"GANGS" MAC Protocol Proposal

Saâd Biaz          Dai Yawen
Computer Science and Software Engineering Department
Auburn University
Auburn, AL 36849-5347, USA
{sbiaz,dyeaia}@eng.auburn.edu
Technical Report # CSSE03-07
June 26, 2003

Abstract
Recent advances in wireless communication and electronics have enabled the development of low-cost sensor networks. Low energy storage is one of the critical features of nodes in these networks. Communication protocols at different layers have been proposed in order to reduce the energy consumption. This paper presents a MAC layer protocol, named "GANGS", as a method of saving energy in sensor networks.

1 Introduction
A sensor network is composed of a large number of sensor nodes that are densely deployed either inside the phenomenon being observed or very close to it. The position of sensor nodes is not necessarily engineered or predetermined. Because of this, sensor network protocols and algorithms must be self-organizing. The sensor network results from the cooperative effort of sensor nodes, which can be categorized as: data source, a node that generates the data; data sink, a node that collects the data, and ordinary node, a node that participates in the data forwarding from a data source to a data sink. Although many protocols and algorithms have been proposed for traditional wireless ad-hoc networks, they are not well suited to the unique features and requirements of sensor networks. The number of sensor nodes in a sensor network is commonly several orders of magnitude higher than that of an ad-hoc network. Sensor nodes are densely deployed, and are limited in power, computational capacity and memory. New protocols and algorithms must be designed with
consideration of these differences, specifically, limited and non renewable power storage of
the sensor nodes.

Energy in sensor nodes is consumed by the functions of multiple components such as
processing units, radios, sensors, actuators and power supplies. Actuators consume the
most energy, followed by radios. Processor and sensor power consumption are usually less
important [8]. Shutdown and scaling are the main techniques used to minimize energy
consumption for radios. The idea behind shutdown is that the node operates at a fixed
transmission rate and power level, and shuts down the radio after the transmission, thus
there is no superfluous energy consumption. Scaling is based on the relation between
performance and energy requirements. The node varies properties such as modulation and
error coding, trading off energy consumption for transmission time. Figure 1 part (a)
shows the relationship between transmission time on the x-axis and required energy per bit
reliably transmitted on the y-axis. For a typical system, there is a minimal energy per bit
($E_e$), at point E. If the transmission time is longer or shorter than $T_e$, the energy consumed
for each bit is more than $E_e$. If the node is allowed to send data for a shorter time than
$T_e$ ($T_a < T_e$), no energy can be saved by scaling. If the node is allowed to send data for a
longer period than $T_e$ ($T_a > T_e$), then the node can use scaling to keep the transmission time
near $T_e$, and shutdown for the rest of the allowed time. Scaling can be done by altering
modulation or coding. A modulated radio signal consists of different symbols, which can be
of different shape, different frequency, etc. Each symbol represents several bits of data. In
modulation scaling, the number of bits represented by each symbol is varied. By decreasing
the number of bits per symbol, the transmission time for a fixed amount of data can be
increased, and the transmission power can be decreased. In code scaling, the amount of
coding overhead, such as error detection or correction, is varied. By increasing the coding
overhead, the transmission time for a fixed amount of data can be increased. Both scaling
methods are trying to adjust the transmission time to meet the optimal transmission time
requirement. [9].

The choice of the radio management method is based on the energy-transmission time
curve of the system. The shape of the curve depends on the relative importance of RF
and electronics, and is a function of the transmission range, as shown on figure 1 part (b).
For long-range systems, the energy consumed per bit decreases as the transmission time
(a) Energy Vs. transmission Time  
(b) Performance Vs. Range

Figure 1: Scaling and Shutdown

increases, so these systems have an operational region where they benefit from scaling. For short-range systems, the minimal required energy per bit point occurs at a short transmission time point, so these systems have an operational region where scaling is not beneficial and the best strategy is to transmit as fast as possible and shutdown.

Another approach to energy efficient sensor nodes is to exploit information such as location, timing, specific features of the application, or neighbor information to enable cross-layer optimizations.

Currently, energy efficiency in sensor networks is achieved by using the following techniques:

1. Pre-process raw data before transmission, trading off communication energy against computation energy. This is called data aggregation.
2. Shut down some unnecessary neighbor nodes.

