Sensor Network Information Theory

Gregory J. Pottie, Huiyu Luo, and Ameesh Pandya
UCLA Electrical Engineering Department
Los Angeles, CA, USA

CONTENTS

1. INTRODUCTION

1.1Wireless Sensor Networks

Wireless sensor networks include in each node some combination of sensors, signal processing, data storage and wireless communications capability (e.g. radio), together with an energy supply [1,2]. The package size and the capabilities of these components vary widely among the different applications, and often networks will be composed of a heterogeneous mix of nodes of differing capabilities. It is anticipated that such networks will be deployed on very large scales for applications including industrial, security, agricultural and medical monitoring. A common requirement for these networks is energy efficiency, since battery technology is evolving more slowly than the other components and contributes a large fraction of the system size and cost. Additionally, battery recharging often is not available because of difficult environments (e.g., shaded regions) or the logistical issues associated with massive numbers of sensors being deployed. Another characteristic that distinguishes sensor networks from general communication networks is that the data being transmitted from different sensor nodes are observations of related physical phenomena and thus can have high correlation. This should be exploited to reduce the communication data rate and the transmission power.

In sensor networks, greatest efficiency is achieved when signal processing and networking are viewed as connected processes. The traffic generated from some region of the network depends not only on the sources being observed but also the signal processing strategy, which includes how the data is routed for cooperative processing. The traditional approach of simply shipping raw data back to some gateway is neither scalable (due to congestion) nor energy efficient. Clearly this implies making decisions or otherwise compressing the data as near the source as possible, but it is not clear how to actually accomplish this nor what are the fundamental limits on the efficiency of the intertwined processes. Some of the issues are illustrated in Figure 1. Gray scale intensity denotes the likelihood of high activity, whether in terms of signal processing or communications, in response to any particular physical event. For example, nodes may always have an energy detector active, which will then bring other sensors and signal processing to higher levels of activity internally. With somewhat lower probability, the node may trigger formation of a local network for such purposes as data fusion or beamforming. This will require a local exchange of information, not least to learn if the data are sufficiently correlated for subsequent cooperation to be worthwhile. However, after fusion or beamforming is accomplished, only the decisions need be relayed further. These may be combined with other information collected by the network as the data propagates back to the gateway. Most of the time, raw data only propagates within the local cooperative network, but the gateway may on occasion not be satisfied with the source identification decision and may request more information to be sent over the entire multi-hop link. This might be done either in support of a learning algorithm, or to provide more details at the request of an end user. With this hierarchical approach, the network preserves the capability of extracting selected high-resolution data, but for routine situations little long range traffic is generated and most decisions are local. This provides scalability, and conserves energy.

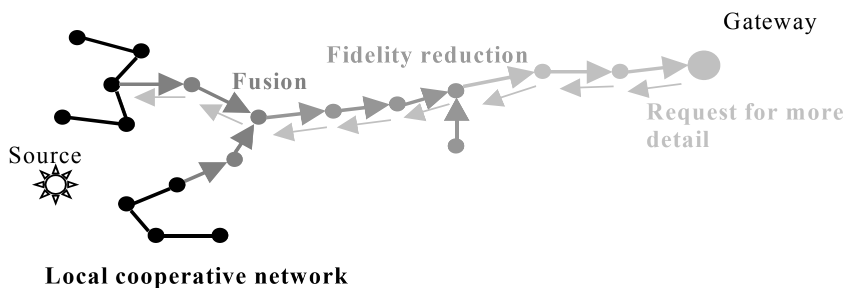

Figure 1. Interaction of Signal Processing and Networking in Sensor Networks

The question naturally arises as to what is the fundamental limit on the energy required to transmit observed data to the gateway such that it can recover these data within fidelity constraints or make reliable decisions, given a network topology, channel transition matrix, source distribution, observation noise, and the definition of estimation fidelity and decision reliability. This question can be approached from the direction of network information theory, and formulated as a mathematical optimization problem.

1.2. Information theory

In this section basic notions from information theory are introduced followed by a description of the types of network information theory problems arising in sensor networks.

1.2.1 *Entropy and the noiseless source coding theorem*

Shannon's measure of information is termed the entropy [3]. Let X be a random variable (RV) taking values in the set $\boldsymbol{X}=\{x_1,x_2,\ldots, x_n\}$ with the probability of element i being $P(x_i)$. Then the entropy in bits is given by

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log P(x_i) \qquad (1)$$

where the logarithm is base 2. When X takes on binary values the entropy achieves its maximum value of 1 bit when the outcomes are equally likely, and a minimum of 0 when either outcome is certain.

According to the noiseless source coding theorem, the entropy represents the minimum number of bits required to represent a sequence of discrete-valued RVs with complete fidelity (i.e., the ability to reproduce exactly the original sequence). Many practical source coding techniques have been developed that get very close to the entropy limit, provided one is dealing with long sequences of random variables.

The weak law of large numbers states that for iid RVs $X_1, X_2,\ldots, X_n$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mathrm{E}[X] \text{ as } n \rightarrow \infty \qquad (2)$$

One consequence [4,5] is the asymptotic equipartition property (AEP) which states that if $P(x_1, x_2,\ldots, x_n)$ is the probability of observing the sequence $x_1, x_2,\ldots, x_n$, then $-1/n \log P(X_1, X_2,\ldots, X_n)$ is close to $H(X)$ for large n. As a result, a small set of the sequences known as the typical set will consume almost all of the observations. Further, the members of the typical set are equally likely, and the number of typical sequences is approximately 2^{nH}. The remaining sequences, although large in number, are collectively only rarely observed.

Shannon's universal source coding method allocates one bit to labeling whether a sequence is a member of the typical set or not. The typical set members are enumerated and then encoded using their indices as the codewords. Since there are 2^{nH} typical set members, the index length will be at most the integer above nH, plus 1 for the indication of typical set membership. For the atypical set members, at most $n\log|\boldsymbol{X}| +1$ bits will be required; they are sent in raw form. The expected number of bits required to describe a message is only nH+1 bits because the atypical set has vanishing probability as n gets large. Thus, the number of bits per symbol required will approach the entropy H, the smallest possible according to the noiseless source coding theorem. Note that the coding scheme is fully reversible, there being a one to one correspondence between codewords and the original data. Practical schemes can closely approach H [6].

Thus, coding effort is spent on the typical set, while it is unimportant how the more numerous but low probability atypical set members are dealt with. Given that the conclusion depends only upon the weak law of large numbers, this is a design principle with broad implications.

1.2.2 *Mutual information and channel capacity*

The joint entropy between two variables X and Y is

$$H(X,Y) = -\sum_{x\in \boldsymbol{X}}\sum_{y\in \boldsymbol{Y}} P(x,y)\log P(x,y) \qquad (3)$$

while the conditional entropy is defined as

$$H(Y \mid X) = \sum_{x\in \boldsymbol{X}} P(x)H(Y \mid X = x) = -\sum_{x\in \boldsymbol{X}}\sum_{y\in \boldsymbol{Y}} P(x,y)\log P(Y \mid X). \qquad (4)$$

The joint and conditional entropies are related by the chain rule:

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \qquad (5)$$

Notice that it is not generally the case for $H(X\mid Y)$ to be the same as $H(Y\mid X)$.

A question that naturally arises is the extent to which knowledge of X (e.g. a sensor reading) assists in knowing Y (e.g., a reading from another sensor observing the same source). This is measured by the mutual information:

$$I(X;Y) = H(X) - H(X \mid Y) \qquad (6)$$

or equivalently,

$$I(X;Y) = H(Y) - H(Y \mid X). \qquad (7)$$

The mutual information can thus be thought of as either the reduction in uncertainty of X due to knowledge of Y, or the reduction in the uncertainty of Y due to knowledge of X. If X and Y are independent, the mutual information is zero. Mutual information plays a central role in defining channel capacity, in problems of lossy compression [7,8], and relates directly to the problem of optimal data fusion.

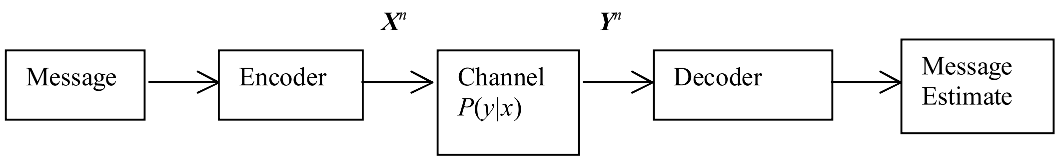

Figure 2. Discrete communication problem

The communications problem is depicted in Figure 2. A message sequence is encoded to produce the vector $\boldsymbol{X}^n$, and then a channel produces output sequence $\boldsymbol{Y}^n$ according to probability transition matrix $P(y|x)$. The decoder then estimates the original sequence. The goal is to maximize the rate of transmission while exactly reproducing the original sequence. This maximum rate is known as the Shannon channel capacity. It is given by

$$C = \max_{P(x)} I(X;Y) \qquad (8)$$

where the maximum is over all possible input distributions $P(x)$. The proof [4] uses the concept of typical sets, and reveals that noise fundamentally limits capacity rather than transmission quality.

Capacity can also be defined for channels with discrete inputs but with output vectors drawn from the reals. For the Gaussian channel, signals of power P and bandwidth W are passed through a channel that adds white Gaussian noise of power spectral density $N_o/2$. Then

$$C = W \log(1 + P/WN_o) \text{ bits/s.} \quad (9)$$

That is, capacity (not reliability) increases linearly with bandwidth (the dimensions available) and logarithmically with signal to noise ratio (SNR). Viewed another way, given fixed bandwidth and data rate requirements, there is some minimum SNR that is required for reliable transmission. Below this SNR, reliable communication is not possible at the desired transmission rate—the error rate very sharply escalates to unusable levels. The (largely achieved) aim of channel coding is to approach this limit with reasonable computational delay and complexity [9-11]. For radio channels, space-time coding (taking advantage of multiple antennas) is also advancing towards the fundamental limits [12].

1.2.3 *Rate distortion coding*

Quantization is an example of lossy source coding—the conversion of a sequence of random variables into a more compact representation of rate *R*, where there is some distortion *D* between the original and reconstructed sequence (e.g., the quantization noise). The aim may be either to minimize *R* subject to a constraint on the maximum distortion *D*, or minimize *D* subject to a constraint on the maximum rate *R* [13]. Suppose the distortion measure is the squared error between the input sequence of RVs ***X*** (e.g., drawn from the reals) and the reconstructed sequence ***Y*** (some finite set). Let *D* be the maximum permitted distortion. The rate distortion theorem states that the minimum achievable rate is

$$R(D) = \min_{f(y \mid x)\,:\, \mathrm{E}[(Y-X)^2] \le D} I(X;Y) \qquad (10)$$

Most such problems are difficult to solve and result in bounds rather than exact solutions. An exception is when the source generates a sequence of iid Gaussian RVs with zero mean and variance σ^2, for which the result is

$$R(D) = \frac{1}{2}\log\frac{\sigma^2}{D},\ 0 \le D \le \sigma^2, \text{and 0 otherwise.} \qquad (11)$$

This can also be formulated as distortion as a function of rate, $D(R) = \sigma^2 2^{-2R}$. Thus, each additional bit reduces the distortion by a factor of 4, or 6 dB. Vector quantizers can closely approach this rate distortion bound. In lossy source coding, in practice there are generally a number of steps. First some heuristic is applied to remove most of the correlation of the source, using knowledge of the source statistics, and in the case of video or audio signals, knowledge of which features have the largest perceptual impact. Then rate-distortion coding or lossless source coding is applied to remove the remaining correlation in the sequence. Additionally, while separately treating source and channel coding is optimal in terms of the achievable rates, in practice combined approaches can offer in some cases lower delay or complexity [14-20].

1.2.4 *Data processing inequality*

Consider now the information processing system depicted in Figure 3.

