International Journal of Computer Network and Security(IJCNS)
Vol 4. No 1. Jan-Mar 2012 ISSN: 0975-8283
www.ijcns.com

Deploying Storage Node Aiming to Minimize the Communication Cost of Sink in Sensor Networks

Vijayalakshmi.G[1], Tamilarasi.D[2]
[1]*Department of ECE, Vel Tech Engineering College, Affiliated to Anna University, Chennai, India*
[2]*Department of ECE, Vel Tech Engineering College, Affiliated to Anna University, Chennai, India*
[1]vijigs.vijaya@gmail.com, [2]tamilarasi8884@yahoo.co.in

***Abstract*__Data storage has became very important issue in Sensor networks for future information retrieval. Data Storage is happens via the Forwarding nodes and Storage nodes. Forwarding nodes are regular sensors and they always forward the data received by other nodes or generated by themselves along a path toward the sink. Storage nodes are much larger capacity than the regular sensors. They store all the data received by other nodes or generated by themselves. The sink itself considered as a storage node. Storage nodes are introduced in this paper to store collected date from the sensors in their proximities, It reduce the energy cost and communication cost induced by network query. Aim of the project is to deploy the storage nodes with fixed tree model and dynamic tree model. If any failure occurs in storage nodes can replace another storage node or back up those data's to sink via storage node.**

***Keywords*: Wireless Sensor Networks, Data Storage, Forwarding Nodes, Storage Nodes, Data Query.**

I. Introduction

A wireless sensor network is a collection of nodes organized into a cooperative network. Each node consists of processing capability (one or more microcontrollers, CPUs or DSP chips), may contain multiple types of memory (program, data and flash memories), have a RF transceiver (usually with a single omni-directional antenna), have a power source (e.g.,batteries and solar cells), and accommodate various sensors and actuators. The nodes communicate wirelessly and often self-organize after being deployed in an ad hoc fashion. Systems of 1000s or even 10,000 nodes are anticipated. Such systems can revolutionize the way we live and work. Currently, wireless sensor networks are beginning to be deployed at an accelerated pace. It is not unreasonable to expect that in 10-15 years that the world will be covered with wireless sensor networks with access to them via the Internet. This can be considered as the Internet becoming a physical network. This new technology is exciting with unlimited potential for numerous application areas including environmental, medical, military, transportation, entertainment, crisis management, homeland defense, and smart spaces. Since a wireless sensor network is a distributed real-time system.

Sensors integrated into structures, machinery, and the environment, coupled with the efficient delivery of sensed information, could provide tremendous benefits to society. Potential benefits include: fewer catastrophic failures, conservation of natural resources, improved manufacturing productivity, improved emergency response, and enhanced homeland security. However, barriers to the widespread use of sensors in structures and machines remain. Bundles of lead wires and fiber optic "tails" are subject to breakage and connector failures. Long wire bundles represent a significant installation and long term maintenance cost, limiting the number of sensors that may be deployed, and therefore reducing the overall quality of the data reported.

Wireless sensing networks can eliminate these costs, easing installation and eliminating connectors. The ideal wireless sensor is networked and scaleable, consumes very little power, is smart and software programmable, capable of fast data acquisition, reliable and accurate over the long term, costs little to purchase and install, and requires no real maintenance. Selecting the optimum sensors and wireless communications link requires knowledge of the application and problem definition. Battery life, sensor update rates, and size are all major design considerations. Examples of low data rate sensors include temperature, humidity, and peak strain captured passively. Examples of high data rate sensors include strain, acceleration, and vibration. Recent advances have resulted in the ability to integrate sensors, radio communications, and digital electronics into a single integrated circuit (IC) package. This capability is enabling networks of very low cost sensors that are able to communicate with each other using low power wireless data routing protocols.

A wireless sensor network (WSN) generally consists of a base station (or "gateway") that can communicate with a number of wireless sensors via a radio link. Data is collected at the wireless sensor node, compressed, and transmitted to the gateway directly or, if required, uses other wireless sensor nodes to forward data to the gateway. The transmitted data is then presented to the system by the gateway connection. The purpose of this chapter is to provide a brief technical introduction to wireless sensor networks and present a few applications in which wireless sensor networks are enabling.

