Tree-based Data Aggregation Algorithms in Wireless Sensor Networks: A Survey

Mohammad Abdus Salam
Department of Computer Science
Southern University and A&M College
Baton Rouge, Louisiana 70813, USA

Tanjima Ferdous
Department of Computer Science
Southern University and A&M College
Baton Rouge, Louisiana 70813, USA

Abstract

Wireless sensor networks (WSN) are resource-constrained self-organizing networks that are often deployed in hostile and inaccessible environments in order to collect data. The reliability of WSN is affected by faults that may occur due to various reasons such as malfunctioning hardware, software glitches, dislocation, or environmental hazards. Appropriate fault tolerance mechanism will mitigate network failure and increase the entire network reliability. Moreover, WSN suffers from many constraints, including low computation capability, small memory, limited energy resources, and the use of insecure wireless communication channels. To overcome these limitations and achieve reliable data, many data aggregation algorithms have been proposed. The main goal of these data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that the network lifetime can be increased. In this paper we present an elaborate survey on different data aggregation algorithms based on tree architecture and compare and contrast different algorithms on the basis of performance measures such as lifetime, energy consumption, synchrony, and robustness. We conclude with possible future research directions.

Keywords
Wireless sensor networks, security, tree-based algorithm, cryptography, synchrony

1. Introduction
Wireless sensor networks consist of low-cost, low-power, multifunctional tiny sensor nodes. These sensor nodes have sensing, data processing, and communicating capabilities. WSN promises researchers a powerful instrument for observing various phenomena and it has been deployed in many application domains such as habitat monitoring (Salam et al 2010) and scientific exploration (Tolle et al 2005). WSN nodes are prone to be failure due to energy depletion, hardware failure, communication link errors, malicious attack, and so on. Therefore, the reliability and accuracy of individual sensor node's reading is very crucial. Many protocols have been developed to ensure secure data routing and aggregation for WSN throughout the past couple of years (Akkaya and Younis 2005), (Tilak et al 2002). Many researchers have also addressed the potential of collaboration among sensor nodes in data gathering and aggregating. However, sensor nodes are constrained in energy supply, computing capability, and bandwidth. Thus, innovative routing techniques have an important consideration for the reduction of these constraints. Routing in WSN is very challenging and has distinguishing characteristics compared to other wireless networks.

Figure 1 shows a sensor field and the components of a sensor node. A sensor node is composed of four basic components: sensing unit, processing unit, transceiver unit, and power unit. Each sensor node has the capability to collect and route data back to the base station (BS). Data can be routed back to the BS by using various routing protocols. The BS may communicate with the user using available Internet connection.

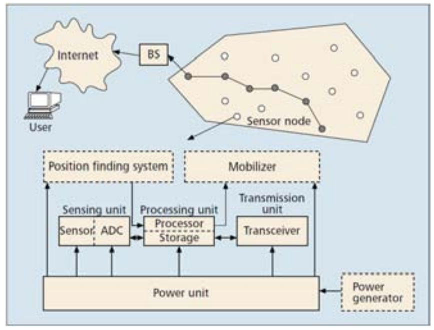

Figure 1: The components of a sensor node (Source: Akyildiz et al 2002)

In our paper, we mainly focus on the tree-based secure data routing techniques. The data gathering of these protocols can be characterized as i) systematic collection and transmission of sensed data from multiple sensor nodes and ii) processing them at a remote location for final decision making. On the other hand, data aggregation is a process of aggregating the data from multiple sensor nodes to eliminate redundant transmission and provide fused information to the base station. Security in data gathering and aggregation is an important design issue for wireless sensor networks. Secure data aggregation of information is a challenging task when sensors are deployed in hostile environments and are susceptible to physical attacks. In this paper, we describe some recent tree-based routing protocols which discuss the secure data aggregation problem and also explain some of the main issues involved in implementing security in sensor networks.

2. Tree-based Routing Protocols in WSN

In this section, we survey tree-based routing protocols for WSN. Although there are some previous works for surveying the characteristics, applications, and communication protocols in WSN (Akkaya and Younis 2005), (Tilak et al 2002), (Nakamura et al 2007), the scope of the survey presented in this paper is restricted only for the following tree-based routing protocols.

