6

ROUTING PROTOCOLS FOR WIRELESS SENSOR NETWORKS

6.1 INTRODUCTION

WSNs are extremely versatile and can be deployed to support a wide variety of applications in many different situations, whether they are composed of stationary or mobile sensor nodes. The way these sensors are deployed depends on the nature of the application. In environmental monitoring and surveillance applications, for example, sensor nodes are typically deployed in an ad hoc fashion so as to cover the specific area to be monitored (e.g., C1WSNs). In health care–related applications, smart wearable wireless devices and biologically compatible sensors can be attached to or implanted strategically within the human body to monitor vital signs of the patient under surveillance. Once deployed, sensor nodes self-organize into an autonomous wireless ad hoc network, which requires very little or no maintenance. Sensor nodes then collaborate to carry out the tasks of the application for which they are deployed.

Despite the disparity in the objectives of sensor applications, the main task of wireless sensor nodes is to sense and collect data from a target domain, process the data, and transmit the information back to specific sites where the underlying application resides. Achieving this task efficiently requires the development of an energy-efficient routing protocol to set up paths between sensor nodes and the data sink. The path selection must be such that the lifetime of the network is maximized. The characteristics of the environment within which sensor nodes typically operate, coupled with severe resource and energy limitation, make the routing problem very challenging.

Wireless Sensor Networks: Technology, Protocols, and Applications, by Kazem Sohraby, Daniel Minoli, and Taieb Znati
Copyright © 2007 John Wiley & Sons, Inc.

Our objective in this chapter is to discuss issues central to routing in WSNs and describe different strategies used to develop routing protocols for these networks. To this end, we first discuss a representative class of sensor applications. The goal is to highlight the unique and distinctive features of the nature of the traffic typically generated in WSNs. In the second part of the chapter we provide a brief taxonomy of the basic routing strategies used to strike a balance between responsiveness and energy efficiency. Achieving this balance brings about new challenges that span the network layers in a manner that differs from infrastructured as well as ad hoc wireless networks. In the third part of the chapter we review a number of protocols that address the problem of routing in today's WSNs. Although the field is in its infancy and routing in WSNs remains largely relegated to research, multiple strategies have emerged as workable solutions to the routing problem. As the application of WSNs to different fields becomes more apparent, advances in network hardware and battery technology will pave the way to practical cost-effective implementations of these routing protocols.

6.2 BACKGROUND

WSNs have created new opportunities across the spectrum of human endeavors, including engineering design and manufacturing, monitoring and control of environmental systems, forest fire tracking, health care, battlefield surveillance, disaster management, and critical infrastructure protection. Several applications involving sensed data collection and dissemination are depicted in Figure 6.1, along with

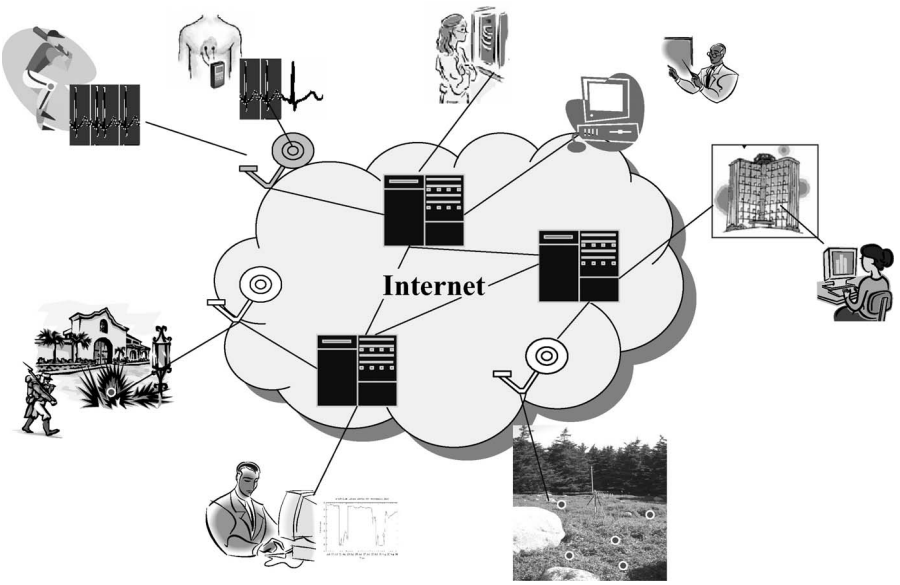

Figure 6.1 Wireless sensor network applications.

