Technical Report, IDE0749, June 2007

State of the Art
on Energy-Efficient and Latency-Constrained Networking Protocols for
Wireless Sensor Networks

Annette Böhm

CONTENTS:

1 INTRODUCTION

Advances in the development and integration of small battery-powered sensing devices have enabled the design of applications where a group of sensors cooperate in monitoring their environment. Wireless sensor networks (WSN) are today used in a wide variety of areas like e.g. habitat monitoring, battle field surveillance, disaster management, health monitoring or industrial control. Their applicability is often reduced by limitations in the sensor nodes power supply, communication bandwidth, processing capabilities and buffer size. Many researchers have put effort in overcoming these shortcomings, with a special focus on maximizing the battery lifetime of a single node as well as the lifetime of the network as a whole. A list of possible applications can be found in [Akyildiz].

As applications become more and more mission-critical, it is crucial that the collected sensor data arrive at the sink within a specified time limit. Guaranteeing a certain quality of service (QoS) to a user or an application is difficult because of the unpredictable nature of the wireless link and the often unstable topology of the sensor network (due to node failure or mobility). Very little research has been done in the field of QoS for WSNs [Chen] [Wang] and many interesting research questions are still unanswered.

This state-of-the-art document aims at giving an overview on routing and medium access issues in WSNs and at summarizing some of the most interesting solutions. The requirements on WSN protocols are very dependent on the application in which the WSN will be used. Aspects as energy-efficiency, latency, Quality of Service (QoS), mobility, distribution density or cost all influence the choice of protocol and its parameters. There is therefore no single protocol that matches all types of WSN and the best results can only be achieved by tailoring the protocol for a specific application or scenario.

In this state-of-the-art report, we make certain assumptions, loosely based upon the requirements of a surveillance application:

- A (possibly large) number of sensor nodes, distributed over the area or object in question, are used to collect data and/or register events.
- Energy in the sensor nodes is a scarce resource and should be saved as much as possible
- Data is collected in sinks (possibly nodes with looser energy constraints and better processing and memory capabilities than normal sensor nodes in the network)
- Communication to and from the sink or between different sensor nodes cannot always be done in a single hop. Multi-hop capabilities are therefore required.
- In case of an event, data must be able to flow to the sink within tight timing restrictions and high requirements on dependability.
- The nodes have either relatively low mobility or are completely static.

If the WSN is then used for the surveillance of a property, a habitat, a volcano or something else is left unsaid.

2 General constraints and requirements

Due to the characteristics of typical WSN applications, the network has to be as self-maintaining as possible. Robustness to node failure or changes in the environment are vital, as well as the ability to adapt to different network sizes (scalability). But the most important requirements for many (surveillance) applications are a long network lifetime (i.e. a low level of energy consumption) and a high level of Quality of Service (QoS).

2.1 Energy consumption

In typical WSN scenarios space and cost constraints make the use of large batteries impossible. Nevertheless, it is often not feasible to change batteries on a regular basis. It is therefore vital that sensor node save as much energy as possible and prolong the lifetime of an individual node as well as the network lifetime. In general, network lifetime can be defined in different ways – e.g. as the time until the first node dies, the time until a certain percentage of nodes has died or the time until the network is partitioned in disconnected islands – and the definition is very much application dependent [Willig].

Major sources of energy consumption in a WSN are [Ye]:

- ***Idle listening*** – Whenever the radio is listening to the channel without receiving actual data, energy is wasted. In order to save energy, nodes should be turned of during the time they are not needed for sensing, communication or processing. Restricting a sensor node's duty cycle comes of course with the challenge of waking the node up from its sleep mode at the right time to fulfill its job.
- ***Communication*** is by far the most energy consuming aspect in WSNs [Ganesan]. One of the main goals to save energy is therefore to reduce communication to a minimum, i.e. both data and control traffic.
    - *Low control overhead* - Protocols that need low volumes of control traffic to work are therefore to be favored over control traffic intensive ones.
    - *Data aggregation* - Data traffic can be reduced by processing and aggregating data as early as possible. This, of course, puts high requirements on the individual sensors' processing and memory capacities.
    - *Data filtering* - Alternatively, data can be selectively transmitted as users or applications might only need to know changes in sensor reading values instead of each single reading. Furthermore, there is often a correlation between the readings of neighboring sensors and it might be sufficient to send one reading, representative of a whole region.
- ***Overhearing*** – In dense or broadcast-based WSNs nodes continuously receive packets that are not destined for it and that have to be partly processed and then discarded.

2.2 Quality of Service (QoS)

Often sensor networks are employed in hostile environments where collecting data manually is not feasible and where even the deployment of the sensors, their maintenance and organization are not always straight forward. WSNs are therefore expected to be self-creating and self-organizing and use intermediate nodes to reach the sink in a multi-hop fashion. This makes QoS support more difficult and in this point WSNs are comparable to wireless ad hoc networks. The following list introduces the major limiting factors to QoS support:

- ***Resource limitations*** - A sensor node is restricted by limited processing capabilities, memory and buffer size. The bandwidth of the wireless channel is limited as well as the transmission power of the sensor nodes, but the most severe constraint is the limited access to battery power. Traditional QoS routing and MAC protocols are badly suited due to their large communication overhead and demand for processing, buffering and memory.
- **Network instability** - Due to link failure, power failure, mobility or the fact that certain nodes are in sleep mode to save energy, the network topology might change frequently. Routing and medium access under these unstable conditions is challenging.
- **Unpredictable traffic patterns** - QoS support in traditional networks is often dependent on a certain periodicity of the data traffic. In WSNs periodic sensor readings (e.g. taking a temperature sample once a second) represent just one of three possible data models [Chen]. The other two are query-driven and event-driven applications, where sensor readings are pulled by the sink or pushed to the sink, respectively.
- **Redundancy** - An event (an intruder, a change in temperature, a chemical substance) is often detected by more than one sensor and redundant data is produced. This redundancy can be valuable from a QoS point of view compensating for lost packets or failed links but is often eliminated by data aggregation to save energy.
- **Uneven traffic distribution** - The deployment of sensor nodes does usually not follow a regular pattern, often a huge number of nodes are scattered over an area of interest [Wang]. The high density and uneven distribution must be taken into account when e.g. designing QoS-aware routing protocols as well as the fact that data mainly flows from a large number of sensor nodes to a small number of sinks, putting higher demands on nodes close to the sink [Chen].
- **Heterogeneous traffic types** - Applications might need access to heterogeneous data collected by different types of sensors with different sampling rates [Chen] [Al-Karaki]. This heterogeneous environment makes QoS support more complex.
- **Non end-to-end applications** - Opposed to most other types of networks, end-to-end parameters like delay, throughput or packet loss from one single source to one single destination are not sufficient to define QoS in WSNs [Chen]. In typical applications, the performance of an individual sensor node or communication link is not as interesting as the collective performance of a group of sensors. This makes QoS handling more complex and difficult to measure.

3 Routing in Wireless Sensor Networks

In WSNs, several basic communication scenarios can be considered. Firstly, a sensor node sends periodic or event-driven data towards the sink. Here we have a unicast situation. Secondly, a sink wants to query a certain group of sensor nodes. Alternatively, the sink or any other node wants to reach parts of the network (possibly the neighborhood). In both cases, we need multicast or broadcast capabilities. In this part of the survey, routing protocols focusing on energy-efficiency and QoS are described. Position-based routing protocols are an interesting group as they make use of the node's location information.

The amount of existing WSN routing protocols is huge. As most of those have energy-efficiency as an aspect, there will only be a short overview of the most important representatives of different WSN routing techniques. Protocols that even take QoS into consideration are described in more detail in the last part of this chapter.

3.1 Proactive vs. reactive routing

Proactive routing protocols offer up-to-date routes to any destination in the network at any given time by frequently updating the routing tables. The main advantage of this approach is the minimal delay an application experiences when it wants to send information across the network. Keeping track of all the routes through the network, on the other hand, introduces a large protocol overhead. The routing table updates are usually periodic but can be extended by event-driven updates to quickly react to network changes.

Reactive routing means that routing information is not gathered in advance, but first when requested by an application. This saves control overhead but introduces the need for a discovery phase each time an application needs to send data over the network and increases the communication delay. In the route discovery phase, the network is partially or entirely flooded with route request messages to find the shortest path to the destination. In reactive routing, the scalability is better because routing information is only exchanged when needed and a node can be certain to use the most recent routing information, which leads to a higher transmission success than with proactive routing where routes might be outdated. Flooding an entire WNS means considerable communication efforts and the flooding should be restricted to a particular area of interest to minimize the energy loss.

The scale of a WSN, its network dynamics and energy constraints are factors in favor of reactive routing protocols using local information [Ganesan]. This enables the network to react quickly to local changes or bursty traffic and does not require the global transmission of routing information. On the other hand, when information has to be transmitted quickly from nodes to the sink (e.g. when an alarm was triggered), a predetermined path from source to sink reduces the delay considerably.

3.2 Flat routing, hierarchical routing, georouting

Routing in WSNs can also be divided based on their network structure [Al-Karaki]. In flat routing, nodes have equal roles in the routing process and collaborate in packet forwarding. Hierarchical routing is based on a clustered structure, where certain nodes play coordinating roles. The creation of clusters can contribute to the scalability of the network and improve overall network lifetime. Georouting (even known as location-based, position-based or geographic routing) uses the knowledge of the nodes' positions (either global or relative to each other) in the routing process.

3.2.1 Flat routing

- ***Directed Diffusion [Intanagonwiwat]***
  Energy savings in directed diffusion are due to a higher level naming procedure. A node names data it has to offer with high level attributes and spread these attributes throughout the network. When a sink is interested in a certain type of data, it sends out queries that are compared to the attributes a node has knowledge of and a path to the node with the proper answer to the query is established. Energy savings are due to the fact that knowledge is spread without having to send the data itself.

- ***Energy-aware routing protocol [Shah]***
  In the energy-aware routing protocol, network lifetime is increased by maintaining different paths from source to sink instead of one optimal one. Depending on the energy level of the nodes along the paths, the one that does the least damage to the overall network lifetime is chosen. By distributing the energy consumption as evenly as possible over the network, the network lifetime is increased.

3.2.2 Hierarchical routing

- ***Low Energy Adaptive Clustering Hierarchical protocol (LEACH) [Heinzelman]***
  In the LEACH protocol, clusters of nodes are built around randomly selected cluster heads. The role as the cluster head is rotated periodically to evenly distribute the energy load. Cluster heads collect data from the sensor nodes in its cluster and perform data aggregation in order to reduce the amount of data that has to travel through the network to the sink. This protocol is most useful for applications where data from monitoring is sent periodically.

- ***Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [Lindsey]***
  PEGASIS is not really based on the formation of clusters, but more on the formation of a communication chain. Each node only communicates with its nearest neighbor which aggregates the data with its own and sends it on to its nearest neighbor. One node is selected to communicate the results of a chain to the sink. If the transmitted power can be adapted to only reach the nearest neighbor, energy can be saved compared to transmitting on full power to reach as many as possible. By changing the order of the chain and the role of communicating with the sink, the load is balanced over the network.