2.1 Secure Aggregation Tree (SAT) (Dreef et al 2006)**:** The authors proposed a secure data aggregation algorithm. It mostly depends on its topological constraints such as a node may hear messages being received by its father node and grandfather node. Every node is considered to be behaving honestly unless any cheating behavior is detected. In this algorithm, first, a secured aggregation tree is built considering those constraints. Second, when there is a doubt that an aggregation node is cheating, it is confirmed by using a weighted voting scheme. Third, once the misbehaving node is detected, a local recovery scheme is presented to re-build SAT so that the misbehaving node is excluded from the aggregation tree. The most important feature is no dependency on cryptography but still providing secured communication. This facilitates less calculation thus less energy consumption in sensor nodes. But, on the other hand, depending mostly on the topological constraints is a limitation. Same time, every node can hear each other which could cause energy consumption as well.

2.2 Power –Efficient Data Gathering and Aggregation Protocol (PEDAP) (Korpeoglu and Tan 2003): In this paper two algorithms are proposed under the name of Power –Efficient Data Gathering and Aggregation Protocol (PEDAP). These are routing schemes based on near optimal minimum spanning tree. These two algorithms are same but one is the power-aware version of the other. First, while the tree starts to build, a node is selected as base station. Then, in each iteration the minimum weighted edge from a vertex in the tree to a vertex not in the tree is selected and added to the tree. Here, the newly added node will send its data through the indicated edge. This procedure is repeated until all nodes are added to the tree. In this work, the sensors are in fixed location. These are in direct communication range of each other and can transmit to and receive from the base station. The sensors periodically sense data from environment and send to base station in round basis. Before sending to base station, the nodes aggregates their data with other data received from other nodes. In this work, the main consideration is wireless sensor produce a single packet to transmit to the base station. This proposal shows that it can save much energy and much improved than LEACH (Heinzelman et al 2000) and PEGASIS (Lindsey and Raghavendra 2002) protocols.

2.3 Dynamic Convoy Tree-based Collaboration (DCTC) (Cao and Zhang 2004)**:** This paper is focusing on detecting and tracking a mobile target. For this purpose they have created a framework naming dynamic convoy tree-based collaboration. The tree structure proposed in this algorithm is called convoy tree. This structure provides sensor nodes around the moving target (Figure 2). When the target moves, the tree is dynamically reconfigured to add and prune nodes. As the target moves, nodes around it awake and establish a tree. One node in this tree becomes the root node by the collaboration of the whole convoy tree. This node collects data from all the other nodes and refines that data for more accurate information. As the target moves, the tree is reconfigured with some new nodes. The root may also be changed for optimized overhead communication. The root is also responsible to send the refined information to the base station and awaking nodes that becomes nearer to the target as the target approaches, since most of them are asleep. Extensive experiments have shown that this local reconfiguring scheme provides balanced energy conservation too.

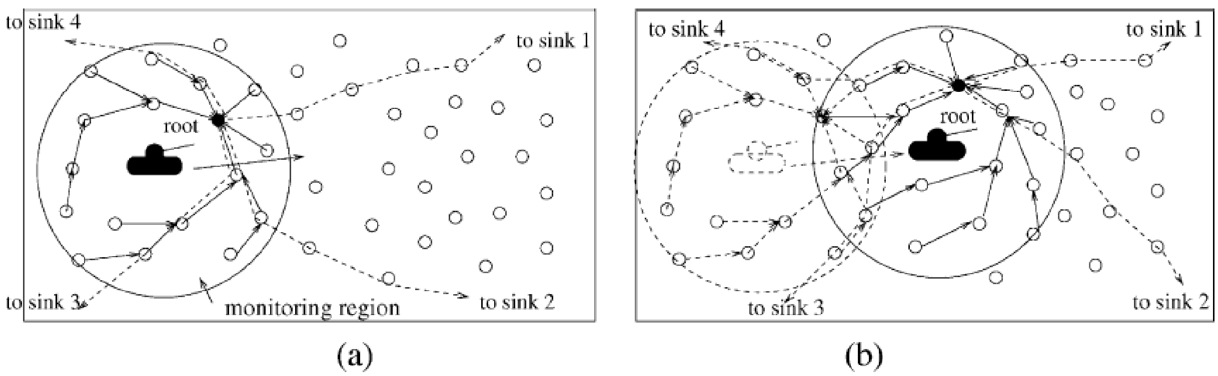

Figure 2: Using convoy tree to track the target. (a) Data collection (b) Tree reconfiguration (Source: Cao and Zhang 2004, Akyildiz et al 2002)