A wireless sensor node (or simply sensor node) consists of sensing, computing, communication, actuation, and power components. These components are integrated on a single or multiple boards, and packaged in a few cubic inches. A WSN usually consists of tens to thousands of such nodes that communicate through wireless channels for information sharing and cooperative processing. WSNs can be deployed on a global scale for environmental monitoring and habitat study, over a battle field for military surveillance and reconnaissance, in emergent environments for search and rescue, in factories for condition based maintenance, in buildings for infrastructure health monitoring, in homes to realize smart homes, or even in bodies for patient monitoring. After the initial deployment (typically ad hoc), sensor nodes are responsible for self-organizing an appropriate network infrastructure, often with multi-hop connections between sensor nodes. The onboard sensors then start collecting acoustic, seismic, infrared or magnetic information about the environment, using either continuous or event driven working modes.

Sensor networks have a wide variety of applications and systems with vastly varying requirements and characteristics. The sensor networks can be used in Military environment, Disaster management, Habitat monitoring, Medical and health care, Industrial fields, Home networks, detecting chemical, Biological, radiological, nuclear, and explosive material etc. Deployment of a sensor network in these applications can be in random fashion (e.g., ropped from an airplane) or can be planted manually (e.g., fire alarm sensors in a facility).

For example, in a disaster management application, a large number of sensors can be dropped from a helicopter. Networking these sensors can assist rescue operations by locating survivors, identifying risky areas, and making the rescue team more aware of the overall situation in the disaster area.

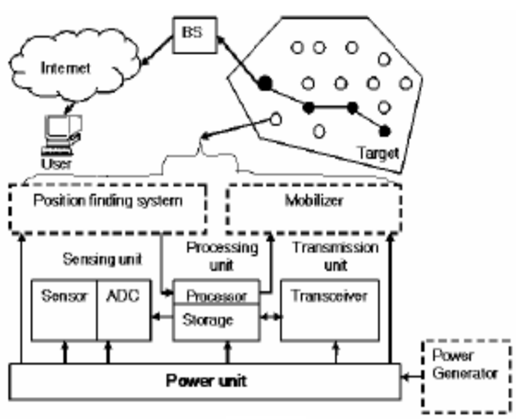


Fig 1. Structural view of Sensor network

Figure 1 shows the schematic diagram of sensor node components in which sensor nodes are shown as small circles. Basically, each sensor node comprises sensing, processing, transmission, mobilizer, position finding system, and power units (some of these components are optional like the mobilizer). The same figure shows the communication architecture of a WSN. Sensor nodes are usually scattered in a sensor field, which is an area where the sensor nodes are deployed. Each of these scattered sensor nodes has the capability to collect and route data either to other sensors or back to an external base station(s). A base-station may be a fixed node or a mobile node capable of connecting the sensor network to an existing communications infrastructure or to the Internet where a user can have access to the reported data. In general, classification of a WSN routing methodology can be done into two main categories; based on network structure or based on the protocol operation. Depending on the network structure, different routing schemes fall into this category. A sensor network can be non hierarchical or flat in the sense that every sensor has the same role and functionality. Therefore the connections between the nodes are set in short distance to establish the radio communication. Alternatively, a sensor network can be hierarchical or cluster-based hierarchical model, where the network is divided into clusters comprising of number of nodes.

II. RELATED WORK

In the literature, several schemes have introduced an intermediate tier between the sink and sensors. LEACH is a clustering-based routing protocol, where cluster heads can fuse the data collected from its neighbors to reduce communication cost to the sink. However, LEACH does not address storage problem. Data-centric storage schemes, as another category of the related work, store data on different places according to different data types. In the authors propose a data-centric storage scheme based on Geographic Hash Table, where the home site of data is obtained by applying a hash function on the data type. Another practical improvement is proposed in by removing the requirement of point-to-point routing. Ahn and Krishnamachari analyze the scaling behavior of data-centric query for both unstructured and structured (e.g., GHT) networks and derive some key scaling conditions. GEM is another approach that supports data-centric storage and applies graph embedding technique to map data to sensor nodes. In general, the data centric storage schemes assume some understanding about the collected data and extra cost is needed to forward data to the corresponding keeper nodes.

