Volume 2, No. 7, July 2011
Journal of Global Research in Computer Science
RESEARCH PAPER
Available Online at www.jgrcs.info
ISSN-2229-371X

DATA DIFFUSION AND GATHERING IN WIRELESS SENSOR NETWORKS

P. Balamurugan[*1] and Dr. K. Duraiswamy[2]
[1]Assistant Professor/CSE, K.S.R.College of Engineering Tiruchengode, Namakkal ,India
balamurugan_tlr@yahoo.co.in
[2]Dean, K.S.Rangasamy College of Technology, Tiruchengode, Namakkal, India

Abstract- A sensor network is a set of small autonomous systems, called sensor nodes which cooperate to solve at least one common problem. Their tasks include the perception of physical parameters. One of the most important applications for wireless sensor networks (WSNs) is Data Collection, where sensing data are collected at sensor nodes and forwarded to a central base station for further processing. Since using battery powers and wireless communications, sensor nodes can be very small and easily attached at specified locations without disturbing surrounding environments. In this paper, we review recent advances in this research area. We first highlight the special features of data collection in WSNs. we then discuss issues and prior solutions on the data gathering protocol design and the data dissemination protocol design. Our discussion also covers different protocol for data gathering, which is a critical component for energy efficient data gathering and greatly affects the overall performance of a data collection WSN system.

Keywords: Wireless Sensor Networks, LEACH, PEGASIS, TREEPSI, Data Collection, Data Dissemination, and Data Gathering

INTRODUCTION

Wireless sensor networks have been applied to many applications. Among them, one of the most important applications is data collection, where sensing data are continuously collected at each sensor node and forwarded through wireless communication to a central base station for further processing. In a WSN, each sensor node is powered by a battery and uses wireless communications. This results in the small size of a sensor node and makes it easy to be attached at any location with little disturbances to the surrounding environment. Such flexibility greatly eases the costs and efforts for deployment and maintenance and makes wireless sensor network a promising approach for data collection comparing with its wired counterpart.

The unique features of WSNs, however, also bring many new challenges. For instance, the lifetime of a sensor node is constrained by the battery attached on it, and the network lifetime in turn depends on the lifetime of sensor nodes, thus, to further reduce the costs of maintenance and redeployment, the consideration of energy efficiency is often preferred in a WSN design. Moreover, these challenges are complicated by the wireless losses and collisions when sensor nodes communicate with each other [1-4].

In addition, the requirements specified by data collection applications also raise issues that need to be considered in the network design. First of all, to accurately acquire different types of data (such as temperature, light, and vibration), different sensors with different sampling rates may be deployed at different locations. Also as being relayed toward the base station, more and more sensing data will be accumulated along the delivery path. These issues may cause unbalanced energy consumptions over a WSN and significantly shorten the network lifetime if not handling carefully.

In this paper, we first highlight the special features of data collection in WSNs. With these features in mind, we then discuss issues and previous works on the data gathering protocol design. In addition, we discuss different protocols for data gathering such as Direct Transmission, Binary Scheme, LEACH, PEGASIS and TREEPSI, which acts as an indispensable protocol for energy efficient data gathering and greatly affects the overall performance of a data collection WSN system.

DATA COLLECTION

In sensor network each sensor node has capability of sensing capacity. The sensor nodes are randomly deployed for data collection at specified location. First of all, a sensor node after deployed is expected to work for days, weeks, or even years without further interventions. Since it is powered by the attached battery, high efficient energy utilization is necessary, which is different from the Internet as well as wireless mesh and mobile ad hoc network, where either constant power sources are available or the expected lifetime is several order of magnitude lower than it is for WSNs.

The collected data are then forwarded back to a central base station for further processing. Traditionally, these sensors are connected by wires which are used for data transmission and power supply. However, the wired approach is found to need great efforts for deployment and maintenance. To avoid disturbing the ambient environment, the deployment of the wires has to be carefully designed. And a breakdown in any wire may make the whole network out of service and enormous time and efforts may be taken to find out and replace the broken line. In addition, the sensing environment itself may make the wired deployment and its maintenance very difficult, if not impossible. For example, the environments near a volcano or a wildfire scene, where the hot gases and steams can damage a wire easily. Indeed, even in a less harsh environment like wild habitat or a building, the threats from rodents are still critical and make the protection of wires much more difficult than that of sensors. All these issues make wireless sensor network a pleasant choice as it emerges with technology advances [5, 6].