2.4 Weighted Spanning Tree on the basis of LEACH (WST-LEACH) (Chen et al 2010)**:** This paper proposed WST-LEACH algorithm based on the well-known LEACH protocol (Heinzelman et al 2000). This elaborates the benefits provided in LEACH protocol as well as spanning it. It has concerns for energy savings and discovering a way of efficient data aggregation. It selects cluster heads in its network to build a data sinking path. It is done not only in a random manner but also considering the remaining energy, distribution density of the nodes and the distance from cluster heads to the base station. Then it establishes a weighted spanning tree using the cluster heads which does the weighted calculation of the weighted value. This calculation contains factors such as remaining energy of the cluster heads, distribution of surrounding nodes and the distance to the other cluster heads. Then, after aggregating the data, it is sent to the base station through the tree.

2.5 Energy-Aware Distributed heuristic to generate the Aggregation Tree (EADAT) (Cheng et al 2003)**:** In this paper, the authors have proposed an algorithm called EADAT. Here, no assumption is made on local network topology. It is mostly depending on the residual power. In this structure, it is considered that sensors are distributed randomly in a target area. There is one gateway that connects this micro-sensor network to the outside system. This gateway would be there at the boundary of the monitoring area and at least some sensors would be able to reach this boundary to communicate with the gateway. Here, micro-sensors are called data source or event source and the gateway is called data sink or event sink. Then, while transmitting the data, it is pre-processed. In this protocol, every node does data aggregation by using aggregation function like SUM, AVG, MEAN, MAX etc. Then they transmit it to their parents. Thus, they are doing data-centric routing. In this protocol, energy saving concern is great. Here network life time increment problem is solved by energy-aware data centric routing.

2.6 Computing aggregates for monitoring wireless sensor networks (Govindan et al 2003)**:** This paper describes architecture for monitoring the sensor networks. This architecture also focuses on doing continuous computation of aggregates (sum, average, count) of network properties (loss rates, energy-levels etc., packet counts). First, a novel tree construction algorithm is proposed that enables energy-efficient computation of some classes of aggregates. Then through actual implementation and experiments it is shown that even relatively benign environments can significantly impact the computation of these aggregate properties. The authors are claimed to be the first to articulate architecture for sensor network monitoring.

2.7 Distributed version of Kruskal's Minimum Spanning Tree (MST) (Hussain et al 2009): This paper is an approach to find an efficient routing of sensor network tree which would be able to save energy efficiently as well as to provide good structure of communication for data aggregation. This paper is a modification of a previously established algorithm called Minimum Spanning Tree (MST) search algorithm by Kruskal (Kruskal 1956). It is based on distributed search by hierarchical clusters. Here the spanning tree is constructed in an undirected weighted simple graph, which is represented by a weighted adjacency matrix. To make efficient search of the distributed spanning tree, the input graph is preliminary clustered by recursive hierarchical clustering in two parts. These two equal parts intends to minimize the intra-cluster distances and to maximize the inter-cluster distance. Simulation results show that this approach is able to extend the functional lifetime of a wireless sensor network 3-4 times in terms of sensor transmission energy.

2.8 Converge-casting Tree Construction and Channel Allocation Algorithm (CTCCAA) (Gupta et al 2003)**:** This paper proposed a converge-casting algorithm called CTCCAA. The same tree is also used efficiently for broadcasting. This algorithm ensures collision free converge-casting which provides schedule for nodes. In converge-casting, a tree is constructed and a schedule is generated for converge-casting. Time slot is compared by calculating total duration of aggregation by doing slot allocation. It is proved here that slot duration is high if the broadcast tree is used instead of constructing a new converge-cast tree. This tree-based algorithm is efficient for reducing energy consumption.

2.9 Directed Diffusion (Govindan et al 2003)**:** This is a well known distributive algorithm called Directed Diffusion for data propagation and aggregation. This algorithm does not depend on any topological constraint. This algorithm is a data centric algorithm. It works mostly with how query or interest is going to be diffused throughout the network and how is it going to draw back the required data. The whole process is divided into four basic steps. First step is *naming*, In this step what queries are needed are illustrated by some attribute-value pairs. A sensing task is disseminated throughout the sensor networks as an interest for these named data. Second step is the i*nterests & gradients*. After the interest is received and data are sensed according to the interest, the sensing nodes create a path with a direction as well as a data sending rate to send the data to the sink, which did the query for the data. In fact this step gives primary information, called exploratory data, that whether the data asked for does really exist or not. Then the real data, if available, is gathered in the next step. Third step is *data propagation* where the exploratory data are sent through the established gradient. Last step is *reinforcement* for path establishment and truncation where the actual data is received through a more sophisticated, reliable, short path with high data rate. Figure 3 illustrates these steps. This tends to be a better design as it provides communication between neighbor to neighbor, not end to end. It is different from many conventional algorithms and solves many problems associated with them.

