

DETECTION IN DECENTRALIZED SENSOR NETWORKS

Saeed A. Aldosari, José M. F. Moura

Department of Electrical and Computer Engineering, Carnegie Mellon University
Pittsburgh, U.S.A

{saldosar, moura}@ece.cmu.edu

ABSTRACT

Advances in integrated technologies are making networks of many inexpensive deployable autonomous sensors a reality. Individually, each sensor may not accomplish much, but working cooperatively they have for example the potential to monitor large areas, detect the presence or absence of targets, or track moving objects. These sensors operate under constraints imposed by scarce power and other limited resources like bandwidth or computing capacity. The paper considers detection in such a distributed sensors environment. We investigate the impact on detection performance, as measured by the probability of error, of such parameters as number of sensors, number of quantization levels at each sensor, or signal to noise ratio, under a rate constraint on the common access communications channel. We optimize the local detectors when the number of sensors is large. We show that the performance loss due to quantization decays exponentially fast as the number of bits per sensor increases and that the choice between hard versus soft local detectors depends not only on the noise distribution and the quantization rate, but also on the SNR under which the sensors operate.

1. INTRODUCTION

Sensor networks are becoming a reality due to the increasing levels of integration offered by technology. In scenarios of interest, the sensors are autonomous and resource starved: power, bandwidth, computing are all scarce commodities. Individual sensors provide negligible performance, but as an aggregate they have the potential to achieve reliably the desired application goals. There are many issues of interest in such sensing network scenarios. We focus here on detection performance when the number of sensors is large (asymptotic analysis) and on the tradeoffs among parameters of interest like the number of sensors, how many bits per local decision, and SNR. We address these tradeoffs when the sensors operate under a global rate constraint imposed by their common access communications channel.

Understanding these tradeoffs is important since it provides answers to important questions, like: how many sensors of a particular type should be deployed in certain environments and whether or not it is advisable to use hard versus soft detectors. This paper considers these questions when the number of sensors N is very large. Under this asymptotic regime, we present the structure of the optimal detector, show that the optimal performance of the sensor network with quantized local detectors of b bits approaches exponentially fast with b the performance of infinite-bandwidth sys-

tems (i.e., with unquantized local detectors), and establish, under a channel rate constraint, when should we opt for fewer higher quality sensors (more bits per sensor) rather than more lower quality sensors (fewer bits per sensor). This particular question is also addressed by [1].

We consider here a very specific architecture for the sensor network, a parallel architecture, see Fig. 1, where there is no communication among the local sensors, and the local detectors feed their quantized decisions to a single fusion center. This decentralized detection architecture has been considered by many authors since the pioneering work of Tenney and Sandell [2] and Tsitsiklis [3]. Due to lack of space, we rely on [3] as well as the book by Varshney [4] for an introduction and overview to the area of decentralized detection, although since their publication, many other relevant papers have also appeared in the literature.

2. MODEL AND PROBLEM STATEMENT

We consider here the following distributed detection problem for the sensor network: N sensors gather T measurements $y_{n,t}$ per sensor n , make a local decision $u_{n,t}$ per measurement and send the decisions $u_{n,t}$ to a single fusion center γ_0 through an error-free multiple-access channel (MAC) as shown in Fig. 1. This particular architecture is often referred to as a parallel fusion network [4]. We consider this problem when the MAC has a rate constraint R . The fusion center makes a global decision \tilde{H} about the true state of nature H based on the collection of the local decisions gathered from all sensors. In the version of the problem we consider here explicitly, the fusion center does not sense measurements directly.

We cast the problem as a binary detection problem with hypotheses H_0 and H_1 . In addition, we assume that the observations $\{y_{n,t} : n = 1, 2, \dots, N, t = 1, 2, \dots, T\}$ are, conditioned on H , independent and identically distributed with conditional densities $f_0(y) = f(y|H_0)$ and $f_1(y) = f(y|H_1)$.