Data collection requires all sensing data are correctly and accurately collected and forwarded to the base station, since the processing of these data needs global knowledge and is much more complex than that in other applications like target tracking. Thus, the major traffic in data collection is the reported data from each sensor to the base station. Such "many-to-one" traffic pattern, if not carefully handled, will cause high unbalanced and inefficient energy consumption in the whole network.

In practice, after a data collection WSN is deployed, network setup/management and/or collection command messages are disseminated from the base station to all sensor nodes by the message dissemination protocol. Then based on the information indicated by the disseminated messages, sensing data are gathered from different sensors and delivered to the base station through the data gathering protocol. It is worth noting that in a data collection system, the above process may work repeatedly, so that after one round of data collection, new setting/command messages are disseminated, thus, starting a new round of collection.

DATA DISSEMINATION

Data dissemination is a process by which data and queries for data are routed in the sensor network. In a scope of data dissemination, a *source* is the node that generates the data and an *event* is the information to be reported. A node that is interested in data is called *sink* and the *interest* is a descriptor for some event that node is interested in. Thus, after source receives an interest from the sink, the event is transferred from the source to the sink. As a result, data dissemination is a two-step process. First, the node that is interested in some events, like temperature or air humidity, broadcasts its interests to its neighbors periodically. Interests are then propagated through the whole sensor network. In the second step, nodes that have requested data send back data after receiving the request. Intermediate nodes in the sensor network also keep a cache of received interests and data [1].There exist several different data dissemination methods. In this paper flooding, gossiping, and SPIN are covered in more detail.

Flooding

In flooding method each sensor node that receives a packet broadcasts it to its neighbors assuming that node itself is not the destination of the packet and the maximum hop count is not reached. This ensures that the data and queries for data are sent all over the network [4].Flooding is a very simple method, but is has several disadvantages. In flooding duplicate messages can be sent to the same node which is called implosion. This occurs when a node receives the same message for several neighbors. In addition, the same event may be sensed by several nodes, and thus when using flooding, neighbors will receive duplicate reports of the same event, this situation is called overlap. Finally, many redundant transmissions occur when using flooding and flooding does not take into account available energy at sensor nodes. This wastes a lot of network's resources and decreases the lifetime of the network significantly.

Gossiping

Gossiping method is based on flooding, but node that receives the packet forwards it only to a single randomly selected neighbor instead of sending it to all neighbors. The advantage of gossiping is that it avoids the problem of implosion and it does not waste as much network resources as flooding. The biggest disadvantage of gossiping is that since the neighbor is selected randomly, some nodes in the large network may not receive the message at all. Thus, gossiping is not a reliable method for data dissemination [4].

SPIN

Sensor Protocols for Information via Negotiation (SPIN) use negotiation and resource adaption to address the disadvantages of basic flooding. SPIN uses data-centric routing, nodes are advertising their data and they will send the data after receiving a reply from interested nodes.

SPIN uses three types of messages: ADV, REQ, and DATA. The sensor node that has collected some data sends an ADV message containing meta-data describing the actual data. If some of node's neighbors is interested in the data, the neighbor sends a REQ message back. After receiving the REQ message, the sensor node sends the actual DATA. The neighbor also sends ADV message forward to its neighbors, thus data is disseminated through the network. Figure1 below describes ADV-REQ-DATA exchange of SPIN.

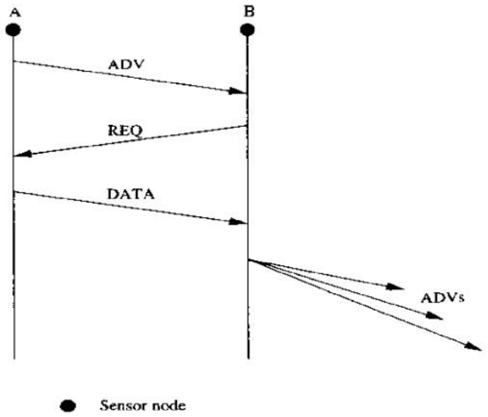

Figure 1: An example of SPIN

In the figure 1, node A advertises its data using an ADV message, its neighbor node B replies with a REQ message and thus node A sends actual data to the B. Node B also forwards ADV messages to its neighbors. Improved version of SPIN, SPIN-2 uses an energy or resource threshold to reduce participation of nodes. Thus, only those nodes that have sufficient amount of resources participate in ADV-REQ-DATA exchange [4][6].

