Wireless Distributed Sensor Networks for In-situ Exploration of Mars[*]

Craig Ulmer & Sudhakar Yalamanchili
Critical Systems Laboratory
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250
E-mail: {ulmer,sudha}@ee.gatech.edu

Leon Alkalai
Center for Integrated Space Microsystems
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
E-mail: Leon.Alkalai@jpl.nasa.gov

1 Introduction

In many scientific applications it is important to obtain precise data measurements over time for a large geographic region of interest. While these measurements can sometimes be performed at a distance using *remote sensing* techniques, it is often necessary to capture data using *in-situ sensing* where sensors are placed directly in the region of interest. The recent images of the Mars terrain captured by the Mars Pathfinder are a graphic example of how in-situ sensing can have a dramatic impact on the way that data is both collected and perceived [1]. Although the Mars Rover was capable of taking multiple measurements within its environment, it was limited to collecting data points at one location at a time.

From the perspective of advancing scientific knowledge the next step is to be able to capture geographically distinct measurements simultaneously for extended periods of time. Such data acquisition presents several challenges notably in ensuring that data points can be correlated in both time and space. The potential solution is the construction of large multi-point sensor networks comprised of nodes capable of measurement, (elementary) processing, and communication. These systems can employ hundreds to thousands of such *sensor nodes* that are interconnected by a flexible communication substrate such as a wireless network. These networks present a significant opportunity for measuring scientific phenomena in-situ with more precision than previously possible. In this role as a massively parallel, accurate, and reliable data acquisition system sensor networks can enable scientific investigations that were previously infeasible due to the lack of data sets necessary to address relevant scientific queries. The goal of this paper is to convey the essential characteristics of sensor networks in general, for scientific data acquisition in particular, and describe the challenges facing their construction and deployment.

NASA has several planned interplanetary missions over the next twenty years to explore environments such as Mars and Europa. Sensor devices are vital to these missions and will be central to a number of scientific experiments. Results of these experiments will help provide planetary information such as climate history, atmospheric content variations, soil toxicity to humans, and the presence of water. These environmental measurements are essential in order for future manned missions to such locations to be safe and realizable. Therefore this paper utilizes the study of planetary environments as a primary motivation for sensor networks.

The network is a central component and source of many technical challenges in sensor networks. The nodes employed in these networks utilize minimal hardware, have limited power, and communicate via low-bandwidth wireless interfaces. These characteristics render traditional schemes for managing such a large network infeasible due to large overheads. The challenge therefore is in the development of communication schemes that are well suited for such fine-grained, inexpensive devices so as to produce a global behavior that is scientifically productive. In this paper we view distributed sensor networks through the lens of internodal communication. Our underlying focus is the need to be able to organize these sensor nodes into a network that reliably captures and relays data for analysis by external systems. We employ a model of sensor networks where nodes are distributed throughout an environment, perform measurements, and then relay the data through neighboring sensor nodes to reach a small number of network exit points.

---

[*] This work is supported by a fellowship from the NASA Jet Propulsion Laboratory's Graduate Student Research Program under grant NGT5-50146.

This paper begins with a brief tour of existing research involving distributed in-situ sensing. This work is then abstracted into a general model for both nodal and network architectures, defining elements that are of particular relevance to sensor networks for planetary exploration. The communication aspects of these architectures are then discussed in greater detail, followed by scientific requirements that must be considered in the design of autonomous sensor networks. An example of the expected mode of operation for an interplanetary exploration mission is presented. Finally remarks about future work in the area of sensor networks conclude this paper.

1.1 Applications of Sensor Networks

An in-situ sensor node is an embedded device placed in an environment to monitor the scientific events of the region. Such devices typically have enough on-board hardware resources to make local data measurements and then transmit the data to an external system for offline processing. By utilizing a number of sensor nodes it is possible to construct a sensor network where a number of geographically distinct samples are collected simultaneously for the region. Nodes in the network collaborate using a flexible communication substrate such as wireless transceivers. Sensor networks can be applied in a number of diverse scenarios including the following:

- **Scatter Probes**: One technique for obtaining distributed samples is simply to scatter a number of low-cost sensors throughout a region of interest and form a data collection network. Scatter probes in such networks must be able to manage the irregular network topologies caused by the sensors’ random physical distribution, as well as handle the presence of faulty components expected of low-power, limited-lifetime sensors. The knowledge that the positions of scatter probe do not change after distribution may be used to optimize network management protocols and reduce power consumption.
- **Sensor Clouds**: The analysis of fluid spaces found in the atmosphere or oceans is significantly improved if sensor probes are injected into the medium to measure physical properties such as the motion of currents. Sensor clouds are similar to scatter probe networks in that sensors are distributed randomly, but are more complex in that sensor positions are constantly changing. The dynamic nature of sensor clouds implies sophisticated sensor nodes that are capable of maintaining accurate position information as well as a dynamic interpretation of the communication network at any given time.
- **Fixed-Point Safety Monitors**: In order to improve internal diagnosis in space vehicles, sensor devices can be placed around the hull of the vehicle to detect abnormalities such as excessive heat or displacement. Observing that wiring inside of a space vehicle can contribute significantly to the overall mass, wireless sensor networks would be essential for improving reliability without negatively affecting the vehicle's mass budget. In a similar situation, wireless communication can be used to overcome routing limitations of wired networks. Often embedded devices have strict device placement constraints that make it difficult to route signals via wires. For example, routing traces in a planar surface may result in a costly multi-layered board for wired communication. If wireless communication is instead used to connect the necessary fixed-point devices, the devices may be placed anywhere within transmission range of each other.
- **Intelligent Exploration Devices**: Future space exploration efforts will use multiple drone devices to study a region. These drones will be sophisticated embedded devices capable of exploring their environment without outside assistance. While such devices would utilize a more sophisticated form of a sensor network, the general principals of reliability and data collection apply.
- **Biological Monitoring Networks**: One future application of sensor networks would be to use nano-sensors to monitor conditions within the human body. Such sensors would move through an organism's natural systems, take measurements, and either transmit or record observations as needed.

1.2 Characteristics of Sensor Networks

From the preceding applications it is evident that the characteristics of sensor networks can be distinguished between those that are common across applications and those that are unique to an application. These characteristics are important to identify because they define the design space for the construction of a sensor network for an application. The following is a general list of such characteristics:

- **Sensor Cost**: The target cost of an individual sensor in a sensor network is a significant design constraint. The cost of an individual sensor is generally proportional to the sophistication of the sensor hardware. Therefore networks with low-cost sensors leave little to be invested in inter-node communication necessitating simplified network construction and management techniques. For a fixed system cost, the cost of an individual sensor effectively determines the number of nodes.
- **Sensor Topology**: The physical arrangement of sensors can play a significant role in how routing is performed in the sensor network. If sensors are distributed in a regular or periodic fashion, routing algorithms can take advantage of the topology to produce optimized solutions. Sensors distributed randomly must incorporate relatively more complex algorithms for irregular topologies that are also resilient to node failures.
- **Sensor Mobility**: Sensor mobility plays a significant role in how communication is handled in sensor networks. If sensor network nodes frequently change position, routing algorithms must be able to recognize changes to the topology and adapt accordingly. Conversely, networks with relatively stable topologies should use routing algorithms that exploit position information to improve communication efficiency.
- **Types of Network Elements**: Availability of additional resources such as routers may increase the overall performance and survivability of a network but come with extra design costs and must be handled with special attention so as to maximize their role in the network.
- **External Signal Availability**: External signals may significantly aid in making more precise data measurements available. Should Global Positioning System (GPS) information be available to a sensor network, then the extraction of position and time information can be greatly simplified. Since position and time are key elements in any data measuring systems, such simplifications can lead to major improvements in performance.
- **Deployment Mechanisms**: The means by which sensor nodes are deployed can have a number of consequences for network communication. If all nodes are deployed at the same time, network configuration occurs only once and at the beginning. However if the network is periodically “re-seeded” with the addition of new nodes then the network must support dynamically reconfiguration.

Scenario	Sensor Cost / Sophistication	Number of Sensors in a Typical Network	Network Topology	Sensor Mobility	Homogeneous
Scatter Probes	Low	100 – 10,000	Irregular	Low	Yes
Sensor Clouds	Medium	5 – 100	Irregular	High	Yes
Safety Monitors	Low	10 – 1,000	Regular	None	Yes
Intelligent Drone	High	3 – 50	Irregular	Medium	No
Biological	Low	10 – 100	Irregular	High	Yes

Table 1 : Comparison of Different Sensor Network Scenarios

The range of characteristics leads to many possible types of sensor networks. Initially sensor networks for in-situ exploration of Mars will focus on using minimal hardware in sensor nodes to accomplish distributed scientific measurements. From the list of scenarios described in section 1.1, the design style that best matches the requirements of the first generation of Mars sensor network is the use of scatter probes. Therefore this paper focuses on a scatter probe style of sensor network as the motivating design style. However it is important to keep in mind that the lessons learned from scatter probe networks are often applicable in several other sensor network scenarios.