In our paper, we do not assume any prior knowledge about the data: indeed in many applications, raw data may not be easily categorized into different types. To transmit the collected data to a remote location is also considered expensive because the total collected data may be in a very large quantity. To facilitate data query, Ganesan et al. propose a multi resolution data storage system, DIMENSIONS, where data are stored in a degrading lossy model, i.e., fresh data are stored completely while long-term data are stored lossily. In comparison, our scheme is more general without any assumption about the data correlation. PRESTO is a recent research work on storage architecture for sensor networks. A proxy tier is introduced between sensor nodes and user terminals and proxy nodes can cache previous query responses. Compared to the storage nodes in this paper, proxy nodes in PRESTO have no resource constraints in terms of power, computation, storage, and communication.

A. *The Proposed Approach(CBHRP)*

In this paper an optimum energy efficient cluster based hierarchical routing protocol for wireless sensor network is proposed, which is a two layer protocol where a number of cluster cover the whole region. Proposed protocol introduces a concept of head-set instead of a cluster head. At one time, only one member of head-set is active and the remaining are in sleep mode. Several states of a node are found in this protocol such as- ***candidate state, non-candidate state, active state, associate state, and passive associate state***. This protocol divides the network into a few real clusters including an active cluster head and some associate cluster heads. For a given number of data collecting nodes, the headset members are systematically adjusted to reduce the energy consumption, which increases the network life.

B. *Communication in LEACH Protocol*

In LEACH the operation is divided into rounds, during each round a different set of nodes are cluster-heads (CH). Nodes that have been cluster heads cannot become cluster heads again for P rounds. Thereafter, each node has a 1/p probability of becoming a cluster head in each round. At the end of each round, each node that is not a cluster head selects the closest cluster head and joins that cluster to transmit data.

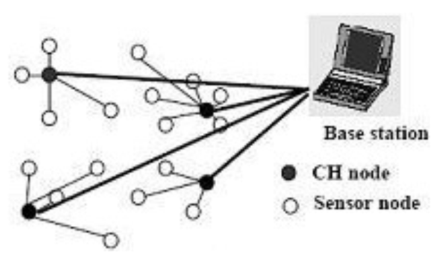


Fig 2. Communication in LEACH Protocol

The cluster heads aggregate and compress the data and forward it to the base station, thus it extends the lifetime of major nodes. In this algorithm, the energy consumption will distribute almost uniformly among all nodes and the non-head nodes are turning off as much as possible. LEACH assumes that all nodes are in wireless transmission range of the base station which is not the case in many sensor deployments. In each round, LEACH has cluster heads comprising 5% of total nodes. It uses Time Division Multiple Access (TDMA) as a scheduling mechanism which makes it prone to long delays when applied to large sensor networks.

C. *Architecture of CBHRP*

In the proposed model, the number of clusters k and nodes n are pre-determined for the wireless sensor network. Iteration consists of two stages: an election phase and a data transfer phase. At the beginning of election phase, a set of cluster heads are chosen on random basis. These cluster heads send a short range advertisement broadcast message. The sensor nodes receive the advertisements and choose their cluster heads based on the signal strength of the advertisement messages. Each sensor node sends an acknowledgment message to its cluster head. Moreover, in each iteration, the cluster heads choose a set of associate heads based on the signal strength of the acknowledgments. A head-set consists of a cluster head and the associates. The head-set member is responsible to send messages to the base station. Each data transfer phase consists of several epochs. Each member of head-set becomes a cluster head once during an epoch. A round consists of several iterations. In one round, each sensor node becomes a member of head-set for one time. All the head-set members share the same time slot to transmit their frames.

Each sensor node joins the network as a candidate. At the start of each iteration, a fixed number of sensor nodes are chosen as cluster heads; these chosen cluster heads acquire the active state. By the end of election phase, a few nodes are selected as members of the head-sets within a cluster; these nodes acquire associate state where one of them is in active state and the remaining is in associate state. In an epoch of a data transfer stage, the active sensor node transmits a frame to the base station and goes to the passive associate state. At that time the next associate member acquires the active state. Therefore, during an epoch, the head-set members are distributed as follows: one member is in active state, a few members are in associate state, and remaining are in passive associate state. At the time of last frame transmission of an epoch, one member is active and the remaining are passive associates; there is no member in an associate state. For the start of next epoch, one head-set member acquires active state and the remaining are associate.

