A Characterization of the Error Exponent for the Byzantine CEO Problem

Oliver Kosut and Lang Tong
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email: {oek2,lt35}@cornell.edu

***Abstract*— The discrete CEO Problem is considered when the agents are under Byzantine attack. That is, a malicious intruder has captured an unknown subset of the agents and reprogrammed them to increase the probability of error. Two traitor models are considered, depending on whether the traitors are able to see honest agents' messages before choosing their own. If they can, bounds are given on the error exponent with respect to the sum-rate as a function of the fraction of agents that are traitors. The number of traitors is assumed to be known to the CEO, but not their identity. If they are not able to see the honest agents' messages, an exact but uncomputable characterization of the error exponent is given. It is shown that for a given sum-rate, the minimum achievable probability of error is within a factor of two of a quantity based on the traitors simulating a false distribution to generate messages they send to the CEO. This false distribution is chosen by the traitors to increase the probability of error as much as possible without revealing their identities to the CEO. Because this quantity is always within a constant factor of the probability of error, it gives the error exponent directly.**

***Index Terms*—Distributed Source Coding. Byzantine Attack. Sensor Fusion. Network Security.**

I. Introduction

Distributed systems are more likely to be vulnerable to physical attack. In particular, a malicious intruder might seize a set of nodes, then reprogram them to cooperatively obstruct the goal of the network, launching a so-called Byzantine attack [1], [2]. A useful application which could come under threat of Byzantine attack is distributed source coding. The simplest form of this is the problem of Slepian-Wolf [3], in which a common decoder attempts to reconstruct all the source values from a number of encoders. The Slepian-Wolf problem under Byzantine attack is studied in [4]. The main drawback to this problem, however, is that we cannot expect a reprogrammed node to transmit any useful information about its measurement. Thus it is unreasonable to expect to recover all the data perfectly, as can be done in the non-Byzantine problem.

However, this is not as catastrophic as it might first appear. For instance, one application is a sensor network, in which a fusion center receives data from a large number of sensors to gain some knowledge about the environment. Because there are so many sensors reporting data, any individual sensor's data is not so important. What the fusion center is really interested in recovering is not sensor measurements themselves, but rather some underlying phenomenon that is correlated with these measurements. Hence, the fact that a Byzantine attack removes the fusion center's access to certain sensors' measurements is not so damaging.

One approach to solving this problem would be to use the techniques of [4] to decode the sensors' measurements, even though some of them might by incorrect, then post-process these measurements using the methods of [5], which studies distributed detection under Byzantine attack but without coding. However, this strategy is not rate optimal, since perfectly reconstructing all the measurements as in [4] is hardly necessary. It is our goal in this paper to combine these two steps into one, thereby reducing the rate.

The problem we wish to solve is the CEO Problem [6], which makes the additional assumption that measurements are conditionally independent given the underlying phenomenon. We also assume that conditional distributions are identical across sensors, an assumption that was partially relaxed in [6], but we have not done so here for simplicity. To be precise, we assume there are L agents, where agent i has access to the sequence $\{Y_i(t)\}_{t=1}^{\infty}$, and the CEO (common decoder or fusion center) is interested in recovering the sequence $\{X(t)\}_{t=1}^{\infty}$. These random variables compose a temporally memoryless source with distribution

$$p(x)\prod_{i=1}^{L} W(y_i|x).$$

We assume that a fraction β of the L agents are reprogrammed. These we call *traitors*, and the rest we call *honest*. The quantity β is assumed to be known prior to design of the code, though the exact identity of the traitors is unknown to the CEO. It is shown in [6] that even without traitors, the probability of error cannot be arbitrarily reduced for any finite total communication rate even when the number of agents and the block length go to infinity, rather the best possible probability of error falls exponentially with increasing sum-rate. As in [6], we are interested in the error exponent associated with this drop in probability of error, but now as a function of β.

In this paper we investigate two different traitor models. In the first, which we call *strong traitors*, the traitors are able to observe the messages that the honest agents send to the CEO, and may use this information to decide what to send themselves. The other model we refer to as *weak*

traitors, in which the traitors cannot observe these messages. In both these models, we assume the traitors have complete access to all the sources, as well as the code, so the main difference between strong and weak traitors is that with weak traitors, the honest agents may use independent randomness to construct their codewords, and this randomness is unknown to the traitors. Hence, even though weak traitors know an agent's measurement and the manner in which it chooses its transmission, they may not know the transmission itself. As we will show, this difference has a profound effect on the resulting error exponent.

The main results of this paper give computable bounds on the error exponent for strong traitors, and an uncomputable but exact characterization of the error exponent for weak traitors. The specification of the model is completed, and the results are stated, in Section II. The upper bound for strong traitors is proved in Section III. Section IV contains the proof of achievability for weak traitors, and Section V the converse. Finally, Section VI gives some concluding thoughts.

II. Model and Results

Given block length n and rates R_i for $i = 1, \dots, L$, the encoding function for agent i is given by

$$f_i : \mathcal{Y}_i^n \to \{1, \dots, 2^{nR_i}\}$$

where in general f_i may be a random function. The decoding function for the CEO is given by

$$\phi : \prod_{i=1}^{L} \{1, \dots, 2^{nR_i}\} \to \mathcal{X}^n.$$

Denote by C_i the codeword from the set $\{1, \dots, 2^{nR_i}\}$ sent by agent i to the CEO. Honest agents choose their transmissions by setting $C_i = f_i(Y_i^n)$. If i is a traitor, then it may select C_i in any manner it chooses, based on the following constraints. The traitors may cooperate, and they have access to all the sources $X^n, Y_1^n, \dots, Y_L^n$, and to f_i and ϕ. This assumption that the traitors have access to much more than those same agents if they were honest is perhaps overly pessimistic, but we err on the side of giving the traitors more power rather than less to ensure robustness. As discussed above, strong traitors may base their choose of transmission on C_i for honest i, while weak traitors may not. Finally, the CEO produces its estimate of X^n by setting $\hat{X}^n = \phi(C_1, \dots, C_L)$.

