



Comprehensive Survey of Routing Techniques in Wireless Sensor Networks

Nidhi Sharma, Shikha Sharma,
Assistant bProfessor (CSE), SDDIET, Barwala Panchkula, Haryana, India,
Sharma.sunidhi2000@gmail.com
M.tech Research Sclar, , CSE Deptt. SSIET, Derabasi, PTU, Jalandhar, Punjab, India
SharmaShikha299@gmail.com

ABSTRACT

Wireless Sensor Networks (WSNs) consist of small nodes with sensing, computation, and wireless communications capabilities. Many routing, power management, and data dissemination protocols have been specifically designed for WSNs where energy awareness is an essential design issue. The focus, however, has been given to the routing protocols which might differ depending on the application and network architecture. In this paper, we present a survey of the state-of-the-art routing techniques in WSNs. We first outline the design challenges for routing protocols in WSNs followed by a comprehensive survey of different routing techniques. Overall, the routing techniques are classified into three categories based on the underlying network structure: flat, hierarchical, and location-based routing. Furthermore, these protocols can be classified into multipath-based, query-based, negotiation-based, QoS-based, and coherent-based depending on the protocol operation. We study the design tradeoffs between energy and communication overhead savings in every routing paradigm. We also highlight the advantages and performance issues of each routing technique.

Keywords

WSNs(Wireless sensor networks), QoS(Quality of service), Routing Protocols.

Council for Innovative Research

Peer Review Research Publishing System

Journal: [International Journal of Management & Information Technology](#)

Vol. 5, No. 2

editor@cirworld.com

www.cirworld.com, member.cirworld.com

Computer Networks

A **computer network** or **data network** is a telecommunications network that allows computers to exchange data. The connections (network links) between networked computing devices (network nodes) are established using either cable media or wireless media. The best-known computer network is the Internet. Network devices that originate, route and terminate the data are called network nodes.^[1] Nodes can include hosts such as servers and personal computers, as well as networking hardware. Two devices are said to be networked when a process in one device is able to exchange information with a process in another device. Computer networks support applications such as access to the World Wide Web, shared use of application and storage servers, printers, and fax machines, and use of email and instant messaging applications.

WIRELESS SENSOR NETWORKS

A **wireless sensor network (WSN)** consists of spatially distributed autonomous sensors to *monitor* physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a main location. The more modern networks are bi-directional, also enabling *control* of sensor activity. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance; today such networks are used in many industrial and consumer applications, such as industrial process monitoring and control, machine health monitoring, and so on. The WSN is built of "nodes" – from a few to several hundreds or even thousands, where each node is connected to one (or sometimes several) sensors. Each such sensor network node has typically several parts: a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting. A sensor node might vary in size from that of a shoebox down to the size of a grain of dust, although functioning "motes" of genuine microscopic dimensions have yet to be created. The cost of sensor nodes is similarly variable, ranging from a few to hundreds of dollars, depending on the complexity of the individual sensor nodes. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth. In computer science and telecommunications, wireless sensor networks are an active research area with numerous workshops and conferences .

I. Introduction

Due to recent technological advances, the manufacturing of small and low cost sensors became technically and economically feasible. A Wireless Sensor Network (WSN) contain hundreds or thousands of these sensor nodes. These sensors have the ability to communicate either among each other or directly to an external base-station (BS). A greater number of sensors allows for sensing over larger geographical regions with greater accuracy. Figure 1 shows the schematic diagram of sensor node components. Basically, each sensor node comprises sensing, processing, transmission, mobilize, position finding system, and power units (some of these components are optional like the mobilize). 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. Sensor nodes coordinate among themselves to produce high-quality information about the physical environment. Each sensor node bases its decisions on its mission, the information it currently has, and its knowledge of its computing, communication, and energy resources. 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.

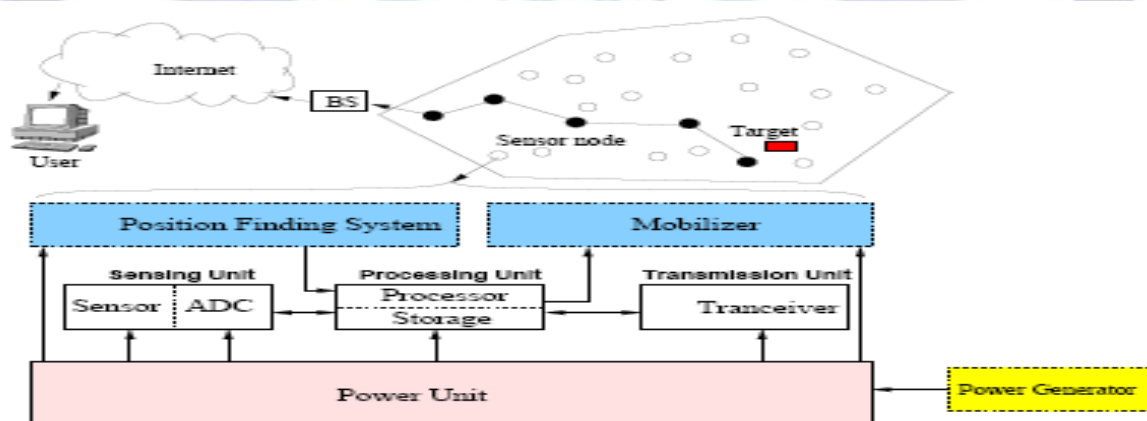


Figure 1: The components of a sensor nodes

Routing in WSNs is very challenging due to the inherent characteristics that distinguish these networks from other wireless networks like mobile ad hoc networks or cellular networks

Although there are some previous efforts for surveying the characteristics, applications, and communication protocols in WSNs [4, 37], the scope of the survey presented in this paper is distinguished from these surveys in many aspects. The surveys in [1] and [16] addressed several design issues and techniques for WSNs describing



the physical constraints on sensor nodes, applications, architectural attributes, and the protocols proposed in all layers of the network stack. However, these surveys were not devoted to routing only. Due to the importance of routing in WSNs and the availability of a significant body of literature on this topic, a detailed survey becomes necessary and useful at this stage. Our work is a dedicated study of the network layer, describing and categorizing the different approaches for data routing.

II. Routing Challenges and Design Issues in WSNs

Despite the innumerable applications of WSNs, these networks have several restrictions, e.g., limited energy supply, limited computing power, and limited bandwidth of the wireless links connecting sensor nodes. One of the main design goals of WSNs is to carry out data communication while trying to prolong the lifetime of the network and prevent connectivity degradation by employing aggressive energy management techniques. The design of routing protocols in WSNs is influenced by many challenging factors.

