Prize-Collecting Data Fusion for Cost-Performance Tradeoff in Distributed Inference

Animashree Anandkumar*, Meng Wang*, Lang Tong*, and Ananthram Swami†

*ECE Dept., Cornell University, Ithaca, NY 14853, USA. Email: {aa332,mw467,lt35}@cornell.edu
†Army Research Laboratory, Adelphi, MD 20783, USA. Email: a.swami@ieee.org

***Abstract*—A novel formulation for sensor selection and in-network fusion for distributed inference known as the prize-collecting data fusion (PCDF) is proposed, and its optimality is studied in terms of the tradeoff between the costs of aggregating the selected set of sensor measurements and the resulting inference performance at the fusion center. For i.i.d. measurements in large networks, PCDF reduces to the prize-collecting Steiner tree (PCST) with the Kullback-Leibler distance as node penalties. The problem is then extended to a correlation model specified by a Markov random field (MRF) with dependency graph. Under some conditions, PCDF reduces to PCST on an augmented graph with a penalty function separable in the cliques of the MRF. An approximation algorithm for PCDF is given with the approximation ratio depending only on the number of profitable cliques in the dependency graph. Simulation results show the influence of the correlation structure, the tradeoff factor and the node placement on the performance of our heuristics.**

***Index Terms*—Optimal Node Selection, Sensor Networks, In-network Aggregation, Detection, Prize-Collecting Steiner Tree.**

I. Introduction

Consider a sensor network deployed in an area taking measurements for distributed inference. Here, a designated fusion center collects the sensor measurements and makes a final decision about the underlying signal field. The classical works on this topic are concerned with optimal inference rules [1], and the role of network constraints is not considered.

Sensor networks have many resource constraints (e.g., energy). Hence, it may not be feasible to use all the sensor measurements for inference. It is then crucial for the fusion center to select an efficient set of sensor measurements based on the tradeoff between the routing costs, and the resulting inference performance at the fusion center. Intuitively, it is cheaper to select nearby sensors and more useful to select "informative" sensors.

Efficient sensor selection presents several challenges in deriving an analytical solution. It should be apparent that the brute force approach of searching over all possible sensor subsets for selection is not feasible even for moderate-sized networks. On the other hand, the cost of data fusion and the performance of distributed inference are intertwined, leading to a complex optimization problem. Are there reasonable heuristics for sensor selection? Is it possible to provide guarantees for the heuristics? How do factors such as the correlation model, node placement strategy and the tradeoff factor affect these heuristics? How do we aggregate[1] data at intermediate nodes to reduce routing costs and yet provide guaranteed inference performance? We address these issues in this paper.

This work was supported in part through the collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and by the National Science Foundation under Contract CNS-0435190. The third author was partially supported by the DARPA ITMANET program. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

A. Summary of Results

This paper considers selection of sensors to achieve optimal cost-performance tradeoff. The costs are incurred in routing and aggregating the selected subset of measurements and the performance is in terms of the resulting error probability. The contributions are three fold. First, we propose a formulation for sensor selection and fusion known as the *prize-collecting data fusion* (PCDF). Second, we prove its reduction to a known optimization problem for certain correlation structures. Third, we propose heuristics for any correlation structure and study their performance through simulations.

When the measurements are i.i.d. and the network size goes to infinity, PCDF reduces to an optimization problem known as the *prize-collecting Steiner tree* (PCST) [2]. It is defined as the sub-tree rooted at a specified vertex that minimizes the sum of edge costs in the tree plus the penalties of the nodes not spanned by it. For i.i.d. data, the node penalties are given by the Kullback-Leibler (KL) distance.

We then consider correlated measurements incorporated via the statistical model of a *Markov random field* (MRF) with (undirected) *dependency graph* [3]. Under some conditions, PCDF reduces to PCST on an augmented graph. The augmentation involves adding new nodes and edges to account for increase in aggregation costs due to the presence of correlation. In general, finding the PCDF is NP-hard and we resort to approximations. The *approximation ratio* ρ of any polynomial-time algorithm guarantees that its output is no worse than ρ times the optimal value. We give an approximation algorithm for PCDF with the approximation ratio depending only on the number of profitable cliques in the dependency graph.

We develop group selection heuristics based on the PCST reduction and study their performance through simulations. It is observed that they perform substantially better than the selection scheme with no data fusion. Indeed, it is very expensive to route all the raw measurements to the fusion center. We then study the influence of node configuration and

[1]The terms aggregation and fusion are used interchangeably.

observe that a clustered node placement achieves a better tradeoff compared to uniform placement at sparse spatial dependencies. These have direct implications on designing good placement strategies.

B. Related Work

Sensor selection algorithms have been considered in a variety of contexts: for target tracking [6], multimedia streams [7], fixed number selection [8], region selection [4], for information maximization [9] and so on. See survey [10]. None of the above works consider selection for an inference application. In [11], selection for hypothesis testing in a dynamical system is considered. But the influence of data aggregation and correlation model are not studied.

A limited number of works consider energy-efficient inference in sensor networks, e.g., [12]–[14], under some special correlation models. More relevant here is the notion of in-network aggregation, considered for specific function computation in [15]. However, the mechanism to aggregate correlated data with guaranteed inference performance is nontrivial and depends on the statistical model of the sensor measurements.

In [16], [17], we consider minimum cost data aggregation of all the sensor measurements under the Markov random field model for optimal inference but did not deal with the issue of sensor selection. In [18], we consider optimal node density for inference leading to sleeping strategies to meet the energy constraints. In contrast, this work uses the approach of careful sensor selection to achieve energy efficiency.

To the best of our knowledge, the problem of optimal node selection has not been considered in conjunction with in-network aggregation before (see section below). Indeed in single-hop networks, there is no need for data aggregation. But most large networks are multi-hop and substantial savings in routing costs can be obtained through aggregation, as seen in simulations in Section V.

Many works on selection consider perfect sensing of a region (e.g., [4]). In contrast, our work explicitly models correlated imprecise measurements via the Markov random field model, and this model is used to select sensors that are expected to be "informative" for inference. Indeed, there is the issue of accuracy in learning the statistical model. Conceding this limitation, we aim to gain insights through the model-based framework.

We first introduced the problem of optimal node selection in [5]. We have new results in this paper on the asymptotic optimality of our approach and improved approximation guarantees. We additionally propose new heuristics and provide extensive simulation results.

II. System Model & Problem Formulation

In this paper, we will consider various graphs. Chiefly among these are *dependency graphs* specifying the correlation structure of sensor measurements, *network graphs* denoting feasible links for communication, and *fusion graphs* denoting links used by a policy to route and aggregate data.