The probability of error is given by

$$P_e = \frac{1}{n} d_H(X^n, \hat{X}^n) \tag{1}$$

where d_H is the Hamming distance. Observe the the probability of error depends on the actions of the traitors, and indeed the identity of the traitors. Let $P_e(f_1, \dots, f_L, \phi)$ be the probability of error as given in (1) where $f_1, \dots, f_L$ and ϕ are the coding functions, but maximized over all possible sets of βL traitors, and all possible actions of those traitors. Let $P_e(R, L)$ be the minimum of $P_e(f_1, \dots, f_L, \phi)$ over all choices of coding functions with $\sum_{i=1}^{L} R_i \le R$. Also let

$$P_e(R) = \lim_{L \to \infty} P_e(R, L).$$

As is shown in [6], $P_e(R)$ is positive for all values of R, but it falls exponentially fast with increasing R. Hence, our quantity of interest is the error exponent given by

$$E(p, W, \beta) = \lim_{R \to \infty} \frac{-\log P_e(R)}{R}.$$

Observe that E is a function of the distribution p, W and also the fraction of traitors β.

We now state our results. The first gives computable bounds on the error exponent for strong traitors. These bounds meet and match the result of [6] at $\beta = 0$. The second theorem gives uncomputable bounds on the probability of error for weak traitors. As these bounds are a factor of two apart, they give the error exponent exactly.

Theorem 1: In addition to X and Y, we introduce two auxiliary random variables U and J. The variable J is independent of (X, Y) with marginal distribution $P_J(j)$, and $X \to (Y, J) \to U$ is a Markov chain. The conditional distribution of U is given by $Q(u|y, j)$, and we define for convenience

$$\tilde{Q}(u|x, j) = \sum_y W(y|x) Q(u|y, j).$$

We also introduce the vector γ_j for all $j \in \mathcal{J}$. Let

$$F(P_J, Q, \gamma) = \frac{\min_{x_1, x_2} \sum_j \gamma_j D(\tilde{Q}_{\lambda,j} \| \tilde{Q}(u|x_1, j))}{I(Y; U|X, J)} \tag{2}$$

where

$$\tilde{Q}_{\lambda,j} = \frac{\tilde{Q}^{1-\lambda}(u|x_1, j) \tilde{Q}^{\lambda}(u|x_2, j)}{\sum_u \tilde{Q}^{1-\lambda}(u|x_1, j) \tilde{Q}^{\lambda}(u|x_2, j)} \tag{3}$$

and λ is chosen so that

$$\sum_j \gamma_j D(\tilde{Q}_{\lambda,j} \| \tilde{Q}(u|x_1, j)) = \sum_j \gamma_j D(\tilde{Q}_{\lambda,j} \| \tilde{Q}(u|x_2, j)). \tag{4}$$

For strong traitors,

$$\max_{P_J, Q} \min_{\gamma} F(P_J, Q, \gamma) \le E(\beta) \le \min_{\gamma} \max_{P_J, Q} F(P_J, Q, \gamma) \tag{5}$$

where on both sides we impose the constraints that

$$\sum_j \gamma_j \ge 1 - 2\beta \quad \text{and} \quad \gamma_j \le P_J(j) \text{ for all } j \in \mathcal{J}. \tag{6}$$

Theorem 2: Consider a block of k independent copies of X and Y^L denoted X^k and Y_i^k for $i = 1, \dots, L$. We introduce auxiliary random variables U_i for $i = 1, \dots, L$, where U_i is conditionally independent of all other variables given Y_i^k. Denote the conditional distribution $Q(u_i|y_i^k)$. Given sets $H, S \subset \{1, \dots, L\}$ with $|H| = |S| = (1 - \beta)L$ and conditional distributions $q(u_{H^c}|y_H^k)$ and $q(u_{S^c}|y_S^k)$, define the following two distributions:

$$P_1(x^k, u^L) = \sum_{y_H^k} p(x^k) W(y_H^k|x^k) Q(u_H|y_H^k) q(u_{H^c}|y_H^k),$$

$$P_2(x^k, u^L) = \sum_{y_S^k} p(x^k) W(y_S^k|x^k) Q(u_S|y_S^k) q(u_{S^c}|y_H^k).$$

Let

$$\tilde{P}_e(R) = \min_{k,L,Q} \max_{H,S,q} \frac{1}{k} \sum_{t=1}^{k} \sum_{\substack{x^n,\hat{x}^n,u^L: \\ x(t)\neq\hat{x}(t)}} \frac{P_1(x^k,u^L)P_2(\hat{x}^k,u^L)}{P(u^L)}$$

where the following constraints are imposed on Q and q:

$$R \geq \frac{1}{k}\sum_{i=1}^{L} I(Y_i^k; U_i|X^k), \tag{7}$$

$$P_1(u^L) = P_2(u^L). \tag{8}$$

For weak traitors,

$$\tilde{P}_e(R) \geq P_e(R) \geq \frac{1}{2}\tilde{P}_e(R). \tag{9}$$

Therefore

$$E(\beta) = \lim_{R\to\infty} \frac{-\log \tilde{P}_e(R)}{R}.$$

This problem with weak traitors was previously studied in [7], which gave computable but non-matching bounds on the error exponent. The lower bound in (5) was one of those bounds, and while this result was proved for weak traitors in [7], the proof given there does not rely on this, so we do not repeat it here. The upper bound in (5) is proved in Section III. Achievability for Theorem 2 is proved in Section IV. and the converse in Section V.