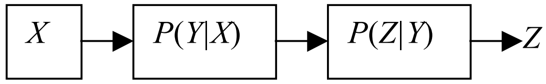

Figure 3. An information processing system

Here a source produces data X that passes through a block (e.g. a channel) which produces Y according to the conditional distribution $P(Y\,|X)$. A processing step produces Z according to the conditional distribution $P(Z\,|\,Y)$. The data processing inequality states that

$$I(X;Y) \geq I(X;Z)\,. \qquad (12)$$

That is, processing cannot increase the mutual information between the received sequence $\boldsymbol{Y}$ and the source sequence $\boldsymbol{X}$; it can at best keep it constant and may destroy information. It is in particular true when Z=$g(Y)$ where g is an arbitrary function. Thus, all the information is in the original signal, and it is likely all downhill from there as it is processed. What processing can do is apply information-preserving transformations that present the data in a more convenient form. If a more compact representation is required due to constraints, then a reasonable goal is to maximize the mutual information subject to rate or distortion constraints.

1.2.5 *Network information theory*

In the general network information theory problem [4], some collection of nodes gather information from some source, perform processing, and convey the eventual decision across the network to some information sink. The processing or source coding can be integrated with channel coding, with groups of nodes involved in the combined process. Determination of the minimum resources (e.g., communications bandwidth) required to reproduce the source at the sink at the desired level of fidelity (distortion) is an extremely difficult problem. For a single communications link in which only one node has access to sensor measurements, the separation theorem applies: source coding can be performed entirely at the sensor node without taking into account the communications channel, and then channel coding can be applied neglecting the source coding procedure. However, with a network, source and channel coding considered separately are known to be suboptimal. Even worse, apart from a small number of special cases even the cooperative source and channel coding problems when considered separately result in performance bounds rather than complete definition of the region of achievable rates [21-26]. Consequently, it is unclear what level of suboptimality is obtained with different strategies for cooperatively performing source coding and conveying information across the network.

Many different combinations are possible [27-42]. For example, considering the source coding problem alone, one could select the nearest node to the source and perform conventional rate distortion coding (the gray node in Figure 4). Better results may be obtained by combining information from nodes within the second tier. This can involve some combination of processing of sensor information at the individual nodes, fusion as data is moved among these nodes, and final fusion at the gray node [43,44]. Because of this possibility of processing along the way, simply applying shortest path routing will not in general be optimal. Similarly, in considering the process of sending this fused information to a data sink (the white node), one could simply route along the shortest path. However, cooperative communications whereby multiple nodes help in a series of relay steps may result in higher capacity communications. Finally, a combined cooperative source and channel coding approach may provide still further efficiency gains.

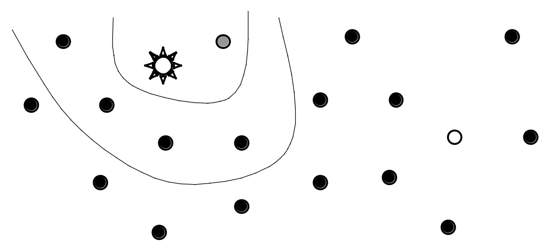

Figure 4. Network information theory problems

Table 1 provides a classification of some network services problems into network information theory problems. In all of these problems there are in practice constraints on latency (delay) and in the rate of transmission (due to limited access to communications, possibly due to bandwidth or energy constraints). In the time distribution (synchronization) problem [45,46], different nodes in the network need to know what time it is relative to one or more neighbors within some time ambiguity Δt. This requirement may vary among the different nodes within the network, or for different purposes in the same node (e.g., frame synchronization for communication vs. time stamps for coherent combining of data). The basic trade is the use of minimum resources (rate) to meet the time ambiguity (distortion) target. In the data transmission problem, information from one or more nodes must be transmitted across the network and be received with a low probability of error $P(e)$, subject to resource constraints such as bandwidth and transmission power. The usual question is the maximum rate that can be achieved subject to these constraints, while in practice delay is usually also considered. For sensor measurements (including position) the concern is the quality of reconstruction in terms of the spatial and temporal dimensions (Δr, Δt), subject to delay and resource (e.g. rate) constraints. As this involves both processing and communication, it can in its most general form be a combination of network rate distortion and capacity problems. With relatively abundant bandwidth but concern for energy consumption, it becomes a rate distortion problem (albeit potentially with network channel coding), with energy playing the role of rate.

Table 1: Network information theory problems

Services	Quality Measure	Constraints	Problem Type
Synchronization	Δt	Delay, rate	Rate distortion
Data transport	$P(e)$	Bandwidth, power	Channel capacity
Measurements	Δr, Δt	Delay, rate	Combined

Delay arises in networks due to congestion and the inherent protocol delays in multi-hop transmissions. It can be ameliorated using higher bandwidth and greater directionality in antennas to reduce interference and allow longer hops. Rate constraints are due in part to restrictions on bandwidth, noise sources that reduce the signal to noise ratio (SNR), and transmission power constraints. Some combination of multiple antennas, processing at source, energy re-supply, and increased densities of nodes (to allow more hops and thus increased frequency re-use) can relax these constraints. However, there are in any practical situation limits to how far one can push in any of these directions, still leaving the question of what is the best performance that can be achieved given these constraints.

1.3 Overview

The structure of remainder of this article is as follows. In section 2 we address the conditions under which the network is scalable, that is, for which information extraction remains possible as the number of nodes tends to infinity. It will be seen that given fidelity constraints, networks with finite resources can still be scalable using even simple forms of local cooperation, and further that the local cooperation operations will dominate overall resource usage. Section 3 then addresses more complicated forms of local interactions with the view towards establishing some information theoretic limits for interactions among small groups of nodes. Section 4 addresses the combined problem of processing and routing. In section 5 we briefly consider how mobile nodes and other infrastructure within the network can dramatically change the set of communications and sensing tradeoffs available. Section 6 presents some conclusions and areas for further research.

2. NETWORK SCALABILITY

A network is scalable in the number of nodes n if the quality of service available to each node does not depend on the network size. In communication networks this implies that the message rate achievable by individual nodes over some distance should not decline as numbers increase. This should be distinguished from the network capacity, the total transmission rate. For example, in a radio network, due to the decay of signals with distance, node pairs that are sufficiently far apart can communicate with interference levels below the threshold that would produce unacceptable performance. Thus frequencies or channels can be re-used so that the capacity of a large radio network will tend to infinity with the number of nodes, even with finite bandwidth.

Unfortunately, general ad hoc networks with resource constraints (energy, bandwidth) are usually not scalable [47-52]. Suppose *n* nodes in a network are equally likely to serve as a source or destination node for any other node within the network, and further that communications signal strength decays with distance. One way to achieve communications to distant nodes while limiting interference is to use a multi-hop relay strategy, allowing frequency re-use [47]. Lower energy transmissions can be used, generating interference over a smaller radius, and thus permitting other nodes to communicate. As the network grows in size the average number of hops will grow as the square root of *n*, and consequently an increased fraction of the traffic a node conveys will have originated at other nodes. The ability of each node to ship its own data in this naïve scheme thus declines as the network size grows. Unfortunately, cooperation in transmission does not make things much better than simple relay strategies [53]. In sensor networks, the communications signals decay fairly sharply with distance so that the frequency re-use implied by use of simple relays is a good high-level strategy for managing interference.

For this reason, we will largely pass over the prospect of cooperative communications. Classic versions of the cooperation problem include the multiple access, relay, and interference channels [4]. The solutions that approach channel capacity require a high degree of coordination among the users. For example, coherent combination of radio communications requires precise synchronization, which while feasible in asymmetric situations such as sensor node to satellite communications [54] is difficult to achieve at low energy in a many-to-many peer situation. This makes use of high-capacity methods such as space-time codes [55,56] a difficult prospect, even neglecting the overhead of communicating among the transmitter and receiver groups. Cooperation among peers would appear to be mainly useful for dealing with reliability issues (i.e. overcoming occasional gaps in the multi-hop network [57]), which in many circumstances can more simply be handled by either increasing node density or through use of the methods described in section 5. The main action in communications cooperation will rather be at the level of routing, to be discussed in section 4.

The next alternative to consider is use of a lossless compression scheme that can exploit all the correlation in sensor data [58,59]. Of course, there is no truly lossless method for observing nature, but one can take the output of the A/D converter on each sensor as the quantity to be reproduced. As the sensor density increases, the correlations of nearby sensors increase until limited by the detector noise. Thus after a certain point increased density serves only to linearly increase the amount of noise that is to be perfectly reproduced (since it is uncorrelated among sensors), and thus the network traffic. The network therefore remains not scalable.

The only solution to the scalability problem is to change the source-destination probabilities so that most traffic is local, i.e., traverses relatively few hops [60]. In telecommunications networks, this is achieved through hierarchy and increased resources: local traffic is aggregated, and then is transported over longer distances using larger bandwidth pipes. Progressively higher capability and longer range layers are added on. Within any given layer, traffic only moves some limited number of hops [61]. The same result can be obtained by having data sinks throughout the network, with their number being some constant fraction of the number of nodes. The sensor network approach to this problem is to perform signal processing such that relatively little traffic is even transported over a long distance. This is only feasible given data reproduction fidelity constraints, in effect, lossy source coding. Having such constraints is very natural for sensor networks, as for any given purpose there will be specifications for spatial and temporal resolution, and the number of bits of A/D conversion.

Consider first the situation depicted in Figure 5. In 5(a) the objective is to sense some point source while in 5(b) the objective is to map the lines of constant value of some distributed phenomenon such as temperature or light intensity. In both cases assume there is a data fidelity constraint. This could for example be the requirement that the point source is measured above some signal to noise ratio threshold, its location determined to within some number of meters, or a decision made on it belonging to some class of objects according to some error probability constraint. Similarly one might require the distributed phenomenon measurements to be made with some precision, or the positions of the lines of constant temperature within some location error.

(a) (b)

Figure 5. Sensing of (a) point and (b) distributed sources

Now consider what happens when node density is increased from having the black nodes only to also include the white nodes. In Figure 5(a) clearly there is a white node closer to the source than the nearest black node. It may have been the case that the black nodes needed to cooperate (e.g., perform coherent data fusion such as beamforming [62]) to make a sufficiently reliable estimate. Having a closer node will decrease the number of such nodes that must cooperate until in the limit of very high density not only will the SNR for the closest node exceed any required threshold but its own location will sufficiently approximate that of the source. Thus, as node density increases, eventually the simple strategy of selecting the closest node will meet fidelity constraints. Moreover, the amount of information to extract from the network is unchanged: this depends only the number of sources and the fidelity constraints, neither of which change with node density. However, communications relaying capacity increases with the number of nodes, as interference radius decreases and therefore more message streams can independently propagate through the network. If there is still insufficient communication relay capacity, further increasing the density of relays will eventually solve the problem, since while network capacity can grow without bound the network traffic will reach a plateau due to the fidelity constraints and sensor selection protocol. Thus, with fidelity constraints, even very simple cooperative processing strategies (e.g. node selection) render the network scalable.

For distributed sources [63], spatial fidelity constraints allow definition of a grid, much as an image can be divided into pixels. So long as node density is high enough to have at least one node within each pixel in this scene, then adding further nodes does not increase the amount of information to be extracted but does increase the ability of the network to relay that information out. Thus, given a mechanism to select which nodes to use for detection, both the point and distributed source sensor network problems are scalable with the number of nodes.

The basic conclusion is that for scalability some mechanism must be found such that the most resource intensive interactions are local. It then follows that longer range interactions will occur with lower probability, and be less important in the overall network optimization. In this the problem is analogous to that of universal source coding, with the local interactions in the role of the typical set and the long-range interactions in the role of the atypical set. This is fortunate in two respects. First, the full-blown network optimization problems are analytically intractable, but there exists a voluminous set of practical solutions for dealing with information transport in large networks. Making use of these solutions will thus not only save in development/design time but with judicious choices may have little impact on the overall optimality of the design compared to a custom implementation. That is, a hierarchical solution may yield most of the benefits of a one-layer solution with far less cost. Second, local problems by their nature involve relatively small numbers of nodes. Consequently, many of the optimization problems will be of a scale such that analysis can lend practical insight, and for which the optimization problems [64] will at least be computationally feasible.