Node deployment: Node deployment in WSNs is application dependent and affects the performance of the routing protocol. The deployment can be either deterministic or randomized. In deterministic deployment, the sensors are manually placed and data is routed through pre-determined paths. However, in random node deployment, the sensor nodes are scattered randomly creating an infrastructure in an ad hoc manner. If the resultant distribution of nodes is not uniform, optimal clustering becomes necessary to allow connectivity and enable energy efficient network operation.

Energy consumption without losing accuracy: sensor nodes can use up their limited supply of energy performing computations and transmitting information in a wireless environment. As such, energy- conserving forms of communication and computation are essential. Sensor node lifetime shows a strong dependence on the battery lifetime. In a multihop WSN, each node plays a dual role as data sender and data router. The malfunctioning of some sensor nodes due to power failure can cause significant topological changes and might require rerouting of packets and reorganization of the network.

Data Reporting Model: Data sensing and reporting in WSNs is dependent on the application and the time criticality of the data reporting. Data reporting can be categorized as either time-driven (continuous), event-driven, query-driven, and hybrid. The time-driven delivery model is suitable for applications that require periodic data monitoring. As such, sensor nodes will periodically switch on their sensors and transmitters, sense the environment and transmit the data of interest at constant periodic time intervals. In event-driven and query-driven models, sensor nodes react immediately to sudden and drastic changes in the value of a sensed attribute due to the occurrence of a certain event or a query is generated by the BS. The routing protocol is highly influenced by the data reporting model with regard to energy consumption and route stability.

Node/Link Heterogeneity: In many studies, all sensor nodes were assumed to be homogeneous, i.e., having equal capacity in terms of computation, communication, and power. However, depending on the application a sensor node can have different role or capability. The existence of heterogeneous set of sensors raises many technical issues related to data routing. These special sensors can be either deployed independently or the different functionalities can be included in the same sensor nodes. Even data reading and reporting can be generated from these sensors at different rates, subject to diverse quality of service constraints, and can follow multiple data reporting models. For example, hierarchical protocols designate a cluster-head node different from the normal sensors. These cluster heads can be chosen from the deployed sensors or can be more powerful than other sensor nodes in terms of energy, bandwidth, and memory. Hence, the burden of transmission to the BS is handled by the set of cluster-heads.

Fault Tolerance: Some sensor nodes may fail or be blocked due to lack of power, physical damage, or environmental interference. The failure of sensor nodes should not affect the overall task of the sensor network. If many nodes fail, MAC and routing protocols must accommodate formation of new links and route to the data collection base stations. This may require actively adjusting transmit powers and signaling rates on the existing links to reduce energy consumption, or rerouting packets through regions of the network where more energy is available. Therefore, multiple levels of redundancy may be needed in a fault-tolerant sensor network.

Scalability: The number of sensor nodes deployed in the sensing area may be in the order of hundreds or thousands, or more. Any routing scheme must be able to work with this huge number of sensor nodes. In addition, sensor network routing protocols should be scalable enough to respond to events in the environment. Until an event occurs, most of the sensors can remain in the sleep state, with data from the few remaining sensors providing a coarse quality.

Network Dynamics: Most of the network architectures assume that sensor nodes are stationary. However, mobility of both BS's or sensor nodes is sometimes necessary in many applications. Routing messages from or to moving nodes is more challenging since route stability becomes an important issue, in addition to energy, bandwidth etc. Moreover, the sensed phenomenon can be either dynamic or static depending on the application, e.g., it is dynamic in a target detection/tracking application, while it is static in forest monitoring for early fire prevention. Monitoring static events allows the network to work in a reactive mode, simply generating traffic when reporting. Dynamic events in most applications require periodic reporting and consequently generate significant traffic to be routed to the BS.

Transmission Media: In a multi-hop sensor network, communicating nodes are linked by a wireless medium.

The traditional problems associated with a wireless channel (e.g., fading, high error rate) may also affect the operation of the sensor network. In general, the required bandwidth of sensor data will be low, on the order of 1-100 kb/s. Related to the transmission media is the design of medium access control (MAC). One approach of MAC design for sensor networks is to use TDMA based protocols that conserve more energy compared to contention based protocols like CSMA (e.g., IEEE 802.11).

Connectivity: High node density in sensor networks precludes them from being completely isolated from each other. Therefore, sensor nodes are expected to be highly connected. This, however, may not prevent the network topology from being variable and the network size from being shrinking due to sensor node failures. In addition, connectivity depends on the, possibly random, distribution of nodes.

Coverage: In WSNs, each sensor node obtains a certain view of the environment. A given sensor's view of the environment is limited both in range and in accuracy; it can only cover a limited physical area of the environment. Hence, area coverage is also an important design parameter in WSNs.

Data Aggregation: Since sensor nodes may generate significant redundant data, similar packets from multiple nodes can be aggregated so that the number of transmissions is reduced. Data aggregation is the combination of data from different sources according to a certain aggregation function, e.g., duplicate suppression, minima, maxima and average. This technique has been used to achieve energy efficiency and data transfer optimization in a number of routing protocols. Signal processing methods can also be used for data aggregation. In this case, it is referred to as data fusion where a node is capable of producing a more accurate output signal by using some techniques such as beamforming to combine the incoming signals and reducing the noise in these signals.

Quality of Service: In some applications, data should be delivered within a certain period of time from the moment it is sensed, otherwise the data will be useless. Therefore bounded latency for data delivery is another condition for time-constrained applications. However, in many applications, conservation of energy, which is directly related to network lifetime, is considered relatively more important than the quality of data sent. As the energy gets depleted, the network may be required to reduce the quality of the results in order to reduce the energy dissipation in the nodes and hence lengthen the total network lifetime. Hence, energy-aware routing protocols are required to capture this requirement.

III. Routing Protocols in WSNs

In this section, we survey the state-of-the-art routing protocols for WSNs. In general, routing in WSNs can be divided into flat-based routing, hierarchical-based routing, and location-based routing depending on the network structure. In flat-based routing, all nodes are typically assigned equal roles or functionality. In hierarchical-based routing, however, nodes will play different roles in the network. In location-based routing, sensor nodes' positions are exploited to route data in the network. A routing protocol is considered adaptive if certain system parameters can be controlled in order to adapt to the current network conditions and available energy levels. Furthermore, these protocols can be classified into multipath-based, query-based, negotiation-based, QoS-based, or coherent-based routing techniques depending on the protocol operation. In addition to the above, routing protocols can be classified into three categories, namely, proactive, reactive, and hybrid protocols depending on how the source finds a route to the destination. In proactive protocols, all routes are computed before they are really needed, while in reactive protocols, routes are computed on demand. Hybrid protocols use a combination of these two ideas. When sensor nodes are static, it is preferable to have table driven routing protocols rather than using reactive protocols. A significant amount of energy is used in route discovery and setup of reactive protocols. Another class of routing protocols is called the cooperative routing protocols. In cooperative routing, nodes send data to a central node where data can be aggregated and may be subject to further processing, hence reducing route cost in terms of energy use. Many other protocols rely on timing and position information. We also shed some light on these types of protocols in this paper. In order to streamline this survey, we use a classification according to the network structure and protocol operation (routing criteria). The classification is shown in Figure 2 where numbers in the figure indicate the references

