Distributed Source Coding in the Presence of Byzantine Sensors

Oliver Kosut, *Student Member, IEEE*, and Lang Tong, *Fellow, IEEE*

***Abstract*—The distributed source coding problem is considered when the sensors, or encoders, are under Byzantine attack; that is, an unknown group of sensors have been reprogrammed by a malicious intruder to undermine the reconstruction at the fusion center. Three different forms of the problem are considered. The first is a variable-rate setup, in which the decoder adaptively chooses the rates at which the sensors transmit. An explicit characterization of the variable-rate achievable sum rates is given for any number of sensors and any groups of traitors. The converse is proved constructively by letting the traitors simulate a fake distribution and report the generated values as the true ones. This fake distribution is chosen so that the decoder cannot determine which sensors are traitors while maximizing the required rate to decode every value. Achievability is proved using a scheme in which the decoder receives small packets of information from a sensor until its message can be decoded, before moving on to the next sensor. The sensors use randomization to choose from a set of coding functions, which makes it probabilistically impossible for the traitors to cause the decoder to make an error. Two forms of the fixed-rate problem are considered, one with deterministic coding and one with randomized coding. The achievable rate regions are given for both these problems, and it is shown that lower rates can be achieved with randomized coding.**

***Index Terms*—Byzantine attack, distributed source coding, network security, sensor fusion, Slepian–Wolf coding.**

I. Introduction

We consider a modification to the distributed source coding problem in which an unknown subset of sensors are taken over by a malicious intruder and reprogrammed. We assume there are m sensors. Each time slot, sensors i for $i = 1, \ldots, m$ observe random variables X_i according to the joint probability distribution $p(x_1 \cdots x_m)$. Each sensor encodes its observation independently and transmits a message to a common decoder, which attempts to reconstruct the source values with small probability of error based on those messages. A subset of sensors are *traitors*, while the rest are *honest*. Unbeknownst to the honest sensors or the decoder, the traitors have been reprogrammed to cooperate to obstruct the goal of the network, launching a so-called Byzantine attack. To counter this attack, the honest sensors and decoder must employ strategies so that the decoder can correctly reconstruct source values no matter what the traitors do.

It is obvious that observations made by the traitors are irretrievable unless the traitors choose to deliver them to the decoder. Thus the best the decoder can hope to achieve is to reconstruct the observations of the honest sensors. A simple procedure is to ignore the statistical correlations among the observations and collect data from each sensor individually. The total sum rate of such an approach is $\sum_i H(X_i)$. One expects however that this sum rate can be lowered if the correlation structure is not ignored.

Without traitors, Slepian–Wolf coding [1] can be used to achieve a sum rate as low as

$$H(X_1 \cdots X_m). \qquad (1)$$

However, standard Slepian–Wolf coding has no mechanism for handling any deviations from the agreed-upon encoding functions by the sensors. Even a random fault by a single sensor could have devastating consequences for the accuracy of the source estimates produced at the decoder, to say nothing of a Byzantine attack on multiple sensors. In particular, because Slepian–Wolf coding takes advantage of the correlation among sources, manipulating the codeword for one source can alter the accuracy of the decoder's estimate for other sources. It will turn out that for most source distributions, the sum rate given in (1) cannot be achieved if there is even a single traitor.

In this paper, we are interested in the lowest achievable sum-rate such that the decoder can reconstruct observations of the honest sensors with arbitrarily small error probability. In some cases, we are also interested in the rate region. We note that although the problem setup does not allow the detector to distinguish traitors from the honest sensors, an efficient scheme that guarantees the reconstruction of data from honest sensors is of both theoretical and practical interest. For example, for a distributed inference problem in the presence of Byzantine sensors, a practical (though not necessarily optimal) solution is to attack the problem in two separate phases. In the first phase, the decoder collects data from sensors over multiple access channels with rate constraints. Here we require that data from honest sensors are perfectly reconstructed at the decoder even though the decoder does not know which piece of data is from an honest sensor. In the second step, the received data is used for statistical inference. The example of distributed detection in the presence of Byzantine sensors is considered in [2]. The decoder may also have other side information about the content of the messages that allows the decoder to distinguish messages from the honest sensors.

Manuscript received February 16, 2007; revised February 15, 2008. This work is supported in part by the National Science Foundation under Award CCF-0635070, the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program DAAD19-01-2-0011, and TRUST (The Team for Research in Ubiquitous Secure Technology) sponsored by the National Science Foundation under award CCF-0424422.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: oek2@cornell.edu; lt35@cornell.edu.)

Communicated by U. Maurer, Guest Editor for Special Issue on Information Theoretic Security.

Digital Object Identifier 10.1109/TIT.2008.921867

A. Related Work

The notion of Byzantine attack has its root in the Byzantine generals problem [3], [4] in which a clique of traitorous generals conspire to prevent loyal generals from forming consensus. It was shown in [3] that consensus in the presence of Byzantine attack is possible if and only if less than a third of the generals are traitors.

Countering Byzantine attacks in communication networks has also been studied in the past by many authors. See the earlier work of Perlman [5] and also more recent review [6], [7]. An information theoretic network coding approach to Byzantine attack is presented in [8]. In [9], Awerbuch *et al.* suggest a method for mitigating Byzantine attacks on routing in ad hoc networks. Their approach is most similar to ours in the way they maintain a list of current knowledge about which links are trustworthy, constantly updated based on new information. Sensor fusion with Byzantine sensors was studied in [10]. In that paper, the sensors, having already agreed upon a message, communicate it to the fusion center over a discrete memoryless channel. Quite similar results were shown in [11], in which a malicious intruder takes control of a set of links in the network. The authors show that two nodes can communicate at a nonzero rate as long as less than half of the links between them are Byzantine. This is different from the current paper in that the transmitter chooses its messages, instead of relaying information received from an outside source, but some of the same approaches from [11] are used in the current paper, particularly the use of randomization to fool traitors that have already transmitted.

B. Redefining Achievable Rate

The nature of Byzantine attack require three modifications to the usual notion of achievable rate. The first, as mentioned above, is that small probability of error is required only for honest sources, even though the decoder may not know which sources are honest. This requirement is reminiscent of [3], in which the lieutenants need only perform the commander's order if the commander is not a traitor, even though the lieutenants might not be able to decide this with certainty.

The next modification is that there must be small probability of error no matter what the traitors do. This is essentially the definition of Byzantine attack.

The final modification has to do with which sensors are allowed to be traitors. Let $\mathcal{H}$ be the set of honest sensors, and $\mathcal{T} = \{1, \dots, m\} \backslash \mathcal{H}$ the set of traitors. A statement that a code achieves a certain rate must include the list of sets of sensors that this code can handle as the set of traitors. That is, given such a list, we say that a rate is achieved if there exists a code with small probability of error when the actual set of traitors is in fact on the list. Hence a given code may work for some lists and not others, so the achievable rates will depend on the specified list. It will be more convenient to specify not the list of allowable sets of traitors, but rather the list of allowable sets of honest sensors. We define $\mathfrak{H} \subset 2^{\{1,\dots,m\}}$ to be this list. Thus small probability of error is required only when $\mathcal{H} \in \mathfrak{H}$. One special case is when the code can handle any group of at most t traitors. That is,

$$\mathfrak{H} = \mathfrak{H}_t \triangleq \{\mathcal{S} \subset \{1, \dots, m\} : |\mathcal{S}| \geq m - t\}.$$

Observe that achievable rates depend not just on the true set of traitors but also on the collection $\mathfrak{H}$, because the decoder's willingness to accept more and more different groups of traitors allows the true traitors to get away with more without being detected. Thus we see a trade off between rate and security—in order to handle more traitors, one needs to be willing to accept a higher rate.

C. Fixed-Rate Versus Variable-Rate Coding

In standard source coding, an encoder is made up of a single encoding function. We will show that this fixed-rate setup is suboptimal for this problem, in the sense that we can achieve lower sum rates using variable-rate coding. By variable-rate we mean that the number of bits transmitted per source value by a particular sensor will not be fixed. Instead, the decoder chooses the rates at "run time" in the following way. Each sensor has a finite number of encoding functions, all of them fixed beforehand, but with potentially different output alphabets. The coding session is then made up of a number of transactions. Each transaction begins with the decoder deciding which sensor will transmit, and which of its several encoding functions it will use. The sensor then executes the chosen encoding function and transmits the output back to the decoder. Finally, the decoder uses the received message to choose the next sensor and encoding function, beginning the next transaction, and so on. Thus a code is made up of a set of encoding functions for each sensor, a method for the decoder to choose sensors and encoding functions based on previously received messages, and lastly a decoding function that takes all received messages and produces source estimates.

Note that the decoder has the ability to transmit some information back to the sensors, but this feedback is limited to the choice of encoding function. Since the number of encoding functions need not grow with the block length, this represents zero rate feedback.

In variable-rate coding, since the rates are only decided upon during the coding session, there is no notion of an m-dimensional achievable rate region. Instead, we only discuss achievable sum rates.

D. Traitor Capabilities

An important consideration with Byzantine attack is the information to which the traitors have access. First, we assume that the traitors have complete knowledge of the coding scheme used by the decoder and honest sensors. Furthermore, we always assume that they can communicate with each other arbitrarily. For variable-rate coding, they may have any amount of ability to eavesdrop on transmissions between honest sensors and the decoder. We will show that this ability has no effect on achievable rates. We assume with fixed-rate coding that all sensors transmit simultaneously, so it does not make sense that traitors could eavesdrop on honest sensors' transmissions before making their own, as that would violate causality. Thus we assume for fixed-rate coding that the traitors cannot eavesdrop.

The key factor, however, is the extent to which the traitors have direct access to information about the sources. We assume the most general memoryless case, that the traitors have access to the random variable W, where W is i.i.d. distributed with $(X_1 \cdots X_m)$ according to the conditional distribution $r(w|x_1 \cdots x_m)$. A natural assumption would be that W always includes X_i for traitors i, but in fact this need not be the case. An important special case is where $W = (X_1, \dots, X_m)$, i.e., the traitors have perfect information.

We assume that the distribution of W depends on who the traitors are, and that the decoder may not know exactly what this distribution is. Thus each code is associated with a function $\mathcal{R}$ that maps elements of $\mathfrak{H}$ to sets of conditional distributions r. The relationship between r and $\mathcal{R}(\mathcal{H})$ is analogous to the relationship between $\mathcal{H}$ and $\mathfrak{H}$. That is, given $\mathcal{H}$, the code is willing to accept all distributions $r \in \mathcal{R}(\mathcal{H})$. Therefore a code is designed based on $\mathfrak{H}$ and $\mathcal{R}$, and then the achieved rate depends at run time on $\mathcal{H}$ and r, where we assume $\mathcal{H} \in \mathfrak{H}$ and $r \in \mathcal{R}(\mathcal{H})$. We therefore discuss not achievable rates R but rather achievable rate functions $R(\mathcal{H}, r)$. In fact, this applies only to variable-rate codes. In the fixed-rate case, no run time rate decisions can be made, so achievable rates depend only on $\mathfrak{H}$ and $\mathcal{R}$.

E. Main Results

The main results of this paper give explicit characterizations of the achievable rates for three different setups. The first, which is discussed in the most depth, is the variable-rate case, for which we characterize achievable sum rate functions. The other two setups are for fixed-rate coding, divided into deterministic and randomized coding, for which we give m-dimensional achievable rate regions. We show that randomized coding yields a larger achievable rate region than deterministic coding, but we believe that in most cases randomized fixed-rate coding requires an unrealistic assumption. In addition, even randomized fixed-rate coding cannot achieve the same sum rates as variable-rate coding.

We give the exact solutions later, but describe here some intuition behind them. For variable-rate, the achievable rates, given in Theorem 1, are based on alternate distributions on $(X_1 \cdots X_m)$. Specifically, given W, the traitors can simulate any distribution $\bar{q}(x_\mathcal{T}|w)$ to produce a fraudulent version of $X_\mathcal{T}^n$, then report this sequence as the truth. Suppose that the overall distribution $q(x_1 \cdots x_m)$ governing the combination of the true value of $X_\mathcal{H}^n$ with this fake value of $X_\mathcal{T}^n$ could be produced in several different ways, with several different sets of traitors. In that case, the decoder cannot tell which of these several possibilities is the truth, which means that from its point of view, many sensors might be honest. Since the error requirement described in I-B stipulates that the decoder must produce a correct estimate for every honest sensor, it must attempt to decode the source values associated with each potentially honest sensor. Thus the sum rate must be at least the joint entropy, when distributed according to q, of the sources associated with all potentially honest sensors. The supremum over all possible simulated distributions is the achievable sum rate.

For example, suppose $\mathfrak{H} = \mathfrak{H}_{m-1}$. That is, at most one sensor is honest. Then the traitors are able to create the distribution $q(x_1 \cdots x_m) = p(x_1) \cdots p(x_m)$ no matter which group of $m-1$ sensors are the traitors. Thus every sensor appears as if it could be the honest one, so the minimum achievable sum rate is

$$H(X_1) + \cdots + H(X_m). \tag{2}$$

In other words, the decoder must use an independent source code for each sensor, which requires receiving $nH(X_i)$ bits from sensor i for all i.

