Data Fusion and Topology Control in Wireless Sensor Networks

VRINDA GUPTA and RAJOO PANDEY
Department of Electronics and Communication Engineering
National Institute of Technology (Deemed University)
Kurukshetra. 136119. Haryana.
INDIA.
vrindag16@yahoo.com and rajoo_ pandey @ rediffmail.com

Abstract: - The design of large-scale sensor networks interconnecting various sensor nodes has spurred a great deal of interest due to its wide variety of applications. Data fusion is a critical step in designing a wireless sensor network as it handles data acquired by sensory devices. Wireless sensor networks allow distributed sensing and signal processing while collaborating during energy efficient operations. Wireless sensor networks are battery powered; therefore prolonging the network lifetime through an energy aware node organization is highly desirable. The main goal of a topology control scheme in wireless sensor networks is to reduce power consumption in order to extend network lifetime. Our aim is to provide a better understanding of the current research issues in this field. The paper provides a more detailed look at some existing data fusion and topology management algorithms. The most important design issues of data fusion and topology control are also highlighted.

Key-words:- Wireless sensor networks, data aggregation, data fusion, topology control, protocols.

1 Introduction

Recent advances in micro-fabrication and wireless communication technologies have spurred a great deal of interest in the use of large-scale wireless sensor networks. Research and commercial interest in the area of wireless sensor networks are currently growing exponentially, which is manifested in the number of web pages (Google: 1,180,000 hits for sensor networks; 3,110,000 for wireless sensor networks in April 2008).

A wireless sensor network (WSN) is a special type of ad hoc network, which consists of a large number of sensor nodes that may be randomly and densely deployed. These tiny sensor nodes consist of data processing and communicating components in addition to sensing component. A WSN may be designed with different objectives. It may be designed to gather and process data from the environment in order to have a better understanding of the behavior of the monitored entity. It may also be designed to monitor an environment for the occurrence of a set of possible events, so that the proper action may be taken whenever necessary. These features ensure a wide range of applications for sensor networks. Some of the applications areas are environmental monitoring, industrial and manufacturing automation, health-care, and military.

Dust Inc., Berkeley, CA, (Smart Dust research project) [1] at the University of California, Berkeley is building MEMS sensors that can sense and communicate and yet are tiny enough to fit inside a cubic millimeter. A Smart Dust optical mote uses MEMS to aim sub millimeter-sized mirrors for communications. A wireless network of these ubiquitous, low-cost, disposable micro sensors can provide close in sensing capabilities in many novel applications.

1.1 Applications

1.1.1 Agriculture and Environmental Monitoring Center for Embedded Network Sensing (CENS), LA, California, has a focus on environmental and habitat monitoring. On a very large scale, the system for the vigilance of the Amazon (SIVAM) provides environmental monitoring, drug trafficking monitoring and air traffic control for the Amazon Basin. Sponsored by the government of Brazil, this large sensor network consists of different types of interconnected sensors including radar, imagery, and environmental sensors. The imagery sensors are space based; radars are located on aircraft, and environmental sensors mostly on the ground [2].

- *Precision agriculture*: Crop and livestock management and precise control of fertilizer concentrations are possible. Precision farming requires analysis of spatial data to determine crop response to varying properties such as soil type. The ability to embed sensor nodes in a field at strategic locations could give farmers detailed soil analysis to help maximize crop yield or possibly alert them when soil and crop conditions attain a predefined threshold.

- *Planetary exploration*: Exploration and surveillance in inhospitable environments such as remote geographic regions or toxic locations can take place.
- *Geophysical monitoring*: Seismic activity can be detected at a much finer scale using a network of sensors equipped with accelerometers.
- *Monitoring of freshwater quality*: The field of hydrochemistry has a compelling need for sensor networks because of the complex spatio-temporal variability in hydrologic, chemical, and ecological parameters and the difficulty of labor-intensive sampling, particularly in remote locations or under adverse conditions.
- *Zebra net*: The Zebra net project at Princeton aims at tracking the movement of Zebras in Africa.
- *Habitat monitoring*: Researchers at UC Berkeley and the college of the Atlantic in Bar Harbor deployed sensors on Great Duck Island in Maine to measure humidity, pressure, temperature, infrared radiation, total solar radiation, and photo synthetically active radiation. The primary purpose of the sensor network was to monitor the microclimates in and around nesting burrows used by the Leach's Storm Petrel. Thus, researchers could take multiple measurements of biological parameters at frequent interval with minimal disturbance to the breeding colony. Monitoring of the birds can then proceed without direct human contact. Similarly, the PODS project [3] at the University of Hawaii uses WSNs to observe the growth of endangered species of plants. In this particular WSN application, two types of data are collected: weather data, which are collected every 10 minutes, and high-resolution images, which are collected every hour.
- *Disaster detection*: Forest fire and floods can be detected early and causes can be localized precisely by densely deployed sensor networks. Disaster relief efforts such as the ALERT flood – detection system makes use of remote field sensors to relay information to a central computer system in real time. Typically, an ALERT installation comprises several types of sensors, such as rainfall sensors, water-level sensors, and other weather sensors. Data from each set of sensors are gathered and relayed to a central base station.

Wang and Meng [4] have proposed a wireless sensor network paradigm for real-time forest fire detection. The forest fire is a fatal threat in the world: it is reported that a total of 77,534 wildfires burned 6,790,692 acres in USA for 2004. The wireless sensor network can detect and forecast forest fire more promptly than the traditional satellite-based detection approach.

1.1.2 Civil Engineering

- *Monitoring of structures*: Sensors will be placed in bridges to detect and warn of structural weakness and in water reservoirs to spot hazardous materials. The reaction of tall buildings to wind and earthquakes can be studied and material fatigue can be monitored easily.

- *Urban planning*: Urban planners will track ground water patterns and how much CO_2 cities are expelling, enabling them to make better land-use decisions.
- *Disaster recovery*: Buildings razed by an earthquake may be infiltrated with sensor robots to locate signs of life.

1.1.3 Military Applications

As with many technologies, defense applications have been a driver for research and development in sensor networks. During the cold war, the sound surveillance system (SOSUS), a system of acoustic sensors (hydrophones) on the ocean bottom, was deployed at strategic locations to detect and track quiet Soviet submarines. Over the years, other more sophisticated acoustic networks have been developed for submarine surveillance. SOSUS is now used by the National Oceanographic and Atmospheric Administration (NOAA) for monitoring events in the ocean, e.g., seismic and animal activity. Also during the cold war, networks of air defense radars were developed and deployed to defend the continental US and Canada. This air defense system has evolved over the years to include aerostats as sensors and Airborne Warning and Control System (AWACS) planes [2].

- *Asset monitoring and management*: Commanders can monitor the status and location of troops, weapons, and supplies to improve military command, control, communications and computing (C4).
- *Surveillance and battle-space monitoring*: Vibration and magnetic sensors can report vehicle and personnel movement, permitting close surveillance of opposing forces. Sensor nodes can be programmed to send notifications whenever movement through a particular region is detected.
- *Urban warfare*: Sensors are deployed in buildings that have been cleared to prevent reoccupation; movements of friend and foe are displayed in PDA-like devices carried by soldiers.

In chemical and biological warfare, close proximity to ground zero is needed for timely and accurate detection of the agents involved. Sensor networks deployed in friendly regions can be used as early-warning systems to raise an alert whenever the presence of toxic substances is detected. Deployment in an area attacked by chemical or biological weapons can provide detailed analysis, such as concentration levels of the agents involved, without the risk of human exposure.

An example of network-centric warfare is the Cooperative Engagement Capability (CEC) developed by the U.S. Navy. This system consists of multiple radars collecting data on air targets. Other military sensor networks include acoustic sensor arrays for antisubmarine warfare such as the Fixed Distributed System (FDS) and the Advanced Deployable System (ADS), and unattended ground sensors (UGS) such as the Remote Battlefield Sensor System (REMBASS) and the Tactical Remote Sensor System (TRSS) [2].

1.1.4 Health monitoring and surgery

The Smart Sensors and Integrated Microsystems (SSIM) project at Wayne State University and the Kresge Eye Institute are working on developing an artificial retina. One of the project goals is to build a chronically implanted artificial retina that allows a visually impaired individual to see at an acceptable level. Currently, smart sensor chips equipped with 100 micro sensors exist. The smart sensor comprises an IC with transmit and receive capabilities and an array of sensors.

- *Medical Sensing*: Physiological data such as body temperature, blood pressure and pulse are sensed and automatically transmitted to a computer or physician, where it can be used for health status monitoring and medical exploration [5].

Glucose-level monitoring is a potential application suitable for WSNs. Individuals with diabetes requires constant monitoring of blood sugar levels to lead healthy, productive lives. Embedding a glucose meter within a patient with diabetes could allow the patient to monitor trends in blood-sugar levels and also alert the patient whenever a sharp change in blood-sugar levels is detected.

1.1.5 Commercial applications

CSIRO [6] is using wireless sensor network technology for “tiny agents”, deployed as autonomous controllers for individual pieces of electrical load/generation equipment in a distributed energy system including heating, ventilation, and air-conditioning (HVAC) systems. Smart Dust [1] project is exploring innovative methods of interacting with the environment, providing more information from more places less intrusively. Smart Dust, a technology developed at UC Berkeley will enable a rich collection of diverse applications such as

- building virtual keyboards;
- managing inventory control;
- monitoring product quality;
- constructing smart-office spaces; and
- providing interfaces for the disabled.

