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 CEO Problem is considered when a subset of the agents are under Byzantine attack; that is, they have been taken over and reprogrammed by a malicious intruder. Inner and outer bounds are given for the error exponent with respect to the sum rate, as a function of the fraction of reprogrammed, or traitor, agents. The inner bound is proved by introducing a coding scheme that takes advantage of the fact that the set of honest (non-traitor) agents will report jointly typical information. The CEO looks for a group with the same size as the set of honest agents that appear to do so. Even if not all the agents in this group are honest, the fact that they all agree keeps the probability of error in check. The outer bound is given in two parts, based on two different possible attacks by the traitors. The first is a black hole attack, in which the traitors simply transmit no information at all. The second is one in which they fabricate false data such that the CEO cannot determine which of two possibilities is the truth.**

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

I. Introduction

Distributed systems are more likely to be vulnerable to physical assault. In particular, a malicious intruder could 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, as can be done in the non-Byzantine problem.

However, this is not as catastrophic as it might first appear. One application, for instance, is a sensor network, in which a fusion center receives data from a large number of sensors to gain some knowledge about the environment. In this case, the multitude and noisiness of the sensors makes the measurements taken by any individual sensor essentially irrelevant. What the fusion center is really interested in recovering is not sensor measurements themselves, but rather some underlying phenomenon that controls the distribution of 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 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. As in [6], we are interested in the error exponent associated with the sum rate given a large number of agents, but now as a function of β.

The main results of this paper give inner and outer bounds on the error exponent. The specification of the model is completed, and the bounds are stated, in Section II. The inner bound is proved in Section III using a coding scheme that takes advantage of the fact that the honest agents are expected to transmit jointly typical codewords to the CEO. The outer bound, proved in Section IV, has two parts, based on two different potential attacks by the traitors. The first results from the traitors performing a black hole attack; that is, they send no information at all to the CEO, unveiling their identities but forcing the CEO to use only the honest sensors' transmissions to estimate X^n. The second part results from the traitors

producing fraudulent transmissions such that the CEO cannot determine which of two possibilities is the truth. Section V gives some concluding thoughts.

II. Model and Results

Any code is associated with a block length n and a set of rates R_i for $i = 1, \cdots, L$. Each agent has a (possibly random) encoding function

$$f_i : \mathcal{Y}^n \to \{1, \cdots, 2^{nR_i}\}.$$

Let $C_i \in \{1, \cdots, 2^{nR_i}\}$ be the codeword sent from agent i to the CEO. Obviously the selection of the transmitted codeword depends on whether agent i is a traitor or not. We assume that the identity of the traitors is fixed before coding begins, but it is unknown to the CEO and the code must accommodate any possible group of traitors and any actions they take. If agent i is honest, then $C_i = f_i(Y_i^n)$. The traitors may cooperatively choose the codewords they transmit in any manner they like based on all the sources $X^n, Y_1^n, \cdots, Y_L^n$. That is, we assume that the traitors have access not only to all the measurements but also the underlying source X. This assumption is perhaps overly pessimistic, but to ensure robust performance we err on the side of giving traitors more power rather than less.

The CEO then has a decoding function

$$g : \prod_{i=1}^{L} \{1, \cdots, 2^{nR_i}\} \to \mathcal{X}^n$$

and it produces its estimate by $\hat{X}^n = g(C_1, \cdots, C_L)$. We define the probability of error as

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

where d_H is the Hamming distance. Of course, the probability of error depends on the actions of the traitors, so let P_e^* be the probability of error as defined in (1), maximized over all possible groups of βL traitors, and all possible actions taken by those traitors. Hence if for a certain code, $P_e^* \leq \epsilon$, we can guarantee that no matter what the traitors do, the probability of error will not exceed ϵ.

We define the minimum probability of error over all codes with any number of agents and sum rate less than R as

$$P_e(R) = \lim_{L \to \infty} \min_{\sum_{i=1}^{L} R_i \leq R} P_e^*.$$

Finally, the quantity of interest is the error exponent with respect to the sum rate R, defined as

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

Note again that the error exponent is a function of the fraction of traitors β.

Following are the main theorems of this paper, giving inner and outer bounds on $E(p, W, \beta)$. Like the result of [6], our results involve two auxiliary random variables U and J, where the distribution of X, Y, U, and J is given by

$$p(x)W(y|x)P_J(j)Q(u|y, J = j)$$

for distributions P_J and Q. We also define for convenience

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

In the interest of space, we have chosen not to exhibit cardinality bounds on these variables in our theorems, but they do exist, ensuring computability.