- ***Threshold-Sensitive Energy Efficient Protocol (TEEN) [Manjeshwar] and Adaptive Periodic TEEN (APTEEN) [Manjeshwar2]***
  TEEN reduces the number of data transmission from source to sink. The sink specifies threshold values for interesting data and sends them to the sensor nodes. One threshold specifies e.g. what kind of sensed data that is interesting to the sink at the moment. The other threshold defines with how much the sensed value must differ from the last sent value in order to be interesting to the sink. Thereby, only significant changes trigger communication.

3.2.3 Georouting

In geographic routing the knowledge of a node's position plays a vital role in the routing process. The topic can generally be divided into two areas: position-based routing and geocasting [Willig]. In position-based routing the position of source, destination and the intermediate nodes has to be known to help the routing process. Routing tables are decreased or omitted entirely as the geographic location of a node carries information about its next hop neighbors and their vicinity to the destination node. Here energy savings can be made, but, on the other hand, determining each node's current position introduces itself processing and communication needs.

Geocasting is a concept that fits the nature of typical WSN applications very well. The sink is usually interested in events and activities in certain geographic areas. Which individual sensor that delivers the event data is not really of importance. Based on this concept, geocasting aims at addressing all nodes in a certain geographic region rather than a single sensor node.

Recent research in localization has shown that cheap and precise localization is possible, even in nodes that cannot accommodate a GPS receiver due to space and energy constraints [Ganesan]. As applications using WSN data often are interested in information from a certain geographic region and sensor data is likely to be geographically correlated, routing protocols based upon position information have great potential. Surveys of geographic routing methods for WSNs can be found in [Willig], [Seada], [Maihöfer] and [Mauve].

3.2.3.1 Position-based routing

Position-based routing (other frequently used names are georouting, location-based or geographical routing) consists of two stages [Mauve]. First a location service is responsible for providing a source with the position of the desired destination. In the second stage, the forwarding phase, the route from source to destination is determined.

Location Services

Popular location services are ***Distance Routing Effect Algorithm for Mobility (DREAM)*** [Basagni] and ***Grid Location Service (GLS)*** [Li].

- ***DREAM***
  In DREAM, each node maintains a location database with the positions of every other node in the network. The database entries are periodically broadcasted to all nodes within a predefined radius and the location databases are updated.

DREAM does not scale very well. On the other hand, a source node can access a destination's position without requesting information from other nodes, which reduces the communication delay and makes it a candidate for local communication between cars in emergency situations [Mauve].

- ***GLS***
  GLS uses a different strategy. It divides the network into hierarchical squares. Each node keeps a table with location information of all other nodes in the same first-order square (the smallest square in the hierarchy). In each of the first-order squares surrounding the source node, the node with an ID closest (but higher) than the source node's ID is chosen as location server and stores the location information. In each of the surrounding second-order squares a location server is chosen in the same fashion and so on. Closer to the node, the location server density is high but the further away you get, the scarcer the location servers get. This strategy scales better but suffers from the delay introduced by finding the right location server when a node's location has to be determined.

Forwarding strategies

Greedy packet forwarding

The sending node includes the position of the destination in the data packet and forwards it to a node closer to the destination. Each intermediate node does the same, i.e. it chooses a next hop that brings the packet closer to the destination. Different greedy routing strategies are described in [Mauve].

- *Most forward within r* relays the packet to the node within the transmission range that is closest to the destination.
- *Nearest with forward progress* chooses the node closest to the sender but still closer to the destination than the sender itself. This strategy pays off in terms of energy savings when the transmission range is adaptable.
- *Compass routing* forwards the data to the node closest to a straight line between sender and receiver and thereby minimizing the spatial distance the packet travels.
- Even a randomly chosen next hop has been proposed.

The disadvantage of greedy packet forwarding is that it cannot be guaranteed that a path can be found even if it actually exists. The ***Greedy Parameter Stateless Routing Protocol (GPSR)*** [Karp] defines a recovery mode when this case occurs and constructs a planar subgraph of the network between sender and receiver. If a path between sending and destination node does exist, it can be found along the subgraph. As soon as a node closer to the destination than the sender is reached, the recovery mode is stopped and normal greedy forwarding is reentered.

Restricted Directional Flooding

In contrast to greedy forwarding, where a packet was forwarded to a single node only, in restricted flooding the message is flooded to all nodes within a given region. There a different ways to determine such a region.

- The ***DREAM*** protocol mentioned as a location service, even defines one possible restricted flooding strategy. An *expected region* is defined by a circle around the position where the destination is expected to be found. The radius can be adapted based on the known speed of the destination and the

freshness of the information. The more likely, the location information is outdated, the larger the radius. Two straight lines between the sender and each side of the circle define the area to be flooded. See figure 1.
- Location Aided Routing (***LAR***) [Ko] uses a similar strategy. It is not a routing protocol itself but it aims at restricting the flooding of route requests in the route discovery phase of a reactive protocol. In LAR, a circular *expected zone* around the destination is created based on available position information and a *request zone* is defined as the smallest rectangle containing both the source and the expected zone. The route request is then flooded within the request zone.

3.2.3.2 Geocast

In general, two different strategies for geocast can be found. Either data is forwarded according to a unicast protocol until it reaches a geographic destination region where flooding is used to reach the targeted node(s), or a geographically restricted flooding strategy is employed from the beginning [Willig].