the way that data flow from its source to the anticipated sink [6.1–6.7]. Many applications (see Chapter 3) exhibit strong similarity in the way that sensors are used to collect and disseminate data to carry out the objectives for which they are deployed. In these applications, sensors are designed primarily to sense the environment and to record and possibly process the sensor readings before forwarding the information collected, through the base station toward the data sink and eventually, to where the application resides. These readings may be light levels, temperature, vital signs, or levels of environmental disturbance. The process of data collection and forwarding is either triggered by the occurrence of specific events in the environment where the sensors are deployed or is initiated in response to a query issued by the application supported. It is worth noting that in many cases it is useful to aggregate data collected by various sensors before forwarding the data to the base station. Data aggregation reduces the number of messages transmitted, leading to a significant decrease in energy consumption due to communication.

6.3 DATA DISSEMINATION AND GATHERING

The way that data and queries are forwarded between the base station and the location where the target phenomena are observed is an important aspect and a basic feature of WSNs. A simple approach to accomplishing this task is for each sensor node to exchange data directly with the base station. A single-hop-based approach, however, is costly, as nodes that are farther away from the base station may deplete their energy reserves quickly, thereby severely limiting the lifetime of the network. This is the case particularly where the wireless sensors are deployed to cover a large geographical region or where the wireless sensors are mobile and may move away from the base station.

To address the shortcomings of the single-hop approach, data exchange between the sensors and the base stations is usually carried out using multihop packet transmission over short communication radius. Such an approach leads to significant energy savings and reduces considerably communication interference between sensor nodes competing to access the channel, particularly in highly dense WSNs. Data forwarding between the sensors where data are collected and the sinks where data are made available is illustrated in Figure 6.2. In response to queries issued by the sinks or when specific events occur within the area monitored, data collected by the sensors are transmitted to the base station using multihop paths. It is worth noting that depending on the nature of the application, sensor nodes can aggregate data correlated on their way to the base station.

In a multihop WSN, intermediate nodes must participate in forwarding data packets between the source and the destination. Determining which set of intermediate nodes is to be selected to form a data-forwarding path between the source and the destination is the principal task of the routing algorithm. In general, routing in large-scale networks is inherently a difficult problem whose solution must address multiple challenging design requirements, including correctness, stability, and optimality with respect to various performance metrics. The intrinsic properties

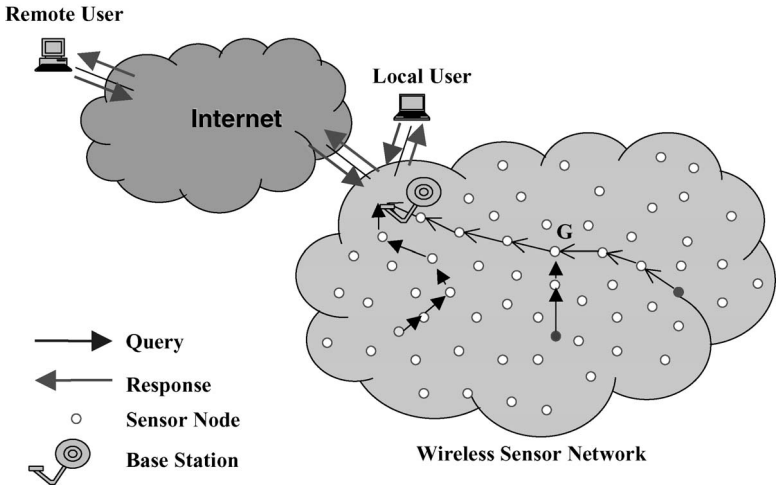

Figure 6.2 Multihop data and query forwarding.

of WSNs, combined with severe energy and bandwidth constraints, bring about additional challenges that must be addressed to satisfy the traffic requirements of the application supported, while extending the lifetime of the network.