In addition to these variables, the bounds involve the values γ_j for all $j \in \mathcal{J}$. The constraints put on γ_j vary somewhat among the bounds, and are always listed. The intuition behind γ_j is that it represents how the traitors choose to devote their resources. The variable J can be thought of as splitting the agents into separate groups, with about $P_J(j)L$ in the jth group. As the constraint that $\gamma_j \leq P_J(j)$ for all $j \in \mathcal{J}$ is always in place, γ_j represents a part of the jth group of agents of size $\gamma_j L$.

Theorem 1 (Inner Bound): Given p, W, and β

$$E(p, W, \beta) \geq \max_{P_J, Q} \frac{\min\limits_{\substack{\gamma_j, x_1, x_2: \\ \sum_j \gamma_j \geq 1-2\beta \\ \gamma_j \leq P_J(j)}} \sum\limits_j \gamma_j D(\tilde{Q}_{\lambda,j} \| \tilde{Q}(u|x_1, j))}{I(Y; U|X, J)} \tag{2}$$

where

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

and λ is chosen so that

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

Theorem 2 (Outer Bound): Firstly,

$$E(p, W, \beta) \leq \max_{P_J, Q} \frac{\min\limits_{\substack{\gamma_j, x_1, x_2: \\ \sum_j \gamma_j \geq 1-\beta \\ \gamma_j \leq P_J(j)}} \sum\limits_j \gamma_j D(\tilde{Q}_{\lambda,j} \| \tilde{Q}(u|x_1, j))}{I(Y; U|X, J)} \tag{5}$$

where $\tilde{Q}_{\lambda,J}$ is defined by (3) and (4). Secondly,

$$E(p, W, \beta) \leq \max_{P_J, Q} \frac{\min\limits_{x_1, x_2} (1 - 2\beta) D(W_\lambda \| W(y|x_1))}{I(Y; U|X, J)} \tag{6}$$

where we consider only those P_J and Q such that

$$\min_{x_1, x_2} \sum_j P_J(j) D(\tilde{Q}_{\lambda,j} \| \tilde{Q}(u|x_1, j)) \leq \min_{x_1, x_2} (1 - \beta) D(W_\lambda \| W(y|x_1)) \tag{7}$$

and where

$$W_\lambda(y) = \frac{W(y|x_1)^{1-\lambda} W(y|x_2)^{\lambda}}{\sum\limits_y W(y|x_1)^{1-\lambda} W(y|x_2)^{\lambda}} \tag{8}$$

and λ is such that $D(W_\lambda \| W(y|x_1)) = D(W_\lambda \| W(y|x_2))$.

The inner and outer bounds meet at $\beta = 0$, where they match the result of [6], and at $\beta = 1/2$, where the error exponent is 0. Observe that (2) and (5) only differ in that $1 - 2\beta$ is replaced by $1 - \beta$. The result from [7] giving the capacity of a channel composed of many parallel links, some controlled by an adversary, has a similar two part form to that of our outer bound, in that as the number of adversarial links increases, the capacity drops as would result from a black hole attack. Then the capacity breaks to 0 when half the links are adversarial.

III. Inner Bound

As in [6], we present our inner bound proof in two steps, the second a generalization of the first. In addition, our coding scheme will be identical to that of [6] in many respects, so we endeavor to summarize the method given there when necessary and focus on the differences. First we state Lemma 1, the looser inner bound constituting the first step of our proof. Section III-A contains some notations that we will use in our proof. To prove Lemma 1, Section III-B introduces the proposed coding scheme, then Section III-C evaluates its probability of error. Finally, Section III-D tightens the bound to conclude the proof of Theorem 1.

Lemma 1: Let U be a random variable such that $X \to Y \to U$ is a Markov Chain and the distribution of U is given by $Q(u|y)$. Let

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

Given p, W, and β,

$$E(p, W, \beta) \geq \max_Q \frac{\min_{x_1,x_2} (1-2\beta)D(\tilde{Q}_\lambda \| \tilde{Q}(u|x_1))}{I(Y;U|X)}$$

where

$$\tilde{Q}_\lambda(u) = \frac{\tilde{Q}(u|x_1)^{1-\lambda}\tilde{Q}(u|x_2)^\lambda}{\sum_u \tilde{Q}(u|x_1)^{1-\lambda}\tilde{Q}(u|x_2)^\lambda} \tag{9}$$

and λ is such that $D(\tilde{Q}_\lambda \| \tilde{Q}(u|x_1)) = D(\tilde{Q}_\lambda \| \tilde{Q}(u|x_2))$.