III. Upper Bound for Strong Traitors

We denote by C_i the codeword transmitted by agent i, and $Q(c_i|y_i^n)$ the distribution used by agent i, if it is honest, to generate C_i from Y_i^n. Of course, C_i may be deterministic given Y_i^n, but we assume in general that it may be randomized. Define a distribution on X^n and C^L as

$$P(x^n, c^L) = \sum_{y^{nL}} p(x^n) \prod_{i=1}^{L} W(y_i^n|x^n)Q(c_i|y_i^n).$$

We will refer to various marginals and conditionals of this distribution as well.

Let $\tilde{X}_t = (X(1),\ldots,X(t-1),X(t+1),\ldots,X(n))$. For any t and $\tilde{x}_t$, define $U_i(t,\tilde{x}_t)$ to be a random variable distributed with $X(t)$ and $Y_i(t)$ such that

$$\Pr(X(t)=x, Y_i(t)=y, U_i(t,\tilde{x}_t)=c) = p(x)W(y|x)\Pr(C_i=c|Y_i(t)=y, \tilde{X}_t=\tilde{x}_t).$$

Note that $X(t) \to Y(t) \to U_i(t,\tilde{x}_t)$ is a Markov chain.

Suppose the traitors perform the following attack. They select a set $S \subset \{1,\ldots,L\}$ with $|S| = (1-\beta)L$ and $|H\cap S| = (1-2\beta)L$, where H is the true set of honest agents. The set S is the traitors' target set, that they endeavor to fool the CEO into thinking may be the true set of honest agents. They generate a sequence X'^n from the distribution $P(x^n|c_{H\cap S})$. Finally, they construct $C_{S\setminus H}$ just as honest agents would if X'^n were the truth. That is, from X'^n, they generate $C_{S\setminus H}$ from the distribution $P(c_{S\setminus H}|x^n)$, and transmit this $C_{S\setminus H}$ to the CEO.

Observe that X^n, X'^n, C^L will be distributed according to

$$P(x^n,c_H)P(x'^n|c_{H\cap S})P(c_{S\setminus H}|x'^n) = \frac{P(x^n,c_H)P(x'^n,c_S)}{P(c_{H\cap S})}.$$

This distribution is symmetric in x^n and x'^n. In particular, if S were the true set of honest agents, and the traitors performed an analogous attack selecting the set H as their target set, then precisely the same distribution among X^n, X'^n, C^L would result, except that X^n and X'^n would switch roles. Hence, if the CEO achieves a probability of error of P_e; that is, if $\hat{X}^n$ is such that $P_e \geq \frac{1}{n}d_H(X^n,\hat{X}^n)$, then it must also be that $P_e \geq \frac{1}{n}d_H(X'^n,\hat{X}^n)$, because the CEO can only generate one estimate, but it must work in both situations. Therefore

$$\begin{aligned} P_e &\geq \frac{1}{2n}[d_H(X^n,\hat{X}^n) + d_H(X'^n,\hat{X}^n)] \\ &\geq \frac{1}{2n}d_H(X^n,X'^n) \qquad (10) \\ &= \frac{1}{2n}\sum_{t=1}^{n}\Pr(X(t)\neq X'(t)) \\ &= \frac{1}{2n}\sum_{t=1}^{n}\sum_{x(t)\neq x'(t),c^L} \frac{P(x(t),c_H)P(x'(t),c_S)}{P(c_{H\cap S})} \\ &= \frac{1}{2n}\sum_{t=1}^{n}\underbrace{\sum_{x(t)\neq x'(t),c_{H\cap S}} \frac{P(x(t),c_{H\cap S})P(x'(t),c_{H\cap S})}{P(c_{H\cap S})}}_{P_e(t)} \qquad (11) \end{aligned}$$

where we used the triangle inequality in (10). The expression in (11) can be shown to be concave in P. We may write

$$\begin{aligned} &P(x(t),c_{H\cap S}) \qquad (12) \\ &= \sum_{\tilde{x}_t,y_H^n} p(x^n) \prod_{i\in H\cap S} W(y_i^n|x^n)Q(c_i|y_i^n) \\ &= \sum_{\tilde{x}_t} p(x^n) \prod_{i\in H\cap S}\sum_{y} W(y|x(t)) \cdot \Pr(C_i=c_i|\tilde{X}_t=\tilde{x}_t, Y_i(t)=y) \\ &= \mathbb{E}_{\tilde{X}_t} p(x(t)) \prod_{i\in H\cap S}\sum_{y} W(y|x(t)) \cdot \Pr(U_i(t,\tilde{X}_t)=c_i|Y_i(t)=y) \\ &= \mathbb{E}_{\tilde{X}_t} p(x(t)) \prod_{i\in H\cap S} \Pr(U_i(t,\tilde{X}_t)=c_i|X(t)=x(t)). \qquad (13) \end{aligned}$$

Define for convenience

$$P(x,u_{H\cap S}|t,\tilde{X}_t) = p(x)\prod_{i\in H\cap S}\Pr(U_i(t,\tilde{X}_t)=u_i|X(t)=x). \tag{14}$$

