JOINT COMPRESSION, DETECTION, AND ROUTING IN CAPACITY CONSTRAINED WIRELESS SENSOR NETWORKS

Onur G. Guleryuz and Ulaş C. Kozat

DoCoMo USA Labs
181 Metro Dr, Ste 300, San Jose, CA 95110
{guleryuz,kozat}@docomolabs-usa.com

ABSTRACT

This paper considers an important class of sensor networks where the ultimate goal is not necessarily to collect each individual measurement but rather a potentially smaller set of statistics. Considering link capacity constrained topologies, we derive results that optimally allocate rate/distortion to information collected by the sensors. As a key contribution, we determine how the flow of information emanating from the sensors should be managed, yielding optimal routing algorithms and jointly optimized networks. Our analysis encompasses the typical scenarios that are widely observed in sensor networks, and over these scenarios, we quantify the gains offered by sending the statistics rather than the measurement data itself. Our results reveal bottleneck situations over various scenarios, where directly performing bandwidth allocation over the statistics does not provide the desired gains.

We start the analysis from a simple scenario, where a fixed node aggregates all the information in the sensor network and relays the information to a remote control center. We obtain close form expressions and illustrate how allocating bandwidth for each individual measurement *(Case-1)* performs compared to allocating it for each desired statistic *(Case-2)* in different *bottleneck* situations. Then, we extend this scenario to the case where we optimally select a number of aggregators from a cloud of sensor nodes. In this second scenario, under well defined bandwidth constraints, we look at the optimum clustering problem, in which the goal is to select the best aggregation nodes to minimize the total distortion of the desired statistics at the remote control node. We provide an algorithmic solution that returns the optimum aggregation points and the optimum size of each cluster under some mild assumptions. We finally turn our attention to the general routing problem and provide an algorithm that performs routing and bandwidth allocation jointly. We also study the performance and behavior of bandwidth allocation for both Case-1 and Case-2.

1. INTRODUCTION

Wireless sensor networks have attracted significant attention in recent years. The rich investigation is due to the fact that these networks are application specific and hence, they are amenable to highly engineered network design that removes layer separation. Most part of the literature independently treats transmission capacity, sensing capacity, energy efficiency, routing, compression, and detection issues [1, 2, 3, 4, 5, 6, 7, 8, 9, 18]. Some of the prior art investigates the inter-dependence of different routing strategies or topology connectivity constraints and particular sensor applications [10, 11, 12]. Another important set of works can be categorized under joint source-channel coding strategies for sensor networks [13, 14, 15].

In this paper, we look at an important class of sensor networks, where the ultimate goal is not necessarily to collect each individual measurement, but rather a potentially smaller set of statistics. These statistics are often expressed as weighted sum of individual measurements and used for hypothesis testing or classification purposes. In such scenarios, we encounter several interesting problems that cross rate allocation, routing, and clustering, which we investigate under the gross picture of network detection. Starting from simpler yet tractable scenarios and moving to more involved ones, we investigate a fundamental subset of these problems.

In the sensor network model we consider, every node except for the remote control point is capable of: (i) data measurement, (ii) simple data processing, and (iii) relaying. Due to data processing capabilities, sensor nodes can decode, re-encode, and aggregate received information with the knowledge of local measurements.

We first analyze a relatively simple scenario (see Figure 1), where we have a single relay node. Each sensor node transmits its own quantized measurements to a local aggregation node (node 1). The aggregation node, in return, relays the measurements to a remote location (node 0). As a natural constraint, each transmission is subject to point-to-point throughput capacity constraints. We obtain optimal transmission strategies for two different cases: without and with signal processing. In the former case, aggregation node performs rate allocation to minimize the total distortion under the given capacity constraints. Whereas in the latter case, aggregation node has the knowledge of desired statistics (i.e. useful information) at the control point and can instead transmit a representation of these statistics. We provide analytical results for the total distortion seen at the controller node[1]. Our closed form expressions reveal bottleneck situations, where signal processing no longer achieves the desired gains.

We then shift our attention to a more general topology (see Figure 2(a)), where each sensor can directly transmit its data to the remote central node or it can pass its local measurement to another peer node that will eventually relay all the data to the remote node. Hence, we pose the problem as optimal clustering that selects the cluster heads and cluster sizes such that the total distortion at the central node is minimized under the capacity constraints. Again, we investigate the situation both with and without additional signal processing at the cluster heads. After some

[1] In order to accommodate general applications that use weighted sum statistics, we will use mean squared error (MSE) as the fidelity metric. We note however that MSE can be translated to the probability of detection/classification error under mild assumptions in order to state results in terms of other quality criteria when necessary.