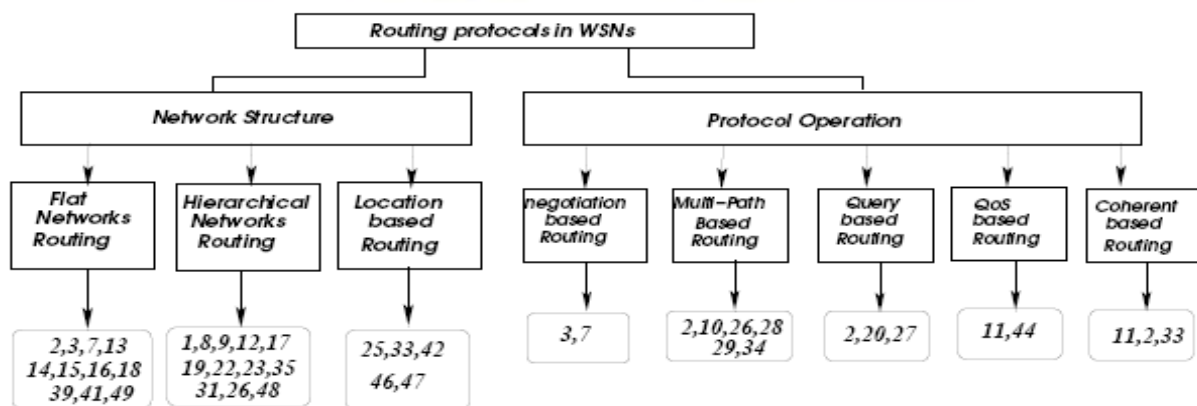


Figure 2: Routing protocols in WSNs



In the rest of this section, we present a detailed overview of the main routing paradigms in WSNs. We start with network structure based protocols.

3.1 Network Structure Based Protocols

The underlying network structure can play significant role in the operation of the routing protocol in WSNs. In this section, we survey in details most of the protocols that fall below this category.

3.1.1 Flat Routing

The first category of routing protocols are the multihop flat routing protocols. In flat networks, each node typically plays the same role and sensor nodes collaborate together to perform the sensing task. Due to the large number of such nodes, it is not feasible to assign a global identifier to each node. This consideration has led to data centric routing, where the BS sends queries to certain regions and waits for data from the sensors located in the selected regions. Since data is being requested through queries, attribute-based naming is necessary to specify the properties of data. Early works on data centric routing, e.g., SPIN and directed diffusion were shown to save energy through data negotiation and elimination of redundant data. These two protocols motivated the design of many other protocols which follow a similar concept. In the rest of this subsection, we summarize these protocols and highlight their advantages and their performance issues.

Sensor Protocols for Information via Negotiation (SPIN): Heinzelman et.al. in [3] and [7] proposed a family of adaptive protocols called Sensor Protocols for Information via Negotiation (SPIN) that disseminate all the information at each node to every node in the network assuming that all nodes in the network are potential base-stations. This enables a user to query any node and get the required information immediately. The SPIN family is designed to address the deficiencies of classic flooding by negotiation and resource adaptation. The SPIN family of protocols is designed based on two basic ideas:

1. Sensor nodes operate more efficiently and conserve energy by sending data that describe the sensor data instead of sending all the data; for example, image and sensor nodes must monitor the changes in their energy resources.

2. Conventional protocols like flooding or gossiping based routing protocols waste energy and bandwidth when sending extra and un-necessary copies of data by sensors covering overlapping areas. The drawbacks of flooding include implosion, which is caused by duplicate messages sent to the same node, overlap when two nodes sensing the same region will send similar packets to the same neighbor, and resource blindness by consuming large amounts of energy without consideration for the energy constraints. Gossiping avoids the problem of implosion by just selecting a random node to send the packet to rather than broadcasting the packet blindly. However, this causes delays in propagation of data through the nodes.

Directed Diffusion: In [2], C. Intanagonwiwat et.al. proposed a popular data aggregation paradigm for WSNs, called directed diffusion. Directed diffusion is a data-centric (DC) and application-aware paradigm in the sense that all data generated by sensor nodes is named by attribute-value pairs. The main idea of the DC paradigm is to combine the data coming from different sources enroute (in-network aggregation) by eliminating redundancy, minimizing the number of transmissions; thus saving network energy and prolonging its lifetime. Unlike traditional end-to-end routing, DC routing finds routes from multiple sources to a single destination that allows in-network consolidation of redundant

Rumor routing: Rumor routing [4] is a variation of directed diffusion and is mainly intended for applications where geographic routing is not feasible. In general, directed diffusion uses flooding to inject the query to the entire network when there is no geographic criterion to diffuse tasks. However, in some cases there is only a little amount of data requested from the nodes and thus the use of flooding is unnecessary. An alternative approach is to flood the events if the number of events is small and the number of queries is large. The key idea is to route the queries to the nodes that have observed a particular event rather than flooding the entire network to retrieve information about the occurring events. In order to flood events through the network, the rumor routing algorithm employs long-lived packets, called agents. When a node detects an event, it adds such event to its local table, called events table, and generates an agent. Agents travel the network in order to propagate information about local events to distant nodes. When a node generates a query for an event, the nodes that know the route, may respond to the query by inspecting its event table. Hence, there is no need to flood the whole network, which reduces the communication cost. On the other hand, rumor routing maintains only one path between source and destination as opposed to directed diffusion where data can be routed through multiple paths at low rates. Simulation results showed that rumor routing can achieve significant energy savings when compared to event flooding and can also handle node's failure. However, rumor routing performs well only when the number of events is small. For a large number of events, the cost of maintaining agents and event-tables in each node becomes infeasible if there is not enough interest in these events from the BS. Moreover, the overhead associated with rumor routing is controlled by different parameters used in the algorithm such as time-to-live (TTL) pertaining to queries and agents. Since the nodes become aware of events through the event agents, the heuristic for defining the route of an event agent highly affects the performance of next hop selection in rumor routing.