Substituting (13) and (14) into (11) and using concavity gives

$$P_e(t) \geq \mathbb{E}_{\tilde{X}_t} \sum_{\substack{x_1 \neq x_2 \\ u_{H\cap S}}} \frac{P(x_1, u_{H\cap S}|t, \tilde{X}_t) P(x_2, u_{H\cap S}|t, \tilde{X}_t)}{\sum_{x_3} P(x_3, u_{H\cap S}|t, \tilde{X}_t)}$$
$$\geq |X|^{-1} \mathbb{E}_{\tilde{X}_t} \max_{x_1 \neq x_2} \sum_{u_{H\cap S}} \frac{P(x_1, u_{H\cap S}|t, \tilde{X}_t) P(x_2, u_{H\cap S}|t, \tilde{X}_t)}{\max_{x_3} P(x_3, u_{H\cap S}|t, \tilde{X}_t)}$$

Let

$$\mathcal{U}_x = \left\{ u_{H\cap S} : x = \operatorname*{argmax}_{x'} p(x') \prod_{i \in H\cap S} \tilde{Q}(u_i(t, \tilde{X}_t)|x') \right\}.$$

Then

$$P_e(t) \geq |X|^{-1} \mathbb{E}_{\tilde{X}_t} \max_{x_1 \neq x_2} \sum_{x_3} \sum_{u_{H\cap S} \in \mathcal{U}_{x_3}} \frac{P(x_1, u_{H\cap S}|t, \tilde{X}_t) P(x_2, u_{H\cap S}|t, \tilde{X}_t)}{P(x_3, u_{H\cap S}|t, \tilde{X}_t)}$$
$$\geq |X|^{-1} \mathbb{E}_{\tilde{X}_t} \max_{x_1 \neq x_2, x_3} \sum_{u_{H\cap S} \in \mathcal{U}_{x_3}} \frac{P(x_1, u_{H\cap S}|t, \tilde{X}_t) P(x_2, u_{H\cap S}|t, \tilde{X}_t)}{P(x_3, u_{H\cap S}|t, \tilde{X}_t)}. \quad (15)$$

For fixed x_3, if both x_1 and x_2 are different from x_3, we can always increase the value in (15) by making x_1 or x_2 equal to x_3. Hence, we need only consider cases in which either $x_1 = x_3$ or $x_2 = x_3$. Thus

$$P_e(t) \geq |X|^{-1} \mathbb{E}_{\tilde{X}_t} \max_{x_1 \neq x_2} \sum_{u_{H\cap S} \in \mathcal{U}_{x_2}} P(x_1, u_{H\cap S}|t, \tilde{X}_t)$$
$$= |X|^{-1} \mathbb{E}_{\tilde{X}_t} \max_{x_1 \neq x_2} p(x_1) \Pr(\mathcal{U}_{x_2}|x_1, \tilde{X}_t).$$

Using ideas from [6], we have that

$$\Pr(\mathcal{U}_{x_2}|x_1, \tilde{X}_t) \geq 2^{-\sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{X}_t)|x_1)) - o(L)}$$

where

$$Q_\lambda^{(i)}(u) = \frac{\Pr^{1-\lambda}(U_i(t, \tilde{X}_t) = u|x_1) \Pr^{\lambda}(U_i(t, \tilde{X}_t) = u|x_2)}{\Delta_\lambda^{(i)}} \quad (16)$$

with $\Delta_\lambda^{(i)}$ a normalizing constant and λ chosen such that

$$\sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{X}_t)|x_1)) = \sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{X}_t)|x_2)). \quad (17)$$

Hence

$$P_e(t) \geq \mathbb{E}_{\tilde{X}_t} 2^{-\min_{x_1,x_2} \sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{X}_t)|x_1)) - o(L)}. \quad (18)$$

Putting (18) back into (11) gives

$$-\log P_e$$
$$\leq -\log \frac{1}{2n} \sum_{t=1}^{n} \mathbb{E}_{\tilde{X}_t} \cdot 2^{-\min_{x_1,x_2} \sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{X}_t)|x_1)) - o(L)}$$
$$\leq \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\tilde{X}_t} \min_{x_1,x_2} \sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{X}_t)|x_1)) + o(L) \quad (19)$$

where we have used Jensen's inequality in (19).

A chain of standard inequalities (see [6]) yields

$$R = \sum_{i=1}^{L} R_i \geq \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\tilde{X}_t} \sum_{i=1}^{L} I(Y_i(t); U_i(t, \tilde{X}_t|X(t)). \quad (20)$$

Putting (19) together with (20) and using the fact that

$$\frac{\sum_i A_i}{\sum_i B_i} \leq \max_i \frac{A_i}{B_i}$$

for any nonnegative A_i and B_i, we get

$$\frac{-\log P_e}{R}$$
$$\leq \max_{t, \tilde{x}_t} \frac{\min_{x_1,x_2} \sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \Pr(U_i(t, \tilde{x}_t)|x_1)) + o(L)}{\sum_{i=1}^{L} I(Y_i(t); U_i(t, \tilde{x}_t)|X(t))}$$
$$\leq \max_{U_i : X \to Y_i \to U_i} \frac{\min_{x_1,x_2} \frac{1}{L} \sum_{i \in H\cap S} D(Q_\lambda^{(i)} \| \tilde{Q}(u_i|x_1))}{\frac{1}{L} \sum_{i=1}^{L} I(Y_i; U_i|X)} + \epsilon. \quad (21)$$

Observing that the choices of H and S could have been made differently by the traitors, we introduce a vector γ_i for $i = 1, \ldots, L$ under the constraints

$$\gamma_i \in \left\{ 0, \frac{1}{L} \right\} \quad \text{and} \quad \sum_i \gamma_i = 1 - 2\beta. \quad (22)$$