A. Preliminaries

Given a set $S \subset \{1, \cdots, L\}$, y_S denotes the set of y_i with $i \in S$. The same for u_S, Y_S, etc. Observe that when referring to the complete set, we use the superscript, e.g. y^L.

Let $T_\epsilon^n(X)$ be the strongly typical set defined by ϵ of sequences with length n.

Given a sequence x^n, let $t(x^n)$ be the type of x^n. Given a type t, let $\Lambda_t^n(X)$ be the set of sequences X^n with type t. Note that there are two "directions" of sequences that we will discuss: temporal sequences, such as x^n, denoting the values of a single variable across time, and spatial sequences, such as y^L, denoting the values of a group of variables at a single time. We will use types in both these directions, and the meaning should be clear from context.

B. Coding Scheme

No positive error exponent can be achieved if $\beta \geq 1/2$, so we assume that $\beta < 1/2$.

1) Random Code Structure: The codebook construction is exactly the same as in [6]. Given $Q(u|y)$, we randomly construct a codebook $\mathcal{C}_i^{(n)} = \{c_i(1), \cdots, c_i(M)\}$ of $M = \lceil 2^{n(I(Y;U)+\delta)} \rceil$ codewords from the distribution

$$Q^n(u^n) = \sum_{x^n, y^n} p^n(x^n)W^n(y^n|x^n)Q^n(u^n|y^n).$$

These codewords are then randomly placed into $N = \lceil 2^{n(I(Y;U|X)+2\delta)} \rceil$ bins denoted by $B_i = \{b_i(1), \cdots, b_i(N)\}$. This procedure is done independently for each of the L agents.

2) Encoding Rule: Encoding is also done identically to [6]. If agent i observes sequence y_i^n, a codeword $\tilde{u}_i^n$ is randomly selected from the set $\mathcal{C}_i^{(n)} \cap \Lambda^n_{s[t(y_i^n)]}(U|y_i^n)$ according to the uniform distribution, where $s(t)$ is a function that takes a type on Y and gives a conditionally typical joint type on Y and U. The agent then transmits to the CEO the index of the bin b_i that contains $\tilde{u}_i^n$. If the above set is empty, then the agent transmits the symbol 0.

This is of course only performed by honest agents. Traitors can choose their transmission any way they like.

3) Decoding and Estimation Rules: This step constitutes the major difference from [6]. The key observation is that since there are $(1-\beta)L$ honest agents, the CEO should expect to find a set of $(1-\beta)L$ agents whose transmitted bin indices appear to agree (that is, they represent jointly typical codewords). The CEO looks for such a group of $(1-\beta)L$ agents, then estimates X^n based on only the transmissions from these agents. Even if the traitors have engineered their transmissions so that there is more than one such group, any group of $(1-\beta)L$ agents must contain at least $(1-2\beta)L$ honest agents, which is enough to guarantee some accuracy.

The exact estimation procedure is as follows. Let $b_1, \cdots, b_L$ be the bins whose indices were transmitted to the CEO from the agents. For all sets of agents $S \subset \{1, \cdots, L\}$, define

$$B_S = \prod_{i \in S} b_i$$

the Cartesian product of the bins sent by the agents in S. The CEO finds a set S with $|S| = (1-\beta)L$ such that $B_S \cap T_\epsilon^n(U_S)$ is not empty. It sets $\hat{u}_S^n$ to be an element of this set and sets $\hat{x}^n$ to be element randomly selected from the set $T_\epsilon^n(X|\hat{u}_S^n)$ according to the uniform distribution. Both these typical sets are with respect to the distribution

$$p(x)\prod_{i \in S} \tilde{Q}(u_i|x).$$

C. Probability of Error

Let S be the set of size $(1-\beta)L$ chosen by the CEO from which to estimate X^n, and H be the true set of honest sensors. Let $\tilde{X}^n$ be a random sequence constructed by choosing an element from $T_\epsilon^n(X|\hat{U}^n_{S\cap H})$ with respect to the uniform distribution. Of course, the CEO itself does not have access

to this sequence, since it does not know H, but it exists in principle. Note that

$$P_e = \frac{1}{n}\mathbb{E}d_H(X^n,\hat{X}^n) \le \underbrace{\frac{1}{n}\mathbb{E}d_H(X^n,\tilde{X}^n)}_{P_e(1)} + \underbrace{\frac{1}{n}\mathbb{E}d_H(\tilde{X}^n,\hat{X}^n)}_{P_e(2)}.$$

It will be enough to bound each of $P_e(1)$ and $P_e(2)$.

