The CEO Problem

Anthony Ekpenyong

February 18, 2003

1 Introduction

It may be advantageous in many situations to monitor or observe the state of a system in a decentralized fashion. Sensor networks are used in a variety of applications where a centralized detector/estimator would not be feasible, is too expensive or cannot achieve the space-time resolution or probability of detection required. This is an active area of research in the computer science, control and detection fields. In decentralized detection, sensor nodes forward their decisions to a decicion fusion center which then computes a global decision on what the state of the system is.

The CEO problem though similar to decentralized detection views the problem from an information theoretic perspective. It is a relatively new topic that can be viewed as a multiterminal source coding problem. A major difference from decentralized detection is that instead of making instantaneous decision about the state, it can make a sequence of decisions based on a set of successive observations. This should be familiar to information theorists: we may achieve less distortion by encoding a block of observations instead of single observations for the same rate [1]. We may now state the problem:

The CEO of a firm (or the Central Estimating Officer in a sensor network) is interested in the output from a source $\{X(t)\}_{t=1}^{\infty}$ which it cannot observe directly. He therefore employs a group of L agents who observe independently corrupted versions of $\{X(t)\}_{t=1}^{\infty}$ and transmit their observations to the CEO under a finite sum rate constraint, R. The main restriction here is that the agents are not allowed to communicate with each other before informing the CEO of their decision. If such a data convention was allowed they would be able to smooth out their observation noises in the limit as $L \to \infty$ and send a joint message to the CEO which can then reconstruct the source with a fidelity D(R) where $D(\cdot)$ is the distortion rate function of the source $X(t)$. This is not just a mathematical abstraction but is rooted in reality. In a battleground scenario for example, there may be L nodes tracking the movement of an enemy tank and sending this information to command headquarters. An exchange of data would increase the time required to reach a decision which could be critical for mission success. Also data exchange means a great deal more communications which increases the susceptibility to an intercept or discovery. In a different setting, the nodes could be geographically dispersed such that a data convention is impractical.

Berger et al [2] introduced the classical formulation of the problem for a discrete memoryless source X. They determined the asymptotic behavior of the minimal error frequency as both L and R $\to \infty$. As a special case Viswanathan and Berger [3] considered the continuous case where the source produces a sequence of independently and identically distributed (i.i.d) Gaussian random variables. The aim then is to study the asymptotic performance of distortion as L and then R $\to \infty$. We shall follow the problem development in [3].

2 Problem Statement

The source produces an i.i.d Gaussian sequence $\{X(t)\}_{t=1}^{\infty}$ with variance σ_x^2. Each agent observes

$$Y_i(t) = X(t) + N_i(t) \tag{1}$$

where $N_i(t) \sim \mathcal{N}(0, \sigma_N^2)$ and are i.i.d $\forall\, i = 1, \cdots, L$ and $t = 1, \cdots, n, \cdots$

The observations $Y_i(t)$ are conditionally independent given the source[1] and the joint density may be written as

$$p(x, y_1, \cdots, y_L) = p(x) \prod_{i=1}^{L} W(y_i|x)$$

The ith agent separately encodes a length n sequence Y_i^n from her observations and transmits

[1]This is a common assumption also used in distributed detection literature to derive optimal decision rules at the sensors and the fusion center (see [4])

to the CEO at a rate

$$R_i = \frac{1}{n} \log |\mathcal{C}_i^n|$$

where $\mathcal{C}_i^n$ is its codebook. The CEO uses a decoding function

$$\Phi_L : \mathcal{C}_1^n \times \cdots \times \mathcal{C}_L^n \to \mathcal{X}^n$$

producing an estimate $\hat{X}^n$ in order to minimize the distortion defined as

$$D(X^n, \hat{X}^n) = \frac{1}{n} E \sum_{t=1}^{n} \left(X(t) - \hat{X}(t) \right)^2 \tag{2}$$

as shown in Figure 1.