3. LOCAL COOPERATIVE PROCESSING

3.1 Cooperative Group Selection

To enable this hierarchical strategy, some procedure is required for determining in the first place which collection of nodes will be involved in cooperative decisions. Since for network scalability the most intensive communications will need to be local, careful design at this level can have a large impact on the overall network usage. Two components are required: selection of the group, and then cooperative processing and communication within this group. By breaking down the large network optimization problem so that most processing will involve small collections of nodes it is possible to more readily solve the computationally difficult combined processing/communication problem, in the sense that non-scalable brute force techniques become feasible. While this procedure is in general highly suboptimal, when signals decay strongly with distance it is only the nodes in the general neighborhood of the phenomenon whose sensors will produce readings with significant information about the source. This is true both for primary sensors and communications receivers. Thus sensor network problems naturally lead to a consideration of the selection of some subgroup of nodes that are sufficient to meet the fidelity constraints.

Figure 6 illustrates one scenario for subgroup selection. Lines of constant value (e.g. temperature) for a distributed phenomenon are shown. Where the lines are far apart, readings of adjacent sensors will be highly correlated and thus contribute little additional information, as compared to sensors in regions of rapid change. Thus a lower sampling density is required in order to reproduce the phenomenon at the desired spatial fidelity. Consequently, resources can be conserved by taking readings according to a variable sampling density over the study area, e.g., by activating only the darker nodes [65,66]. This produces a scalable approach to data collection.

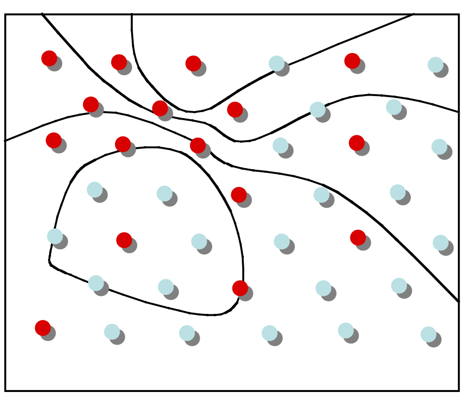

Figure 6. Adaptive sampling to meet fidelity constraint

One can alternatively pursue a source coding strategy that exploits this local correlation. The highest correlation will be with the nearest neighbors, and thus the greatest gains in source compression will occur with groups that cooperate in local neighborhoods. A hybrid strategy may also be considered in which some selection of nodes is made to make correlations among the nodes to be below some threshold, with local cooperation to achieve reconstruction with fewer nodes than would be required in a pure selection strategy.

In another scenario, nodes may be deployed with some fixed density, and the goal of selection will be to find the subset that can achieve the desired fidelity with minimal resource usage. Through cooperation, it may be possible to meet these fidelity constraints over a larger fraction of events, or to deploy at some lower density. We now consider some means by which nodes can cooperate.

3.2 Data Fusion

A variety of criteria have been used for making decisions or estimates based upon the data observed by one or more sensors [67]. Among these are the maximum likelihood, maximum a posteriori (MAP), minimum mean squared error and Neyman-Pearson criteria. Bayes methods are perhaps the most influential. They amount to choosing the hypothesis or estimate that maximizes the posterior probability, weighted by some cost function. Other criteria are resorted to when the information required for using Bayes methods are absent (e.g. prior probabilities). Thus while in the sequel the focus is upon

Bayes methods, it should be recognized that there are many practical circumstances in which other approaches are mandated. A broad overview of methods for fusing data from multiple sensors is found in [68]. Decentralized data fusion and connections to information theory are explored for example in [69-71]

We now consider a number of ways to implement local cooperation, selection of the "best" node being a limiting case. Other strategies can lead to meeting the desired performance targets at lower sensor node densities, or alternatively, meet performance targets with higher likelihoods at a given deployment density. Selection is part of a family of weighted voting schemes for making decisions. In its case, a weight of one is assigned to the best (e.g., highest SNR or greatest likelihood of correct decision) node and weights of zero to the others. This is a reasonable approach when the best node is far more advantaged than the others. Majority logic on the other hand assumes the nodes involved in the decision are all similarly advantaged, assigning equal weight to all members of the group. Maximal ratio combining provides for maximum likelihood decision-making when the only impairment is additive noise, and when the channel between source and receiver can be modeled by a single complex coefficient. Let the multiplicative coefficient for channel k be $\alpha_k e^{-j\theta_k}$ and the noise added be denoted by w_k. If the noise variance is the same for each channel, and noise is independent from one channel to another, then the received signals x_k are combined to form the decision variable

$$x = \sum_{k=1}^{N} c_k x_k = \sum_{k=1}^{N} \alpha_k e^{j\theta_k} (\alpha_k e^{-j\theta_k} s + w_k) = \sum_{k=1}^{N} \alpha_k^2 s + \alpha_k e^{j\theta_k} w_k \qquad (13)$$

where s is the signal [72]. Notice that the signals add up in phase with this choice of c_k, with more weight given to signals that were transmitted over channels with high gain. That is, signals transmitted over channels with better SNR get more weighting. It amounts to a coherent combination of signals that maximizes the SNR in the decision variable. To generalize to the situation of unequal noise variance in the various channels, the weighting would be adjusted so that SNRs follow the same ratio as in the above maximal ratio combining formula.

When multiple sensing modes are used for detecting or estimating sources it is often unclear how the data should be fused for increased decision reliability. For example, audio and video signals can both give information on the location of a moving object, but the signal processing involved is very different (e.g., coherent beamforming for the audio signal, change detection in combination with geometric optics for the video signals). The information theoretic approach is to maximize the mutual information between the fused data and the source. This principle may also be applied to the problem of node cluster selection for source identification or localization [73,74]. We show following [75] the equivalence of maximizing mutual information to the Bayes criterion.

Let $\boldsymbol{Z}^r$ denote a set of observations $\{z(1), z(2), \ldots z(r)\}$ to time r. Then the recursive form of the Bayes estimator is

$$f_X(x \mid \boldsymbol{Z}^r) = \frac{f_{\boldsymbol{Z}^r}(z(r) \mid x) f_X(x \mid \boldsymbol{Z}^{r-1})}{f_Z(z(r) \mid \boldsymbol{Z}^{r-1})} \qquad (14)$$

These calculations may be carried out at multiple nodes, either using information from individual sensors or the results of coherent combining such as from maximal ratio combining for groups of sensors of the same type. Weighting of the results so that the final result is the overall posterior probability can be accomplished using the likelihood opinion pool estimator, depicted in Figure 7. It implements the recursive fusion equation,

$$f(x \mid \boldsymbol{Z}^r) = \alpha f(x \mid \boldsymbol{Z}^{r-1}) \prod_j f(z_j(r) \mid x) \qquad (15)$$

where α is a normalizing constant. An information theoretic interpretation of this equation may be obtained by taking logarithms and then expectations on both sides:

$$\mathrm{E}\{\ln[f(x \mid \boldsymbol{Z}^r)]\} = \mathrm{E}\{\ln\{\alpha f(x \mid \boldsymbol{Z}^{r-1})]\} + \sum_j \mathrm{E}\left\{\ln\left[\frac{f(z_j(r) \mid x)}{f(z_j(r) \mid \boldsymbol{Z}^{r-1})}\right]\right\}. \qquad (16)$$

The terms can be interpreted as the posterior information being equal to the prior information plus the observation information. Thus maximization of the posterior probability will also maximize the mutual information.

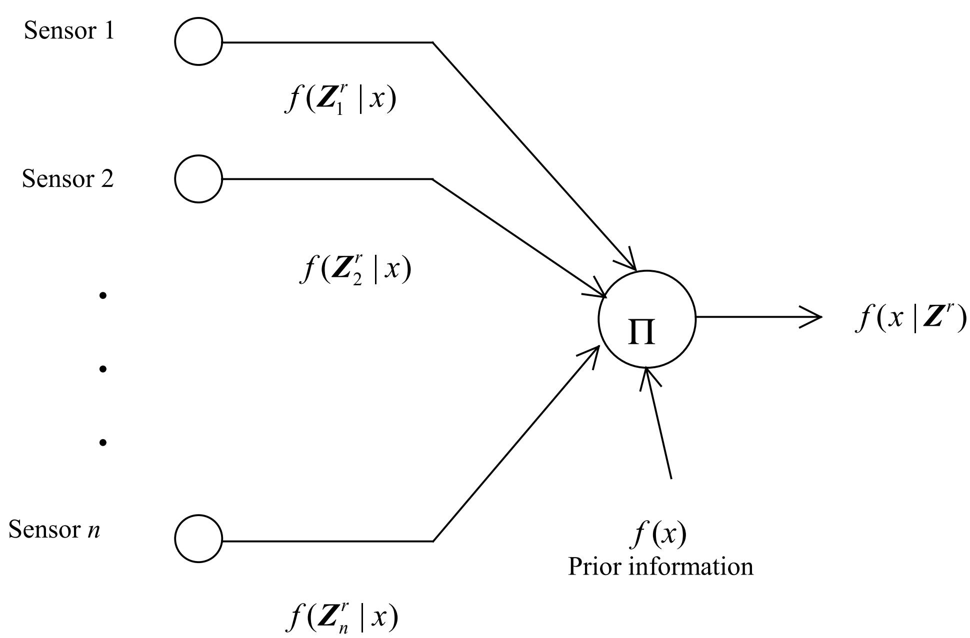

Figure 7. Likelihood opinion pool

If the prior probabilities are known, this provides a clear answer to the question of what processing to do for each sensing mode and how to weight the outputs. Any procedure that results in computation of the posterior probabilities will allow optimal combining according to the criterion of maximizing mutual information. In practice, some approximation to this procedure may be used when the relevant quantities are unknown or too difficult to compute exactly.

3.3 Fundamental Limits in Local Cooperation

In this section we examine the question of the minimum resources required in making some fusion decision. Consider for example the problem of a cluster of nodes that observe some source, with a node in their neighborhood designated as the central node (CN) to which the processed data must be reported. A variety of algorithms may be used to elect this node [76,77]. A simple question to pose is the minimum number of bits that must be sent by the CN to describe the source to a given level of fidelity. Suppose the source is Gaussian and perturbed by Gaussian noise. The local sensors send measurements to the CN which are weighted using maximal ratio combining to achieve an SNR larger than the desired fidelity. In the case of a mean squared error distortion criterion, the minimum rate is then given by the Gaussian rate distortion function,

$$R(D) \geq \frac{1}{2}\log_2 \frac{\sigma_x^2}{D} \text{ bits/symbol} \qquad (17)$$

where σ_x^2 is the variance of the (fused) signal variable sequence. This is the minimum rate at which data must be sent to a user to enable reconstruction of the source to the desired fidelity.

Consider for example a scenario of free space propagation losses (power dropping as the inverse square of distance) and uniform deployment of nodes. All nodes at a distance less than a quantity denoted as the fusion radius will participate in fusing their information to obtain a better SNR. It is readily shown that each time this radius is doubled the expected SNR from fusion will increase by 3 dB. One fusion approach is for all nodes within this region to send information at the same resolution. However, nodes that are further away will have in general a lower SNR and thus less mutual information to the source. Thus even with completely independent readings they may convey their measurements at lower resolution than sensors closer to the source. If correlation is taken into account the number of bits can be lowered further.