The achievable fixed-rate regions, given in Theorem 2, are based on the Slepian–Wolf achievable rate region. For randomized fixed-rate coding, the achievable region is such that for all $\mathcal{S} \in \mathfrak{H}$, the rates associated with the sensors in $\mathcal{S}$ fall into the Slepian–Wolf rate region on the corresponding random variables. Note that for $\mathfrak{H} = \{\{1, \dots, m\}\}$, this is identical to the Slepian–Wolf region. For $\mathfrak{H} = \mathfrak{H}_{m-1}$, this region is such that for all i, $R_i \geq H(X_i)$, which corresponds to the sum rate in (2). The deterministic fixed-rate achievable region is a subset of that of randomized fixed-rate, but with an additional constraint stated in Section VI.

F. Randomization

Randomization plays a key role in defeating Byzantine attacks. As we have discussed, allowing randomized encoding in the fixed-rate situation expands the achievable region. In addition, the variable-rate coding scheme that we propose relies heavily on randomization to achieve small probability of error. In both fixed and variable-rate coding, randomization is used as follows. Every time a sensor transmits, it randomly chooses from a group of essentially identical encoding functions. The index of the chosen function is transmitted to the decoder along with its output. Without this randomization, a traitor that transmits before an honest sensor i would know exactly the messages that sensor i will send. In particular, it would be able to find fake sequences for sensor i that would produce those same messages. If the traitor tailors the messages it sends to the decoder to match one of those fake sequences, when sensor i then transmits, it would appear to corroborate this fake sequence, causing an error. By randomizing the choice of encoding function, the set of sequences producing the same message is not fixed, so a traitor can no longer know with certainty that a particular fake source sequence will result in the same messages by sensor i as the true one. This is not unlike Wyner's wiretap channel [12], in which information is kept from the wiretapper by introducing additional randomness. See in particular Section V-D for the proof that variable-rate randomness can defeat the traitors in this manner.

The rest of the paper is organized as follows. In Section II, we develop in detail the case that there are three sensors and one traitor, describing a coding scheme that achieves the optimum sum rate. In Section III, we formally give the variable-rate model and present the variable-rate result. In Section IV, we discuss the variable-rate achievable rate region and give an analytic formulation for the minimum achievable sum rate for some special cases. In Section VI, we give the fixed-rate models and present the fixed-rate result. In Sections V and VII, we prove the variable-rate and fixed-rate results respectively. Finally, in Section VIII, we conclude.

II. Three Sensor Example

A. Potential Traitor Techniques

For simplicity and motivation, we first explore the three-sensor case with one traitor. That is, $m = 3$ and

$$\mathfrak{H} = \{\{1,2\},\{2,3\},\{1,3\}\}.$$

Suppose also that the traitor has access to perfect information (i.e., $W = (X_1, X_2, X_3)$). Suppose sensor 3 is the traitor. Sensors 1 and 2 will behave honestly, so they will report X_1 and X_2 correctly, as distributed according to the marginal distribution $p(x_1 x_2)$. Since sensor 3 has access to the exact values of X_1 and X_2, it may simulate the conditional distribution $p(x_3|x_2)$, then take the resulting X_3 sequence and report it as the truth. Effectively, then, the three random variables will be distributed according to the distribution

$$q(x_1 x_2 x_3) \triangleq p(x_1 x_2) p(x_3|x_2).$$

The decoder will be able to determine that sensors 1 and 2 are reporting jointly typical sequences, as are sensors 2 and 3, but not sensors 1 and 3. Therefore, it can tell that either sensor 1 or 3 is the traitor, but not which one, so it must obtain estimates of the sources from all three sensors. Since the three streams are not jointly typical with respect to the source distribution $p(x_1 x_2 x_3)$, standard Slepian–Wolf coding on three encoders will not correctly decode them all. However, had we known the strategy of the traitor, we could do Slepian–Wolf coding with respect to the distribution q. This will take a sum rate of

$$H_q(X_1 X_2 X_3) = H(X_1 X_2 X_3) + I(X_1; X_3|X_2)$$

where H_q is the entropy with respect to q. In fact we will not do Slepian–Wolf coding with respect to q but rather something slightly different that gives the same rate. Since Slepian–Wolf coding without traitors can achieve a sum rate of $H(X_1 X_2 X_3)$, we have paid a penalty of $I(X_1; X_3|X_2)$ for the single traitor.

We supposed that sensor 3 simulated the distribution $p(x_3|x_2)$. It could have just as easily simulated $p(x_3|x_1)$, or another sensor could have been the traitor. Hence, the minimum achievable sum rate for all $\mathcal{H} \in \mathfrak{H}$ is at least

$$R^* \triangleq H(X_1 X_2 X_3) + \max\{I(X_1; X_2|X_3), \\ I(X_1; X_3|X_2), I(X_2; X_3|X_1)\}. \quad (3)$$

In fact, this is exactly the minimum achievable sum rate, as shown below.

B. Variable-Rate Coding Scheme

We now give a variable-rate coding scheme that achieves R^*. This scheme is somewhat different from the one we present for the general case in Section V, but it is much simpler, and it illustrates the basic idea. The procedure will be made up of a number of rounds. Communication from sensor i in the first round will be based solely on the first n values of X_i, in the second round on the second n values of X_i, and so on. The principle advantage of the round structure is that the decoder may hold onto information that is carried over from one round to the next.

In particular, the decoder maintains a collection $\mathfrak{V} \subset \mathfrak{H}$ representing the sets that could be the set of honest sensors. If a sensor is completely eliminated from $\mathfrak{V}$, that means it has been identified as the traitor. We begin with $\mathfrak{V} = \mathfrak{H}$, and then remove a set from $\mathfrak{V}$ whenever we find that the messages from the corresponding pair of sensors are not jointly typical. With high probability, the two honest sensors report jointly typical sequences, so we expect never to eliminate the honest pair from $\mathfrak{V}$. If the traitor employs the q discussed above, for example, we would expect sensors 1 and 3 to report atypical sequences, so we will drop $\{1, 3\}$ from $\mathfrak{V}$. In essence, the value of $\mathfrak{V}$ contains our current knowledge about what the traitor is doing.

The procedure for a round is as follows. If $\mathfrak{V}$ contains $\{1, 2, 1, 3\}$, do the following.

1) Receive $nH(X_1)$ bits from sensor 1 and decode x_1^n.
2) Receive $nH(X_2|X_1)$ bits from sensor 2. If there is a sequence in $\mathcal{X}_2^n$ jointly typical with x_1^n that matches this transmission, decode that sequence to x_2^n. If not, receive $nI(X_1; X_2)$ additional bits from sensor 2, decode x_2^n, and remove $\{1, 2\}$ from $\mathfrak{V}$.
3) Do the same with sensor 3: Receive $nH(X_3|X_1)$ bits and decode x_3^n if possible. If not, receive $nI(X_1; X_3)$ additional bits, decode, and remove $\{1, 3\}$ from $\mathfrak{V}$.

If $\mathfrak{V}$ is one of the other two subsets of $\mathfrak{H}$ with two elements, perform the same procedure but replace sensor 1 with whichever sensor appears in both elements in $\mathfrak{V}$. If $\mathfrak{V}$ contains just one element, then we have exactly identified the traitor, so ignore the sensor that does not appear and simply do Slepian–Wolf coding on the two remaining sensors.

Note that the only cases when the number of bits transmitted exceeds nR^* are when we receive a second message from one of the sensors, which happens exactly when we eliminate an element from $\mathfrak{V}$. Assuming the source sequences of the two honest sensors are jointly typical, this can occur at most twice, so we can always achieve a sum rate of R^* when averaged over enough rounds.

C. Fixed-Rate Coding Scheme

In the procedure described above, the number of bits sent by a sensor changes from round to round. We can no longer do this with fixed-rate coding, so we need a different approach. Suppose sensor 3 is the traitor. It could perform a black hole attack, in which case the estimates for X_1^n and X_2^n must be based only on the messages from sensors 1 and 2. Thus, the rates R_1 and R_2 must fall into the Slepian–Wolf achievability region for X_1 and X_2. Similarly, if one of the other sensors was the traitor, the other pairs of rates also must fall into the corresponding Slepian–Wolf region. Putting these conditions together gives

$$R_1 \geq \max\{H(X_1|X_2), H(X_1|X_3)\}$$
$$R_2 \geq \max\{H(X_2|X_1), H(X_2|X_3)\}$$

$$\begin{aligned} R_3 &\geq \max\{H(X_3|X_1), H(X_3|X_2)\} \\ R_1 + R_2 &\geq H(X_1X_2) \\ R_1 + R_3 &\geq H(X_1X_3) \\ R_2 + R_3 &\geq H(X_2X_3). \end{aligned} \tag{4}$$

If the rates fall into this region, we can do three simultaneous Slepian–Wolf codes, one on each pair of sensors, thereby constructing two estimates for each sensor. If we randomize these codes using the method described in Section I-F, the traitor will be forced either to report the true message, or report a false message, which with high probability will be detected as such. Thus either the two estimates for each sensor will be the same, in which case we know both are correct, or one of the estimates will be demonstrably false, in which case the other is correct.

We now show that the region given by (4) does not include sum rates as low as R^*. Assume without loss of generality that $I(X_1;X_2|X_3)$ achieves the maximum in (3). Summing the last three conditions in (4) gives

$$\begin{aligned} &R_1 + R_2 + R_3 \\ &\geq \frac{1}{2}\big(H(X_1X_2) + H(X_1X_3) + H(X_2X_3)\big) \\ &= H(X_1X_2X_3) + \frac{1}{2}\big(I(X_1;X_2|X_3) + I(X_1X_2;X_3)\big). \end{aligned} \tag{5}$$

If $I(X_1X_2;X_3) > I(X_1;X_2|X_3)$, (5) is larger than (3). Hence, there exist source distributions for which we cannot achieve the same sum rates with even randomized fixed-rate coding as with variable-rate coding.

If we are interested only in deterministic codes, the region given by (4) can no longer be achieved. In fact, we will prove in Section VII that the achievable region reduces to the trivially achievable region where $R_i \geq H(X_i)$ for all i when $m = 3$, though it is nontrivial for $m > 3$. For example, suppose $m = 4$ and $\mathfrak{H} = \mathfrak{H}_1$. In this case, the achievable region is similar to that given by (4), but with an additional sensor. That is, each of the 6 pairs of rates must fall into the corresponding Slepian–Wolf region. In this case, we do three simultaneous Slepian–Wolf codes for each sensor, construct three estimates, each associated with one of the other sensors. For an honest sensor, only one of the other sensors could be a traitor, so at least two of these estimates must be correct. Thus we need only take the plurality of the three estimates to obtain the correct estimate.

III. Variable-Rate Model and Result

Notation

Let X_i be the random variable revealed to sensor i, $\mathcal{X}_i$ the alphabet of that variable, and x_i a corresponding realization. A sequence of random variables revealed to sensor i over n timeslots is denoted X_i^n, and a realization of it $x_i^n \in \mathcal{X}_i^n$. Let $\mathcal{M} \triangleq \{1,\dots,m\}$. For a set $\mathcal{S} \subset \mathcal{M}$, let $X_{\mathcal{S}}$ be the set of random variables $\{X_i\}_{i\in\mathcal{S}}$, and define $x_{\mathcal{S}}$ and $\mathcal{X}_{\mathcal{S}}$ similarly. By $\mathcal{S}^c$ we mean $\mathcal{M}\backslash\mathcal{S}$. Let $T_\epsilon^n(X_{\mathcal{S}})[q]$ be the strongly typical set with respect to the distribution q, or the source distribution p if unspecified. Similarly, $H_q(X_{\mathcal{S}})$ is the entropy with respect to the distribution q, or p if unspecified.

A. Communication Protocol

The transmission protocol is composed of L transactions. In each transaction, the decoder selects a sensor to receive information from and selects which of K encoding functions it should use. The sensor then responds by executing that encoding function and transmitting its output back to the decoder, which then uses the new information to begin the next transaction.

For each sensor $i \in \mathcal{M}$ and encoding function $j \in \{1,\dots,K\}$, there is an associated rate $R_{i,j}$. On the lth transaction, let i_l be the sensor and j_l the encoding function chosen by the decoder, and let h_l be the number of $l' \in \{1,\dots,l-1\}$ such that $i_{l'} = i_l$. That is, h_l is the number of times i_l has transmitted prior to the lth transaction. Note that i_l, j_l, h_l are random variables, since they are chosen by the decoder based on messages it has received, which depend on the source values. The jth encoding function for sensor i is given by

$$f_{i,j} : \mathcal{X}_i^n \times \mathcal{Z} \times \{1,\dots,K\}^{h_l} \to \{1,\dots,2^{nR_{i,j}}\} \tag{6}$$

where $\mathcal{Z}$ represents randomness generated at the sensor. Let $I_l \in \{1,\dots,2^{nR_{i_l,j_l}}\}$ be the message received by the decoder in the lth transaction. If i_l is honest, then $I_l = f_{i_l,j_l}(X_{i_l}^n, \rho_{i_l}, J_l)$, where $\rho_{i_l} \in \mathcal{Z}$ is the randomness from sensor i_l and $J_l \in \{1,\dots,K\}^{h_l}$ is the history of encoding functions used by sensor i_l so far. If i_l is a traitor, however, it may choose I_l based on W^n and it may have any amount of access to previous transmissions $I_1,\dots,I_{l-1}$ and polling history $i_1,\dots,i_{l-1}$, $j_1,\dots,j_{l-1}$. But, it does not have access to the randomness ρ_i for any honest sensor i. Note again that the amount of traitor eavesdropping ability has no effect on achievable rates.