Gradient-Based Routing: Schurgers et al. [5] proposed another variant of directed diffusion, called Gradient-Based Routing (GBR). The key idea in GBR is to memorize the number of hops when the interest is



diffused through the whole network. As such, each node can calculate a parameter called the height of the node, which is the minimum number of hops to reach the BS. The difference between a node's height and that of its neighbor is considered the gradient on that link. A packet is forwarded on a link with the largest gradient. GBR uses some auxiliary techniques such as data aggregation and traffic spreading in order to uniformly divide the traffic over the network. When multiple paths pass through a node, which acts as a relay node, that relay node may combine data according to a certain function. In GBR, three different data dissemination techniques have been discussed (1) Stochastic Scheme, where a node picks one gradient at random when there are two or more next hops that have the same gradient, (2) Energy-based scheme, where a node increases its height when its energy drops below a certain threshold, so that other sensors are discouraged from sending data to that node, and (3) Stream-based scheme, where new streams are not routed through nodes that are currently part of the path of other streams. The main objective of these schemes is to obtain a balanced distribution of the traffic in the network, thus increasing the network lifetime. Simulation results of GBR showed that GBR outperforms directed diffusion in terms of total communication energy.

Information-driven sensor querying (IDSQ) and Constrained anisotropic diffusion routing (CADR:) Two routing techniques, namely, information-driven sensor querying (IDSQ) and constrained anisotropic diffusion routing (CADR) were proposed in [16]. CADR aims to be a general form of directed diffusion. The key idea is to query sensors and route data in the network such that the information gain is maximized while latency and bandwidth are minimized. CADR diffuses queries by using a set of information criteria to select which sensors can get the data. This is achieved by activating only the sensors that are close to a particular event and dynamically adjusting data routes. The main difference from directed diffusion is the consideration of information gain in addition to the communication cost. In CADR, each node evaluates an information/cost objective and routes data based on the local information/cost gradient and end-user requirements. Estimation theory was used to model information utility measure. In IDSQ, the querying node can determine which node can provide the most useful information with the additional advantage of balancing the energy cost. However, IDSQ does not specifically define how the query and the information are routed between sensors and the BS. Therefore, IDSQ can be seen as a complementary optimization procedure. Simulation results showed that these approaches are more energy-efficient than directed diffusion where queries are diffused in an isotropic fashion and reaching nearest neighbors first.

Energy Aware Routing: The objective of energy-aware routing protocol, a destination initiated reactive protocol, is to increase the network lifetime. Although this protocol is similar to directed diffusion, it differs in the sense that it maintains a set of paths instead of maintaining or enforcing one optimal path at higher rates. These paths are maintained and chosen by means of a certain probability. The value of this probability depends on how low the energy consumption of each path can be achieved. By having paths chosen at different times, the energy of any single path will not deplete quickly. This can achieve longer network lifetime as energy is dissipated more equally among all nodes. Network survivability is the main metric of this protocol. The protocol assumes that each node is addressable through a class-based addressing which includes the location and types of the nodes. The protocol initiates a connection through localized flooding, which is used to discover all routes between source/destination pair and their costs; thus building up the routing tables. The high-cost paths are discarded and a forwarding table is built by choosing neighboring nodes in a manner that is proportional to their cost. Then, forwarding tables are used to send data to the destination with a probability that is inversely proportional to the node cost. Localized flooding is performed by the destination node to keep the paths alive. When compared to directed diffusion, this protocol provides an overall improvement of 21.5% energy saving and a 44% increase in network lifetime. However, the approach requires gathering the location information and setting up the addressing mechanism for the nodes, which complicate route setup compared to the directed diffusion.

Routing Protocols with Random Walks: The objective of random walks based routing technique [20] is to achieve load balancing in a statistical sense and by making use of multi-path routing in WSNs. This technique considers only large scale networks where nodes have very limited mobility. In this protocol, it is assumed that sensor nodes can be turned on or off at random times. Further, each node has a unique identifier but no location information is needed. Nodes were arranged such that each node falls exactly on one crossing point of a regular grid on a plane, but the topology can be irregular. To find a route from a source to its destination, the location information or lattice coordination is obtained by computing distances between nodes using the distributed asynchronous version of the well-known Bellman-Ford algorithm. An intermediate node would select as the next hop the neighboring node that is closer to the destination according to a computed probability. By carefully manipulating this probability, some kind of load balancing can be obtained in the network. The routing algorithm is simple as nodes are required to maintain little state information. Moreover, different routes are chosen at different times even for the same pair of source and destination nodes. However, the main concern about this protocol is that the topology of the network may not be practical.

3.1.2 Hierarchical Routing

Hierarchical or cluster-based routing, originally proposed in wireline networks, are well-known techniques with special advantages related to scalability and efficient communication. As such, the concept of hierarchical routing is also utilized to perform energy-efficient routing in WSNs. In a hierarchical architecture, higher energy nodes can be used to process and send the information while low energy nodes can be used to perform the sensing in the proximity of the target. This means that creation of clusters and assigning special tasks to cluster heads

can greatly contribute to overall system scalability, lifetime, and energy efficiency. Hierarchical routing is an efficient way to lower energy consumption within a cluster and by performing data aggregation and fusion in order to decrease the number of transmitted messages to the BS. Hierarchical routing is mainly two-layer routing where one layer is used to select clusterheads and the other layer is used for routing. However, most techniques in this category are not about routing, rather on "who and when to send or process/aggregate" the information, channel allocation etc., which can be orthogonal to the multihop routing function.

LEACH protocol: Heinzelman, et. al. [1] introduced a hierarchical clustering algorithm for sensor networks, called Low Energy Adaptive Clustering Hierarchy (LEACH). LEACH is a cluster-based protocol, which includes distributed cluster formation. LEACH randomly selects a few sensor nodes as clusterheads (CHs) and rotate this role to evenly distribute the energy load among the sensors in the network. In LEACH, the clusterhead (CH) nodes compress data arriving from nodes that belong to the respective cluster, and send an aggregated packet to the base station in order to reduce the amount of information that must be transmitted to the base station. LEACH uses a TDMA/CDMA MAC to reduce inter-cluster and intra-cluster collisions. However, data collection is centralized and is performed periodically. Therefore, this protocol is most appropriate when there is a need for constant monitoring by the sensor network. A user may not need all the data immediately. Hence, periodic data transmissions are unnecessary which may drain the limited energy of the sensor nodes. After a given interval of time, a randomized rotation of the role of the CH is conducted so that uniform energy dissipation in the sensor network is obtained. The authors found, based on their simulation model, that only 5% of the nodes need to act as cluster head.