The formal question is the minimum set of local rates, or more practically, the minimum energy for communicating data to fuse it so that a distortion of less than or equal to D is obtained. This involves (1) identifying the set of sensor nodes involved in fusion (2) determining what kind of fusion will take place and (3) establishing the fusion graph (who communicates with who at what rates, and what fusion operations take place along the way). Independent coding does not permit exploitation of correlation in the measurements at the different sensors. Thus it is of interest to bound how much correlation can help. This is in general a very difficult problem to solve, and as usual with information theoretic problems demands that relatively long sequences of data are sent so that overhead to learn the correlations can be neglected. We now consider a number of versions of this problem.

3.3.1 ***Lossless source coding***

One distortion limit is D=0, i.e., perfect reproduction. In the distributed lossless source coding problem, the goal is to reproduce the sources represented by random variables X and Y with minimum transmission rate, given the sources may be correlated. The nodes that observe X and Y need only know the joint distribution $f\ (x,y)$ in order for the following rate region to be achievable using Slepian-Wolf coding [58]:

$$R_1 \geq H(X \mid Y) \tag{18}$$

$$R_2 \geq H(Y \mid X) \tag{19}$$

$$R_1 + R_2 \geq H(X,Y) \tag{20}$$

where H is the entropy, and R denotes the rate from a given node. That is, the total number of bits transmitted is as small as if each of the individual sensors had observed the other sequence. The theorem extends easily to m sources so that the conclusion remains that the total rate is lower bounded by the joint entropy [4]. There are many practical methods that come close to this bound in various scenarios. Lossless source coding of this type is of interest in practice when all the bits resulting from A/D conversion are thought to be significant (e.g., high SNR)

A related problem is source coding with side information [78]. Again the variables X and Y are independently encoded, but only source X needs to be recovered, with the information from the second source being viewed as side information. If X is encoded at rate R_1 and Y at rate R_2, then X can be recovered with an arbitrarily small probability of error if and only if

$$R_1 \geq H(X \mid U) \quad (21)$$

$$R_2 \geq I(Y;U) \quad (22)$$

for some joint probability mass function $f(x,y)f(u \mid y)$, where $|\boldsymbol{U}| \leq |\boldsymbol{Y}| + 2$, with $u \in \boldsymbol{U}, y \in \boldsymbol{Y}$. This problem arises in practice when two sensors are observing the same phenomenon and we are interested only in reproducing the higher SNR set of measurements. The total information rate can be made lower than in the Slepian-Wolf coding problem through careful selection of U.

3.3.2 *Lossy source coding*

The problem of coding with side information was first extended to lossy compression (that is, with a fidelity criterion) by Wyner and Ziv, who found the relevant rate distortion function [79]. A variety of other problems have since been formulated [57, 80-86]. In many cases the problems are analytically intractable absent the assumption of Gaussian sources. Nevertheless they provide insight into the types of limits that will arise in network rate distortion coding. In the sequel we consider two such problems: the n-helper and Gaussian CEO problems.

Consider the problem of a cluster of n+1 nodes that observe some source, as shown in Figure 8 for n=3.

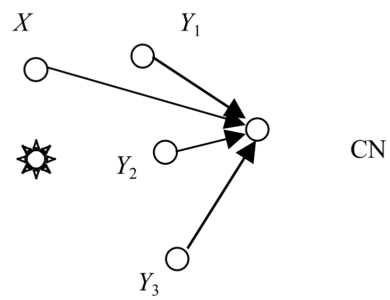

Figure 8. The *n*-helper problem

After a boot-up procedure to identify the nodes that have observed the source with useful SNR, one then picks the CN and the node X with the highest SNR, with the remaining n nodes being "helpers" that generate correlated data. The question is then the minimum data rate from X required to reproduce the data observed at X at the CN to the desired level of fidelity at some distant location (e.g., the gateway), using the data streams of rates R_k from the helper nodes as side information. Note that a combined rate of up to

$$R = R_X + \sum_{k=1}^{n} R_k \qquad (23)$$

is sent to the CN to represent the observations to the desired level of fidelity. In this particular problem, the noise in the measurements is neglected, with all distortion being due to constraints on the rate of communication.

For Gaussian sources, this can be formulated as a rate distortion problem following [87]. The covariance matrix of the observations (X, Y_1,…,Y_n) is given by

$$\boldsymbol{R} = \begin{bmatrix} \sigma_X^2 & \rho_{XY_1}\sigma_X\sigma_{Y_1} & \cdots & \rho_{XY_n}\sigma_X\sigma_{Y_n} \\ \rho_{XY_1}\sigma_X\sigma_{Y_1} & \sigma_{Y_1}^2 & & \rho_{Y_1Y_n}\sigma_{Y_1}\sigma_{Y_n} \\ \vdots & & \ddots & \vdots \\ \rho_{Y_nX}\sigma_{Y_n}\sigma_X & \rho_{Y_1Y_n}\sigma_{Y_1}\sigma_{Y_n} & \cdots & \sigma_{Y_n}^2 \end{bmatrix} \tag{24}$$

The distortions of the sources Y are not of interest, and so the goal is to choose the rate region to minimize R_X given distortion D_X. That is, neglecting the cost of the data streams from the helpers, how small can R_X be if the correlations are exploited? This bounds how much exploiting this correlation can help, since in a fusion problem we would also have to minimize the aggregate rate R, and not just R_X. This problem may be approached by considering the information provided by Y_k given Y_{k-1}, Y_{k-2}, …,X as being provided by some new source. In this way the problem at each step is similar to a one-helper problem. It is found that

$$R_X(D_X) \geq \frac{1}{2}\log^+\left\{\frac{\sigma_x^2}{D_X}\prod_{i=1}^{n}\left(1-\rho_i^2\right)\Gamma_i\right\} \tag{25}$$

where $\Gamma_i = 1-\rho^2_{XY_i|(Y_1,\ldots,Y_{i-1})} + \rho^2_{XY_i|(Y_1,\ldots,Y_{i-1})}2^{-2R_i}$, $\rho_i^2 = \dfrac{\sigma^2_{X|Y_1,\ldots,Y_{i-1}}}{\sigma_X^2}$ and $\log^+(x)=\max(\log(x), 0)$.

This collapses to the rate distortion function for a single Gaussian source in the absence of helpers. It is found in practical situations that a relatively small number of helpers provides most of the improvement in R_X, when the propagation loss exponent is two or greater. However, the aggregate rate R is always greater than if a single sensor were to observe the source, since X is chosen as the one with the best SNR. Viewed another way, the value of helpers is to extend the achievable rate region to lower distortion values.

The n-helper problem arises in practice in sensor networks when there is a higher cost to communicate from some particular sensor node than others, thus providing incentive to get a higher aggregate rate by using the lower cost links. The question of the best aggregate rate is answered in the separate coding problem. This is also illustrated in Figure 8, where now the Y's are assumed to be peers of X and the problem is to minimize the total rate R, again considering all distortion is due to the limited rate rather than noise. The n-helper results are useful in proving an upper bound on $R(D_0,D_1,...,D_n)$, where the distortion vector is the difference between the sequence observed at $(X,Y_1,...,Y_n)$ and that reconstructed at the fusion center. Here as well each sensor node is aware of the correlation matrix, so that separate coding differs from independent coding in exploiting the correlation. The fairly complicated general expression simplifies in the case of two nodes to

$$R(D_0,D_1) \geq \frac{1}{2}\log\left[(1-\rho^2)\frac{\sigma_X^2\sigma_Y^2}{D_X^2 D_Y^2}\right] \qquad (26)$$

The separate coding problem is generalized with the Gaussian CEO problem [84,88], in which correlated measurements are again made but now the independent noise cannot be neglected. The fusion center or CEO then attempts to reconstruct the source. Here again the sensor nodes only need to know the correlation matrix to achieve gains compared to independent encoding. When the issue is communications resources as opposed to pure rate (e.g., energy to support transmission at some particular rate), the same machinery can operate to allow optimization. The general form of the problem is illustrated in Figure 9.

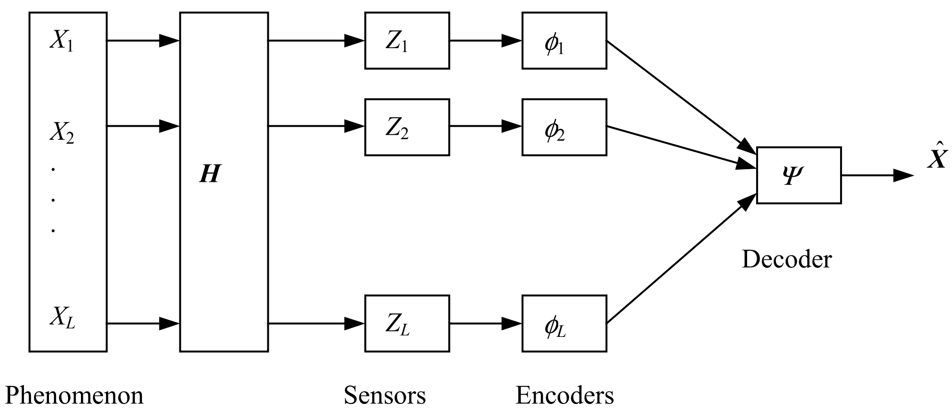

Figure 9. The distributed Gaussian CEO system

Let $\{X_1(t),...,X_L(t)\}_{t=1}^{\infty}$ represent an L-dimensional source, in which the sequences are assumed to be memoryless and stationary (with respect to time). The vector at any given instant t is modeled as a zero-mean Gaussian random variable $\boldsymbol{X}(t)$ with a non-singular covariance matrix $\boldsymbol{R_X}$. The channel, producing attenuation between each source and each sensor is designated by the positive definite matrix $\boldsymbol{H}$. The sensor readings $Z_i(t)$ additionally include noise so that we may write

$$\boldsymbol{Z} = \boldsymbol{HX}+\boldsymbol{N} \quad (27)$$

where $\boldsymbol{N}$ is an additive white zero mean Gaussian noise vector with covariance matrix $\boldsymbol{R_N}$. If at this point the problem were merely to fuse the sensor data then the techniques of section 3.1 would apply. Here however the idea is minimize transmission rate subject to the distortion constraint, where the sensor nodes have knowledge of the joint distribution. Data sequences from the sensors are therefore independently encoded according to their separate encoder functions ϕ_i producing transmitted rates R_i. From all of these sequences the decoder produces an estimate of the original source vector, with the objective that the distortion is less than some value D, while minimizing the sum of the rates R. It may be shown that [89]:

$$R(D) \geq \frac{1}{2}\log^{+}\left\{ \frac{|\Theta \boldsymbol{H}^T \boldsymbol{R}_N^{-1}|^2 \, |\boldsymbol{R}_N| \left[\prod \lambda_i\right] |\boldsymbol{R}_X|}{|\Theta| \left(\frac{D}{L} |\boldsymbol{R}_X|^{1/L} - \left[\prod \lambda_i\right]^{1/L} |\Theta|^{1/L} \right)^L} \right\} \tag{28}$$

where $\Theta = \left(\boldsymbol{R}_X^{-1} + \boldsymbol{H}^T \boldsymbol{R}_N^{-1} \boldsymbol{H}\right)^{-1}$, and the λ_i are the eigenvalues of $\boldsymbol{R}_X$. While not immediately obvious from this expression, higher correlation values among the measurements result in lower required total rate for a given distortion. As noise increases, the required rate also increases and moreover the rate region itself shrinks as low values of distortion are no longer achievable. In the absence of noise, (or more practically, very high SNR) the problem collapses to that of separate encoding.

These types of problems establish useful bounds against which practical algorithms can be tested under differing assumptions on SNR and source correlation. For example, in the adaptive fidelity problem, a sensor field is thinned so that the measurements are nearly uncorrelated. (This terminology originated in the related problem of thinning communications relays to conserve energy [90]). In the context of the above discussion, the strategy is to set the rates R_k to zero when the cross-correlation is high, accepting some loss of fidelity in reconstructed measurements in exchange for lower rates. Thus the adaptive algorithms in this domain can be compared to the rate distortion limits to see how closely they approach the optimal (Gaussian CEO) strategy. Similarly, algorithms that explicitly include cooperation among nodes can be compared, and issues such as the number of sensors that should cooperate (and how they should cooperate) can be studied in a rigorous context.

