INFORMATION FLOW OPTIMIZATION IN INFERENCE NETWORKS

Aditya Deshmukh⋆, Jing Liu⋆, Venugopal V. Veeravalli⋆, Gunjan Verma†

⋆University of Illinois at Urbana-Champaign †Army Research Laboratory

ABSTRACT

The problem of maximizing the information flow through a sensor network tasked with an inference objective at the fusion center is considered. The sensor nodes take observations, compress and send them to the fusion center through a network of relays. The network imposes capacity constraints on the rate of transmission in each connection and flow conservation constraints. It is shown that this rate-constrained inference problem can be cast as a Network Utility Maximization problem by suitably defining the utility functions for each sensor, and can be solved using existing techniques. Two practical settings are analyzed: multi-terminal parameter estimation and binary hypothesis testing. It is verified via simulations that using the proposed formulation gives better inference performance than the Max-Flow solution that simply maximizes the total bit-rate to the fusion center.

Index Terms— Flow optimization, sensor networks, statistical inference, Internet of Things

1. INTRODUCTION

Consider a sensor network comprised of sensor nodes, relay/intermediate nodes and a fusion center (FC). Each sensor node takes a single observation, which is a random variable related to the state of nature/target of interest. Assume that the sensor nodes cannot communicate with each other, and that they compress their observation and send it to the intermediate nodes. The intermediate nodes just act as relays and route messages towards the fusion center. The fusion center is tasked with inferring the state of nature. The relay network is considered to be capacitated and satisfies flow conservation at each relay node, i.e. number of units of information received is equal to the number of units of information transmitted by each relay node. So, the problem is a rate-constrained inference problem. Note that the trade-off here is between the rate assigned to each sensor versus the quality of compressed observations. We show that if the inference objective can be modeled as a sum of utility functions, which capture the information content in the compressed observation of each sensor according to the rate assigned to them, then the problem can be cast as a Network Utility Maximization problem. We analyze two inference tasks: parameter estimation and detection (specifically binary hypothesis testing), and explicitly formulate the utility functions for these tasks.

There is considerable literature on multi-terminal parameter estimation, either from one-shot or having time series (e.g., [1–7]). To the best of our knowledge, they at most consider the sum rate constraint for the sensors or characterizing their rate region for a given distortion constraint, without considering the actual network structure and link capacity constraints (e.g., whether the desired rates for each sensor are achievable by the actual network).

We focus on the one-shot problem in this paper. The closet work to our multi-terminal (vector) parameter estimation is Sani and Vosoughi [8], in which it is assumed that the unknown vector is zero-mean Gaussian with known covariance matrix. We do not need this assumption and treat the vector as deterministic and unknown. Another distinguishing factor is that [8] assumes there is an overall sum rate constraint on the network for these sensors, while our information flow optimization scheme takes into account the actual structure and link capacity constraints of the network.

Prior work that studies detection in sensor networks, includes Tay and Tsitsiklis [9], in which the one-shot problem of decentralized detection in tree structured sensor networks is considered. They characterized the optimal error exponent under a Neyman-Pearson formulation, but did not analyze capacitated networks. Another line of work was initiated by Han and Amari [10], in which the problem of multi-terminal hypothesis testing is studied with one-hop rate constraints to a fusion center and sequential observations. The emphasis of this work is on extending rate distortion theory to this distributed setting.

2. PROBLEM SETUP

We model the network as an undirected graph $G = (V, E)$. In this graphical model, each edge $(u, v) \in E$ has an integral capacity c_{uv} associated with it, which is the maximum possible rate in that connection(either way). Let the rate in edge $(u, v) \in E$ be r_{uv}. We define the rates to be anti-symmetric, i.e. $r_{uv} = -r_{vu}$. The rates obey flow conservation at each intermediate node. Let S denote the set of sensor nodes. Let $N = |S|$ be the number of sensor nodes. Let t denote the fusion center. The performance of the graph is measured in terms of the information content regarding the state of nature (to be inferred) in the messages sent by the sensor nodes. In

this regard we assume there exists a concave utility function $g_s : \mathbb{R} \to \mathbb{R}$, that captures the notion of information content in the message from sensor node s. The goal is to maximize the sum utilities of the messages received at the fusion center. Let the sum rate from a node $u \in V$ be denoted by