TABLE I
NUMBER OF COMPONENTS

r_d / s_d: rate of data generation / size of each data
r_q / s_q: rate of user queries / size of each query message
α: data reduction rate (query reply size / raw data size)
n: total number of sensors
k: total number of storage nodes (for LIMITED problem)
T_i: the subtree rooted at node i
d_i: the depth of node i
c_i: the number of node i's children
$e(i)$: energy cost of node i
$E(i)$: energy cost of all the nodes in T_i
λ / λ_s: density of regular sensors / storage nodes
e_{tr} / e_{re}: energy cost for transmitting / receiving a unit data

III. Problem Formulation

In this paper, we consider an application in which sensor networks provide real-time data services to users. A sensor network is given with one sensor identified as the sink (or base station) and each sensor generating (or collecting) data from its environment. Users specify the data they need by submitting queries to the sink and they are usually interested in the latest readings generated by the sensors1. There are two
types of sensors (or nodes) in this hybrid network, defined as follows.

• **Storage nodes**: This type of nodes stores all the data it has received from other nodes or generated by themselves. The sink only sends queries to storage nodes. According to the query description, storage nodes obtain the results needed from the raw data they are holding and then send these results back to the sink. The sink itself is considered as a storage node.

• **Forwarding nodes**: Each forwarding node is associated with a storage node. A forwarding node always forwards the data generated by itself to the associated storage node. Since forwarding nodes are not aware of queries, the forwarding operation is independent of queries and there is no data processing at these nodes.

Since storage nodes hold raw data sent from nearby forwarding nodes, it requires a large local disk space (flash memory), which makes storage nodes more expensive than normal forwarding nodes. Considering the total budget of a sensor network, we probably can afford only a limited number of storage nodes (a small fraction of all the deployed sensors). Thus, given an input parameter k, our goal is to strategically allocating at most k storage sensors in a sensor network to minimize the energy cost (power consumption) associated with raw data transfers, query diffusion, and query replies. In the deployment, we first deploy normal forwarding nodes. After collecting their location information, we select at most k of them to be storage nodes. We can attach large flash memory to these selected forwarding nodes or replace them by deploying more powerful storage nodes at the same locations.

We also associate each forwarding node with a storage node which will hold the raw data from the forwarding node. We broadcast the association information to the network in the initial phase. In this model (shown in Fig. 6), queries are only diffused to every storage node. Thus, in
the following of this paper, energy cost includes transmission cost of the raw data and query reply cost but not query diffusion cost.

IV. Simulation Results

In our simulation, we consider a network of sensors deployed on a disk of radius 5 with the sink placed at the center. One thousand sensor nodes (n ¼ 1; 000) are deployed to the field randomly following two-dimensional spatial Poisson process. Node transmission range is set to 0.65. After all nodes are deployed, a routing tree rooted at the sink is constructed by flooding a message from the sink to all the nodes in the network. As we mentioned in Section 3, the message carries the number of hops it travels so that each node chooses among its neighbors the node that has the minimum number of hops to be its parent. This tree topology is needed in the simulation of the fixed tree model. This step, however, can be skipped for the dynamic tree model.

- FT-DD: It represents the fixed tree model with deterministic deployment. In FT-DD, the storage nodes are deployed by following the dynamic programming algorithm according to the known tree topology.
- FT-RD: It represents the fixed tree model with random deployment. In FT-RD, we randomly select a certain number of nodes in the network to be storage nodes.
- DT-RD: It represents the dynamic tree model with random deployment. In this algorithm, the storage nodes are randomly deployed. After that, each forwarding node selects the best storage node to deliver data and each storage node replies to query by following the shortest path to the sink.
- ST-RD: It represents semi dynamic tree model with random deployment, which is the enhanced version of FT-RD with a local adjustment. When a sensor i is upgraded to storage node in a tree structure, its siblings' children will try to set i as their parents if I is within their communication range.
- Greedy: It represents a greedy algorithm where the most heavily loaded sensors will be upgraded to storage nodes. Usually, those sensors close to the sink will become storage nodes in this algorithm.