4. COMBINED ROUTING AND PROCESSING

While the solution to the overall sensor network optimization problem, which requires a complete network information theory, remains a distant goal, various suboptimal versions of the problem have been considered recently. One of these is to jointly design data fusion and routing such that some objective function is optimized. A key issue in any such approaches is consideration of the correlation of the data among the nodes.

Many energy-aware routing algorithms for wireless sensor networks have been proposed and analyzed. Some research focuses on judiciously turning on and off nodes such that energy consumption is minimized while at the same time the network connectivity is maintained [91-93]. Some researchers proposed energy-aware routing algorithm such that the routes preferentially use nodes with more residual battery power, in order to maximize network lifetime [94,95]. Concerns over the sacrifice of QoS have prompted some to propose QoS based energy-aware routing [96]. More relevant to the combined routing and processing problem is the joint optimization of data aggregation and routing structure, which has been explored in [97-100]. In particular, [97] proposed a data-centric routing paradigm for sensor networks; [98] discussed the joint optimization of correlated source coding and routing structure, which is restricted to a tree. It was recognized that under lossless source coding the two subproblems, source coding and routing , are decoupled. For joint optimization of data fusion and routing, cluster-based algorithms have been one of the major solutions discussed for example in [101-103].

Suppose now that some local fusion strategy is pursued (e.g., none at all, or something elaborate), and there remains some residual correlation in the results among the clusters so formed. A natural question is then how to combine routing and processing to minimize resource usage. Before plunging into the problem formulation, we first review some concepts and notation in graph theory, network information theory and source coding.

4.1 Graphs

Networks are often modeled as graphs where the nodes represent sensors and the edges indicate the connectivity among sensors [104]. Often the edge is assigned some weight to indicate the cost of transmitting along the link. Sometimes the network is directed, meaning transmission can only go one-way, or different costs are associated with the communications in the two directions (e.g., due to different interference conditions or asymmetric costs of using transmitters).

Network flow theory has been one of the many successful applications of mathematical programming [105]. Such problems concern how to route the flows in the network such that an objective function is optimized. Many routing algorithms are based on these results, including the max-flow min-cut shortest path algorithms.

4.2 Network Information Theory

Network information theory answers the following question: given many senders and receivers, a power budget, and a channel transition matrix that describes the interference and noise, determine whether the (what kind of) sources can be transmitted across the channel and recovered subject to a distortion constraint [106]. The problem can also be formulated using power budget as a variable: given the source distribution, whether the (what kind of) power budget can achieve the same goal. It is usually raised as a decision type of question. Asked in parentheses are the corresponding questions of determining the sets of admissible sources and power allocations. Use the idea of time sharing, it can be shown that the admissible sets of source and power allocation are convex. For instance, if $P = (P_1, P_2, \cdots, P_n)$ and $Q = (Q_1, Q_2, \cdots, Q_n)$ are two admissible power allocations, then $\lambda P + (1-\lambda)Q$ is admissible as well.

4.3 Data processing

Data fusion at sensor nodes can involve source coding or the computation of likelihood functions. Here, we will deal mostly with source coding. As a suboptimal approach, we will consider source and channel coding separately. Even with this sacrifice of optimality, the problem is still in many cases NP-hard. In general, the coding problem in sensor networks is a lossy one, but at high fidelity it is approximately treated as lossless coding. That is, after some subset of the nodes have been selected as having useful data, the information they generate must all be retrieved. Sometimes, sensor nodes carry out data fusion, in which only a likelihood function needs to be computed or a decision made. In this case, the data rate is very dependent on what type of data fusion is carried out. Many researchers have chosen to use a constant data rate. In other words, as soon as the data are fused, the data rate required to communicate with the fusion center is given as a constant. Under this assumption, clustering algorithms have been extensively investigated for use in local fusion.

4.4 Problem Formulation

The problem we address is what is the *optimal* energy required for the sensor network to transmit observed data to the fusion center such that it either can reproduce these data according to a fidelity constraint or make reliable decisions, given a network topology, channel transition matrix, source distribution, observation noise, and the definition of estimation fidelity or decision reliability. We prefer the term optimal energy rather than minimum energy because in a distributed system the collective minimum is not necessarily the optimal solution. For example, minimizing aggregate energy may cause an uneven distribution of energy load in the network, and drain some sensor nodes' batteries more quickly than others. This can result in parts of the network becoming disconnected. Hence, in the next section, we will discuss cost functions of both aggregate and distributed natures, and consider not only energy consumption but also data rate, delay, and battery lifetime.

4.4.1 *Cost functions*

Various cost functions have been formulated for sensor networks. One cost function based upon the aggregate data rate is

$$C = \sum_{(i,j)\in E} f_{ij} \tag{29}$$

in which f_{ij} is the data rate across link (i, j). This cost function assumes that all the links have equal capacities. This is however often not the case as the capacity of the wireless link is affected by distance, fading, node transmission power etc. Suppose the capacity of link (i, j) is given as u_{ij}. A normalized cost function can be formed to take this into account:

$$C = \sum_{(i,j)\in E} f_{ij} / u_{ij} \ . \tag{30}$$

This amounts to placing weights on communication links.

A cost function for the aggregate energy consumption is

$$C = \sum_{(i,j)\in E} c_{ij} f_{ij} \tag{31}$$

where f_{ij} is the data rate across link (i, j), and c_{ij} is energy per bit on link (i, j). Here it is assumed that on each link consumed energy is a linear function of data rate. This is a reasonable approximation to reality if increased data rate is due to increased radio operation time. However, if the data rate increase is done by switching modulation schemes, energy consumption on each link cannot then be thought of as a pure link function of data rate. Moreover, state of the art radios can possess multiple antennas, and operate on different sets of sub-channels. Therefore, a better characterization of energy consumption is a piece-wise linear function of the data rate:

$$C = \sum_{(i,j)\in E} \left(c_{ij}^1 f_{ij}^1 + c_{ij}^2 f_{ij}^2 + \cdots + c_{ij}^n f_{ij}^n \right). \qquad (32)$$

The data rate across link (i, j) is given by $\sum_{k=1}^{n} f_{ij}^k$. An example of such a piece-wise linear function is depicted in Figure 10. If the coefficients satisfy $c_{ij}^1 < c_{ij}^2 < \cdots < c_{ij}^n$, which is the case in the example, the cost function is a convex function of data rate, and it can be represented by multiple links with a linear cost function.

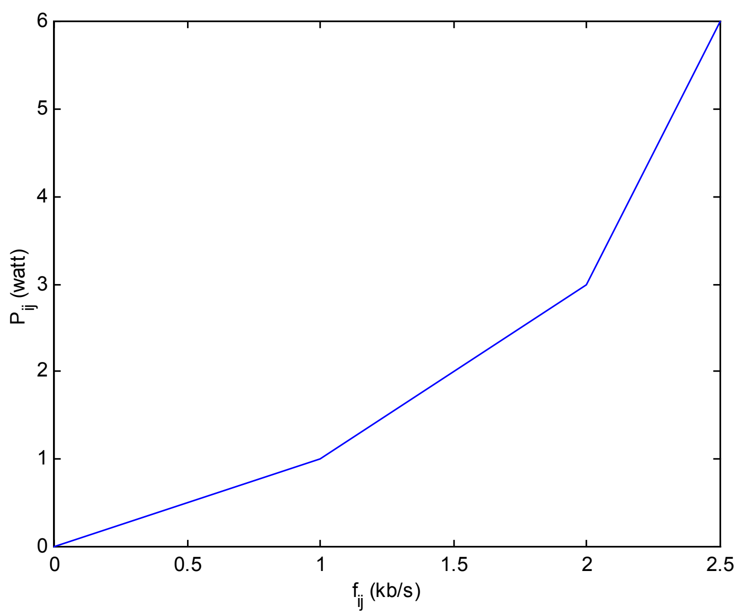

Figure 10. Power consumption (Watts) as a function of data rate (kb/s) on link (i, j).

Besides the nonlinearity, additional energy may be required to power up the circuit. This represents a non-zero initial cost when sensors switch from the stand-by mode to a transmission mode. This can be modeled by a binary variable x_i defined as

$$x_i = \begin{cases} 1 & \text{power on} \\ 0 & \text{standby} \end{cases}$$

$$C = \sum_{(i,j)\in E}\left(c_{ij}^1 f_{ij}^1 + c_{ij}^2 f_{ij}^2 + \cdots + c_{ij}^n f_{ij}^n\right) + \sum_{i\in N} c_{ij}^0 x_i \qquad (33)$$

where c_{ij}^0 is the additional energy needed to power up the circuit.

More insight can be gained by looking at the energy consumption at individual nodes rather than links, and the processing power can be taken into account as well. The energy consumed at sensor i is

$$E_i = c_i^t \sum_{j\in O(i)} f_{ij} + c_i^r \sum_{j\in I(i)} f_{ji} + c_i^p b_i^c + c_i^0 x_i \qquad (34)$$

where c_i^t and c_i^r are the transmission and reception powers per bit at node *i*, c_i^p is the processing cost (power) per bit at node *i* (i.e. the amount of processing power to reduce one bit of data rate), and b_i^c is the number of bits being compressed due to processing. Using flow conservation, this term disappears in equation (35). Hence the total energy is

$$\begin{aligned} C &= \sum_{i\in N} E_i \\ &= \sum_{(i,j)\in E}\left(c_i^t + c_j^r + c_j^p - c_i^p\right) f_{ij} + \sum_{i\in N} c_i^0 x_i \\ &= \sum_{(i,j)\in E} c_{ij} f_{ij} + \sum_{i\in N} c_i^0 x_i \end{aligned} \qquad (35)$$

Another metric of interest is delay. If the capacity of link (i, j) is given as u_{ij}, the queuing delay at this link can be modeled as

$$\tau_{ij} = \frac{\alpha f_{ij}}{u_{ij} - f_{ij}} \tag{36}$$

Ignoring propagation and processing delay, we can formulate the aggregate delay of the network as

$$C = \sum_{(i,j)\in E} \tau_{ij} \tag{37}$$

Many researchers have discussed algorithms to maximize the battery lifetime of sensor nodes. The lifetime of battery can be modeled as a function of incoming and outgoing data rate:

$$t_i = t_i\left(f_{ij}, f_{ki}, j \in O(i), k \in I(i)\right) \tag{38}$$

The strategy is then to route more messages through the sensor nodes with higher battery power so that the system lifetime is lengthened.

As mentioned earlier, aggregate values may not be a good indication of optimality. Often we need to consider the min-max and max-min optimization. For example, we may want to minimize the maximum energy consumption or maximize the minimum lifetime of sensor nodes.

$$\begin{aligned} &\min C \\ &\min \max E_i \\ &\min \max \tau_{ij} \\ &\max \min t_i \end{aligned}$$

4.4.2 *Complete problem*

Suppose there are n sensors and a fusion center in the network G. Each sensor produces a stream of data X_i, (X_i can be empty, which means no data is produced) and transmit them to the fusion center through the network. The data are the result of observing some physical phenomenon, and they satisfy some ergodic condition so that the results of statistical probability theory can be applied here. The goal is to optimize the given cost function C while recovering these data streams $(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_n)$ subject to a distortion constraint $d(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_n) \le D$. For the particular instance of minimizing the aggregate power, the problem can be formulated as follows:

$$\min \sum_{i=1}^{n} P_i \qquad (39)$$

$$\begin{aligned} &\text{subject to}: (P_1, P_2, \cdots, P_n) \text{ is admissible for the network } G \\ &\qquad\qquad \text{i.e. } d(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_n) \le D \text{ is achievable} \end{aligned}$$

As we have indicated before that the set of admissible power allocations is convex, this constitutes a convex optimization problem. Fully solving this optimization problem requires network information theory, which remains a distant goal, so instead we take a suboptimal approach: consider channel coding, source coding, and routing separately.