$$r_u = \sum_{v \in V:(u,v)\in E} r_{uv} \tag{1}$$

We pose the optimization problem as follows

$$\max_{r_{uv}} \sum_{s \in S} g_s(r_s) \tag{2a}$$

$$\text{subject to } \forall (u,v) \in E, \ r_{uv} \leq c_{uv} \tag{2b}$$

$$\forall (u,v) \in E, \ r_{uv} = -r_{vu} \tag{2c}$$

$$\forall u \in V \setminus \{S \cup t\}, \ r_u = 0. \tag{2d}$$

In this problem formulation, we only consider the one-shot problem, i.e. only a single observation (or a single block of observations) is collected, compressed and sent through the network by each sensor. Assuming the rate is real-valued, Problem (2) is a convex program and therefore is straightforward to solve. It is worth mentioning that, many decentralized or distributed methods have been proposed to solve this kind of Network Utility Maximization problem, e.g., [11–14] and the references therein. For example, it is straightforward to use Dual Decomposition method to solve the above convex problem [11, 15]. If the rates and capacities are constrained to be integer-valued, we can use the solution of the real-valued relaxation to obtain an integral solution in polynomial time [16, Theorem 5], such that the integral rates $r_{uv}^{(I)}$ are close to the real-valued rates r_{uv} in the sense that $\sum_{s \in S} r_s^{(I)} = \left\lfloor \sum_{s \in S} r_s \right\rfloor$.

We now discuss two instances of inference tasks in sensor networks: multi-terminal parameter estimation and binary hypothesis testing.

3. FLOW OPTIMIZATION FOR PARAMETER ESTIMATION

We consider a sensor network with N spatially distributed sensors. Each sensor i takes a linear measurement of the unknown deterministic vector x, and quantizes the corresponding scalar measurement y_i into certain number of bits. Then the sensors want to send those bits to the fusion center through a capacitated relay network. Finally, the fusion center estimates x based on the quantized data it received from all the sensors. Mathematically, we assume the following linear observation model on each sensor i:

$$y_i = \boldsymbol{a}_i^T \boldsymbol{x} + \eta_i, i = 1, ..., N, \tag{3}$$

which can be jointly written as:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta}, \tag{4}$$

where the vector $\boldsymbol{y} \in \mathbb{R}^N$ are the overall measurements from all the N sensors, and the sensing matrix $\boldsymbol{A} \in \mathbb{R}^{N \times q}$ is known. The vector $\boldsymbol{\eta} \in \mathbb{R}^N$ models the i.i.d. *bounded* noise with zero mean and variance σ^2.

For simplicity, we assume that the value of y_i is within the sensor input range, and the quantizer uniformly divides that range into small intervals for quantization. For each y_i, we denote its quantized version as d_i, and the corresponding quantization noise is $\epsilon_i = d_i - y_i$. So we have

$$d_i = \boldsymbol{a}_i^T \boldsymbol{x} + \eta_i + \epsilon_i, i = 1, ..., N, \tag{5}$$

For small quantization intervals Δ_i, the quantization noise ϵ_i is approximately uncorrelated with the input y_i and has zero-mean with variance $\Delta_i^2/12$ [17].

The fusion center wants to accurately estimate x from all the quantized data it received via the least squares $\hat{x} = \boldsymbol{A}^\dagger \boldsymbol{d}$. The question is: *Given the limited capacity of the network, to minimize the estimation error at the fusion center, how should we optimize the network flows and how many bits should we allocate to each sensor?*

The well studied max-flow method may not give the optimal solution for this problem, even though it finds the maximal possible number of bits that can be sent from those sensors to the FC.

Our goal here is not to minimize the reconstruction error of y, but rather minimizing the mean squared error $\mathbb{E}[\|\hat{x} - x\|_2^2]$ via optimizing the network *information* flows. We have that

$$\mathbb{E}[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2] \tag{6}$$

$$= \mathbb{E}[\|\boldsymbol{A}^\dagger \boldsymbol{d} - \boldsymbol{x}\|_2^2] \tag{7}$$

$$= \mathbb{E}[\|\boldsymbol{A}^\dagger (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta} + \boldsymbol{\epsilon}) - \boldsymbol{x}\|_2^2] \tag{8}$$

$$= \mathbb{E}[\|\boldsymbol{A}^\dagger (\boldsymbol{\eta} + \boldsymbol{\epsilon})\|_2^2] \tag{9}$$