Wireless sensor networks produce a large amount of data that needs to be processed according to the application objectives. The way these data are manipulated by the sensor nodes is a fundamental issue. Although many protocols and algorithms have been proposed for wireless and ad hoc networks, they are not well suited for the unique features and application requirements for sensor networks. Alternative approaches are required. These are due to following reasons: -

- Since sensor nodes are randomly deployed, so they do not fit into any regular topology. Once deployed, they usually don’t require any human intervention. Therefore, all routing and maintenance algorithms need to be distributed.
- Also, due to the relatively large number of sensor nodes, it is not possible to build a global addressing scheme for the deployment of a large number of sensor nodes, as the overhead of ID maintenance is high. Thus traditional IP-based protocols may not be applied to WSNs.
- Sensors rely on battery for power, which is difficult to be replaced or recharged. As sensor nodes being tightly constrained in energy, processing and storage capabilities, energy efficient protocols should be designed. So, sensor networks require careful resource management.
- Sensor networks are dense, neighbor nodes may be very close to each other. Hence, multihop communication is expected to consume less power than the traditional single hop communication.
- Almost all applications of sensor networks require the flow of sensed data from multiple sources to a particular Base Station.
- Sensor networks are application-specific. The design requirements of a sensor network change with application.
- Position awareness of sensor nodes is important, since data collection is normally based on the location.
- WSNs are data centric networks, as data is requested based on certain attributes.
- Data collected by many sensors in WSNs is typically based on common phenomenon, so there is a high probability that this data has some redundancy. Such redundancy needs to be exploited.

As these sensor nodes are typically constrained in energy and communication bandwidth, it is desirable to minimize the number of messages relayed because radio transmissions can quickly consume battery power. A reduction in communication and energy costs is possible if collected sensor data is aggregated prior to relaying. In this context, data fusion arises as a discipline that is concerned with how data gathered by sensors can be processed to increase the relevance of such a mass of data. Data fusion can be defined as a process of combining data or information to estimate or predict entity states.

Collaborative signal processing algorithms are another enabling technology for WSNs. While sensor nodes collect raw data from the environment, only useful information is of importance. Hence raw data need to be properly processed locally at sensing node, and only processed data is sent back to the end users. Since computation is much more energy efficient than wireless communication, this avoids wasting energy on sending large volumes of raw data. Such signal processing is often required to be performed by a set of sensor nodes in proximity, due to the weak sensing and processing capabilities of each individual node.

In practice, data fusion operation has been incorporated into a wide range of existing wireless sensor network designs. For example, simple aggregation techniques (e.g., maximum, minimum, and average) have been used to reduce the overall data traffic to save energy [7][8][9]. Additionally, data fusion techniques have been applied to WSNs to improve location estimates of sensor nodes [10], detect routing failures [11], and collect link statistics for routing protocols [12].

A number of physical layer parameters have found their role in MAC and routing. Among these parameters, channel state and residual energy are perhaps the most relevant to the efficiency of sensor networks. The benefit of exploiting CSI and REI has been recognized in [13]. Using channel state information (CSI) in transmission and networking is the fundamental idea behind opportunistic strategies. Also, the residual energy information (REI) of individual nodes plays a crucial role in network lifetime. Various sensor placement schemes, routing, and transmission protocols that utilize REI have been proposed.

Sensor networks may exhibit a wide range of variations in traffic load and traffic pattern, from quiescent sensing state to emergency response. It is highly desirable to have traffic-adaptive MAC and routing that are reconfigurable based on network operating conditions. For example, at times when an emergency arises resulting in a rush of data toward certain parts of the network, routing protocols should be proactive, maintaining network connection to ensure rapid and energy efficient data delivery. When the network is in a quiescent sensing state, routing protocols should be reactive, establishing links and connections only when necessary. A fundamental challenge in achieving this traffic-adaptive routing is to develop signal-processing techniques for traffic estimation and change detection. Such signal processing techniques should be distributed to ensure scalability and avoid extra data flows.

For query processing in sensor networks, the Tiny DB [14] and Cougar projects [15] support various operations in an SQL-like language. These are further generalized in [16] to include median, consensus value, histogram of the data distribution, as well as range queries.

Another way of optimizing the energy consumption in sensor networks is by selectively switching off the radio of sensor nodes based on the availability of alternate routing paths. Switching off the radio of the sensor nodes is only possible if the topology is configured in such a way that the network is not partitioned due to those inactive nodes. Thus effectively controlling the topology of the network emerges as a solution to the problem of energy conservation for wireless sensor networks.

Given the importance of data fusion and topology control for WSNs, this work surveys the state-of-the-art related to data fusion and topology control and how it has been used in WSNs. This background is presented in the following structure. Section 2 raises the issue concerned with energy efficient requirements in WSNs. The main methods for data fusion and the common classification are presented in Section 3. Section 4 provides a detailed investigation of the current proposed topology management algorithms. Section 5 presents our final remarks with a look at directions for future research.

2 Energy Efficiency Requirements in WSN

Focus is on applications demanding higher peak power or longer lifetime in an environment where changing batteries is impractical or impossible. Therefore requiring a renewable energy source. Research into energy scavenging suggests that microsensors can utilize energy harvested from the environment. Energy harvesting schemes convert ambient energy into electrical energy, which is stored and utilized by the node. The most familiar sources of ambient energy include solar power, thermal gradients, radio-frequency (RF), and mechanical vibration. It is expected that 10s of microwatts of power to be harvested from ambient energy. Coupling energy harvesting techniques with some form of energy storage can theoretically extend microsensor node lifetimes indefinitely.

At the architectural level, designing for energy awareness can allow a sensor node to minimize energy consumption in the variable environment of a microsensor network. Having energy awareness in every aspect of design and operation can do this.

Fig. 1 shows the schematic diagram of sensor node components. A typical sensor network is generally composed of four components: power supply unit, a sensing unit, a computing / processing unit, and a communicating unit [17].

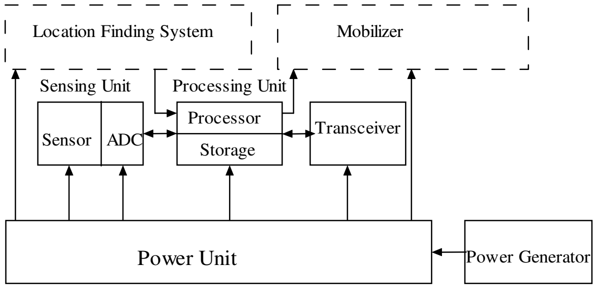

Fig.1 Components of a sensor node.

- *Computing Energy*

The computing/processing unit is a micro controller unit (MCU) or microprocessor with memory. The MCU is responsible for control of the sensors, and execution of communication protocols and signal processing algorithms on the gathered sensor data. Commonly used MCUs are Intel's Strong ARM SA-1100 microprocessor and Atmel's AVR micro controller. MCU's usually support various operating modes, including Active, Idle, and Sleep modes for power management purposes, with each mode characterized by a different amount of power consumption. But transitioning between these operating modes involves a power overhead and therefore transition costs should be included while considering the total energy consumption of the sensor node. A scheme of energy saving on computation is dynamic voltage scaling (DVS) explored in [18][19]. It adaptively adjusts operating voltage and frequency to meet the dynamically changing workload without degrading performance. Few papers have considered algorithmic transformations on multiple-voltage power minimization. The proposed approach [20] optimizes the power saving for DSP applications when the resources and the latency are constrained.

On-board ROM and RAM are included for storage of sampled and processed data, signal processing tasks, and the operating system. A simple energy model can be used to model the active energy dissipation of the SA-1100 as a function of supply voltage

$$E_{comp} = NC\, V_{dd}^{2} \quad (1)$$

Where N is the number of clock cycles per task, C is the average capacitance switched per cycle, and V_{dd} is the supply voltage [21].

- *Communicating Energy*

The communicating unit in a sensing node mainly consists of a short–range RF circuit that performs data transmission and reception. Radios can operate in four distinct modes of operation, namely Transmit, Receive, Idle, and Sleep modes. An important observation in the case of most radios is that, operating in idle mode results in significantly high power consumption, almost equal to the power consumed in the Receive mode [22]. Thus, it is important to completely shutdown the radio rather than transitioning to Idle mode, when it is not transmitting or receiving data. While power management of individual sensor nodes reduces energy consumption, it is important for the communication between nodes to be conducted in an energy efficient manner as well. The energy required for radio communication scales with distance as d^{n}, where d is the distance and n is the path loss exponent, which typically ranges between 2 and 4. Dividing a long-distance transmission into several shorter ones may reduce communication energy. Also, it is seen that as the distance to the end-user increases and as processor energy is much cheaper than communication energy, it becomes more energy efficient to perform signal processing locally at the sensor node.

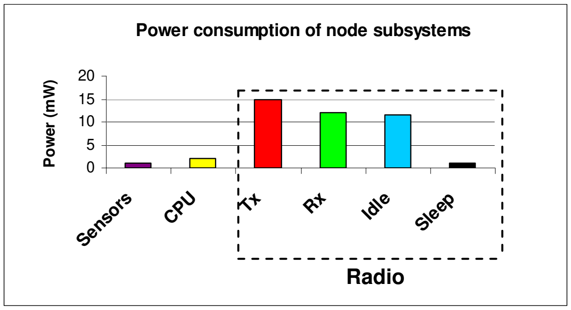

Fig.2 Power consumption of node subsystems

- *Sensing Energy*

The sensing unit in a sensor node includes the embedded sensor and / or actuator and the analog-digital converter. It links the node to the physical environment. Energy consumed for sensing includes (1) physical signal sampling and conversion to electrical signals (2) signal conditioning, and (3) analog to digital conversion. Sensing energy represents only a small percentage of the total power consumption in a WSN. It can be reduced by using low power hardware as well as by interval sensing.