Consider the distribution on $X, \tilde{X}, \hat{X}, U_S$ given by

$$q(\cdot) = p(x)\tilde{Q}(u_{S\cap H}|x)\frac{p(\tilde{x})\tilde{Q}(u_{S\cap H}|\tilde{x})}{\Pr(u_{S\cap H})}\frac{p(\hat{x})\tilde{Q}(u_S|\hat{x})}{\Pr(u_S)}$$

where

$$\Pr(u_S) = \sum_x p(x)\tilde{Q}(u_S|x).$$

We claim that with high probability $(X^n,\tilde{X}^n)$ and $(\tilde{X}^n,\hat{X}^n)$ are each jointly typical with respect to q. Therefore in the limit as $n\to\infty$,

$$\begin{aligned} &P_e(1) = P_e(2) \\ &= \sum_{x_1,x_2:x_1\neq x_2}\sum_{u_{S\cap H}} \frac{p(x_1)\tilde{Q}(u_{S\cap H}|x_1)p(x_2)\tilde{Q}(u_{S\cap H}|x_2)}{\Pr(u_{S\cap H})}. \end{aligned} \tag{10}$$

First we show the joint typicality properties with respect to q. By the same arguments given in [6],

$$\Pr(\hat{U}^n_{S\cap H} \neq \tilde{U}^n_{S\cap H}) \le 2^{-n\sigma} \tag{11}$$

for sufficiently large L, where σ is positive and goes to 0 as ϵ does. This applies because $|S\cap H| \ge (1-2\beta)L$ which grows with L since $\beta < 1/2$. The q marginal of X and $U_{S\cap H}$ is just $p(x)\tilde{Q}(u_{S\cap H}|x)$, so by the encoding method, $(X^n,\tilde{U}^n_{S\cap H})$ is jointly typical with respect to q with high probability. By (11), the same holds for $(X^n,\hat{U}^n_{S\cap H})$. Applying the Markov Lemma to $X\to U_{S\cap H}\to\tilde{X}$ in q and using the definition of $\tilde{X}^n$ yields that the pair $(X^n,\tilde{X}^n)$ is jointly typical with respect to q with high probability. Further applying the Markov Lemma to both $\tilde{X}\to U_{S\cap H}\to U_{S\setminus H}$ and $\tilde{X}\to U_S\to\hat{X}$ yields that $(\tilde{X}^n,\hat{X}^n)$ is also jointly typical with respect to q.

All that remains is to evaluate the error exponent of (10). Let $\gamma = |S\cap H|/L$. Certainly $\gamma \ge 1-2\beta$. Note that if $t(u_{S\cap H}) = t$, then

$$\begin{aligned} \frac{p(x_1)\tilde{Q}(u_{S\cap H}|x_1)}{\Pr(u_{S\cap H})} &= \frac{p(x_1)2^{-\gamma L[D(t\|\tilde{Q}(u|x_1))+H(t)]}}{\sum_x p(x)2^{-\gamma L[D(t\|\tilde{Q}(u|x))+H(t)]}} \\ &\le 2^{-\gamma L[D(t\|\tilde{Q}(u|x_1))-\min_x D(t\|\tilde{Q}(u|x))-\delta]} \end{aligned}$$

for any $\delta>0$ and sufficiently large L. Therefore

$$\begin{aligned} &\sum_{u_{S\cap H}\in\Lambda_t^{\gamma L}(U)} \frac{p(x_1)\tilde{Q}(u_{S\cap H}|x_1)p(x_2)\tilde{Q}(u_{S\cap H}|x_2)}{\Pr(u_{S\cap H})} \\ &\le 2^{\gamma L H(t)}2^{-\gamma L[D(t\|\tilde{Q}(u|x_1))-\min_x D(t\|\tilde{Q}(u|x))-\delta]} \\ &\quad\cdot p(x_2)2^{-\gamma L[D(t\|\tilde{Q}(u|x_2))+H(t)]} \\ &\le 2^{-\gamma L[D(t\|\tilde{Q}(u|x_1))+D(t\|\tilde{Q}(u|x_2))-\min_x D(t\|\tilde{Q}(u|x))-\delta]} \end{aligned}$$