$$= \mathbb{E}[\text{Tr}\{\boldsymbol{A}^\dagger(\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{A}^\dagger(\boldsymbol{\eta} + \boldsymbol{\epsilon}))^T\}] \tag{10}$$

$$= \mathbb{E}[\text{Tr}\{\boldsymbol{A}^\dagger(\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta} + \boldsymbol{\epsilon})^T(\boldsymbol{A}^\dagger)^T\}] \tag{11}$$

$$= \text{Tr}\{\boldsymbol{A}^\dagger \mathbb{E}[(\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta} + \boldsymbol{\epsilon})^T](\boldsymbol{A}^\dagger)^T\} \tag{12}$$

$$\approx \text{Tr}\{\boldsymbol{A}^\dagger \text{diag}(\sigma^2 + \Delta_1^2/12, ..., \sigma^2 + \Delta_N^2/12)(\boldsymbol{A}^\dagger)^T\} \tag{13}$$

$$= \sum_{i=1}^N \left[(\sigma^2 + \Delta_i^2/12)\|\boldsymbol{A}^\dagger(:,i)\|_2^2 \right] \tag{14}$$

Since $\Delta_i \propto \frac{1}{2^{r_i}}$, where r_i is the number of bits allocated to sensor i, minimizing (14) is equivalent to maximizing the following:

$$-\sum_{i=1}^N \|\boldsymbol{A}^\dagger(:,i)\|_2^2/4^{r_i}, \tag{15}$$

Now the overall problem is cast as Problem (2).

4. FLOW OPTIMIZATION FOR DETECTION

We first discuss some known results of optimal quantizers for detection in the classical setting before introducing the sensor network setting. Let $(\mathcal{Y}, \mathcal{F})$ be a measurable space. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be two probability measures defined on $(\mathcal{Y}, \mathcal{F})$. Under hypothesis H_i, $i = 0, 1$, a $\mathcal{Y}$-valued random variable Y has the distribution $\mathbb{P}_i$. We define a deterministic n-level quantizer as an $\mathcal{F}$-measurable function that maps $\mathcal{Y} \to [n] := \{1, 2, \ldots, n\}$. Let Γ_n be the set of all randomized n-level quantizers. Let the distribution of $\gamma(Y)$ under hypothesis H_i be $Q_i(\gamma)$, for $\gamma \in \Gamma_n$ for some n. Consider the problem of finding a randomized n-level quantizer to maximize the Kullback-Leibler(KL) divergence between $Q_1(\gamma)$ and $Q_0(\gamma)$, denoted by $D(Q_1(\gamma)||Q_0(\gamma))$. It is well-known that the optimal error exponent achieved in Neyman-Pearson testing with access to quantized observations is $D(Q_1(\gamma)||Q_0(\gamma))$. In this regard, we define a utility function $f_{\mathbb{P}_1, \mathbb{P}_0} : \mathbb{N} \to \mathbb{R}$ as follows:

$$f_{\mathbb{P}_1, \mathbb{P}_0}(n) := \sup_{\gamma \in \Gamma_n} D(Q_1(\gamma)||Q_0(\gamma)) \qquad (16)$$

Assume for simplicity that $\mathbb{P}_0$ and $\mathbb{P}_1$ have densities p_0 and p_1 with respect some common measure μ. The likelihood ratio is a measurable function $L : \mathcal{Y} \to [0, \infty]$ given by

$$L(y) = \begin{cases} p_1(y)/p_0(y), & p_1(y) \neq 0, p_0(y) \neq 0 \\ 0, & p_1(y) = 0, p_0(y) \neq 0 \\ \infty & p_1(y) \neq 0, p_0(y) = 0. \end{cases} \qquad (17)$$

We define the *threshold set* T_n to be the set of vectors $\boldsymbol{t} = (t_0, t_1, \ldots, t_n) \in [0, \infty]^{n+1}$ satisfying $0 = t_0 \leq t_1 \leq \cdots \leq t_{n-1} \leq t_n = \infty$. We define a *likelihood ratio quantizer* $\gamma_{\boldsymbol{t}}$ with threshold vector $\boldsymbol{t} \in T_n$, as a quantizer that satisfies the following $\forall \ell \in [n], \forall i \in \{0, 1\}$:

$$\mathbb{P}_i(\gamma_{\boldsymbol{t}}(Y) = \ell \text{ and } L(Y) \notin [t_{\ell-1}, t_\ell]) = 0. \qquad (18)$$