An energy model for the communication sub-system has been developed to model the energy dissipated by a sensor node when transmitting and receiving data [21]. The radio module energy dissipation can be characterized into two types. The first is given by E_{elec} (J/b), the energy dissipated to run the transmit or receive electronics and the second is given by ε_{amp} (J/b/m^2), the energy dissipated by the transmit power amplifier to achieve an acceptable E_b / N_o at the receiver. To transmit a k-bit packet a distance, d, the energy dissipated is

$$E_{Tx}(k, d) = E_{elec} * k + \varepsilon_{amp} * k * d^{\lambda} \quad (2)$$

$$E_{Rx}(k) = E_{elec} * k \quad (3)$$

Where λ is the path loss exponent and $\lambda \geq 2$.

2.1 Network Lifetime

Sensor nodes typically run on batteries, which make effective power management a key challenge in operating system design. Optimizing energy consumption has been the focus of recent research in sensor networks. Power consumption of the node varies from 3.5mW in the deepest sleep state up to almost 2W (1.1 W of which goes into the transmitter power amplifier) with the processor running at the fastest clock rate and the radio transmitting at the highest power level.

Because battery technology is not improving very fast, primarily reducing power consumption, rather than increasing supply increases the lifetime of a sensor node. Large–scale increases in node lifetime are obtained by turning components off during times when they are not needed. For this reason, most microsensor networks duty cycle, or shut down unused components whenever possible. Here, duty cycling refers generically to alternating between an active mode and a low power sleep mode. Although duty cycling helps to extend sensor lifetimes, it does not remove the energy constraint placed by the battery. A 1cm^3 Lithium battery can continuously supply 10 μW of power for five years [23].

The “duty cycle” of a node is the fraction of the time that the node’s high power components are active, and may be on the order of 1%.

The lifetime of a sensor network is the average time span from the deployment to the instant when the network can no longer perform the task. The performance measure of network lifetime is particularly relevant to sensor networks where battery-powered, dispensable sensors are deployed to collectively perform a certain task. Much has been said about maximizing network lifetime. It is seen both CSI and REI are critical to maximizing network lifetime. The role of REI is to balance the energy consumption across the network by prioritizing nodes with more residual energy energy-consuming tasks such as transmission.

There is a simple law that governs the network lifetime for all applications, under any network configuration. It is shown in [24], that the network lifetime α defined as the average span from the deployment to the instant when the network is considered dead is given by

$$\alpha = \frac{\varepsilon_0 - \varepsilon_w}{\lambda \varepsilon_r} \qquad (4)$$

where ε_0 is the total energy over the network, ε_w is the expected wasted energy, λ the average sensor reporting rate, defined as the number of data collections per unit time, and ε_r the expected reporting energy consumed by all sensors in a randomly chosen data collection.

Specifically, a lifetime-maximizing protocol should aim at reducing the average wasted energy ε_w and the average reporting energy ε_r. To reduce ε_w, the protocol should exploit residual energy information (REI) of individual sensors to achieve balanced energy consumption across the network. To reduce ε_r, the protocol should exploit channel state information (CSI) to prioritize sensors with better channels for transmission so that energy consumed in transmission is reduced. Since channel realizations are independent of the residual energies, the sensor with the best channel may not have the most residual energy. A lifetime-maximizing protocol needs to optimally trade off CSI and REI.

3 Data Fusion in WSN

Data fusion has received attentions for both military and commercial applications over the past two decades. Data fusion is an important topic for collaborative signal processing. Since sensor readings are usually imprecise due to strong variations of the monitored entity or interference from the environment, data fusion can be used to process data from multiple sensors in order to filter noise measurements and provide more accurate interpretations of the information generated by a large number of sensor nodes. A rich set of techniques is applicable in this context, including Kalman filtering, Bayesian interference, neural networks, and fuzzy logic. For Wsns, data can be fused with at least two objectives: accuracy improvement and energy saving. The definition of data fusion was given as, “a process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from multiple sources”[25]. The techniques use the observations of events from multiple sensors as its input, and integrate the information to achieve improved accuracies and more specific inferences than could be achieved by the use of a single source alone [26]. For WSNs, data can be fused with at least two objectives: accuracy improvement and energy saving.

Data fusion can be categorized based on several aspects such as purpose, parameters, and type of data. According to the relationship among the sources, data fusion can be classified as complementary, redundant, or cooperative [27].

- *Complementary fusion* consists in fusing data from sensor nodes that describes the whole sensor field, and hence achieves broader information [28][29].

- *Redundant fusion* might be used to provide high quality information and prevent sensor nodes from transmitting redundant information.

- *Cooperative fusion* exists when information provided by two independent sources is fused into new information, usually more complex than original data e.g., computation of a target location based on angle and distance information.

Data fusion can be performed with different objectives such as inference, estimation, aggregation, and compression.

3.1 Inference

Inference methods are applied in decision fusion. A decision is taken based on the knowledge of the perceived situation. Decision–making paradigms include Bayesian decision-making, Dempster-Shafer Inference, fuzzy logic. Information fusion based on Bayesian Inference offers a formalism to combine evidence according to rules of probability theory. The uncertainty is represented in terms of conditional probabilities describing the belief, and it can assume values in the [0,1] interval, where 0 is absolute disbelief and 1 is absolute belief. Bayesian inference is based on Baye's rule, which states that:

$$P_r\left(\frac{Y}{X}\right)=\frac{P_r\left(\frac{X}{Y}\right)P_r(Y)}{P_r(X)} \tag{5}$$

where the posterior probability Pr (Y/X) represents the belief of hypothesis Y given the information X. This probability is obtained by multiplying Pr (Y), the prior probability of the hypothesis Y, by Pr (X/Y), the probability of receiving X, given that Y is true; Pr (X) can be treated as a normalizing constant. The main issue regarding Bayesian Inference is that the probabilities Pr (X) and Pr (X/Y) have to be estimated since they are unknown. Bayesian Inference has been used to solve the localization problem. Biswas et al. [30] model the sensor network as a Bayesian network. A distributed and localized Bayesian algorithm for detecting and correcting measurement faults has been developed in [31]. This work is further extended in [32] where both measurement errors and sensor faults in the detection task are considered.

Dempster-Shafer Inference is based on Theory of Evidence, which is a mathematical theory introduced by Dempster and Shafer. It generalizes the Bayesian theory and deals with beliefs. Dempster-Shafer theory can be used to fuse data provided by different types of sensors in contrast to inference with a Bayesian method. In [11], Topology Rebuilding Algorithm is proposed as an improvement to tree-based routing algorithms. It analyzes data traffic and uses the Dempster-Shafer inference to detect routing failures, and trigger a topology reconstruction only when necessary.

Fuzzy Logic generalizes probability and deals with approximate reasoning to draw conclusions. Each quantitative input is fuzzyfied by a membership function. The fuzzy rules of an inference system produce fuzzy outputs, which, in turn, are defuzzyfied by a set of output rules. Fuzzy reasoning is used in [33] for deciding the best cluster heads in a WSN. It considers three features: node concentration, energy level, and node centrality with respect to the entire cluster. To optimize energy usage in WSNs, fuzzy logic is also used for efficient routing [34]. This assumes a cluster-based architecture and study gateway centralized intercluster routing. Transmission energy, remaining energy, rate of energy consumption, queue size, distance from the gateway, and current status are considered as input variables, the cost is the fuzzy output. Another use a fuzzy system to infer the ability of each node to transmit data using its battery power and the type of data being forwarded; and during route discovery, the output of the fuzzy logic controller is used to decide whether or not to forward a packet [35]. Another data fusion algorithm based on fuzzy logic methods, Mamdani and Tsukamoto-Sugeno inference method is proposed [36]. Both methods are completed in four phases: fuzzification, rule evaluation, combination or aggregation of rules, and defuzzification. It is observed that the Mamdani method gives better results than the Tsukamoto approach and can be applied to WSNs to provide optimal data fusion and ensure maximum sensor lifetime and minimum time delay. In the data fusion domain, neural networks have also been used by classification and recognition tasks. A key feature of neural networks is the ability of learning from examples of input/output pairs in a supervised fashion. Neural networks have been applied to data fusion mainly for Automatic Target Recognition (ATR) using multiple complementary sensors [37].

3.2 Estimation

Estimation methods such as maximum likelihood, Least squares, Kalman filter are used in WSN.

Xiao et al. [38] propose robust, distributed and localized maximum likelihood estimation, where every node computes a local unbiased estimate that converges towards the global maximum likelihood solution. The authors in [39] further extended this method to support asynchronous and timely delivered measurements. Other distributed implementation of maximum likelihood estimators for WSNs include the decentralized Expectation Maximization (EM) algorithm [40] and the Local Maximum Likelihood Estimator [41] that relax the requirement of sharing all the data. MLE is used in the network tomography domain to estimate per-node loss rates during the aggregation and reporting of data from source to sink nodes. Such a strategy may be useful for routing algorithms to bypass lossy areas or when designing robust fault-tolerant protocols [42]. The MLE is commonly used to solve location discovery problems i.e. to obtain accurate distance or direction, angle estimations [43] [44].