In the deployment, we first deploy normal forwarding nodes. After collecting their location information, we select at most k of them to be storage nodes. We can attach large flash memory to these selected forwarding nodes or replace them by deploying more powerful storage nodes at the same locations. We also associate each forwarding node with a storage node which will hold the raw data from the forwarding node. We broadcast the association information to the network in the initial phase.

We make the following assumptions about the characteristics of data generation, query diffusion, and query reply.

- First, for data generation, assume that each node generates rd readings per time unit and the data size of each reading is sd.
- Second, for query rate, assume that rq queries of the same type are submitted from users per time unit.
- Third, for query reply, assume that the size of data needed to reply a query is a fraction α of that of the raw data.

Specifically, we define a data reduction function f for query reply. For input x, which is the size of the raw data generated by a set of nodes, function $f(x) = \alpha x$ for $\alpha \in (0, 1]$ returns the size of the processed data needed to reply the query. In this paper, we consider multi-hop communication for relaying data.

We assume the data routing between a pair of sensors, e.g., a normal sensor and a storage node, a storage node and the sink, follows the geographic routing algorithm, which looks for the shortest path connecting them. Thus, the energy cost model is simplified by the assumption that the transmission cost is proportional to the data size and the hop distance between the sender and the receiver. In a densely deployed sensor network, the hop distance between two sensors is proportional to the Euclidean distance. Therefore, in this paper, we use

· **Euclidean distance X Data size**

to measure the energy consumed to send data.

Fig.3 Energy cost with varying number of storage nodes

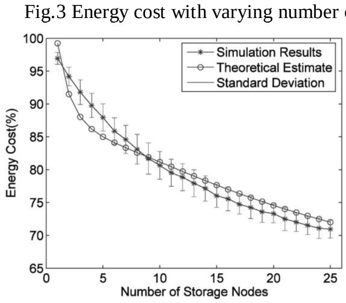


Fig. 4 The impact of data reduction rate

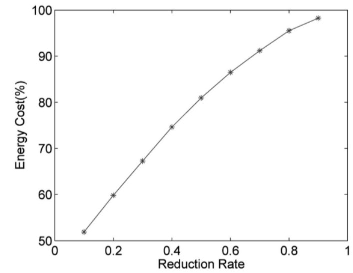
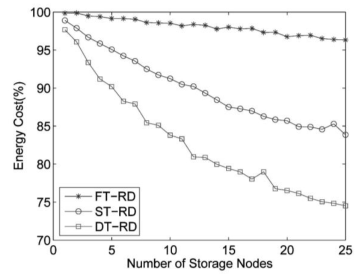


Fig. 5 Comparison of energy cost with varying Number of storage nodes (k)



V. Conclusion

This paper considers the storage node placement problem in a sensor network, introducing storage nodes into the sensor network alleviates the communication burden of sending all the raw data to a central place for data archiving and facilitates the data collection by transporting data from limited number of storage nodes. In this paper, we examine how to place storage nodes to save energy for data collection and data query and how to back up the data's when the storage node got failure due to some environmental defects. We also implement the algorithm and conduct simulation on different network parameters. Our simulation shows that the performance of our approximation algorithm is very close to optimal when the number of storage nodes is small. Our future work includes how to optimize query reply and replacing the storage nodes in a sensor network and how to solve the storage node placement problem in terms of other performance metrics.

Acknowledgment

We wish to extend our gratitude individually to all those who have helped us to complete the work successfully.

References

[1] P. Gupta and P.R. Kumar, "The Capacity of Wireless Networks," IEEE Trans. Information Theory, vol. 46, no. 2, pp. 388-404, Mar. 2000.