This allows us to tighten (21) to

$$\frac{-\log P_e}{R}$$
$$\leq \min_{\gamma_i} \max_{U_i : X \to Y_i \to U_i} \frac{\min_{x_1,x_2} \sum_{i=1}^{L} \gamma_i D(Q_\lambda^{(i)} \| \tilde{Q}(u_i|x_1))}{\frac{1}{L} \sum_{i=1}^{L} I(Y_i; U_i|X)} + \epsilon. \quad (23)$$

we claim that the value of (23) does not change if we replace (22) with

$$\gamma_i \leq \frac{1}{L} \quad \text{and} \quad \sum_i \gamma_i \geq 1 - 2\beta. \quad (24)$$

This is because we may use arbitrarily large L, so any γ_i satisfying (22) can be closely approximated by a γ_i satisfying (24). Furthermore, we introduce a variable I with values in $\{1,\dots,L\}$ such that

$$\Pr(U=u|I=i,Y=y)=\Pr(U_i=u|Y=y)$$

and maintaining the condition $\gamma_i \le P_I(i)$ for all $i=1,\dots,L$. Doing so gives

$$\begin{aligned}\frac{-\log P_e}{R} &\le \min_{\gamma_i}\max_{P_I,Q}\frac{\min\limits_{x_1,x_2}\sum\limits_i \gamma_i D(\tilde{Q}_{\lambda,i}\|\tilde{Q}(u|x_1,i))}{I(Y;U|X,I)}\\ &= \min_{\gamma_i}\max_{P_I,Q} F(P_I,Q,\gamma).\end{aligned}$$

Replacing I with a variable J over an arbitrary alphabet proves the upper bound in (5). Note that in this process (16), (17), and (24) have become (3), (4), and (6) respectively.

IV. Achievability for Weak Traitors

We first prove the upper bound in (9) for $k=1$, and then extend it to higher k. Descriptions of the codebook, and the encoding and decoding rules follow in Section IV-A. An error analysis is conducted in Section IV-B.

A. Coding Method

1) Random Code Structure: Each agent i forms its codebook in the following way. Given $Q(u_i|y_i)$, it generates $2^{n(I(Y_i;U_i)+\delta)}$ n-length codewords from the marginal distribution of U_i. Let $\mathcal{C}_i^{(n)}$ be this codeword set. These codewords are then uniformly at random placed into $2^{n(I(Y_i;U_i|X)+2\delta)}$ bins.

2) Encoding Rule: Upon receiving Y_i^n, agent i selects uniformly at random an element of

$$\mathcal{C}_i^{(n)}\cap T_\epsilon^{(n)}(U_i|Y_i^n).$$

This random selection is performed at run time, not in the codebook generation. Recall that this randomization is unknown to the weak traitors, and is the main way in which honest agents can do better with weak traitors than with strong. Call the selected sequence U_i^n. Agent i then sends to the CEO the index of the bin containing U_i^n. Observe that the sum rate is

$$\sum_{i=1}^{L}[I(Y_i;U_i|X)+2\delta]$$

so (7) is satisfied as $\delta\to 0$.

3) Decoding Rule: For each $S\subset\{1,\dots,L\}$ with $|S|=(1-\beta)L$, the CEO looks for a sequence in $T_\epsilon^{(n)}(U_S)$ that matches the received bins from all agents in S. If there is exactly one such a sequence, call it $\hat{U}_i^n[S]$ for all $i\in S$. Otherwise, define this to be null.

For all i, if there is exactly one non-null value of $\hat{U}_i^n[S]$ for all $S\ni i$, then call this sequence $\hat{U}_i^n$. If all the values of $\hat{U}_i^n[S]$ are null or they are inconsistent, then leave $\hat{U}_i^n$ undefined. Let R be the set of agents with $\hat{U}_i^n$ defined.

The CEO looks for a set S and a distribution $q(u_{R\setminus S}|y_S)$ such that $\hat{U}_R^n$ is typical with respect to the distribution

$$P_2(x,u_R)=\sum_{y_S}p(x)W(y_S|x)Q(u_S|y_S)q(u_{R\setminus S}|y_S). \quad (25)$$

If there are more than one such pair (S,q), choose between them arbitrarily. Finally, form $\hat{X}^n$ by simulating the distribution $P_2(x|u_R)$ with $\hat{U}_R^n$ as the input sequence.

B. Error Analysis

Consider the following error events:

1) Agent i can find no conditionally typical codewords given the sequence Y_i^n. That is, the set
$$\mathcal{C}_i^{(n)}\cap T_\epsilon^{(n)}(U_i|Y_i^n)$$
is empty.
2) The sequence U_H^n is not jointly typical, where H is the true set of honest agents.
3) There is another typical sequence u_H^n in the same bin as U_H^n.
4) For some $S\ne H$ and $i\in H\cap S$, $\hat{U}_i^n[S]\ne U_i^n$.
5) The complete sequence $(X^n,\hat{U}_R^n)$ is not typical with respect to the distribution
$$\sum_{y_H}p(x)W(y_H|x)Q(u_H|y_H)q(u_{R\setminus H}|y_H)$$
for any $q(u_{R\setminus H}|y_H)$.

We will consider each of these error events in turn, starting with event (1). The probability that a particular typical sequence u_i^n is chosen as an agent i codeword is

$$\frac{2^{n(I(Y_i;U_i)+\delta)}}{2^{nH(U_i)}}=2^{-n(H(U_i|Y_i)-\delta)}.$$

Since given Y_i^n, the number of jointly typical sequences U_i^n is about $2^{nH(U_i|Y_i)}$, with high probability there will be at least one conditionally typical codeword (indeed, on average there will be $2^{n\delta}$). That is, event (1) occurs with small probability. By the Markov Lemma, event (2) also occurs with small probability.