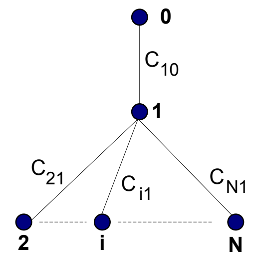
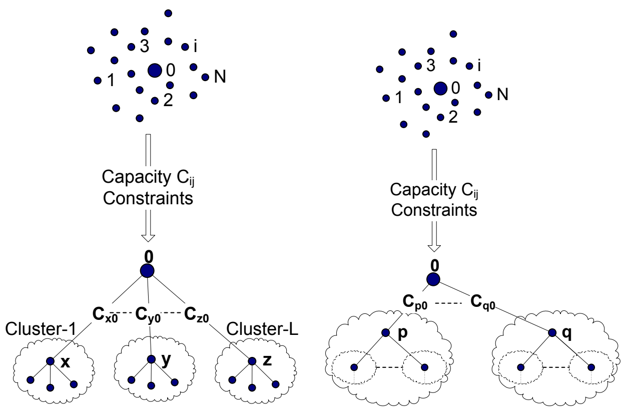


Fig. 1. Basic framework: Single cluster with fixed aggregation point and depth-2 routing tree.



(a) Routing over depth-2 tree (b) Routing over arbitrary tree

Fig. 2. Left hand side depicts the clustering problem: Select the optimum aggregation points, in the rate-distortion (**R**-**D**) sense, over a depth-2 routing tree and under capacity constraints. Right hand side removes the depth-2 routing constraint and poses a minimum distortion loop-free routing problem.

mild assumptions, such as identical observation variance and well-defined intra-cluster capacity scaling as a function of cluster size, we provide a polynomial-time solution to the problem by reducing the question to dynamic programming.

As our last contribution with this paper, we also look at the most general topology with arbitrary link capacities and observation variances. We present a greedy heuristic for our algorithmic solution, which returns a tree hierarchy solving the routing, rate allocation, and compression problems jointly. As before we compare distortion performance of both cases, i.e. when we apply and preclude signal processing at the aggregation points, via simulations.

We should explicitly note here that, in our analysis, we do not assume any special (spatial) correlation structure that can unfold itself into an efficient (and natural) routing strategy. We are also not interested in an asymptotic regime where the sensors are densely located. We assume that sensor nodes have very simple digital communication skills with low data rate and they cannot coordinate joint transmission strategies (i.e. transmissions are point to point). All these *practical* physical constraints together pose unique challenges that are not yet addressed in the literature for an arbitrary topology and our study constitutes a fundamental step towards quantifying the gains one can achieve by performing routing, rate allocation, and signal processing together.

The rest of the paper is organized as follows. In Section 2, we present the notation that is used in the rest of the paper. In Section 3, we provide our basic framework and corresponding analytical results. Section 4 extends the basic framework and provides exact as well as approximate algorithmic solutions for different problem instances. We finally deliver some simulation results and conclude in Sections 5 and 6.

2. NOTATION

In the bulk of this paper we will be encoding scalar random variables using simple uniform quantizers and entropy coders. In order to obtain convenient formulas for the resulting average rate in bits, R, and the mean squared quantization error, D, let us consider a zero-mean random variable y having variance σ_y^2. In general, R and D are complicated functions of the probability density function (pdf) of y. Regardless, for purposes of tractable optimization, we are interested in encoding strategies that produce $R = f_1(\sigma_y, \Delta)$, $D = f_2(\sigma_y, \Delta)$, where Δ is the uniform quantizer stepsize, and $f_.(.,.)$ are simple functions independent of the pdf of y.

Consider the entropy coding operation that takes the quantized version of y and losslessly encodes it as a sequence of bits [16]. By using an entropy code that is optimized for a zero-mean Gaussian random variable of variance $\sigma^2 \geq \sigma_y^2$, we can obtain

$$R \leq max(K_1, K_2 + \log(\sigma) - \log(\Delta)) \quad (1)$$

$$D \leq (\Delta/2)^2 \quad (2)$$

for positive constants K_1 and K_2 under modest assumptions on the regularity of the pdf of y ([17], section IV). Noting that $D = \sigma_y^2$ can be accomplished with $R = 0$ bits, and always choosing $\sigma = \sigma_y$, it is clear that we can define sufficiently large constants $K_2, K_3 > 0$ so that this simple encoding scheme obtains

$$R \leq K_2 + 1/2 \log(\frac{\sigma_y^2}{\Delta^2}) \quad (3)$$

$$D \leq K_3 \Delta^2 \quad (4)$$

for $D \leq \sigma_y^2$. Note also that in the important special case where y is a Gaussian random variable, the bound in Equation (3) becomes tighter, and in high rate regimes, we can find constants that replace inequalities with approximate equalities. In the rest of this paper we will summarize these relationships into a bound

$$D(R, \sigma_y^2) \leq \mathcal{D}(R, \sigma_y^2). \quad (5)$$

We will typically obtain y as a linear projection of a random vector x ($N \times 1$), via $y = h^T x$, for a given deterministic vector h. In this paper, x will be a conditionally Gaussian random vector conditioned on a state variable s. The state variable $s \in \{1, \ldots, \mathcal{S}\}$ determines the mean vector of x while the conditional covariance matrix of x is set to a fixed diagonal matrix, i.e., x becomes the observation of one of s deterministic signals under additive independent Gaussian noise. In much of the paper we will further assume that the noise is independent and identically distributed, i.e., the conditional covariance matrix is a multiple of identity.