Since the fusion center makes the final decision, the output of the fusion rule γ_0 is binary, i.e., either H_0 or H_1 . Local sensors, on the other hand, are not restricted to binary outputs: each sensor classifies each measurement $y_{n,t} \in \mathcal{Y}$ into $L = 2^b$ classes, where b is the number of transmitted bits per sensor per measurement. We can think of each classifier as a mapping from the observation space \mathcal{Y} to the classification space \mathcal{U} , i.e., $\gamma_n : \mathcal{Y} \rightarrow \mathcal{U}$. Similarly, the fusion rule maps NT local decisions $\{u_{n,t} : n = 1, 2, \dots, N, t = 1, 2, \dots, T\}$ into one of two classes, i.e., $\gamma_0 : \mathcal{Y}^{NT} \rightarrow \{0, 1\}$.

There are two main problems that we address in this paper:

1. For a fixed number N of sensors, we investigate the impact of the number of bits b per local classifier on the overall de-

Partially supported by DARPA Integrated Sensing and Processing (ISP) Program under grant ARO DAAD 19-02-1-0180

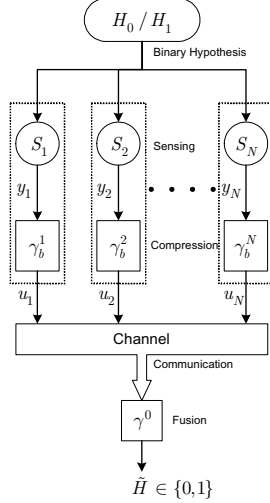


Fig. 1: Parallel fusion network

tection performance represented by the probability of error $P_e = \Pr(\tilde{H} \neq H)$. This, in turn, allows for making decisions about how large the number of bits per sensor should be in order to meet certain performance requirements.

2. For a fixed MAC rate constraint R , we assess the role of SNR in deciding whether to use many low-cost detectors as opposed to using few high-quality sensors.

3. OPTIMIZATION: ASYMPTOTIC ANALYSIS

A major source of difficulty with the general problem in section 2 is that optimization is performed over all possible classification rules $\gamma_n \in \Gamma$ and all possible fusion rules $\gamma_0 \in \Gamma_0$. However, the conditional independence assumption simplifies the problem greatly since, in this case, optimal local classifiers are likelihood ratio tests characterized by a finite number of thresholds. Another simplification can be introduced by assuming that the likelihood ratio $f_1(y)/f_0(y)$ is monotonic in y , in which case, we are allowed to quantize the measurements directly rather than their likelihood ratios. In summary, for a fixed fusion rule γ_0 , the optimization problem amounts to finding the set of optimal thresholds $\lambda = \{\lambda_{n,i}, n = 1, 2, \dots, N, i = 1, 2, \dots, L - 1\}$ given by

$$\lambda^* = \arg \min_{\lambda \in \mathbb{R}^{N(L-1)}} P_e(T, N, b, \lambda, \gamma_0) \quad (1)$$

where λ^* is a vector containing the optimal local thresholds and $P_e(T, N, b, \lambda, \gamma_0)$ is the probability of error associated with the specific number of samples T per sensor, number of sensors N , number of bits per sample b , and local thresholds λ .

It should be noted that the spatio-temporal conditional independence assumption results into the indistinguishability of space and time. Under such assumption, what affects the probability of error $P_e(T, N, b, \lambda, \gamma_0)$ is the total number of samples NT and not the individual values of N and T . In other words, the system described above with N sensors and T samples per sensor would be equivalent to a system with NT sensors each of which is transmitting a single sample. For this reason, without loss of generality,

we will assume that $T = 1$ throughout the rest of the paper and drop the dependence on T .

Up to this point, we have not discussed the structure of optimal fusion rules. Although, this issue is crucial to the overall design process, we adopt an asymptotic approach ($N \rightarrow \infty$), which abstracts the details of the fusion center and, therefore, simplifies the optimization. Design of the fusion center is addressed in [5], where we use a non-asymptotic approach and show that optimal fusion rules have an interesting structure similar to the majority voting rule. Also, in [5], we show how the SNR affects convergence towards the asymptotes.