Tsitsiklis [18] proved that the optimal quantizer for maximizing $D(Q_1(\gamma)||Q_0(\gamma))$ for a given number of quantization levels is the likelihood ratio quantizer, i.e.

$$\max_{\gamma \in \Gamma_n} D(Q_1(\gamma)||Q_0(\gamma)) = \max_{\boldsymbol{t} \in T_n} D(Q_1(\gamma_{\boldsymbol{t}})||Q_0(\gamma_{\boldsymbol{t}})). \qquad (19)$$

Thus, we get $f_{\mathbb{P}_1, \mathbb{P}_0}(n) = \max_{\boldsymbol{t} \in T_n} D(Q_1(\gamma_{\boldsymbol{t}})||Q_0(\gamma_{\boldsymbol{t}}))$.

We now consider binary hypothesis detection in the sensor network setting. Let $\mathbb{P}_i^{(j)}$ denote the distribution of observation $Y_j \in \mathcal{Y}_j$ from sensor j under the hypotheses H_i, $i = 0, 1$. As before, assume that $\mathbb{P}_0^{(j)}$ and $\mathbb{P}_1^{(j)}$ are absolutely continuous with respect to some common measure $\mu^{(j)}$. The fusion center is tasked with the objective of finding the true hypothesis. In this case, the Neyman-Pearson problem with the rate conservation and capacity constraints is an intractable problem. We consider the problem of maximizing the KL-divergence between the distributions of the quantizer outputs under the two hypothesis,

$$D\left(Q_1^{(1)} \otimes Q_1^{(2)} \otimes \ldots Q_1^{(n)}||Q_0^{(1)} \otimes Q_0^{(2)} \otimes \ldots Q_0^{(N)}\right) \quad (20)$$

$$= \sum_{j=1}^{N} D\left(Q_1^{(j)}||Q_0^{(j)}\right), \qquad (21)$$

where $Q_i^{(j)}$ is a shorthand notation for $Q_i^{(j)}(\gamma_j)$, the distribution of the quantized observation $\gamma_j(Y_j)$ from the jth sensor. Note that (20) is the optimal error exponent achieved in Neyman-Pearson testing with access to quantized observations from all sensors.

In light of (19), our objective can be re-written as follows:

$$\max_{r_j, \gamma_j \in \Gamma_{r_j}} \sum_{j=1}^{N} D\left(Q_1^{(j)}(\gamma_j)||Q_0^{(j)}(\gamma_j)\right) \qquad (22)$$

$$= \max_{r_j, \boldsymbol{t_j} \in T_{r_j}} \sum_{j=1}^{N} D\left(Q_1^{(j)}(\gamma_{\boldsymbol{t_j}})||Q_0^{(j)}(\gamma_{\boldsymbol{t_j}})\right) \qquad (23)$$

$$= \max_{r_j} \sum_{j=1}^{N} f_{\mathbb{P}_1^{(j)}, \mathbb{P}_0^{(j)}}(r_j) \qquad (24)$$

Now the problem posed can be cast as Problem (2) with $g_s(r_s) = f_{\mathbb{P}_1^{(s)}, \mathbb{P}_0^{(s)}}(r_s)$, where $r_s = \log_2(n_s)$ and n_s are the number of quantization levels for sensor s. Note that the domain of the function (16) is discrete and so we use linear-interpolation to get a surrogate objective function for the real-valued relaxation of Problem (2). It can be easily shown that (16) is increasing, but it is difficult to show that it is concave. We verified via simulations that it is indeed concave for many cases of distributions. For example, Fig. 1 shows the plot of $f_{\mathbb{P}_1, \mathbb{P}_0}(n)$ for the case when p_0 and p_1 are unit variance Gaussian densities with means 0 and 3 respectively. Fig. 2 shows the plot of $f_{\mathbb{P}_1, \mathbb{P}_0}(n)$ for the case when p_0 and p_1 are exponential densities with rates 0.5 and 1 respectively.
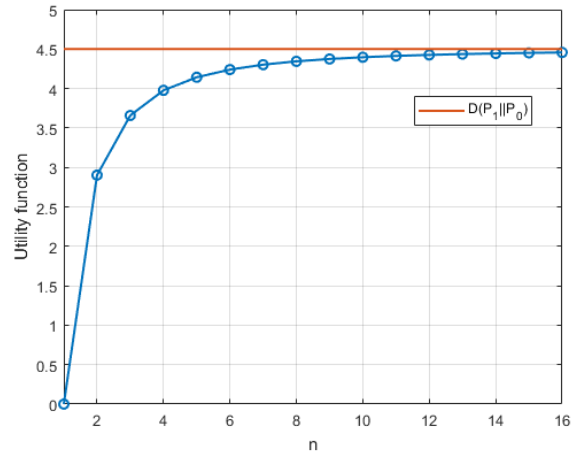