2 Related Work

There are a number of research efforts both prior and ongoing that address challenges similar to those described of in-situ sensor networks. These research efforts cover a variety of applications including oceanography, seismology, space exploration, as well as commercial and military based systems. While inter-node communication is a common issue in all of these applications, future in-situ systems warrant special attention.

2.1 Scientific Distributed Sensor Networks

The scientific community has used sensor networks over the last thirty years for a number of applications. Perhaps the best-established work with a large number of sensor nodes comes from oceanographic studies. Oceanographers have been collecting data about the oceans using specially designed buoys. These buoys are typically sophisticated devices that are equipped with environmental sensors, generators to produce power, and transmitter devices to communicate with observation stations via satellites in low earth orbits (LEO) [2,3]. The overall position of a buoy may change depending on whether it is anchored or free floating, and in some cases the buoy may adapt its depth using onboard submersion equipment [4]. While buoy networks capture regional data, they are different from the sensor networks discussed in this paper due to communication requirements. Since commercial satellite links are economically available, there is no need for buoys to communicate with other buoys. Therefore while buoy networks illustrate a means of monitoring an area, there is little buoy literature available to suggest how future sensor networks can intelligently capture and propagate information.

A number of other scientific projects present solutions similar to buoy networks in capturing data. In the Free-Flying Magnetometer experiment [5], a number of hockey-puck sized devices are released far above the Earth's surface to independently capture information about the magnetosphere. NASA's New Millennium Program Deep Space 2 (DS-2) featured two independent penetrating probes that were designed to capture information about Mars and transmit data from the landing sites back to Earth. Similarly the Mars Pathfinder mission used the Sojourner microrover to capture multiple data points throughout a region and then transmitted the data to Earth. Numerous experiments have been performed with aerobots and weather balloons on Earth. All of these scientific efforts to study a region of interest have used either multiple sensor probes or multiple data readings to capture distributed information about a region. However, like the buoy networks most of these efforts relay information directly to observation stations and have not needed to construct inter-node collaborative networks.

An interesting example of a scientific sensor network that resembles the form of future sensor networks can be found in modern underwater acoustic networks [6]. In this work researchers place a number of sensor nodes on the ocean floor to observe environmental features over a long duration of time. Given the infrequency at which sensor nodes can be retrieved, it is important that sensor nodes relay captured data to neighboring nodes to reduce the consequences of a failed node. Information in these networks is propagated at very low data rates (100 bps) and infrequently (less than five messages transmitted per node per hour) due to the difficulty of underwater acoustical modem design [7]. While the underwater communication properties have limited the amount of networking operations that can be implemented in this application, this application captures many of the basic characteristics of future sensor networks: distributed sensing, nearest neighbor communication, and the need for reliability.

2.2 Military Applications

Military research groups have demonstrated an active interest in wireless communication networks for a number of years. One key goal of military efforts is to create a distributed wireless information network to facilitate a digital battlefield [8]. In a digital battlefield military resources such as soldiers or artillery are equipped with wireless communication devices. Resources relay position and status information to military command centers as well as other in-field agents. Building a reliable network for this application is challenging because of the high mobility and likelihood of failure in individual network points. Security is another concern in such networks since it is possible that an enemy can either intercept wireless transmissions or inject its own transmissions into the network to compromise battlefield information. Given the risk that a poorly constructed battlefield network can negate its benefits, researchers have primarily focused on networks with sufficient power and hardware to survive the rigors of combat. While this hardware sophistication exceeds that of general-purpose sensor networks, the lessons of fault tolerance and routability from military network projects are important for future sensor networks.

2.3 Ad Hoc Routing Efforts

Another research area closely related to sensor networks is ad hoc routing. With ad hoc routing a network can be constructed out of a number of mobile computers that are equipped with wireless network interfaces. These networks are self-forming, operate without the assistance of a fixed communication infrastructure, and manage units that drift in and out of transmission range of the network. Work in ad hoc networking has focused on providing routing that can be used by traditional host-level applications. A number of routing algorithms have been proposed as summarized in [9]. One example of these algorithms that is noted for its simplicity and performance is dynamic source routing (DSR) [10]. In DSR a sender employs controlled flooding of the network to discover a path to the destination. All nodes cache routing information locally as space permits and messages are source routed after a suitable route is discovered.

While the general principles of ad hoc routing can be leveraged, fundamental assumptions predicating ad hoc routing render the current generation of protocols unsuitable for sensor networks. First ad hoc protocols often use costly network flooding mechanisms that are inefficient in static networks where there is relatively little mobility. Second ad hoc routing protocols are often designed for a small network of capable hosts. We are studying large networks of nodes with minimal hardware resources. Many ad hoc algorithms do not scale well, nor are the processing demands suitable for the limited capabilities of the anticipated sensor nodes. Finally ad hoc network protocols are designed to provide a seamless interface for networking traditional host level applications. Given the specialized nature of sensor networks, this abstraction of general connectivity may be less important than one that is less flexible, but more attuned to the actual application requirements.

2.4 Commercial Wireless Communication

Wireless communication is the basis for modern communications including cellular phones, pagers, personal digital assistants, and resource tracking agents. Primarily due to reliability concerns most commercial wireless systems are based on a form of fixed network infrastructure. In these systems dedicated base stations or routers are spread throughout a region so that mobile nodes are always in communication range of a fixed network access point. While a network with a fixed infrastructure provides a stable notion of topology and access, it comes at the cost of installing and maintaining physical network access points. While these installations are suitable for wide commercial use such as in cellular phone networks, they are generally unavailable for low-cost embedded sensor networks for space exploration missions.

Recent commercial ventures have demonstrated technologies that are just beginning to operate without a fixed network infrastructure. The Bluetooth standard is a commercial wireless interface that allows embedded devices to communicate with each other [11]. Bluetooth devices will be capable of short-range transmissions (10-100 meters) at moderate data rates (1Mbps). The standard includes provisions for multimedia traffic with Quality of Service (QoS). The first generation of Bluetooth devices will primarily be aimed at *piconet* style networking, where a piconet consists of one master device and a small number of slaves. While this form of piconet does not inherently scale well to the large size of sensor networks, it is possible to see how this technology could be extended to meet the needs of future large-scale sensor networks.

3 Structure and Operation of In-Situ Sensor Networks

Sensor networks discussed in this paper follow the large-scale scatter probe methodology presented in section 1.1. We focus on a sensor network where a large number of low cost sensor nodes are dispersed within a region, with nodes undergoing little mobility during their limited lifetimes. This section presents an overview of the hardware architecture for such nodes and the basic communication requirements for the overall network.

3.1 Node Architecture

Individual sensor nodes for the scatter probe style of operation must be constructed to accomplish specific tasks: scientific measurement, communication with other nodes, simple data processing, and the ability to be reprogrammed to meet the needs of diverse applications. An example of a node’s architecture is illustrated in Figure 1.

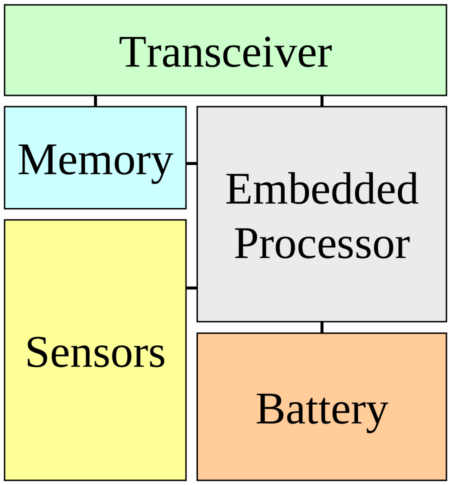

Figure 1: Architecture Features of a Generic Sensor Node

In this view, sensor nodes are equipped with five main features:

- **Embedded Processor**: Nodes typically use a low-performance embedded microprocessor to co-ordinate all sensor node activities. The ability to reprogram the microprocessor's firmware before a mission allows the node to function in a variety of applications.
- **Memory**: A small amount (16KB-1MB) of memory is used for data capture, network queuing, and as storage space for the microprocessor's applications.
- **Sensors**: A probe may contain multiple sensors for scientific measurements, such as temperature, pressure, acceleration, light intensity, and magnetic fields.
- **Wireless Transceiver**: The wireless transceiver is an interface that allows the node to communicate with nearby nodes. While radio frequency (RF) transceivers are common in sensor nodes, recent research projects have examined the use of optical transceivers [14].
- **Battery and Power Management**: Since sensor nodes are self-contained, they must be equipped with their own sources of energy. In addition to batteries, nodes can employ power management hardware as well as solar cells for recharging [15].