After the decoder receives I_l, if $l < L$ it uses $I_1,\dots,I_l$ to choose the next sensor i_{l+1} and its encoding function index j_{l+1}. After the Lth transaction, it decodes according to the decoding function

$$g : \prod_{l=1}^{L}\{1,\dots,2^{nR_{i_l,j_l}}\} \to \mathcal{X}_1^n \times \cdots \times \mathcal{X}_m^n.$$

Note that we impose no restriction whatsoever on the size of the total number of transactions L. Thus, a code could have arbitrary complexity in terms of the number of messages passed between the sensors and the decoder. However, in our below definition of achievability, we require that the communication rate from sensors to decoder always exceeds that from decoder to sensors. Therefore while the number of messages may be very large, the amount of feedback is diminishingly small.

B. Variable-Rate Problem Statement and Main Result

Let $\mathcal{H} \subset \mathcal{M}$ be the set of honest sensors. Define the probability of error

$$P_e \triangleq \Pr\left(X_{\mathcal{H}}^n \neq \hat{X}_{\mathcal{H}}^n\right)$$

where $(\hat{X}_1^n,\dots,\hat{X}_m^n) = g(I_1,\dots,I_L)$. The probability of error will in general depend on the actions of the traitors. Note again that we only require small probability of error on the source estimates corresponding to the honest sensors.

We define a rate function $R(\mathcal{H},r)$ defined for $\mathcal{H}\in\mathfrak{H}$ and $r\in\mathcal{R}(\mathcal{H})$ to be *α-achievable* if there exists a code such that, for all pairs $(\mathcal{H},r)$ and any choice of actions by the traitors, $P_e\leq\alpha$

$$\Pr\left(\sum_{l=1}^{L}R_{i_l,j_l}\leq R(\mathcal{H},r)\right)\geq 1-\alpha$$

and $\log K\leq\alpha nR_{i,j}$ for all i,j. This last condition requires, as discussed above, that the feedback rate from the decoder back to the sensors is arbitrarily small compared to the forward rate. A rate function $R(\mathcal{H},r)$ is *achievable* if for all $\alpha>0$, there is a sequence of α-achievable rate functions $\{R'_k(\mathcal{H},r)\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty}R'_k(\mathcal{H},r)=R(\mathcal{H},r).$$

Note that we do not require uniform convergence.

The following definitions allow us to state our main variable-rate result. For any $\mathcal{H}\in\mathfrak{H}$ and $r\in\mathcal{R}(\mathcal{H})$, let $\tilde{r}(w|x_{\mathcal{H}})$ be the distribution of W given $X_{\mathcal{H}}$ when W is distributed according to $r(w|x_{\mathcal{M}})$. That is

$$\tilde{r}(w|x_{\mathcal{H}})=\sum_{x_{\mathcal{H}^c}\in\mathfrak{X}_{\mathcal{H}^c}}p(x_{\mathcal{H}^c}|x_{\mathcal{H}})r(w|x_{\mathcal{H}}x_{\mathcal{H}^c}).$$

The extent to which W provides information about $X_{\mathcal{H}^c}$ is irrelevant to the traitors, since in order to fool the decoder they must generate information that appears to agree only with $X_{\mathcal{H}}$ as reported by the honest sensors. Thus it will usually be more convenient to work with $\tilde{r}$ rather than r. For any $\mathcal{S}\in\mathfrak{H}$ and $r'\in\mathcal{R}(\mathcal{S})$, let

$$\mathcal{Q}_{\mathcal{S},r'}\triangleq\left\{p(x_{\mathcal{S}})\sum_{w}\tilde{r}'(w|x_{\mathcal{S}})\bar{q}(x_{\mathcal{S}^c}|w):\forall\bar{q}(x_{\mathcal{S}^c}|w)\right\}.\quad(7)$$

If $\mathcal{S}^c$ were the traitors and W were distributed according to r', then $\mathcal{Q}_{\mathcal{S},r'}$ would be the set of distributions q to which the traitors would have access. That is, if they simulate the proper $\bar{q}(x_{\mathcal{S}^c}|w)$ from their received W, this simulated version of $X_{\mathcal{S}}$ and the true value of $X_{\mathcal{S}^c}$ would be jointly distributed according to q. For any $\mathfrak{V}\subset\mathfrak{H}$, define

$$\mathcal{Q}(\mathfrak{V})\triangleq\bigcap_{\mathcal{S}\in\mathfrak{V}}\bigcup_{r'\in\mathcal{R}(\mathcal{S})}\mathcal{Q}_{\mathcal{S},r'}$$

$$\mathcal{U}(\mathfrak{V})\triangleq\bigcup_{\mathcal{S}\in\mathfrak{V}}\mathcal{S}.$$

That is, for some distribution $q\in\mathcal{Q}(\mathfrak{V})$, for every $\mathcal{S}\in\mathfrak{V}$, if the traitors were $\mathcal{S}^c$, they would have access to q for some $r'\in\mathcal{R}(\mathcal{S})$. Thus any distribution in $\mathcal{Q}(\mathfrak{V})$ makes it look to the decoder like any $\mathcal{S}\in\mathfrak{V}$ could be the set of honest sensors, so any sensor in $i\in\mathcal{U}(\mathfrak{V})$ is potentially honest.

Theorem 1: A rate function $R(\mathcal{H},r)$ is achievable if and only if, for all $(\mathcal{H},r)$,

$$R(\mathcal{H},r)\geq R^*(\mathcal{H},r)\triangleq\sup_{\mathfrak{V}\subset\mathfrak{H},\ q\in\mathcal{Q}_{\mathcal{H},r}\cap\mathcal{Q}(\mathfrak{V})}H_q(X_{\mathcal{U}(\mathfrak{V})}).\quad(8)$$

See Section V for the proof.

We offer the following interpretation of this result. Suppose we placed the following constraint on the traitors' behavior. Given W^n, they must produce a value of $X_{\mathcal{T}}^n$ in an i.i.d. fashion, then report it as the truth. That is, they choose a value of $X_{\mathcal{T}}$ at time τ based only on W at time τ, making each choice in an identical manner. Then each traitor i takes the produced value of X_i^n and behaves for the duration of the coding session exactly as if it were honest and this was the true source sequence. We can now easily classify all possible behaviors of the traitors simply by specifying the manner in which they generate $X_{\mathcal{T}}$ from W, which is given by some distribution $\bar{q}(x_{\mathcal{T}}|w)$. The joint distribution of $X_{\mathcal{H}}$ and $X_{\mathcal{T}}$ will be given by

$$q(x_{\mathcal{M}})=p(x_{\mathcal{H}})\sum_{w}\tilde{r}(w|x_{\mathcal{H}})\bar{q}(x_{\mathcal{T}}|w).\quad(9)$$

By (7), $q\in\mathcal{Q}_{\mathcal{H},r}$. If q is also contained in $\mathcal{Q}_{\mathcal{S},r'}$ for some $\mathcal{S}\in\mathfrak{H}$ and $r'\in\mathcal{R}(\mathcal{S})$, then again by (7), there exists a distribution $\bar{q}'(x_{\mathcal{S}}|w)$ such that

$$q(x_{\mathcal{M}})=p(x_{\mathcal{S}})\sum_{w}\tilde{r}'(w|x_{\mathcal{S}})\bar{q}'(x_{\mathcal{S}}|w).\quad(10)$$

Since (9) and(10) have exactly the same form, the decoder will not be able to determine whether $\mathcal{H}$ is the set of honest sensors with W distributed according to r, or $\mathcal{S}$ is the set of honest sensors with W distributed according to r'. On the other hand, if for some $\mathcal{S}\in\mathfrak{H}$, $q\notin\mathcal{Q}_{\mathcal{S},r'}$ for all $r'\in\mathcal{R}(\mathcal{S})$, then the decoder should be able to tell that $\mathcal{S}$ is not the set of honest sensors. We have not yet said how it might know, but intuition suggests that it should be possible. Hence, if there is no $\mathcal{S}$ containing a certain sensor i for which

$$q\in\bigcup_{r'\in\mathcal{R}(\mathcal{S})}\mathcal{Q}_{\mathcal{S},r'}\quad(11)$$

then the decoder can be sure that i is a traitor and it may be ignored. Let $\mathfrak{V}$ be the collection of all $\mathcal{S}\in\mathfrak{H}$ for which (11) holds. Every sensor in $\mathcal{U}(\mathfrak{V})$ looks to the decoder like it could be honest; all the rest are surely traitors. Thus, in order to make sure that the decoder reconstructs honest information perfectly, it must recover X_i^n for all $i\in\mathcal{U}(\mathfrak{V})$, which means the sum rate must be at least $H_q(X_{\mathcal{U}(\mathfrak{V})})$. Observe that

$$q\in\bigcap_{\mathcal{S}\in\mathfrak{V}}\bigcup_{r'\in\mathcal{R}(\mathcal{S})}\mathcal{Q}_{\mathcal{S},r'}=\mathcal{Q}(\mathfrak{V}).$$

As already noted, $q\in\mathcal{Q}_{\mathcal{H},r}$, so $q\in\mathcal{Q}_{\mathcal{H},r}\cap\mathcal{Q}(\mathfrak{V})$. Moreover, for any $\mathfrak{V}\subset\mathfrak{H}$, every element of $\mathcal{Q}_{\mathcal{H},r}\cap\mathcal{Q}(\mathfrak{V})$ can be produced with the proper choice of $\bar{q}(x_{\mathcal{T}}|w)$. Hence $H_q(X_{\mathcal{U}(\mathfrak{V})})$ can be as high as

$$\sup_{\mathfrak{V}\subset\mathfrak{H},\ q\in\mathcal{Q}_{\mathcal{H},r}\cap\mathcal{Q}(\mathfrak{V})}H_q(X_{\mathcal{U}(\mathfrak{V})})=R^*(\mathcal{H},r)$$

but no higher. Thus it makes sense that this rate and no better can be achieved if we place this constraint on the traitors. Therefore Theorem 1 can be interpreted as stating that constraining the traitors in this manner has no effect on the set of achievable rates.

IV. Properties of the Variable-Rate Region

It might at first appear that (8) does not agree with (3). We discuss several ways in which (8) can be made more manageable, particularly in the case of perfect traitor information (i.e., $W=X_{\mathcal{M}}$), and show that the two are in fact identical. Let R^*

be the minimum rate achievable over all $\mathcal{H} \in \mathfrak{H}$ and $r \in \mathcal{R}(\mathcal{H})$. Thus by (8), we can write

$$R^* = \sup_{\mathcal{H}\in\mathfrak{H}, r\in\mathcal{R}(\mathcal{H})} R^*(\mathcal{H}, r) = \sup_{\mathfrak{V}\subset\mathfrak{H},\ q\in\mathcal{Q}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})}). \tag{12}$$

This is the quantity that appears in (3). Note also that for perfect traitor information

$$\mathcal{Q}_{\mathcal{S},r'} = \{q(x_{\mathcal{M}}) : q(x_{\mathcal{S}}) = p(x_{\mathcal{S}})\}. \tag{13}$$

This means that $\mathcal{Q}_{\mathcal{H},r} \cap \mathcal{Q}(\mathfrak{V}) = \mathcal{Q}(\mathfrak{V} \cup \{\mathcal{H}\})$. Therefore (8) becomes

$$R^*(\mathcal{H}, r) = \sup_{\mathfrak{V}\subset\mathfrak{H}:\mathcal{H}\in\mathfrak{V},\ q\in\mathcal{Q}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})}).$$

The following lemma simplifies calculation of expressions of the form $\sup_{q\in\mathcal{Q}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})})$.