In the following sections, we first discuss the primary routing challenges and design goals of routing in WSNs. We then discuss various strategies, approaches, and techniques that have been proposed to design efficient routing protocols for WSNs. Following that, we survey the state of the art in WSN routing protocols.

6.4 ROUTING CHALLENGES AND DESIGN ISSUES IN WIRELESS SENSOR NETWORKS

Although WSNs share many commonalities with wired and ad hoc networks, they also exhibit a number of unique characteristics which set them apart from existing networks. These unique characteristics bring to sharp focus new routing design requirements that go beyond those typically encountered in wired and wireless ad hoc networks. Meeting these design requirements presents a distinctive and unique set of challenges. These challenges can be attributed to multiple factors, including severe energy constraints, limited computing and communication capabilities, the dynamically changing environment within which sensors are deployed, and unique data traffic models and application-level quality of service requirements.

6.4.1 Network Scale and Time-Varying Characteristics

Sensor nodes operate with limited computing, storage, and communication capabilities under severe energy constraints, as discussed in Chapter 4. Due to the large

number of conceivable sensor-based applications, the densities of the WSNs may vary widely, ranging from very sparse to very dense. Furthermore, in many applications, the sensor nodes, in some cases numbering in the hundreds if not thousands, are deployed in an ad hoc and often unsupervised manner over wide coverage areas. In these networks, the behavior of sensor nodes is dynamic and highly adaptive, as the need to self-organize and conserve energy forces sensor nodes to adjust their behavior constantly in response to their current level of activity or the lack thereof. Furthermore, sensor nodes may be required to adjust their behavior in response to the erratic and unpredictable behavior of wireless connections caused by high noise levels and radio-frequency interference, to prevent severe performance degradation of the application supported.

6.4.2 Resource Constraints

Sensor nodes are designed with minimal complexity for large-scale deployment at a reduced cost. Energy is a key concern in WSNs, which must achieve a long lifetime while operating on limited battery reserves. Multihop packet transmission over wireless networks is a major source of power consumption. Reducing energy consumption can be achieved by dynamically controlling the duty cycle of the wireless sensors. The energy management problem, however, becomes especially challenging in many mission-critical sensor applications. The requirements of these applications are such that a predetermined level of sensing and communication performance constraints must be maintained simultaneously. Therefore, a question arises as to how to design scalable routing algorithms that can operate efficiently for a wide range of performance constraints and design requirements. The development of these protocols is fundamental to the future of WSNs.

6.4.3 Sensor Applications Data Models

The data model describes the flow of information between the sensor nodes and the data sink. These models are highly dependent on the nature of the application in terms of how data are requested and used. Several data models have been proposed to address the data-gathering needs and interaction requirements of a variety of sensor applications [6.8,6.9]. A class of sensor applications requires data collection models that are based on periodic sampling or are driven by the occurrence of specific events. In other applications, data can be captured and stored, possibly processed and aggregated by a sensor node, before they are forwarded to the data sink. Yet a third class of sensor applications requires bidirectional data models in which two-way interaction between sensors and data sinks is required [6.10,6.11].

The need to support a variety of data models increases the complexity of the routing design problem. Optimizing the routing protocol for an application's specific data requirements while supporting a variety of data models and delivering the highest performance in scalability, reliability, responsiveness, and power efficiency becomes a design and engineering problem of enormous magnitude.