SPIN is more efficient than flooding since the negotiation reduces the implosion and overlap. Resource adaptation in SPIN-2 prolongs the lifetime of the network: sensor nodes with low resources do not have to participate in ADV-REQ-DATA exchange and as a result they can collect data for a longer time.

DATA GATHERING

The aim of data gathering is to transmit data that has been collected by the sensor nodes to the base station. Data gathering algorithms aim to maximize the amount of rounds of communication between nodes and the base station, one round means that the base station has collected data from all sensor nodes. Thus, data gathering algorithms try to minimize power consumption and delay of the gathering process. Data gathering may seem similar to data dissemination, but there are some differences. In data dissemination, also other nodes beside the base station can request the data while in data gathering all data is transmitted to the base station. In addition, in data gathering data can be transmitted periodically, while in data dissemination data is always transmitted on demand[1][7,8].Various data gathering approaches like direct transmission, PEGASIS, and binary scheme [5] will be covered here in more detail.

Direct Transmission

In direct transmission method all sensor nodes send their data directly to the base station. While direct transmission is a simple method, it is also very ineffective. Some sensor nodes may be very far away from the base station, thus amount of energy consumed can be extremely high. In addition, sensor nodes must take turns when transmitting data to the base station to avoid collision. Thus, the delay is also very high. Overall, direct transmission method performs very poorly since the aim of data gathering approaches is to minimize both the energy consumption and the delay [2].

Pegasis

Power-Efficient Gathering for Sensor Information Systems (PEGASIS) is a data gathering protocol that assumes that all sensor nodes know the topology of the whole network. PEGASIS aims to minimize the transmission distances over the whole sensor network, minimize the broadcast overhead, minimize the amount of messages that are sent to the base station, and to distribute the energy consumption equally between all nodes.

In PEGASIS a chain of sensor nodes is constructed using a greedy algorithm starting from the node farthest from the base station. This chain is constructed before the data transmission begins and is reconstructed if nodes die out. During the data transmission, nodes aggregate the data and only one message is forwarded to the next node. The node that is selected as a leader then transmits all the data to the base station in a single message. The delay of messages reaching the base station is *O (N)* where *N* is the amount of sensor nodes in the network. An example of PEGASIS is shown in Figure 2. Data is transmitted from both ends of the chain to the leader, which sends all data to the base station [7].

PEGASIS achieves its goals: Transmission distances over the whole network are short, overhead is relatively small, only one message is sent to the base station and energy is distributed quite equally between all nodes, since almost all nodes will send and receive exactly one message. Disadvantages of PEGASIS include high delay, in large sensor networks the chain becomes very long and a high amount of hops is required to forward data from the end points of the chain to the base station. In addition, PEGASIS assumes that every node has topology information about the network and this assumption is not always valid in sensor networks [8-10].

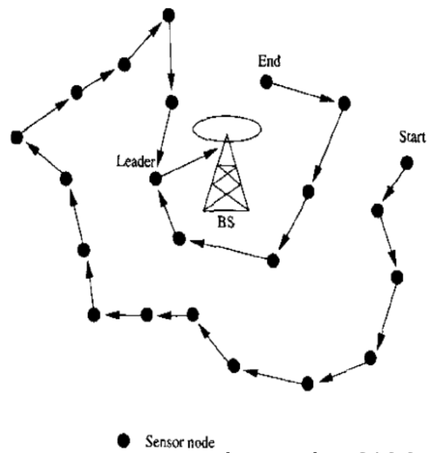

Figure 2. Data gathering with PEGASIS

Binary Scheme

Binary scheme is also a chain-based scheme like PEGASIS. It classifies nodes into different levels. All nodes that receive message at one level rise to the next level where the amount of nodes is halved. Transmission on a one level occurs simultaneously to reduce delay. An example of the binary scheme is shown in Figure 3. Nodes s1, s3, s5 and s7 receive messages on the first level and thus they rise to the next level. On the second level nodes s3 and s7 receive messages and finally node s7 forwards all data to the base station.