4.5 A Divided Approach

Channel coding can be represented by the link weight and capacity. This can be interpreted as the amount of power that is required to achieve some particular transmission capacity given the channel transition matrix. If the source has only a discrete alphabet, the optimal network source coding scheme is given by Slepian and Wolf [58]. However, in general, the source coding in sensor networks is lossy. Network rate-distortion coding is an open problem, so we consider an explicit side information transmission scheme. In other words, only when the side information is known at both the encoder and decoder will it be used to reduce the data rate. The resulting problem is how to choose the appropriate network routing structure based on the network topology and source correlation such that the given cost function is optimized. Even with such a large simplification, the problem is still difficult to solve. For example, consider the network configuration in Figure 11.

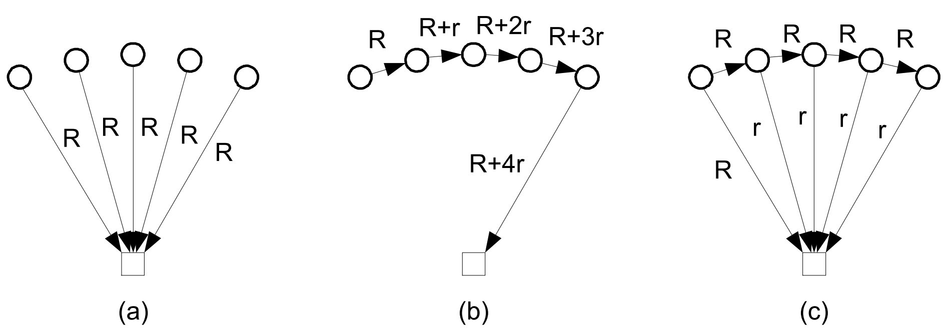

Figure 11. An explicit side information example.

Five sensor nodes (circles) are equally spaced on an arc centered at the fusion center (the square). The edges between adjacent sensors have weights d, and the edges between sensors and the fusion center have weight D ($d << D$). Suppose we use an explicit side information model. Without any explicit side information, the data rate at each sensor is

R. If there is side information from an adjacent sensor the rate is r ($r < R$). When the shortest path algorithm is used, the total cost is

$$E_a = 5RD \quad (40)$$

If the tour in (b) is used, the total cost is

$$E_b = RD + 4Rd + 4rD + 6rd \quad (41)$$

If the help information and the data going to the fusion center are transmitted separately as depicted in (c), the total cost is

$$E_c = RD + 4Rd + 4rD \quad (42)$$

Which scheme is optimal depends on specific values of $r, R, d,$ and D. When r is zero, (b) and (c) are optimal. If r is the same as R, (a) has the lowest cost. In real networks, the configuration can in general be much more complicated and the problem more difficult.

To formulate the optimization problem, we first need a data reduction model, in other words, a model to capture the effect of explicit side information. For lossy source coding, we turn to a practical algorithmic approach: a spatial-temporal DPCM coding scheme. [107] The system block diagrams of the encoder and decoder are given in Figures 12 and 13.

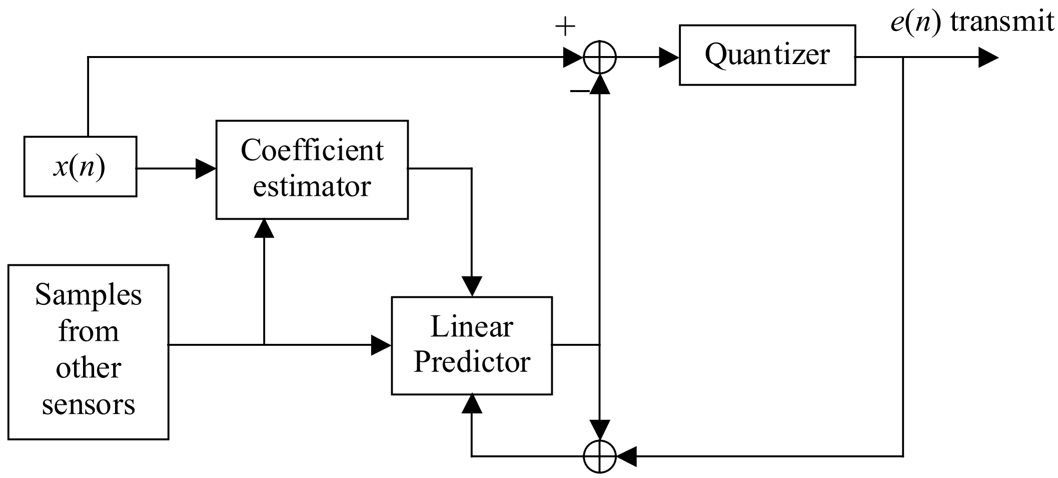

Figure 12. A spatial-temporal DPCM encoder

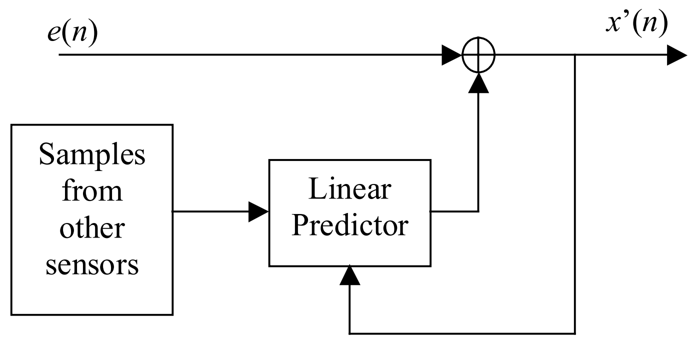

Figure 13. A spatial-temporal DPCM decoder

The data rate reduction at a sensor node is a function of source correlation and available data from neighboring sensors. In practice, this can be estimated by sending a sample sequence around, and running the DPCM coder to see how much the data variance can be reduced. The data reduction at sensor nodes is certainly a complicated nonlinear function of incoming side information. To make the problem tractable, we model it as a piecewise linear function of the number of helping sensors as depicted in Figure 14. A few assumptions were made in this model. First, adding information from one more sensor brings diminishing reduction in the data rate, and after four helping sensors the reduction saturates. Second, side information from each helping sensor has the same data reduction effect, so only the number of helping sensors is used as a variable instead of which sensor comes to help.

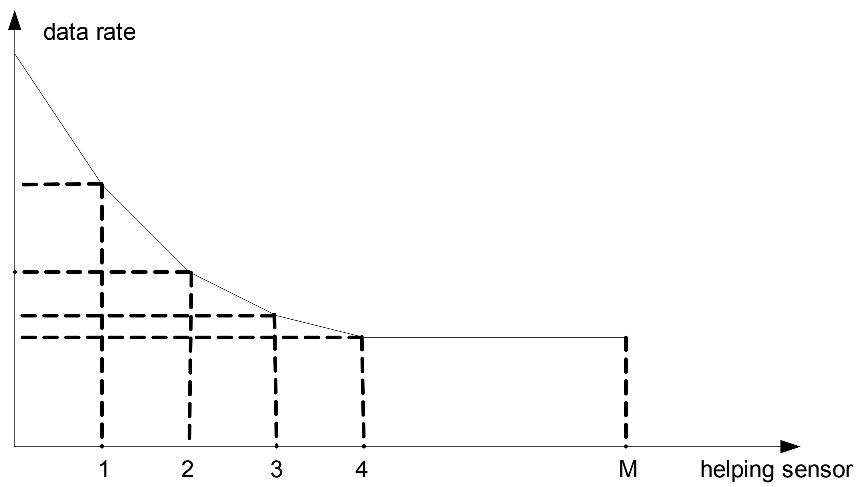

Figure 14. Data rate with help information.

If the routing structure is prescribed to be a tree, the problem can be formulated as a mixed integer programming problem:

$$\min \sum_{e \in A} g_e w_e \tag{43}$$

subject to

$$0 \le g_e \le u_e y_e; \tag{44}$$

$$\sum_{e \in \delta^+(i)} g_e - \sum_{e \in \delta^-(i)} g_e = b_i, \quad e \in A, \ i \neq r; \tag{45}$$

$$b_i = \sum_{j=0}^{J-1} b_i^j \lambda_i^j; \tag{46}$$

$$\sum_{k \in H(i)} \sum_{e \in \delta^+(i)} f_e^k = \sum_{j=0}^{J-1} h_i^j \lambda_i^j; \tag{47}$$

$$\sum_{j=0}^{J-1} \lambda_i^j = 1, \ \lambda_i^j \geq 0; \tag{48}$$

$$\sum_{e \in \delta^+(r)} f_e^k - \sum_{e \in \delta^-(r)} f_e^k = -1, \quad k \neq r; \tag{49}$$

$$\sum_{e \in \delta^+(i)} f_e^k - \sum_{e \in \delta^-(i)} f_e^k = -1, \quad k \neq r; \tag{50}$$

$$\sum_{e \in \delta^+(v)} f_e^k - \sum_{e \in \delta^-(v)} f_e^k = 0, \quad v \neq d, \ v \neq r; \tag{51}$$

$$f_e^k \leq y_e, \quad e \in A; \tag{52}$$

$$\sum_{e \in A} y_e = n - 1; \tag{53}$$

$$f_e^k \geq 0, \ g_e \geq 0, \ y_e = \{0, 1\}. \tag{54}$$

Constraint (45) is the flow conservation equation; (46)-(48) use the piecewise linear data reduction model; (49)-(53) build a directed tree pointing to the fusion center using a multicommodity flow model [108].

The relation between the afore mentioned optimization problem and other inherently difficult network problems has been identified by many researchers. For example, [98] identifies the connection to the traveling salesman problem, and [109] recognizes under certain assumptions the optimal solution is a Steiner tree. Both are well known NP hard problems.

4.6 Solutions

There are no known efficient algorithms for solving the mixed integer programming problem. Popular approaches include branch and bound and cutting plane [110], but neither of these are guaranteed to terminate in polynomial time. As a special case, when the source coding is lossless, the solution is known. With the global knowledge of source statistics, the Slepian-Wolf random binning scheme does not require explicit exchange of side information among sensors. As a result, the source coding and routing are decoupled. The problem is solved by linear programming [64], and the result is given by [98]:

$$\begin{aligned}
R_1 &= H(X_1), \\
R_2 &= H(X_2 \mid X_1), \\
&\vdots \\
R_N &= H(X_N \mid X_{N-1}, X_{N-2}, \cdots, X_1).
\end{aligned} \tag{55}$$

in which $X_1, X_2, \cdots, X_N$ are ordered such that their distances to the fusion center increase.

Here we summarize some practical algorithms that have been proposed to approximate the optimal solution.

Directed diffusion: In this data-centric routing paradigm, data generated by sensor nodes are named by attribute-value pairs. Various entities can express "interest" in different types of data by sending requests for named data. Data matching the interest are drawn towards the nodes that request them. The degree of interest decides how far the data goes. "Interest gradients" are set up to control message routing, and local processing can be organized to reduce the quantity of data that travel a long distance.

Center at nearest source: Use the source node that is nearest to the fusion center as the aggregation point. All the data are routed to this node, aggregated, and transmitted to the fusion center.

Shortest path tree: In this algorithm, each source routes its data to the fusion center along the shortest path. Wherever the paths merge, the data are combined and recoded.

Greedy incremental tree: This algorithm is an approximation to the Steiner tree problem. The routes are constructed sequentially. First, the tree consists of only the shortest path connecting the nearest source and the fusion center. Then, at each step the shortest path to the source that is closest to the current tree is added until all the sources are connected to the center.

Leaves deletion: This scheme is built on top of shortest path tree. After constructing the shortest path tree, the leaf nodes look around in their respective neighborhoods. A leaf node will switch its parent to another node if there is a cost reduction by doing this. This continues until no further improvement is possible.