A conditionally Gaussian x will in turn make y a mixture Gaussian random variable, i.e., given the mixture variable $s \in \{1, \ldots, \mathcal{S}\}$, y will be distributed conditionally Gaussian with mean $m_{y,s}$ and variance $\sigma_{y,s}^2$. In order to accommodate a wide range of means without a significant performance penalty, our encoder will first determine a state variable $\hat{s}$ as the state that maximizes the aposteriori probability of the quantized $y - m_{y,\hat{s}}$, encode $\hat{s}$ with $\sim \lceil \log(\mathcal{S}) \rceil$ bits, and then encode the quantized $y - m_{y,\hat{s}}$

as above using $\sigma = \sigma_{y,\hat{s}}$. This will increase the constant K_2 in Equation (3) by $\sim \log(\mathcal{S})$ bits but it will not effect our main results to order N as we will have $\mathcal{S} << N$. Observe that since any component of x can be obtained as a linear projection, we will be using the same encoding strategy when encoding the components of x. In addition, as means are handled by the encoder/decoder, in what follows we will restrict ourselves to zero-mean analysis without affecting Equations (3) and (4). With this caveat we will let σ_i^2 denote the variance of the i^{th} component of the vector x, without worrying about mixtures or the particulars of the state variable s.

Throughout, we will assume that relevant statistics (variances, means, etc.) are known at the encoder and decoder. In order to avoid technicalities with respect to cross correlations among quantized values and respective quantization errors we will assume that the uniform quantizer reproduction levels are at the centroids. For a given random variable y we will use the notation y^{Δ} to denote its quantized version using a quantizer of stepsize Δ. G will denote an $(M \times N)$ matrix whose application to a vector x results in M statistics that the central node is interested in.

3. BASIC FRAMEWORK AND COMPUTATION

In this section, we consider a relatively simple scenario where we have one local aggregation node that collects information from the sensors within its reach and combines with its own measurement to relay to a distant central node. The scenario is depicted in figure 1, where C_{ij} represents the channel capacity between nodes i and j. Indices 0 and 1 are reserved for the central and local aggregation nodes respectively. Node 1 can perform relaying in two different ways: (i) Encode each node's information, i.e. each entry of $x = (x_1, x_2^{\Delta_{21}}, \ldots, x_N^{\Delta_{N1}})^T$, separately, or (ii) encode each desired statistic, i.e. each element in $y = Gx$, separately. We compare these two cases in terms of their optimally achievable total distortion.

Case 1: Optimal transmission strategy for this case can be expressed as a distortion (equivalently *rate*) allocation problem, which is formally defined as:

Problem P1:

$$\min_{D_1,\ldots,D_N} \left\{ D_T = \sum_{i=1}^{N} D_i \right\} \quad (6)$$

$$\sum_{i=1}^{N} R_i = C_{10} \ , \quad (7)$$

$$D_i \leq \sigma_i^2 \ , \ i = 1, \ldots, N \ , \quad (8)$$

$$D_i \geq \mathcal{D}(C_{i1}, \sigma_i^2) \ , \ i = 2, \ldots, N \ , \quad (9)$$

where $\mathcal{D}(C_{i1}, \sigma_i^2) = K_3 \sigma_i^2 \exp(2K_2 - 2C_{i1})$, which is the distortion incurred by compressing a random variable with variance σ_i^2 using C_{i1} bits.

Solution of P1: When we consider the problem by assuming measurements from nodes 2 to N are available with no quantization, i.e. constraints in (9) are omitted, the solution is well-known and follows a "reverse water-filling" argument (see [16], pp. 348-349 for details). In a similar vein, by adding additional Kuhn-Tucker conditions, we can find the optimal distortion values (denoted as D_i^*'s) as:

$$D_i^* = \begin{cases} \mathcal{D}(C_{i1}, \sigma_i^2), & if\ \lambda \leq \mathcal{D}(C_{i1}, \sigma_i^2) \ , \\ \lambda, & if\ \mathcal{D}(C_{i1}, \sigma_i^2) < \lambda < \sigma_i^2 \ , \\ \sigma_i^2, & if\ \lambda \geq \sigma_i^2 \ . \end{cases} \quad (10)$$