Another estimation method known as Least Squares is applied in the WSN domain. A mathematical optimization technique, Least Squares method searches for a function that best fits a set of input measurements. Different Square – error metrics can be used such as the ordinary squared error, the Huber Loss function [45], and the root mean squared error [46]. In noisy environments, although the ordinary Least Squares algorithm quickly converges to the expected value, the variance is strongly affected. Therefore in such cases where noisy measurements might be frequent, Huber Loss function is more suitable [45]. Instead of transmitting the complete data stream from source to sink, a dual prediction scheme based on Least Squares filters is used both in the source and sink [47]. Only when the predicted value differs from the actual value by more than a given error, the value is transmitted to the sink. A robust and interactive Least Squares algorithm that explicitly considers noisy measurements is proposed for node localization in which, at each iteration, nodes are localized [48].

The Kalman filter is a very popular fusion method. The Kalman filter estimates the state x of a discrete-time controlled process that is ruled by the state-space model

$$x[k]) = A.x[k-1] + w[k-1] \qquad (6)$$

The system is influenced by process noise denoted w. The state dynamics determine the linear operator A. The state contributes to the observation y, which also includes a stochastic, additive measurement noise v:

$$y[k] = C.x[k] + v[k] \qquad (7)$$

The process and measurement noises are assumed to be Normal processes with known variances W and V.

Now assume that we have an estimate $\hat{x}[k-1]$ of the state, and also an estimate of the error co-variance P [k-1] in the estimate, at step k-1. The Kalman filter uses these estimates, the observation y [k] at sample k, and A, C, W and V to form an estimate of the state and its error covariance at step k:

$$\hat{x}[k] = K[k].(y[k] - \hat{y}[k]) \qquad (8)$$

$$\hat{P}[k] = (I - K[k].C).\tilde{P}[k].(I - K[k].C)^T \qquad (9)$$

where

$$\hat{y}[k] = C.A.\hat{x}[k-1] \qquad (10)$$

$$K[k] = \tilde{P}[k].C^T \Big/ \left(V + C.\tilde{P}[k]C^T\right) \qquad (11)$$

$$\tilde{P}[k] = A.\hat{P}[k-1].A^T + W \qquad (12)$$

The estimated system state x [k] is thus completely determined by the observation y [k], the estimated state at step k-1, the system dynamics, and the statistical properties of the process and measurement noise. The error in the estimate x [k] falls with k, converging upon a limiting error covariance that is fully determined by {A, C, W, V}. Correspondingly, we can choose any initial estimate of x and P and the filter will, after several iterations, adjust the state estimate and error accordingly.

Kalman filter is used in WSN in many schemes. One is in which solution is computed based on reaching an average consensus among sensor nodes [49]. Another is concerned with data loss due to the unreliable communication channels in WSNs [50]. Kalman filter has been applied to refine location and distance estimates [51], and track different sources [44]. A dual Kalman filter approach has also been proposed in which both source and sink nodes predict the sensed value so the source node sends data only when it knows the sink prediction is incorrect [52]. Author in [53] explores the effectiveness of the Kalman filter in monitoring with sensor networks. The aim is to produce an accurate spatial picture of a certain physical process, while making an efficient use of resources. It observes that the data fusion with feedback improves quality of monitoring in sensor-based networks.

To depict the geographical distribution of resources or activity of a WSN, Zhao et al. [28] implemented Network Scans known as eScan. These scans provide a summarized view of the resource distribution instead of providing detailed information of each sensor node. eScan retrieves information about the residual energy in the network in a distributed in-network fashion.

3.3 Aggregation

Data aggregation has been applied to eliminate redundancy in neighboring nodes [7] [54]. It applies a novel data-centric approach to replace the traditional address-centric approach in data forwarding [12]. When data are measured or arrive from a neighbor, the sensor needs to decide whether or not they are important enough to forward them. The coding techniques used need to minimize the number of forwarded bits. The new data may also be combined with other received data, in order to minimize the number of bits to forward.

Also, the information gathered by neighboring sensors is often redundant and highly correlated, and that the energy is much more constrained, necessitates the need for data fusion. Instead of transmitting all the data to a centralized node for processing, data are processed locally and a concise digest is forwarded to sinks. Data fusion reduces the number of packets to be transmitted among sensors, and thus the usage in bandwidth and energy. For a network with n sensors, the centralized approach takes O $(n^{3/2})$ bit-hops, while data fusion takes only O (n) bit–hops to transmit data [45]. There are two types of data aggregation: *"Snapshot aggregation"* is data fusion for a single event, such as tracking a target, while *"periodic aggregation"* periodically executes the data-fusion function, such as monitoring an environment parameter periodically [55].

Kulik et al. [56] define data aggregation as a technique used to overcome two problems: implosion and overlap. In the former, data sensed by one node is duplicated in the network due to data routing strategy (e.g., flooding). The overlap problem happens when two different nodes disseminate the same data. This might occur when the sensors are redundant – they sense the same property in the same place. In both cases, redundancy occurs and has negative impact i.e., waste of energy and bandwidth. The use of data aggregation in WSNs and its impact on energy consumption is the subject for further research.

Several data aggregation algorithms have been reported in the literature. The most straight forward is duplicate suppression i.e. if multiple sources send the same data; the intermediate node will only forward one of them [57]. It suppresses redundant data by discarding duplicates. Using a maximum or minimum function is also possible. Heinzelman et al. [57] and Kulik & colleagues [56] proposed SPIN to realize traffic reduction for information dissemination using metadata negotiations between sensors to avoid redundant and/ or unnecessary data propagation through the network. The greedy aggregation approach [58] can improve path sharing and attain significant energy savings when the network has higher node densities compared with the opportunistic approach. Krishnamachari and colleagues [8] described the impact of source-destination placement on the energy costs and delay associated with data aggregation. They also investigated the complexity of optimal data aggregation. Boulis et al. [55] discuss the trade off between energy consumption and accuracy when aggregation functions are used to summarize data from a WSN.

In-network data aggregation is a complex problem that involves many layers of the protocol stack and different aspects of protocol design, and a characterization and classification of concepts and algorithms is still lacking in the literature.

In-network aggregation deals with this distributed processing of data within the network. In-network data aggregation can be considered a relatively complex functionality, since the aggregation algorithms should be distributed in the network and therefore require coordination among nodes to achieve better performance. Also, the data size reduction through in-network processing shall not hide statistical information about the monitored event. In-network aggregation could be defined as follows:

In-network aggregation is the global process of gathering and routing information through a multi-hop network, processing data at intermediate nodes with the objective of reducing resource consumption thereby increasing lifetime.

Two approaches to in-network aggregation are:

- In-network aggregation with size reduction refers to the process of combining and compressing data coming from different sources in order to reduce the information to be sent over the network.
- In-network aggregation without size reduction refers to the process of merging packets coming from different sources into the same packet without data processing.

The first approach is better able to reduce the amount of data to be sent over the network but it may also reduce the accuracy with which the gathered information can be recovered at the sink. After the aggregation operation, it is usually not possible to perfectly reconstruct all of the original data.

The second approach, instead, preserves the original information i.e. at the sink; the original data can be perfectly reconstructed.

One of the most important functionalities that in-network aggregation techniques should provide is the ability to combine data coming from different nodes. There are several types of aggregation functions and most of them are closely related to the specific sensor application. Aggregation functions can compress and merge data according to either a lossy or a loss less approach. In the first case the original values cannot be recovered after having merged them by means of the aggregation function. In contrast, the second approach (loss less) allows compressing the data by preserving the original information. This means that all readings can be perfectly reconstructed from their aggregate at the receiver side.

The efficiency of these algorithms depends on the correlation among the data generated by different information sources (sensor units). Such a correlation can be spatial, when the values generated by close-by sensors are related, temporal, when the sensor readings change slowly over time, or semantic, when the contents of different data packets can be classified under the same semantic group (e.g., the data is generated by sensors placed in the room). The gains of in-network data aggregation can be best demonstrated in the case when data generated by different sources can be combined into a single packet (e.g., when the sources generate identical data). If there are K sources all close to each other and far away from the sink, the combination of their data into a single packet leads, on average, to a K-fold reduction in transmissions with respect to the case where all data are sent separately.

3.4 Compression

In WSNs, data can be compressed by exploiting spatial correlation among sensor nodes in a distributed fashion demanding no extra communication except the dissemination of the sensed data [59]. This is possible by considering that two neighbors provide correlated measurements.

Distributed Source Coding (DSC) effectively makes routing and coding decisions independent of each other. On the downside, however, this solution increases the computational complexity and requires the collection of information about joint statistics, which may not always be easy in practice.

Distributed Source Coding (DSC) [60] refers to the compression of multiple correlated sources, physically separated, that do not communicate with each other. These sources can send their compressed outputs to a central unit e.g., a sink node for joint decoding. Distributed Source Coding Using Syndrome (DISCUS) framework has been proposed for data compression in WSNs [61]. Tang et al. [62] propose a DSC scheme for data compression based on a cost function that considers the energy necessary for encoding, transmitting, and decoding the bit stream being compressed. In [63], the authors employed distributed source coding, in particular Slepian-Wolf coding. Slepian-Wolf coding [64] is a promising distributed source coding technique that can completely remove the data redundancy caused by the spatially correlated observations in WSNs. Under this scheme, all sources can be coded with a total rate equal to the joint entropy H (X.Y) without explicit communication between each other, as long as their individual rates are at least equal to their respective conditional entropies H(X/Y) and H(Y/X). Slepian – Wolf coding has been studied for data aggregation in cluster-based WSNs [65],[66]. In [65], applying Slepian-Wolf coding locally within each cluster is shown to be able to overcome the effect of node and relay failures on the data reconstruction at the remote sink. The authors in [66] proposed a distributed optimal-compression clustering protocol (DOC^2) and described the procedures to perform Slepian – Wolf coding with an optimal intra-cluster rate allocation.