Balanced SPT/TSP tree: This algorithm attempts to combine the shortest path tree (SPT) and traveling salesman paths (TSP). It constructs the shortest path subtree for nodes that are less than a radius ρ away from the fusion center. The paths for the rest of the nodes are built from the leaves of the current subtree in their subregions.

Comparisons of these suboptimal approaches have been conducted in [97,98,109]. In [97], the directed diffusion is shown to consume less energy than omniscient multicast and flooding. For some sensor fields, the dissipated energy of directed diffusion is only 60% of omniscient multicast because of in-network processing. This is expected since the latter two methods do not take advantage of data correlation. The performance of address centric (AC, no data aggregation), center at nearest source (CNSDC), shortest path tree SPTDC), and greedy incremental tree algorithms (GITDC) are compared in [109]. A few sensor nodes are selected to act as sources. Each source generates one data packet to be relayed to the fusion center. The aggregation data model is defined such that intermediate nodes in the route transmit only one packet even if multiple packets are received. In this setup, the average number of transmissions required decreases in the order of AC, CNSDC, SPTDC, and GITDC. The performance of shortest path tree without in-network data processing, leaves deletion approximation, and SPT/TSP balanced tree algorithms were investigated in [98]. An exponential correlation model and entropy coding were used, and it was shown that the shortest path tree without in-network data processing consumes the most energy, while the SPT/TSP balanced tree algorithm requires the least.

5. MOBILITY AND INFRASTRUCTURE

In contrast to static ad hoc networks in which the per-node capacity declines as the size of the network grows [47], mobile networks can actually have scalable per-node capacity. Consider for example a network in which a collection of nodes randomly move within some bounded region. Nodes within some small radius *r* of each other are assumed to be able to achieve very high speed communications (perhaps even having touched). The nodes may also store large quantities of data, so that they can relay information around the network. In a bounded region, eventually any node will get within *r* of any other and could therefore directly exchange all required information. However, a better strategy is for nodes to act as relays, as this increases the fraction of time nodes within transmission range will actually have information to exchange. Information can be conveyed to all nodes that pass within the transmission range. These nodes will only relay once to the destination node to avoid multiplication of transmissions. With nodes acting as relays in this fashion, the per node transmission capacity becomes independent of network size, while average message transmission delay depends only on the mobility time scale of nodes traversing the network [111,112]. Worst-case delay can however be very large, motivating tradeoffs between capacity and delay [113].

Directed motion of a limited number of nodes in an otherwise static network can also lead to capacity that scales with network size. In contrast to networks in which the mobile components move randomly or in some predictable but not network-controlled pattern, this enables bounded message delivery delay for any topology. Most simply, if all nodes are mobile then a source node can simply move towards the destination, achieving full-bandwidth communication with a delay that depends only on the physical dimension of the network and the speed of the mobile. More practically, some fraction of the nodes will be put in the role of longer-range transport of information and the static nodes will communicate with these nodes using single or multi-hop communication. With only single hops, and with the number of mobile nodes m being some constant fraction a of the total number of nodes n, then it may be shown that the per node capacity scales with network size and the worst-case delay remains constant [57]. A trade between capacity and transmission delay can be obtained by allowing multi-hop communications to mobile nodes, while also varying the fraction of nodes that are mobile. In this situation, it is possible to achieve delay of O(1) with constant per-node capacity even if m is a small fraction of n. This assumes of course that the routes can be perfectly computed, which is unfortunately not possible as the problem can be reduced to the NP-complete optimization known as the general pickup and delivery problem (GPDP) [114]. Thus, some over-provisioning of mobile nodes coupled with use of one of the several heuristic solutions to this problem would be needed.

Analogous behavior can be obtained for static networks by instead adding fixed long-range communications infrastructure. Wired connections or mobile nodes that physically move data act as traffic attractors, causing the mean number of hops within the lowest static level of the network to be independent of the overall network size. Since the number of hops is bounded, the capacity per node does not depend on the network size provided sufficient capacity is available in the longer-distance transport mechanisms.

Clearly there will be different actual latency values for electronic versus physical long-range transport of data, but what matters for scalability is the average number of bits that can be conveyed a given distance. In either case, if capacity is insufficient, additional layers with greater capability can be added.

Hybrid combinations of static and mobile sensing elements are also possible. One could sample a region using either type of sensor node alone. A small number of mobile nodes can patrol some region to provide dense spatial sampling. However, sampling will not necessarily be closely spaced temporally for any given location. This is by contrast easily achieved with static nodes, but larger numbers are required for dense spatial sampling. A hybrid combination where there is also some possibility of adaptively moving static nodes using the robotic elements can provide the capability of meeting fidelity constraints with fewer resources [115,116].

Finally, it may be noted that infrastructured elements are not subject to the same resource constraints as untethered nodes. The resource cost has already been paid in creating the infrastructure, and consequently it should be used to best advantage. Infrastructure components will attract processing tasks and communications. These components may also include longer range or more vigilant sensing elements, and provide services such as timing and position location. Infrastructure such as roadways or cable-lines further provides lower-energy paths for mobility, while power stations together with mobile elements provide for the possibility of automated recharging of untethered elements. The combination of infrastructure with more resource-constrained elements opens up interesting new resource optimization questions in sensing [117].

6. CONCLUSION

Wireless sensor networks were shown to be scalable to large numbers provided a mechanism exists to determine participation in decision-making based upon meeting fidelity constraints. Maximization of mutual information between source and fused decision was shown to be equivalent to use of the Bayes criterion, and thus can be one practical technique for determining which group of nodes should participate and how their decisions can be fused. While the general multi-terminal information theory problem is daunting in its difficulty, the requirements for scalability dictate that most action will take place on a local level. Consequently, information theoretic bounds become tighter and the various optimization problems that combine source coding and routing become computationally feasible. This holds promise for the discovery of practical techniques that will match the progress made in point-to-point transmission problems in the domains of both source and channel coding. Mobile elements add further richness to the set of optimization problems.

While energy efficiency can (at least in principle) be maximized by tightly coupling processing and communication, this does not necessarily imply abandonment of standard network protocols such as TCP/IP in sensor networks. When most communication is local, the marginal efficiencies achieved on the longer haul portions become small. There are in such situations many good reasons to use standard solutions. The question of scale is always important; at larger scales the situation is more akin to that of more traditional networks for which the richly developed theory and practical approaches are appropriate. Exactly how to make the transition between behaviors at some local level to this global scale remains an area of research.

GLOSSARY

Ad hoc network: a network without infrastructure, typically of irregular topology, where nodes must relay traffic of other nodes

Bayes criterion: selection of the most likely hypothesis as being the true one, or when cost functions are employed, selection of the hypothesis with the maximum product of likelihood and revenue

CEO problem: data fusion at one node in the presence of noise, with the objective of minimizing communication rates to that fusion center with distortion less than some constraint.

Channel capacity: the maximum rate of error free transmission, given the resource constraints (typically, bandwidth and power)

Channel coding: a set of methods to achieve reliable communications in the presence of noise, while meeting bandwidth and power constraints

Data fusion: combination of information from two or more sources to achieve better estimation accuracy or likelihood of correct decisions

Delay: also referred to as latency, the time taken between the start and finish of a transaction such as encoding/decoding or transmission of information across a network

DPCM: differential pulse code modulation, a method for compressing data that employs linear predictive filters to exploit correlation among successive data samples

Entropy: a measure of the information of random variables proposed by C.E. Shannon, equal to the negative of the sum of the probabilities of the events represented multiplied by the logarithms of the probabilities of those events

Fidelity constraint: a requirement by some end user that events must be distinguishable according to some combination of spatial and temporal resolution, and to within some accuracy or probability of error

Gaussian channel: a sensing or communications channel in which the only impairment is additive white (broadband and independent) Gaussian noise (that is, noise with a Gaussian distribution of sample values)

Routing: methods for computing desirable paths through networks between the sending and destination nodes, for example to optimize some combination of delay and resource usage

SNR: the ratio of the signal power to the noise power at the decision point of a communications or sensing system, often used as an indicator of the quality of reception; interference from other sources may sometimes be considered to be part of the noise

Quantization: the process of converting real numbers (e.g. analog voltages) into discrete levels to permit representation in digital machines (e.g. computers or communication receivers); a non-reversible process that introduces an error between the original and quantized values

Rate distortion function: the minimum possible rate at which some random process can be represented such that the distortion (error) between the original and reproduced process is less than some value; the function is convex with all rate/distortion pairs greater than or equal to the function being achievable by some encoding scheme.

Scalability: the ability of a network to continue to provide some service (e.g. transmission between pairs of nodes) without the addition of resources, as the number of nodes in the network grows without bound

Source coding: a set of methods to represent data at a rate lower than that of the original data, with or without reproduction error when decoding is performed (lossy and lossless coding respectively)

Synchronization: a set of methods to provide a common time base between two or more nodes

REFERENCES

[1] G. Pottie and W. Kaiser, *Comm. ACM*, 43(5), 51 (2000)

[2] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. *IEEE Comm. Magazine*, 40(8), 102, (2002)

[3] C.E. Shannon, *Bell System Tech. J.,* 27, 379, (1948)

[4] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York (1991)

[5] R. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York (1968)

[6] J. Ziv and A. Lempel, *IEEE Trans. Inform. Theory*, 23, 337 (1977)

[7] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[8] T. Berger. *IEEE Trans. Inform. Theory*, 44, 2693 (1998)

[9] T. Berger, *The Information Theory Approach to Communications*, Vienna/New York, Springer-Verlag, 1978.

[10] A. Glavieux, C. Berrou, and P. Thitimajshima, *Proc. of IEEE Inter. Conf. Comm.*, 1, 1064 (1993)

[11] G. Montorsi, F. Pollara, S. Benedetto, and D. Divsalar, *IEEE Trans. Inform. Theory*, 44, 909 (1998)

[12] E. Telatar, *European Transactions on Telecommunications*, 10, 585 (1999)

[13] C. Shannon, *Institute of Radio Engineers, Intn'l Convention Record,* 7, 142 (1959)

[14] R. Blahut, *IEEE Trans. Inform Theory*, 18, 14 (1972)

[15] M. Bystrom and J. Modestino, *IEEE Journal on Selected Areas in Communications*, 18, 880 June 2000.

[16] M. Gastpar and M. Vetterli, *Second International Workshop on Information Processing in sensor Networks*, 1, 162 (2003)

[17] A. Goldsmith and M. Effros, *Proc. IEEE Inter. Conf. Comm*. 1, 319, (1997)

[18] J. Modestino and D. Daut, *IEEE Trans. Comm.,* 27, 1644 (1979)

[19] L. Qian, D. Jones, K. Ramchandran, and S. Appadwedula, *Proceedings of Data Compression Conference*, 1, 414 (1999)

[20] M. Ruf and J. Modestino, *Proceedings of IEEE International Conference on Image Processing*, 1, 77 (1995)

[21] R. Benzel. *IEEE Trans. on Inform. Theory*, 25, 228 (1979)

[22] A. Carleial, *IEEE Trans. Inform Theory*, 29, 602 (1983)

[23] T. Cover, *IEEE Trans. Inform Theory*, 18, 2 (1972)

[24] T. Cover, *IEEE Trans. Inform Theory* , 21, 399 (1975)

[25] T. Cover and A. Gamal, *IEEE Trans. Inform Theory*, 25, 572 (1979)

[26] U. Mitra and A. Sabharwal, *Proceedings of the third international symposium on Information processing in sensor networks*, 1, 301 (2004)

[27] R. Ahlswede, *Ann. Prob.*, 1, 805 (1974)

[28] T. Cover and C. Leung, *IEEE Trans. Inform Theory*, 27, 292 (1981)

[29] E. Van der Meulen, *IEEE Trans. Inform Theory* , 23, 1 (1977)

[30] T. Flynn and R. Gray, *IEEE Trans. Inform Theory* , 33, 773 (1987)

[31] T. Gaarder and J. Wolf, *IEEE Trans. Inform Theory*, 21, 100 (1975)

[32] A. Gamal, *IEEE Trans. Inform Theory*, 24, 379 (1978)

[33] M. Gastpar, *Proc. of 37th Asilomar Conference on Signals, Systems, and Computers*, 1, 9 (2003)

[34] T. Han and K. Kobayashi, *IEEE Trans. Inform Theory*, 26, 277 (1980)

[35] T. Han and K. Kobayashi, *IEEE Trans. Inform Theory*, 27, 49 (1981)

[36] K. Housewright, PhD thesis, University of California, Los Angeles 1977.