Lemma 1: Suppose the traitors have perfect information. For any $\mathfrak{V} \subset \mathfrak{H}$, the expression

$$\sup_{q\in\mathcal{Q}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})}) \tag{14}$$

is maximized by a q satisfying(13) for all $\mathcal{S} \in \mathfrak{V}$ such that, for some set of functions $\{\sigma_{\mathcal{S}}\}_{\mathcal{S}\in\mathfrak{V}}$

$$q(x_1 \cdots x_m) = \prod_{\mathcal{S}\in\mathfrak{V}} \sigma_{\mathcal{S}}(x_{\mathcal{S}}). \tag{15}$$

Proof: By (13), we need to maximize $H_q(X_{\mathcal{U}(\mathfrak{V})})$ subject to the constraints that for each $\mathcal{S} \in \mathfrak{V}$ and all $x_{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}$, $q(x_{\mathcal{S}}) = p(x_{\mathcal{S}})$. This amounts to maximizing the Lagrangian

$$\Lambda = -\sum_{x_{\mathcal{U}(\mathfrak{V})}\in\mathcal{X}_{\mathcal{U}(\mathfrak{V})}} q(x_{\mathcal{U}(\mathfrak{V})}) \log q(x_{\mathcal{U}(\mathfrak{V})}) + \sum_{\mathcal{S}\in\mathfrak{V}} \sum_{x_{\mathcal{S}}\in\mathcal{X}_{\mathcal{S}}} \lambda_{\mathcal{S}}(x_{\mathcal{S}}) \big(q(x_{\mathcal{S}}) - p(x_{\mathcal{S}})\big).$$

Note that for any $\mathcal{S} \subset \mathcal{U}(\mathfrak{V})$

$$\frac{\partial q(x_{\mathcal{S}})}{\partial q(x_{\mathcal{U}(\mathfrak{V})})} = 1.$$

Thus, differentiating with respect to $q(x_{\mathcal{U}(\mathfrak{V})})$ gives, assuming the log is a natural logarithm

$$\frac{\partial\Lambda}{\partial q(x_{\mathcal{U}(\mathfrak{V})})} = -\log q(x_{\mathcal{U}(\mathfrak{V})}) - 1 + \sum_{\mathcal{S}\in\mathfrak{V}} \lambda_{\mathcal{S}}(x_{\mathcal{S}}).$$

Setting this to 0 gives

$$q(x_{\mathcal{U}(\mathfrak{V})}) = \exp\Big(-1 + \sum_{\mathcal{S}\in\mathfrak{V}} \lambda_{\mathcal{S}}(x_{\mathcal{S}})\Big) = |\mathcal{X}_{\mathcal{U}(\mathfrak{V})^c}| \prod_{\mathcal{S}\in\mathfrak{V}} \sigma_{\mathcal{S}}(x_{\mathcal{S}})$$

for some set of functions $\{\sigma_{\mathcal{S}}\}_{\mathcal{S}\in\mathfrak{V}}$. Therefore setting

$$q(x_1 \cdots x_m) = \frac{q(x_{\mathcal{U}(\mathfrak{V})})}{|\mathcal{X}_{\mathcal{U}(\mathcal{V})^c}|}$$

satisfies (15), so if $\sigma_{\mathcal{S}}$ are such that (13) is satisfied for all $\mathcal{S} \in \mathfrak{V}$, q will maximize $H_q(X_{\mathcal{U}(\mathfrak{V})})$. □

Suppose $m = 3$ and $\mathfrak{H} = \mathfrak{H}_1$. If $\mathfrak{V} = \{\{1,2\},\{2,3\}\}$, then $\tilde{q}(x_1x_2x_3) = p(x_1x_2)p(x_3|x_2)$ is in $\mathcal{Q}(\mathfrak{V})$ and by Lemma 1 maximizes $H_q(X_1X_2X_3)$ over all $q \in \mathcal{Q}(\mathfrak{V})$. Thus

$$\begin{aligned}\sup_{q\in\mathcal{Q}(\mathfrak{V})} H_q(X_1X_2X_3) &= H_{\tilde{q}}(X_1X_2X_3)\\ &= H(X_1X_2X_3) + I(X_1; X_3|X_2).\end{aligned}$$

By similar reasoning, considering $\mathfrak{V} = \{\{1,2\},\{1,3\}\}$ and $\mathfrak{V} = \{\{1,3\},\{2,3\}\}$ results in (3). Note that if $\mathfrak{V}_1 \subset \mathfrak{V}_2$, then $\mathcal{Q}(\mathfrak{V}_1) \supset \mathcal{Q}(\mathfrak{V}_2)$, so $\mathfrak{V}_2$ need not be considered in evaluating (8). Thus we have ignored larger subsets of $\mathfrak{H}_1$, since the value they give would be no greater than the others.

We can generalize to any collection $\mathfrak{V}$ of the form $\{\{\mathcal{S}_1,\mathcal{S}_2\},\{\mathcal{S}_1,\mathcal{S}_3\},\ldots,\{\mathcal{S}_1,\mathcal{S}_k\}\}$, in which case

$$\sup_{q\in\mathcal{Q}(\mathfrak{V})} = H(X_{\mathcal{S}_1}X_{\mathcal{S}_2}) + H(X_{\mathcal{S}_3}|X_{\mathcal{S}_1}) + \cdots + H(X_{\mathcal{S}_k}|X_{\mathcal{S}_1}).$$

Employing this, we can rewrite (12) for $\mathfrak{H} = \mathfrak{H}_t$ and certain values of t. For $t = 1$, it becomes

$$R^* = H(X_1 \cdots X_m) + \max_{i,i'\in\mathcal{M}} I(X_i; X_{i'}|X_{\{i,i'\}^c}).$$

Again, relative to the Slepian–Wolf result, we always pay a conditional mutual information penalty for a single traitor. For $t = 2$

$$\begin{aligned}R^* = {} & H(X_1 \cdots X_m)\\ & + \max\left\{\max_{\mathcal{S},\mathcal{S}'\subset\mathcal{M}:|\mathcal{S}|=|\mathcal{S}'|=2} I(X_{\mathcal{S}}; X_{\mathcal{S}'}|X_{(\mathcal{S}\cup\mathcal{S}')^c}),\right.\\ & \left.\max_{i,i',i''\in\mathcal{M}} I(X_i; X_{i'}; X_{i''}|X_{\{i,i',i''\}^c})\right\}\end{aligned}$$

where

$$\begin{aligned}I(X;Y;Z|W) = {} & H(X|W) + H(Y|W)\\ & + H(Z|W) - H(XYZ|W).\end{aligned}$$

For $t = m-1$, R^* is given by (2). There is a similar formulation for $t = m - 2$, though it is more difficult to write down for arbitrary m.

With all these expressions made up of nothing but entropies and mutual informations, it might seem hopeful that (14) can be reduced to such an analytic expression for all $\mathfrak{V}$. However, this is not the case. For example, consider $\mathfrak{V} = \{\{1,2,3\},\{3,4,5\},\{5,6,1\}\}$. This $\mathfrak{V}$ is irreducible in the sense that there is no subset $\mathfrak{V}'$ that still satisfies $\mathcal{U}(\mathfrak{V}') = \{1,\ldots,6\}$, but there is no simple distribution $q \in \mathcal{Q}(\mathfrak{V})$ made up of marginals of p that satisfies Lemma 1, so it must be found numerically. Still, Lemma 1 simplifies the calculation considerably.

V. Proof of Theorem 1

A. Converse

We first show the converse. Fix $\mathcal{H} \in \mathfrak{H}$ and $r \in \mathcal{R}(\mathcal{H})$. Take any $\mathfrak{V} \subset \mathfrak{H}$, and any distribution $q \in \mathcal{Q}_{\mathcal{H},r} \cap \mathcal{Q}(\mathfrak{V})$. Since $q \in \mathcal{Q}_{\mathcal{H},r}$, there is some $\bar{q}(x_{\mathcal{T}}|w)$ such that $X_{\mathcal{H}}$ and $X_{\mathcal{T}}$ are distributed according to q. Since also $q \in \mathcal{Q}_{\mathcal{S},r'}$ for all $\mathcal{S} \in \mathfrak{V}$

and some $r' \in \mathcal{R}(\mathcal{S})$, if the traitors simulate this $\bar{q}$ and act honestly with these fabricated source values, the decoder will not be able to determine which of the sets in $\mathfrak{V}$ is the actual set of honest sensors. Thus, the decoder must perfectly decode the sources from all sensors in $\mathcal{U}(\mathfrak{V})$, so if $R(\mathcal{H}, r)$ is an α-achievable rate function, $R(\mathcal{H}, r) \geq H_q(X_{\mathcal{U}(\mathfrak{V})})$.

B. Achievability Preliminaries

Now we prove achievability. To do so, we will first need the theory of types. Given $y^n \in \mathcal{Y}^n$, let $t(y^n)$ be the type of y^n. Given a type t with denominator n, let $\Lambda_t^n(Y)$ be the set of all sequences in $\mathcal{Y}^n$ with type t. If t is a joint y, z type with denominator n, then let $\Lambda_t^n(Y|z^n)$ be the set of sequences $y^n \in \mathcal{Y}^n$ such that $(y^n z^n)$ have joint type t, with the convention that this set is empty if the type of z^n is not the marginal of t.

We will also need the following definitions. Given a distribution q on an alphabet $\mathcal{Y}$, define the η-ball of distributions

$$B_\eta(q) \triangleq \left\{ q' : \forall x \in \mathcal{Y} : |q(x) - q'(x)| \leq \frac{\eta}{|\mathcal{Y}|} \right\}.$$

Note that the typical set can be written

$$T_\epsilon^n(X) = \{x^n : t(x^n) \in B_\epsilon(p)\}.$$

We define slightly modified versions of the sets of distributions from Section III-B as follows:

$$\breve{\mathcal{Q}}_{\mathcal{S},r'}^\eta \triangleq \bigcup_{q \in \mathcal{Q}_{\mathcal{S},r'}} B_\eta(q),$$
$$\breve{\mathcal{Q}}^\eta(\mathfrak{V}) \triangleq \bigcap_{\mathcal{S} \in \mathfrak{V}} \bigcup_{r' \in \mathcal{R}(\mathcal{S})} \breve{\mathcal{Q}}_{\mathcal{S},r'}^\eta.$$

These sets are nearly the same as those defined earlier. We will eventually take the limit as $\eta \to 0$, making them identical to $\mathcal{Q}_{\mathcal{S},r'}$ and $\mathcal{Q}(\mathfrak{V})$, but it will be necessary to have slightly expanded versions for use with finite block length.

Finally, we will need the following lemma.

Lemma 2: Given an arbitrary n length distribution $q^n(x^n)$ and a type t with denominator n on $\mathcal{X}$, let $q_i(x)$ be the marginal distribution of q^n at time i and $\bar{q}(x) = \frac{1}{n}\sum_{i=1}^n q_i(x)$. If X^n is distributed according to q^n and $\Pr(X^n \in \Lambda_t^n(X)) \geq 2^{-n\zeta}$, then $D(t\|\bar{q}) \leq \zeta$.

Proof: Fix an integer $\tilde{n}$. For $\tilde{i} = 1, \ldots, \tilde{n}$, let $X^n(\tilde{i})$ be independently generated from q^n. Let Γ be the set of types t^n on superletters in $\mathcal{X}^n$ with denominator $\tilde{n}$ such that $t^n(x^n) = 0$ if $x^n \notin \Lambda_t^n(X)$. Note that

$$|\Gamma| \leq (\tilde{n}+1)^{|\mathcal{X}|^n}.$$

If $X^{n\tilde{n}} = (X^n(1), \ldots, X^n(\tilde{n}))$, then

$$\begin{aligned}\Pr\left(X^{n\tilde{n}} \in \bigcup_{t^n \in \Gamma} \Lambda_{t^n}^{\tilde{n}}(X^n)\right) &= \Pr(X^n(\tilde{i}) \in \Lambda_t^n(X), \forall \tilde{i}) \\ &\geq 2^{-n\tilde{n}\zeta}.\end{aligned}$$

But

$$\begin{aligned}\Pr\Big(X^{n\tilde{n}} \in \bigcup_{t^n \in T^n} \Lambda_{t^n}^{\tilde{n}}(X^n)\Big) &= \sum_{t^n \in \Gamma} \Pr(X^{n\tilde{n}} \in \Lambda_{t^n}^{\tilde{n}}(X^n) \\ &\leq \sum_{t^n \in \Gamma} 2^{-\tilde{n}D(t^n\|q^n)} \\ &\leq (\tilde{n}+1)^{|\mathcal{X}|^n} 2^{-\tilde{n}\min_{t^n \in \Gamma} D(t^n\|q^n)}.\end{aligned}$$

For any $t^n \in \Gamma$, letting t_i be the marginal type at time i gives $\frac{1}{n}\sum_{i=1}^n t_i = t$. Therefore

$$\begin{aligned}\zeta + \frac{1}{n\tilde{n}}|\mathcal{X}|^n \log(\tilde{n}+1) &\geq \min_{t^n \in \Gamma} \frac{1}{n} D(t^n\|q^n) \\ &\geq \min_{t^n \in \Gamma} \frac{1}{n} \sum_{i=1}^n D(t_i\|q_i) \qquad (16) \\ &\geq D(t\|\bar{q}) \qquad (17)\end{aligned}$$

where (16) holds by [13, Lemma 4.3] and (17) by convexity of the Kullback–Leibler distance in both arguments. Letting $\tilde{n}$ grow proves the lemma. □

The achievability proof proceeds as follows. Section V-C describes our proposed coding scheme for the case that traitors cannot eavesdrop. In Section V-D, we demonstrate that this coding scheme achieves small probability of error when the traitors have perfect information. Section V-E shows that the coding scheme achieves the rate function $R^*(\mathcal{H}, r)$. In Section V-F, we extend the proof to include the case that the traitors have imperfect information. Finally, Section V-G gives a modification to the coding scheme that can handle eavesdropping traitors.

C. Coding Scheme Procedure

Our basic coding strategy is for a sensor to transmit a sequence of small messages to the decoder until the decoder has received enough information to decode the sensor's source sequence. After receiving one of these messages, the decoder asks for another small message only if it is unable to decode the sequence. If it can, the decoder moves on to the next sensor. This way, the rate at which a sensor transmits is as small as possible. Once each sensor's source sequence has been decoded, the decoder attempts to use them to accumulate information about which sensors could be traitors. It is in this step that it uses its knowledge of the power of the traitors to tell the difference between a sensor that could be honest under some circumstances and one that is surely a traitor. After this, the decoder goes back across all the sensors again, repeating the same procedure for the next block of source values and ignoring those sensors that it knows to be traitors. The decoder repeats this again and again, gathering more information about which sensors could be traitors each time. The precise description of the coding strategy follows.