A. Measurements: Correlation & Inference Model

We assume that the sensor measurements are drawn from a Markov random field. Let $\mathbf{Y}_V = [Y_i, i \in V]^T$ denote the random vector of measurements in set V. If $\mathbf{Y}_V$ is a MRF with dependency graph $\mathrm{DG}(V)$, then its joint pdf f_V is given by the Hammersley-Clifford theorem [3],

$$-\log f_V(\mathbf{Y}_V; \Upsilon) = \sum_{c \in \mathcal{C}} \psi_c(\mathbf{Y}_c), \tag{1}$$

where $\mathcal{C}$ is the collection of (maximal) cliques[2] in $\mathrm{DG}(V)$ and the function ψ_c is known as the *potential* for clique c. Hence, the tuple $\{\mathrm{DG}(V), \mathcal{C}, \psi\}$ represents a MRF. For a discussion on the use of MRF for spatial correlation, see [17].

We consider the *binary hypothesis-testing* problem with null hypothesis $\mathcal{H}_0$ and alternative $\mathcal{H}_1$. Under either hypothesis, we assume that the measurements are drawn from distinct MRFs,

$$\mathcal{H}_0 : \{\mathrm{DG}_0(V), \mathcal{C}_0, \psi_0\} \; ; \; \mathcal{H}_1 : \{\mathrm{DG}_1(V), \mathcal{C}_1, \psi_1\}. \tag{2}$$

In order to quantify inference performance, we consider the *Neyman-Pearson* criterion, where for a fixed false-alarm probability (type-I error), the detector at the fusion center is optimal in terms of the type-II error probability P_M.

B. Network and Cost Model

The network is connected via a *network graph* of feasible links with given costs. For optimization of routing costs, we only need to work with the *metric closure*[3] of the network graph, denoted by $G_n(V)$, and the metric cost for each link (i, j), denoted by $C(i, j)$. For any graph G, let $\mathrm{C}(G)$ denote the total cost of using all its links. Communication between the nodes is perfect and scheduled so as to avoid interference.

Nodes communicate in the form of packets. Each packet contains bits for at most one (quantized) real variable and other overhead bits. The quantization error is assumed to be small and ignored here. A node can function as an aggregator (combines its own measurement) or a router (forwarding without combination). An aggregation scheme consists of the transmitter-receiver pairs, the transmission sc'1hedule, and the aggregation algorithm. It is modeled by a *fusion digraph* G_f consisting of directed links used by the scheme.

C. Problem Formulation

The goal of this paper is to select an optimal sensor subset[4] $V_s \subset V$, given the entire set V, and to incorporate in-network aggregation of the measurements in V_s before delivery to the fusion center. It is not possible to quantify inference performance under arbitrary aggregation.

We limit ourselves to aggregation schemes which guarantee the same inference performance as if the fusion center had

[2] A clique refers to a maximal clique unless otherwise mentioned.

[3] The metric closure on graph G, is defined as the complete graph where the cost of each edge (i, j) in the metric closure is the cost of the shortest path between i and j in G.

[4] The unselected nodes can still function as routers and forward data.

direct access to the selected raw measurements. In this case, there is no performance loss due to aggregation. Assuming that the optimal Neyman-Pearson (NP) detector is used at the fusion center, inference performance is measured by the NP type-II error P_M.

In statistical theory, a *sufficient statistic* is a well-behaved function of the data, which is as informative as the raw data for inference [19]. Hence, a scheme which delivers a sufficient statistic results no loss in inference performance due to data aggregation.

We are thus interested in subset selection $V_s \subset V$ and design of aggregation scheme $\Gamma(V_s)$ delivering a sufficient statistic of its measurements such that optimal linear tradeoff is achieved between the total routing costs $\mathrm{C}(\Gamma(V_s))$ and a *penalty* function π, based on the NP type-II error $P_M(V_s)$,

$$\mathrm{opt}(V,\mathrm{C},\gamma\pi):=\min_{V_s\subset V,\Gamma(V_s)}\Big[\mathrm{C}(\Gamma(V_s))+\gamma\pi(V\backslash V_s)\Big],\ \gamma>0 \quad (3)$$

where $V\backslash V_s:=\{i: i\in V, i\notin V_s\}$ and π is given by

$$\pi(V\backslash V_s):=\log\frac{P_M(V_s)}{P_M(V)}>0,\quad \forall V_s\subset V. \quad (4)$$

When we select all the sensors ($V_s = V$), (4) evaluates to zero and there is no loss in performance since none of the measurements are dropped. On the other hand, when we do not select all the sensors ($V_s \subsetneq V$), we incur a loss in performance and hence, pay a penalty in terms of the fraction of increase in error P_M due to non-selection of nodes in $V\backslash V_s$.

The parameter γ is known as the *tradeoff factor*, and is used to adjust the relative importance of cost and performance. Denote the objective function in (3) as

$$\mathrm{obj}(V_s,\Gamma(V_s);V,C,\gamma\pi):=\Big[\mathrm{C}(\Gamma(V_s))+\gamma\pi(V\backslash V_s)\Big], \quad (5)$$

and the optimal node set and fusion scheme by

$$[V_*,\Gamma_*(V_*)]:=\arg\min_{V_s\subset V,\Gamma(V_s)}\mathrm{obj}(V,\mathrm{C},\gamma\pi). \quad (6)$$

When the tradeoff factor γ is sufficiently large, the optimal tradeoff problem in (3) reduces to minimum cost fusion, considered in [16], where optimal inference is required and hence, all the nodes are selected, and the goal is to find the fusion scheme which minimizes the total routing costs while ensuring delivery of a sufficient statistic of all the measurements. On the other hand, when γ is sufficiently small, none of the nodes are selected.

$$\lim_{\gamma\to 0}V_*(V,\mathrm{C},\gamma\pi)\to\emptyset,\quad \lim_{\gamma\to\infty}V_*(V,\mathrm{C},\gamma\pi)\to V.$$

The optimization in (3) is the Lagrangian dual for the problem of finding a minimum energy fusion scheme under a constraint on the inference performance or vice versa. Hence, once we have an algorithm to find the (approximate) solution to (3), we can use it in the constrained optimization problems. This aspect is however not dealt in this paper and we will limit to finding solutions to (3). We will henceforth refer to the above formulation of node selection and aggregation as the prize-collecting data fusion (PCDF).