Design restrictions such as cost, size, mass, and power can have a significant impact on the architecture and operation of sensor nodes. With current and next generation fabrication technology, cost and power restrictions have the most significant impact on the design of sensor networks.

3.2 Network Architecture

For sensor networks that contain more than a singe node, it is important to study the communication aspects of the overall network architecture. By network architecture we refer to the communication topology of the sensor network, the algorithms for routing data through the network, and the systems software that ensures reliable, efficient, and sustained internodal communication.

Two factors constrain the communication topology of the network architecture: the physical location of sensor nodes and the characteristics of the transmission technology, which in this case is low-bandwidth wireless communication. Wireless transmission is fundamentally a broadcast operation to multiple nearby nodes. The ability to directly transmit information between two nodes is based on the distance between the two nodes and the amount of power applied at the transmitter. Therefore a "traditional" representation of a network can be constructed by defining links between communicating pairs of nodes. Links in this representation carry a weighting factor, corresponding to the amount of energy that must be exerted to accomplish direct communication.

Routing algorithms are a part of the network architecture that facilitates communication between non-neighboring nodes. Such algorithms are multi-hop in nature, using information from the network topology to determine how intermediate nodes can be utilized to relay messages between nodes that cannot communicate directly. The routing algorithms and the network software that implements them must be robust enough to accommodate the random distribution of nodes in the physical environment as well as the likelihood of node failures.

Higher-level functions implemented in the network software can dynamically mange the network connectivity, determining which neighbors are "close" and which are "far", with the latter requiring expensive power expenditures for communication. These factors can be applied to routing decisions dynamically in order to improve the battery lifetime of the system. The broadcast form of wireless communication can be exploited at the system level to realize efficient means of distributing information throughout the network. In general trade-offs in power, performance, and complexity are possible leading to wide ranges of solutions for a sensor network. Being that such tradeoffs can be managed by system software, sensor network hardware can be adapted to fit a diverse number of applications and situations.

Knowledge of the communication patterns of sensor networks can be applied at the system level as an optimization in the network architecture. Many sensor networks primarily function as *data capture networks*, where data is collected from distributed sensor nodes and relayed to specific exit points in the network for uploaded to external systems for processing. Communication in these data capture networks is predictable. As illustrated in Figure 2, sensor nodes capture individual measurements and then channel data through established paths to a base station. A common optimization in this environment is to provide *data fusion* in the network, combining multiple data points into a single message. Transmission power is non-linear in the size of the message data payload. Thus data fusion can reduce power dissipation by reducing the number of transmissions. Research projects such as LEACH [12] have demonstrated that organized approaches to data fusion can lead to significant power savings. As the number of nodes in the network increases, tailoring network operations to fit application behavior can lead to significant gains in efficiency.

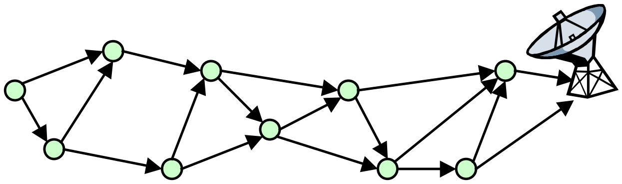

Figure 2: Data Collection Network

3.3 Sensor Network Examples

A number of researchers have built hardware prototypes that illustrate both current sensor network technology as well as the potential of future architectures. Some examples include the following. One of the most longstanding research projects in this area is the Wireless Integrated Network Sensors (WINS) project at UCLA and the Rockwell Science Center. This project uses relatively sophisticated node hardware featuring significant in-network processing and a custom built low-power wireless interface. Seismic sensors were used in this architecture to demonstrate the sensor network's ability to detect military vehicles. The smart dust project [13] at the University of California at Berkeley is another leading effort in future sensor networks. As a first step in building sensor nodes that are light enough to float in the air, researchers have constructed a number of prototype architectures using commercial parts [14]. These devices can be equipped with magnetometer, accelerometer, temperature, and pressure sensors, and use either radio frequency (RF), optical, or infrared (IR) transceivers. Nodes use an 8-bit Atmel microcontroller operating at 8MHz with 4KB of memory.

The sensor web project at NASA's Jet Propulsion Laboratory is another example of how sensor node hardware can be effectively constructed and utilized to capture in-situ data [15, 16]. This project uses commercial hardware components, including a PIC embedded microcontroller, moderate amounts of memory, a solar cell array for battery charging, and a 20 Kbps commercial wireless transceiver. Sensing devices include those for temperature, pressure, light, and trace gasses. Outside of the physical design of sensor nodes, the sensor web project is valuable because it addresses the concerns of distributed, intelligent data collection in the presence of failures. The networking aspect of this scheme uses a form of intelligent flooding to distribute captured data throughout the network. As a result this approach does not scale to large numbers of nodes. However, it preserves the significant advantages of fault-tolerance, simplicity, and the ability to easily incorporate new nodes into the network.

3.4 Commercial vs. Custom Designs

As with any new technology a topic of debate is the use of commercial components vs. custom designs. Embedded processors continue to evolve with increasing features sets. Modern embedded processors include sufficient computation facilities, support for power management, analog-to-digital converters, and memory in a single chip package at a low cost ($10). A number of wireless transceivers such as Bluetooth [11] are also available at a low cost ($5) due to the demand for wireless consumer products. We can expect continued improvements in commercial technologies and economics of scale clearly argue for the use of commercial components.

However when operating at the leading edge, custom hardware has a number of significant advantages primarily due to the fact that it can be designed to accomplish specifically what a sensor network requires with minimal size, mass, and power characteristics. Adaptations of commercial parts may simply be infeasible for meeting physical design constraints. Perhaps most importantly, custom hardware design can include advanced technologies that might not be commercially available, e.g., micro-electro-mechanical systems (MEMs) technology.

Perhaps the most promising solution can be found in the way modern chip design is evolving. Modern digital ASIC work has shifted away from custom implementations and has shifted more towards the use of core libraries. Each year more companies are releasing core architectures for commercial parts that can easily be integrated into a chip design. Other efforts have resulted in CAD tools that automatically synthesize an architecture of cores complete with a data path that matches the needs of the application [17]. Given the automation of the ASIC design process into physical hardware [18], it is not unrealistic to expect that future sensor network architectures can be fabricated rapidly using both well-known commercial cores and necessary custom logic.

4 Communication Properties of Wireless Sensor Networks

In order to design an effective sensor network, it is necessary to consider the properties of each layer of network abstraction. This section outlines the lower four layers of the OSI 7-layer network abstraction [19] as they relate to sensor networks.

4.1 Physical Layer

4.1.1 Transmission Model

At the lowest level of abstraction in the OSI model is the physical layer, which facilitates the physical transmission of data from one node to another. In wireless sensor networks we abstract the wireless communication interface into a broadcast radio model. This model assumes three basic characteristics in order to provide a suitable abstraction for most wireless transmission schemes: limited range, broadcast nature, and half-duplex operation.

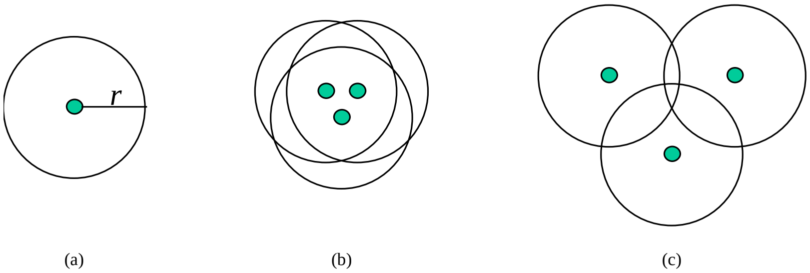

Figure 3: (a) Transmission Range for a Node (b) Three Nodes in Transmission Range (c) Three Nodes out of Transmission Range

The first characteristic of the communication model is that nodes have a limited transmission range as illustrated in Figure 3(a). This range of transmission has a maximum value **r** and is proportional to the amount of energy expended by the transmitter. While the nodes of Figure 3(b) are close enough to observe each other's transmissions, the nodes of Figure 3(c) are out of range and cannot directly communicate. A second property of the model is that communication is broadcast in nature for all nodes that are within transmission range. If a single node in Figure 3(b) transmits a message, the two remaining nodes can potentially receive the message. However, if two of the nodes in Figure 3(b) simultaneously transmit, the third node perceives the transmissions as a corrupted message due to the broadcast nature of the medium. The third characteristic of this wireless communication model is that nodes cannot simultaneously receive and transmit messages. Since the signal power of the transmitter is significantly greater than that of incoming transmissions, it is difficult to reliably isolate the incoming signal while transmitting. This trait has important implications in protocol design since it makes collision detection at the sending node unlikely.