In what follows, we are primarily concerned with the asymptotic regime ($N \rightarrow \infty$). This regime provides an alternative measure of performance in terms of the error decay rate. Specifically, for any reasonable fusion rule γ_0 and local classification rules, the associated probability of error decays exponentially fast as the number of samples grows to ∞ . Although optimal local detectors are not necessarily identical even when the observations are identically distributed, it has been shown in [3] that the performance loss due to this assumption is negligible especially when the number of sensors is large. We study the problem here assuming identical local detectors. In this case, one appropriate way of representing the performance of the system is through the error decay rate given by

$$C_b(\lambda) = -\lim_{N \rightarrow \infty} \frac{1}{N} \log P_e(N, b, \lambda, \gamma_0), \quad (2)$$

where $\lambda \in \mathbb{R}^{L-1}$ here is a vector containing the $L - 1$ thresholds of the identical local detectors. Thus, we only need to optimize the error decay rate $C_b(\lambda)$, which is much easier to compute than the exact probability of error $P_e(T, N, b, \lambda, \gamma_0)$. Chernoff's theorem states that the best achievable exponent in the probability of error at the fusion center is given by

$$C_b(\lambda) = \max_{s \in [0,1]} C_b(\lambda, s), \quad (3)$$

$$C_b(\lambda, s) = -\log \sum_{u \in \mathcal{U}} [P_0(u)]^s [P_1(u)]^{1-s}, \quad (4)$$

where u represents the decisions produced by the local classifiers, $\mathcal{U} = \{0, 1, \dots, L - 1\}$ is the classification set, and $P_i(u), i \in \{0, 1\}$ are the conditional probabilities of deciding u given that H_i is true. Since u results from quantizing the measurement y , its probability is related to the pdf of y through the quantization thresholds λ as follows

$$P_i(u) = P(u|H_i) = \int_{\lambda_u}^{\lambda_{u+1}} f_i(y) dy = F_i(\lambda_{u+1}) - F_i(\lambda_u), \quad (5)$$

where λ_0 and λ_L are defined to be $-\infty$ and ∞ , respectively, and $F_i(x)$ is the conditional cumulative density function (ccdf) of y defined as

$$F_i(x) = \int_{-\infty}^x f_i(y) dy. \quad (6)$$

As long as the thresholds λ are designed optimally, the error decay rate $C_b(\lambda)$ should increase as we increase the number of bits per sensor b . In the extreme case, when b goes to infinity, the raw (unquantized) measurements are sent to the fusion center and, hence, the system becomes a centralized detection network. In this case, the error decay rate C_∞ is independent of the thresholds λ

and, from Chernoff theorem, it is given by

$$C_\infty = - \min_{s \in [0,1]} \log \int_{\mathcal{Y}} [f_0(y)]^s [f_1(y)]^{1-s} dy, \quad (7)$$

Usually, the minimization step in the last equation does not cause much difficulty since it is a one dimensional optimization problem. In contrast, a decentralized detection system that quantizes its local measurements to b bits per sample would have to be optimized with respect to the error exponent $C_b(\boldsymbol{\lambda})$. Since we have to optimize $L - 1$ thresholds, evaluating the best error exponent involves an L -dimensional optimization and, therefore it is much harder to do. Since $C_b(\boldsymbol{\lambda})$ is always less than C_∞ due to information loss caused by the quantization process, we will be most interested on the ratio $\zeta_b(\boldsymbol{\lambda}) = C_b(\boldsymbol{\lambda})/C_\infty < 1$. Also, it plays an important role in the comparison of systems with different quantizers. For example, when the rate of the MAC channel is limited to R bits per second, it was shown in [1] that using R binary sensors is guaranteed to be optimal as long as $\zeta_2(\boldsymbol{\lambda}) \geq \frac{1}{2}$. We optimize the objective function $V_b(\boldsymbol{\lambda}) = \log(1 - \zeta_b(\boldsymbol{\lambda}))$. To formulate this optimization problem we show the explicit dependence of V_b on s and rewrite it as $V_b(\boldsymbol{\lambda}, s) = \log(1 - \zeta_b(\boldsymbol{\lambda}, s))$, where $\zeta_b(\boldsymbol{\lambda}, s) = C_b(\boldsymbol{\lambda}, s)/C_\infty$. The constraints can be simplified by optimizing over the inter-threshold distances rather than the thresholds themselves. For this purpose, we define the vector of inter-threshold distances $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_{L-1})$, where $\delta_1 = \lambda_1$ and $\delta_k = \lambda_k - \lambda_{k-1}, k = 2, 3, \dots, L-1$, and write the optimization problem as follows