It can be shown (for example, in [8]) that event (3) occurs with small probability if for all $A\subset H$,

$$\sum_{i\in A}R_i\ge I(U_A;Y_A|U_{H\setminus A})$$

where $R_i=I(Y_i;U_i|X)+2\delta$. That is, we need to show that

$$2\delta|A|\ge I(Y_A;U_A|U_{H\setminus A})-\sum_{i\in A}I(Y_i;U_i|X). \quad (26)$$

Observe that

$$\begin{aligned}I(Y_A;U_A|U_{H\setminus A})&-\sum_{i\in A}I(Y_i;U_i|X)\\ &= I(Y_A;U_A|U_{H\setminus A})-I(Y_A;U_A|X)\\ &= I(X;U_A|U_{H\setminus A})\\ &\le H(X|U_{H\setminus A}). \quad (27)\end{aligned}$$

If $|A| \leq |H|/2$, then $|H \setminus A| \to \infty$ as $L \to \infty$, so $H(X|U_{H\setminus A}) \to 0$. Hence (26) holds for sufficiently large L. If $|A| \geq |H|/2$, then using (27) again gives

$$\frac{1}{|A|}\left[I(Y_A;U_A|U_{H\setminus A}) - \sum_{i\in A} I(Y_i;U_i|X)\right] \leq \frac{1}{|A|}H(X|U_{H\setminus A}) \leq \frac{1}{|A|}H(X) \leq \frac{2H(X)}{|H|} \leq 2\delta$$

for sufficiently large L, so again (26) holds, meaning event (3) occurs with low probability. Note that if events (1)–(3) do not occur, $\hat{U}_i^n[H] = U_i^n$ for all $i \in H$.

Event (4) occurs only if the bin associated with agents $S \setminus H$ sent by the traitors contains a sequence $u_{S\setminus H}^n$ that is jointly typical with some sequence $u_{S\cap H}'^n$ different from the true $U_{S\cap H}^n$ but in the same $S \cap H$ bin. However, since weak traitors have access only to $Y_{S\cap H}^n$ and not $U_{S\cap H}^n$, in order to cause this event to occur with significant probability, they must choose a $S \setminus H$ bin containing a corresponding $u_{S\setminus H}^n$ for each possible $U_{S\cap H}^n$.

For a given $U_{S\cap H}^n$, we first calculate the probability that a certain $S \setminus H$ bin contains an element jointly typical with an element in the same bin as $U_{S\cap H}^n$. The probability that a given pair of $S \cap H$ and $S \setminus H$ codewords are jointly typical is

$$\frac{2^{nH(U_S)}}{\prod_{i\in S} 2^{nH(U_i)}} = 2^{n(H(U_S)-\sum_{i\in S} H(U_i))}.$$

The average number of codewords in an agent i bin is about

$$2^{n(I(Y_k;U_i)-I(Y_k;U_i|X)-\delta)} = 2^{n(I(X;U_i)-\delta)}$$

so the probability that a $S \setminus H$ bin contains any codeword jointly typical with an element of a given $S \cap H$ bin other than $U_{S\cap H}^n$ is

$$\begin{aligned}
&2^{n(H(U_S)-\sum_{i\in S} H(U_i))} \\
&\qquad \cdot 2^{n\sum_{i\in S\setminus H}(I(X;U_i)-\delta)}(2^{n\sum_{i\in S\cap H}(I(X;U_i)-\delta)} - 1) \\
&\leq 2^{n(H(U_S)-\sum_{i\in S} H(U_i)+\sum_{i\in S}(I(X;U_i)-\delta))} \\
&\leq 2^{n(H(X)+H(U_S|X)+\sum_{i\in S}(-H(U_i|X)-\delta))} \\
&= 2^{n(H(X)-|S|\delta)} \\
&\leq 2^{-n\epsilon}
\end{aligned}$$

for L sufficiently large. The expected size of

$$\mathcal{C}_i^{(n)} \cap T_\epsilon^{(n)}(U_i|Y_i^n)$$

is $2^{n\delta}$, and most of these sequences will be in different bins. Hence, the probability that a certain $S \setminus H$ bin contains sequences jointly typical with a large fraction of those $S \cap H$ bins is at most $(2^{-n\epsilon})^{2^{n\delta}}$. The probability that any of the $S\setminus H$ bins has this property is therefore at most

$$2^{n(\sum_{i\in S\setminus H}(I(Y_i;U_i|X)+2\delta)-\epsilon 2^{n\delta})}$$

which is vanishingly small. Thus, event (4) occurs with small probability. Note that if events (1)–(4) do not occur, $\hat{U}_i^n$ will be defined and equal to U_i^n for all $i \in H$.

To evaluate the probability of event (5), consider some agent $i \in R\setminus H$. It will be enough to show that there exists a function $g_i : \mathcal{Y}_H^n \to \mathcal{U}_i^n$ such that with high probability, $g_i(Y_H^n) = \hat{U}_i^n$. That is, it is not just that the traitors choose a bin based on Y_H^n, in fact they choose the exact value of $\hat{U}_i^n$ that will be recovered by the CEO. If there exist such functions g_i for all $i \in R \setminus H$, then it is not hard to show that $Y_H^n, \hat{U}_{R\setminus H}^n$ are typical with respect to the distribution

$$P(y_H)q(u_{R\setminus H}|y_H)$$

for some q. Since $(X^n, U_H^n) - Y_H^n - \hat{U}_{R\setminus H}^n$ is a Markov chain, by the Markov lemma $(X^n, Y_H^n, U_H^n, \hat{U}_{R\setminus H}^n)$ is typical with respect to

$$p(x)W(y_H|x)Q(u_H|y_H)q(u_{R\setminus H}|y_H)$$

with high probability. Since we have already shown in our analysis of events (1)–(4) that with high probability $\hat{U}_H^n = U_H^n$, we have that event (5) occurs with vanishing probability.