D. Preliminary Observations & Results

For binary hypothesis testing, the log-likelihood ratio (LLR) is *minimally* sufficient and represents the maximum reduction in dimensionality of raw data, and is given by

$$\mathrm{LLR}(\mathbf{Y}_{V_s}):=\log\frac{f_{V_s}(\mathbf{Y}_{V_s};\mathcal{H}_0)}{f_{V_s}(\mathbf{Y}_{V_s};\mathcal{H}_1)}, \quad (7)$$

where $f_{V_s}(\mathbf{Y}_{V_s};\mathcal{H}_j)$ is the pdf of the measurements $\mathbf{Y}_{V_s}$ under hypothesis $\mathcal{H}_j$. Hence, we seek an aggregation scheme Γ computing $\mathrm{LLR}(V_s)$ that achieves the minimum routing cost $\mathrm{C}(\Gamma(V_s))$ among all schemes delivering a sufficient statistic to the fusion center. Hence, the optimal aggregation scheme in (3) is a scheme which enables the computation of the LLR at the fusion center while minimizing the sum routing costs.

For the penalty function in (4), in general, the error P_M does not have a closed form and hence, is not tractable. We focus on the large-network scenario, where it can be approximated by the *error exponent*. When the type-II error $P_M(V)$ decays exponentially with the sample size $|V|$ for a fixed type-I error, the NP error exponent is given by

$$\mathcal{D}:=-\lim_{|V|\to\infty}\frac{1}{|V|}\log P_M(V). \quad (8)$$

Hence, a larger exponent implies faster decay of the error probability with increasing sample size.

III. IID Measurements

In this section, we focus on the special case when all the sensor measurements are i.i.d., $Y_i \overset{i.i.d.}{\sim} f(Y;\mathcal{H}_j)$, for $j=0,1$. We first solve a modified optimization problem based on (8) and then prove its convergence to (3).

For i.i.d. data, from Stein's Lemma [19], the exponent $\mathcal{D}$ in (8) is given by the single-letter Kullback-Leibler (KL) distance

$$\mathcal{D}=D(f(Y;\mathcal{H}_0)||f(Y;\mathcal{H}_1)). \quad (9)$$

We now consider a modified penalty function which assigns uniform penalty to each node equal to the KL-distance D.

$$\pi^{iid}(V\backslash V_s):=[|V|-|V_s|]D, \quad (10)$$

We first solve the optimization with π^{iid} instead of π in (4)

$$\mathrm{opt}(V,\mathrm{C},\gamma\pi^{iid}):=\min_{V_s\subset V,\Gamma(V_s)}\Big[\mathrm{C}(\Gamma(V_s))+\gamma[|V|-|V_s|]D\Big]. \quad (11)$$

Before we tackle the above problem, we consider optimal selection under no aggregation with penalty function π^{iid}. The minimum cost scheme here is the shortest path routing (SPR) of all raw data to the fusion center v_0. Hence, both the cost

and the penalty are separable in the nodes and the optimal selection is simply the set of "profitable" nodes

$$V_*^{\text{SPR}}(V, C, \gamma\pi^{iid}) = \{i : i \in V, \gamma D > C(i, v_0)\}, \qquad (12)$$

where C is the cost of shortest path. Although the above solution is simple, simulations in Section V suggest that it is not efficient compared to selection with aggregation.

In order to incorporate in-network aggregation, we need an explicit form for $\text{LLR}(\mathbf{Y}_{V_s})$. For i.i.d. data, it is

$$\text{LLR}(\mathbf{Y}_{V_s}) = \sum_{i \in V_s} \frac{f(Y_i; \mathcal{H}_0)}{f(Y_i; \mathcal{H}_1)}, \quad \forall V_s \subset V. \qquad (13)$$

In the theorem below, we prove that the optimal solution is the prize-collecting Steiner tree.

Theorem 1 (Selection & aggregation of i.i.d. data): The optimal solution in (11) is aggregation along the prize-collecting Steiner tree rooted at the fusion center v_0, and edges directed towards v_0: each node i in the PCST computes and transmits q_i to its immediate successor, given by

$$q_i = \text{LLR}(Y_i) + \sum_{j \in \mathcal{N}_p(i)} q_j, \qquad (14)$$

where $\mathcal{N}_p(i)$ is the set of immediate predecessors of i in the directed PCST.

Proof: Once a node is selected for aggregation, it should be connected to the fusion center in the fusion digraph. By definition, the PCST is the tree spanning a set of nodes achieving optimality in (11). The aggregation algorithm in (14) delivers the LLR of the selected set in PCST as required. □

Hence, the optimal aggregation for i.i.d. data is along the directed PCST. A schematic of the scheme is shown in Fig.1. In general, finding the PCST is NP-hard. The *approximation ratio* ρ of any polynomial-time algorithm guarantees that its output is no worse than ρ times the optimal value. In [2], an approximation algorithm for the PCST with ratio $2 - (|V| - 1)^{-1}$ for any node set V was proposed and is referred to as the *Goemans-Williamson* (GW) algorithm.

Theorem 1 establishes the optimality of PCST for the penalty function π^{iid} in (10). In the theorem below, we prove that the PCST its optimality for the penalty function π in (4) when the network size goes to infinity.

Theorem 2 (Asymptotic optimality of PCST for i.i.d. data): Under bounded link costs for π in (4) and π^{iid} in (10),

$$\lim_{|V| \to \infty} \frac{\text{opt}(V, C, \gamma\pi)}{\text{opt}(V, C, \gamma\pi^{iid})} \to 1, \quad \forall \gamma > 0. \qquad (15)$$

Proof: See [20, Sec. 1]. □

Hence, PCDF in (3) reduces to aggregation along the PCST, as the network size goes to infinity. The GW-algorithm approximates the PCST with a proven guarantee. On the other hand, there are no known approximations for a general PCDF in (3). Hence, for large networks, it is reasonable to use the KL-distance D in (9) instead of the error probability P_M.

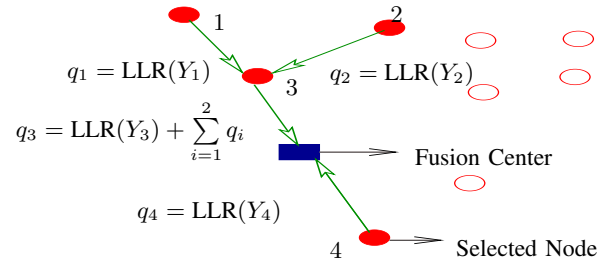


Fig. 1. Aggregation of i.i.d. measurements along the PCST.

IV. Correlated Measurements: MRF Model

We now generalize the results to the case when the measurements are correlated according to the Markov random field model, described in Section II-A. Several challenges arise here. First, the LLR is no longer a simple sum function as in the i.i.d. case in (13). Hence, the structure of fusion schemes computing the LLR is not clear. Second, the error exponent $\mathcal{D}$ is no longer the single-letter KL-distance as in (9), and hence, the exponent-based penalty may not be separable in the nodes. Third, nodes cannot be assigned uniform penalties since they affect inference performance differently. We will see that it is tractable to consider certain groups of nodes as candidates for selection rather than all the possible node subsets.