The basic forwarding technique is simple flooding, where a message is sent to all nodes in the transmission range, which in their turn broadcast the information further until the destination is reached. Nodes that already received the same packet before disregard it. As a result of the flooding process, each node in the network will have received the packet at least once, including the destination nodes in a specified geocasting region. This strategy introduces of course unwanted redundancy and thereby waste of energy and bandwidth.

Below are examples of geocast protocols. [Maihöfer] gives a survey of current geocast protocols.

- ***Location-based Multicast [Ko]***
  Location-based multicast (LBM) is based on the concepts of location-based routing (LAR) and directional flooding. It defines a destination region and a flooding zone and discards packets outside this zone. As in LAR, the flooding zone is a rectangle including at least the sender and the destination area. Another approach is to only allow intermediate nodes that are closer than a certain distance threshold from the destination area. A node calculates its geographical distance from the center of the destination region and determines if it is to forward the packet or to drop it. The distance is decreased with each hop.

- ***GeoGRID [Liao]***
  GeoGRID enhances the LBM approach by dividing the forwarding region into grids where only preselected gateway nodes are relaying the packets until the destination region is reached where the packets are flooded.

- ***Unicast Routing with Area Delivery***
  In his survey [Maihöfer], Maihöfer proposes to use any unicast georouting protocol of choice to reach the destination area and simply flood the region. He calls this suggestion Unicast Routing with Area Delivery.

3.3 Routing protocols with focus on QoS

Finding and maintaining paths in a WSN is complex due to the frequent and unpredictable changes in the topology. Many specially designed routing protocols for WSNs have been proposed. In most cases the communication overhead was reduced and the network lifetime prolonged by using hierarchical structures combined with data aggregation or location based protocols that target traffic at specific parts of the network. [Al-Karaki] gives an overview over recent research efforts in routing techniques for WSNs and points out that QoS issues need to be addressed in a far greater extent to support applications with real-time demands. Nevertheless, a few efforts in QoS routing can be found and will be presented in this section.

3.3.1 Protocol overview

- ***Sequential Assignment Routing (SAR) [Sohrabi]***
  The Sequential Assignment Routing (SAR) protocol is one of the first efforts to achieve QoS in a WSN. Multiple trees are created from all one-hop neighbors to the sink by successively adding nodes with higher and higher hop distances from the sink. Nodes with low QoS and energy reserves are omitted. In the end each node has multiple paths to the sink, each path associated with an energy resource parameter and an additive QoS metric (the lower the QoS metric the higher the QoS that can be provided). Taking into account the energy reserve parameter, QoS metric and the priority level of a packet, a source node chooses the most suitable path to a destination. A weighted QoS metric is calculated as the product of the QoS metric and a weight coefficient connected to the priority level of the packet. By minimizing the average weighted QoS metric throughout the network, both QoS and energy demands are taken care of.

  Path tables are maintained by periodic updates of the routing tables and enforcing routing table consistency between upstream and downstream nodes on each path. This creates overhead, but on the other hand, SAR is able to maintain multiple paths between nodes and a sink and ensures thereby fault tolerance and easy recovery from node failure. The algorithm has been shown to outperform protocols focused on energy-efficiency from both a QoS and an energy point of view.

- ***SPEED [Tian]***
  Speed is one of the first attempts to address real-time guarantees in wireless sensor networks. By maintaining a constant delivery speed throughout the network, an end-to-end delay can be calculated and a path can therefore be determined to provide soft real-time guarantees. Each node maintains information about its neighbors and estimates the end-to-end delay based on this information and the acknowledgement arrival time as a response to a transmitted packet. The neighbor that meets the end-to-end speed requirements is chosen as next hop. Apart from the routing protocol called Stateless Nondeterministic Geographic Forwarding (SNGF), SPEED comprises several modules that reduce network congestion. SPEED was compared to common, reactive ad-hoc routing protocols like DSR and AOVD and show better performance in terms of end-to-end delay and deadline-miss-ratio.

- ***Multi-path and multi-speed routing protocol (MMSPEED) [Felemban]***
  MMSPEED addresses both real-time and reliability requirements in a decentralized fashion, offering multiple levels of delivery speed and redundancy. The protocol relies on local routing decisions where the individual node’s knowledge is restricted

to its direct neighbors' location and link quality. The SPEED protocol [Tian] is used to guarantee a network-wide speed.
In the time domain, there is the choice between several speed levels that guarantee a high probability of reaching the sink in a certain time span. Each node maintains delay estimations for the transmission to the sink via its neighbors and compares the resulting speed to the chosen speed level and forwards the packet to the most promising neighbor in terms of delivery speed and closeness to the end destination. Each node has one queue per speed level, where the queue with the highest speed requirements always gets priority.
MMSPEED handles several reliability levels in a similar fashion. At the source a packet is assigned a certain reliability level which is connected to the probability of the packet reaching the sink. The more redundant forwarding paths used on the way, the lower the probability of loosing the packet to contention or transmission error. Even here, each intermediate node compares the requested level of reliability to the estimated packet loss rate through different neighbors to the destination and chooses one or more paths (that fulfill the timing requirements) to forward the packet closer to the sink.
In both cases there is an expected discrepancy between the locally made estimations and the actual transmission speed or delivery probability. An intermediate node can determine that by comparing a time estimation made by the previous node (contained in the packet) and the actual arrival time. By changing a packet's speed or delivery level, the node can dynamically compensate for estimation errors.
In order to handle prioritized access to the shared medium, multicast and the support of measurements of average delay and loss rate, the MMSPEED protocol needs adaptations on the MAC layer. Minor changes to the IEEE 802.11e MAC protocol suffice to handle the different priority levels. Shorter inter-frame spacing leads to a higher probability of high priority packets to get access to the shared medium.