We now prove the existence of the functions g_i. Since $i \in R$, there must be some S such that $i \in S$ and the bins transmitted by the agents in S contain a jointly typical element $\hat{U}_S^n[S]$. Furthermore, all estimates of U_i^n must have been consistent, so $\hat{U}_i^n = \hat{U}_i^n[S]$. We consider two cases. First, suppose the $S \setminus H$ bin selected by the traitors contains an element typical with $Y_{S\cap H}^n$ according to the non-traitor distribution

$$\sum_{x,y_{S\setminus H}} p(x)W(y_S|x)Q(u_{S\setminus H}|y_{S\setminus H}).$$

In this case, let $g_i(Y_H^n)$ be this typical element. The Markov lemma implies that with high probability $(U_{S\setminus H}^n, g_i(Y_H^n)) \in T_\epsilon^{(n)}(U_S)$. Since we have assumed that $\hat{U}_S^n[S]$ exists, this sequence must be the unique jointly typical sequence in the transmitted bins, meaning $g_i(Y_H^n) = \hat{U}_i^n$.

Now consider the case that the $S\setminus H$ bin contains no element typical with $Y_{S\cap H}^n$. We will show that if so, it is highly unlikely that any element of the bin could be jointly typical with $U_{S\cap H}^n$. Given jointly typical $y_{S\cap H}^n$ and $u_{S\cap H}^n$, we first determine the probability that a $S \setminus H$ codeword is jointly typical with $u_{S\cap H}^n$ given that it is not typical with $y_{S\cap H}^n$. If we let U_i^n be selected i.i.d. from $P(u_i)$ for all $i \in S \setminus H$, independently from each other. Then

$$\begin{aligned}
&\Pr(U_{S\setminus H}^n \in T_\epsilon^{(n)}(U_{S\setminus H}|u_{S\cap H}^n)|U_{S\setminus H}^n \notin T_\epsilon^{(n)}(U_{S\setminus H}|y_{S\cap H}^n)) \\
&= \frac{\Pr(U_{S\setminus H}^n \in T_\epsilon^{(n)}(U_{S\setminus H}|u_{S\cap H}^n) \setminus T_\epsilon^{(n)}(U_{S\setminus H}|y_{S\cap H}^n))}{\Pr(U_{S\setminus H}^n \notin T_\epsilon^{(n)}(U_{S\setminus H}|y_{S\cap H}^n))} \\
&\leq \frac{\Pr(U_{S\setminus H}^n \in T_\epsilon^{(n)}(U_{S\setminus H}|u_{S\cap H}^n))}{\Pr(U_{S\setminus H}^n \in \prod_{i\in S\setminus H} T_\epsilon^{(n)}(U_i) \setminus T_\epsilon^{(n)}(U_{S\setminus H}|y_{S\cap H}^n))} \\
&\leq \frac{2^{-n(\sum_{i\in S\setminus H} H(U_i)+\epsilon)}}{2^{-n(\sum_{i\in S\setminus H} H(U_i)-\epsilon)}} \\
&\qquad \cdot \frac{|T_\epsilon^{(n)}(U_{S\setminus H}|u_{S\cap H}^n)|}{|\prod_{i\in S\setminus H} T_\epsilon^{(n)}(U_i) \setminus T_\epsilon^{(n)}(U_{S\setminus H}|y_{S\cap H}^n)|}
\end{aligned}$$

$$\leq \frac{2^{n(H(U_{S\setminus H}|U_{S\cap H})+3\epsilon)}}{2^{n(\sum_{i\in S\setminus H} H(U_i)-\epsilon)} - 2^{n(H(U_{S\setminus H}|Y_{S\cap H})+\epsilon)}}$$
$$\leq 2^{n(H(U_{S\setminus H}|U_{S\cap H})-\sum_{i\in S\setminus H} H(U_i)+5\epsilon)}$$

Hence, the probability that any codeword in a given $S\setminus H$ bin is jointly typical with $u^n_{S\cap H}$ given that they are all not typical with $y^n_{S\cap H}$ is at most

$$2^{n(H(U_{S\setminus H}|U_{S\cap H})-\sum_{i\in S\setminus H} H(U_i)+5\epsilon)} 2^{n(\sum_{i\in S\setminus H}(I(X;U_i)-\delta))}$$
$$= 2^{n(I(X;U_{S\setminus H}|U_{S\cap H})-|S\setminus H|\delta+5\epsilon)} \leq 2^{-n\epsilon}$$

for sufficiently large L. Therefore, it is highly unlikely that any $S\setminus H$ bin without a sequence jointly typical with $y^n_{S\cap H}$ contains a sequence jointly typical with a large fraction of possible values of $U^n_{S\cap H}$.