$$\begin{aligned} & \min_{\boldsymbol{\delta}, s} V_b(\boldsymbol{\lambda}, s) \text{ subject to} \\ & \delta_k > 0, k = 2, 3, \dots, L-1 \text{ and } 0 \leq s \leq 1. \end{aligned} \quad (8)$$

4. ALGORITHM

The problem presented in (8) is an L -dimensional constrained non-linear optimization problem. This makes it extremely difficult, if not impossible, to find analytical solutions, especially when the number of bits per sample b is large. Therefore, we propose a gradient-based numerical approach to solve this optimization problem. Due to the large number of constraints, instead of moving in the direction of the L -dimensional gradient, we propose an algorithm similar to the *cyclic coordinate descent* explained in [6]. In this algorithm, each optimization step involves moving along the direction of the one-dimensional gradient with respect to one of the variables as long as the constraints are satisfied. The optimization is then carried out cyclically over all variables.

Each step in the algorithm involves the computation of the gradient $\Delta_i = \frac{\partial V_b(\boldsymbol{\theta})}{\partial \theta_i}$ with respect to a variable θ_i in the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_L)$, then moving in the direction of that gradient according to $\theta_i \leftarrow \theta_i - \alpha_i \Delta_i$ as long as the constraints are not violated. In our work, the arrangement of the variables is such that $\boldsymbol{\theta} = (\delta_1, \delta_2, \dots, \delta_{L-1}, s)$ then the algorithm loops over all variables sequentially. The algorithm keeps searching until all gradients $\Delta_i, i = 1, 2, \dots, L$ approach zero or if a maximum number of iterations has been exceeded.

Proper initialization plays another crucial role in the convergence of the algorithm. In the following section, we will provide a guideline on how to initialize the thresholds. In addition, our results show that a good initial value for s is 0.5. In fact, in all of the cases studied in the following section, the algorithm converged to

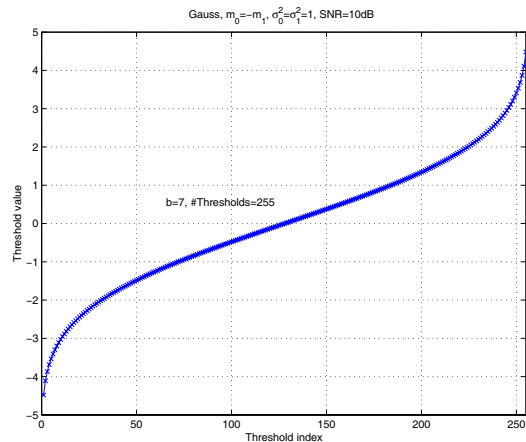


Fig. 2: Optimized thresholds for $b = 8$ (255 thresholds)

0.5. This is interesting since it suggests that it is reasonable to replace the Chernoff rate in our optimization with the simpler Bhattacharyya coefficient with little or no loss in performance. Also, it is interesting to observe that the Chernoff information, which is the exponential rate of the average probability of error, is equivalent to the Renyi α -entropy when $\alpha = 0.5$.