6.5 ROUTING STRATEGIES IN WIRELESS SENSOR NETWORKS

The WSN routing problem presents a very difficult challenge that can be posed as a classic trade-off between responsiveness and efficiency. This trade-off must balance the need to accommodate the limited processing and communication capabilities of sensor nodes against the overhead required to adapt to these. In a WSN, overhead is measured primarily in terms of bandwidth utilization, power consumption, and the processing requirements on the mobile nodes. Finding a strategy to balance these competing needs efficiently forms the basis of the routing challenge. Furthermore, the intrinsic characteristics of wireless networks gives rise to the important question of whether or not existing routing protocols designed for ad hoc networks are sufficient to meet this challenge [6.12].

Routing algorithms for ad hoc networks can be classified according to the manner in which information is acquired and maintained and the manner in which this information is used to compute paths based on the acquired information. Three different strategies can be identified: proactive, reactive, and hybrid [6.13,6.14]. The *proactive strategy*, also referred to as *table driven*, relies on periodic dissemination of routing information to maintain consistent and accurate routing tables across all nodes of the network. The structure of the network can be either flat or hierarchical. Flat proactive routing strategies have the potential to compute optimal paths. The overhead required to compute these paths may be prohibitive in a dynamically changing environment. Hierarchical routing is better suited to meet the routing demands of large ad hoc networks.

Reactive routing strategies establish routes to a limited set of destinations on demand. These strategies do not typically maintain global information across all nodes of the network. They must therefore, rely on a dynamic route search to establish paths between a source and a destination. This typically involves flooding a route discovery query, with the replies traveling back along the reverse path. The reactive routing strategies vary in the way they control the flooding process to reduce communication overhead and the way in which routes are computed and reestablished when failure occurs.

Hybrid strategies rely on the existence of network structure to achieve stability and scalability in large networks. In these strategies the network is organized into mutually adjacent clusters, which are maintained dynamically as nodes join and leave their assigned clusters. Clustering provides a structure that can be leveraged to limit the scope of the routing algorithm reaction to changes in the network environment. A hybrid routing strategy can be adopted whereby proactive routing is used within a cluster and reactive routing is used across clusters. The main challenge is to reduce the overhead required to maintain the clusters.

In summary, traditional routing algorithms for ad hoc networks tend to exhibit their least desirable behavior under highly dynamic conditions. Routing protocol overhead typically increases dramatically with increased network size and dynamics. A large overhead can easily overwhelm network resources. Furthermore, traditional routing protocols operating in large networks require substantial internodal coordination, and in some cases global flooding, to maintain consistent and accurate information, which is necessary to achieve loop-free routing. The use of

these techniques increases routing protocol overhead and convergence times. Consequently, although they are well adapted to operate in environments where the computation and communications capabilities of the network nodes are relatively high compared to sensor nodes, the efficiency of these techniques conflict with routing requirements in WSNs. New routing strategies are therefore required for sensor networks that are capable of effectively managing the trade-off between optimality and efficiency.

6.5.1 WSN Routing Techniques

The design of routing protocols for WSNs must consider the power and resource limitations of the network nodes, the time-varying quality of the wireless channel, and the possibility for packet loss and delay. To address these design requirements, several routing strategies for WSNs have been proposed. One class of routing protocols adopts a flat network architecture in which all nodes are considered peers. A flat network architecture has several advantages, including minimal overhead to maintain the infrastructure and the potential for the discovery of multiple routes between communicating nodes for fault tolerance.

A second class of routing protocols imposes a structure on the network to achieve energy efficiency, stability, and scalability. In this class of protocols, network nodes are organized in clusters in which a node with higher residual energy, for example, assumes the role of a cluster head. The cluster head is responsible for coordinating activities within the cluster and forwarding information between clusters. Clustering has potential to reduce energy consumption and extend the lifetime of the network.

A third class of routing protocols uses a data-centric approach to disseminate interest within the network. The approach uses attribute-based naming, whereby a source node queries an attribute for the phenomenon rather than an individual sensor node. The interest dissemination is achieved by assigning tasks to sensor nodes and expressing queries to relative to specific attributes. Different strategies can be used to communicate interests to the sensor nodes, including broadcasting, attribute-based multicasting, geo-casting, and anycasting.