[37] T. Berger and R. Yeung, *IEEE Trans. Inform. Theory*, 35, 228 (1989)

[38] P. Bergmans, *IEEE Trans. Inform Theory* , 19, 197 (1973)

[39] H. Liao. PhD thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[40] T. Oates, M. Prasad, and V. Lesser, *IEE Proceedings of Software Engineering*, 144, 72 (1997)

[41] J. Omura and K. Housewright, *Proc. IEEE Inter. Conf. on Comm.*, 1, 237 (1977)

[42] H. Sato, *IEEE Trans. Inform Theory*, 27, 786 (1981)

[43] J. Manyika and H. Durrant-Whyte, *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*, Prentice Hall PTR, 1995.

[44] L. Klein, *Sensor and Data Fusion Concepts and Applications*, SPIE Press, 1993.

[45] A. Hu and S. Servetto, *Proc. 2nd ACM inter. Conf. Wireless sensor networks and applications*, 1, 1, (2003)

[46] J. Elson, Time Synchronization in Wireless Sensor Networks, Ph.D. Dissertation, UCLA CS Dept. (2003)

[47] P. Gupta and P.R. Kumar, *IEEE Trans. Inform. Theory,*46, 388, (2000)

[48] J. Li et al., *Proceedings of the 7th annual international conference on Mobile computing and networking*, 1, 61 (2001)

[49] S. Toumpis and A. Goldsmith, *IEEE Transactions on Wireless Communications*, 2, 736 (2003)

[50] L. Xie and P. Kumar, *IEEE Trans. Inform. Theory*, 50, 748 (2004)

[51] S. Yi, Y. Pei, and S. Kalyanaraman, *Proc. 4th ACM international symposium on Mobile ad hoc networking & computing*, 1, 108 (2003)

[52] C. Zeng, F. Kuhlmann, and A. Buzo, *IEEE Trans. Inform. Theory*, 35, 1160 (1989)

[53] M. Gastpar and M. Vetterli, *Proc. IEEE INFOCOM 2003*, 3, 1577 (2002)

[54] Y.S. Tu and G.J. Pottie, *Proc. IEEE Vechic. Tech. Conf.,* 1, 130 (2002)

[55] V. Tarokh, A.F. Naguib, N. Seshadri, and A.R. Calderbank, *IEEE Trans. Inform. Theory,* 45, 1121 (1999)

[56] D. Aktas, H. El Gamal and M.P. Fitz, *IEEE Trans. Comm.,* 51, 854 (2003)

[57] A. Pandya. On Fundamental Limits of Scalable Sensor Networks. *Ph.D. Dissertation, UCLA EE Department,* (2004).

[58] D. Slepian and J. K Wolf, *IEEE Trans. Inform Theory*, 19, 471 (1973)

[59] S. Servetto, *Proceedings of the 40th Annual Allerton Conference on Communication, Control, and Computing*, Urbana, IL, 1, **PAGES**, (2002)

[60] A. Pandya and G.J. Pottie, *CENS Technical Report 17*, (2003)

[61] U. Kozat and L. Tassiulas, *Proceedings of the 9th annual international conference on Mobile computing and networking*, 1, 55 (2003)

[62] K. Yao, R.E. Hudson, C.W. Reed, D. Chen and F. Lorenzelli, *IEEE J. Select. Areas in Comm.,* 16, 1555 (1998)

[63] H. Luo, A. Pandya, and G.J. Pottie, *CENS Technical Report 20,* (2003)

[64] S. Boyd and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, 2004.

[65] D. Ganesan, S. Ratnasamy, H. Wang, and D. Estrin, *SIGCOMM Computer Communication Review*, 34(1), 125, (2004)

[66] C. Huang and Y. Tseng, *Proc. 2nd ACM inter. Conf. Wireless sensor networks and applications*, 1, 115 (2003)

[67] H.L. Van Trees. *Detection, Estimation, and Modulation Theory Part I.* Wiley, New York (1968)

[68] P.K. Varshney. *Distributed Detection and Data Fusion.* Springer-Verlag, New York (1997)

[69] I. Csiszar and J. Korner, *IEEE Trans. Inform. Theory*, 26, 155 (1980)

[70] E. Gamal and T.M. Cover, "Multiple user information theory," *Proc. IEEE*, 68, 1466 (1980)

[71] J.N. Tsitsiklis, *IEEE Trans. Automatic Control*, 30, 440 (1985)

[72] J. Proakis, *Digital Communications 4^{th} Ed.*, McGraw-Hill, New York (2001)

[73] F. Zhao and L. Guibas. *Wireless Sensor Networks: an Information Processing Approach*. Morgan Kaufman, San Francisco (2004)

[74] H. Wang, K. Yao, G. Pottie and D. Estrin, *Proc. 3rd ACM-IEEE IPSN,* 1, 36 (2004)

[75] M. Ahmed and G. Pottie, in *Frontiers in Distributed Sensor Networks,* Ed. R. Brooks and S. Iyengar, CRC Press, 2004.

[76] M. Gerla and J. Tsai, *ACM/Baltzer Journal of Wireless Networks*, 1, 255 (1995)

[77] J.L. Gao. *Energy Efficient Routing for Wireless Sensor Networks*. Ph.D. Dissertation, UCLA EE Dept., (2000)

[78] A. Wyner, *Inform. Contr.*, 38(7), 60 (1978)

[79] A. Wyner and J. Ziv, *IEEE Trans. Inform. Theory,* 26, 1 (1976)

[80] Y. Oohama, *IEEE Trans. Inform. Theory*, 43, 1912 (1997)

[81] Y. Oohama, *IEEE Trans. Inform. Theory*, 44, 1057 (1998)

[82] Y. Oohama, *Proceedings of IEEE Information Theory and Communications Workshop*, 1, 100 (1999)

[83] T. Berger, K. Houswright, J. Omura, S. Tung, and J. Wolfowitz, *IEEE Trans. Inform. Theory* 25, 664 (1979)

[84] T. Berger, Z. Zhang, and H. Vishwanathan, *IEEE Trans. Inform. Theory,* 42, 887 (1996)

[85] M. Ahmed and G. Pottie. *Proc. IEEE Inter. Symp. on Inform. Theory*, 1, 436 (2000)

[86] S. Pradhan and K. Ramchandran, *IEEE Trans. Inform. Theory*, 49, 626 (2003)

[87] A. Pandya, A. Kansal, G. Pottie and M. Srivastava, *CENS Technical Report 27*, (2003)

[88] H. Vishwanathan and T. Berger, *IEEE Trans. Inform. Theory,* 43, 1549 (1997)

[89] A. Pandya, A. Kansal, G.J. Pottie, and M.B. Srivastava, *CENS Technical Report 42,* (2004)

[90] Y.Xu, J. Heideman and D. Estrin, *USC/Information Sciences Institute Research Report 527,* (2000)

[91] M. R. Pearlman, J. Deng, B. Liang, and Z. J. Haas, *Proc. IEEE Globecom* 1, 17 (2002)

[92] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," *Proceedings of the 7th Annual ACM/IEEE International Conf. on Mobile Computing and Networking,* 1, 70 (2001)

[93] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, *Proc. the 7th Annual ACM/IEEE International Conf. on Mobile Computing and Networking,* 1, 481 (2001)

[94] A. Misra and S. Banerjee, *IEEE Wireless Communications and Networking Conference*, 217-21, 800 (2002)

[95] J. Aslam, Q. Li, and D. Rus, *Wireless Communications and Mobile Computing,* 3(2), 187 (2003)

[96] Kemal Akkaya, Mohamed Younis, *23rd International Conference on Distributed Computing Systems,* 1, 710 (2003)

[97] C. Intanagonwiwat, R. Govindan, D. Estrin, and John Heidemann, *IEEE/ACM Trans. Networking,* 11(2), 2 (2003)

[98] R. Cristescu. Efficient Decentralized Communications in Sensor Networks. *D.Sc. Dissertation, EPFL.* (2004)

[99] S. Lindsey, C. Raghavendra, K. Sivalingam, *Proceedings of the 15th International Parallel & Distributed Processing Symposium,* 1, 2001 (2001).

[100] A. Scaglione and S. Servetto, *Proceedings of the 8th annual international conference on Mobile computing and networking*, 1, 140 (2002)

[101] S. Bandyopadhyay and E. J. Coyle, *Proceedings of IEEE Infocom,* 1, 1713 (2003)

[102] W. R. Heinzelman, A. Chandrakasan, and H Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proceedings of the Hawaii International Conference on System Sciences,* 1, 8020 (2000)

[103] S. Ghiasi, A. Srivastava, X. Yang, and M. Sarrafzadeh, *Sensors* 2002, 2, 258 (2002)

[104] Bela Bollobas, *Modern Graph Theory*, Springer Verlag, New York (1998)

[105] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows,* Prentice Hall, Englewood Cliffs, NJ, (1993)

[106] R. Zamir and T. Berger, *IEEE Trans. on Inf. Theory,* 45, 106 (1999)

[107] H. Luo, Y. C. Tong, and G. Pottie, *CENS Technical Report 45*, (2004)

[108] T. L. Magnanti and L. A. Wolsey, in *Handbooks in Operations Research and Management Science vol. 7: Network Models*, Elsevier, New York, (1995)

[109] B. Krishmamachari, D. Estrin, and S. Wicker, *USC Computer Engineering Technical Report CENG 02-14,* (2002)

[110] L. A. Wolsey. *Integer Programming.* Wiley-Interscience, New York (1998)

[111] M. Grossglauser and D. Tse, *Proc. IEEE Infocom 2001*, 3, 1360 (2001)

[112] M. Grossglauser and D. Tse, *IEEE/ACM Trans. Networking*, 10, 477 (2002)

[113] D. Shah, A. El Gamal, J. Mammen and B. Prabhakar, *Proc. IEEE Infocom 2004,* (2004)

[114] M. Savelsbergh and M. Sol. "The general pickup and delivery problem," *Transportation Science,* 29, 17

[115] A. Kansal, M. Rahimi, D. Estrin, W.J. Kaiser, G.J. Pottie and M.B. Srivastava, *Proc. IEEE SECON 2004*, (2004)

[116] A. Kansal, E. Yuen, W.J. Kaiser, G.J. Pottie and M. Srivastava, *Proc. 3rd ACM-IEEE IPSN,* 1, 388 (2004)

[117] W.J. Kaiser, G.J. Pottie, M.B. Srivastava, G.S. Sukhatme, J. Villasenor and D. Estrin, *CENS Technical Report 31*, (2003).

Figure 1. Interaction of Signal Processing and Networking in Sensor Networks
Figure 2. Discrete communication problem

Figure 3. An information processing system

Figure 4. Network information theory problems

Figure 5. Sensing of (a) point and (b) distributed sources

Figure 6. Adaptive sampling to meet fidelity constraint

Figure 7. Likelihood opinion pool

Figure 8. The n-helper problem

Figure 9. The distributed Gaussian CEO system

Figure 10. Energy consumption as a function of data rate.

Figure 11. An explicit side information example.

Figure 12. A spatial-temporal DPCM encoder

Figure 13. A spatial-temporal DPCM decoder

Figure 14. Data rate with help information.