for sufficiently large L. Therefore, using the fact that the number of types t is polynomial in L,

$$\begin{aligned} \frac{-\log P_e(1)}{L} &\ge \min_{x_1,x_2:x_1\neq x_2}\min_t \gamma\big[D(t\|\tilde{Q}(u|x_1)) \\ &\qquad + D(t\|\tilde{Q}(u|x_2)) - \min_x D(t\|\tilde{Q}(u|x)) - \delta\big] \\ &= \min_t \min_x 2\,\gamma D(t\|\tilde{Q}(u|x)) - \delta \end{aligned} \tag{12}$$

where $\min 2$ takes the second smallest value. It can be shown that this term involving the second smallest value of x is the minimum Chernoff Information. That is,

$$\frac{-\log P_e(1)}{L} \ge \min_{x_1,x_2}\gamma D(\tilde{Q}_\lambda\|\tilde{Q}(u|x_1)) - \delta$$

where $\tilde{Q}_\lambda$ and λ are defined by (9). Recalling that $\gamma\ge 1-2\beta$ and taking the limit as $\delta\to 0$ gives

$$\lim_{L\to\infty}\frac{-\log P_e(1)}{L} \ge \min_{x_1,x_2}(1-2\beta)D(\tilde{Q}_\lambda\|\tilde{Q}(u|x_1)).$$

This proves Lemma 1 since $R/L \approx I(Y;U|X)$.

D. Tighter Bound

Now we improve this bound by introducing the additional auxiliary random variable J. As in [6], we alter the coding scheme described in Section III-B so that the agents are split into groups, each with a different method of quantization. Partition $\{1,\cdots,L\}$ into disjoint sets R_j such that $\big||R_j| - P_J(j)L\big| \le 1$ for all j. For all $i\in R_j$, the encoder for agent i follows the procedure described in Section III-B where U is distributed by $Q(u|y,J=j)$. Observe that the rate for agent $i\in R_j$ becomes $I(Y;U|X,J=j)$, so

$$\frac{R}{L} = \frac{1}{L}\sum_j |R_j| I(Y;U|X,J=j) \le I(Y;U|X,J) + \mathcal{O}(\tfrac{1}{L}). \tag{13}$$

The decoder and estimation rules are the same, except the CEO now looks for a set S of size $(1-\beta)L$ such that u_S^n is typical with respect to the distribution

$$\sum_x p(x)\prod_{j\in\mathcal{J}}\prod_{i\in S\cap R_j}\tilde{Q}(u_i|x,J=j).$$

Let $\gamma_j = |R_j\cap S\cap H|/L$. Then

$$\sum_j \gamma_j \ge 1-2\beta \text{ and } \gamma_j \le P_J(j)\ \forall j\in\mathcal{J}. \tag{14}$$

Observe that if $t(u_{R_j\cap S\cap H}) = t_j$ for all j, then

$$\tilde{Q}(u_{S\cap H}|x) = \prod_j 2^{-L\gamma_j[D(t_j\|\tilde{Q}(u|x,j))+H(t_j)]}.$$

Applying this to (10) yields

$$\frac{-\log P_e}{L} \ge \min_{x_1,x_2}\sum_j \gamma_j D(\tilde{Q}_{\lambda,j}\|\tilde{Q}(u|x_1,j)) \tag{15}$$

where $\tilde{Q}_{\lambda,j}$ is given by (3) and (4). Extending (15) to minimize over all γ_j satisfying (14), then combining the result with (13) completes the proof of Theorem 1.

IV. Outer Bound

The two parts of the outer bounds result from two different actions the traitors can take. First, the traitors may perform a black hole attack. Evaluating the probability of error of such a strategy is complicated by the fact that the traitors may intelligently choose which agents to remove from the useful pool. Therefore this decision is made "after" the choice of the code. By an argument similar to that used in the converse proof of [6], $E(p, W, \beta)$ can be upper bounded by

$$\max_{U_i: X \to Y_i \to U_i} \frac{\min\limits_{\substack{H, x_1, x_2: \\ |H|=(1-\beta)L}} \frac{1}{L} \sum\limits_{i \in H} D(\tilde{Q}_{\lambda,i} \| \tilde{Q}_i(u|x_1))}{\frac{1}{L} \sum\limits_{i=1}^{L} I(Y_i; U_i|X)} \tag{16}$$

where H represents the set of honest sensors left over after the traitors prevent agents H^C from transmitting. The numerator of (16) can be rewritten to

$$\min_{\gamma_i, x_1, x_2} \sum_{i=1}^{L} \gamma_i D(\tilde{Q}_{\lambda,i} \| \tilde{Q}_i(u|x_1)) \tag{17}$$

under the condition that

$$\sum_{i=1}^{L} \gamma_i = 1 - \beta \text{ and } \gamma_i \in \{0, \tfrac{1}{L}\} \ \forall i \in \{1, \cdots, L\}. \tag{18}$$