Power-Efficient Gathering in Sensor Information Systems (PEGASIS): In [6], an enhancement over LEACH protocol was proposed. The protocol, called Power-Efficient Gathering in Sensor Information Systems (PEGASIS), is a near optimal chain-based protocol. The basic idea of the protocol is that in order to extend network lifetime, nodes need only communicate with their closest neighbors and they take turns in communicating with the base-station. When the round of all nodes communicating with the base-station ends, a new round will start and so on. This reduces the power required to transmit data per round as the power draining is spread uniformly over all nodes. Hence, PEGASIS has two main objectives. First, increase the lifetime of each node by using collaborative techniques and as a result the network lifetime will be increased. Second, allow only local coordination between nodes that are close together so that the bandwidth consumed in communication is reduced. Unlike LEACH, PEGASIS avoids cluster formation and uses only one node in a chain to transmit to the BS instead of using multiple nodes.

Threshold-sensitive Energy Efficient Protocols (TEEN and APTEEN): Two hierarchical routing protocols called TEEN (Threshold-sensitive Energy Efficient sensor Network protocol), and APTEEN (Adaptive Periodic Threshold-sensitive Energy Efficient sensor Network protocol) are proposed in [2], respectively. These protocols were proposed for time-critical applications. In TEEN, sensor nodes sense the medium continuously, but the data transmission is done less frequently. A cluster head sensor sends its members a hard threshold, which is the threshold value of the sensed attribute and a soft threshold, which is a small change in the value of the sensed attribute that triggers the node to switch on its transmitter and transmit. Thus the hard threshold tries to reduce the number of transmissions by allowing the nodes to transmit only when the sensed attribute is in the range of interest. The soft threshold further reduces the number of transmissions that might have otherwise occurred when there is little or no change in the sensed attribute. A smaller value of the soft threshold gives a more accurate picture of the network, at the expense of increased energy consumption. Thus, the user can control the trade-off between energy efficiency and data accuracy. When cluster-heads are to change (see Figure 5(a)), new values for the above parameters are broadcast. The main drawback of this scheme is that, if the thresholds are not received, the nodes will never communicate, and the user will not get any data from the network at all.

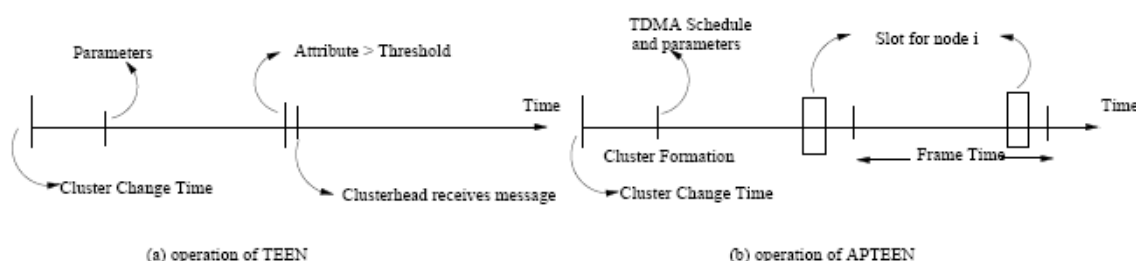


Figure 5: Time line for the operation of (a) TEEN and (b) APTEEN

Self Organizing Protocol (SOP): Subramanian et al. [12] describes a self-organizing protocol and an application taxonomy that was used to build architecture used to support heterogeneous sensors. Furthermore, these sensors can be mobile or stationary. Some sensors probe the environment and forward the data to a designated set of nodes that act as routers. Router nodes are stationary and form the backbone for communication. Collected data are forwarded through the routers to the more powerful BS nodes. Each



sensing node should be able to reach a router in order to be part of the network. A routing architecture that requires addressing of each sensor node has been proposed. This protocol, however, is not an on-demand protocols especially in the organization phase of algorithm. Therefore, introducing extra overhead. Another issue is related to the formation of hierarchy. It could happen that there are many cuts in the network, and hence the probability of applying reorganization phase increases, which will be an expensive operation.

Sensor Aggregates Routing: In [15], a set of algorithms for constructing and maintaining sensor aggregates were proposed. The objective is to collectively monitor target activity in a certain environment (target tracking applications). A sensor aggregate comprises those nodes in a network that satisfy a grouping predicate for a collaborative processing task. The parameters of the predicate depend on the task and its resource requirements. The formation of appropriate sensor aggregates were discussed in [15] in terms of allocating resources to sensing and communication tasks. Sensors in a sensor field is divided into clusters according to their sensed signal strength, so that there is only one peak per cluster. Then, local cluster leaders are elected. One peak may represent one target, multiple targets, or no target in case the peak is generated by noise sources. To elect a leader, information exchanges between neighboring sensors are necessary. If a sensor, after exchanging packets with all its one-hop neighbors, finds that it is higher than all its one-hop neighbors on the signal field landscape, it declares itself a leader. This leader-based tracking algorithm assumes the unique leader knows the geographical region of the collaboration.

Virtual Grid Architecture routing (VGA): An energy-efficient routing paradigm is proposed in [13] that utilizes data aggregation and in-network processing to maximize the network lifetime. Due to the node stationarity and extremely low mobility in many applications in WSNs, a reasonable approach is to arrange nodes in a fixed topology as was briefly mentioned in [10]. A GPS-free approach is used to build clusters that are fixed, equal, adjacent, and non-overlapping with symmetric shapes. In [13], square clusters were used to obtain a fixed rectilinear virtual topology. Inside each zone, a node is optimally selected to act as clusterhead. Data aggregation is performed at two levels: local and then global. The set of clusterheads, also called Local Aggregators (LAs), perform the local aggregation, while a subset of these LAs are used to perform global aggregation. However, the determination of an optimal selection of global aggregation points, called Master Aggregators (MAs), is NP-hard problem.

Hierarchical Power-aware Routing (HPAR): In [11], a hierarchical power-aware routing was proposed. The protocol divides the network into groups of sensors. Each group of sensors in geographic proximity are clustered together as a zone and each zone is treated as an entity. To perform routing, each zone is allowed to decide how it will route a message hierarchically across the other zones such that the battery lives of the nodes in the system are maximized. Message are routed along the path which has the maximum over all the minimum of the remaining power, called the max-min path. The motivation is that using nodes with high residual power may be expensive as compared to the path with the minimal power consumption. An approximation algorithm, called the max-min zPmin algorithm, was proposed in [11]. The crux of the algorithm is based on the tradeoff between minimizing the total power consumption and maximizing the minimal residual power of the network. Hence, the algorithm tries to enhance a max-min path by limiting its power consumption as follows.

First, the algorithm finds the path with the least power consumption (P_{min}) by using the Dijkstra algorithm. Second, the algorithm finds a path that maximizes the minimal residual power in the network. The proposed algorithm tries to optimize both solution criteria. This is achieved by relaxing the minimal power consumption for the message to be equal to zPmin with parameter $z \geq 1$ to restrict the power consumption for sending one message to zPmin. The algorithm consumes at most zPmin while maximizing the minimal residual power fraction.