3. Only forward data to a specific neighbor or set of neighbors according to location information or some other metric.

4. Use indirect routes instead of direct routes. This is based on the fact that under some conditions, using indirect paths consumes less energy than using a longer but direct path. Suppose there are three nodes on one line, A, B and C, in that order. A sends data to C. In order for C to receive the data, the energy of the signal must at least be e. Since the power of the signal decreases at the magnitude of square of the distance, A must send out a signal of power $e \cdot AC^2$. If A sends the data to B and B sends the data to C, then the power needed would be $e \cdot AB^2 + e \cdot BC^2$, which is less than $e \cdot AC^2$. Thus the indirect path is more efficient.

5. Increase the packet size. This is based on the fact that turning on a transmitter consumes energy. If the energy to turn on a transmitter is significant compared to the transmission energy, then sending a longer packet would save energy.

6. Avoid contention on the communication medium. Less contention implies less retransmissions.

2 Related Work

Several protocols have been proposed to optimize the energy consumption of sensor networks. They cover the areas of network layer protocols, MAC layer protocols and cross-layer protocols.

Diffusion [4], Rumor [7], and GSPR [1] are candidates for network layer optimization. Diffusion protocol is a task-based network layer protocol. Information in a diffusion based sensor network consists of interest packets and data packets. Interest packets are packets informing sensor nodes about the data that the sink wants to collect. The data sink broadcasts interest packets. Data packets are sent back to the sink by the source when interest packets reach the source that can provide the desired information. Each node that participates in the data forwarding remembers the paths that data packets arrive by, and records the shortest path. The shortest path between the source and the sink is found by reversing the shortest path that is recorded by the intermediate nodes. After
the shortest path is found, data packets are primarily transmitted on that path. Diffusion avoids flooding of data packets, and thus saves energy.

In a sensor network in which Rumor protocol is deployed, information consists of queries and events. Events are the data that needs to be collected. Queries are messages sent out to retrieve events. Rumor is a logical compromise between flooding queries and flooding event notifications. Event flooding creates a network-wide gradient field. When a query is generated it can be sent on a random walk until it finds the event path, instead of being flooded through the network.

GPSR heavily uses geography to achieve scalability. It assumes that all wireless routers know their positions and node sources can determine the locations of node destinations. The main idea is to forward packets to the neighbor closest to the destination.

While the network layer optimizations attempt to optimize the network topology to minimize data flooding and thus decrease energy consumption, MAC layer optimizations focus on decreasing the contention and certain type of unnecessary energy consumption.

MAC protocols address four main sources of energy consumption: collision, overhearing, control overhead, and idle listening. Collisions occur when neighbor nodes transmit at the same time and data get garbled. Overhearing results from nodes wasting energy listening to data not meant for them. Control overhead consists the hand-shaking (RTS/CTS) signal sent out to make sure the transmission media is available before sending data. Idle Listening means that nodes have the radio on when there is no data transmission occurring.

There are two categories of existing MAC protocol. In the first category includes contention-based MAC protocols such as IEEE802.11 [5]. The main problem with these protocols is that they consume energy by idle listening. PAMAS [6] is based on IEEE 802.11, and uses two different radio channels for signaling and transmitting data. Since it performs signaling on a different channel than the data transmission channel, nodes know whether or not the data is for them. Thus it avoids overhearing among the neighbor nodes, but it does not address the idle listening problem. In the second category are contention free MAC protocols such as TDMA. TDMA reduces the energy consumption because it eliminates contention. Two problems with the TDMA protocol are that it does not support scalability and it requires centralized control of all nodes. S-MAC [11] protocol is designed specifically for sensor networks. It uses RTS/CTS to avoid collisions. It handles overhearing
by turning off a node’s radio when a transmission is not meant for it. Control overhead is handled by message passing. Only one pair of RTS/CTS along with some ACKs are sent during a burst of data transmission. It uses periodic sleeping and listening to reduce idle listening. LEACH [12] is a cluster-based MAC protocol. Nodes elect themselves periodically and randomly as cluster-head, and the nodes in each cluster adopt a TDMA scheme. LEACH does not address inter-cluster communication, so it is not very practical for sensor network. ASCENT [3] is a sub-layer protocol which is designed to work between the network layer protocol and the underlying MAC protocol. Nodes in the sensor network in which ASCENT is deployed select one set of neighbors to be active. Passive neighbors only listen, and do not transmit. If nodes experience a degraded performance, they send help messages to some passive neighbors to wake them up. The awakened nodes begin to participate in data forwarding.