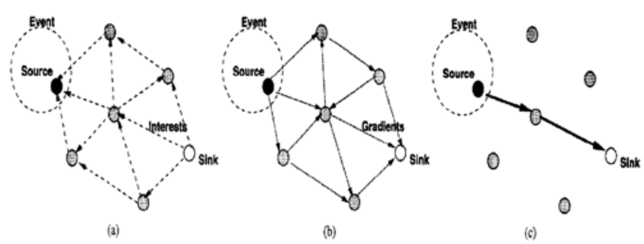

Figure 3: Simplified schematic for directed diffusion. (a) Interest propagation. (b) Initial gradients setup. (c) Data delivery along reinforced path. (Source: Govindan et al. 2003)

2.10 Hierarchical Scheme for Data Aggregation in Sensor Network (Ngoh 2004)**:** This paper is an enhanced version of directed diffusion (Govindan et al 2003). It claims an effective enhancement of directed diffusion which will provide much betterment. The scheme consists of four stages. First stage is the hierarchy construction of the nodes. In this stage, a node sends an interest message to all its neighbors. Unlike the directed diffusion, interest message will have an extra attribute named 'level'. This field is going to sort out which node is in which level. A node receiving an interest with a level N, will denote itself as in N^{th} level and will record the sender as its parent. Same process will go till the level N+2. Second stage is attribute establishment. In hierarchical tree building, though there were several parents, a node will choose only one parent to transmit the real data. Unlike the directed diffusion, in this stage, this scheme is going to add an extra field in the exploratory data which will identify the data source. Third stage is the data aggregation. In this stage, the sink reinforces the interest to get the real data. The special procedure followed here is called per hop aggregation because data are opportunistically aggregated by each parent. Fourth stage is data distribution. This stage illustrates how data is drawn back in multiple sinks. With multiple sinks, each sink will have its own hierarchical tree with separate interest. This algorithm is a better implementation of directed diffusion by saving 50% of transmission energy and it provides greater reliability.

2.11 Cluster-based Minimal Spanning Tree with Degree-Constrained (CMST-DC) (Huang et al 2008): In this paper, authors have proposed the algorithm called Cluster-based Minimal Spanning Tree with Degree-Constrained (CMST-DC). It is a hierarchical two layered structure that includes both tree structure and cluster structure. It has the features of robustness, low network delay and low time taking to complete a round. In the algorithm, first, clusters are formed. Then, a cluster head is elected in each of the clusters in every round based on residual energy. The clusters are the first layer of the network. Then all the cluster heads build a routing tree together and build the second layer of the network. Finally, one of the cluster heads in that second layer (selected by some criteria) sends the final data to the base station. Lots of design issues are considered before creating the structure of the algorithm that makes it a better approach. This algorithm ensures maximum utilization of network energy which in turns gives longer lifetime. This proposed algorithm compared favorably with LEACH (Heinzelman et al 2000) and GSEN (Tabassum et al 2006) in terms of delay time and total transmission distance.

2.12 Two-party and Three-party Key Agreement Protocols based on Bilinear Pairing (Liehuang et al 2008)**:** This paper focuses on group key authentication for secure communication. The authors proposed two-party and three-party key agreement protocols. Then using those, a group key agreement protocol is proposed which is defined as novel tree-based authenticated group key agreement protocol. In two-party and three-party authenticated key-agreement protocol it is not needed that all users to be authenticated. It is denoted that AKP(l,n) as n-party authenticated key agreement protocol with l users to be authenticated. So, for AKP(2,1) protocol, it is meant that A and B (considered users) want to agree upon a common key and A wants to be authenticated by B. For AKP(2,2), A and B want to agree upon a common key and they want to be authenticated by each other. Similarly AKP(3,3) and AKP (3,2) are described. These protocols are performed by following some certain steps, based on bilinear pairing, for each of the type. After that the authenticated group key agreement protocol, which is denoted as AP3TGDH, is described. It has three steps. First step is to build the key tree and how all the nodes come into a final key authentication decision using private/public keys is described. In the second step, it is described that if any new node joins the tree then how it is going to contribute in the tree. In third step, it is described if any node is kicked out or leaves the tree, then the whole tree is going to be refreshed and the agreement protocol may carry on. This paper is based on many group key authentication papers.