4.1.2 Physical Layer Radio Considerations

There are a number of design issues that are involved in the physical layer of a sensor network. While the majority of common physical layer issues (spread spectrum, code assignment, etc.) are beyond the scope of this paper, there are a few considerations that are particularly relevant to sensor networks for in-situ science. Selection of a basic transmission frequency for the transceivers has a number of effects:

- **Environment**: Radio performance varies on a number of environmental factors, so a transceiver should be matched to its target environment. Surface-to-surface communication on Mars has been suggested to be within a 100-450MHz frequency range [20].
- **Antenna Length**: Antenna length is inversely proportional to transmission frequency. Under sensor node size constraints, transmission frequency should be as high as possible to minimize antenna length.
- **Power Consumption**: Power consumption roughly increases with transmission frequency. This constraint suggests that using lower frequency transmissions may extend battery lifetime [21,22].
- **Fabrication**: Radios utilizing higher frequencies require specialized substrates such as Gallium Arsenide, and are more susceptible to electromagnetic noise. Therefore if a single chip design is preferred to reduce manufacturing costs, lower frequency radios allow simpler designs.

4.1.3 Physical Layer Power Consumption

Power consumption is an important issue to address at the physical layer. Transmission is a large source of power consumption in a sensor node since the transceiver must generate an appropriate analog signal with enough amplification to reach a destination. Minimizing the number and duration of transmissions therefore is a basic form of power savings. The receiving hardware represents a greater challenge for power consumption reduction. While receiving hardware consumes only a fraction of the power required by transmitters, receivers must operate continuously since transmissions can arrive at any time. A common technique employed at the physical level is to move the node into a sleep mode where the receiver hardware is powered down for a period of time. Sleep modes can be initiated for a number of reasons, including scheduled shut down times, the detection of a busy transmission channel, or a general lack of power in a node. Power savings techniques have been demonstrated in commercial wireless networks [22].

4.2 Data Link Layer

Given the potential for transmission errors in the physical layer, it is necessary to address reliability at the data link layer. This layer uses medium access control (MAC) protocols to provide access to the shared transmission medium and logical link control (LLC) protocols for reliable transmission of messages.

4.2.1 Medium Access Control: Carrier Sensing and Control Messages

MAC protocols allow multiple devices to share the same transmission medium in a way that reduces contention. Two techniques are commonly employed in MAC protocols: carrier sensing and the use of control messages to schedule access to the broadcast medium. In carrier sensing a node with a message to send will stall transmission until it detects that no other transmissions are taking place in its neighborhood. This technique by itself can be used to provide on-demand scheduling for a transmission channel and is known as carrier sensing multiple access (CSMA). CSMA is a basic trait of Ethernet and is a simple but effective method that operates on the probability that two nodes within range of each other will not have a message to send at the same time.

Another technique employed by MAC protocols is the use of control messages to synchronize access to the transmission channel. With this technique nodes use a dialog of control messages to actively alert neighboring nodes of their use of the broadcast medium. Figure 4 illustrates an example dialog that takes place during the reliable transmission of a message between two nodes so as to reduce the probability of a collision. In this dialog a node with a message to send first broadcasts a short request to send (RTS) message containing its ID as well as the receiver ID. The receiver node accepts the message and transmits a clear to send (CTS) message when it is capable of accepting the message. Upon receiving the CTS message, the sending node is free to transmit its data message to the destination. Acknowledgement of the reliable receipt of the data message is answered either positively (ACK) or negatively (NACK) by the receiver.

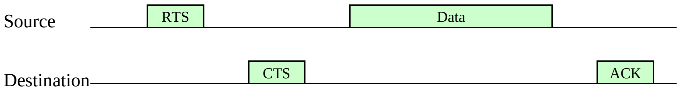

Figure 4: MAC Use of Control Messages for Reliable Delivery

4.2.2 Medium Access Control: Obstacles

Due to the properties of ranged transmissions, there are obstacles that can result in transmission errors. An important issue in radio-based networks is the hidden terminal problem [23]. The hidden terminal problem is the case where a collision in the network occurs because two nearby but out of range nodes are unaware of each other’s actions. An example of this problem is illustrated in Figure 5 where three nodes are at the fringes of each other’s communication ranges. In this arrangement node B can observe transmissions from nodes A and C, but nodes A and C cannot observe each other’s transmissions. Therefore if node A is transmitting to node B, node C will fail to sense the transmission and can begin transmitting, resulting in a transmission collision. While control messages can be used to reduce the consequences of the hidden terminal problem, it is challenging to fully protect against these forms of collisions.

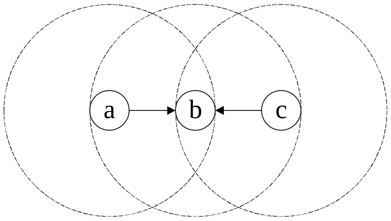

Figure 5: The Hidden Terminal Problem

Another issue in MAC protocols for shared mediums is the exposed node problem. In the exposed node problem, a node unnecessarily delays transmission because it falsely senses that the transmission will interfere with another node’s transmission. An example of this situation can be found in Figure 6, where node B is transmitting a message to A and node C needs to transmit a message to node D. In this case node C senses that the broadcast medium is already in use and therefore delays its message transmission to prevent a collision. However, since node A is outside of the transmission range of node C, node C’s transmission will not interfere with the reception at node A. As this example demonstrates, the carrier sensing mechanisms of the physical layer can be overprotective and lead to unnecessary transmission delays. However, while the exposed node problem affects efficiency, it is not as critical as the hidden terminal problem since it does not directly affect reliability.

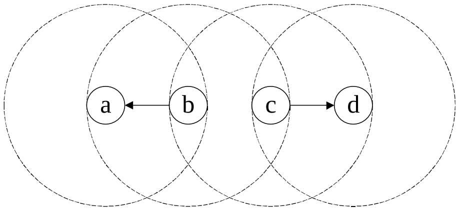

Figure 6: Exposed Terminal Problem

4.2.3 Medium Access Control Protocol Examples

A number of MAC protocols have been designated to deal with the properties of wireless networks. The following list briefly describes a few approaches:

- **CSMA/CA**: This approach adds an RTS/CTS/DATA dialog to CSMA to provide collision avoidance.
- **MACA [24]**: Carrier sensing is removed from CSMA/CA to address the exposed terminal problem. Nodes observing a CTS for another node back off the channel since the receiving node is within transmission range.
- **MACAW [25]**: MACA-Wireless adds a short data-send (DS) message before the transmission of a DATA message and follows delivery with an ACK. These messages help clarify transmissions in the presence of a lossy transmission medium.
- **MACA-BI [26]:** In MACA-By-Invitation, the transmission scheme is reversed. Receivers in the network poll neighbors to determine if they have a message to send. While this operation reduces the number of control messages required to facilitate a transfer, receivers must continuously transmit invitation messages that are wasteful in terms of power.
- **Bluetooth [11]:** The Bluetooth standard employs a MAC scheme similar to MACA-BI in that the master of a piconet polls neighboring nodes for data. This scheme is however more elaborate in that it includes scheduling mechanisms for the transmission of QoS style data streams.
- **Busy-Tone [23]**: This scheme employs two transmission channels, one for data and another as a busy signal. When a node is receiving data on the data channel it broadcasts a signal on the busy channel. If a node observes activity on the busy channel, it knows that it should not transmit because it would interfere with a node receiving data that is in range. While effective this technique uses multiple transmission channels and causes both the sender and receiver to expend transmitter levels of power during a transmission.
- **PAMAS [27]**: The Power Aware Multi-Access protocol with Signaling extends MACA and includes an additional signaling channel. The protocol is constructed so that nodes can move to a powered down state if the transmission channel is in use.