A fourth class of routing protocols uses location to address a sensor node. Location-based routing is useful in applications where the position of the node within the geographical coverage of the network is relevant to the query issued by the source node. Such a query may specify a specific area where a phenomenon of interest may occur or the vicinity to a specific point in the network environment.

In the following sections, several routing algorithms that have been proposed for data dissemination in WSNs are described. The design trade-offs and performance of these algorithms are also discussed.

6.5.2 Flooding and Its Variants

Flooding is a common technique frequently used for path discovery and information dissemination in wired and wireless ad hoc networks. The routing strategy is

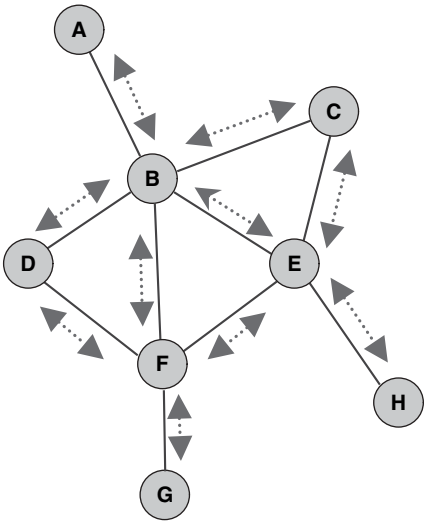

Figure 6.3 Flooding in data communications networks.

simple and does not rely on costly network topology maintenance and complex route discovery algorithms. Flooding uses a reactive approach whereby each node receiving a data or control packet sends the packet to all its neighbors. After transmission, a packet follows all possible paths. Unless the network is disconnected, the packet will eventually reach its destination. Furthermore, as the network topology changes, the packet transmitted follows the new routes. Figure 6.3 illustrates the concept of flooding in data communications network. As shown in the figure, flooding in its simplest form may cause packets to be replicated indefinitely by network nodes.

To prevent a packet from circulating indefinitely in the network, a hop count field is usually included in the packet. Initially, the hop count is set to approximately the diameter of the network. As the packet travels across the network, the hop count is decremented by one for each hop that it traverses. When the hop count reaches zero, the packet is simply discarded. A similar effect can be achieved using a time-to-live field, which records the number of time units that a packet is allowed to live within the network. At the expiration of this time, the packet is no longer forwarded. Flooding can be further enhanced by identifying data packets uniquely, forcing each network node to drop all the packets that it has already forwarded. Such a strategy requires maintaining at least a recent history of the traffic, to keep track of which data packets have already been forwarded.

Despite the simplicity of its forwarding rule and the relatively low-cost maintenance that it requires, flooding suffers several deficiencies when used in WSNs. The first drawback of flooding is its susceptibility to *traffic implosion,* as shown in Figure 6.4. This undesirable effect is caused by duplicate control or data packets being sent repeatedly to the same node. The second drawback of flooding is the

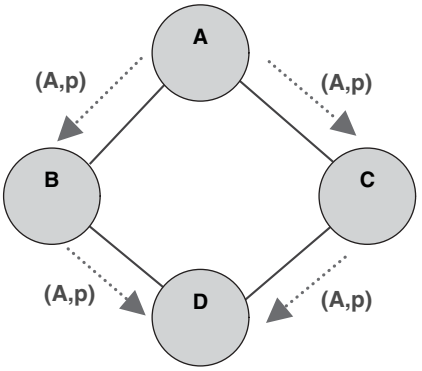

Figure 6.4 Flooding traffic implosion problem.

overlap problem to which it gives rise, as depicted in Figure 6.5. Overlapping occurs when two nodes covering the same region send packets containing similar information to the same node. The third and most severe drawback of flooding is *resource blindness*. The simple forwarding rule that flooding uses to route packets does not take into consideration the energy constraints of the sensor nodes. As such, the node's energy may deplete rapidly, reducing considerably the lifetime of the network.