- ***An Energy-Aware QoS Routing Protocol for WSNs [Akkaya]***
  In [Akkaya], Akkaya and Younis present a QoS and energy-aware routing protocol aimed at meeting the delay requirements of real-time traffic and maximizing throughput of best-effort traffic. The sensor nodes in this approach are clustered and managed by a gateway that communicates data further to the sink. The individual nodes in a cluster are one or more hops from the gateway and are not involved in the routing process, which keeps the overhead down compared to e.g. the SAR algorithm.
  There are separate queues for real-time and best-effort traffic and a dynamic bandwidth ratio value ensures that a certain amount of bandwidth is dedicated to best-effort traffic. An extended version of Dijkstra's algorithm finds a set of least cost paths which are checked against the end-to-end QoS requirements while adapting the bandwidth ratio value to maximize best-effort traffic throughput. Routing tables are sent out within a cluster.
  This approach is mainly targeted for surveillance applications, where imaging and video data demands QoS guarantees while other best effort traffic still has to be maintained to a reasonable extent.

3.3.2 Evaluation of the protocols

Table 1. compares the protocols on several points:
Akkaya et. al. and the authors of SAR take a node's energy reserves into account when choosing a forwarding path. The SPEED and MMSPEED protocols consider energy consumption indirectly by not using unnecessary flooding but do not care about an

individual node’s energy situation. MMSPEED could use its redundant path selection for load balancing, not only for reliability enhancement, and thereby improve the overall network lifetime.
In general, global knowledge means high control overhead, especially in large scale WSNs. Nevertheless, the authors of SAR prefer a table-driven routing approach that requires each node to continuously maintain a global routing table. They argue that the energy consumption at route setup in an on-demand approach exceeds the overall cost for routing table maintenance and is therefore inferior from an energy point of view. The other solutions argue for an on-demand approach.
Transmitting data from the sensing nodes to a sink node is the most frequent case. Nevertheless, a sink might want to query a specific part of the network for data. This transmission direction is not supported by the proposed routing protocols. Considering energy consumption, it is advantageous if requests for data (queries) and the requested sensor data would not have to travel all the way through the network, but could “meet halfway” at intermediate nodes that can match incoming sensor data to queries stored in memory. It would be interesting to look at this publish/subscribe approach from a timing point of view.
Service differentiation is targeted in all protocols but one (SPEED), while mobility is moderately supported by all protocols but the one described by Akkaya et.al., where a static network is assumed. SAR and MMSPEED provide redundant paths that can be used to improve reliability.

Table 1: Comparison of routing protocols with QoS focus

	Energy-awareness			Knowledge		Routing approach		Transmission direction			QoS differentiation		Mobility			Multipath	
	low	mode-rate	high	glo-bal	local	table-driven	on demand	sensors-to-sink	sink-to-sensors	both	no	yes	low	mode-rate	high	no	yes
SAR		x		x		x		x				x		x			x
SPEED	x				x		x	x			x			x		x	
MMSPEED	x				x		x	x				x		x			x
Akkaya et.al.		x			x		x	x				x	x			x	

4 Medium Access Control in Wireless Sensor Networks

Medium access control resides in the Data Link Layer of the OSI stack and is responsible for assisting the nodes to decide who is allowed to send its traffic when. The way in which a MAC protocol assigns channels plays an important role in saving energy and assuring certain QoS levels. Due to some of the special requirements on WSNs (listed above), traditional MAC protocols are not suitable without modifications [Ye].

4.1 Contention-based vs. contention-free MAC protocols

In general, there are two groups of MAC mechanisms, contention-free (scheduled) and contention-based. Contention-free MAC protocols divide the channel into time slots and assign cyclically repeating slots to one single node. Each node knows therefore in advance when it is allowed to send and when it can expect traffic from other nodes and can maximize its time in the energy saving sleep mode. Thereby, even the problem of overhearing is avoided [Ye]. QoS handling is easier as throughput and latency can be calculated in advance based on the channel access schedules. This convenience is of course costly in terms of scalability and adaptability to traffic bursts or topology changes. There is the need for central nodes determining the schedule and nodes need to be tightly synchronized in order to have a common conception of time. This introduces control overhead.

In the contention-based protocols, the nodes compete for a common channel and access is granted on demand. Collisions are frequent and usually resolved by backing of and retransmitting. There are clear advantages in WSN environments. No central node for scheduling is required, peer-to-peer communication is supported and resources are allocated on demand. All this makes the network scale well and adapt to changes in topology and traffic load. There is no need for time-synchronization either. On the downside, idle listening to the channel and retransmissions waste energy and there are no guarantees that a certain level of throughput can be given.

Which type to prefer for a WSN, is application dependent. Surveys on MAC protocols for WSNs can be found in [Ye], [Akyildiz], and [Demirkol].

4.2 MAC protocols with focus on energy-efficiency

4.2.1 Contention-based protocols

- ***Sensor-MAC (S-MAC) [Ye2]***
  S-MAC is a protocol specially designed for WSNs that circumvents the problems of idle listening, collision and overhearing but still keeps the flexibility of a contention-based scheme. Each node has changes between sleep and listening mode periodically. How long these periods are is determined by a schedule that is negotiated between neighboring nodes during a short schedule synchronization phase between each sleep and listening cycle. Neighbors broadcast their schedule to

each other, build virtual clusters, agree on a common schedule and use it to wake up at approximately the same time to communicate. Time synchronization does not have to be as tight as in TDMA protocols as the sleep and listening cycles are much longer than the usual TDMA time slot. During their active phase, nodes content for the medium in ordinary contention-based fashion.