4.2.4 Logical Link Control Protocols

Logical link control (LLC) protocols are built on top of the MAC and serve as a means of guaranteeing that data messages are reliably transmitted from a node to its intended neighbor. While damaged transmissions can be repaired to an extent by using forward error correction (FEC), it is desirable to employ automatic repeat request (ARQ) protocols in the LLC to provide link level reliability. In ARQ protocols a sending node will retransmit a message if it receives either a NACK from a destination or the time period to hear a response has expired. In order to improve performance, ARQ approaches such as *go-back-n* and *selective repeat* relax the tight coupling between sender and receiver and allow multiple acknowledgements to be collapsed into a single message. Unfortunately these optimizations are generally inappropriate for sensor networks due to the network characteristics. Since there is a higher probability that messages will be corrupted during transmission in wireless networks, it is important for the LLC to recognize errors as soon as possible. Additionally, the limited buffer space of the nodes renders it infeasible for the LLC to manage a large number of outstanding messages. Finally, wireless MAC protocols often include built-in forms of acknowledgements that should be used by the LLC whenever possible.

4.3 Network Layer

The limited transmission range for a node's transceiver is an important variable in determining how communication in the network takes place. In the simplest scenario, all nodes are within transmission range of each other, allowing any node to directly communicate with any other node simply by broadcasting in the transmission channel. Unfortunately, practical distributions of sensor nodes lead to the high probability that a given node will only have a small number of neighbors that are within its transmission range. Therefore it is necessary to employ multi-hop routing to allow non-neighboring nodes to communicate. As illustrated in Figure 7, multi-hop routing is the process of using intermediate nodes to relay a message from a source to a destination. By forwarding messages through sets of neighboring nodes, multi-hop routing extends the amount of area that the sensor network can cover while still remaining connected. The main criticism of multi-hop routing is that it adds to system complexity and forces intermediate nodes to do the work of other nodes.

Figure 7: Establishing a Multi-hop Route Between Non-Neighbors

Providing routing for a multi-hop network can be a nontrivial task. Routing schemes must address the specific characteristics of the network, such as mobility, number of nodes, regular communication patterns, and availability of node resources for network management. The simplest form of routing is flooding, where nodes rebroadcast messages received by other nodes until information is dispersed through the entire network. At the most complex end of routing, all nodes maintain complete connectivity lists for the entire network and make routing decisions based on this global knowledge. While both of these routing extremes perform well for a small number of nodes, they fail to scale to networks with a large number of nodes.

Effective routing schemes for large-scale sensor networks must be able to direct traffic efficiently without the benefit of large memory resources. A number of traditional routing schemes are applicable. Table based routing is useful since nodes simply maintain a list of destinations and the next hop to reach the destination. Should nodes primarily need only a few routing destinations (e.g., the exit nodes of the network where data is uploaded to outside computers), distributing routing information is a relatively straightforward operation requiring little overhead and the topic of interest is efficiency given the limited resources of sensor nodes. For example, one approach is to improve flooding algorithms by basing routing decisions partly on the data collected by sensors. Such approaches seek to route the network in a loop-free, fault-tolerant fashion that matches the application requirements of the sensor network.

4.4 Transport Layer

The transport layer in communication systems is a mechanism for monitoring the network as a whole and providing reliable end-to-end message delivery. In traditional networks transport is commonly handled with the transmission control protocol (TCP). TCP essentially fragments a large stream of data into manageable packets, injects the packets into the network at a controlled rate, and then uses control messages to determine which packets should be retransmitted. TCP contains a number of subtle features that make it valuable for traditional networks. A sending node in TCP monitors the rate of acknowledgement messages received from the destination in order to make assumptions about the state of the network. Should acknowledgements be delayed or missing, the protocol assumes network saturation and decreases the injection rate.

While the transport layer is useful in traditional networks, there are a number of factors that suggest that transport protocols are not as valuable to sensor networks. First, sensor networks primarily transport measured data values to known locations in the network. While network protocols should be written to deliver as many data values as possible, it is expected that the network is lossy. Therefore the efforts of end-to-end reliability in transport layer can be wasteful. Second, nodes have a limited amount of memory and processing power to implement transport protocols. Protocols such as TCP manage large caches of state and data. Finally, a fundamental level of reliability is built into the link level of the network. While this reliability does not guarantee that a message will ultimately reach its destination, routing algorithms can be developed to contribute to improved reliability.

Outside of end-to-end reliability, transport layers provide means of monitoring the dynamic state of the network. While a node can gain a fair amount of local information from its neighbors' link-level activity, it is desirable to discover more distant but relevant information. For example, nodes in a sensor network that are close to the ejection points are more likely to observe larger amounts of traffic than distant nodes. Should the nodes close to the ejection point begin to saturate with traffic, feedback information should be dispersed throughout the network so that far away nodes back off from injecting new messages. Feedback information can be adequately spread through the network using broadcast flooding [28].

5 Scientific Aspects of Wireless Sensor Networks

The main goal of the construction and deployment of sensor networks is the enabling of "good science". Therefore a key consideration for architects of future sensor networks is the manner in which the network can support science. In the ideal case a sensor network captures an infinite number of accurate data points in both time and space. Each data point would be labeled with both the time and location of the measurement in order to allow a meaningful interpretation of the data. However, realizable sensor networks contain a finite number of nodes, operate over a limited lifetime, and can suffer from inaccurate sensor measurements. This section discusses techniques that can be applied in sensor networks to improve the quality of science.

5.1 Global Positioning System

One of the most significant technological advances for terrestrial based science is the Global Positioning System (GPS) [29]. By using signals generated by a network of satellites orbiting the Earth, a GPS receiver can accurately discover its location to a meter of resolution [30]. Additionally information in the GPS signal can be used to recover a global clock for temporal synchronization [29]. While initial GPS receivers were large and expensive, the commercial demand for this technology has led to chipsets that are both small (the size of a credit card [30]) and low cost (less than $100). GPS chipsets can easily be adapted to fit a sensor probe's power budget by only periodically enabling the hardware to synchronize the node.

While GPS is a viable technology, there are a number of reasons to consider alternative time and position synchronization methods in sensor networks. Designers must weight the use of GPS against its effects on a node's power, size, and cost budgets. Considering that the hardware is used infrequently, an alternative is to supply only a percentage of the nodes with GPS and use distributed algorithms to synchronize the remaining nodes. Another significant issue is the physical feasibility of GPS. For example, deep sea and non-terrestrial sensor networks cannot make use of GPS since reception is infeasible. Therefore it is beneficial to consider alternative options for temporal and positional identification in sensor networks.

5.2 Temporal Synchronization

Temporal synchronization is a subject studied for a number of applications in distributed computing. The challenge addressed is given a number of discrete, distributed devices, how can all devices accurately observe the same global clock? This question appears in a number of contexts, including clock distribution in chip layout, time of day synchronization among computers on the Internet, and synchronization among distributed telescopes in radio astronomy. A simple mechanism that a number of distributed systems use to provide temporal synchronization is to rely on globally observable events. For example, radio telescopes can be accurately synchronized to pulsars [31]. Likewise sensor networks can be synchronized to a beacon signal generated by a lander or orbiter.

Another technique for clock synchronization is to use completely in-network messaging to distribute global clock information. Much like the clock distribution networks of modern microprocessors, a central timekeeping node periodically broadcasts a timing message that is propagated through the network. At each hop away from the timekeeper, nodes make slight adjustments to the timing message to compensate for the latency of transmission over a distance. Clock messages essentially flood the network periodically and can be efficiently propagated by broadcasting techniques. A more sophisticated version of this technique is to create structured clock distribution hierarchies. Clock messages are transmitted down the hierarchy to leaf nodes, then back up to the master timekeeper. By timing a round sweep through the network, the master timekeeper obtains a maximum limit on the amount of time required to synchronize all nodes.

5.3 Position Estimation

Determining the location of all nodes in a sensor network is an important task that potentially must be accomplished with minimal or no GPS support. By using triangulation techniques first used by ancient Greeks, it is possible to determine a topology for the network by measuring the distances between nodes. In this system the distances between three nodes are used to determine the angles of the triangle formed by their positions. By applying triangulation to a number of node triplets, it is possible to estimate the relative positions of nodes within the overall network. If GPS is available in at least two well-triangulated nodes, the relative positions found by triangulation can further be translated into absolute co-ordinates. While absolute co-ordinates are desirable, it is clear that a large percentage of applications can accomplish scientific goals of distributed measurement with relative position information alone.

A key element in accomplishing triangulation is accurately measuring the distances between nodes. This process is known as ranging and can be accomplished through both digital and analog techniques. Ranging techniques typically operate by measuring the amount of time or effort required for a transmitted signal to reach a destination node. The main challenge in performing ranging is accuracy. Transmission signals propagate at roughly the speed of light (approximately a meter in 3 nanoseconds) and are subject to environmental interference. Therefore latency based ranging techniques must use high-speed clocks (1-10 nanoseconds of resolution) to measure small distances (less than 10 meters). Transmission effort based systems must likewise take into account obstructions and reflections.