Define the sets $P = \{i : \mathcal{D}(C_{i1}, \sigma_i^2) < \lambda < \sigma_i^2\}$, $Q = \{i : D_i^* = \sigma_i^2\}$, and $Z = \{i : D_i^* = \mathcal{D}(C_{i1}, \sigma_i^2)\}$. Then,

$$\lambda = \mathcal{D}\left[\frac{C_{10} - \sum_{i \in Z} C_{i1}}{|P|}, (\prod_{i \in P} \sigma_i^2)^{1/|P|} \right], \quad (11)$$

$$R_i^* = \frac{1}{2} \log \left(\frac{\sigma_i^2}{(\prod_{i \in P} \sigma_i^2)^{1/|P|}} \right) + \frac{C_{10} - \sum_{i \in Z} C_{i1}}{|P|}. \quad (12)$$

Case 2: Intuition suggests that we should be able to do much better in the sense of total distortion by directly working on the desired statistics. Therefore, as a simple scheme, we are interested in optimally transmitting each statistics $y_j = \sum_{i=1}^{N} g_{ji} x_i^{\Delta_{i1}}$ with $x_1^{\Delta_{11}} = x_1$. Under the independence of $x_i^{\Delta_{i1}}$'s, we obtain $E[y_j^2] = \sum_{i=1}^{N} g_{ji}^2 E[(x_i^{\Delta_{i1}})^2]$. Denote quantization error at the ith node as w_i, i.e. $x_i = x_i^{\Delta_{i1}} + w_i$. Since $x_i^{\Delta_{i1}}$ are at the centroids, $E[x_i^{\Delta_{i1}} w_i] = 0$, and we have $\sigma_i^2 = E[x_i^2] = E[(x_i^{\Delta_{i1}})^2] + E[w_i^2]$ or $E[(x_i^{\Delta_{i1}})^2] \leq \sigma_i^2$. Hence,

$$\sigma_{y_j}^2 = E[y_j^2] \leq \sum_{i=1}^{N} g_{ji}^2 \sigma_i^2 \ . \quad (13)$$

Using the upper-bound (13) and the results from case 1, we can compute the optimal transmission rates R_j^* for each statistic y_j. Then, by (3), we can find the quantization step size Δ_j for each transmitted statistic as:

$$\Delta_j = \sigma_{y_j} \exp(K_2 - R_j^*) \ . \quad (14)$$

Since we are actually interested in $\sum_{i=1}^{N} g_{ji} x_i$, we need to compute the overall distortion by:

$$\begin{aligned} D_T^{(j)} &= E\left[\left(\sum_{i=1}^{N} g_{ji} x_i - y_j^{\Delta_j} \right)^2 \right] && (15) \\ &= E\left[\left((\sum_{i=1}^{N} g_{ji} x_i - y_j) + (y_j - y_j^{\Delta_j}) \right)^2 \right] && (16) \\ &= \sum_{i=2}^{N} g_{ji} E\left[(x_i - x_i^{\Delta_{i1}})^2 \right] + E\left[(y_j - y_j^{\Delta_j})^2 \right] \\ &\quad + \sum_{i=2}^{N} 2 g_{ji} E\left[(x_i - x_i^{\Delta_{i1}})(y_j - y_j^{\Delta_j}) \right] && (17) \\ &\leq \sum_{i=2}^{N} g_{ji}^2 \mathcal{D}(C_{i1}, \sigma_i^2) + \mathcal{D}(R_j^*, \sigma_{y_j}^2) \ , && (18) \end{aligned}$$

where cross terms in Equation (17) are zero since the correlation of quantized values and quantization errors is zero. Note that in case 1, we may throw away some of the measurements due to insufficient capacity at node 1 following reverse water filling arguments, whereas in case 2, we effectively compress each statistics of interest by utilizing all the received information provided $g_{ji} > 0, \ \forall i$.

3.1. Comparison of Case 1 and Case 2 in Special Scenarios:

Let's limit ourselves to the scenario where we are only interested in a single statistic that is the sum of all measurements, i.e. $g_{1i} = 1$. We next show that sending statistics rather than individual measurements is beneficial if there is sufficient intra-cluster bandwidth.