5. RESULTS AND APPLICATIONS

We consider a parallel fusion network with N sensors gathering T measurements each and quantizing them using N identical quantizers to b bits per measurement as explained in section 2 and 3. Local observations are conditionally independent and identically distributed and, in addition, are assumed to follow the additive noise model $y = m_i + n$, where m_i is the signal mean under $H_i, i = 0, 1$ and n is a zero-mean noise with known distribution and variance σ^2 . Our main objective here is to see how the error decay rate behaves as a function of the number of bits per sample b . We use the algorithm developed in the previous section to optimize the thresholds $\boldsymbol{\lambda}$ and the s parameter in order to find the minimum $V_b = V_b(\boldsymbol{\lambda}^*, s^*)$ or, equivalently, the maximum error decay rate $C_b = C_b(\boldsymbol{\lambda}^*, s^*)$. Fig. 2 shows these thresholds produced by the optimization algorithm for the case of $b = 8$ bits per sample when the observations are Gaussian with mean $m_0 = -m_1$, unit variance, and $\text{SNR}=10\text{dB}$. It can be seen that all of the 255 thresholds are concentrated around the intersection of $f_0(y)$ and $f_1(y)$, which is not unexpected since this is the region where it is hardest to distinguish between H_0 and H_1 . This also gives a guideline on how to initialize the thresholds in the optimization algorithm.

For the problem of determining the number of bits per sample, we considered three different distributions. In addition to the Gauss, we considered the Laplace and Logistic distributions because they have heavier tails than the Gauss distribution and we are interested on finding how this impacts our results. Fig. 3 illustrates the interaction between SNR, number of bits per samples b , and V_b , which represents the error between the quantized and unquantized error exponent. There are two curves shown per distribution, one for the high SNR case $\text{SNR}_H = 20 \text{ dB}$ and for the low SNR case $\text{SNR}_L = 0 \text{ dB}$. In all cases, it was assumed that $m_0 = -m_1$ and $\sigma^2 = 1$. It is clear from Fig. 3 that V_b decays linearly with increasing b . This means that the decay rate of the de-

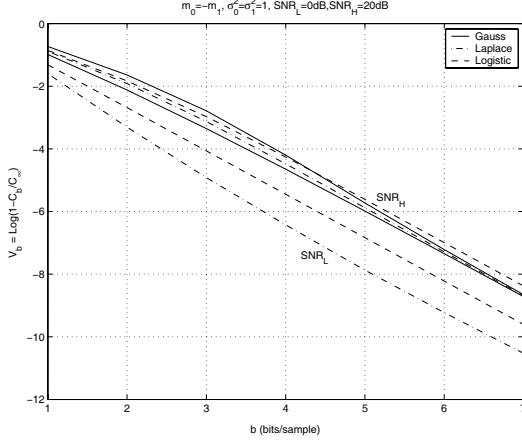


Fig. 3: Logarithm of the error between the quantized and the unquantized error decay rates.

tection error based on quantized measurements approaches that of unquantized measurement exponentially fast as the number of bits per sample b is increased. This, in turn, suggests that it is not advised to go for lower compression rates since the performance gain is expected to be small. In addition, Fig 3 indicates that low SNR values lead to better V_b , which indicates that the ratio of the quantized to the unquantized error decay rate $\zeta_b = C_b/C_\infty$ becomes higher for low SNR conditions. However, it does not suggest that the error rates themselves are improved when the SNR becomes lower since both C_b and C_∞ decrease in this case. In what follows, we demonstrate how the SNR affects the choice of sensor types when the overall communication requirement is fixed.