In WSNs, data fusion is closely related to data communication. Data fusion occurs in different ways depending on the chosen distributed computing paradigm. The In-network aggregation is the most popular distributed–computing paradigm in WSNs. Research on data compression for communication mainly focuses on how to decrease the communication delay or the required transmission bandwidth. Hans and Schafer [67] present an overview of loss less data compression in the context of audio data. Zhang and Li [68] discuss the implementation of compression algorithms for seismic data. This work estimates the energy reduction after compression that is due to data reduction, and considers the energy costs of communication alone or in isolation from the costs of computation. Another [69] focus on compression of acoustic signals and have developed methodology on evaluating energy consumption trade-offs between computation and communication based on the static-version linear predictive coding (SVLPC), dynamic-version linear predictive coding (DVLPC), and dynamic cyclo-static linear predictive coding (DCLPC) compressions methods. The idea is to take advantage of the node computation capacity and perform the desired fusion algorithm while data is routed towards the sink node. This paradigm is also referred to as Data-Centric Routing. Early work on data centric routing (e.g., SPIN and Directed Diffusion [70]) was shown to save energy through data negotiation and elimination of redundant data. Directed Diffusion incorporates in-network data aggregation, data caching, and data-centric dissemination while enforcing adaptation to the empirically best path. The main goal of this protocol is to compute a path robustly from source to sink through the use of attribute-based naming and gradient paths.

Depending on the network organization, in-network aggregation may occur in different ways, according to routing strategy.

- In flat networks, every node is functionally the same and data are routed in a multi-hop fashion. Thus every node that takes part in the routing process should execute data fusion.

Examples of multi-hop communication with in-network aggregation include the directed diffusion family of algorithms [54] and tree-based algorithms [8].

- In hierarchical networks, we usually have a two-hop communication: one hop for the cluster members to reach the CH and another hop for CHs to reach the sink node. In this type of communication, data fusion is performed by CHs that send the results to the sink. The first hierarchical solution for WSNs was the LEACH [57], but several others have been proposed since then [59][33].
- In a hybrid solution, we may combine flat and hierarchical in-network aggregation. Thus, here we have multiple hops connecting source nodes to their CH and/or multiple hops connecting CHs to the sink. The strategy proposed in [71] illustrates a routing algorithm for hybrid networks performing in-network aggregation.

In data aggregation applications, a sink node is interested in collecting aggregated data from a subset of nodes. In this context, data fusion should use as few nodes and resources as possible to ensure the delivery and aggregation of data generated by source nodes. Authors in [8] have evaluated three schemes:

- *Center at Nearest Source (CNS)*: In this, each source sends its data directly to the source closest to the sink.
- *Shortest Paths Tree (SPT)*: In this, each source sends its data to the sink along the shortest path between both nodes.
- *Greedy Incremental Tree (GIT)*: In this, the routing tree starts with the shortest path between the sink and the nearest source, and at each step after that, the source closest to the current tree is included in the tree.

As Krishnamachari et al. [8] show, the GIT method is the best of the three.

When we have data fusion as a leading role, source selection and route selections are problems of major concern. In [72][73], authors propose an information –directed approach in which sources and communicating nodes are chosen by dynamically optimizing the information utility of data for a given cost of communication and computation. In [74], the Energy efficient Protocol for Aggregator selection (EPA) is proposed for selecting nodes that perform data fusion. The authors derive the optimal number of aggregators, and present fully distributed algorithms for the aggregator selection. In [75], cluster-based communication architecture is used where in data aggregation runs parallel to the CHs, improving the energy efficiency via Meta data negotiation. In addition, for each event and each cluster, only one of the cluster members is selected to send data to the cluster head.

In the following paragraphs, we review the main routing approaches based on aggregation trees.

- **TAG** [9] - The Tiny AGgregation (TAG) approach is a data centric protocol. It is based on aggregation trees and is specifically designed for monitoring applications. This means that all nodes should produce relevant information periodically. The implementation of the core TAG algorithm consists of two main phases: 1) the distribution phase, where queries are disseminated to the sensors, and 2) the collection phase, where the aggregated sensor readings are routed up the aggregation tree.

For the distribution phase, TAG uses a tree based routing scheme rooted at the sink node. The sink broadcasts a message asking nodes to organize into a routing tree and then sends its queries. In each message there is a field specifying the level, or distance from the root, of the sending node. It also elects the node from which it receives the message as its parent. Each sensor then rebroadcasts the received message adding its own identifier (ID) and level. This process continues until all nodes have been assigned an ID and a parent. The routing messages are periodically broadcast by the sink in order to keep the tree structure updated. After the construction of the tree, the queries are sent along the structure to all nodes in the network. TAG adopts the selection and aggregation facilities of the database query languages (SQL). In practice, the sink sends a query, where it specifies the quantities that it wants to collect (attrs field), how these must be aggregated (agg(expr)) and the sensors that should be involved in the data retrieval. This last request is specified through the WHERE, GROUP and HAVING clauses [9]. Finally, an EPOCH duration field specifies the time (in seconds) each device should wait before sending new sensor readings.

During the data collection phase, due to the tree structure, each parent has to wait for data from all of its children before it can send its aggregate up the tree. Epochs are divided into shorter intervals called communication slots. The number of these slots equals the maximum depth of the routing tree. As the time is slotted, sensor nodes can be put to sleep until the next scheduled transmission interval. Data aggregation is performed by all intermediate nodes.

As for most tree-based schemes, TAG may be inefficient in case of dynamic topologies or link/device failures: as, trees are particularly sensitive to failures at intermediate nodes as the related sub tree may become disconnected. In addition, as the topology changes, TAG has to re-organize the tree structure and this means high costs in terms of energy consumption and overhead.

- **COUGAR** [15] – Cougar is most suitable for monitoring applications, where nodes produce relevant information periodically. Cougar is basically a clustering scheme. As soon as the cluster-heads receive all data from the nodes in their clusters, they send their partial aggregates to a gateway node. Of course, being similar to LEACH, Cougar is also affected by the same problems in highly dynamic environments.

Cougar differs from the previous clustering based algorithms in the way cluster-heads are elected. Unlike in LEACH, where each node picks its cluster-head based on signal strength measurements, in Cougar the cluster-head selection may be driven by additional metrics. In fact, a node could be more than one hop away from its cluster-head. For this reason, the routing algorithm adopted to exchange packets within clusters is based on the AODV (Ad hoc On demand Distance Vector) technique. As AODV does not generate duplicate data packets, Cougar is particularly suitable to perform in-network aggregation with duplicate sensitive aggregators. The core Cougar algorithm consists of the node synchronization engine, which ensures that data is aggregated correctly. Each cluster-head has a waiting list containing all nodes it expects a message from. The list is updated every time the node receives a record from a node in its cluster. The cluster-head does not report it's reading to the gateway until at time $t_{scnd,}$ it hears from all nodes in its waiting list.

4 Topology Control Algorithms

The network topology varies due to duty cycling, battery depletion, and friendly interference and hostile jamming; traffic assumes various heterogeneous patterns and QoS requirements due to events that are random in space and time.

In contrast to the case of wired networks, the network topology in wireless networks is not fixed and can be changed by varying the node's transmitting range. So, further energy can be saved if the network topology used to route messages is energy-efficient itself. Topology Control (TC) is one of the most important techniques used in WSNs to reduce energy consumption and radio interference. The purpose of traditional topology control has been to balance two contradictory goals - reducing energy consumption and maintaining high connectivity.

The topic of topology control in general ad hoc networks has been studied extensively. Most early topology control protocols adjusted radio settings e.g., transmission power [76], beam forming patterns [77] to maintain connectivity with an optimal set of neighbors. Because it is often power-efficient to relay packets over several short hops than a single long hop, reducing transmission power is an effective means for reducing overall energy consumption. These methods may be very effective in sensors networks where energy consumption is dominated by the energy consumed in transmitting data packets. However, typical power models considered for sensor networks show that receive power and idle power are comparable to transmit power [78]. Based on this observation, further savings can surely be achieved by not only reducing transmission power, but also setting the sensors radios into a sleep state whenever possible.

The various approaches to the topology control problem appear in the literature. One of the classifications is based on constraints we put on the range assignment; and other is based on the type of information, which is available to network nodes.

The first distinction is between homogeneous and non-homogeneous approaches. In *homogeneous* type of topology control, nodes are assumed to use the same transmitting range, and the topology control problem reduces to the one of determining the minimum value of r such that the critical transmitting range is satisfied.

In the *non-homogeneous* case, nodes are allowed to choose different transmitting ranges provided they do not exceed the maximum range. Non-homogeneous is classified into three categories, depending on the type of information that is used to compute the topology.

. In *location-based approaches*, exact node positions are known. This information is either used by a centralized authority to compute a set of transmitting range assignments; which optimizes a certain measure, or it is exchanged between nodes and used to compute an almost optimal topology in a fully distributed manner.