1) Random Code Structure: Fix $\epsilon > 0$. The maximum number of small messages that could be sent by sensor i when transmitting a certain sequence to the decoder is $J_i = \left\lceil \frac{\log|\mathcal{X}_i|}{\epsilon} \right\rceil$. Each of these small messages is represented by a function to be defined, taking the source sequence as input and producing the small message as output. In addition, as we discussed in I-F, it is necessary to randomize the messages at run time in order to defeat the traitors. Thus, sensor i has C different but identically created subcodebooks, each of which is made up of a sequence of J_i functions, one for each small messages, where

C is an integer to be defined. Hence the full codebook for sensor i is composed of CJ_i separate functions. In particular, for $i = 1, \dots, m$ and $c = 1, \dots, C$, let

$$\begin{aligned} \tilde{f}_{i,c,1} &: \mathcal{X}_i^n \to \{1, \dots, 2^{n(\epsilon+\nu)}\} \\ \tilde{f}_{i,c,j} &: \mathcal{X}_i^n \to \{1, \dots, 2^{n\epsilon}\}, \quad j = 2, \dots, J_i \end{aligned}$$

with ν to be defined later. Thus, a subcodebook associates with each element of $\mathcal{X}_i^n$ a sequence of about $n(\log|\mathcal{X}_i| + \nu)$ bits chopped into small messages of length $n(\epsilon + \nu)$ or $n\epsilon$. We put tildes on these functions to distinguish them from the fs defined in (6). The $\tilde{f}$s that we define here are functions we use as pieces of the overall encoding functions f. Each one is constructed by a uniform random binning procedure. Define composite functions

$$\tilde{F}_{i,c,j}(x_i^n) \triangleq (\tilde{f}_{i,c,1}(x_i^n), \dots, \tilde{f}_{i,c,j}(x_i^n)).$$

We can think of $\tilde{F}_{i,c,j}(x_i^n)$ as an index of one of $2^{n(j\epsilon+\nu)}$ random bins.

2) Round Method: Our coding scheme is made up of N rounds, with each round composed of m phases. In the ith phase, transactions are made entirely with sensor i. We denote $X_i^n(I)$ as the Ith block of n source values, but for convenience, we will not include the index I when it is clear from context. As in the three-sensor example, all transactions in the Ith round are based only on $X_{\mathcal{M}}^n(I)$. Thus the total block length is Nn.

The procedure for each round is identical except for the variable $\mathfrak{V}(I)$ maintained by the decoder. This represents the collection of sets that could be the set of honest sensors based on the information the decoder has received as of the beginning of round I. The decoder begins by setting $\mathcal{V}(1) = \mathfrak{H}$ and then pares it down at the end of each round based on new information.

3) Encoding and Decoding Rules: In the ith phase, if $i \in \mathcal{U}(\mathfrak{V}(I))$, the decoder makes a number of transactions with sensor i and produces an estimate $\hat{X}_i^n$ of X_i^n. If $i \notin \mathcal{U}(\mathfrak{V}(I))$, then the decoder has determined that sensor i cannot be honest, so it does not communicate with it and sets $\hat{X}_i^n$ to a null value.

For $i \in \mathcal{U}(\mathfrak{V}(I))$, at the beginning of phase i, sensor i randomly selects a $c \in \{1, \dots, C\}$ according to the uniform distribution. In the first transaction, sensor i transmits $(c, \tilde{f}_{i,c,1}(X_i^n))$. That is, along with the small message itself, the sensor transmits the randomly selected index c of the subcodebook that it will use in this phase. As the phase continues, in the jth transaction, sensor i transmits $\tilde{f}_{i,c,j}(X_i^n)$.

After each transaction, the decoder must decide whether to ask for another transaction with sensor i, and if not, to decode X_i^n. In the random binning proof of the traditional Slepian–Wolf problem, the decoder decides which sequence in the received bin to select as the source estimate by taking the one contained in the typical set. Here we use the same idea, except that instead of the typical set, we use a different set for each transaction, and if there is no sequence in this set that falls into the received bin, this means not that we cannot decode the sequence but rather that we have not yet received enough information from the sensor and must ask for another transaction. The set associated with the jth transaction needs to have the property that its size is less than $2^{n(j\epsilon+\nu)}$, the number of bins into which the source space has been split after j messages, so that it is unlikely for two elements of the set to fall into the same bin. Furthermore, in order to ensure that we eventually decode any sequence that might be chosen by the sensor, the set should grow after each transaction and eventually contain all of $\mathcal{X}_i^n$.

Now we define this set. First let $\mathcal{S}_i \triangleq \{1, \dots, i\} \cap \mathcal{U}(\mathfrak{V}(I))$, the sensors up to i that are not ignored by the decoder, and let $\hat{x}_{\mathcal{S}_{i-1}}^n$ be the source sequences decoded in this round prior to phase i. The set associated with transaction j is

$$T_j(\hat{x}_{\mathcal{S}_{i-1}}^n) \triangleq \{x_i^n : H_{t(\hat{x}_{\mathcal{S}_{i-1}}^n x_i^n)}(X_i | X_{\mathcal{S}_{i-1}}) \leq j\epsilon\}. \tag{18}$$

To be specific, after j transactions, if there are no sequences in $T_j(\hat{x}_{\mathcal{S}_{i-1}})$ matching the received value of $\tilde{F}_{i,c,j}$, the decoder chooses to do another transaction with sensor i. If there is at least one such sequence, choose one to be $\hat{X}_i^n$, deciding between several possibilities arbitrarily.

Observe that

$$|T_j(\hat{x}_{\mathcal{S}_{i-1}}^n)| \leq (n+1)^{|\mathcal{X}_i \times \mathcal{X}_{\mathcal{S}_{i-1}}|} 2^{nj\epsilon}.$$

Hence T_j satisfies the size property that were discussed above. Moreover, it grows with j to eventually become $\mathcal{X}_i^n$. Finally, we have chosen T_j in particular because it has the property that when a sequence x_i^n falls into T_j for the first time, the rate at which sensor i has transmitted to the decoder is close to the entropy of the type of x_i^n. This means that we can relate the accuracy of the decoded sequences to the achieved rate, which will allow us to prove that the coding scheme achieves the claimed rate.

4) Round Conclusion: At the end of round I, the decoder produces $\mathfrak{V}(I+1)$ by setting

$$\mathfrak{V}(I+1) = \left\{ \mathcal{S} \in \mathfrak{V}(I) : t(\hat{x}_{\mathcal{U}(\mathcal{V}(I))}^n) \in \bigcup_{r' \in R(\mathcal{S})} \breve{\mathcal{Q}}_{\mathcal{S},r'}^{\eta} \right\} \tag{19}$$

for η to be defined such that $\eta \geq \epsilon$ and $\eta \to 0$ as $\epsilon \to 0$. As we will show, it is essentially impossible for the traitors to transmit messages such that the type of the decoded messages does not fall into $\breve{\mathcal{Q}}_{\mathcal{H},r}^{\eta}$, meaning that $\mathcal{H}$ is always in $\mathfrak{V}(I)$. This ensures that the true honest sensors are never ignored and their source sequences are always decoded correctly.

D. Error Probability

Define the following error events:

$$\begin{aligned} \mathcal{E}_1(I,i) &\triangleq \{\hat{X}_i^n(I) \neq X_i^n(I)\}, \\ \mathcal{E}_2(I) &\triangleq \{\mathcal{H} \notin \mathfrak{V}(I)\} \\ \mathcal{E}_3(I) &\triangleq \{t(\hat{X}_{\mathcal{U}(\mathfrak{V}(I))}^n(I)) \notin \breve{\mathcal{Q}}_{\mathcal{H},r}^{\eta}\}. \end{aligned}$$

The total probability of error is

$$P_e = \Pr\left(\bigcup_{I=1}^{N} \bigcup_{i \in \mathcal{H}} \mathcal{E}_1(I,i) \right).$$

As we have said but not yet proved, $\mathcal{H}$ will usually be in $\mathfrak{V}(I)$ (i.e., $\mathcal{E}_2(I)$ does not occur), so we do not lose much by writing

$$P_e \leq \Pr\left(\bigcup_{I=1}^{N} \left[\mathcal{E}_2(I+1) \cup \bigcup_{i \in \mathcal{H}} \mathcal{E}_1(I,i) \right] \right).$$

Let

$$\mathcal{A}_I \triangleq \mathcal{E}_2^c(I+1) \cap \bigcap_{i\in\mathcal{H}} E_1^c(I,i)$$

for $I = 1,\dots,N$, so

$$1 - P_e \ge \Pr(\mathcal{A}_1,\dots,\mathcal{A}_N) = \prod_{I=1}^{N} \Pr(\mathcal{A}_I|\mathcal{A}_1,\dots,\mathcal{A}_{I-1}).$$

Observe that $\mathcal{A}_I$ depends only on $\hat{X}^n_{\mathcal{M}}(I)$ and $X^n_{\mathcal{M}}(I)$, both of which are independent of all events before round I given that $\mathcal{H} \in \mathfrak{V}(I)$ (i.e., $\mathcal{E}_2^c(I)$ occurs), since this is enough to ensure that $\hat{X}_i^n(I)$ is non-null. Since $\mathcal{A}_1,\dots,\mathcal{A}_{I-1}$ includes $\mathcal{E}_2^c(I)$, we can drop all conditioning terms expect it. Note also that $\mathcal{E}_2^c(1)$ occurs with probability 1. Therefore

$$\begin{aligned}1 - P_e &\ge \prod_{I=1}^{n} \Pr(\mathcal{A}_I|\mathcal{E}_2^c(I))\\ &= \prod_{I=1}^{n}[1-\Pr(\mathcal{A}_I^c|\mathcal{E}_2^c(I))] \ge 1 - \sum_{I=1}^{n}\Pr(\mathcal{A}_I^c|\mathcal{E}_2^c(I))\end{aligned}$$

so

$$P_e \le \sum_{I=1}^{N} \Pr\left(\mathcal{E}_2(I+1) \cup \bigcup_{i\in\mathcal{H}} \mathcal{E}_1(I,i) \middle| \mathcal{E}_2^c(I)\right).$$

By (19), if $\mathcal{H}$ is in $\mathfrak{V}(I)$ but not in $\mathfrak{V}(I+1)$, then $t(\hat{X}^n_{\mathcal{U}(\mathfrak{V}(I))}(I)) \notin \breve{\mathcal{Q}}^\eta_{\mathcal{H},r}$. Thus

$$\mathcal{E}_2(I+1) \cap \mathcal{E}_2^c(I) \subset \mathcal{E}_3(I) \cap \mathcal{E}_2^c(I)$$

so

$$\begin{aligned}P_e &\le \sum_{I=1}^{N} \Pr\left(\mathcal{E}_3(I) \cup \bigcup_{i\in\mathcal{H}} \mathcal{E}_1(I,i) \middle| \mathcal{E}_2^c(I)\right)\\ &\le \sum_{I=1}^{N} \Pr\left(\mathcal{E}_3(I) \middle| \mathcal{E}_2^c(I), \bigcap_{i\in\mathcal{H}} \mathcal{E}_1^c(I,i)\right)\\ &\quad + \sum_{I=1}^{N} \Pr\left(\bigcup_{i\in\mathcal{H}} \mathcal{E}_1(I,i) \middle| \mathcal{E}_2^c(I)\right)\\ &\le \sum_{I=1}^{N} \Pr\left(\mathcal{E}_3(I) \middle| \bigcap_{i\in\mathcal{H}} \mathcal{E}_1^c(I,i)\right)\\ &\quad + \sum_{I=1}^{N}\sum_{i\in\mathcal{H}} \Pr(\mathcal{E}_1(I,i)|\mathcal{E}_2^c(I)) \qquad (20)\end{aligned}$$

where we have dropped the conditioning on $\mathcal{E}_2^c(I)$ in the first term because it influences the probability of $\mathcal{E}_3(I)$ only in that it ensures that $\hat{X}_i^n$ for $i \in \mathcal{H}$ are non-null, which is already implied by $\bigcap_{i\in\mathcal{H}} \mathcal{E}_1^c(I,i)$.

We first bound the first term in (20) by showing that for all I,

$$\Pr\left(\mathcal{E}_3(I) \middle| \bigcap_{i\in\mathcal{H}} \mathcal{E}_1^c(I,i)\right) \le \frac{\alpha}{2N}. \qquad (21)$$

If the traitors receive perfect source information, then as we have already noted in (13), $\mathcal{Q}_{\mathcal{H},r}$ only puts a constraint on the $X_{\mathcal{H}}$ marginal of distributions, and the same is true of $\breve{\mathcal{Q}}^\eta_{\mathcal{H},r}$. In particular, $t(\hat{X}^n_{\mathcal{U}(\mathfrak{V}(I))}(I)) \in \breve{\mathcal{Q}}^\eta_{\mathcal{H},r}$ is equivalent to $\hat{X}^n_{\mathcal{H}}(I)$ being typical. Conditioning on $\bigcap_{i\in\mathcal{H}} \mathcal{E}_1^c(I,i)$ implies that $\hat{X}^n_{\mathcal{H}}(I) = X^n_{\mathcal{H}}(I)$, so

$$\Pr\left(\mathcal{E}_3(I) \middle| \bigcap_{i\in\mathcal{H}} \mathcal{E}_1^c(I,i)\right) \le \Pr(X^n_{\mathcal{H}}(I) \in T^n_\epsilon(X_{\mathcal{H}}))$$

meaning (21) holds for sufficiently large n by the AEP. Thus (21) is only nontrivial if the traitors receive imperfect source information. This case is dealt with in Section V-F.