3.1.1. $C_{i1} = C > C_{10}/N$ *and* $\sigma_i^2 = \sigma^2 > \lambda$

In this scenario, we have a communication bottleneck between the local aggregation and central nodes. Since measurement variances are same, in case 1, all measurements are transmitted at rate $R_i^* = C_{10}/N$. Therefore, we can express the ratio of total distortion of case 1 to that of case 2 as:

$$\rho = \frac{(D_T)^I}{(D_T)^{II}} = \frac{\exp(-2C_{10}/N)}{\frac{N-1}{N}\exp(-2C) + \exp(-2C_{10})} \,. \qquad (19)$$

When $C \approx C_{10}$, we have an exponential improvement in distortion in terms of C_{10}:

$$\rho = \frac{N}{2N-1}\exp\left\{2\frac{(N-1)}{N}C_{10}\right\}.$$

3.1.2. $C_{i1} = C < C_{10}/N$ *and* $\sigma_i^2 = \sigma^2 > \lambda$

When $C < C_{10}/N$, we observe that transmissions to local aggregation node constitute the bottleneck and there is no incentive in case 1 to encode at a higher rate to the central node. In return, ρ becomes:

$$\rho = \frac{exp(-2C)}{\frac{(N-1)}{N}\exp(-2C) + \exp(-2C_{10})} \approx \frac{N}{(N-1)} \,, \qquad (20)$$

for sufficiently large C_{10}. Hence, when transmissions to local aggregation point form a bottleneck, both schemes perform similar for moderate number of nodes.

4. GENERALIZED FRAMEWORK

In the following two sections, we investigate two different topological settings. In section 4.1, we limit ourselves to a routing tree structure that is of depth two or less and we jointly optimize it under the specified conditions. In section 4.2, we look at the most general scenario, which allows arbitrary topologies, and we search for the optimal routing tree structure.

4.1. Optimal Clustering Under a Simplified Scenario

In this section, we are interested in optimally forming L clusters in a capacity constrained network of N nodes, with every cluster of the particular form shown in Figure 2(a). Observe that the information flow from the nodes to the central node is via a tree of depth two. Hence, in effect, we would like to find the optimal tree that conforms to a network/transmission model, i.e., we are interested in finding the tree of depth less than or equal to two that minimizes the total distortion observed at the central node, given the capacity constraints enforced by the transmission model. In order to obtain optimal solutions, we restrict discussion to the Case-1 scenario identified in Section 3, and $\sigma_i^2 = \sigma^2$, $i = 1, \ldots, N$. We consider the application to a simple form of Case-2 at the end of the section.

Assume that the capacity between node i and the central node 0 is given by C_{i0}, $i = 1, \ldots, N$. We will start with the situation where the C_{i0} are arbitrary but the internode capacities $C_{ij} = C$, $i, j = 1, \ldots, N$, in order to define an algorithm that finds the optimal tree, and then generalize this algorithm to conform to the more general transmission model where C varies for each cluster as an arbitrary function of the number of nodes aggregated by that cluster.

Let node $\gamma(l)$ be the aggregator of cluster l and let $\xi(l)$ be the set of nodes that are aggregated by $\gamma(l)$ so that the information coming out of cluster l is due to the nodes $\xi(l) \cup \{\gamma(l)\}$. Let $\alpha_l = |\xi(l)|$ denote the cardinality of $\xi(l)$ with $\sum_{l=1}^{L} \alpha_l = N - L$. Using Case-1 results, we determine that nodes in $\xi(l) \cup \{\gamma(l)\}$ will encode x_j, $j \in \xi(l) \cup \{\gamma(l)\}$ to achieve the distortions

$$\text{if } \frac{c_{k,0}}{\alpha_l + 1} \geq c \quad \rightarrow \quad \begin{cases} D_j = \mathcal{D}(c_{k,0} - \alpha_l c, \sigma^2), & j = \gamma(l) \\ D_j = \mathcal{D}(c, \sigma^2), & j \in \xi(l) \end{cases}$$
$$\text{otherwise} \quad \rightarrow \quad D_j = \mathcal{D}(\frac{c_{k,0}}{\alpha_l + 1}, \sigma^2),\ j \in \xi(l) \cup \{\gamma(l)\}, \qquad (21)$$

with the total distortion incurred at the central node due to cluster l given by

$$\sum_{j \in \xi(l) \cup \{\gamma(l)\}} D_j. \qquad (22)$$

We are now ready to specify a dynamic programming algorithm that optimally determines the cluster number $L \leq N$, the number of nodes α_l that are aggregated by cluster l, $l = 1, \ldots, L$, and the nodes that act as the cluster aggregators. Due to symmetry, observe that L, α_l, and the indices of the cluster aggregator nodes are sufficient to determine the optimal tree.

Consider the $N + 1$ step trellis, where the states at step i are determined by the possible values of $\mathcal{N}(i)$, the number of nodes left available for clustering to steps $i + 1$ through N. We have $0 \leq \mathcal{N}(i) \leq N$, $\mathcal{N}(N) = 0$, and we define $\mathcal{N}(0) = N$.

Consider a path through this trellis given by the sequence of values $(N, \mathcal{N}(1), \ldots, \mathcal{N}(N-1), 0)$, with $\mathcal{N}(m) \geq \mathcal{N}(n)$ for all $m \leq n$. We will say that node i, $i = 1, \ldots, N$, is *aggregated* if $\mathcal{N}(i-1) = \mathcal{N}(i)$, and *aggregating* otherwise. If node i is aggregating, then the number of nodes it aggregates including itself is given by $\mathcal{N}(i-1) - \mathcal{N}(i)$.