3 GANNS Protocol

We propose to create a cluster-based MAC protocol that takes advantage of the mechanism of both contention based and TDMA protocols. Our assumption is that for most nodes, the forwarded traffic is much heavier than originated traffic. That is, most of the bandwidth is used to forward other node’s data. The purpose of our protocol is to avoid contention for forwarded traffic. The network of sensor nodes is divided into clusters. Each cluster has a cluster head. Cluster heads form the backbone of the sensor network. This backbone carries forwarding data from one cluster to another. Nodes in one cluster only talk to their cluster heads. The backbone of cluster heads can be exploited by a network layer protocol to optimize routing. The TDMA scheme is adopted for communication between cluster heads. A contention-based scheme is used among nodes within each cluster. Time is divided into frames. Each frame is divided into several slots. There are two kinds of slots, contention-free TDMA slots and contention-based slots. A contention-free TDMA slot is dedicated to cluster heads. A contention slot is a piece of time that is shared among all the nodes in the cluster for exchanging data with their cluster heads. Each frame has several TDMA slots and one contention slot. The number of TDMA slots depends on the number of connections with neighboring cluster heads. The radio for each node (other than a cluster head) is turned off during all TDMA slots and turned on during the contention
slot. A cluster head’s radio is always on. Each cluster head communicates with its neighbor cluster heads during TDMA slots. It sends out its data to its neighbor cluster heads during its dedicated TDMA slot and listens to the data from neighbor cluster heads during their TDMA slots. Thus, the bandwidth between cluster heads is reserved, and contention caused by the large traffic among cluster heads will be reduced.

We call our protocol GANGS. Each cluster acts like a gang. The most powerful node (with highest remaining energy) will be elected as cluster/gang head. Each cluster head controls its own cluster, and negotiates with other clusters/gangs through their cluster heads. The gangs construct a network in which the transmission can reach every single cluster. Because cluster heads are doing more work and consuming more energy than other nodes, a cluster head will eventually become less powerful than another node in its cluster. When this happens, the more powerful node will take over the position and cause a reconfiguration of clusters/gangs.

3.1 Scenario of GANGS

On all the figures in this paper, a cluster head is represented with a shadowed square, and the ordinary node is represented with a filled circle. A dashed and external circle represents the radio range. Figure 2 shows the scenario of GANGS protocol.

3.2 Time Frame of GANGS

1. For Cluster Head
   Shown in figure 3.

2. For Ordinary Node
   Shown in figure 3.

3.3 Establishing Clusters

1. Local maximum stage

   When a node is up, (see figure 4) it communicates with its neighbors and provides its energy information. At the beginning of setup, the node that has the maximum energy among all its neighbors, named “local maximum”, claims that it is a cluster
head and sends this claim to its neighbors. Its neighbors decide whether or not they will accept this cluster head. In the following description, a cluster head will simply be referred as a "head". Note that no head is within any others’ range at this stage.

2. *Inter – Cluster stage* Add more cluster heads to construct the back bone.

After the first stage, a node that is not a head will be in one of the following three situations: situation 1, it is in the range of one head and accepts the head; situation 2, it is in the range of multiple heads and needs to choose one head among them, as in figure 5 case 1; or situation 3, it is not in the range of any head, figure 5 case 2. In figure 5 case 1, nodes A and B are in the intersection of two clusters. Both nodes receive claims from heads C1 and C2. These nodes are aware of each others’ energy information. The node that has more power, node A in this case, will claim to be a new head.
Figure 3: Time frame for cluster head/node

In figure 5 case 2, node A is not in the range of any head. Within its own range, which is represented by the dashed hexagon, there is a local maximum node B. B accepted a head during the local maximum stage. In this situation, node A sends a message to B to demand head service. B then proclaims itself a head and the topology changes. Each node that has not had a head yet will still fall into situation 2 or 3 specified above. In this way, the backbone will gradually be constructed. According to the paper," optimum transmission radii for packet radio networks or why six is a magic number" [10], if the network degree is around six, the probability that the network is fully connected is approximately 0.95. So we can assume that if each head has about six neighbor heads, the cluster heads on the backbone will be fully connected, and thus the whole sensor network will be fully connected.