We now consider the second term of (20), involving $\Pr(\mathcal{E}_1(I,i)|\mathcal{E}_2^c(I))$ for honest i. Conditioning on $\mathcal{E}_2^c(I)$ ensures that $i \in \mathcal{U}(\mathfrak{V}(I))$ for honest i, so $\hat{X}_i^n(I)$ will be non-null. The only remaining type of is a decoding error. This occurs if for some transaction j, there is an sequence in $T_j(\hat{X}_{\mathcal{S}_{i-1}})$ different from X_i^n that matches all thus far received messages. That is, if

$$\exists j, x_i'^n \in T_j(\hat{X}^n_{\mathcal{S}_{i-1}})\backslash\{X_i^n\} : \tilde{F}_{i,c,j}(x_i'^n) = \tilde{F}_{i,c,j}(X_i^n).$$

However, $\mathcal{S}_{i-1}$ may contain traitors. Indeed, it may be made entirely of traitors. Thus, we have to take into account that $\hat{X}^n_{\mathcal{S}_{i-1}}$ may be chosen to ensure the existence of such an erroneous $x_i'^n$. The sensor's use of randomizing among the C subcodebooks is the method by which this is mitigated, as we will now prove.

Let

$$\begin{aligned}k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}}) \triangleq |\{c : \exists j, x_i'^n \in T_j(\hat{x}^n_{\mathcal{S}_{i-1}})\backslash\{x_i^n\} :\\ \tilde{F}_{i,c,j}(x_i'^n) = \tilde{F}_{i,c,j}(x_i^n)\}|.\end{aligned}$$

That is, k_1 is the number of subcodebooks that if chosen could cause a decoding error at some transaction. Recall that sensor i chooses the subcodebook randomly from the uniform distribution. Thus, given x_i^n and $\hat{x}^n_{\mathcal{S}_{i-1}}$, the probability of an error resulting from a bad choice of subcodebook is $k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}})/C$. Furthermore, k_1 is based strictly on the codebook, so we can think of it as a random variable defined on the same probability space as that governing the random codebook creation. Averaging over all possible codebooks

$$\Pr(\mathcal{E}_1(I,i)|\mathcal{E}_2^c(I)) \le \mathbb{E} \sum_{x_i^n \in \mathcal{X}_i^n} p(x_i^n) \max_{\hat{x}^n_{\mathcal{S}_{i-1}} \in \mathcal{X}^n_{\mathcal{S}_{i-1}}} \frac{k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}})}{C}$$

where the expectation is taken over codebooks.

Let $\mathcal{C}$ be the set of all codebooks. We define a subset $\mathcal{C}_1$, then show that the probability of error can be easily bounded for any codebook in $\mathcal{C}\backslash\mathcal{C}_1$, and that the probability of a codebook being chosen in $\mathcal{C}_1$ is small. In particular, let $\mathcal{C}_1$ be the set of codebooks for which, for any $x_i^n \in \mathcal{X}_i^n$ and $\hat{x}^n_{\mathcal{S}_{i-1}} \in \mathcal{X}^n_{\mathcal{S}_{i-1}}$, $k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}}) > B$, for an integer $B \le C$ to be defined later. Then

$$\begin{aligned}\Pr(\mathcal{E}_1(I,i)|\mathcal{E}_2^c(I)) &\le \Pr(\mathcal{C}\backslash\mathcal{C}_1) \sum_{x_i^n \in \mathcal{X}_i^n} p(x_i^n) \max_{\hat{x}^n_{\mathcal{S}_{i-1}} \in \mathcal{X}^n_{\mathcal{S}_{i-1}}} \frac{B}{C}\\ &\quad + \Pr(\mathcal{C}_1) \sum_{x_i^n \in \mathcal{X}_i^n} p(x_i^n) \max_{\hat{x}^n_{\mathcal{S}_{i-1}} \in \mathcal{X}^n_{\mathcal{S}_{i-1}}} \frac{C}{C}\\ &\le \frac{B}{C} + \Pr(\mathcal{C}_1). \qquad (22)\end{aligned}$$

Recall that k_1 is the number of subcodebooks that could cause an error. Since each subcodebook is generated identically, k_1

is a binomial random variable with C trails and probability of success P, where P is the probability that one particular subcodebooks causes an error. Thus

$$\begin{aligned}P &= \Pr\left(\exists j, x_i'^n \in T_j(\hat{x}^n_{\mathcal{S}_{i-1}})\backslash\{x_i^n\} :\right.\\ &\qquad \left.\tilde{F}_{i,c,j}(x_i'^n) = \tilde{F}_{i,c,j}(x_i^n)\right)\\ &\le \sum_{j=1}^{J_i} \sum_{x_i'^n \in T_j(\hat{x}^n_{\mathcal{S}_{i-1}})\backslash\{x_i^n\}} \Pr\left(\tilde{F}_{i,c,j}(x_i'^n) = \tilde{F}_{i,c,j}(x_i^n)\right)\\ &\le J_i \left|T_j(\hat{x}^n_{\mathcal{S}_{i-1}})\right| 2^{-n(j\epsilon+\nu)}\\ &\le J_i(n+1)^{|\mathcal{X}_i \times \mathcal{X}_{\mathcal{S}_{i-1}}|} 2^{-n\nu} \le 2^{n(\epsilon-\nu)}\end{aligned}$$

for sufficiently large n. For a binomial random variable X with mean $\bar{X}$ and any κ, we can use the Chernoff bound to write

$$\Pr(X \ge \kappa) \le \left(\frac{e\bar{X}}{\kappa}\right)^{\kappa}. \tag{23}$$

Therefore

$$\Pr(k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}}) > B) \le \left(\frac{eCP}{B+1}\right)^{B+1} \le 2^{nB(\epsilon-\nu)}$$

if $\nu > \epsilon$ and n is sufficiently large. Thus

$$\begin{aligned}\Pr(\mathcal{C}_1) &= \Pr(\exists x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}} : k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}}) > B)\\ &\le \sum_{x_i^n} \sum_{\hat{x}^n_{s_{i-1}}} \Pr(k_1(x_i^n, \hat{x}^n_{\mathcal{S}_{i-1}}) > B)\\ &\le \sum_{x_i^n} \sum_{\hat{x}^n_{\mathcal{S}_{i-1}}} 2^{nB(\epsilon-\nu)}\\ &= 2^{n[\log|\mathcal{X}_i| + \log|\mathcal{X}_{\mathcal{S}_{i-1}}| + B(\epsilon-\nu)]}.\end{aligned} \tag{24}$$

Combining (20) with (21), (22), and (24) gives

$$\begin{aligned}P_e &\le \frac{\alpha}{2} + \sum_{I=1}^{N} \sum_{i \in \mathcal{H}} \left(\frac{B}{C} + 2^{n[\log|\mathcal{X}_i| + \log|\mathcal{X}_{\mathcal{S}_{i-1}}| + B(\epsilon-\nu)]}\right)\\ &\le \frac{\alpha}{2} + Nm\left(\frac{B}{C} + 2^{n[\log|\mathcal{X}_{\mathcal{M}}| + B(\epsilon-\nu)]}\right)\end{aligned}$$

which is less than α for sufficiently large n if

$$B > \frac{\log|\mathcal{X}_{\mathcal{M}}|}{\nu - \epsilon}$$

and

$$C \ge \frac{3NmB}{\alpha} > \frac{3Nm\log|\mathcal{X}_{\mathcal{M}}|}{\alpha(\nu-\epsilon)}.$$

E. Code Rate

The discussion above placed a lower bound on C. However, for sufficiently large n, we can make $\frac{1}{n}\log C \le \epsilon$, meaning it takes no more than ϵ rate to transmit the subcodebook index c at the beginning of the phase. Therefore the rate for phase i is at most $(j+1)\epsilon + \nu$, where j is the number of transactions in phase i. Transaction j must be the earliest one with $\hat{x}_i^n \in T_j(\hat{x}_{\mathcal{S}_{i-1}})$, otherwise it would have been decoded earlier. Thus j is the smallest integer for which

$$H_{t(\hat{x}^n_{\mathcal{S}_{i-1}}\hat{x}^n_i)}(X_i|X_{\mathcal{S}_{i-1}}) \le j\epsilon$$

meaning

$$j\epsilon \le H_{t(\hat{x}^n_{\mathcal{S}_{i-1}}\hat{x}^n_i)}(X_i|X_{\mathcal{S}_{i-1}}) + \epsilon. \tag{25}$$

By (19), for all $\mathcal{S} \in \mathfrak{V}(I+1)$, $t(\hat{x}^n_{\mathcal{U}(\mathfrak{V}(I))}) \in \bigcup_{r' \in \mathcal{R}(\mathcal{S})} \breve{\mathcal{Q}}^{\eta}_{\mathcal{S},r'}$, meaning

$$t(\hat{x}_{\mathcal{U}(\mathfrak{V}(I))}) \in \bigcap_{\mathcal{S} \in \mathfrak{V}(I+1)} \bigcup_{r' \in \mathcal{R}(\mathcal{S})} \breve{\mathcal{Q}}^{\eta}_{\mathcal{S},r'} = \breve{\mathcal{Q}}^{\eta}(\mathfrak{V}(I+1)).$$

Furthermore, from (21) we know that with probability at least $1-\alpha$, $t(\hat{x}_{\mathcal{U}(\mathfrak{V}(I))}) \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r}$. Therefore

$$t(\hat{x}_{\mathcal{U}(\mathfrak{V}(I))}) \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} \cap \breve{\mathcal{Q}}^{\eta}(\mathfrak{V}(I+1)). \tag{26}$$

Combining (25) with (26) gives that with high probability, the rate for all of round I is at most

$$\begin{aligned}&\sum_{i \in \mathcal{U}(\mathfrak{V}(I))} \left[H_{t(\hat{x}^n_{\mathcal{S}_{i-1}}\hat{x}^n_i)}(X_i|X_{\mathcal{S}_{i-1}}) + 2\epsilon + \nu\right]\\ &\quad\le H_{t(\hat{x}_{\mathcal{U}(\mathfrak{V}(I))})}\left(X_{\mathcal{U}(\mathfrak{V})}\right) + m(2\epsilon+\nu)\\ &\quad\le \sup_{q \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} \cap \breve{\mathcal{Q}}^{\eta}(\mathfrak{V}(I+1))} H_q\left(X_{\mathcal{U}(\mathfrak{V})}\right) + m(2\epsilon+\nu)\\ &\quad\le \sup_{q \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} \cap \breve{\mathcal{Q}}^{\eta}(\mathfrak{V}(I+1))} H_q\left(X_{\mathcal{U}(\mathfrak{V}(I+1))}\right)\\ &\qquad + \sup_q H_q(X_{\mathcal{U}(\mathfrak{V}(I))\backslash\mathcal{U}(\mathfrak{V}(I+1))}) + m(2\epsilon+\nu)\\ &\quad\le \sup_{\mathfrak{V} \subset \mathfrak{H},\ q \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} \cap \breve{\mathcal{Q}}^{\eta}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})})\\ &\qquad + \log\left|\mathcal{X}_{\mathcal{U}(\mathfrak{V}(I))\backslash\mathcal{U}(\mathcal{V}(I+1))}\right| + m(2\epsilon+\nu).\end{aligned} \tag{27}$$

Whenever $\mathcal{U}(\mathfrak{V}(I))\backslash\mathcal{U}(\mathfrak{V}(I+1)) \neq \emptyset$, at least one sensor is eliminated. Therefore the second term in (27) will be nonzero in all but at most m rounds. Moreover, although we have needed to bound ν from below, we can still choose it such that $\nu \to 0$ as $\epsilon \to 0$. Thus if N is large enough, the rate averaged over all rounds is no more than

$$R_\epsilon(\mathcal{H}, r) \triangleq \sup_{\mathfrak{V} \subset \mathfrak{H},\ q \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} \cap \breve{\mathcal{Q}}^{\eta}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})}) + \dot{\epsilon}$$

where $\dot{\epsilon} \to 0$ as $\epsilon \to 0$. This is an α-achievable rate function. By continuity of entropy,

$$\lim_{\epsilon \to 0} R_\epsilon(\mathcal{H}, r) = \sup_{\mathfrak{V} \subset \mathfrak{H},\ q \in \mathcal{Q}_{\mathcal{H},r} \cap \mathcal{Q}(\mathfrak{V})} H_q(X_{\mathcal{U}(\mathfrak{V})}) = R^*(\mathcal{H}, r)$$

so $R^*(\mathcal{H}, r)$ is achievable.

F. Imperfect Traitor Information

We now consider the case that the traitors have access to imperfect information about the sources. The additional required piece of analysis is to prove (21). That is

$$\Pr(t(\hat{Y}^n\hat{Z}^n) \notin \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} | \hat{Y}^n = Y^n) \le \frac{\alpha}{2N} \tag{28}$$

where we define for notational convenience $Y \triangleq X_{\mathcal{H}}(I)$ and $Z \triangleq X_{\mathcal{T} \cap \mathcal{U}(\mathfrak{V}(I))}(I)$. Observe that we can drop the hat from $\hat{Y}^n$ if we wish because of the conditioning term.