To address the shortcomings of flooding, a derivative approach, referred to as *gossiping*, has been proposed [6.15,6.16]. Similar to flooding, gossiping uses a simple forwarding rule and does not require costly topology maintenance or complex route discovery algorithms. Contrary to flooding, where a data packet is broadcast to all neighbors, gossiping requires that each node sends the incoming packet to a randomly selected neighbor. Upon receiving the packet, the neighbor selected randomly chooses one of its own neighbors and forwards the packet to the neighbor chosen. This process continues iteratively until the packet reaches its intended

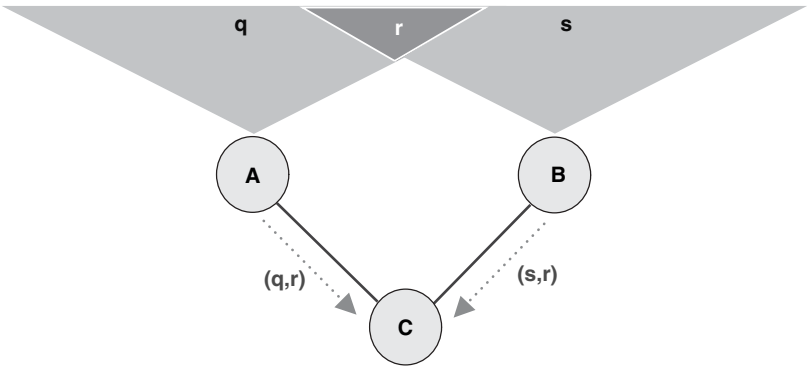

Figure 6.5 Flooding traffic overlapping problem.

destination or the maximum hop count is exceeded. Gossiping avoids the implosion problem by limiting the number of packets that each node sends to its neighbor to one copy. The latency that a packet suffers on its way to the destination may be excessive, particularly in a large network. This is caused primarily by the random nature of the protocol, which, in essence, explores one path at a time.

6.5.3 Sensor Protocols for Information via Negotiation

Sensor protocols for information via negotiation (SPIN) is a data-centric negotiation-based family of information dissemination protocols for WSNs [6.17]. The main objective of these protocols is to efficiently disseminate observations gathered by individual sensor nodes to all the sensor nodes in the network. Simple protocols such as flooding and gossiping are commonly proposed to achieve information dissemination in WSNs. Flooding requires that each node sends a copy of the data packet to all its neighbors until the information reaches all nodes in the network. Gossiping, on the other hand, uses randomization to reduce the number of duplicate packets and requires only that a node receiving a data packet forward it to a randomly selected neighbor.

The simplicity of flooding and gossiping is appealing, as both protocols use simple forwarding rules and do not require topology maintenance. The performance of these algorithms in terms of packet delay and resource utilization, however, quickly deteriorates with the size of the network and the traffic load. This performance drawback is typically caused by traffic implosion and geographical overlapping. Traffic implosion results in multiple copies of the same data being delivered to the same sensor node. Geographical overlapping, on the other hand, causes nodes covering the same geographical area to disseminate, unnecessarily, similar data information items to the network sensor nodes. Simple protocols such as flooding and gossiping do not alter their behavior to adapt communication and computation to the current state of their energy resource. This lack of resource awareness and adaptation may reduce the lifetime of the network considerably, as highly active nodes may rapidly deplete their energy resources.

The main objective of SPIN and its related family members is to address the shortcomings of conventional information dissemination protocols and overcome their performance deficiencies. The basic tenets of this family of protocols are data negotiation and resource adaptation. Semantic-based *data negotiation* requires that nodes running SPIN "learn" about the content of the data before any data are transmitted between network nodes. SPIN exploits data naming, whereby nodes associate *metadata* with data they produce and use these descriptive data to perform negotiations before transmitting the actual data. A receiver that expresses interest in the data content can send a request to obtain the data advertised. This form of negotiation assures that data are sent only to interested nodes, thereby eliminating traffic implosion and reducing significantly the transmission of redundant data throughout the network. Furthermore, the use of meta data descriptors eliminates the possibility of overlap, as nodes can limit their requests to name only the data that they are interested in obtaining.