3. Performance Metrics

We observed that there are some hybrid protocols that fit under more than one category. We summarize tree-based data routing techniques WSN in Table 1. The Table shows different routing protocols and also indicates what different papers have been discussed in their papers. It also compares routing techniques according to many metrics such as robustness, hierarchy, functional lifetime, self-configurable, data-centric, and location centric. In column 2 of Table 1 “use of other algorithms” indicates that the proposed algorithm is based on other existing algorithms.

Some of the algorithms have hierarchy based routing where as others have both hierarchy as well as tree-based routing algorithms. Cluster-based routing protocols group sensor nodes to efficiently relay the sensed data to the sink. Since sensor nodes are energy constrained device, almost all of the discussed algorithms focused on energy consumption and targeted to achieve balance energy consumption. Many authors have performed simulation to compare their proposed algorithms with the existing algorithms. For example, Cluster-based Minimal Spanning Tree with Degree-Constrained (CMST-DC) is compared with other well known algorithms such as LEACH and GSEN.

Data-centric protocols name the data and query the nodes based on some attributes of the data. Location-based protocols utilize the location information and topological deployment of sensor nodes. There are not many numbers of energy-aware location-based approaches in the literature.

Table 1: Comparison of tree-based routing protocols in sensor networks

Paper	Use of other algorithms	Hierarchy	Balanced energy consumption	Comparative study	Extended functional lifetime	recovery scheme/ reconfiguration	Synchrony	Self-Configurable	Robustness	Static/ Mobile	Communication layer	Data centric	Location based
[1]			√	BFT BFT-D		√		√					√
[2]	√		√	√	√					Static			√
[3]	√		√			√		√		Mobile-Target			
[4]	LEACH	√	√	√	√						√		
[5]	Directed Diffusion		√		√		√			Static-sensors		√	
[6]			√						√	Static			√
[7]	MST	√	√		√					Static-sensors			
[8]			√				√						√
[9]			√	√					√	mobile target		√	
[10]	Directed Diffusion	√	√	Directed Diffusion							√	√	
[11]	LEACH GREEDY SHORT MST	√	√	LEACH GSEN	√			√	√		√		√
[12]			√	TGDH									

Legends: [1] (Dreef et al 2006), [2] (Korpeoglu and Tan 2003), [3] (Cao and Zhang 2004), [4] (Chen et al 2010), [5] Cheng et al 2003, [6] (Govindan et al 2003), [7] (Hussain et al 2009), [8] (Gupta et al 2003), [9] (Govindan et al 2003), [10] (Ngoh 2004), [11] (Huang et al 2008), [12] (Liehuang et al 2008)

4. Conclusion

Routing techniques in sensor networks is a challenging area of research. In this paper, we presented a comprehensive survey of tree-based routing techniques in wireless sensor networks which have been published in various literatures. There most common trend among these protocols is the reduction of energy consumption and extend the lifetime of the sensor network.

Since most of the applications in security and environmental monitoring require the data collected from the sensor nodes to be transmitted to a server for further analysis, we focused on those algorithms that address the issue of secure data gathering and aggregation.

Future work includes the enhancement of trust and reliability of the collected data from the sensor nodes. Wireless sensor networks are prone to many security attacks which impede the deployment and data propagation of sensor nodes. Security threats against routing protocols, which are a basic networking service for any sensor network communication, are important area of future research.

Acknowledgement

The authors would like to thank anonymous reviewers for their comments and suggestions which helped to improve this paper. This research is funded by Thurgood Marshall College Fund under research grant of the Department of Energy.

References

Akkaya K., and Younis, M. "A Survey of routing protocols in wireless sensor networks." *The Elsevier Ad Hoc Network Journal* 3/3, pp. 325-349, 2005.

Akyildiz, I., Su, W., Sankarasubramaniam, Y., and Cayirci, E. "A survey on sensor networks." *IEEE Communications Magazine* 40.8, pp. 102-114, 2002.

Cao, G., and Zhang, W. "Weighted spanning tree clustering routing algorithm based on LEACH." *2nd International Conference on Future Computer and Communication (ICFCC), 2010*. IEEE, pp. V2-223 - V2-227, 2010.