Fig. 1: Plot of $f_{\mathbb{P}_1, \mathbb{P}_0}(n)$, p_0 is $\mathcal{N}(0, 1)$ and p_1 is $\mathcal{N}(3, 1)$.
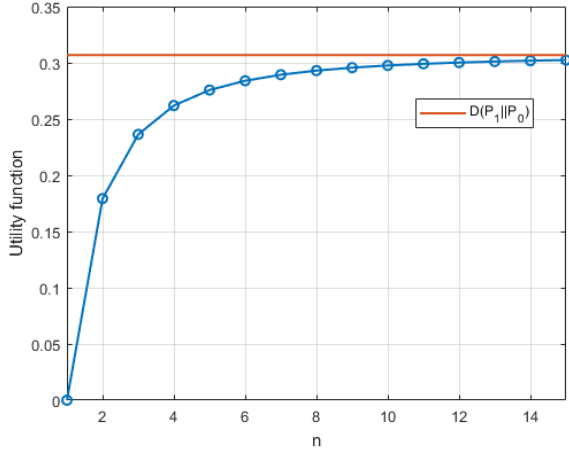
Fig. 2: Plot of $f_{\mathbb{P}_1, \mathbb{P}_0}(n)$, p_0 is Exp(0.5) and p_1 is Exp(1).

5. EXPERIMENTS

We generate a random graph having the following structure: N_1 sensor nodes form the first layer, N_2 intermediate nodes form the second layer, N_3 intermediate nodes form the third layer, and N_4 intermediate nodes, which form the fourth layer, are connected to the fusion center. Each node from first layer is randomly connected to K nodes from the second layer. Now each connected node in the second layer is in turn randomly connected to K nodes from the third layer. Similar connections are established between the third and fourth layer.

5.1. Information Flow Optimization for Parameter Estimation

We set the network structure parameters as follows: $N_1 = 10$, $N_2 = 50$, $N_3 = 30$, $N_4 = 10$ and $K = 4$. Capacities are assigned to the edges uniformly random from $\{1, 2, \ldots, 15\}$. We study the sensing matrix A considered in [19], where there are some sensors close to the target and therefore have stronger signal. The rest of sensors have less proximity to the target and the signal is weak. To simulate 4 sensors with weak signal and 6 sensors with strong signal, we first generate a 10 by 3 matrix whose entries are i.i.d. drawn from the uniform distribution $U(0, 1)$, then we multiply its first 4 rows by a factor α to obtain the sensing matrix $A \in \mathbb{R}^{10 \times 3}$. The unknown vector $x \in \mathbb{R}^3$ is randomly drawn from $\mathcal{N}(0, I)$. The bounded noise η_i is i.i.d. generated from uniform distribution $U(-0.1, 0.1)$. The uniform quantizer with range [-5,5] is employed to quantize y_i into r_i number of bits.

We first generate the network and sensing matrix A with $\alpha = 1$, and report the Mean Squared Error (MSE) of the estimated $\hat{x}$ by the proposed method [1] under 100000 Monte Carlo

runs with different realizations of x and η. Then we multiply the first 4 rows of A by a factor α and redo the simulations. We also compare with the Max-Flow solution, which maximizes the total number of bits send through the network to the Fusion Center, and allocates the corresponding bits for the sensors to quantize their measurements. Table 1 lists the total number of allocated bits and the MSE of the estimated $\hat{x}$ under one realization of the network and A. On this example, the proposed information flow optimization scheme performs much better than the Max-Flow solution[2], even though it may not allocate maximum number of bits. As the value of α decreases, the difference between the sensors increases, and the benefits of the proposed method become more clear.