A. In-network Aggregation of LLR

For hypothesis testing of MRFs in (2), define the combined dependency graph $\text{DG}(V) := \text{DG}_0(V) \cup \text{DG}_1(V)$. Henceforth, we only work with $\text{DG}(V)$. The LLR of all the measurements $\mathbf{Y}_V$ is based on the cliques in $\text{DG}(V)$.

$$\begin{aligned} \text{LLR}(\mathbf{Y}_V) &:= \log \frac{f_V(\mathbf{Y}_V; \mathcal{H}_0)}{f_V(\mathbf{Y}_V; \mathcal{H}_1)}, \\ &= \sum_{a \in \mathcal{C}_0} \psi_{1,a}(\mathbf{Y}_a) - \sum_{b \in \mathcal{C}_1} \psi_{0,b}(\mathbf{Y}_b) \qquad (16) \\ &:= \sum_{c \in \mathcal{C}} \phi_c(\mathbf{Y}_c), \quad \mathcal{C} := \mathcal{C}_0 \cup \mathcal{C}_1. \qquad (17) \end{aligned}$$

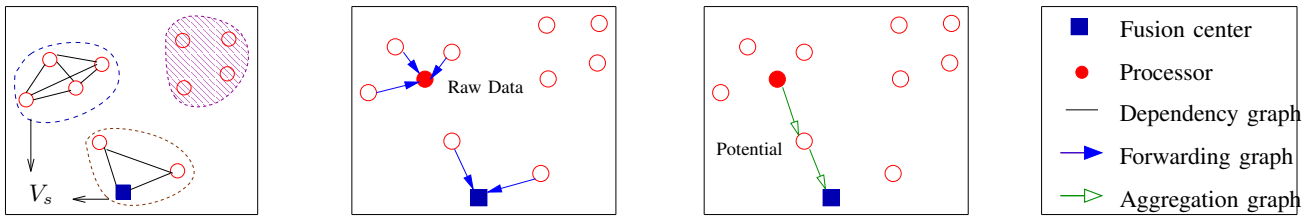
Comparing the above form with that for i.i.d data in (13), correlation increases the complexity of the LLR.

For any subset $V_s \subset V$, its LLR can be similarly expressed based on the clique set $\mathcal{C}'$ of the dependency graph $\text{DG}'(V_s)$

$$\text{LLR}(\mathbf{Y}_{V_s}) = \sum_{c \in \mathcal{C}'} \phi'_c(\mathbf{Y}_c), \qquad (18)$$

where $\text{DG}'(V_s) := \text{DG}'_0(V_s) \cup \text{DG}'_1(V_s)$ and $\text{DG}'_j(V_s)$ is the dependency graph of the marginal pdf $f_{V_s}(\mathbf{Y}_{V_s}; \mathcal{H}_j)$, for $j = 0, 1$. In general, $\text{DG}'(V_s)$ is not a subgraph of $\text{DG}(V)$ and hence, $\mathcal{C}'$ is not contained in $\mathcal{C}$. Hence, the structure of fusion scheme changes depending on the selected set V_s.

We now describe the structure of fusion schemes computing the LLR of a given subset V_s. See Fig.2. The issue of optimal selection of V_s will be considered later. Given the dependency graph $\text{DG}'(V_s)$, the computation is done in two stages. In the



(a) Cliques of dependency graph (b) Forwarding subgraph (c) Aggregation subgraph (d) Legend

Fig. 2. In-network Aggregation for inference: computation of the log-likelihood ratio $\text{LLR}(\mathbf{Y}_{V_s})$ of a given node subset V_s.

first stage, raw data are forwarded from all the members of clique $c \in \mathcal{C}'$ to compute its potential ϕ'_c at an assigned node called its *processor* denoted by $\text{Proc}(c)$. The links used for forwarding raw data here form the *forwarding* subgraph (FG).

In the second stage of LLR computation, all the clique potentials are summed up and delivered to the fusion center. The links used here form the *aggregation* subgraph (AG). The tuple with the forwarding and aggregation subgraphs of a fusion scheme is the referred to as the *fusion digraph*, $G_f := \{\text{FG}, \text{AG}\}$.

Hence, the total routing costs of a fusion scheme is

$$\text{C}(G_f) = \text{C}(\text{FG}) + \text{C}(\text{AG}). \tag{19}$$

B. Error Exponent & Penalty Function

We now provide results for the error exponent $\mathcal{D}$, which it then used to define a penalty function approximating the function π in (4).

Theorem 3 (NP Error Exponent for MRF): When the sequence of normalized log-likelihood ratio variables is uniformly integrable and converges in probability under the null hypothesis $\mathcal{H}_0$, the error exponent in (8) is

$$\mathcal{D} = \text{p} \lim_{n\to\infty} \frac{1}{n} \sum_{c\in\mathcal{C}} \mathbb{E}(\phi_c(\mathbf{Y}_c)|V;\mathcal{H}_0), \tag{20}$$

where ϕ_c is the potential function for clique c, $\mathcal{C}$ is the MRF clique collection in (17) and $\mathbb{E}$ is the expectation under $\mathcal{H}_0$.

Proof: We use the form of LLR in (17). See [20, Sec. 2]. □

Hence, the error exponent is given by the limit of the normalized sum of functions over the dependency cliques. This suggests that inference performance may be separable over the cliques and a tractable optimization problem may be the *optimal clique selection* policy. In this case, based on the exponent in (20), each unselected clique is assigned a penalty

$$\pi^{clq}(c) := \big(\mathbb{E}(\phi_c(\mathbf{Y}_c)|V;\mathcal{H}_0)\big)^+. \tag{21}$$

We provide theoretical guarantees for clique selection for a special class of dependency graphs. This in turn inspires the development of a general class of heuristics for any dependency graph in Section IV-D.

C. Special Case of MRF: Disjoint Cliques

In this section, we consider the special case when all the cliques in the dependency graph $\text{DG}(V)$ are disjoint. This can occur for instance, when nodes are placed according to a cluster process and the dependency graph is given by a disk graph. See Section V. Here, the form of the exponent in (20) is simplified further. Moreover, we simplify the form of LLR in (18) for any node subset V_s spanning a set of cliques.

For disjoint cliques, the dependency graph $\text{DG}'(V_s)$ is a subgraph of $\text{DG}(V)$ for any node subset V_s spanning a sub-collection of cliques $\mathcal{C}_s \subset \mathcal{C}$, and hence,

$$\text{LLR}(\mathbf{Y}_{V_s}) = \sum_{c\in\mathcal{C}_s} \phi_c(\mathbf{Y}_c). \tag{22}$$

Hence, it is simpler to design fusion schemes in this case since the dependency structure does not change for different subsets.