3.1.3 Location based routing protocols

In this kind of routing, sensor nodes are addressed by means of their locations. The distance between neighboring nodes can be estimated on the basis of incoming signal strengths. Relative coordinates of neighboring nodes can be obtained by exchanging such information between neighbors [7], [19], [12]. Alternatively, the location of nodes may be available directly by communicating with a satellite, using GPS (Global Positioning System), if nodes are equipped with a small low power GPS receiver [10]. To save energy, some location based schemes demand that nodes should go to sleep if there is no activity. More energy savings can be obtained by having as many sleeping nodes in the network as possible. The problem of designing sleep period schedules for each node in a localized manner was addressed in [33, 25]. In the rest of this section, we review most of the location or geographic based routing protocols.

Geographic Adaptive Fidelity (GAF): GAF [10] is an energy-aware location-based routing algorithm designed primarily for mobile ad hoc networks, but may be applicable to sensor networks as well. The network area is first divided into fixed zones and form a virtual grid. Inside each zone, nodes collaborate with each other to play different roles. For example, nodes will elect one sensor node to stay awake for a certain period of time and then they go to sleep. This node is responsible for monitoring and reporting data to the BS on behalf of the nodes in the zone. Hence, GAF conserves energy by turning off unnecessary nodes in the network without affecting the level of routing fidelity. Each node uses its GPS-indicated location to associate itself with a point in the virtual grid. Nodes associated with the same point on the grid are considered equivalent in terms of the cost of packet routing. Such equivalence is exploited in keeping some nodes located in a particular grid area



in sleeping state in order to save energy. Thus, GAF can substantially increase the network lifetime as the number of nodes increases. There are three states defined in GAF. These states are discovery, for determining the neighbors in the grid, active reflecting participation in routing and sleep when the radio is turned off. In order to handle the mobility, each node in the grid estimates its leaving time of grid and sends this to its neighbors. The sleeping neighbors adjust their sleeping time accordingly in order to keep the routing fidelity. Before the leaving time of the active node expires, sleeping nodes wake up and one of them becomes active. GAF is implemented both for non-mobility (GAF- basic) and mobility (GAF-mobility adaptation) of nodes. Figure 7 shows an example of fixed zoning that can be used in sensor networks similar to the one proposed in [10].

Geographic and Energy Aware Routing (GEAR): Yu et al. [19] discussed the use of geographic information while disseminating queries to appropriate regions since data queries often include geographic attributes. The protocol, called Geographic and Energy Aware Routing (GEAR), uses energy aware and geographically-informed neighbor selection heuristics to route a packet towards the destination region. The key idea is to restrict the number of interests in directed diffusion by only considering a certain region rather than sending the interests to the whole network. By doing this, GEAR can conserve more energy than directed diffusion. Each node in GEAR keeps an estimated cost and a learning cost of reaching the destination through its neighbors. The estimated cost is a combination of residual energy and distance to destination. The learned cost is a refinement of the estimated cost that accounts for routing around holes in the network. A hole occurs when a node does not have any closer neighbor to the target region than itself. If there are no holes, the estimated cost is equal to the learned cost. The learned cost is propagated one hop back every time a packet reaches the destination so that route setup for next packet will be adjusted.

MFR, DIR, and GEDIR: Stojmenovic and Lin described and discussed basic localized routing algorithms. These protocols deal with basic distance, progress, and direction based methods. The key issues are forward direction and backward direction. A source node or any intermediate node will select one of its neighbors according to a certain criterion. The routing methods, which belong to this category, are MFR (Most Forward within Radius), GEDIR (The Geographic Distance Routing) that is a variant of greedy algorithms, 2-hop greedy method, alternate greedy method and DIR (compass routing method). GEDIR algorithm is a greedy algorithm that always moves the packet to the neighbor of the current vertex whose distance to the destination is minimized. The algorithm fails when the packet crosses the same edge twice in succession. In most cases, the MFR and Greedy methods have the same path to destination. In the DIR method, the best neighbor has the closest direction (that is, angle) toward the destination. That is, the neighbor with the minimum angular distance from the imaginary line joining the current node and the destination is selected. In MFR method, the best neighbor A will minimize the dot product $DA \cdot DS$, where S, D are the source and destination nodes, respectively, and SD represents the Euclidian distance between the two nodes S, D. Alternatively, one can maximize the dot product $SD \cdot SA$. Each method stops forwarding the message at a node for which the best choice is to return the message back to a previous node. GEDIR and MFR methods are loop-free, while DIR method may create loops, unless past traffic is memorized or a time-stamp is enforced.

The Greedy Other Adaptive Face Routing (GOAFR): In a geometric ad-hoc routing algorithm combining greedy and face routing was proposed. We will now briefly review the key points of GOAFR in this section. The greedy algorithm of GOAFR always picks the neighbor closest to a node to be next node for routing. However, it can be easily stuck at some local minimum, i.e. no neighbor is closer to a node than the current node. Other Face Routing (OFR) is a variant of Face Routing (FR). The Face Routing (FR) algorithm is the first one that guarantees success if the source and the destination are connected.

SPAN: Another position based algorithm called SPAN selects some nodes as coordinators based on their positions. The coordinators form a network backbone that is used to forward messages. A node should become a coordinator if two neighbors of a non-coordinator node cannot reach each other directly or via one or two coordinators (3 hop reachability). New and existing coordinators are not necessarily neighbors in which, in effect, makes the design less energy efficient because of the need to maintain the positions of two or three hop neighbors in the complicated SPAN algorithm.

3.2 Routing Protocols based on Protocol Operation

In this section, we review routing protocols that different routing functionality. It should be noted that some of these protocols may fall below one or more of the above routing categories.

3.2.1 Multipath routing protocols

In this subsection, we study the routing protocols that use multiple paths rather than a single path in order to enhance the network performance. The fault tolerance (resilience) of a protocol is measured by the likelihood that an alternate path exists between a source and a destination when the primary path fails. This can be increased by maintaining multiple paths between the source and the destination at the expense of an increased energy consumption and traffic generation. These alternate paths are kept alive by sending periodic messages. Hence, network reliability can be increased at the expense of increased overhead of maintaining the alternate paths.

Directed diffusion is a good candidate for robust multipath routing and delivery. Based on the directed diffusion paradigm, a multipath routing scheme that finds several partially disjoint paths is studied in [10] (alternate routes



are not node disjoint, i.e., routes are partially overlapped). It has been found that the use of multipath routing provides viable alternative for energy efficient recovery from failures in WSN. The motivation of using these braided paths is to keep the cost of maintaining the multipaths low. The costs of alternate paths are comparable to the primary path because they tend to be much closer to the primary path.