Table 1: MSE of the estimated $\hat{x}$ and number of bits allocated under one realization of the network and A.

MSE/bits	α=1	α=0.3	α=0.1
Max-Flow	0.2673 / 74	1.1939 / 74	1.5068 / 74
Proposed	0.0148 / 73	0.0230 / 73	0.0241 / 74

5.2. Information Flow Optimization for Binary Hypothesis Testing

We set the network structure parameters as follows: $N_1 = 10$, $N_2 = 4$, $N_3 = 3$, $N_4 = 2$, and $K = 1$. We consider 3 different settings. For all settings, under H_0, the observations from all sensors follow standard normal distribution. For setting 1, under H_1, the observations from first five sensors follow the distribution $\mathcal{N}(11, 1)$, and the rest follow $\mathcal{N}(2, 1)$. For setting 2, under H_1, the observation from the jth sensor follows the distribution $\mathcal{N}(j + 1, 1)$. For setting 3, under H_1, the observations from first five sensors follow the distribution $\mathcal{N}(2, 1)$, and the rest follow $\mathcal{N}(11, 1)$. Capacities are assigned to the edges uniformly at random from $\{1, 2, 3, 4, 5\}$. Table 2 lists the KL-divergence between the distributions of the quantized observations under H_1 and H_0, achieved by the proposed information flow optimization scheme and Max-Flow algorithm, for one realization of a network. In the obtained realization of the network, the first five nodes have smaller overall capacity than the rest of the nodes. In this case too, we see that the proposed information flow optimization scheme performs consistently better than the max-flow solution.

Table 2: KL-divergence between the distributions of the quantized observations under H_1 and H_0.

Setting#	1	2	3
Max-Flow	49.0015	109.6984	135.6004
Proposed	50.9881	160.1884	198.6157

---

[1]After obtaining real-valued solutions to Problem (2), we simply floor them to get integral value.

[2]Sometimes the benefits of the proposed method are not such significant. It depends on the network and matrix A.

6. REFERENCES

[1] H. Viswanathan and T. Berger, "The quadratic gaussian ceo problem," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.

[2] Y. Oohama, "The rate-distortion function for the quadratic gaussian ceo problem," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[3] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic gaussian ceo problem," in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, June 2004, pp. 119–.

[4] Ersen Ekrem and Sennur Ulukus, "An outer bound for the vector gaussian ceo problem," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6870–6887, 2014.

[5] K. Liu, H. El Gamal, and A. Sayeed, "Decentralized inference over multiple-access channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3445–3455, July 2007.

[6] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized mmse estimation," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5485–5496, Nov 2007.

[7] S Amari et al., "Statistical inference under multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, 1998.

[8] A. Sani and A. Vosoughi, "Distributed vector estimation for power- and bandwidth-constrained wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3879–3894, Aug 2016.

[9] Wee Peng Tay and John N. Tsitsiklis, *Error Exponents for Decentralized Detection in Tree Networks*, pp. 73–92, Springer US, Boston, MA, 2008.

[10] Te Sun Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, Oct 1998.

[11] Daniel Pérez Palomar and Mung Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.

[12] C. Zhang, J. Kurose, Y. Liu, D. Towsley, and M. Zink, "A distributed algorithm for joint sensing and routing in wireless networks with non-steerable directional antennas," in *Proceedings of the 2006 IEEE International Conference on Network Protocols*, Nov 2006, pp. 218–227.

[13] Yuan Xue, Yi Cui, and K. Nahrstedt, "A utility-based distributed maximum lifetime routing algorithm for wireless networks," in *Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE'05)*, Aug 2005.

[14] R. Gallager, "A minimum delay routing algorithm using distributed computation," *IEEE Transactions on Communications*, vol. 25, no. 1, pp. 73–85, January 1977.

[15] Magdalena Turowska, "Maximization of utility in computer network with application of game theory," in *PWT'12*, 2012.

[16] Yin Tat Lee, Satish Rao, and Nikhil Srivastava, "A new approach to computing maximum flows using electrical flows," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 755–764.

[17] Bernard Widrow and István Kollár, *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*, Cambridge University Press, 2008.

[18] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 550–558, April 1993.

[19] J. Liu, V. Veeravalli, G. Ciocarlie, and J. George, "Robust estimation in a sensor network in the presence of adversaries," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Under Review.