Our aim here is to investigate the role of SNR in designing decentralized detection systems with fixed communication constraints. We consider the problem of deciding whether to use $2N$ low-cost hard ($b = 1$) sensors versus N high-quality quaternary ($b = 2$) sensors. Due to their quality difference, each sensor type offers a different SNR value denoted by η_1 and η_2 for the binary and the quaternary sensors, respectively, where $\eta_1 < \eta_2$. This problem is similar to the one addressed by the authors of [1] with the exception that they made the assumption that $\eta_1 = \eta_2$. This assumption led to preference of the binary detectors over any other alternative as long as $\zeta_1 = C_1/C_\infty > 1/2$, which is satisfied in many cases. This is not surprising since, after all, if all alternative sensors provide the same SNR then the sensor independence assumption will drive us into choosing the alternative with the largest number of sensors, which corresponds to larger number of observations. What might not be so obvious is which alternative should we choose if each one offers a different SNR.

We show the dependence of the performance criteria on the SNR and rewrite the ζ_b ratio as $\zeta_b(\eta_b) = C_b(\eta_b)/C_\infty(\eta_b)$. It can be shown that the choice of binary over quaternary sensors should be made whenever

$$\frac{C_\infty(\eta_2)}{C_\infty(\eta_1)} < 2 \frac{\zeta_1(\eta_1)}{\zeta_2(\eta_2)}. \quad (9)$$

In the case where equal SNRs is assumed, the condition reduces to $\zeta_1 > 1/2$, which is satisfied for Gaussian and exponential observations [1]. However, in the general case where $\eta_1 < \eta_2$, we have to check the condition in (9) by explicit evaluation of

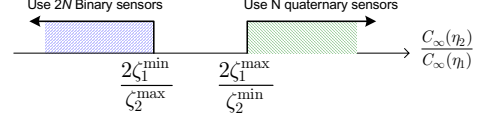


Fig. 4: Decision regions for choosing binary versus quaternary detectors

pdf	b	ζ_b^{\min}	ζ_b^{\max}	β_{binary}	$\beta_{\text{quaternary}}$
Gauss	1	1/2	$2/\pi$	1.14	1.62
	2	0.79	0.88		
Laplace	1	1/2	1	1	2.5
	2	0.8	1		
Logistic	1	1/2	3/4	1.06	2.0
	2	0.73	0.94		

Table 1: Bounds on $\zeta_b(\eta)$

both sides of the inequality. This requires solving two optimization problems to compute $\zeta_1(\eta_1)$ and $\zeta_2(\eta_2)$, in which case we make use of the algorithm presented in the previous section. A set of weaker conditions is obtained by bounding $\zeta_b(\eta)$ such that $\zeta_b^{\min} < \zeta_b(\eta) < \zeta_b^{\max}$, which can be done by finding the limits as $\eta \rightarrow \infty$ and $\eta \rightarrow 0$, respectively. In some cases these limiting values can be found analytically as in the binary case. An approximation to these limits can also be obtained by running the optimization algorithm at very high and very low SNR values. Table 1 shows the bounds on $\zeta_b(\eta)$ for different distributions when $m_0 = -m_1$ and $\sigma^2 = 1$. Using these bounds, the choice of binary versus quaternary sensors can be done according to Fig. 4. The boundary thresholds $\beta_{\text{binary}} = 2\zeta_1^{\min}/\zeta_2^{\max}$ and $\beta_{\text{quaternary}} = 2\zeta_1^{\max}/\zeta_2^{\min}$ for different distributions are in Table 1.

CONCLUSIONS The results in this paper provide a way for designing sensor networks based on not only the sensor compression rate but also the SNR that each sensor type provide. For example, when observations are Gaussian, the analysis provided in this paper enables us to state that it is preferable to use quaternary sensors as long as they can offer 2.1dB more SNR than binary sensors.

6. REFERENCES

- [1] Jean-Francois Chamberland and Venugopal V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Processing*, vol. 51, pp. 407–416, Feb. 2003.
- [2] R. R. Tenney and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 98–101, July 1981.
- [3] John N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr., Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.
- [4] Pramod K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag, New York, 1996.
- [5] Saeed A. Aldosari and José M. F. Moura, "Fusion in sensor networks with communication constraints," Submitted.
- [6] David G. Luenberger, *Introduction to linear and nonlinear programming*, Addison-Wesley, Reading, MA, 1973.