3.2.2 Query based routing

In this kind of routing, the destination nodes propagate a query for data (sensing task) from a node through the network and a node having this data sends the data which matches the query back to the node, which initiates the query. Usually these queries are described in natural language, or in high-level query languages. For example, client C1 may submit a query to node N1 and ask: Are there moving vehicles in battle space region 1?. All the nodes have tables consisting of the sensing tasks queries that they receive and send data which matches these tasks when they receive it. Directed diffusion described in Section 3.1.1 is an example of this type of routing. In directed diffusion, the BS node sends out interest messages to sensors. As the interest is propagated throughout the sensor network, the gradients from the source back to the BS are set up. When the source has data for the interest, the source sends the data along the interests gradient path. To lower energy consumption, data aggregation (e.g., duplicate suppression) is performed enroute.

3.2.3 Negotiation based routing protocols

These protocols use high level data descriptors in order to eliminate redundant data transmissions through negotiation. Communication decisions are also taken based on the resources that are available to them. The SPIN family protocols discussed earlier and the protocols in are examples of negotiation based routing protocols. The motivation is that the use of flooding to disseminate data will produce implosion and overlap between the sent data, hence nodes will receive duplicate copies of the same data. This operation consumes more energy and more processing by sending the same data by different sensors. The SPIN protocols are designed to disseminate the data of one sensor to all other sensors assuming these sensors are potential base-stations. Hence, the main idea of negotiation based routing in WSNs is to suppress duplicate information and prevent redundant data from being sent to the next sensor or the base-station by conducting a series of negotiation messages before the real data transmission begins.

3.2.4 QoS-based routing

In QoS-based routing protocols, the network has to balance between energy consumption and data quality. In particular, the network has to satisfy certain QoS metrics, e.g., delay, energy, bandwidth, etc. when delivering data to the BS. Sequential Assignment Routing (SAR) proposed in [3] is one of the first routing protocols for WSNs that introduces the notion of QoS in the routing decisions. Routing decision in SAR is dependent on three factors: energy resources, QoS on each path, and the priority level of each packet. To avoid single route failure, a multi-path approach is used and localized path restoration schemes are used. To create multiple paths from a source node, a tree rooted at the source node to the destination nodes (i.e., the set of base-stations (BSs)) is built. The paths of the tree are built while avoiding nodes with low energy or QoS guarantees. At the end of this process, each sensor node will be part of multi-path tree. As such, SAR is table-driven multi-path protocol that aims to achieve energy efficiency and fault tolerance. In essence, SAR calculates a weighted QoS metric as the product of the additive QoS metric and a weight coefficient associated with the priority level of the packet. The objective of SAR algorithm is to minimize the average weighted QoS metric throughout the lifetime of the network. If topology changes due to node failures, a path re-computation is needed. As a preventive measure, a periodic re-computation of paths is triggered by the base-station to account for any changes in the topology. A handshake procedure based on a local path restoration scheme between neighboring nodes is used to recover from a failure. Failure recovery is done by enforcing routing table consistency between upstream and downstream nodes on each path. Simulation results showed that SAR offers less power consumption than the minimum-energy metric algorithm, which focuses only the energy consumption of each packet without considering its priority. SAR maintains multiple paths from nodes to BS. Although, this ensures fault-tolerance and easy recovery, the protocol suffers from the overhead of maintaining the tables and states at each sensor node especially when the number of nodes is huge.

3.2.5 Coherent and non-coherent processing

Data processing is a major component in the operation of wireless sensor networks. Hence, routing techniques employ different data processing techniques. In general, sensor nodes will cooperate with each other in processing different data flooded in the network area. Two examples of data processing techniques proposed in WSNs are coherent and non-coherent data processing-based routing [3]. In non-coherent data processing routing, nodes will locally process the raw data before being sent to other nodes for further processing. The nodes that perform further processing are called the aggregators. In coherent routing, the data is forwarded to aggregators after minimum processing. The minimum processing typically includes tasks like time stamping, duplicate suppression, etc. To perform energy-efficient routing, coherent processing is normally selected.

Non-coherent functions have fairly low data traffic loading. On the other hand, since coherent processing generates long data streams, energy efficiency must be achieved by path optimality. In non-coherent processing, data processing incurs three phases: (1) Target detection, data collection, and preprocessing (2) Membership declaration, and (3) Central node election. During phase 1, a target is detected, its data collected and preprocessed. When a node decides to participate in a cooperative function, it will enter phase 2 and declare



this intention to all neighbors. This should be done as soon as possible so that each sensor has a local understanding of the network topology. Phase 3 is the election of the central node. Since the central node is selected to perform more sophisticated information processing, it must have sufficient energy reserves and computational capability.

In [3], a single and multiple winner algorithms were proposed for non-coherent and coherent processing, respectively. In the single winner algorithm (SWE), a single aggregator node is elected for complex processing. The election of a node is based on the energy reserves and computational capability of that node. By the end of the SWE process, a minimum-hop spanning tree will completely cover the network. In the multiple winner algorithm (MWE), a simple extension to the single winner algorithm (SWE) is proposed. When all nodes are sources and send their data to the central aggregator node, a large amount of energy will be consumed and hence this process has a high cost. One way to lower the energy cost is to limit the number of sources that can send data to the central aggregator node. Instead of keeping record of only the best candidate node (master aggregator node), each node will keep a record of up to n nodes of those candidates. At the end of the MWE process, each sensor in the network has a set of minimum-energy paths to each source node (SN). After that, the single winner algorithm is used to find the node that yields the minimum energy consumption. This node can then serve as the central node for the coherent processing. In general, the MWE process has longer delay, higher overhead, and lower scalability than that for non-coherent processing networks.

IV. Routing in WSNs: Future Directions

The future vision of WSNs is to embed numerous distributed devices to monitor and interact with physical world phenomena, and to exploit spatially and temporally dense sensing and actuation capabilities of those sensing devices. These nodes coordinate among themselves to create a network that performs higher-level tasks.

Although extensive efforts have been exerted so far on the routing problem in WSNs, there are still some challenges that confront effective solutions of the routing problem. First, there is a tight coupling between sensor nodes and the physical world. Sensors are embedded in unattended places or systems. This is different from traditional Internet, PDA, and mobility applications that interface primarily and directly with human users. Second, sensors are characterized by a small foot print, and as such nodes present stringent energy constraints since they are equipped with small, finite, energy source. This is also different from traditional fixed but reusable resources. Third, communications is primary consumer of energy in this as thousands-to-millions of operations (known as R^4 signal energy drop-off). environment where sending a bit over 10 or 100 meters consumes as much energy. Although the performance of these protocols is promising in terms of energy efficiency, further research would be needed to address issues such as Quality of Service (QoS) posed by video and imaging sensors and real-time applications. Energy-aware QoS routing in sensor networks will ensure guaranteed bandwidth (or delay) through the duration of connection as well as providing the use of most energy efficient path. Another interesting issue for routing protocols is the consideration of node mobility.