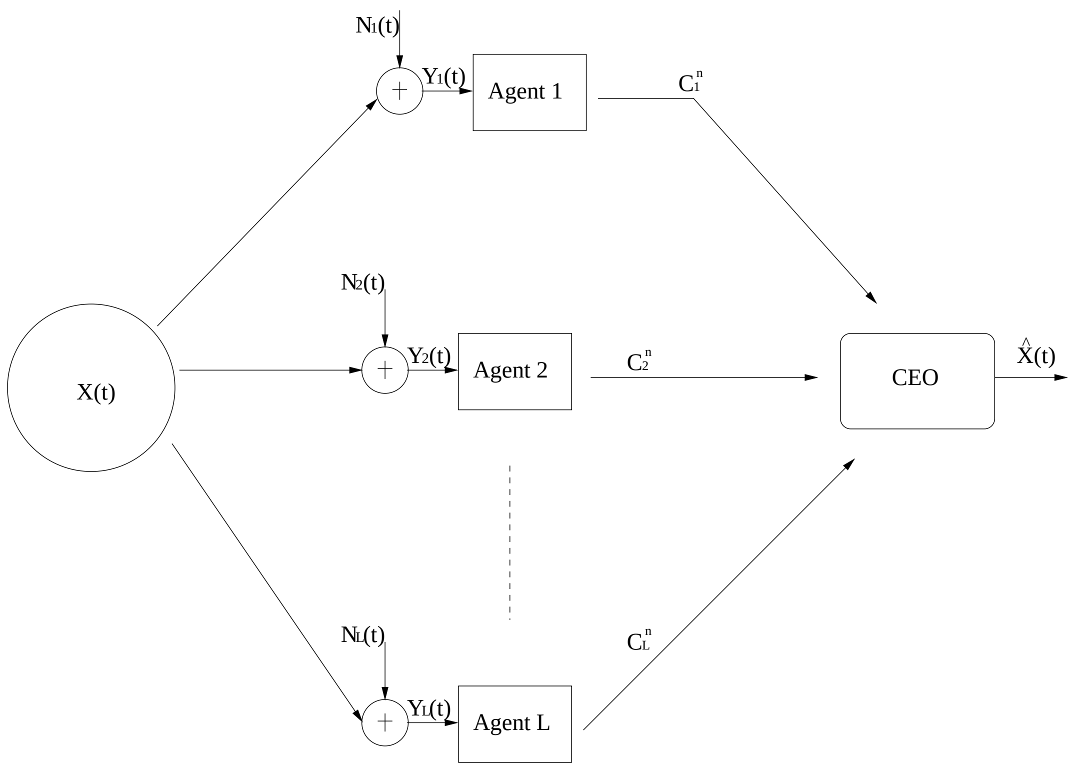

Figure 1: CEO network configuration

The question to be answered then is: what is the mimimal achievable distortion performance (as n $\to \infty$, then L $\to \infty$) given the sum rate constraint $\sum_i R_i \leq R$. Given the codes used by the agents $\mathcal{C}_i^n, i = 1, \cdots, L$, define

$$
\begin{aligned}
D^n(\mathcal{C}_1^n \times \cdots \times \mathcal{C}_L^n) &= \min_{\Phi_L} D^n(X^n, \Phi_L(C_1^n \times \cdots \times C_L^n)) && (3)\\
D^n(L,R) &= \min_{\{\mathcal{C}_i^n\}:\sum_i R_i \le R} D^n(\mathcal{C}_1^n \times \cdots \times \mathcal{C}_L^n) && (4)\\
D(L,R) &= \lim_{n\to\infty} D^n(L,R) && (5)\\
D(R) &= \lim_{L\to\infty} D(L,R) && (6)
\end{aligned}
$$

It may be noticed that due to the assumptions on the distributions we may describe the source by a generic random variable X and the observations as Y since the observations are i.i.d conditioned on X. The main result by Viswanathan and Berger is:

Theorem 1. *Let $Q(u|y)$ be any conditional density on an arbitrary alphabet U and let*

$$\tilde{Q}(u|x) = \int_y W(y|x)Q(u|y)dy$$

then assuming the Cramer-Rao regularity conditions hold

$$\beta(\sigma_x^2, \sigma_N^2) \triangleq \lim_{R\to\infty} R\frac{D(R)}{\sigma_x^2} \ge \inf_{Q(u|y)} \frac{I(Y;U|X)}{\sigma_x^2 E\left[-\frac{\partial^2}{\partial X^2} \log \tilde{Q}(U|X)\right]} > 0 \tag{7}$$

and

$$\beta(\sigma_x^2, \sigma_n^2) \le \frac{\sigma_N^2}{2\sigma_x^2} \tag{8}$$

The authors conjecture that the bounds are actually tight and show this for the case where $Q(u|y)$ is restricted to be Gaussian. Another auxiliary random variable V is defined such that

$$\inf_{U:U=\gamma Y+V, V\sim\mathcal{N}(0,\sigma_v^2)} \frac{I(Y;U|X)}{\sigma_x^2 E\left[-\frac{\partial^2}{\partial X^2} \log \tilde{Q}(U|X)\right]} = \frac{\sigma_N^2}{2\sigma_x^2}$$

where V is independent of (X,Y) and γ is a gain factor [2]. The significance of $\beta(\sigma_x^2, \sigma_n^2) > 0$ as $R \to \infty$ implies that the distortion decays at best as k/R where k is a constant. This is markedly different from the D(R) of a single Gaussian source which may be achieved if the agents are allowed to combine their data. In that case recall that [1]

$$D(R) = \sigma_x^2 2^{-2R}$$

[2] Oohama [5] has shown that the bounds are actually tight.

which has an exponential decay with respect to R and thus the distortion can be made arbitrarily small if $R > H(X)$. The inverse decay for the CEO case as against an exponential decay is the penalty paid by disallowing communications between agents. In the classical CEO case [2] the minimum achievable expected error frequency is shown to decay exponentially with increasing R for most cases. The difference is as a result of the different approaches taken to solve for the rate distortion. In the classical case hypothesis testing is used while in the Gaussian (continuous) case a similar approach would necessitate parameter estimation. It is important to point out that both approaches exploit the relationship between statistics and information theory. Notice in (7) that we are minimizing the ratio of a mutual information to the Fisher information.

3 Outline of Coding Scheme

A three-stage process is used to prove the achievability of the theorem given above. It is divided into coding, decoding and estimation steps. First the generic Gaussian random variables X, Y and the auxiliary U, are discretized to obtain finite alphabets. Let $\tilde{X}, \tilde{Y}, \tilde{U}$ be the quantized random variables, for positive $\delta_0, \delta_1, \delta_2$, there must exist quantization schemes such that[3]

$$\begin{aligned} E(U-\tilde{U})^2 &\leq \frac{\delta_0}{L^2} \\ |I(Y;U) - I(\tilde{Y};\tilde{U})| &\leq \delta_1 \\ |I(X;U) - I(\tilde{X};\tilde{U})| &\leq \delta_2 \end{aligned}$$

In the encoder, a sequence of n observations is mapped to a codeword from a codebook which is identical for all agents.

$$C_i = \tilde{U}_i^n = f^n(\tilde{Y}^n) \tag{9}$$

where C_i is the index of the codeword $\tilde{U}_i^n$. This means that the L codewords $(C_1, C_2, \cdots, C_L)$ from all the agents may be correlated since they all depend on the source sequence X^n. At this point what we have is a situation where we want to encode L correlated sources and send to the decoder.

[3]The condition for the difference in mutual information may be recalled from p.231 of [1]

It may be recalled that the Slepian-Wolf theorem provides the method to do this. A Slepian-Wolf encoder then maps length m blocks of codebook indices on to a smaller index set for each agent as shown in figure 2.