In *direction-based approaches*, it is assumed that nodes do not know their position, but they can estimate the relative direction of each of their neighbors.

Finally, in *neighbor-based techniques*, nodes are assumed to know only the ID of the neighbors and are able to order them according to some criterion (e.g., distance, or link quality).

Most of the approaches presented in the literature are concerned with building and maintaining a connected network topology, as a network partitioning is highly undesirable.

More recently, some authors have considered the problem of building a k-connected network topology (with k>1), i.e., a topology in which there exists at least k distinct paths between any two-network nodes. Guaranteeing k-connectivity of the communication graph is fundamental in all those applications in which a certain degree of fault-tolerance is needed.

Other authors have recently also considered the topology control problem in which nodes alternate between active and sleeping times, and the goal is to build a network topology such that the sub network composed of the active nodes is connected at any time.

Topology control contributes to power saving mainly in two ways in sensor networks:

(1) It allows non-routing nodes or sensing nodes to maintain lower duty cycle because they don't have to receive packets for the routing purpose and

(2) Routing nodes can act as data aggregation points as all the packets are forwarded through these nodes. The former serves to reduce idle listening and overhearing since sensing nodes can simply turn off their radio most times while performing sensing. The latter serves to reduce the amount of traffic on the routing backbone.

Also recently some authors have investigated the topology control problem with the goal of reducing radio interference. In [79], authors have shown that reducing energy consumption and interference might be conflicting goals and present centralized and distributed algorithms to build low-interference topologies. In [80], authors consider several measures of radio interference in the communication graph and propose algorithms for building optimal or near-optimal topologies according to their metrics. The studies presented in [79] and [80] are only initial steps towards a thorough understanding of the interrelationship between range assignment and level of interference generated in the network and further research on this topic is needed.

Topology control protocols can be classified into two groups depending on which network layer information is used for identifying redundant nodes.

(i) Protocols like PAMAS [81], STEM [82] use MAC layer information to identify redundancy in the network.

(ii) Protocols like GAF [83], ASCENT [84], LEACH [57]-use information from the routing layer and above for identifying redundant nodes.

- *PAMAS: Power-aware multiple access protocol*

PAMAS [81] is a contention-based protocol designed for ad hoc networks with energy efficiency as the primary design goal. It uses a second radio channel to monitor neighbor traffic to determine the duty cycle of its main radio channel. A major contribution of the PAMAS protocol is the power savings achieved without sacrificing network throughput and latency. However, a major drawback observed here is that the power consumption of the nodes during excessive switching between the sleep and wake-up states is not given due attention. Power consumption during state switching is significant. Thus, PAMAS method may not perform satisfactorily without appropriate modifications for WSNs.

- *STEM: Sparse topology and energy management*

STEM [82] is a power saving-strategy that does not try to preserve the capacity of the network. STEM works by putting an increasing number of nodes into sleep node, and then encountering the latency to setup a multihop path. Nodes in STEM must have an extra low power radio (paging channel) that does not go into sleeping mode and constantly monitors the network to wake up the node in case of an interesting event. It claims to improve beyond SPAN and GAF in terms of obtaining higher energy savings so as to prolong system lifetime by trading off an increased latency to establish a multihop path.

- *GAF: Geographic adaptive fidelity*

GAF [83] is another power- saving scheme that saves energy by powering off the redundant nodes. GAF identifies the redundant nodes by using the geographic location and a conservative estimate of the radio ranges. It superimposes a virtual grid proportional to the communication radius of the nodes on to the network. Because the nodes in one grid are equal from the routing perspective, the radios of the redundant nodes within a grid can be turned off. The nodes awake within a grid rotate to balance their energy. In this case, little energy is used, so energy consumption can be reduced.

- *ASCENT: Adaptive self-configuring sensor networks topology*

Protocols like ASCENT [84], which use application level information display high energy savings. In ASCENT, neighbor density and packet loss information is used to determine local connectivity and thereafter choose redundant nodes. It uses the redundancy of nodes over time to extend the network lifetime; each node assesses its connectivity and adapts its participation in the multihop network topology based on the measured operating region. A node may reduce its duty cycle if it detects high data losses due to collisions. ASCENT has the potential for significant reduction of packet loss rate and increases in energy savings as well as its mechanisms are responsive and stable under systematically varied conditions. Nodes do not consume energy equally or fairly. ASCENT may employ a load balance policy that allows nodes to switch state from time to time between active and non-active in order to ensure all nodes share the task of providing global connectivity equally and distribute the energy load. In addition, it has too many parameters to be configured, which make it difficult to be optimized.

- *LEACH: Low energy adaptive clustering hierarchy protocol*

LEACH [57] is a clustering based routing protocol that uses randomized rotation of cluster-heads to evenly distribute the energy load among the sensors in the network. In order to avoid the energy drainage of cluster-heads in LEACH, the cluster-head positions are not fixed and are re-elected periodically. LEACH selects routing paths based on the total path energy. However, it is used for proactive application scenarios and does not take the energy consumption for idle sensing of the channel into account, and the formation of clusters is not energy aware. Some efforts have been made to improve its performance. The protocol works in rounds and defines two main phases: 1) a setup phase to organize the clusters and 2) a steady-state phase that deals with the actual data transfers to the sink node. In the first phase the nodes organize themselves into clusters. Within each cluster a node is elected as the cluster-head. At the beginning of the setup phase, each sensor elects itself to be the local cluster-head for the current round. This decision is made according to a distributed probabilistic approach. The aim is to have, on average, a percentage P of the nodes acting as cluster-heads, where P has to be optimally chosen according to the node density. In practice, sensors calculate the following threshold:

$$T(n) = \frac{P}{1 - P \times \left(r \bmod \frac{1}{P} \right)} \quad \text{if n } \varepsilon \text{ G} \tag{13}$$

$$T(n) = 0 \quad \text{otherwise} \tag{14}$$

where P is the desired percentage of cluster-heads, r is the round number and G is the set of nodes that have not been cluster-heads during the last 1/P rounds. A given node *n* picks a random number [0,1] and decides to be a cluster-head if this number is lower than T(n). A cluster-head sends advertisements to its neighbors using a CSMA MAC. Surrounding nodes decide which cluster to join based on the signal strength of these messages. Finally, based on the number of nodes that are willing to be part of the cluster, each cluster-head creates a TDMA schedule to optimally manage the local transmissions.

Also new protocols have been developed with special focus on energy balancing in order to increase the lifetime of network that can be applied in biomedical applications [85]. In this paper, the optimization has been carried out for the *chain protocol* (when nodes are forwarding the packets toward the BS via the neighboring nodes) and for the *shortcut* type of protocols (when a packet may get to the BS by being first transferred in the chain up to a certain node which then sends it directly to the BS). The optimization problem is solved by combinatorial optimization tool.

There are also a number of research efforts that trade off between latency and energy consumption. The power management approach presented in Kravets and Krishnan [86] selectively chooses short periods of time to suspend and shut down the communication unit, they queue the data before suspending the communication.

5 Conclusions

Recent advances in miniaturization and low-cost, low-power electronics have led to active research in large-scale networks of small, low-power sensors. Networks of such small, possibly microscopic sensors embedded in buildings, machinery and even on people, perform automated continual and discrete monitoring [87]. Sensor networks will eventually be integral to our homes and everyday lives in ways that are difficult to imagine today.

The important design issues of data fusion and topology control in wireless sensor networks are highlighted. We have provided background that supports the design of fusion and topology based solutions for different levels of applications in a WSN, such as data routing, target detection. Since sensor networks are deployed for specific applications, which may be signal processing in nature, we provided a signal processing perspective on different aspects of the sensor-networking problem. The research work made in signal processing and networking fields can be joined together to advance the fundamental theory of sensor networks.

However, there are some limitations regarding the methods and the architectures that should be considered. Some methods might be improved to operate in a distributed fashion. For example, authentication of data and sender is crucial in sensor networks; which has not been discussed in this paper. Therefore, security must be explicitly taken into account in the integrated design of sensor networks.

Currently, there is minimal research that looks at handling QoS requirements in a highly energy-constrained environment like sensor networks. Further, research would be necessary to address issues, such as (QoS) proposed by imaging sensors and real-time applications.

Other possible future research for energy-efficient protocols includes the integration of wireless sensor networks with wired networks. To date, much of the research and system construction has been based on a two- or three-tiered architecture. The highest tier is typically a connection to the Internet, where sensor networks can merge with traditional wired, server-based computing. The integration of WSNs and the Internet is becoming more and more important because of the numerous numbers of WSNs that will join the Internet domain [88].

In any network, routing is a topic that arises almost immediately. So, is of course the case with sensor networks. However, there is an important difference in the routing used by sensor networks. Much of the earlier research in ad-hoc wireless networks was building a way of blindly routing packets to a far-away endpoint. In a sensor network, many applications do processing *at each hop* inside the network (e.g. data reduction by aggregating similar data, filtering redundant information, and so forth). Unlike the Internet–style routing, routing in sensor networks must often be integrated with and influenced by the application. Sink mobility brings new challenges to a sensor network routing [89].

We have outlined several directions for further research which we hope will motivate researchers to undertake additional studies in this field. One of the great challenges is to assure temporal and spatial correlation among the sources while the data is fused and disseminated at the same time. Also, more work needs to be done to investigate the effect of mobility on topology control.

To make wireless sensor networks practically useful, we need to develop network protocols for them that meet several unique requirements and constraints. We find that a practical design of sensor networks may require a joint consideration of multiple layers, e.g., physical layer, MAC layer, network layer, or even application layer.