For disjoint cliques, the penalty function in (21) is the KL-distance of measurements in clique c.

$$\pi^{clq}(c) = D(f_c(\mathbf{Y}_c;\mathcal{H}_0)||f_c(\mathbf{Y}_c;\mathcal{H}_1)) := D_c, \tag{23}$$

Hence, if nodes in a clique is not selected, then a penalty equal to its KL-distance is paid.

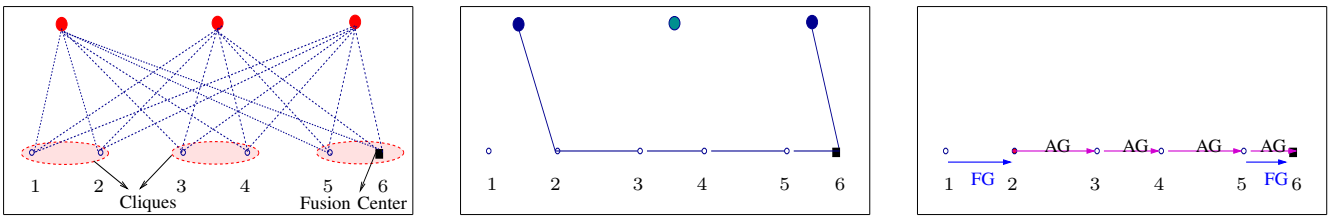
We now formally define the optimal clique selection policy $\text{opt\_clique}(V, C, \gamma\pi)$ as the optimization in (3) but with the additional constraint that the subsets V_s considered are only those that span a sub-collection of cliques $\mathcal{C}_s \subset \mathcal{C}$.

Along the lines of our approach for the i.i.d. case, we first consider the optimal solution $\text{opt\_clique}(V, C, \gamma\pi^{clq})$ with π^{clq} given by (23) and prove its reduction to a PCST.

Theorem 4 (PCST Reduction): $\text{opt\_clique}(V, C, \gamma\pi^{clq})$ has an approximation-ratio preserving prize-collecting Steiner tree reduction.

Proof: By simplifying an integer program [20, Sec. 3]. □

The above result implies that any approximation algorithm for finding the PCST can be adapted to find an approximation for $\text{opt\_clique}(V, C, \gamma\pi^{clq})$, with its approximation ratio preserved. One such instance called the approximate prize-collecting data fusion (Approx_PCDF) is given in Fig.4. It builds an approximate PCST on an augmented graph using the GW-algorithm in [2]. The augmented graph is given by the function Map in Fig.5, where for each non-trivial clique c (size greater than one) of the MRF, it adds a virtual node



(a) Augmented graph using Map (b) PCST on augmented graph (c) Scheme to compute LLR via RevMap

Fig. 3. Illustration of Clique Selection and Data Fusion via PCST Reduction for Binary Cliques.

v_c and connects it to the nodes $v \in V$. The costs of new edges reflect the cost of forwarding raw data to candidate processors to compute the clique potentials. The penalty of each virtual node v_c is $\pi^{clq}(c)$ in (23) and the penalties of all nodes $v \in V$ are zero. After building the approximate PCST on the augmented graph, the function RevMap in Fig.6 maps it to a valid output: the set of selected cliques and the fusion scheme to compute its LLR. An example of the PCST reduction is shown in Fig.3.

The added nodes and edges in the augmented graph using Map function represent additional routing costs in the presence of correlation. This is because of the increased complexity of the LLR to be computed by the aggregation scheme.

The approximation PCST is built on a graph with $|V|+|\mathcal{C}^{nt}|$ number of nodes, where $\mathcal{C}^{nt}$ is the set of non-trivial cliques using the the GW-algorithm. Hence, the approximation ratio of Approx_PCDF(Map) with respect to opt_clique$(V, C, \gamma\pi^{clq})$ is $2-(|V|+|\mathcal{C}^{nt}|-1)^{-1}$. We now improve its approximation ratio based on some simple observations regarding the GW-algorithm. Define the collection of *profitable cliques* $\mathcal{C}_p \subset \mathcal{C}$ as those generating a net "profit" after reducing their scaled KL-distance by the costs of raw-data routing to any processor

$$\mathcal{C}_p := \{c : c \in \mathcal{C}, |c| = 1 \text{ or } |c| > 1 \text{ and } \gamma D_c \geq \min_{i \in V} \sum_{v_k \subset c_j, k \neq i} C(v_i, v_k)\}, \quad (24)$$

and let Map$'$ be the modified version of Map which only adds virtual nodes for non-trivial profitable cliques $c \in \mathcal{C}_p, |c| > 1$. Below, we give the improved approximation ratio.

Theorem 5 (Improved Approx. Ratio): The approximation ratio on using the Map$'$ function is

$$\rho(\text{Approx\_PCDF(Map}')) = 2 - \frac{1}{\max(|\mathcal{C}_p| - I(v_0 \in \mathcal{C}_p), 1)}.$$

Proof: Only profitable cliques can be selected in the optimal solution. See [20, Sec. 4]. □

Hence, the approximation ratio depends only on the number of profitable cliques. So far, we have considered optimality based on the penalty function in (23). We now prove its asymptotic optimality with respect to the function π in (4).

Theorem 6 (Asymptotic Optimality): When the number of cliques grows with network size ($|\mathcal{C}| \to \infty$, as $|V| \to \infty$), and the link costs are bounded, we have

Require: $V = \{v_0, \ldots, v_{|V|-1}\}$ nodes, v_0= Fusion center,
$\mathcal{M} = \{c_0, \ldots, c_{|\mathcal{M}|-1}\}$= Candidate node groups
G_n= Metric closure of network, C = Link costs
Π_m = Penalty of group m, γ = tradeoff factor
$\{G', V_m, \pi\} \leftarrow$ Map$(G_n; \mathcal{M}, C, \Pi, \gamma)$
PCST$(G; C, \pi)$ = (Approx.) Prize-collecting Steiner tree on G using GW algorithm with cost C, node penalty fn. π
DPCST = PCST(G') directed towards v_0
$\{\mathcal{M}_s, \Gamma\} \leftarrow$ RevMap(DPCST; $V_m, V, \mathcal{M}$, Algo)
return $\{\mathcal{M}_s, \Gamma\}$

Fig. 4. Approx_PCDF(Map,Algo): outputs selected groups $\mathcal{M}_s$ and fusion scheme Γ. For Algo=Clique Selection, $\mathcal{M} = \mathcal{C}$ is the clique set of DG(V) and $\Pi = \pi^{clq}$ in (21). For Algo=Component Selection, $\mathcal{M}$ is the set of components of DG(V) and $\Pi = \pi$_cmp in (26).

$$\lim_{|V| \to \infty} \frac{\text{opt\_clique}(V, C, \gamma\pi)}{\text{opt\_clique}(V, C, \gamma\pi^{clq})} = 1, \quad \forall \gamma > 0. \quad (25)$$

Proof: Along the lines of Theorem 2. See [20, Sec. 5]. □

Hence, using the penalty function π^{clq} in (23) is efficient for networks with large number of cliques. An example where this does *not* occur is when the dependency graph is complete, and hence, has a single clique. In other words, we need limited dependencies in for the result in (25) to hold.