5.3.1 Distance Estimation through Time of Flight

One method of performing latency based digital ranging is to measure the round-trip time (RTT) required for a message to travel to and from a destination. As illustrated in Figure 8, node A timestamps and transmits a message to node B. Node B receives the message and transmits a reply that contains both the current time and the time at which node A's message was observed. The transaction is completed when node A records the time at which node B's reply is observed. The actual in-flight time can be recovered from the four timestamps of this procedure by determining the overall round-trip time and removing node B's handling delay. This value can be related to distance by multiplying the one-way in-flight time by the propagation speed of the signal. The precision of the operation may be improved by performing the timing operation over N repetitions of transmissions, with the delay between receiving and transmitting the timing message carefully noted and removed in each iteration.

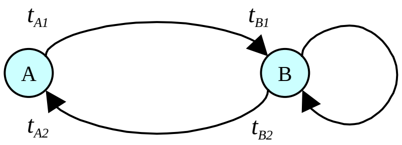

Figure 8: Round-Trip Time Measurement

The total round-trip in-flight time (RTIFT) of the message can therefore be summarized as:

$$RTIFT = T_{A2} – (T_{B2}-T_{B1}) – T_{A1}$$

The RTT method has both advantages and disadvantages compared to other ranging methods. One of the most beneficial aspects is that it is a digital technique that can be accomplished without additional analog hardware. The operation also uses local clocks as opposed to global clocks, allowing two nodes to perform the operation without being synchronized to a global clock. The round-trip nature of the measurement potentially allows for increased precision, since the in-flight time is double that of a one-way transmission. The primary challenge in this form of ranging is dealing with high-speed timers. A variance in clock frequencies between the two nodes results in timing imprecision. Additionally since power consumption increases with clock speed, it is important that these timers be used infrequently and within the node's power budget. Finally, it should be noted that the RTT method expends power at both nodes in the measurement but provides data only to one. Therefore the algorithm should allow for either an odd number of timing transmissions, or force the initiating node to share its observations.

5.3.2 Distance Estimatation using Power Scaling

Another method of performing digital ranging is through power scaling. In a free-space radio transmission between two nodes separated by distance r, the signal strength observed at the receiver is proportional to $(1/r^2)$ for small r and $(1/r^4)$ for larger r [32]. Therefore a node can crudely estimate its distance from another node by discovering the minimal amount of power required for reliable communication with the neighbor. Finding the amount of power necessary to communicate with a neighbor can be accomplished by transmitting a series of advertisement messages with increasing signal strength. A neighboring node completes the discovery process by transmitting an acknowledgement that indicates which advertisement was the first to be received reliably. This process is illustrated in Figure 9, where node A broadcasts two advertisements of different signal strengths. Node B would reply that it observed both transmissions (implying that it is close) while nodes C, D, and E would reply only to the second message (implying a greater distance than B).

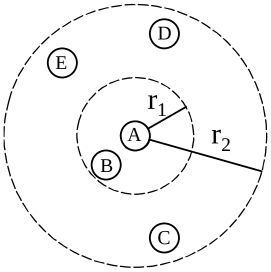

Figure 9: Power Scaling Discovery

Power scaling has both advantages and disadvantages for ranging. The technique requires only the ability to scale transmission power and can be used to find the distance to several nodes at a time. The primary criticism however is that the distances measured by this technique can be inaccurate due to environmental characteristics. A basic assumption of this method is that signals radiate uniformly from a transmitting node. Unfortunately the design and orientation of the antenna can result in non-uniform signal dispersions. Additionally obstructions in the physical environment inhibit transmissions. The net result is that two nodes that are equidistant from a node performing the ranging operation might require different levels of signal strength for reliable communication. While this effect results in distance estimation errors, power scaling can be valuable to other sensor network operations such as routing. Power scaling methods provide an estimate of the amount of *effort* required in transmitting information from one node to another. Therefore network algorithms can use this information to route messages in a power aware fashion.

5.3.3 Distance Estimation using Analog Techniques

In addition to digital ranging techniques, there are analog methods for estimating the distance between two communicating devices. Analog ranging has been used in space applications dating back to the Voyager era, where a satellite's distance from the Earth was estimated by the signal characteristics of its transmission messages. In this form of analog ranging the satellite broadcasts a transmission pulse containing a series of frequencies at known intervals. At the receiving end the pulse is captured and analyzed to observe the shift in phase of the known frequencies. The differences in phase are then used to pinpoint the distance the signal has traveled. The properties of analog ranging techniques make it difficult to apply such algorithms in low-cost sensor probe networks. The primary problem with analog techniques is that there is a certain amount of the associated signal processing that must take place in order to obtain a distance estimate. Using analog circuits to facilitate the processing leads to mixed signal design, which may exceed the scope of a low-cost sensor node. Similarly, sampling the analog signals and performing the computation in the digital domain leads to a large amount of silicon and power that must be devoted to this functionality. Therefore, while analog ranging techniques are useful for certain distance estimation applications, they are generally not appropriate for low-cost sensor networks.

5.4 Regional Sampling

An important consideration in terms of the scientific value of sensor networks is the effect that sensor distribution has on data analysis. From a data processing perspective an ideal sensor network has an evenly spaced distribution of nodes, as depicted in Figure 10(a). This grid-like distribution divides the region of study into discrete, equal-sized territories, with each territory monitored by a node. This ideal distribution is beneficial because it results in a linear sampling of the region, with all territories evenly represented. Unfortunately most sensor distributions are not regular, such as in the random distribution of Figure 10(b). The location of nodes in these networks leads to both non-linear sampling as well as the potential for over- and under-sampled regions. In order to improve the science that such networks are capable of, data analysis programs must relate the randomly placed sample points to a form that resembles the linear ideal sensor network.

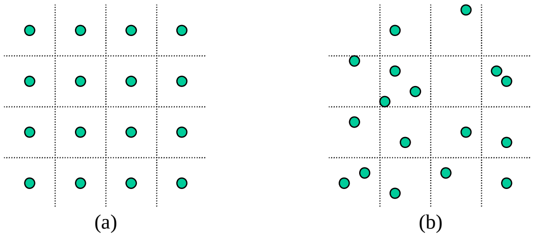

Figure 10: (a) Regular Distribution (b) Random Distribution

The concepts and requirements of regional sampling can be applied within sensor networks to improve the quality and performance of data collection. The overall goal of this optimization is to reduce the number of data points collected by the network while still meeting minimal science requirements. Reducing the number of collected data points can be performed simply by removing redundancies in-network. For example, a grid territory that contains multiple sensor nodes can opt to temporarily shut down a number of sensor nodes to reduce the amount of information captured for the region. In a similar manner, the redundant information for the region can be gathered and averaged into a single data point to provide a better quality measurement. A more challenging aspect is tailoring data collection to address under-sampled regions. For a grid territory without any sensor nodes, the network should require neighboring territories to supply redundant sensor data if available so that analysis programs could better interpolate data for the missing region.

5.5 Temporal Sampling

A second element of sampling is the manner in which data points are collected over time. Assuming that nodes can be synchronized as discussed in section 5.2, the challenge of temporal sampling is to devise a data collection scheme that matches the characteristics of the observed phenomenon over time. One approach to temporal sampling is to continuously capture data points at fixed time intervals. This method results in periodic snapshots of the distributed phenomenon that are useful in general-purpose data analysis. The key to implementing this form of sampling is selecting the time interval between snapshots. Scientifically it is desirable to have as small of a time interval as possible to increase the amount of captured information. However, choosing a time interval that is too small results in network congestion and dropped data since sensor nodes are injecting more data traffic than the network can handle. Additionally high sampling rates are challenging to implement within a limited power budget. While data transmissions can be reduced (e.g., data fusion, compression, or transmitting only significant changes), the sensor network is still in essence a real-time system with deadlines and limited communication capacity.

Another strategy for capturing data over time is an event driven or bursty model. In this form the sensor network is dormant until a scientific event is observed. As soon as a triggering event occurs all sensor nodes come online and capture as much information as possible. Since this form of sampling is used for short bursts of activity, relaying data points through the network for collection can be performed after the event occurrence without having to meet real-time deadlines. Nodes in these networks must have some notion of the significance of sensed data in order to trigger data capture in neighboring nodes. The operational mechanisms of this style of data collection may therefore resemble or benefit from established research fields such as neural networks.

6 Operation of an Example Sensor Network

To bring all of the issues discussed in the preceding sections into focus this section provides a description of the operational scenario of a distributed sensor network on Mars. In this example a large number of mini-science stations are deployed in a region of study. These nodes must work without outside intervention and collaborate among themselves for an extended period of time. Nodes have low mobility and must rely on neighboring nodes to forward messages to base stations capable of transmitting data back to Earth. Due to the characteristics of the network, it is assumed that an in-depth initialization process will be more beneficial than employing dynamic on-demand network management techniques. Therefore initialization is presented as four distinct phases: deployment, activation, local organization, and global organization.