To help explain the task in proving (28), we present a similar argument to the one we used in Section III-B to interpret Theorem 1: we impose a constraint on the traitors, then demonstrate that (28) would be easy to prove under this constraint. Suppose that, given W^n, the traitors apply a function $h : \mathcal{W}^n \to \mathcal{Z}^n$ to get the sequence $\tilde{Z}^n = h(W^n)$, then report this $\tilde{Z}^n$ as the truth. Assuming the decoder successfully decodes $\hat{Z}^n$ so that $\hat{Z}^n = \tilde{Z}^n$, Y^n and $\hat{Z}^n$ would be distributed according to

$$q^n(y^n z^n) = \sum_{w^n} \left[\prod_{\tau=1}^{n} p(y_\tau) r(w_\tau | y_\tau) \right] \mathbf{1}\{z^n = h(w^n)\}.$$

By Lemma 2, the only Y, Z types t that could be generated from this distribution with substantial probability are those for which t is close to $\bar{q}(yz)$. Furthermore, we can write

$$\bar{q}(yz) = p(y) \sum_{w} r(w|y) \bar{q}(z|w)$$

for some $\bar{q}(z|w)$. Thus, $\bar{q}(yz) \in \mathcal{Q}_{\mathcal{H},r}$ by (7), so $t \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r}$ for some small η. This would prove (28).

However, we cannot place any such limitations on the traitors' behavior. Our goal will be to show that for any action, there exists a function h such that the behavior just described produces nearly the same effect. Observe that a transmission made by the traitors is equivalent to a bin, or subset, of $\mathcal{Z}^n$: the set of all sequences that would produce this transmission if the sensors were honest. The decoder will choose an element of this bin as $\hat{Z}^n$, making its decision by selecting one that agrees with Y^n (specifically, by always taking elements in $T_{\mathbf{j}}$). Because the traitors do not know Y^n exactly, they must select their transmitted bin so that for every likely y^n, the bin contains some sequence agreeing with it. That is, each element of the bin agrees with a certain set of y^ns, and the union of all these sets must contain all likely values of y^n given W^n. We will show that the distribution of the sizes of these "agreement sets" is highly nonuniform. That is, even though no single element of the bin agrees with all likely y^n, a small number of elements of the bin agree with many more y^ns that the others. Therefore, transmitting this bin is not much different from choosing one of these "special" elements and reporting it as the truth.

The manner in which the traitors choose a bin based on W^n is complicated by two factors. First, they must choose a subcodebook index c to use for each traitor in $\mathcal{U}(\mathfrak{V}(I))$ before transmitting any information. Second, the exact rate at which each traitor transmits depends on the number of small messages that it takes for the decoder to construct a source estimate, which the traitors will not always know *a priori*. Let $\mathbf{j} \triangleq \{j_i\}_{i \in \mathcal{T} \cap \mathcal{U}(\mathfrak{V}(I))}$ be the vector representing the number of transactions (small messages) that take place with each traitor in $\mathcal{U}(\mathfrak{V}(I))$. There are $J_{\mathcal{T}} \triangleq \prod_{i \in \mathcal{T} \cap \mathcal{U}(\mathfrak{V}(I))} J_i$ different possible values of $\mathbf{j}$. For a given $\mathbf{j}$, each set of messages sent with this number of transactions is represented by a bin. Let $\mathcal{B}_{\mathbf{j}}$ be the set of these bins. Note that we include all choices of subcodebook indices in this set; there are many different binnings for a given $\mathbf{j}$, any of which the traitors may select. Now the traitors' behavior is completely described by a group of potentially random functions $g_{\mathbf{j}} : \mathcal{W}^n \to \mathcal{B}_{\mathbf{j}}$ for all $\mathbf{j}$. That is, if the traitors receive W^n, and the numbers of transactions are given by $\mathbf{j}$, then their transmitted bin is $g_{\mathbf{j}}(W^n)$. Note that when we refer to a bin, we mean not the index of the bin but the actual set of sequences in that bin. Thus $g_{\mathbf{j}}(W^n)$ is a subset of $\mathcal{Z}^n$.

Consider a joint y, z type t. We are interested in the circumstances under which $(Y^n \hat{Z}^n)$ has type t. Recall that in a given phase, the value of j determines what source sequences can be decoded without receiving additional messages from the sensor. In particular, only those sequences in T_j can be decoded. Thus, in order to decode $\hat{Z}^n$ such that $(Y^n \hat{Z}^n)$ has type t, $\mathbf{j}$ must be such that in every phase, sequences of the proper type fall into T_{j_i}. Specifically, by (18), we need for every i

$$H_t(X_i | X_{\mathcal{S}_{i-1}}) \leq j_i \epsilon.$$

Hence

$$\sum_{i \in \mathcal{T} \cap \mathcal{U}(\mathfrak{V}(I))} j_i \epsilon \geq H_t(Z|Y).$$

Let $R(\mathbf{j})$ be the total rate transmitted by all the traitors in $\mathcal{U}(\mathfrak{V}(I))$ given $\mathbf{j}$. The transmitted rate by sensor i is $j_i \epsilon + \nu$, so

$$R(\mathbf{j}) = \sum_{i \in \mathcal{T} \cap \mathcal{U}(\mathfrak{V}(I))} [j_i \epsilon + \nu] \geq H_t(Z|Y) + \nu.$$

Therefore if $(Y^n \hat{Z}^n) \in \Lambda^n_t(YZ)$, then there exists a $\mathbf{j}$ such that $R(\mathbf{j}) \geq H_t(Z|Y) + \nu$ and $g_{\mathbf{j}}(W^n) \cap \Lambda^n_t(Z|Y^n)$ is not empty. Let $\delta \triangleq \frac{\epsilon}{4N}$

$$\delta_{t,\mathbf{j}} \triangleq \Pr((Y^n W^n) \in T^n_\epsilon(YW), g_{\mathbf{j}}(W^n) \cap \Lambda^n_t(Z|Y^n) \neq \emptyset)$$

and

$$\mathcal{P} \triangleq \left\{ t : \max_{\mathbf{j}: R(\mathbf{j}) \geq H_t(Z|Y) + \nu} \delta_{t,\mathbf{j}} \geq \frac{\delta}{(n+1)^{|\mathcal{Y} \times \mathcal{Z}|} J_{\mathcal{T}}} \right\}.$$

We will show that $\mathcal{P} \subset \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r}$, so that

$$\begin{aligned}
&\Pr(t(Y^n \hat{Z}^n) \notin \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r} | \mathcal{H} \in \mathfrak{V}(I)) \\
&\quad \leq \Pr(t(Y^n \hat{Z}^n) \notin \mathcal{P} | \mathcal{H} \in \mathfrak{V}(I)) \\
&\quad \leq \Pr((Y^n W^n) \notin T^n_\epsilon(YW)) \\
&\qquad + \sum_{t \in \mathcal{P}^c} \Pr((Y^n W^n) \in T^n_\epsilon(YW), (Y^n \hat{Z}^n) \in \Lambda^n_t(YZ)) \\
&\quad \leq \delta + \sum_{t \in \mathcal{P}^c} \Pr((Y^n W^n) \in T^n_\epsilon(YW), \exists \mathbf{j} : \\
&\qquad R(\mathbf{j}) \geq H_t(Z|Y) + \nu, g_{\mathbf{j}}(W^n) \cap \Lambda^n_t(Z|Y^n) \neq \emptyset) \\
&\quad \leq \delta + \sum_{t \in \mathcal{P}^c} \sum_{\mathbf{j}: R(\mathbf{j}) \geq H_t(Z|Y) + \nu} \delta_{t,\mathbf{j}} \\
&\quad \leq \delta + (n+1)^{|\mathcal{Y} \times \mathcal{Z}|} J_{\mathcal{T}} \frac{\delta}{(n+1)^{|\mathcal{Y} \times \mathcal{Z}|} J_{\mathcal{T}}} = 2\delta = \frac{\alpha}{2N}
\end{aligned}$$

for sufficiently large n.

Fix $t \in \mathcal{P}$. We show that $t \in \breve{\mathcal{Q}}^{\eta}_{\mathcal{H},r}$. There is some $\mathbf{j}$ with

$$R(\mathbf{j}) \geq H_t(Z|Y) + \nu \tag{29}$$

and $\delta_{t,\mathbf{j}} \geq \frac{\delta}{(n+1)^{|\mathcal{Y} \times \mathcal{Z}|} J_{\mathcal{T}}}$. Any random $g_{\mathbf{j}}$ is a probabilistic combination of a number of deterministic functions, so if this lower bound on $\delta_{t,\mathbf{j}}$ holds for a random $g_{\mathbf{j}}$, it must also hold for some deterministic $g_{\mathbf{j}}$. Therefore we do not lose generality to assume from now on that $g_{\mathbf{j}}$ is deterministic. We also drop the $\mathbf{j}$ subscript for convenience.

Define the following sets:

$$\begin{aligned} &A_\epsilon^n(Y|w^n)\\ &\triangleq \{y^n \in T_\epsilon^n(Y|w^n):\\ &\quad g(w^n) \cap \Lambda_t^n(Z|y^n) \neq \emptyset\}\\ &A_\epsilon^n(W)\\ &\triangleq \Big\{w^n \in T_\epsilon^n(W):\\ &\quad \Pr\left(Y^n \in A_\epsilon^n(Y|w^n)|W^n=w^n\right) \geq \frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}J_{\mathcal{T}}}\Big\}. \end{aligned}$$

Applying the definitions of $\mathcal{P}$ and $\delta_{t,\mathbf{j}}$ gives

$$\begin{aligned} &\frac{\delta}{(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}J_{\mathcal{T}}}\\ &\leq \Pr((Y^nW^n) \in T_\epsilon^n(YW): g(W^n)\cap\Lambda_t^n(Z|Y^n) \neq \emptyset)\\ &= \sum_{w^n \in T_\epsilon^n(W)} p(w^n)\Pr\left(Y^n \in A_\epsilon^n(Y|w^n)|W^n=w^n\right)\\ &\leq \Pr(W^n \in A_\epsilon^n(W)) + \frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}J_{\mathcal{T}}} \end{aligned}$$

meaning $\Pr(W^n \in A_\epsilon^n(W)) \geq \frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}J_{\mathcal{T}}}$. Fix $w^n \in A_\epsilon^n(W)$. Since $A_\epsilon^n(Y|w^n) \subset T_\epsilon^n(Y|w^n)$

$$|A_\epsilon^n(Y|w^n)| \geq \frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}J_{\mathcal{T}}}2^{n(H(Y|W)-\epsilon)}. \tag{30}$$

Note also that

$$\begin{aligned} |A_\epsilon^n(Y|w^n)| &\leq \sum_{y^n \in T_\epsilon^n(Y|w^n)} |g(w^n)\cap\Lambda_t^n(Z|y^n)|\\ &= \sum_{z^n \in g(w^n)} |\Lambda_t^n(Y|z^n)\cap T_\epsilon^n(Y|w^n)|. \end{aligned} \tag{31}$$

Let $k_2(z^n,w^n) \triangleq |\Lambda_t^n(Y|z^n)\cap T_\epsilon^n(Y|w^n)|$. This value is the size of the "agreement set" as described above. Applying (30) and (31) gives

$$\begin{aligned} \sum_{z^n \in g(w^n)} k_2(z^n,w^n) &\geq \frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}J_{\mathcal{T}}}2^{n(H(Y|W)-\epsilon)}\\ &\geq 2^{n(H(Y|W)-2\epsilon)} \end{aligned} \tag{32}$$

for sufficiently large n. We will show that there is actually a single $\tilde{z}^n \in g(w^n)$ such that $k_2(\tilde{z}^n,w^n)$ represents a large portion of the above sum, so $\tilde{z}^n$ itself is almost as good as the entire bin. Then setting $h(w^n)=\tilde{z}^n$ will give us the properties we need. Note that

$$\begin{aligned} \sum_{z^n \in \mathcal{Z}^n} k_2(z^n,w^n) &= \sum_{y^n \in T_\epsilon^n(Y|w^n)} |\Lambda_t^n(Z|y^n)|\\ &\leq 2^{n(H(Y|W)+H_t(Z|Y)+\epsilon)}. \end{aligned} \tag{33}$$