3. Reconfiguration stage
The energy consumption of heads is usually higher than ordinary nodes. After a while, the head may not be the most powerful node in its cluster. When a node is more powerful than its cluster head, and other conditions based on energy information and other metrics are satisfied, reconfiguration will be done. The current local maximum will elect itself and start the reconfiguration.

3.4 Arrange TDMA Schedule

Because of the requirement of synchronization between nodes, TDMA networks are not scalable. In a sensor network, synchronization is not required for the entire network. Only synchronization between neighboring cluster heads is needed. After the clusters are established, we only need to consider the TDMA schedule among the cluster heads that comprise the backbone.

For the situation in figure 6, we give out a sample schedule:
A: AB**
B: ABC*
C: EBCD
D: *FCD
E: EFC*
F: EFGD
Then we arrange a time frame of specific length, call it T:

Shown in figure 7. Each T-4*L period of time is used for contention based traffic. Heads send TDMA schedules to their nodes, and the nodes will shutdown during TDMA slots and wake up during contention slots. Every cluster has the same frame length. It is not necessary that every cluster have the same contention slot length. The actual contention slot length should be based on the connection information of the current head, such as how many neighbor heads it has.

1. Time slot arrangement

Each head knows its neighbors’ information and the total number of neighbors. For example, in figure 6, Node A has one neighbor and Node C has three neighbors. Each
Figure 6: TDMA Schedule between Cluster Heads

head randomly chooses a number from one to the number of its neighbors plus one. Node A randomly chooses a number from one to two and Node C randomly chooses a number from one to four. Node A and C send out the numbers to their neighbors. If their chosen numbers are the same, the head that has either fewer or more neighbors will change the its selection. A good algorithm is required to achieve fast scheduling.

2. Synchronization method

Cluster heads adjust according to the information of their neighbors. Other nodes follow the head to which they belong to. We are still considering whether or not it is necessary to reserve a time slot for signaling.

4 Conclusion

In this proposed protocol, most nodes are working in sleep-listen mode, as in the S-MAC protocol. TDMA will reduce contention more than the RTS/CTS pair in S-MAC, and if the traffic fits our assumption that forwarded traffic is much heavier than originated traffic,
then we have reason to believe it will be superior. The backbone constructed by cluster heads works as a virtual wired network, and the contention time slots give flexibility to network access, i.e. every node can talk during the contention time slot, so we think it might help in handling mobility issues.

5 Further Considerations

1. If there is a data source in a cluster that is not the head, is it necessary to assign a time slot to it?

2. Which proportion of the time frame should the contention time slot occupy?

3. How much energy is consumed on cluster establishment and slot assignment?

4. What is the contention rate after the new MAC protocol is deployed?

5. The parameters such like the length of the packet, node shutdown time, etc, are based on the underlying physical layer. Is it possible or practical to design a general MAC protocol that takes all physical layers into consideration?

6. Considering the head change, how fast can the MAC layer reconfigure?

7. Is it possible for this MAC protocol to fit in mobile sensor networks? If the MAC layer reacts quickly to the head change, then this is likely.
8. During the TDMA time slots, should each cluster listen to every slot when it is not its
turn to send, or just listen to the slots in which it will receive data, as in traditional
TDMA?

9. What is the appropriate energy model to evaluate the protocol?

10. How can we balance latency and energy consumption?

11. In general, can we conceive of a MAC layer that smoothly varies from purely contention
based to contention free? Considering the GANGS protocol proposed above, we might
be able to do it by adjusting the length of contention slot. If the traffic is not heavy,
we can just eliminate the TDMA slots and make the protocol completely contention
based; if the forwarding traffic is heavy and the node density is not low, we can just
adopt GANGS as it is; if both forwarding traffic and data source traffic from the
cluster are heavy and the node density is reasonably low, we can reduce the length of
the contention time slot and increase the number of TDMA slots.

References


gies,” June 2002.