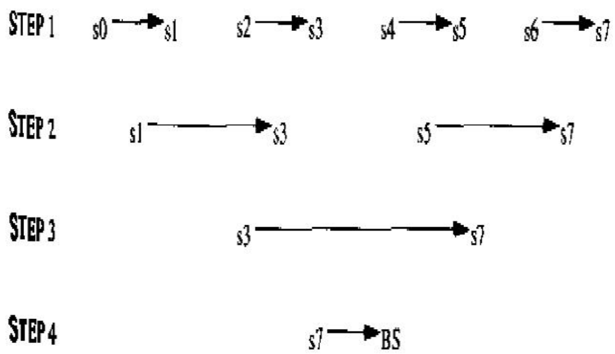

Figure 3. Binary scheme

Biggest advantage of binary scheme is a very low delay of only *O(log2N)*, where the *N* is the amount of nodes. Thus, binary scheme has significantly lower delay than PEGASIS in large sensor networks. However, binary scheme relies on simultaneous transmission which is possible if the nodes communicate using CDMA, but the scheme does not work with all networks. Other disadvantages include non equal distribution of energy consumption; nodes that are active on several levels consume more energy than nodes that are only active at the first level. This might lead to the situation where some of sensor nodes die earlier than others. In addition, transmission distances may become long in high levels, which lead to a high power consumption.

For other networks, similar chain-based three-level scheme has been developed. It divides the chain in groups and within a group only one node is transmits at once. Then the leader of the group rises to the next level where the first level leaders transmit data to a new leader, like in the binary scheme. At the third level, all data is transmitted to a single node that passes it to the base station. Three levels have been found to optimize delay and power consumption.

Leach

LEACH is a self-organizing, adaptive clustering protocol that uses randomization to distribute the energy load evenly among the sensors in the network [2]. Looking back at the old algorithms, one could see how picking a random sensor and having it fixed to be the CH through the system lifetime that it would die very quickly cutting short the lifetime of all other nodes belonging to the cluster. LEACH changes this by randomly rotating among the various sensors in order to not drain the battery of a single sensor. Also, it reduces more energy dissipation and enhancing system lifetime by performing local data fusion to compress the amount of data being sent from the clusters to the base station.

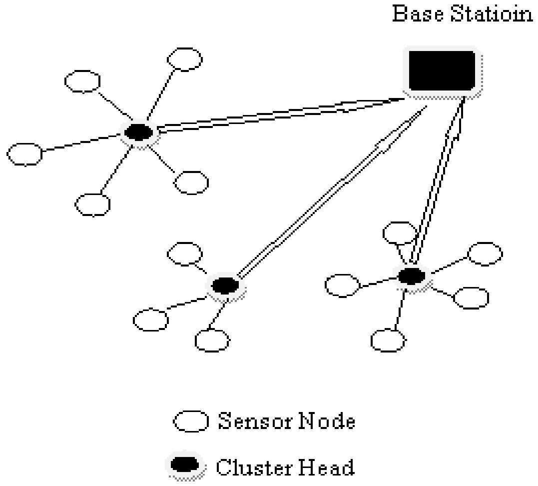

Figure 4. Graph showing construction of cluster in LEACH

Sensors elect themselves to be local CHs at any given time with certain probability and these CH nodes broadcast their status to the other sensors in the network [2][11]. The sensor nodes then chooses a cluster to be a part of by which CH requires the minimum communication energy. Although most of the time a sensor would choose the closest CH that connection could have a barrier interrupting the communication, so joining a cluster where the CH is further off would be much easier. When all of the sensors have been structured inside of each cluster, the CH creates a schedule for them in the cluster. This helps minimize the energy dissipated in the individual sensors, because it enables all non-CHs to shut off their radio components until their transmit time. Each sensor transmits its data to the CH and once the CH collects all of the data it aggregates it and transmits it to the BS. Normally the BS is a great distance away, so it will be high energy transmission. This method is described in figure4.