D. Clique & Component Selection Heuristics: General MRF

The results in the previous section inspire the development of heuristics for a general MRF. We provide two heuristics, viz., clique selection and component selection, summarized in Fig.4, Fig.5, and Fig.6.

The error exponent in (20) for any MRF is separable in its cliques. Hence, using the penalty function in (21), we have the clique selection heuristic for any MRF, as outlined in the previous section. However, some complexities arise.

In Section IV-B, we noted that for a general MRF, the dependency graph DG$'(V_s)$ for any subset $V_s \subset V$ is not, in general, a subgraph of DG(V). Hence, the fusion scheme via PCST reduction may not compute the LLR of the selected subset V_s since the algorithm takes DG(V) as the input. In Fig.6, we add additional lines from (17) to (26) to ensure that LLR(V_s) is indeed computed. For each clique $c \in$ DG$'(V_s)\backslash$DG(V), we ensure that its potential is computed by adding edges from its members to a processor to the

forwarding subgraph (FG). However, since new edges are added, routing costs increase, and we can no longer provide optimality results for the clique selection heuristic for a general MRF, as we did in the previous section.

function MAP($G_n(V); \mathcal{M}, C, \Pi, \gamma$)
  $\mathcal{N}_u(v; G)$ = Neighborhood of v in undirected G
  $G' \leftarrow G_n$, $V_m \leftarrow \emptyset$, $n \leftarrow |V|$, $\pi(v) \leftarrow 0$, $\forall v \in V$
  for $j \leftarrow 0$ to $|\mathcal{M}| - 1$ **do** ▷ Let V and $\mathcal{M}$ be ordered
    if $|m_j| > 1$ **then**
      $V_m \leftarrow v_{n-1+j}$
      Add new node v_{n-1+j} to G'
      Assign penalty $\gamma\pi(v_{n-1+j}) \leftarrow \gamma\Pi_{m_j}$,
      for all $v_i \in V$ **do**
        Add node v_i to $\mathcal{N}_u(v_{n-1+j}; G')$
        $C(v_{n-1+j}, v_i; G') \leftarrow \sum_{v_k \subset c_j, k \neq i} C(v_i, v_k; G_n)$
      end for
    else
      $V_m \leftarrow v_i$, $\pi(v_i) \leftarrow \gamma\Pi_{m_j}$, $v_i \subset m_j$ ▷ 1-groups
    end if
  end for
return $\{G', V_m, \pi\}$
end function

Fig. 5. Map($G_n; \mathcal{M}, C, \Pi, \gamma$) adds nodes for each non-trivial group, and returns augmented graph G' with penalty π and group-representative set V_c.

An alternate strategy is to search over only those subsets V_s for which $\mathrm{DG}'(V_s) \subset \mathrm{DG}(V)$. In such a case, lines from (17) to (26) in Fig.6 are not needed. This holds when V_s spans a sub-collection of components of the original dependency graph $\mathrm{DG}(V)$. We term such policies considering nodes in different components of the dependency graph for selection as *component selection* policies. The penalty for a component is defined by collecting the terms in (20) of all the cliques contained in it, given by

$$\pi\_\mathrm{cmp}(m) := \sum_{c \subset m, c \in \mathcal{C}} \mathbb{E}[\phi_c(\mathbf{Y}_c); \mathcal{H}_0] = D_m, \qquad (26)$$

where D_m is the KL-distance of the component m. In Fig.4, for each selected component, a unique processor is assigned, to which all the component members send raw data. The values from the processors are then summed up and delivered to the fusion center as before. Below, we prove optimality results for component selection, when restricted to a subset of strategies.

Theorem 7 (Optimality of Component Selection): The optimal solution over all policies selecting components of the dependency graph and assigning a unique processor for each selected component is given by the PCST in the augmented graph in Fig.4.

Proof: Along the lines of Theorem 4. □

Hence, we can provide a restricted optimality result for the clique-selection policy. The component and clique selection policies represent group selection of nodes with aggregation for efficient cost-performance tradeoff.

function REVMAP($G'; V_c, V, \mathcal{M}$,Algo)
  $\mathcal{N}_s(v; G)$, $\mathcal{N}_p(v; G)$ = Imm. successor, predecessor
  $< i, j >$= Directed edge from i to j
  Initialize $G \leftarrow G'$, $n \leftarrow |V|$, $\mathcal{M}_s \leftarrow \emptyset$
  for all $v_j \in V_c$ with $\mathcal{N}_s(v_j; G') \neq \emptyset$ **do**
    if $j > n - 1$ **then**
      $k \leftarrow j - n + 1$, $\mathcal{M}_s \leftarrow \mathcal{M}_s \cup m_k$
      Proc(m_k) $\leftarrow \mathcal{N}_s(v_j; G')$, for $m_k \in \mathcal{M}$,
      $V_j \leftarrow c_k \backslash \mathrm{Proc}(m_k)$, Delete $< v_j, \mathrm{Proc}(m_k) >$ in G, add $< V_j, \mathrm{Proc}(m_k) >$, mark them
      if $\mathcal{N}_p(v_j; G) \neq \emptyset$ **then** Replace $< \mathcal{N}_p(v_j), v_j >$ in G with edges $< \mathcal{N}_p(v_j), \mathrm{Proc}(m_k) >$
      end if
    else
      Proc(m_l) $\leftarrow v_j$, for $v_j \subset m_l$, $\mathcal{M}_s \leftarrow \mathcal{M}_s \cup m_l$
    end if
  end for
  FG $\leftarrow$ Marked edges of G, AG $\leftarrow G \backslash$FG
  Retain only one edge in FG if there are parallel links
  Let V(Proc) be set of all processors
  Let $V_s \leftarrow$ nodes in V spanning the groups $\mathcal{M}_s$
  if Algo=Clique Selection **then**
    Let $\mathcal{C}'$ be clique set of $\mathrm{DG}'(V_s)$
    for all $c \in \mathcal{C}' \backslash \mathcal{M}_s$ **do**
      $\mathrm{Proc}(c) \leftarrow \arg \min_{i \in \mathrm{V(Proc)}} \sum_{j: j \subset c, <j,i> \notin \mathrm{FG}} C(i, j)$
      Add $< j, \mathrm{Proc}(c) >$, $j \subset c \backslash \mathrm{Proc}(c)$ to FG if not already present
    end for
    $\mathcal{M}_s \leftarrow \mathcal{C}'$
  end if
  $\Gamma \leftarrow \{\mathrm{Proc}, \mathrm{FG}, \mathrm{AG}\}$
return $\{\mathcal{M}', \Gamma\}$
end function

Fig. 6. RevMap($G; V_c, V, \mathcal{M}$) returns the selected groups $\mathcal{M}_s$ and maps tree G' to fusion scheme Γ with processor assignment Proc, forwarding subgraph FG and aggregation subgraph AG.