References:

[1] B. Warneke, M. Last, B. Liebowitz, and K.S.J. Pister, Smart Dust: Communicating with a cubic-millimeter computer, *IEEE Computer*, 2001, 44-51.

[2] Chee-Yee Chong, and S.P. Kumar, Sensor Networks: Evolution, Opportunities, and Challenges, *Proc. of IEEE,* Vol.91, No.8, Aug. 2003, pp.1247-1256.

[3] E. Biagioni and K. Bridges, The application of remote sensor technology to assist the recovery of rare and endangered species, *Intl. J. of High Performance Computing Applications*, 16(3), Aug.2002.

[4] L. Yu, N. Wang, and X. Meng, Real-time Forest Fire detection with wireless sensor networks, *IEEE*, 2005, pp. 1214-1217.

[5] C. Hofmann, C. Weigand, and J. Bernhard, Wireless medical sensor network with ZigBee, *Proc. of the 5th WSEAS Int. Conf. on Electronics, Hardware, Wireless and Optical Communications*, Spain, Feb. 2006, pp. 12-15.

[6] J. Wall, G. Platt, and P. Valencia, Wireless Sensor Networks as agents for intelligent control of distributed energy resources, *IEEE*, 2007, pp. 547-551.

[7] C. Intanagonwiwat, R. Govindan, and D. Estrin, Directed diffusion: a scalable and robust communication paradigm for sensor networks, *ACM/IEEE MOBICOM 2000*, 2000, 56-67.

[8] B. Krishnamachari, D. Estrin, and S. Wicker, Impact of data aggregation in wireless sensor networks, in *International workshop of Distributed Event Based Systems (DEBS), IEEE,* Vienna, Austria, July 2002, 575-578.

[9] Madden et al., TAG: a Tiny Aggregation service for ad hoc sensor networks, *ACM SIGOPS* Operating Systems Rev. 36, 2002, 131-146.

[10] A. Savvides, C. Han, and M.B. Shrivastva, The n-hop multilateration primitive for node localization, *Mobile Network Application,* 8, 4(Aug), 2003, 443-451.

[11] E.F.Nakamura, F.G.Nakamura, C.M. Figueiredo, and A.A. Loureiro, Using information fusion to assist data dissemination in wireless sensor networks, *Telecomm. System*, 30, 1-3 (Nov.), 2005, 237-254.

[12] A. Woo, T.Tong, and D.Culler, Taming the underlying challenges of reliable multihop routing in sensor networks, *Proc. of the 1st Intl. Conf. on Embedded Network Sensor Systems (SensSys' 03),* 2003,14-27.

[13] Qing Zhao, Ananthram Swami, and Lang Tong, The interplay between signal processing and networking in sensor networks, *IEEE Signal Processing Magazine*, July 2006, pp. 84-93

[14] B. J. Bonfils, and P.Bonnet, Adaptive and decentralized operator placement for in-network query processing, *Proc. 2nd Intl. Wrksp. Info. Processing in Sensor Network (IPSN'03),* Apr. 2003,47-62.

[15] Y.Yao , and J.Gehrke, The Cougar Approach to in-network query processing in sensor networks, *ACM SIGMOD Record,* Vol.31, no.3, 2002.

[16] N.Shrivastava, et al., Medians and Beyond: New aggregation techniques for sensor networks, *Proc. ACM Sensys'04*, Maryland, MD, Nov. 2004.

[17] Rabaey et al., Pico Radio supports ad hoc ultra-low power wireless networking, *IEEE Computer Mag.,* 33(7), 2000, 42-48.

[18] R.Min, J. Furrer, and A. Chandrakasan, Dynamic voltage scaling techniques for distributed microsensor networks, *Proc. of ACM MobiCom' 95*, Aug.1995.

[19] T. A. Pering, T.D. Burd, R.W. Brodersen, The simulation and evaluation of dynamic voltage scaling algorithms, *Proc Intl. Symp. on Low Power Electronics Design (ISLPED' 98),* Aug. 1998, pp.76-81.

[20] H. C. Yang and L. R. Dung, Algorithmic Transformations and Peak Power constraint applied to multiple-voltage low power VLSI signal processing, *WSEAS Transactions on Signal Processing,* Issue 12, vol.3, Dec.2007, 479-486.

[21] Alice Wang and Anantha Chandrakasan, Energy–Efficient DSPs for Wireless Sensor Networks, *IEEE Signal Processing Magazine,* July 2002, 68-78.

[22] Y. Xu, J. Heidemann, and D.Estrin, Geography-informed energy conservation for ad hoc routing, in *Proc. Mobi Com,* 2001.

[23] S. Roundy, P. Wright, and J. Rabaey, A study of low level vibrations as a power source for wireless sensor nodes, *Computer Comm.*, vol.26, no. 11, July 2003, pp.1131-1144.

[24] Y. Chen and Q. Zhao, On the lifetime of wireless sensor networks, *IEEE Commun. Lett.*, Vol. 9, Nov. 2005, pp.976-978.

[25] D.L.Hall, and J. Llinas, *Handbook of multisensor data fusion*, CRC Press, USA, 2001.

[26] D.L. Hall, and J. Llinas , An introduction to multisensor data fusion, *Proc. IEEE* 85,1 (Jan), 1997, 6-23.

[27] H. F. Durrant –Whyte, Sensor Models and Multisensor integration, *Intl. J .Robotics Res.* 7,6 (Dec.), 1988, 97-113.

[28] J. Zhao, R. Govindan, and D. Estrin, Residual energy scans for monitoring wireless sensor networks, *IEEE Wireless Commns. And Networking Conf. (WCNC'02)*, vol.1 IEEE, Orlando, FL., 2002, 356-362.

[29] R. Nowak, U. Mitra, and R. Willette, Estimating in homogeneous fields using wireless sensor networks, *IEEE J., Select. Areas in Comm.,* 22, 6(Aug.), 2004, 999-1006.

[30] Biswas et al., A probabilistic approach to interference with limited information in sensor networks, *Proc. of 3rd Intl. Symp. (IPSN'04)*, 2004, 269-276.

[31] B. Krishnamachari, and S. Iyengar, Distributed Bayesian algorithms for fault-

tolerant event region detection in wireless sensor networks, *IEEE Trans.Comput.*53, 3(Mar.) 2004,241-250.

[32] Luo et al., On distributed fault tolerant detection in wireless sensor networks, *IEEE Trans. Comput.,* 55, 1(Jan.), 2006.

[33] I. Gupta, D. Riordan, and S. Sampalli, Cluster head election using fuzzy logic for wireless sensor networks, *Proc. of the 3rd Annual Commn. Netw. & Services Res. Conf. (CNSR'05), IEEE,* Canada, 2005, 255-260.

[34] M. Yusuf, and T.Haider, Energy-aware fuzzy routing for wireless sensor networks, in *IEEE Intl. Conf. on Emerging Technologies (ICET'05)*, Pakistan, 2005, 63-69.

[35] T. Srinivasan, R. Chandrasekar, and V. Vijay Kumar, A fuzzy, energy efficient scheme for data centric multipath routing in wireless sensor networks, in *IFIP Intl. Conf. on wireless and optical commns. Networks, IEEE,* Bangalore, India, 2006.

[36] Weilian Su and Theodores C. Bougiouklis, Data fusion algorithms in cluster-based wireless sensor networks using fuzzy logic theory, *Proc. of 11th WSEAS Intl. Conf. on Comm.,* Greece, July 2007 pp. 291-299.

[37] M.R. Roth, Survey of neural network technology for automatic target recognition, *IEEE Trans. Neural Netw.* 1, 1 (Mar.), 1990, 28-33.

[38] L. Xiao, S. Boyd, and S. Lall, A scheme for robust distributed sensor fusion based on average consensus, in *Proc. of the 4th Intl. Symp. on Information Processing in Senor Networks (IPSN' 05)*, 2005, 63-70.

[39] L.Xiao, S.Boyd , and S. Lall, A space-time diffusion scheme for peer-to-peer least squares estimation, in *Proc. of 5th Intl. Conf. on Information Processing in sensor networks (IPSN'06)*, 2006,168-176.

[40] R.D. Nowak., Distributed em algorithms for density estimation and clustering in sensor networks, *IEEE Trans. Sig .Proc.* 51,8(Aug), 2003, 2245-2253.

[41] Blatt D. and Hero A., Distributed maximum likelihood estimation for sensor networks in *Proc .of the IEEE Intl Conf. on Acoustics, Speech, and Signal Processing, (ICASSP'04),* vol.3, IEEE, Canada, 2004, 929-932.

[42] G. Hartl and B. Li, Loss inference in wireless sensor networks based on data aggregation, in *Proc. of the 3rd Intl. Symposium on Information Processing in Sensor Networks (IPSN' 04)*, 2004, 396-404.

[43] L. Fang, W. Du , and P. Ning, A beacon-less location discovery scheme for wireless sensor networks in *Proc of the 24th Annual Int. Conf. (INFOCM'05),* 2005, 161- 171.

[44] Li et al., Source localization and tracking using distributed asynchronous sensor, *IEEE Trans. Sig. Proc.* 54, 10(Oct.) 2006, 3991-4003.

[45] M. Rabbat, and R.D. Nowak, Distributed optimization in sensor networks, Proc. of 3rd Intl. Symp. (IPSN'04), 2004,20-27.