It is clear that any tree of depth less than or equal to two can be mapped to a sequence of $\mathcal{N}(.)$'s and vice versa. For example, suppose we have $N = 5$, $L = 2$, and the two aggregating nodes are given by node 1 with $\alpha(1) = 2$, and node 4 with $\alpha(2) = 1$. This corresponds to the sequence $(5, 2, 2, 2, 0, 0)$.

It is also clear that, since the total distortion at the central node is additive over clusters and the optimal cluster distortions can be allocated independently, we can choose the optimal first $i - 1$ steps among all $\mathcal{N}$ sequences that pass through $\mathcal{N}(i) = j$, simply by comparing the distortions they induce at the central node. Hence to carry out step i, we need to only consider the $j + 1$ paths that emanate from $\mathcal{N}(i) = j$ to $\mathcal{N}(i+1) = j, \ldots, \mathcal{N}(i+1) = 0$, and repeat this for all $N + 1$ possible values of $\mathcal{N}(i)$. With N steps, each requiring such $O(N^2)$ path evaluations, we can determine the optimal sequence and thus the optimal tree with a dynamic programming algorithm of complexity $O(N^3)$.[2]

Now, suppose we generalize to the case where the node capacities to the aggregator of cluster l are given by a cluster specific constant $C_l = C(\alpha_l)$, i.e., $C_{i\gamma(l)} = C(\alpha_l)$ if $i \in \xi(l)$. For example in the transmission scenario where bandwidth inside each cluster is allocated as a function of α_l we can have $C(\alpha_l) = W/\alpha_l$. This changes the distortion allocation specifics in Equation (21)

[2]When intra-cluster (i.e. inter-node) capacities are scaled as a monotonically non-increasing function of the cluster size, we can find optimal solution faster than dynamic programming in $O(N^2)$ steps by greedily filling up each cluster after sorting the capacities to the central node.

but the cluster additive nature of the total central node distortion and the independent allocation of optimal cluster distortions do not change. Hence the above outlined dynamic program will again find the optimal tree. Note that the algorithm can also be utilized to solve a specific instance of Case-2 with a single statistic that is the summation of all measurements, i.e., $g_{1i} = 1$. However, the solution for general G remain difficult. Section 5 includes simulation results that compare Case-1 solutions with the specific Case-2 scenario.

4.2. Efficient topology formation with tree-growing heuristic

In this section, we utilize our earlier results to construct an efficient topology in the distortion sense under an arbitrary setting, where capacity constraints C_{ij}'s between node pairs $\{i, j\}$ and σ_i's are arbitrary. Our objective is to deliver the desired statistics to node 0 with minimal total distortion D_T^* by altering the routing tree, which in return determines a hierarchy of aggregation nodes with node 0 at the root.

We start from an initial tree T_0 with depth one, i.e. all nodes 1 through N directly send their individual measurements to node 0. Our algorithm iterates over the trees: at kth iteration, it inputs T_k and outputs T_{k+1}. We also define the total distortion due to tree T_k as $D(T_k)$ and denote the subtree rooted at i as $ST(i)$. Below, we provide the pseudo-code of our algorithm.

while $T_{k+1} \neq T_k$
  for all $i \in \{1, \ldots, N\}$
    $p = arg\min_{j \notin ST(i)} D(f(T_k, i, j))$;
    $T_{k+1}(i) = f(T_k, i, p)$;
  end
  $T_{k+1} = \min_i T_{k+1}(i)$;
  $k = k + 1$;
end

Function $f(T_k, i, j)$ in this procedure constructs a new tree by changing the *parent* node of i to j by keeping the subtree of i intact. We can compute $D(T_k)$ for both transmission schemes as examined in Case-1 and Case-2.

5. SIMULATION RESULTS

Figures 3 and 4 present the results for our dynamic programming solution for the topology discussed in section 4.1, with $N = 40$ and C_{i0}'s randomly picked between 0 and C_{0max} bits, $i = 1, \ldots, N$. The intra-cluster capacities are scaled as $C_l = W/\alpha_l$. 10000 simulations are done at each C_{0max} to obtain average distortions and cluster sizes. The first plot shows an initial exponential gain of Case-2 over Case-1 in average distortion as a function of C_{0max}. This is similar to our results derived for the basic framework over a single cluster. However, as C_{0max} exceeds a certain threshold, the advantage of allocating bandwidth only to the statistic diminishes. Although we included the results for the case $C_l = W/\alpha_l$, the behavior remains same for $C_l = C$. The second plot shows the average optimum cluster sizes for $W = 2.5$. As a function of C_{0max}, the cluster sizes converge to different levels and Case-2 consistently results in less number of clusters in non-bottleneck environment, hence aggregating more sensor nodes.