V. NUMERICAL ANALYSIS

A. Simulation Environment

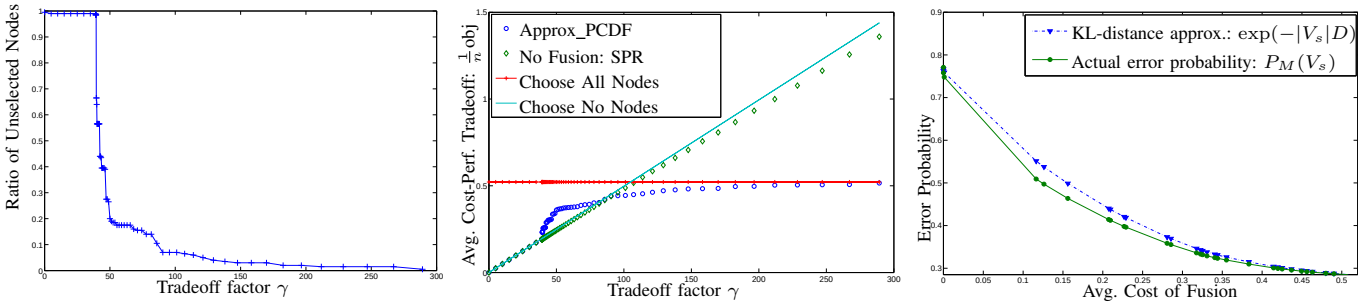
We assume that the sensor measurements are Gaussian under either hypothesis with the same covariance matrix

$$\mathbf{Y}_V \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_V), \quad \text{under } \mathcal{H}_i, i = 0, 1. \qquad (27)$$

This scenario arises when the sensors measure a deterministic signal with additive (correlated) Gaussian noise under each hypothesis. The KL-distance D and the type-II error probability P_M have closed forms for Gaussian variables.

In our setup, n (expected) number of nodes are distributed in a square. We consider two node placement distributions: uniform and Matern cluster process[5] [21]. See Fig.8.

[5]Here, a parent Poisson process first generates points. A child Poisson process then generates nodes in a disc around each point of the parent process.



(a) % of Nodes Not Selected vs. Tradeoff Factor (b) Objective Value obj vs. Tradeoff Factor (c) Cost vs. Performance for Selected Set V_s

Fig. 7. Cost-performance tradeoff for i.i.d. measurements under uniform placement for $n = 200$ nodes. See Theorem 1. For objective function obj, see (5).

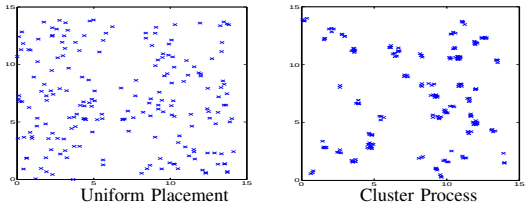


Fig. 8. Samples of i.i.d uniform placement and Matern cluster process.

The routing cost between any two nodes i and j for direct transmission is given by the power-weighted distance $\text{dist}(i,j)^\nu$. We present the result when the set of feasible direct connections is the complete graph and and $\nu = 2$: similar trends were observed for any connected graph and $\nu \in [2,4]$.

B. Results: IID Measurements

We first consider the case when all the measurements are i.i.d. conditioned on each hypothesis and fix $\mathbf{\Sigma}_V = \mathbf{I}$.

Fig.7a shows that more nodes are selected as the tradeoff factor γ increases since the penalty is given by $\gamma\pi$. Fig.7b compares four different schemes: selection with fusion (Approx_PCDF), selection with no fusion based on shortest-path routing in (12), choosing all nodes (MST-based fusion) and choosing none of the nodes (paying penalty for all the nodes). We find that the tradeoff function obj in (5) increases almost linearly with γ for selection with no fusion. On the other hand, obj for Approx_PCDF tends to a constant equal to that of MST-based fusion as γ increases and performs better than all the other schemes for a wide range of γ. Hence, in-network fusion results in an efficient cost-performance tradeoff.

In Fig.7c, we plot the average (per-node) routing cost for aggregation of selected measurements versus the resulting error probability. We see that the KL-distance based approximation is close to the actual error probability P_M. The entire range of cost-performance tradeoff is achieved by varying γ.

C. Results: Correlated Measurements

We employ the GMRF model in [22], where the dependency graph $\text{DG}(V)$ is a disk graph[6] with radius δ and the coefficients of the potential matrix $\mathbf{A}_V := \mathbf{\Sigma}_V^{-1}$ are given by

$$A_V(i,j) = \begin{cases} 1 - \sum_{k:(i,k)\in \text{DG}(V)} A(i,k), & i = j, \\ -2(1 - \frac{\text{dist}(i,j)}{\delta}), & j \neq i, \text{dist}(i,j) \leq \delta, \\ 0, & \text{o.w.} \end{cases} \quad (28)$$

For Gaussian measurements, the maximum clique size is two and higher order clique potentials are zero [17]. Hence, the clique selection heuristic in Fig.4 reduces to selection of the dependency edges, and is called the *edge selection policy*.

In Fig.9a, we compare the component and edge selection heuristics under uniform placement. We fix the disk radius $\delta = 1.2$ and here, the disk graph is connected (single component). We find[7] that the edge selection heuristic performs better and selects some nodes while the component selection heuristic selects none of the nodes.

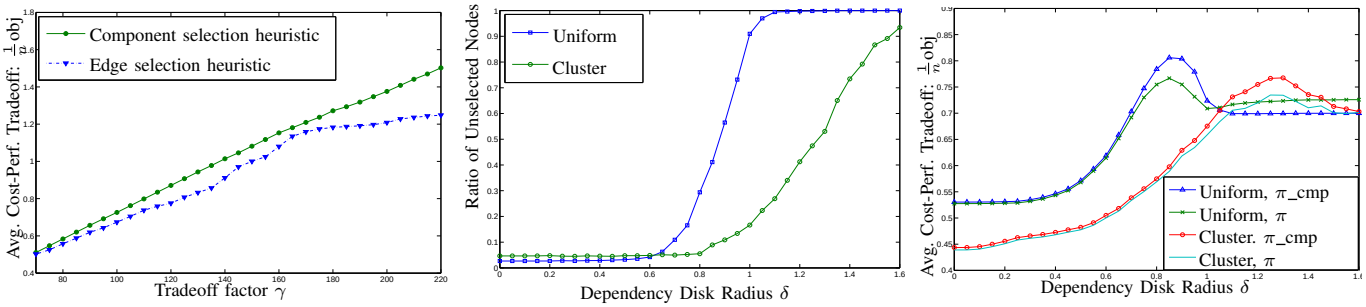
In Fig.9b, for component selection heuristic, we observe that more nodes are chosen under clustered placement compared to uniform placement. In Fig.9c, we find that the component selection heuristic performs better for clustered placement than uniform placement in terms of the tradeoff function obj in (5). As in the i.i.d. case, the KL-distance based penalty π_cmp is close to π, based on the error probability.