6.1 Deployment

Sensors can be deployed in a number of ways. The most precise means of deployment is to hand place sensor nodes. Given the recent advances in remote exploration vehicles, hand placement is more of a viable means of deployment than would appear at first glance. For example, a rover could be equipped to automatically release a sensor node every time it travels a specific distance. This form of deployment is useful because it can be added to other missions with minimal effort and can at the same time provide well-placed distributions of sensor nodes. The primary disadvantage of this approach is that it is difficult to cover a two-dimensional region with sensors without multiple rover passes. Base stations for uploading sensor data in this style of distribution are assumed to be either the rovers or nearby landers.

A more general form of deployment is for sensor nodes to be launched from a self-contained deployment vehicle. This vehicle could effectively scatter sensor nodes during its descent, as demonstrated in the Free-Flying Magnetometer (FFM) project [5]. While sensor distributions are likely to be random, this form of deployment is beneficial because it deposits sensors across a region with little effort. A downside to this technique is that sensor nodes must be designed to handle the shock of impact. To complete the system, it is expected that the deployment vehicle will also serve as the base station for uploading captured data.

6.2 Activation

In order to reduce power consumption and RF interference during transit, sensor nodes reside in a sleep state until they are deployed. Therefore it is necessary for sensors to undergo an activation phase after they are scattered in the region of interest. Activation may be either implicit or explicit. In implicit activation sensors use environmental information to activate. For example, a sensor node with solar cells could be activated when its solar cells have detected sunlight and recharged the node’s battery. Other environmental keys may also be used to activate the node, including temperature, air pressure, or the shock of impact. The challenge of implicit techniques is that sensors may wake up prematurely or not at all due to false readings or the absence of required stimuli. In contrast explicit techniques rely on generated events to trigger activation. For example a deployment vehicle could be designed to broadcast an RF beacon signal that explicitly directs nodes to activate. The challenge in designing such a system is doing so in a power efficient manner, since nodes consume power simply listening for messages.

After activation a node must perform a number of duties before attempting to become part of the sensor network. First a node should perform a self-diagnosis to determine if it is healthy enough to participate in the network. It is important that unfit nodes be disabled from network operations as early as possible since faulty behavior at a node may affect the operation of neighboring nodes. After a node establishes that it is healthy it should check to make sure that it was not activated by mistake. Sensor nodes may be falsely awakened during transit. Simple communication with the deployment vehicle can help determine if the node is still in transit or not. Finally, an activated node should make preliminary observations about its surroundings. These measurements can include how much sunlight the solar cells are capable of receiving, the amount of general RF noise at the receiver, and whether or not the node experiences any extreme temperature variances. This information can be used in the remaining phases of initialization to estimate which nodes in the network are the most reliable.

6.3 Local Organization

The next phase in a sensor node's initialization is to determine information about the node's local communication environment. The first step in this operation is neighbor discovery, where a node advertises its presence and discovers nearby nodes. Neighbor discovery must be robust enough to address a number of implementation issues. First, sensor nodes are likely to awaken at different times. Therefore discovery protocols must address the fact that some nodes will start the discovery process later than others. Second, since there is no organization in the network and messages are broadcast, it is likely that some advertisement messages will have collisions and be lost. Finally, the discovery process is under a fixed time and power budget. Therefore protocols must reach a conclusion within a finite amount of time, and must discover neighbors with as few transmissions as possible.

After nodes become aware of their neighbors, network organization can proceed at the local level. Nodes that reside within a common region can group together into a cluster in order to define how the nodes will operate collectively. Nodes in a cluster can evaluate the resources available to the cluster and establish local rules and schedules to manage regional tasks. For example, a cluster with a large number of members may decide to power off certain members for a period of time since there are more nodes available than are needed. Clusters also provide a means of handling regional issues such as faults without having to disturb the entire communication network.

Constructing a cluster is a problem for which many solutions exist [33, 34, 35]. Selecting an appropriate clustering algorithm for a sensor network involves examining the characteristics of the network. The algorithm must be available in a distributed form and work under a variety of topological conditions. It must also reach a solution in bounded time using hardware with low processing and communication capabilities. Empirically we have observed that it is desirable to have clustering algorithms that construct a number of non-overlapping clusters with similar geographic distributions and are three to six network hops in diameter.

6.4 Global Organization

The final phase of initialization for sensor networks is to establish the global communication network. For clustered networks this implies that communication between cluster edges is established and global cluster knowledge is shared. The most basic requirement in constructing the global communication network is providing routing between sensor nodes and the node(s) capable of uploading data to external systems. This form of routing can be accomplished by having clusters find the shortest points to ejection nodes. Based on the required functionality of the sensor network, it may be necessary to provide additional routing in the network. For example, some of the advanced sampling techniques discussed in section 5.5 require the ability for nodes to communicate with nodes in neighboring clusters.

After the network is established, initialization is complete and sensor network applications can begin operation. The lower level network protocols used in initialization therefore move into a maintenance mode and are responsible for monitoring the health of the network. As faults are detected in the network, these protocols attempt to re-route the network as best as possible to provide continual network service. As with all communication aspects of sensor networks, fault management software should be optimized to meet the scientific requirements of the overall network.

7 Future Work

Researchers have encapsulated a significant amount of functionality into current day sensor networks. As the research evolves, concurrently sensor networks will continue to grow in size, node processing power, and as a result, scientific value. This section identifies some important research issues that will be encountered as sensor networks become increasingly large.

7.1 Network Deadlock

A fundamental topic in communication networks that is often overlooked in sensor networks is deadlock. Deadlock in networks occurs when two or more nodes have a cyclic dependency that prevents the nodes from making forward progress in delivering messages. For store-and-forwards networks such as the multi-hop sensor network, deadlock can result from the use of finite buffer space and unrestricted routing. For example, consider the case where two neighboring nodes each have n message buffers. Consider the case where both nodes have each buffered n messages that must be routed through the neighboring node. Neither node can make progress because in order for a node to accept a new message, it would have to transmit a message. This cyclic dependency prevents the nodes from making forward progress.

A number of techniques have been proposed to prevent [36] or recover from [37] network deadlock. A simple recovery mechanism is to allow the network to drop packets. While this technique is simple to implement and breaks cyclic dependencies, it results in unreliable network operation and lower performance. For example, it can be argued that every data value being transmitted through the network is significant since a number of nodes have already consumed power in capturing and routing the data value. Discarding of data nullifies this investment in data capture and reduces the overall bandwidth/watt that can be achieved. Alternatively deadlock prevention techniques, which employ routing restrictions, may be more useful in the context of sensor networks. For example, routing restrictions that result in tree structured flow of data avoids or prevents cyclic dependencies and thereby deadlock. Since data collection networks can be easily structured as spanning trees, it is possible to apply this technique without significant implementation penalties.

A primary question for research in deadlock-free routing in sensor networks is whether deadlock is in fact an issue in realistic implementations. Deadlock has not been an issue in sensor network research because the networks have not required wide scale multi-hop routing. Nodes have sufficient buffer space and networking protocols permit the dropping of data points. However, as the responsibilities of future sensor networks increase so does the potential for deadlock. Nodes are self contained and embedded systems, which often cannot be manually reset. Combined with a network where traffic flows are unrestricted increases the probability of deadlock. While the probability of occurrence of deadlock has been empirically and analytically studied in the context of regular networks and wormhole switched routers [38], we are unaware of any similarly body of knowledge for large-scale sensor networks. Given the consequences of deadlock, at the very least a better understanding of the chances of occurrence as well as the availability and cost of the solutions is warranted.

7.2 Routing

With respect to routing protocols we find three primary areas in which current and future sensor large scale networks must focus: power awareness, fault-tolerance, and science facilitation. Given the node designs, and the cost/performance constraints discussed earlier in the paper it is clear that power aware routing is critical in optimizing overall power dissipation. Metrics themselves need to be defined. For example, is it important to optimize power dissipated per message or power dissipated per node? While transceivers are continuously being improved, this does not change the amount of work that is performed in the network in terms of the number of messages being transmitted and the number of hops necessary to perform the transmissions. The major saving in power consumption for sensor networks is in intelligent management of network transactions rather than in optimizations in the physical transceiver components. A particularly useful aspect of power aware routing is graceful handling of decaying networks. Routing in these networks is performed so as to minimize communication hotspots within the network and distribute power-consuming workloads fairly and evenly among all sensor nodes.