Most of the current protocols assume that the sensor nodes and the BS are stationary. However, there might be situations such as battle environments where the BS and possibly the sensors need to be mobile. In such cases, the frequent update of the position of the command node and the sensor nodes and the propagation of that information through the network may excessively drain the energy of nodes. New routing algorithms are needed in order to handle the overhead of mobility and topology changes in such energy constrained environment. Future trends in routing techniques in WSNs focus on different directions, all share the common objective of prolonging the network lifetime. We summarize some of these directions and give some pertinent references as follows:

Exploit redundancy: typically a large number of sensor nodes are implanted inside or beside the phenomenon. Since sensor nodes are prone to failure, fault tolerance techniques come in picture to keep the network operating and performing its tasks. Routing techniques that explicitly employ fault tolerance techniques in an efficient manner are still under investigation.

Tiered architectures (mix of form/energy factors): Hierarchical routing is an old technique to enhance scalability and efficiency of the routing protocol. However, novel techniques to network clustering which maximize the network lifetime are also a hot area of research in WSNs.

Exploit spatial diversity and density of sensor/actuator nodes: Nodes will span a network area that might be large enough to provide spatial communication between sensor nodes. Achieving energy efficient communication in this densely populated environment deserves further investigation. The dense deployment of sensor nodes should allow the network to adapt to unpredictable environment.

Achieve desired global behavior with adaptive localized algorithms (i.e., do not rely on global interaction or information). However, in a dynamic environment, this is hard to model.

Leverage data processing inside the network and exploit computation near data sources to reduce communication, i.e., perform in-network distributed processing. WSNs are organized around naming data, not nodes identities. Since we have a large collections of distributed elements, localized algorithms that achieve system-wide properties in terms of local processing of data before being sent to the destination are still needed. Nodes in the network will store named data and make it available for processing. There is a high need to create



efficient processing points in the network, e.g., duplicate suppression, aggregation, correlation of data. How to efficiently and optimally find those points is still an open research issue.

Time and location synchronization: energy-efficient techniques for associating time and spatial coordinates with data to support collaborative processing are also required.

V. Conclusions

Routing in sensor networks is a new area of research, with a limited, but rapidly growing set of research results. In this paper, we presented a comprehensive survey of routing techniques in wireless sensor networks which have been presented in the literature. They have the common objective of trying to extend the lifetime of the sensor network, while not compromising data delivery.

Overall, the routing techniques are classified based on the network structure into three categories: flat, hierarchical, and location based routing protocols. Furthermore, these protocols are classified into multipath-based, query-based, negotiation-based, or QoS-based routing techniques depending on the protocol operation. We also highlight the design tradeoffs between energy and communication overhead savings in some of the routing paradigm, as well as the advantages and disadvantages of each routing technique. Although many of these routing techniques look promising, there are still many challenges that need to be solved in the sensor networks. We highlighted those challenges and pinpointed future research directions in this regard.

References

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communicatio Magazine*, Volume: 40 Issue: 8, pp.102-114, August 2002.
- [2] A. Manjeshwar and D. P. Agarwal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," In 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, April 2001.
- [3] K. Sohrabi, J. Pottie, "Protocols for self-organization of a wireless sensor network", *IEEE Personal Communications*, Volume 7, Issue 5, pp 16-27, 2000.
- [4] D. Braginsky and D. Estrin, "Rumor Routing Algorithm for Sensor Networks," in the Proceedings of the First Workshop on Sensor Networks and Applications (WSNA), Atlanta, GA, October 2002.
- [5] C. Schurgers and M.B. Srivastava, "Energy efficient routing in wireless sensor networks", in the MILCOM Proceedings on Communications for Network-Centric Operations: Creating the Information Force, McLean, VA, 2001.
- [6] S. Lindsey, C. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems", *IEEE Aerospace Conference Proceedings*, 2002, Vol. 3, 9-16 pp. 1125-1130.
- [7] N. Bulusu, J. Heidemann, D. Estrin, "GPS-less low cost outdoor localization for very small devices", Technical report 00-729, Computer science department, University of Southern California, Apr. 2000.
- [8] <http://www.ieee802.org/15/>
- [9] V. Rodoplu and T. H. Meng, "Minimum Energy Mobile Wireless Networks", *IEEE Journal Selected Areas in Communi- cations*, vol. 17, no. 8, Aug. 1999, pp. 1333-44.
- [10] Y. Xu, J. Heidemann, D. Estrin, "Geography-informed Energy Conservation for Ad-hoc Routing," In Proceedings of the Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking 2001, pp. 70-84.
- [11] Q. Li and J. Aslam and D. Rus, "Hierarchical Power-aware Routing in Sensor Networks", In Proceedings of the DIMACS Workshop on Pervasive Networking, May, 2001.
- [12] S. Capkun, M. Hamdi, J. Hubaux, "GPS-free positioning in mobile ad-hoc networks", Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001 pp. 3481-3490.
- [13] Jamal N. Al-Karaki, Raza UI-Mustafa, Ahmed E. Kamal, "Data Aggregation in Wireless Sensor Networks - Exact and Approximate Algorithms", Proceedings of IEEE Workshop on High Performance Switching and Routing (HPSR) 2004, April 18-21, 2004, Phoenix, Arizona, USA.
- [14] <http://www.ieee802.org/15/>
- [15] Q. Fang, F. Zhao, and L. Guibas, "Lightweight Sensing and Communication Protocols for Target Enumeration and Aggre- gation", Proceedings of the 4th ACM international symposium on Mobile ad hoc networking and computing (MOBIHOC), 2003, pp. 165-176.
- [16] S. Tilak, N. Abu-Ghazaleh, W. Heinzelman, "A taxonomy of wireless micro-sensor network models", *ACM SIGMOBILE Mobile Computing and Communications Review*, Volume 6, Issue 2 (April 2002), pp 28-36.
- [17] N. Sadagopan et al., The ACQUIRE mechanism for efficient querying in sensor networks, in the Proceedings of the First International Workshop on Sensor Network Protocol and Applications,



Anchorage, Alaska, May 2003.

- [18] C. Karlof, D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures", *Ad Hoc Networks*, Vol.1 (2003) pp. 293315.
- [19] Y. Yu, D. Estrin, and R. Govindan, "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, May 2001.
- [20] S. Servetto and G. Barrenechea, "Constrained Random Walks on Random Graphs: Routing Algorithms for Large Scale Wireless Sensor Networks", *proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, Atlanta, Georgia, USA, 2002