D. Implications

The relative performance of component and edge selection heuristics is a function of the dependency graph and node placement. As we saw in Section IV-C, the two heuristics are the same and have guaranteed approximation ratio when the dependency graph has disjoint maximal cliques. This occurs in our simulation setup for the cluster process and small radius δ of the dependency graph. The simulations in this regime match our theory. However, as the radius δ increases, the dependency graph becomes connected with a single component. Here, the component selection heuristic has only one candidate for selection, while the edge selection heuristic has many choices. Depending on the tradeoff factor, the component selection heuristic may not choose any of the nodes while the clique selection chooses some nodes, and hence, has a better performance.

[6]A disk graph has edges between nodes within δ inter-node distance.

[7]Not shown here for the lack of space.



(a) Component/Edge Selection $\delta = 1.2, n = 50$ (b) Component Selection, $\gamma = 140, n = 200$ (c) Component Selection, $\gamma = 140, n = 200$

Fig. 9. Cost-performance tradeoff for correlated measurements for model in (28) under 60 simulation runs. See (4), (5), and (26).

Our simulations also provide insights on the influence of node placement. At low dependency (small values of disk radius δ), clustering the nodes is more efficient than uniform placement. This is because at low δ, clustering does not add too many new edges to the disk graph, and the routing costs are also significantly reduced since nodes are nearer. The situation is however reversed at high disk radius δ. Here, clustering adds more edges to the disk graph than uniform placement and hence, increases the aggregation costs. This results in a more complex sufficient statistic that needs to be computed by an aggregation scheme and hence, the routing costs increase.

Hence, the cluster process of node placement is a good candidate for achieving efficient cost-performance tradeoff at sparse spatial dependencies and our heuristics have good performance here.

VI. Conclusion

In this paper, we considered optimal node selection for tradeoff between routing costs and inference performance. We explicitly incorporated the effect of correlation between the sensor measurements via the dependency graph of a Markov random field model and considered in-network aggregation of measurements to reduce routing costs. We provided theoretical and numerical results to show the efficiency of our schemes for node selection and data aggregation.

There are many future directions to pursue such as the development of better algorithms. We have only considered offline and centralized sensor selection and extension to local selection and coordination is of interest. Also, the effect of quantization on tradeoff warrants investigation.

Acknowledgment

The authors thank Prof. A. Ephremides for helpful comments.

References

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York, NY: Springer, 1997.

[2] M. Goemans and D. Williamson, "A General Approximation Technique for Constrained Forest Problems," *SIAM Journal on Computing*, vol. 24, p. 296, 1995.

[3] P. Brémaud, *Markov Chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer, 1999.

[4] Y. Nakamura, K. Tei, Y. Fukazawa, and S. Honiden, "Region-Based Sensor Selection for Wireless Sensor Networks," 2008, pp. 326–331.

[5] A. Anandkumar, L. Tong, A. Swami, and A. Ephremides, "Cost-Performance Tradeoff in Multi-hop Aggregation for Statistical Inference," in *Proc. of IEEE ISIT*, Toronto, Canada, July 2008.

[6] V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," in *Proc. IPSN*, April 2005, pp. 151–158.

[7] P. Atrey, M. Kankanhalli, and J. Oommen, "Goal-oriented optimal subset selection of correlated multimedia streams," *ACM Tran. on Multimedia Comp. Comm. & App.*, vol. 3, no. 1, 2007.

[8] S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *Accepted to IEEE Tran. on Signal Proc.*, 2008.

[9] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: maximizing information while minimizing communication cost," in *Proc. of IPSN*, 2006, pp. 2–10.

[10] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. La Porta, "A survey of sensor selection schemes in wireless sensor networks," in *Proc. of SPIE*, 2007.

[11] R. Debouk, S. Lafortune, and D. Teneketzis, "On an Optimization Problem in Sensor Selection," *Discrete Event Dynamic Systems*, vol. 12, no. 4, pp. 417–445, 2002.

[12] Y. Sung, S. Misra, L. Tong, and A. Ephremides, "Cooperative Routing for Signal Detection in Large Sensor Networks," *IEEE J. Select. Area Comm.*, vol. 25, no. 2, pp. 471–483, 2007.

[13] J. Chamberland and V. Veeravalli, "How Dense Should a Sensor Network Be for Detection With Correlated Observations?" *Information Theory, IEEE Transactions on*, vol. 52, no. 11, pp. 5099–5106, 2006.

[14] L. Yu, L. Yuan, G. Qu, and A. Ephremides, "Energy-driven detection scheme with guaranteed accuracy," in *Proc. IPSN*, 2006, pp. 284–291.

[15] A. Giridhar and P. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Selec. Area Comm.*, vol. 23, no. 4, pp. 755–764, 2005.

[16] A. Anandkumar, L. Tong, A. Swami, and A. Ephremides, "Minimum Cost Data Aggregation with Localized Processing for Statistical Inference," in *Proc. of INFOCOM*, Phoenix, USA, April 2008, pp. 780–788.

[17] A. Anandkumar, A. Ephremides, A. Swami, and L. Tong, "Routing for statistical inference in sensor networks," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and R. Liu, Eds. John Wiley & Sons, 2008, ch. 25.

[18] A. Anandkumar, L. Tong, and A. Swami, "Optimal Node Density for Detection in Energy Constrained Random Networks," *accepted to IEEE Trans. Signal Proc.*, vol. 56, no. 10, Oct. 2008.

[19] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[20] "Appendix for INFOCOM '09 submission." [Online]. Available: http://acsp.ece.cornell.edu/papers/infocom09_appendix.pdf

[21] A. Lawson and D. Denison, *Spatial Cluster Modelling*. CRC Press, 2002.

[22] A. Pettitt, I. Weir, and A. Hart, "A Conditional Autoregressive Gaussian Process for Irregularly Spaced Multivariate Data with Application to Modelling Large Sets of Binary Data," *Statistics and Computing*, vol. 12, no. 4, pp. 353–367, 2002.