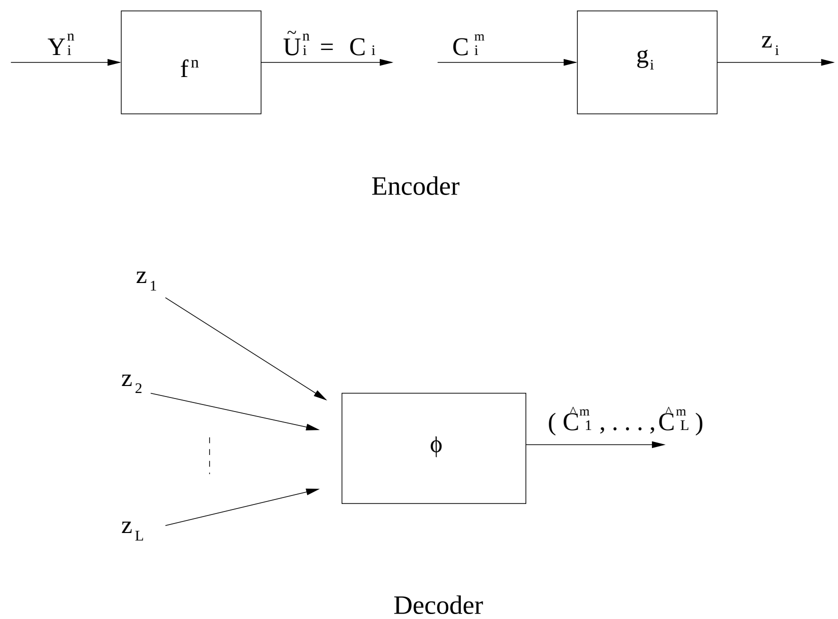

Figure 2: CEO encoder/decoder diagram

$$g_i : \mathcal{U}_i^{mn} \rightarrow \{0, 1, \cdots, N_i - 1\} \qquad for\, 1 \leq i \leq L \tag{10}$$

where N_i and m are chosen sufficiently large so that the index decoding error is small. The complete encoder function is then

$$h_i : f^n \circ g_i \, : \, \mathcal{Y}_i^{mn} \rightarrow \{0, 1, \cdots, N_i - 1\} \tag{11}$$

and the encoder output is $Z_i = h(\tilde{Y}^{nm})$. The total rate is then

$$R = \frac{1}{nm} \sum_{i=1}^{L} \log N_i \tag{12}$$

.

The CEO receives $(Z_1, \cdots, Z_L)$ and decodes to obtain the super codewords $(\hat{C}_1^m, \cdots, \hat{C}_L^m)$for the L agents. It now estimates the source at each time instant from the decoded codeword $(\hat{C}_1^n, \cdots, \hat{C}_L^n)$ obtained from the super codeword $(\hat{C}_1^m(1), \cdots, \hat{C}_1^m(n), \cdots, (\hat{C}_L^m(1), \cdots, \hat{C}_L^m(n))$.

4 Questions and Directions

Some of the assumptions made in the solution of the CEO problem may not be justifiable in a practical sense. In [2, 3], the authors feel that the characterization of the rate distortion region for a finite number of agents would be extremely difficult. Draper [6] has proposed successive coding algorithms for the Gaussian CEO problem for the case of a finite number of agents. It is assumed that the CEO also observes the source. We can think of the CEO as being the (L+1)th agent. Then this problem may be viewed as decoding with side information. A serial CEO problem is also proposed in [6] where the agents are connected in tandem with the last agent being the CEO. All the agents including the CEO observe X and thus each agent (from the second to the CEO) has as input its observation and the encoded form of the previous agent's observation. If all agents know the joint distribution of the source and the observations, then intuitively a sensor can only improve his encoding scheme given the encoded sequence from the previous agent.

The assumption of uncorrelated observations at the agents may not hold. However removing this assumption may significantly increase the difficulty of the problem as is the case in distributed detection [4]. In addition it may not necessarily be true that all agents observe the same source. The power received by an agent is functionally dependent on its distance from the source and hence if our agents are power detectors the measurements would vary.

As this is a relatively new area, not much work has been done on practical applications. The area of sensor networks seems to be a ready-made application for any work done.

References

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

[2] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO Problem," *IEEE Trans. Information Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[3] H. Viswanathan and T. Berger, "The Quadratic Gaussian CEO Problem," *IEEE Trans. Information Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[4] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I - Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, January 1997.

[5] Y. Oohama, "The rate distortion function for the Quadratic Gaussian CEO Problem," *IEEE Trans. Information Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[6] S. C. Draper, "Successive structuring of source coding algorithms for data fusion, buffering and distribution in networks," Ph.D. dissertation, M.I.T, Cambridge, MA, June 2002.