Chen, P., Gong, S., Zhang, H. "Weighted spanning tree clustering routing algorithm based on LEACH." *2nd International Conference on Future Computer and Communication (ICFCC'10)*. IEEE, pp. V2-223 - V2-227, 2010.

Cheng, X., Xue, G., and Ding, M. "Aggregation tree construction in sensor networks." *IEEE 58th Vehicular Technology Conference*. IEEE, pp. 2168-2172, 2003.

Dreef, D., Sun, B., Xiao, Y., and Wu, K. "Secure data aggregation without persistent cryptographic operations in wireless wensor networks." *25th IEEE International Performance, Computing, and Communications Conference (IPCCC'06)*, IEEE,pp. 635-640, 2006.

Govindan, R., Estrin, D., and Zhao, J. "Computing aggregates for monitoring wireless sensor networks." *IEEE International Workshop on Sensor Network Protocols and Application*, IEEE, pp. 139-148, 2003.

Govindan, R., Estrin, D., Heidemann, J., Silva. F., and Intanagonwiwat, C. "Directed diffusion for wireless sensor networking." *IEEE/ACM Transactions on Networking* 11.1, pp. 2-16, 2003.

Gupta, S., Schwiebert, L., and Annamalai, V. "On tree-based convergecasting in wireless sensor networks." *IEEE Wireless Communications and Networking (WCNC)*, pp. 1942-1947, 2003.

Heinzelman, W., Chandrakasan, A., and Balakrishnan, H. "Energy efficient communication protocol for wireless microsensor networks." *33rd Annual Hawaii International Conference on System Sciences*, pp. 3005-3014, 2003.

Huang, Y., Lin, J., and Liang, C. "An energy efficient routing scheme in wireless sensor networks." *22nd International Conference on Advanced Information Networking and Applications - Workshops*, IEEE, pp. 916-921, 2008.

Hussain, S., Yang, L., and Gagarin, A., "Distributed search for balanced energy consumption spanning trees in wireless sensor networks." *International Conference on Advanced Information Networking and Applications Workshops*, IEEE, pp. 1037 - 1042, 2009.

Korpeoglu, I., and Tan, H., "Power efficient data gathering and aggregation in wireless sensor networks." *ACM SIGMOD Record* 32.4, pp. 66-71, 2003.

Kruskal, J., "On the shortest spanning sub-tree of a graph and the traveling salesman problem." *Proceedings of the American Mathematical Society*, JSTOR, 48-50, 1956.

Liehuang, Z., Yuanda, C., Dazhen, W., and Hong, T., "A novel tree-based authenticated dynamic group key agreement protocol for wireless sensor network." *International Symposium on Electronic Commerce and Security*, IEEE, pp. 540-544, 2008.

Lindsey, S., and Raghavendra, C., "Pegasis: Power-efficient gathering in sensor information systems." *IEEE Aerospace Conference*, IEEE, 2002.

Nakamura, E., Loureiro, A., and Frery, A., "Information fusion for wireless sensor networks: methods, models, and classifications." *ACM computing surveys* 39.3, 2007.

Ngoh, L., Lee, B., Fu, C., and Zhou, B., "A hierarchical scheme for data aggregation in sensor network." *12th IEEE International Conference on Networks (ICON'04)*, pp. 525-529, 2004.

Salam, M., Soysal, O., and Schneider, H., "Integration of wireless sensor networks in geographical information systems: a survey." *The International Conference on Modeling, Simulations & Visualization Methods*. Las Vegas, Nevada, USA, July 12-15, 2010.

Tabassum, N., Haque, A., and Urano, Y. "GSEN: An efficient energy consumption routing scheme for wireless sensor network." *International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL'06),* IEEE, pp. 117-117, 2006.

Tilak, S., Abu-Ghazaleh, N., and Heinzelman, W. "A taxonomy of wireless micro-sensor network models." *ACM SIGMOBILE Mobile Computing and Communications Review* 6.2, 28-36, 2002.

Tolle, G., Polastre, J., Szewczyk, R., Turner, N., Tu, K., Burgess, S., Gay, D., Buonadonna, P., Hong, W., Dawson, T., and Culler, D. "A macroscope in the redwoods." *SenSys '05 Proceedings of the 3rd international conference on Embedded networked sensor systems*. New York, NY, USA: ACM, 2005.