As networks become larger reliability is another key topic. While a large amount of fault-tolerant routing literature is available, a common assumption is that designers have control over the network topology. Given that sensor networks are likely to have irregular topologies that can change over time, network designers must consider routing algorithms for arbitrary, and time-varying topologies. In solving this problem, under the minimal power consumption requirements, the network architect can take advantage of properties such as the broadcast nature of transmissions. Based on some probability of reception by a neighboring node, a form of controlled flooding of the network can be employed that must be coupled with deadlock avoidance/recovery techniques to ensure that the message is received by the destination node. Exploiting the natural redundancy in sensor networks is key to adding to overall network reliability.

Ultimately all operations in the sensor network must be relevant to the task of meeting science requirements. In current and future sensor networks routing protocols must be matched to the network’s data traffic patterns. Data capture networks have known communication patterns that can be exploited to reduce the amount of routing state that an individual node must maintain. This reduction in state frees memory resources, which can be allocated to storing more data points locally. In future sensor networks the recharging of nodes will play a more significant role for meeting science requirements. Routing in such networks will require that routing algorithms incorporate power down modes of nodes for battery recharging. While distributed recharging schemes allow for increased science by providing continual sensor node coverage of an event, the routing algorithms must be constructed to dynamically route around sleeping nodes.

7.3 Physical Implementation

Current generation small-scale sensor networks can economically be constructed using commercial components. However, large-scale sensor networks of the future can benefit from custom design to meet both performance as well as manufacturability constraints. Such design largely falls in the domain of systems-on-a-chip (SoC) research. These designs endeavor to fulfill computational, sensing, and wireless transmission requirements all within a single chip design. Low cost packaging technology begins to play a key role both from the perspective of the physical constraints of the target environment as well as the design constraints such as hosting digital and analog designs on the same substrate. Finally this research must be constrained by economics, realizing that the large quantity of nodes in future sensor networks dictates a low individual device expense. For example, applications have been identified that require costs on the order of pennies per device.

8 Conclusions

Sensor networks represent a valuable technology that can be applied in a number of scientific, military and commercial applications. While traditional system and network protocols can be utilized in these networks, it is clear that a number of benefits can be obtained by allowing sensor network software and hardware to be customized to meet specific application goals. Such optimizations represent value in increasing both the lifetime of the system as well as improving the amount of achievable science that sensor networks can obtain.

References

[1] P. Smith et.al., “Results from the Mars Pathfinder Camera,” in *Science*, Dec. 1997.

[2] M. Hadjitheodosiou, “Store-and-Forward Data Communications Using Small Terminals and Microsatellites,” in *Proceedings of Third IEEE Symposium on Computers and Communications*, pp. 79-83, 1998.

[3] J. McCall, “Advances in National Data Buoy Center Technology,” in *Conference Proceedings of OCEANS ’98*, Vol. 1, pp. 544-548, 1998.

[4] W. Macha and K. Rogers, “Submersible Pop-Up Buoy for RF Communication,” in *Proceedings of OCEANS ’97*, Vol. 1, pp. 186-191, 1997.

[5] J. Willis, K. Leschly, H. Javadi, B. Blaes, and K. Boykins, “Free Flying Magnetometers As a Demonstration of Micro-spacecraft Technologies,” in *Proceedings of 13th AIAA/USU Conference on Small Satellites*, 1999.

[6] E. Sozer, M. Stojanovic, J. Proakis, “Underwater acoustic networks,” in *IEEE Journal of Oceanic Engineering*, Vol. 25, Issue 1, pp. 72 –83, Jan. 2000.

[7] E. Sozer, J. Proakis, R. Stojanovic, J. Rice, A. Benson, M. Hatch, “Direct sequence spread spectrum based modem for under water acoustic communication and channel measurements,” in *OCEANS '99 MTS/IEEE. Riding the Crest into the 21st Century*, Vol. 1, pp. 228 –233, 1999.

[8] P. Sass and L.Gorr, “Communications for the digitized battlefield of the 21st century,” in *IEEE Communications Magazine*, Vol. 33, Issue 10, Oct. 1995.

[9] E. Royer and C-K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," in *IEEE Personal Communications*, Vol. 6, No. 2, pp. 46-55, Apr. 1999.

[10] D. Johnson, D. Maltz, "The Dynamic Source Routing Protocol for Mobile Ad Hoc networks," IETF Internet draft, draft-itef-manet-dsr-01.txt, Dec. 1998.

[11] Bluetooth Standard Industry Website: http://www.bluetooth.com

[12] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," in *Proceedings of the Hawaii Conference on System Sciences*, Jan. 2000.

[13] J. Kahn, R. Katz, and K. Pister, "Mobile Networking for Smart Dust," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking*, Aug. 1999.

[14] S. Hollar "COTS Dust," Masters Thesis (work in progress), University of California at Berkeley.

[15] K. Delin and S. Jackson, "Sensor Web for In Situ Exploration of Gaseous Biosignatures," in *Proceedings of IEEE Aerospace Conference*, Mar. 2000.

[16] K. Delin, S. Jackson, and R. Some, "Sensor Webs," in *NASA Tech Brief*, Vol.23, No. 10, Oct. 1999.

[17] FlexSilicon Corporate Website: http://www.flexsilicon.com

[18] Chip Express Corporate Website: http://www.chipexpress.com

[19] International Standards Organisation Description of Basic Reference Model for Open System Interconnection. ISO 7498, 1984.

[20] Mars Microrover Telecommunications Web Page: http://lunar.ksc.nasa.gov/mars/rovercom/radio.html

[21] L. Taylor, "White Paper: HIPERLAN Type1 Technology Overview," TTL Communications Ltd., June 1999.

[22] H. Woesner, J. Ebert, M. Schlager, and A. Wolisz, "Power-Saving Mechanisms in Emerging Standards for Wireless LANs: The MAC Level Perspective," in *IEEE Personal Communications*, pp. 40-48, Jun. 1998.

[23] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II – the hidden terminal problem in carrier sense multiple-access modes and the busy tone solution," *in IEEE Transactions on Communication*, Vol. 23, No. 12, pp. 1417-1433, 1975.

[24] P. Karn, "MACA: A New Channel access Method for Packet Radio," in *Proceedings of 9th Computer Networking Conference*, pp. 134-140, 1990.

[25] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A Medium Access Protocol for Wireless LAN's," in *Proceedings of ACM SIGCOMM 94*, pp. 212-25, 1994.

[26] F. Talucci and M. Gerla, "MACA-BI (MACA by invitation). A Wireless MAC Protocol for High Speed Ad hoc Networking," in *Proceedings of IEEE International Conference on Universal Personal Communication*, 1997.

[27] S. Singh, C.S. Raghavendra, "Power efficient MAC protocol for multihop radio networks," in *Proceedings of the 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Vol. 1, pp. 153-157, 1998.

[28] J. Wieselthier, G. Nguyen, and A. Ephremides, "On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Networks," in *Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 2, pp. 585 –594, 2000.

[29] W. Lewandowski, J. Azoubib, and W. Klepczynski, "GPS: Primary Tool for Time Transfer," in *Proceedings of the IEEE*, Vol. 87, No. 1, Jan. 1999.

[30] Aztec OEM 4000 Differential GPS Receiver Product Description, Ag Com Ab Corporate Website: http://www.ag-com.se

[31] A. Lyne, "Pulsars as clocks**"** in *European Frequency and Time Forum*, pp. 1-7, 1996.

[32] T. Rappaport, *Wireless Communications: Principles and Practice,* Prentice Hall, New Jersey 1996.

[33] A. Amis, R. Prakash, T. Vuong and D. Huynh, "Max-Min D-Cluster Formation in Wireless Ad Hoc Networks," in *Proceedings of IEEE INFOCOM'2000*, Mar. 2000.

[34] J. Tsai and M. Gerla, "Multicluster, Mobile, Multimedia Radio Network," in *ACM-Baltzer Journal of Wireless Networks*, Vol.1, No.3, pp. 255-65, 1995.

[35] B. Everitt, *Cluster Analysis*. Heinemann Educational Books Ltd., 1974.

[36] W. Dally and C. Seitz, "Deadlock-free message routing in multiprocessor interconnection networks," in *IEEE Transactions on Computers,* Vol. C-36, No. 5, pp. 547-553, May 1987.

[37] T. Pinkston, "Flexible and Efficient Routing Based on Progressive Deadlock Recovery," in *IEEE Transactions on Computers,* Vol. 48, No. 7, pp. 649-669, 1999.

[38] T. Pinkston and S. Warnakulasuriya, "On Deadlocks in Interconnection Networks," in *Proceedings of the 24th International Symposium on Computer Architecture*, Jun. 1997.