[46] Guestrin et al., Distributed regression: an efficient framework for modeling sensor network data in *Proc. of 3rd Intl. Symp. (IPSN'04)*, 2004, 1-10.

[47] S. Santini, and K. Romer, An adaptive strategy for quality-based data reduction in wireless sensor networks, in *Proc. of 3rd Intl. Conf. on Networked Sensing Systems (INSS'06)*, Chicago, Il, 2006, 29-36.

[48] J. Liu, Y. Zhang, and F. Zhao, Robust distributed node localization with error management, in *Proc. of the 7th ACM Intl. Symp. on (Mobi Hoc'06), ACM*, Italy, 2006, 250-261.

[49] R. Olfati-Saber, Distributed kalman filter with embedded consensus filters, in *44th IEEE Conf. (CDC'05), IEEE*, Spain, 2005, 8179-8184.

[50] B. Sinopoli et al., Kalman filtering with intermittent observations, *IEEE Trans. Autom. Conf.,* 49, 9 (Sept.) 2004, 1453-1464.

[51] Savvides et al., The n-hop multilateration primitive for node localization, *Mobile Netw. Appl.* 8, 4 (Aug.) 2003, 443-451.

[52] A. Jain, E.Y. Chang, and Y.F. Wang, Adaptive stream resource management using

Kalman filters, in *Proc. of the 2004 ACM SIGMOD Intl. Conference on Management of Data (SIGMOD'04)*, ACM, Paris, France, 2004, 11-22.

[53] R. Wasniowski, Monitoring processes using sensor networks and an extended Kalman filter, *WSEAS Intl. Conf. on Dynamical Systems and Control*, Italy, Nov.2005, pp 540-544.

[54] C. Intanagonwiwat et al. Directed diffusion for wireless sensor networking, *IEEE/ACM Trans.Networking,* 11(1), 2003, 2-16.

[55] Boulis et al., Aggregation in sensor networks: An energy–accuracy trade-off, in *Proc .of 1st IEEE Intl. Workshop on Sensor Network Protocols and Applns. (SNPA'03)*, Alaska, May 2003.

[56] J.Kulik, W. Heinzelman, and H. Balakrishnan, Negotiation-based protocols for disseminating information in wireless sensor networks, *Wireless Networks,* 8, 2002, 169-185.

[57] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy efficient communication protocol for wireless microsensor networks, in *Proceedings on 33rd Hawaii international conferences system sciences (HICSS),* vol. 8, Jan. 2000, 8020-8029.

[58] C. Intanagonwiwat, D. Estrin, and R. Govindan, J. Heidemann , Impact of network density on data aggregation in wireless sensor networks, in *Proc. of the 22nd IEEE International Conf. Distributed Computing Systems (ICDCS'02) IEEE*, Vienna, Austria, 2002, 457-458.

[59] A.T. Hoang, and M. Motani, Collaborative broadcasting and compression in cluster-based wireless sensor networks, in *Proc. of the 2nd Eur. workshop on wireless sensor networks (EWSN'05), IEEE,* Istanbul, Turkey, 2005, 196-206.

[60] Xiong et al., Distributed source coding for sensor networks, *IEEE Sig. Proc.* Mag.21, 5(Sept.) 2004, 80-94.

[61] S.S. Pradhan, and K. Ramchandran, Distributed source coding using syndromes (DISCUS): Design and construction, *IEEE Trans., Information Theory,* vol.49, no.3, Mar.2003, pp.626-643.

[62] C. Tang, C.S. Raghavendra, and V.K. Prasanna, An energy efficient adaptive distributed source-coding scheme in wireless sensor networks in *Proc. of the IEEE Intl. Conf. on commns. (ICC'03),* vol.1, IEEE, AK, 2003, 732-737.

[63] R. Critescu et al., Networked slepian-wolf: theory, algorithms, and scaling laws, *IEEE Trans. on Inform. Theory*, vol.51, no.12, Dec.2005, pp.4057-4073.

[64] Slepian D., and Wolf J., Noiseless encoding of correlated information sources, *IEEE Trans. Information Theory*, vol. IT-19, Jul.1973, pp.471-480.

[65] Marco D., and Neuhoff D.L., Reliability vs. efficiency in distributed source coding for field gathering sensor networks, in *Proc. of the 3rd Intl. Symp. on Inform. Processing in Sensor Networks (IPSN'04)*, Berkeley, CA, Apr. 26-27, 2004, pp.161-168.

[66] Pu Wang, Li Cheng, and Jun Zheng, Distributed data aggregation using clustered slepian-wolf coding in wireless sensor networks, in *Proc. of the IEEE Intl. Conf. on commns. (ICC'07), 2007, 3616-3622.*

[67] M. Hans and R.W. Schafer, Lossless compression of digital audio, IEEE Signal Processing, vol.18,no.4, pp.21-32, July2001.

[68] Y. Zhang and J. Li, Efficient seismic response data storage and transmission using arx model-based sensor data compression algorithm, Earthquake Engineering and Structural Dynamics, vol.35, 2006, pp, 781-788.

[69] S. Puthenpurayil; R.Gu and S.S.Bhattacharya, Energy–aware data compression for wireless sensor networks, *Proc. of the Intl. Conf. on Acoustics Speech, and Signal Processing (ICASSP)*, vol. 2, Hawaii, Apr. 2007, pp. 45-48.

[70] R.C. Shah and J. Rabaey, Energy-aware routing for low energy ad hoc sensor networks, *IEEE WCNC*, Orlando, Fl , Mar., 2002, 350-355.

[71] E.F. Nakamura et al., On demand role assignment for event detection in sensor networks, in *Proc. of the 11th IEEE Intl. Symp. on Computers and Commn. .(ISCC'06)*, Italy, 2006, 941-947.

[72] Zhao F., Shin J., and Reich J., Information–driven dynamic sensor collaboration for tracking applications, *IEEE Signal Process. Mag.*, 19(2), 2002, 61-72.

[73] Zhao F. et al., Collaborative signal and information processing: an information directed approach, *Proc. IEEE* 91, (8), 2003, 1199-1209.[74] Y.P. Chen, A.L. Liestman, and J.Liu, A hierarchical energy efficient framework for data aggregation in wireless sensor networks, *IEEE Trans. Vehic. Tech.* 55, 3, 2006, 789-796.

[75] H.Chen, H. Mineno, and T.Mizuno, A meta-data based data aggregation scheme in clustering wireless sensor networks, in *Proc. of the 7th Intl. Conf. on Mobile Data Management (MDM'06) IEEE*, Japan, 2006, 154-154.

[76] R. Ramanathan and R. Hain, Topology control of multihop wireless networks using transmit power adjustment, in *Proc. Of the 19th Intl. Annual Jt. Conf. of the IEEE Comp. & Commns. Societies (INFOCOM),* 2000.

[77] Q. Li and D. Rus, Sending messages to Mobile users in disconnected ad hoc wireless networks, in *Proc. of the 6th Annual Intl. Conf. on Mobile Computing and Networking,* 2000.

[78] M. Stemm. and R. Katz, Reducing power consumption of network interfaces in hand-held devices, in *Proc. of the 3r d Workshop on Mobile Multimedia Commns. (MoMuC-3)*, Princeton, NJ, 1996.

[79] M. Burkhart et al., Does topology control reduce interference? In *Proc. of ACM Mobihoc'04*, 9-19, 2004.

[80] K. Maoveni-Nejad, and X. Li, Low-interference topology control for wireless ad hoc networks, *Ad hoc sensor networks*, 2005.

[81] S. Singh and G. Raghavendra, PAMAS: Power aware multi-access protocol with signaling for ad hoc networks, *Proc. of ACM Comp.Comm.Review*, 1998, 5- 26.

[82] C. Schurgers, V. Tsiatsis, and M. Srivastava, STEM: Topology management for energy efficient sensor networks, *Proc. of IEEE Aerospace Conf. Networks*, 91(7), 2003, 1002-1022.

[83] Y. Xu, S. Bien, Y. Mori, J. Heidemann, and D.Estrin, Topology control Protocols to conserve energy in wireless ad hoc networks, *CENS Tech. Report UCLA*, No.6, USA, 2003.

[84] A. Cerpa and D.Estrin, ASCENT: Adaptive self-configuring sensor network topologies, in *Trans. on Mobile Computing, IEEE J.* 2004, 272-285.

[85] J.Levendovszky, A. Bojarszky, B. Karlocai, A. Olah, Energy balancing by combinatorial optimization for wireless sensor networks, ***WSEAS** Transactions on Communications,* issue 2, vol. 7, Feb.2008.

[86] R. Kravets and P.Krishnan, Application-driven power management for mobile communication, in *Proc. of the 4th Annl. ACM/IEEE Intl.Conf.on Mobile Computing and Networking (Mobi Com)*, Dallas, Tx, Oct.1998.

[87] F. Rahman and N. Shabana, Wireless Sensor Network based Personal Health monitoring *Proc. of the 11th WSEAS Int. Conf. on Computers* system, WSEAS Transactions on Systems, Issue 5, vol. 5, May 2006, pp.966-972.

[88] W. Su and B. Almaharmeh, QoS-enabled integration of wireless sensor networks and the Internet, , Greece, July 2007, pp.214-219

[89] H. Lee, E. Kim, K. Cho, N. Ha, Y. Choi, An energy-efficient data dissemination using cross topology in wireless sensor network, *Proc. of the 8th WSEAS Intl. Conf. on Neural Networks*, Canada, June 2007, pp.48-52.