Treepsi

Tree-based Efficient Protocol for Sensor Information (TREEPSI) is a tree-based protocol that is different from the above-mentioned protocols. Before the data transmission phase, the WSNs select a root node among all of the sensor nodes [3]. The root is identified by id = *j*. There are two ways to build the tree path: the first is computing the path centrally using the sink and broadcasting the path information to the network. The second can be the same tree structure using a common algorithm in each node. At the initial phase, the root will create a data gathering process for the child nodes using any standard tree traversal algorithm [9][12]. The data transmission phase begins after the tree is built. All of the leaf nodes will start sending sensed data towards their parent nodes. The parent nodes will collect the received data together with their own data that is then sent to their parents. The transmission process will be repeated until all of the data received by the root node is sent. After the root node has aggregated the data, it sends the collected data directly to the sink. The process will repeat until the root node has no more data to send. The WSN will then reselect a new root node. The new root identification number would be $j + 1$. The initial phase is then repeated and the tree path will not change until the root node is dead. TREEPSI and PEGASIS use the same method to transmit data from the leaf node to the chain/root head. The path length from the end leaf node to the root/chain node in TREEPSI is shorter than that in PEGASIS. For this reason, TREEPSI can use less data transmission power less than PEGASIS. TREEPSI has better performance by about 30% than PEGASIS, but it still has a problem involving binary tree algorithm restriction in which the path must make a detour in the topology.

CONCLUSION

Wireless sensor networks have been applied to many applications since emerging. In this paper, we presented an in-depth survey on recent advances in the design issues and solutions for data collection systems using WSNs. Specifically, we first highlighted the special features of data collection in WSNs, by wireless data collection network and other applications using WSNs. Bearing these features in mind, we discussed issues and solutions on the design of data gathering protocols and data dissemination.

Simple approach to collect data from sensor nodes is direct approach where each sensor nodes transmit the data directly to the base station (BS) which is located far away. Cost to transmit data from each sensor node to BS is very high, thus nodes die quickly and hence reducing the lifetime of the network. Therefore to utilize energy efficiently goal is to use as few transmissions as possible. LEACH Protocol is designed where sensor nodes are organized to form local cluster with one node in cluster selected as cluster head. Sensor nodes from one cluster send data to its cluster head where data is aggregated and fused data is transmitted to BS. Cluster heads are chosen randomly and achieve a factor of 8 improvements compared to direct approach. Although LEACH protocol reduces energy consumption by factor 8, energy is consumed is forming cluster. In LEACH 5% of the nodes are the head nodes at the same time that also amounts to energy consumption. PEGASIS is the improved protocol where only one node is chosen a head node which sends the fused data to the BS per round. This achieves factor of 2 improvement compared to LEACH protocol. PEGASIS protocol requires formation of chain which is achieved in two steps: Chain construction and data gathering. In chain construction construct the chain we start from the furthest node from the BS and then greedy approach is used to construct the chain. In Gathering data Leader of each round is selected randomly. Randomly selecting head node also provides benefit as it is more likely for nodes to die at random locations thus providing robust network. When a node dies chain is reconstructed to bypass the dead node.

Many issues still need to be further explored and possibly considered jointly so as to lead to a more efficient and long-lifetime data collection system. In the future it will focus on PEGASIS because it is better than LEACH.

REFERENCES

[1] Akkaya, K., and Younis, M., 2005,"A Survey on Routing Protocols for Wireless Sensor Networks," Elsevier Ad Hoc Networks Journal, Vol.3, No.3, pp.325-349.

[2] W.Heinzelman, A, Chandrakasan, and H, Balakrishnan. "Energy Efficient Communication Protocols for Wireless Micro Sensor Networks". In proceedings of the Hawaii conference on system science, Jan.2000, Vol2.pp.1-10.

[3] Yun-Sheng Yen, Kai-Chun Huang, Han-Chieh Chao, and Jong Hyuk Park," Tree-Clustered Data Gathering Protocol for Wireless Sensor Networks, " JOURNAL of the Chinese Institute of Engineers,vol.32.No.7,pp.1025-1036(2009).

[4] Siva Ram Murthy,C., and Manoj,B.S.,"Ad Hoc Wireless Networks Architectures and Protocols,"Pearson Education,Inc, 2004.

[5] Stephanie Lindsey, Cauligi S. Raghavendra, and Sivalingam,K.M., "Data Gathering Algorithms in Sensor Networks Using Energy Metrics.," IEEE Transactions on Parallel and Distributed Systems,Vol.13,No.9,pp.924-935,Semtember 2002.