- ***Sparse Topology and Energy Management (STEM) [Schurgers]***
  STEM was developed for event triggered networks. Two channels are needed, one data channel, which is always asleep and one wake up channel, which wakes up periodically and listens for an indication that someone wants to communicate. This indication may be a busy-tone (STEM-T) on the wake-up channel or periodic beacons (STEM-B) sent out by communication nodes.

- Other contention-based MAC protocols that address the energy issue have been proposed, amongst other the ***Piconet*** [Bennett] (not to be confused with the piconet concept in Bluetooth) or ***PAMAS*** (Power-aware Multiaccess with Signaling) [Singh]. Piconet avoids unnecessary listening by letting each node sleep an individual amount of time and beacon their ID each time they wake up. Another node wishing to communicate has to listen to the channel until it hears the beacon and can begin to send data. Even PAMAS uses sleep periods and introduces a separate control channel. After waking up, a node probes the control channel to find out about ongoing transmissions blocking the channel and their duration. If there aren’t any it starts to send, in case the channel is occupied it goes back to sleep.

4.2.2 Contention-free protocols

- ***Power Aware Cluster TDMA (PACT) [Pei]***
  PACT uses clustering to take advantage of a dense topology to prolong both battery time and network lifetime. The energy reserves of the individual nodes are taken into consideration and nodes with the biggest reserves are chosen as cluster heads and gateways between clusters. As soon as the ratio of energy reserves change, new cluster heads are chosen. Time slots are allocated to cluster heads and gateway nodes until their demand is satisfied; the remaining slots are distributed amongst the other nodes. In that way energy levels are kept even throughout the network and the total network lifetime is maximized. PACT even puts nodes to sleep in times of low traffic intensity.

- ***Distributed Energy aware MAC (DE-MAC) [Kalidindi]***
  Kalidindi et al. introduce the DE-MAC protocol and approach energy consumption from a network perspective. They optimize energy savings by treating sensor nodes individually depending on their energy status. A node with an energy level under a certain threshold can initiate an election process where time slots are allocated depending on energy levels – the lower the energy remains, the more time slots. There are predetermined sleep phases, but this election scheme reduces the time a node with low energy reserves has to listen to the channel for incoming messages.

4.3 MAC protocols with focus on QoS

Very little research deals with the QoS aspect of MAC protocols in WSNs. Adapting the work done on QoS MAC for wireless ad-hoc networks usually fails from an energy point of view. The need to combine energy-efficient and QoS-aware medium access has recently lead to some WSN-specific solutions. An overview is given below.

4.3.1 Protocol overview

- ***QoS-aware Medium Access Control (Q-MAC) [Yang]***
  The Q-MAC protocol handles service differentiation by using different priority queues in each node. Packets are assigned different priorities based on their content, the number of hops it already has traveled in the network and an energy metric. The access of different nodes to the common medium is coordinated by a RTS/CTS (ready to send/clear to send) phase, followed by a data transmission phase. Time synchronization is needed for all the nodes to follow the same time scheme.

- ***EDF-based Access Protocol [Caccamo]***
  The EDF-based MAC protocol proposed by Caccamo et al., assumes a cellular network structure with one router node per cell that has enhanced capabilities and can be reached in one hop by the ordinary sensor nodes in the cell. Periodic traffic with bounded delay is another assumption made by the authors. Messages are categorized depending on their destination; messages within the same cell (intra-cell messages) are scheduled by Earliest Deadline First (EDF), using TDMA (time division multiple access), messages to any other cell (inter-cell messages) are sent via the router nodes in a hop-by-hop fashion through the network using FDMA (frequency division multiple access). Time slots are reserved for inter and intra-cell communication according to a predefined pattern. Network utilization is increased by using FRASH, a frame sharing technique, where frames unused by the scheduled messages can instead be claimed by other nodes. EDF scheduling removes the overhead introduced by control traffic which is advantageous from an energy point of view. A simulation shows a lower latency and a higher throughput than comparable TDMA or contention based protocols in the case of high traffic and a dense network topology.

- ***Traffic-Adaptive Medium Access Protocol (TRAMA) [Rajendran]***
  TRAMA is a traffic-adaptive MAC protocol which aims to achieve efficient energy management by avoiding receiver collisions at the packet level through TDMA scheduling, and by employing a low power mode for nodes which are not scheduled to send or receive. The usage of this low power mode is dynamically determined and adapted according to the current traffic pattern. Schedules are exchanged in a one-hop neighborhood in a contention-based phase while the actual data transmission phase is contention-free. Each node is informed about the schedule for the next time slot, making it possible for idle nodes to enter their low-power mode. Simulation results prove an increase in throughput for TRAMA compared to well-known contention-based protocols, mostly due to the collision avoidance mechanism. However, the authors point out that the poor delay behavior makes this protocol unfeasible for sensor networks that are delay sensitive.