[2] E.J. Duarte-Melo and M. Liu, “Data-Gathering Wireless Sensor Networks: Organization and Capacity,” Computer Networks (COMNET), vol. 43, no. 4, pp. 519-537, Nov. 2003.
[3] R.C. Shah, S. Roy, S. Jain, and W. Brunette, “Data MULEs: Modeling a Three-Tier Architecture for Sparse Sensor Networks,” Proc. First IEEE Int’l Workshop Sensor Network Protocols and Applications (SPNA), May 2003.
[4] B. Sheng, Q. Li, and W. Mao, “Data Storage Placement in Sensor Networks,” Proc. ACM MobiHoc, pp. 344-355, 2006.
[5] S. Madden, M.J. Franklin, J.M. Hellerstein, and W. Hong, “TAG: A Tiny Aggregation Service for Ad-Hoc Sensor Networks,” SIGOPS Opererating Systems Rev., vol. 36, no. SI pp. 131-146, 2002.
[6] S. Madden, M.J. Franklin, J.M. Hellerstein, and W. Hong, “The Design of an Acquisitional Query Processor for Sensor Networks,” Proc. ACM SIGMOD, pp. 491-502, 2003.
[7] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-Efficient Communication Protocols for Wireless Microsensor Networks,” Proc. Int’l Conf. System Sciences, Jan. 2000.
[8] S. Shenker, S. Ratnasamy, B. Karp, R. Govindan, and D. Estrin, “Data-Centric Storage in Sensornets,” SIGCOMM Computer Comm. Rev., vol. 33, no. 1, pp. 137-142, 2003.
[9] S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, R. Govindan, L. Yin, and F. Yu, “Data-Centric Storage in Sensornets with GHT, a Geographic Hash Table,” Mobile Networks and Applications, vol. 8, no. 4, pp. 427-442, 2003.
[10] J. Newsome and D. Song, “GEM: Graph Embedding for Routing and Data-Centric Storage in Sensor Networks without Geographic Information,” Proc. First Int’l Conf. Embedded Networked Sensor Systems, pp. 76-88, 2003.
[11] Y.-J. Kim, R. Govindan, B. Karp, and S. Shenker, “Geographic Routing Made Practical,” Proc. Second USENIX Symp. Networked Systems Design and Implementation (NSDI ’05), May 2005.
[12] C.T. Ee, S. Ratnasamy, and S. Shenker, “Practical Data-Centric Storage,” Proc. Third USENIX Symp. Networked Systems Design and Implementation (NSDI ’06), May 2006.
[13] J. Ahn and B. Krishnamachari, “Fundamental Scaling Laws for Energy-Efficient Storage and Querying in Wireless Sensor Networks,” Proc. Seventh ACM Int’l Symp. Mobile Ad Hoc Networking and Computing, pp. 334-343, 2006.
[14] D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann, and R. Govindan, “Multiresolution Storage and Search in Sensor Networks,” ACM Trans. Storage, vol. 1, no. 3, pp. 277-315, 2005.
[15] M. Li, D. Ganesan, and P. Shenoy, “PRESTO: Feedback-Driven Data Management in Sensor Networks,” Proc. Third USENIX Symp. Networked Systems Design and Implementation (NSDI ’06), May 2006.
[16] A. Demers, J. Gehrke, R. Rajaraman, N. Trigoni, and Y. Yao, The Cougar Project: A Work-in-Progress Report,” SIGMOD Record, vol. 32, no. 4, pp. 53-59, 2003.
[17] J. Gehrke and S. Madden, “Query Processing in Sensor Networks,” IEEE Pervasive Computing, vol. 3, no. 1, pp. 46-55, Jan. 2004.
[18] G. Mathur, P. Desnoyers, D. Ganesan, and P. Shenoy, “Ultra-Low Power Data Storage for Sensor Networks,” Proc. Fifth Int’l Conf. Information Processing in Sensor Networks, pp. 374-381, 2006.
[19] D. Zeinalipour-Yazti, S. Lin, V. Kalogeraki, D. Gunopulos, and W.A. Najjar, “MicroHash: An Efficient Index Structure for Flash- Based Sensor Devices,” Proc. Fourth USENIX Conf. File and Storage Technologies, 2005.
[20] G. Mathur, P. Desnoyers, D. Ganesan, and P. Shenoy, “CAPSULE: An Energy-Optimized Object Storage System for Memory-Constrained Sensor Devices,” Proc. Fourth Int’l Conf. Embedded Networked Sensor Systems, 2006.