As we have shown, with high probability, events (1)–(5) do not occur. Hence, there exists a distribution $q(u_{R\setminus H}|y_H)$ such that (X^n, U^n_R) is typical with respect to

$$P_1(x,u_R) = \sum_{y_H} p(x)W(y_H|x)Q(u_H|y_H)q(u_{R\setminus H}|y_H).$$

Recall that the CEO's estimation strategy is to find a set S and distribution $q(u_{R\setminus S}|y_S)$ such that U^n_R is typical with respect to $P_2(x,u_R)$ as defined in (25). This means that U^n_R is strongly typical with respect to both these distributions, so

$$|P_1(u_R) - P_2(u_R)| \leq \frac{2\epsilon}{\prod_{i\in R}|\mathcal{U}_i|} \tag{28}$$

for all u_R. Hence, in the limit as $\epsilon\to 0$, these two marginal distributions are equal (i.e. (8) holds). Furthermore, since the CEO generates $\hat{X}^n$ from $P_2(x|u_R)$, with high probability $(X^n,\hat{X}^n)$ is typical with respect to $P_1(x,u_R)P_2(\hat{x}|u_R)$. Hence with high probability

$$\frac{1}{n}d_H(X^n,\hat{X}^n) \leq \sum_{\substack{x\neq\hat{x}\\ u_R}} P_1(x,u_R)P_2(\hat{x}|u_R) + \frac{\epsilon}{|X|}$$
$$\leq \max_{H,S,q} \sum_{\substack{x\neq\hat{x}\\ u^L}} \frac{P_1(x,u^L)P_2(\hat{x},u^L)}{P_2(u^L)} + \frac{\epsilon}{|X|}. \tag{29}$$

Where we have replaced R with $\{1,\ldots,L\}$ in (29) because it cannot decrease the probability of error. Furthermore, we may assume that $P_1(u_R) = P_2(u_R)$, because by continuity and (28), it does not change the value in (29) for small ϵ. Taking the limit as $\epsilon\to 0$ and noting that the honest agents may choose L and Q however they like, we see that

$$P_e(R) \leq \min_{L,Q}\max_{H,S,q} \sum_{\substack{x\neq\hat{x}\\ u^L}} \frac{P_1(x,u^L)P(\hat{x},u^L)}{P(u^L)}.$$

We have proved achievability of Theorem 2 for $k=1$. To prove it for $k>1$, we need only modify the coding scheme to use distributions of the form $Q(u_i|y_i^k)$ to generate U_i^n sequences. That is, each agent treats each k Y_i values as a single letter, and degrades those letters to U_i as before. It is easy to modify the proof given above to show that $P_e(R)\leq \tilde{P}_e(R)$.

V. Converse for Weak Traitors

Consider any coding scheme used by the honest agents and the CEO that achieves a probability of error of P_e. Let $Q(c_i|y_i^n)$ be the distribution with which agent i would honestly generate its codeword C_i from the measurement Y_i^n. Observe that

$$R = \sum_{i=1}^L \frac{1}{n}\log|C_i| \geq \frac{1}{n}\sum_{i=1}^L H(C_i) \geq \frac{1}{n}\sum_{i=1}^L I(Y_i^n;C_i|X^n).$$

Suppose the traitors perform the following attack. They choose a set S with $|S| = (1-\beta)L$ and a distribution $q(c_{H^c}|y_H^n)$ such that there exists a $q(c_{S^c}|y_S^n)$ for which if we define the distributions

$$P_1(x^n,c^L) = \sum_{y_H^n} p(x^n)W(y_H^n|x^n)Q(c_H|y_H^n)q(c_{H^c}|y_H^k),$$
$$P_2(x^n,c^L) = \sum_{y_S^n} p(x^n)W(y_S^n|x^n)Q(c_S|y_S^n)q(c_{S^c}|y_H^k)$$

then

$$P_1(c^L) = P_2(c^L). \tag{30}$$

From Y_H^n, the traitors then use the distribution $q(c_{H^c}|y_H^n)$ to generate C_{H^c}. Because this attack has a mirror image when S is the true set of honest agents and the traitors use $q(c_{S^c}|y_S^n)$, in order to achieve P_e, the probability of error must be no more than P_e in both cases. Hence, by an argument along the lines of that leading up to (11),

$$P_e \geq \frac{1}{2n}\sum_{t=1}^n \sum_{\substack{x^n,\hat{x}^n,c^L:\\ x(t)\neq\hat{x}(t)}} \frac{P_1(x^n,c^L)P_2(\hat{x}^n,c^L)}{P(c^L)}.$$

Replacing n with k and C with U results in the lower bound in (9).

VI. Conclusion

We looked at the Byzantine CEO Problem for two traitor models. For neither one are our results ideal. It would be desirable to find exact computable characterization of the error exponent for both models, but doing so may be, especially for weak traitors, highly challenging. It does appear, however, that in Byzantine multiterminal source coding, exactly what the traitors are able to observe has a significant impact on the resulting performance.

References

[1] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Transactions on Programming Languages and Systems*, vol. 4, pp. 382–401, July 1982.

[2] D. Dolev, "The Byzantine generals strike again," *Journal of Algorithms*, vol. 3, no. 1, pp. 14–30, 1982.

[3] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.

[4] O. Kosut and L. Tong, "Distributed source coding in the presence of Byzantine sensors," *IEEE Trans. Inform. Theory*, vol. IT-54, pp. 2550–2565, 2008.

[5] S. Marano, V. Matta, and L. Tong, "Distributed inference in the presence of Byzantine sensors," in *Proc. 40th Annual Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct 29–Nov 1 2006.

[6] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 887–902, May. 1996.

[7] O. Kosut and L. Tong, "The CEO problem," in *Proc. Int. Symp. Inf. Theory*, Toronto, Canada, 2008.

[8] P. Viswanath, "Sum rate of a class of Gaussian multiterminal source coding problems," in *Advances in Network Information Theory*, ser. DIMACS in Discrete Mathematics and Theoretical Computer Science, P. Gupta, G. Kramer, and A. J. van Wijngaarden, Eds. AMS, vol. 66, pp. 43–60, 2004.