4.3.2 Evaluation of the protocols

The protocols are compared in Table 2.
A TDMA (Time Division Multiple Access) scheme is advantageous for minimizing a node's duty cycle but requires tight synchronization and periodic (or at least predetermined) traffic pattern. When reporting spontaneous events, a WSN can cause bursts of data which can not be accounted for in reservation-based schemes. This fact speaks against the TRAMA and MAC with EDF protocols. Contention-based protocols, on the other hand cannot give any real-time guarantees. All solutions but TRAMA offer different QoS levels and maintain different priority queues. This means that urgent packets are more likely to meet their deadlines and the need of retransmissions is reduced.
Without addressing energy efficiency, a MAC protocol for WSNs will most probably not meet the common requirements. Therefore, protocols that avoid control overhead and allow nodes to safely enter a sleep state, are to be preferred. TRAMA offers a sleep state and has moderate control and synchronization traffic as long as the network density is low. At high densities, maintaining information about the 2-hop neighborhood becomes a burden.

Table 2: Comparison of MAC protocols with QoS focus

	Energy-awareness			MAC approach		QoS differentiation		Synchronization			Control packets	
	low	moderate	high	reservation-based	contention-based	no	yes	network wide	locally	not needed	needed	not needed
QMAC		x			x		x	x			x	
TRAMA			x	x		x			x		x	
EDF-based		x		x			x	x				x

5 Conclusion

Research on MAC and (especially) routing in WSNs is extensive and the variety of approaches and solutions reflects the variety of application areas for WSNs. Despite the approached discussed above, the area of QoS in WSNs is still a rather unexplored field. The main question is how to find the best middle way between meeting QoS requirements and using the least possible amount of energy. Sensor networks might differ hugely in terms of size, density, communication frequency and type, demands on timeliness, energy efficiency, reliability, mobility, scalability etc. General solutions are probably impossible to find and protocols adapted to very specific application scenarios have the best chances to provide the best solutions.

References

[Adlakha]
S. Adlakha, S. Ganeriwal, C. Schurgers and M.B. Srivastava, *Density, Accuracy, Delay and Lifetime Tradeoffs in Wireless Sensor Networks – A Multidimensional Design Perspective*, Proceedings of the ACM International Conference on Sensor Systems 2003 (SenSys '03), Los Angeles, CA, USA, Nov. 2003.

[Akkaya]
K. Akkaya and M. Younis, *An energy-aware QoS Routing Protocol for Wireless Sensor Networks*, Proceedings of the 23rd International Conference on Distributed Computing Systems 2003, pp. 710-715, May 2003.

[Akyildiz]
I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, *Wireless Sensor Network: A Survey,* Computer Networks, vol. 38, issue 4, pp. 393-422, March 2002.

[Al-Karaki]
J.N. Al-Karaki and A.E. Kamal, *Routing Techniques in Wireless Sensor Networks: A Survey*, IEEE Journal on Wireless Communications, vol. 11, issue 6, pp. 6-28, Dec. 2004.

S. Basagni, I. Chlamtac, V.R. Syrotiuk, B.A. Woodward, *A Distance Routing Effect Algorithm for Mobility (DREAM),* Proceedings of the 4th ACM/IEEE International Conference on Mobile Computing and Networking MOBICOM '98, pp. 76-84, Dallas, TX, USA, 1998.

[Bennett]
F. Bennett, D. Clarke, J.B. Evans, A. Hopper, A. Jones and D. Leask, *Piconet: Embedded Mobile Networking, IEEE Personal Communications Magazine, vol.4, no. 5, pp. 8-15, Oct. 1997.*

[Caccamo]
M. Caccamo, L.Y. Zhang, S. Lui and G. Butazzo, *An Implicit Prioritized Access Protocol for Wireless Sensor Networks*, Proceedings of the 23rd IEEE Real-Time Systems Symposium 2002 (RTSS '02), pp. 39-48, Dec. 2002.

[Chen]
D. Chen and P.K. Varshney, *QoS Support in Wireless Sensor Networks: A Survey*, Proceedings of the International Conference on Wireless Networks 2004 (ICWN '04), Las Vegas, Nevada, USA, June 2004.

[Demirkol]
I. Demirkol, C. Ersoy and F. Alagoz, *MAC Protocols for Wireless Sensor Networks: A Survey*, IEEE Communications Magazine, vol. 44, issue 4, pp. 115-121, April 2006.

[Felemban]
E. Felemban, C.-G. Lee, E. Ekici, R. Boder and S. Vural, *Probabilistic QoS Guarantee in Reliability and Timeliness Domains in Wireless Sensor Networks*, Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Society (INFOCOM 2005), vol. 4, pp. 2646-2657, March 2005.

[Ganesan]
D. Ganesan, A. Cerpa, W. Ye, Y. Yu, J. Zhao and D. Estrin, *Networking Issues in Wireless Sensor Networks,* Journal on Parallel and Distributed Computing, vol. 64, 2004.

[Heinzelman]
W. Heinzelman, A. Chandrakasan and H. Balakrishnan, *Energy-Efficient Communication Protocol for Wireless Microsensor Networks,* Proceeding of the 33rd Hawaiian International Conference of System Sciences, Jan. 2000.

[Intanagonwiwat]
C. Intanagonwiwat, R. Govindan and D. Estrin, *Directed Diffusion: a Scalable and Robust Communication Paradigm for Sensor Networks,* Proceedings of the ACM MobyCom 2000, pp. 56-67, Boston, MA, USA, 2000.

[Kalidindi]
R. Kalidindi, L. Ray, R. Kannan and S.S. Iyengar, *Distributed Energy Aware MAC Layer Protocol for Wireless Sensor Networks,* International Conference on Wireless Networks ICWN '03, Las Vegas, NV, USA, June 2003.

[Karp]
B. Karp and H.T. Kung, *Greedy Perimeter Stateless Routing for Wireless Networks,* Proceedings of the 6th ACM International Conference on Mobile Computing and Networking, MOBICOM '00, pp. 243-254, Boston, MA, USA, Aug. 2000.