We claim that the value of (17) does not change if we relax the condition to

$$\sum_{i=1}^{L} \gamma_i \geq 1 - \beta \text{ and } 0 \leq \gamma_i \leq \tfrac{1}{L} \ \forall i \in \{1, \cdots, L\}. \tag{19}$$

This is because we may use L arbitrarily large, so any γ_i satisfying the conditions of (19) can be approximated arbitrarily closely by one satisfying (18). Replacing the numerator of (16) with (17) under the conditions of (19), then continuing the argument from [6] yields (5).

The traitors may also perform the following more complicated attack. The idea is to produce a fraudulent X' that agrees with some honest measurements but not others. If they do this properly, they CEO will be unable to tell which of X and X' is the truth. In particular, if the traitors are H^C, they choose another set S with $|S| = (1-\beta)L$ and $|H \cap S| = (1-2\beta)L$. Then they take the value $Y^n_{H \cap S}$ and use it simulate the conditional distribution on variables X' and $Y_{S \setminus H}$ given by

$$\frac{p(x')W(y_{S \setminus H}|x')W(y_{H \cap S}|x')}{\Pr(y_{H \cap S})}$$

for each time t. Finally, they take the resulting $Y^n_{S \setminus H}$ and report this as the truth, using the encoding rule—no matter what it is—just as an honest agent would. Thus for any t, $X(t)$, $X'(t)$, and $Y^L(t)$ are distributed according to

$$\frac{p(x)W(y_{H \setminus S}|x)W(y_{H \cap S}|x)p(x')W(y_{S \setminus H}|x')W(y_{H \cap S}|x')}{\Pr(y_{H \cap S})}.$$

This distribution is symmetrical in X and X', so the CEO cannot know which of X^n and X'^n is the truth, meaning $P_e \geq \frac{1}{n}\mathbb{E}d_H(X^n, X'^n)$. An argument along the lines of that evaluating (10) shows that

$$\frac{-\log P_e}{L} \leq \min_{x_1, x_2} (1-2\beta) D(W_\lambda \| W(y|x_1)) \tag{20}$$

where W_λ is given by (8). Observe that the error exponent given by the right-hand side of (20) is completely independent from the code. Hence, if the optimum error exponent without traitors is less than this quantity, then the traitors, by employing this attack, have no effect at all on the overall error exponent, meaning that $E(p, W, \beta) = E(p, W, 0)$. However, we have already shown in our proof of (5) that $E(p, W, \beta) < E(p, W, 0)$ for any positive β. Therefore, we need only consider values of P_J and Q for which the error exponent without traitors is greater than the right-hand side of (20), i.e. those for which (7) holds. Further applying arguments from the converse of [6] proves (6).

V. Conclusion

It is yet unclear which, if either, of our inner and outer bounds is tight. The inner bound fails to take into account the fact that the traitors do not have access to the codewords transmitted by honest agents; they must choose their transmissions based only on measurements Y. If this constraint were relaxed, then the inner bound would be tight, but taking advantage of it ought to increase the achievable error exponent. On the other hand, the outer bound perhaps takes too much into account that the traitors must base their transmissions on the measurements Y. They should be able to do something somewhat more insidious than the attacks described in Section IV by using their knowledge of how the honest codewords are generated. However, these arguments are entirely intuitional, and attempts to find better computable bounds formally have failed. Indeed, demonstrating that one of our bounds is tight would be a surprising and interesting result.

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. Information Theory*, vol. IT-19, pp. 471–480, 1973.

[4] O. Kosut and L. Tong, "Distributed source coding in the presence of Byzantine sensors," to appear in *IEEE Trans. Inform. Theory*, 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. 42, pp. 887–902, May. 1996.

[7] T. H. S. Jaggi, M. Langberg and M. Effros, "Correction of adversarial errors in networks," in *Proceedings of International Symposium in Information Theory and its Applications*, (Adelaide, Australia), 2005.