[6] Akyildiz,I.F,M,Su,M,Sangarasubramanian,Y.,and Cayirci,E.,"IEEE Communications Magazine,Vol.40, No.8,pp.102-105,2002"

[7] Stephanie Lindsey, Cauligi S. Raghavendra,"PEGASIS: Power Efficient Gathering in Sensor Information System," In Proceedings of IEEE ICC 2001, Vol.3, pp.1125-1130, June 2001.

[8] Ye,F..Chen,A., and Zhang,L.,"A Scalable Solution to Minimum Cost Forwarding in Large Sensor Networks ,"In Proceedings of IEEE ICCCN 2001,pp.304-309,October 2001.

[9] A. Manjeshwar, and D. P. Agrawal. "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks," in Proceedings of IEEE 2001, 2001.

[10] Stephane Lindsey, Cauligi S. Ragavendra , and Krishna Sivalingam," Data Gathering in Sensor Networks using Energy Delay Metrics", Proceedings of the International Parallel & Distributed Processing Symposium, pp.188,2001.

[11] Rahul C. Shah, Jan M. Rabaey, "Energy Aware Routing For Low Energy Ad Hoc Sensor Networks", Proc of IEEE WCNC, March 2002.

[12] W. Ye, J. Heidemann, and D. Estrin, "An Energy-Efficient MAC Protocol for Wireless Sense Networks," in Proceedings of IEEE INFOCOM 2002, New York, June 2002.

[13] G. Huang, X. Li, and J. He, "Dynamical Minimal Spanning Tree Routing Protocol for Large Wireless Sensor Networks," in Proceedings of IEEE ICIEA 2006, May 2006.

[14] L.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless Sensor Networks : A Survey", Computer Networks, vol.38, pp.393-422, 2002.

[15] J.A. Stankovic, F. Abdelzaher, C.Y. Lu, L. Sha and J.C. Hou, "Real-Time Communication and Coordination in Embedded Sensor Networks", Proceedings of the IEEE, Vol. 91, No. 7, pp. 1002- 1022, July 2003.

[16] Sung-Min Jung, Young-Ju Han, Tai-Myoung Chung, "The Concentric Clustering Scheme for Efficient Energy Consumption in the Pegasis", the 9th International Conference on Advanced Communication Technology, Vol. 1, pp. 260-265, Feb. 2007.

[17] Yan Yu, Ramesh Govindan, Deborah Estrin, "Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks", UCLA Computer Science Department Technical report UCLA/CSD-TR-01-0023, May2001.

[18] C. Intanagonwiwat, R. Govindan, and D. Estrin. "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks". In Proceedings of the Sixth Annual International Conference on Mobile Computing and Networks (MobiCOM 2000), Boston, Massachusetts, August 2000.

[19] Schwiebert L, Gupta S K S, Weinmann J, etal, "Research Challenges in Wireless Networks of Biomedical Sensors". Mobi COM'01, 2001, pp.151-165.

[20] Wendi Rabiner Heinzelman, Joanna Kulik, and Hari Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks", Proceedings of the 5^{th} annual ACM/IEEE international conference on Mobile computing and networking, pp. 174 – 185, 1999.

[21] Jung-Eun Lee, Keecheon Kim, "Diamond-Shaped Routing Method for Reliable Data Transmission in Wireless Sensor Networks", IEEE International Symposium on Parallel and Distributed Processing with Applications, pp.799-801, 2008.

[22] Nahdia Tabassum, Quazi Ehsanul Kabir Mamun and Yoshiyori Urano, "COSEN: A Chain Oriented Sensor Network for Efficient Data Collection", Third International Conference on Information Technology: New Generations (ITNG'06), pp. 262-267, 2006.

[23] Jie Zheng, Shujie Guo, Yugui Qu, Baohua Zhao,"A low Delay Energy Equalizing Routing Scheme in Wireless Sensor Network ", IEEE 2007, pp.2667-2670.

[24] Mulik, W. B. Heinzelman, and H. Balakrishnan. "Negotiation-based protocols for disseminating information in wireless sensor networks",Wireless Networks, 2002(8):109-185.