[Ko]
Y.B. Ko and N.H Vaidya, *Geocasting in Mobile Ad Hoc Networks: Location-Based Multicast Algorithms,* Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications WMCSA99, pp. 101-110, New Orleans, LA, USA, Feb. 1999.

[Li]
J. Li, J. Jannotti, D.S.J De Couto, D.R.Karger and R. Morris, *A Scalable Location Service for Geographic Ad Hoc Routing,* Proceedings of the 6th ACM International Conference on Mobile Computing and Networking, MOBICOM '00, pp. 120-130, Boston, MA, USA, Aug. 2000.

[Lindsey]
S. Lindsey and C. Raghavendra, *PEGASIS: Power-Efficient Gathering in Sensor Information Systems,* Proceedings of the IEEE Aerospace Conference, vol. 3, pp. 1125-1130, Big Sky, MT, USA, March 2002.

[Liao]
W.H. Liao, Y.C. Tseng, K.L Lo and J.P. Sheu, *GeoGRID: A Geocasting Protocol for Mobile Ad Hoc Networks based on GRID,* Journal of Internet Technology, vol. 1, no. 2, pp. 196-213, Dec. 2002.

[Lu]
C. Lu, B.M. Blum, T.F. Abdelzaher, J.A. Stankovic and T. He, *RAP: A Real-Time Communication Architecture for Large-Scale Wireless Sensor Networks*, Proceedings of the 8th IEEE Real-Time and Embedded Technology and Applications Symposium 2002, pp. 55-66, Sept. 2002.

[Manjeshwar]
A. Manjeshwar and D.P. Agrawal, *TEEN: a Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks,* Proceedings of the 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, April 2001.

[Manjeshwar2]
A. Manjeshwar and D.P. Agrawal, *APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks,* Proceedings of the International Parallel and Distributed Processing Symposium, pp. 195-202.

[Mauve]
M. Mauve, A. Widmer, H. Hartenstein, *A Survey on Position-Based Routing in Mobile Ad Hoc Networks,* IEEE Network, vol. 15, issue 6, pp. 30-39, Nov.-Dec. 2001.

[Maihöfer]
C. Maihöfer, *A Survey of Geocast Routing Protocols,* IEEE Communications Surveys & Tutorials, vol. 6, no. 2, pp. 32-42, 2004.

[Pei]
G. Pei and C. Chien, *Low Power TDMA in Large Wireless Sensor Networks,* IEEE Military Communications Conference MILCOM '01, vol. 1, pp. 347-351, Okt. 2001.

[Rajendran]
V. Rajendran, K. Obraczka and J.J. Garcia-Luna-Aceves, *Energy-Efficient, Collision-Free Medium Access Control for Wireless Sensor Networks*, ACM Journal on Wireless Networks, vol. 12, issue 1, pp. 63-78, Feb. 2006.

[Schurgers]
C. Schurgers, W. Tsiatsis, S. Ganerival and M. Srivastava, *Optimizing Sensor Networks in the Energy-Latency-Density Design Space,* IEEE Transactions on Mobile Computing, vol.1, issue 1, pp. 70-80, 2002.

[Shah]
R.C. Shah and J Rabaey, *Energy-Aware Routing for Low Energy Ad Hoc Sensor Networks,* Proceedings of the IEEE Wireless Communications and Networking Conference WCNC 2002, vol. 1, pp. 350-355, Orlando, FL, USA, March 2002.

[Singh]
S. Singh and C.S. Raghavendra, *PAMAS: Power Aware, Multi-Access Protocol with Signalling for Ad Hoc Networks,* ACM Computer Communication Review, vol. 28, no. 3, pp. 5-26, July 1998.

[Sohrabi]
K. Sohrabi, J. Gao, V. Ailawadhi and G.J. Pottie, *Protocols for Self-Organization of a Wireless Sensor Network*, IEEE Journal of Personal Communications, vol. 7, issue 5, pp. 16-27, Oct. 2000.

[Tian]
H. Tian, J.A. Stankovic, L. Chenyang, T. Abdelzaher, *SPEED: A Stateless Protocol for Real-Time Communication in Sensor Networks*, Proceedings of the 23rd International Conference on Distributed Computing Systems 2003, pp. 46-55, May 2003.

[Wang]
Y. Wang, X. Liu and J. Yin, *Requirements of Quality of Service in Wireless Sensor Networks*, Proceeding of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies 2006 (ICN/ICONS/MCL '06), April 2006.

[Yang]
L. Yang, I. Elhanany and Q. Hairong, *An Energy-Efficient QoS-Aware Media Access Protocol for Wireless Sensor Networks*, Proceedings of the International Conference on Mobile Adhoc and Sensor Networks 2005, Nov. 2005.

[Ye]
W. Ye and J. Heidemann, *Medium Access Control in Wireless Sensor Networks,* USC/ISI Technical Report ISI-TR-580, Okt. 2003.
(Available at: http://www.isi.edu/~johnh/PAPERS/Ye03c.html)

[Ye2]
W. Ye, J. Heidemann and D. Estrin, *An Energy-Efficient MAC Protocol for Wireless Sensor Networks,* Proceedings of the IEEE Infocom, pp. 1567-1576, New York, NY, USA, June 2002.

[Younis]
M. Younis, K. Akkaya, M. Eltoweissy and A. Wadaa, *On Handling QoS Traffic in Wireless Sensor Networks*, Proceedings of the 37th Annual Hawaii International Conference on System Sciences 2004, Jan. 2004.