Figure 5 summarizes the performance results obtained by our *tree-growing* heuristic. We placed the sensor nodes randomly over a 5×5 topology with the remote control point at the center. We set the capacity within 1 unit distance to 3.5 bits and applied inverse square law for longer distances reflecting the capacity reduction with communication distance between transceiver pairs. The observations have unit variance. For Case-2, we also assumed that the desired statistics are sums of the observations collected inside concentric strips from the remote control point. Case-2 provides as high as 7dB gain over Case-1, when the number of statistics is one. As the number of statistics increases, the performance and information flow of Case-2 approaches to that of Case-1.

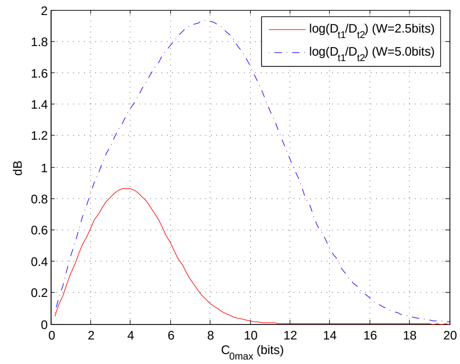


Fig. 3. Comparison of total distortions for Case-1 and Case-2 in Optimal Clustering scenario of Section 4.1.

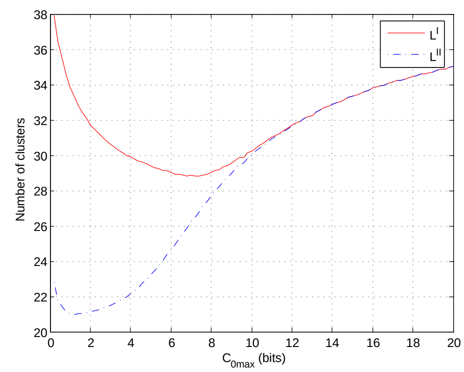


Fig. 4. Average number of clusters for Case-1 and Case-2.

6. CONCLUSION

In this paper, we considered sensor networks under different link capacity and topology constraints. The main focus has been on gathering new insights in the design of new techniques that jointly optimize rate allocation and routing to minimize distortion of the desired statistics collected at a central node. Our analysis concentrated on two cases of information collection. In the first case (i.e. Case-1), the network was optimized to convey all the measurements collected at different sensor nodes under a total fidelity criterion. The measurements of each sensor were individually compressed according to the bandwidth constraints and optimal routing path, and sometimes discarded to minimize the total distortion. Nonetheless, most of the time, popular sensor network applications (e.g. detection and target tracking) require information on a smaller subset of desired statistics, which are often expressed as linear sums of measurements collected at different sensors. Therefore, in the second case (i.e. Case-2), we modified optimization strategy and allowed intermediate nodes to compress the desired statistics directly.

We compared both strategies under three different topologies. In the first topology, there was only one aggregation node that collected the measurements from the rest of the network via direct links. Besides being a good toy example, where we could provide close form expressions for optimal solutions of both cases, this scenario was also interesting in quantifying the bottleneck conditions. Our results showed that Case-2 significantly outperformed Case-1, provided that in-network bandwidth was sufficiently large. Confirming our initial expectations, we also observed that, unlike the first case, the second case optimization was also much less willing to discard individual sensor measurements for nontrivial statistics.

In the second topology, we introduced more degrees of freedom by allowing more than one aggregation node (i.e. cluster head) and allowed these to be freely selected from the available sensor nodes. In essence, through this scenario, we introduced the routing problem into the picture but limited ourselves to a depth-2 routing tree. Under mild assumptions such as letting in-network capacities be an arbitrary function of the network size and restricting ourselves to the same observation variances with a single desired statistic, we were able to provide optimal algorithmic solutions to the posed problem for both cases. Although the problem is relaxed significantly from the first scenario, we still observed that there exists critical thresholds, after which the gains obtained by Case-2 started diminishing.

Finally, we removed the depth-2 tree constraint, and designed a locally optimal algorithm for solving the most general scenario. Case-2 had significant performance gains (as high as 7dB) over randomly deployed networks even when more than one statistic was desired.

It is noteworthy to state that although not directly involved, as a dual problem, our solutions can also be used to address the problem of sensor-net lifetime maximization under a given fidelity criterion by gauging the capacity constraints inside the network.

7. REFERENCES

[1] J.-F. Chamberland and V. V. Veeravalli, “Decentralized detection in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.

[2] J-J. Xiao and Z-Q. Luo, “Decentralized detection in a bandwidth constrained sensor network,” in *Proc. of Globecom*, 2004.

[3] E. J. Duarte-Melo and M. Liu, “Data-gathering wireless sensor networks: Organization and capacity,” *Computer Networks*, vol. 43, no. 4, pp. 519–537, November 2003.

