yin, L., Yu, F.. Data-centric storage in sensornets with GHT, a geographic hash table. *Mobile Netw. Appl*. 8(2003), 427–442.

[7] Miranda H, Leggio S, Rodrigues L, Raatikainen K. A power-aware broadcasting algorithm. Proceedings of the 17th IEEE Symposium on Personal, Indoor and Mobile Radio Communications, Helsinki, Finland(2006).

Authors: *prof. Xiaoya Hu (Corresponding Author), Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China, E-mail:* huxy@mail.hust.edu.cn; *dr Chao Gao, Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China, E-mail:* gaochao0521@gmail.com.