[4] J. Faruque and A. Helmy, “Rugged: Routing in fingerprint gradients in sensor networks,” in *Proc. of ICPS*, 2004.

[5] L. Vasudevan, A. Ortega, U. Mitra, “Application-specific compression for time delay estimation in sensor networks,” in *Proc. of SenSys*, 2003.

[6] A. Deligiannakis, Y. Kotidis, N. Roussopoulos, “Compressing historical information in sensor networks,” in *Proc. of Sigmod*, 2004.

[7] Y. Rachlin, R. Negi, and P. Khosla, “Sensing capacity for target detection,” in *Proc. of ITW*, 2004.

[8] C. Gui and P. Mohapatra, “Power conservation and quality of surveillance in target tracking sensor networks,” in *Proc. of Mobicom*, 2004.

[9] S. D. Servetto, “Lattice quantization with side information: Codes, asymptotics, and applications in sensor networks,” preprint IEEE Tran. on Information Theory.

[10] A. Scaglione and S. Servetto, “On the interdependence of routing and data compression in multi-hop sensor networks,” in *Proc. of Mobicom*, 2002.

[11] S. Pattem,B. Krishnamachari, R. Govindan, “The impact of spatial correlation on routing with compression in wireless sensor networks,” in *Proc. of IPSN*, 2004.

[12] M. Alanyali, S. Venkatesh, O. Savas, S. Aeron, “Distributed bayesian hypothesis testing in sensor networks,” in *Proc. of ACC*, 2004.

[13] Prakash Ishwar, Rohit Puri, Kannan Ramchandran, S. Sandeep Pradhan, “On rate-constrained distributed estimation in unreliable sensor networks,” in *IEEE JSAC*, 2005.

[14] M. Gastpar and M. Vetterli, “Power-bandwidth-distortion scaling laws for sensor networks,” in *Proc. of IPSN*, 2004.

[15] A. Sayeed W. Bajwa and R. Nowak, “Matched source-channel communication for field estimation in wireless sensor networks,” in *Proc. of IPSN*, 2005.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.

[17] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, “On the importance of combining wavelet-based nonlinear approximation with coding strategies,” *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1895–1921, July 2002.

[18] R. Willett, A. Martin, and R. Nowak ”Backcasting: adaptive sampling for sensor networks,” in *Proc. of IPSN*, 2004.

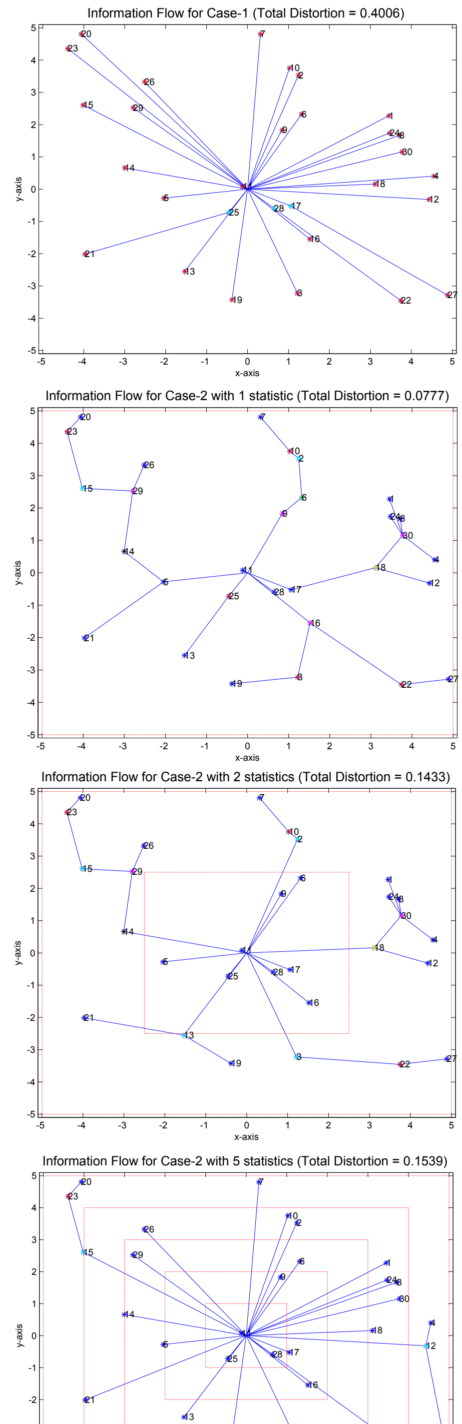


Fig. 5. Optimal information flow over a random topology for (1) Case-1, (2) Case-2 with 1 statistic, (3) Case-2 with 2 statistics, and (4) Case-2 with 5 statistics from left to right.