Moreover

$$k_2(z^n,w^n) \leq |T_\epsilon^n(Y|w^n)| \leq 2^{n(H(Y|W)+\epsilon)}$$

so if for all z^n we let $l(z^n)$ be the integer such that

$$\begin{aligned} 2^{n(H(Y|W)-l(z^n)\epsilon)} &< k_2(z^n,w^n)\\ &\leq 2^{n(H(Y|W)-(l(z^n)-1)\epsilon)}. \end{aligned} \tag{34}$$

then $l(z^n) \geq 0$ for all z^n. Furthermore, if $k_2(z^n,w^n) > 0$, then $l(z^n) \leq L \triangleq \lceil\frac{H(Y|W)}{\epsilon}\rceil$. Let $M(l) = |\{z^n \in \mathcal{Z}^n : l(z^n) = l\}|$. Then from (33), for some l

$$\begin{aligned} 2^{n(H(Y|W)+H_t(Z|Y)+\epsilon)} &\geq \sum_{z^n \in \mathcal{Z}^n} k_2(z^n,w^n)\\ &\geq \sum_{z^n \in \mathcal{Z}^n : l(z^n)=l} k_2(z^n,w^n)\\ &\geq M(l)2^{n(H(Y|W)-l\epsilon)} \end{aligned}$$

giving

$$M(l) \leq 2^{n(H_t(Z|Y)+(l+1)\epsilon)}. \tag{35}$$

For any bin $b \in \mathcal{B}_{\mathbf{j}}$, let $\tilde{M}(l,b) \triangleq |\{z^n \in b : l(z^n) = l\}|$. Observe that when the bin b was created, it was one of $2^{nR(\mathbf{j})}$ bins into which all sequences in $\mathcal{Z}^n$ were placed. Thus the probability that any one sequence was placed in b was $2^{-nR(\mathbf{j})}$. Hence $\tilde{M}(l,b)$ is a binomial random variable with $M(l)$ trials and probability of success $2^{-nR(\mathbf{j})}$. Hence by (29) and (35)

$$\begin{aligned} \mathbb{E}\tilde{M}(l,b) &\leq M(l)2^{-nR(\mathbf{j})}\\ &\leq 2^{n(H_t(Z|Y)+(l+1)\epsilon)}2^{-n(H_t(Z|Y)+\nu)}\\ &= 2^{n((l+1)\epsilon-\nu)}. \end{aligned}$$

We want to disregard all codebooks for which $\tilde{M}(l,b)$ is much larger than its expectation. In particular, let $\mathcal{C}_2$ be the set of codebooks such that for any group of sensors, subcodebooks, type t, transactions $\mathbf{j}$, sequence $w^n \in \mathcal{W}^n$, bin b and integer l, either $\tilde{M}(l,b) \geq 2^{n\epsilon}$ if $(l+1)\epsilon-\nu \leq 0$ or $\tilde{M}(l,b) \geq 2^{n((l+2)\epsilon-\nu)}$ if $(l+1)\epsilon-\nu > 0$. We will show that the probability of $\mathcal{C}_2$ is small, so we may disregard it. Again using (23), if $(l+1)\epsilon-\nu \leq 0$

$$\Pr(\tilde{M}(l,b) \geq 2^{n\epsilon}) \leq \left[\frac{e}{2^{n(-l\epsilon+\nu)}}\right]^{2^{n\epsilon}} \leq 2^{-2^{n\epsilon}}$$

and if $(l+1)\epsilon-\nu > 0$

$$\begin{aligned} \Pr(\tilde{M}(l,b) \geq 2^{n((l+2)\epsilon-\nu)}) &\leq \left[\frac{e}{2^{n\epsilon}}\right]^{2^{n((l+2)\epsilon-\nu)}}\\ &\leq 2^{-2^{n((l+2)\epsilon-\nu)}} \end{aligned}$$

both for sufficiently large n. Therefore

$$\begin{aligned} \Pr(\mathcal{C}_2) \leq\ & 2^mC^m(n+1)^{|\mathcal{X}_{\mathcal{M}}|}J_{\mathcal{T}}|\mathcal{W}|^n2^{n(|\mathcal{X}_{\mathcal{M}}|+\nu)}\\ &\cdot\left[\sum_{0\leq l\leq\frac{\nu}{\epsilon}-1} 2^{-2^{n\epsilon}} + \sum_{\frac{\nu}{\epsilon}-1<l\leq L} 2^{-2^{n((l+2)\epsilon-\nu)}}\right] \end{aligned}$$

which vanishes as n grows.

We assume from now on that the codebook is not in $\mathcal{C}_2$, meaning in particular that $\tilde{M}(l,g(w^n)) \leq 2^{n\epsilon}$ for $(l+1)\epsilon-\nu \leq 0$ and $\tilde{M}(l,g(w^n)) \leq 2^{n((l+2)\epsilon-\nu)}$ for

$(l+1)\epsilon - \nu > 0$. Applying these and (34) to (32) and letting $\tilde{l}$ be an integer defined later

$$\begin{aligned}
2^{-n2\epsilon} &\leq 2^{-nH(Y|W)} \sum_{z^n \in g(w^n)} k_2(z^n, w^n) \\
&\leq \sum_{l=0}^{L} \tilde{M}(l, g(w^n)) 2^{-n(l-1)\epsilon} \\
&= \sum_{0 \leq l < \tilde{l}} \tilde{M}(l, g(w^n)) 2^{-n(l-1)\epsilon} \\
&\quad + \sum_{\tilde{l} \leq l \leq \frac{\nu}{\epsilon}-1} \tilde{M}(l, g(w^n)) 2^{-n(l-1)\epsilon} \\
&\quad + \sum_{\frac{\nu}{\epsilon}-1 < l \leq L} \tilde{M}(l, g(w^n)) 2^{-n(l-1)\epsilon} \\
&\leq \sum_{0 \leq l < \tilde{l}} \tilde{M}(l, g(w^n)) 2^{n\epsilon} + \sum_{\tilde{l} \leq l \leq \frac{\nu}{\epsilon}-1} 2^{n\epsilon} 2^{-n(\tilde{l}-1)\epsilon} \\
&\quad + \sum_{\frac{\nu}{\epsilon}-1 < l \leq L} 2^{n((l+2)\epsilon - \nu)} 2^{-n(l-1)\epsilon} \\
&\leq \sum_{0 \leq l < \tilde{l}} \tilde{M}(l, g(w^n)) 2^{n\epsilon} + L 2^{n(-\tilde{l}+2)\epsilon} + L 2^{n(3\epsilon - \nu)}.
\end{aligned}$$

Therefore

$$\sum_{0 \leq l < \tilde{l}} \tilde{M}(l, g(w^n)) \geq 2^{-n3\epsilon} \left(1 - L 2^{n(-\tilde{l}+4)\epsilon} - L 2^{n(5\epsilon - \nu)}\right).$$

Setting $\tilde{l} = 5$ and $\nu > 5\epsilon$ ensures that the right hand side is positive for sufficiently large n, so there is at least one $z^n \in g(w^n)$ with $|T_\epsilon^n(Y|w^n) \cap \Lambda_t^n(Y|z^n)| \geq 2^{n(H(Y|W)-4\epsilon)}$. Now we define $h : \mathcal{W}^n \to \mathcal{Z}^n$ such that $h(w^n)$ is such a z^n for $w^n \in A_\epsilon^n(W)$ and $h(w^n)$ is arbitrary for $w^n \notin A_\epsilon^n(W)$. If we let $\tilde{Z}^n = h(W^n)$, then

$$\begin{aligned}
&\Pr((Y^n \tilde{Z}^n) \in \Lambda_t^n(YZ)) \\
&\geq \sum_{w^n \in A_\epsilon^n(W)} p(w^n) \Pr(Y^n \in \Lambda_t^n(Y|h(w^n)) | W^n = w^n) \\
&\geq \sum_{w^n \in A_\epsilon^n(W)} p(w^n) \\
&\quad \cdot \Pr\left(Y^n \in T_\epsilon^n(Y|w^n) \cap \Lambda_t^n(Y|h(w^n)) | W^n = w^n\right) \\
&\geq \Pr(W^n \in A_\epsilon^n(W)) 2^{-n(H(Y|W)+\epsilon)} 2^{n(H(Y|W)-4\epsilon)} \\
&\geq \frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}} 2^{-n5\epsilon}.
\end{aligned}$$

The variables $(Y^n W^n \tilde{Z}^n)$ are distributed according to

$$q^n(y^n w^n z^n) = \left[\prod_{\tau=1}^{n} p(y_\tau) r(w_\tau | y_\tau)\right] \mathbf{1}\{z^n = h(w^n)\}.$$

Let $q_\tau(ywz)$ be the marginal distribution of $q^n(y^n w^n z^n)$ at time τ. It factors as

$$q_\tau(ywz) = p(y) r(w|y) q_\tau(z|w).$$

Let $\bar{q}(yz) \triangleq \frac{1}{n}\sum_\tau q_\tau(yz)$ and $\bar{q}(z|w) \triangleq \frac{1}{n}\sum_\tau q_\tau(z|w)$. Then

$$\bar{q}(yz) = p(y) \sum_w r(w|y) \bar{q}(z|w)$$

so by Lemma 2

$$\begin{aligned}
D&\left(t \middle\| p(y) \sum_w r(w|y) \bar{q}(z|w)\right) \\
&\qquad\qquad \leq -\frac{1}{n} \log\left(\frac{\delta}{2(n+1)^{|\mathcal{Y}\times\mathcal{Z}|}}\right) + 5\epsilon.
\end{aligned}$$

Therefore $t \in \breve{\mathcal{Q}}_{\mathcal{H},r}^{\eta}$ for sufficiently large n and some η such that $\eta \to 0$ as $\epsilon \to 0$.

G. Eavesdropping Traitors

We consider now the case that the traitors are able to overhear communication between the honest sensors and the decoder. If the traitors have perfect information, then hearing the messages sent by honest sensors will not give them any additional information, so the above coding scheme still works identically. If the traitors have imperfect information, we need to slightly modify the coding scheme, but the achievable rates are the same.

The important observation is that eavesdropping traitors only have access to messages sent in the past. Thus, by permuting the order in which sensors are polled in each round, the effect of the eavesdropping can be eliminated. In a given round, let $\mathcal{H}'$ be the set of honest sensors that transmit before any traitor. Since the additional information gain from eavesdropping will be no more than the values of $X_{\mathcal{H}'}^n$, the rate for this round, if no sensors are eliminated (i.e., $\mathcal{U}(\mathfrak{V}(I+1)) = \mathcal{U}(\mathfrak{V}(I))$), will be no more than the rate without eavesdropping when the traitors have access to $W'^n = (W^n, X_{\mathcal{H}'}^n)$. The goal of permuting the transmission order is to find an ordering in which all the traitors transmit before any of the honest sensors, since then the achieved rate, if no sensors are eliminated, will be the same as with no eavesdropping. It is possible to determine when such an order occurs because it will be the order that produces the smallest rate.

More specifically, we will alter the transmission order from round to round in the following way. We always choose an ordering such that for some $\mathcal{S} \in \mathfrak{H}$, the sensors $\mathcal{S}^c$ transmit before $\mathcal{S}$. We cycle through all such orderings until for each $\mathcal{S}$, there has been one round with a corresponding ordering in which no sensors were eliminated. We then choose one $\mathcal{S}$ that never produced a rate larger than the smallest rate encountered so far. We perform rounds in a order corresponding to $\mathcal{S}$ from then on. If the rate ever changes and is no longer the minimum rate encountered so far, we choose a different minimizing $\mathcal{S}$. The minimum rate will always be no greater than the achievable rate without eavesdropping, so after enough rounds, we achieve the same average rate.

VI. Fixed-Rate Coding

Consider an m-tuple of rates $(R_1, \ldots, R_m)$, encoding functions $f_i : \mathcal{X}_i^n \to \{1, \ldots, 2^{nR_i}\}$ for $i \in \mathcal{M}$, and decoding function

$$g : \prod_{i=1}^{m} \{1, \ldots, 2^{nR_i}\} \to \mathcal{X}_1^n \times \cdots \times \mathcal{X}_m^n.$$

Let $I_i \in \{1, \ldots, 2^{nR_i}\}$ be the message transmitted by sensor i. If sensor i is honest, $I_i = f_i(X_i^n)$. If it is a traitor, it may choose I_i arbitrarily, based on W^n. Define the probability of error $P_e \triangleq \Pr\left(X_{\mathcal{H}}^n \neq \hat{X}_{\mathcal{H}}^n\right)$ where $\hat{X}_{\mathcal{M}}^n = g(I_1, \ldots, I_m)$.

We say an m-tuple $(R_1, \ldots, R_m)$ is *deterministic-fixed-rate achievable* if for any $\epsilon > 0$ and sufficiently large n, there exist coding functions f_i and g such that, for any choice of actions by the traitors, $P_e \leq \epsilon$. Let $\mathcal{R}_{\text{dfr}} \subset \mathbb{R}^m$ be the set of deterministic-fixed-rate achievable m-tuples.

For randomized fixed-rate coding, the encoding functions become

$$f_i : \mathcal{X}_i^n \times \mathcal{Z} \to \{1, \ldots, 2^{nR_i}\}$$

where $\mathcal{Z}$ is the alphabet for the randomness. If sensor i is honest, $I_i = f_i(X_i^n, \rho_i)$, where $\rho_i \in \mathcal{Z}$ is the randomness produced at sensor i. Define an m-tuple to be *randomized-fixed-rate achievable* in the same way as above, and $\mathcal{R}_{\text{rfr}} \subset \mathbb{R}^m$ to be the set of randomized-fixed-rate achievable rate vectors.

For any $\mathcal{S} \subset \mathcal{M}$, let $\text{SW}(X_{\mathcal{S}})$ be the Slepian–Wolf rate region on the random variables $X_{\mathcal{S}}$. That is

$$\text{SW}(X_{\mathcal{S}}) \triangleq \left\{ R_{\mathcal{S}} : \forall \mathcal{S}' \subset \mathcal{S} : \sum_{i \in \mathcal{S}'} R_i \geq H(X_{\mathcal{S}'} | X_{\mathcal{S} \setminus \mathcal{S}'}) \right\}.$$

Let

$$\begin{aligned}\mathcal{R}_{\text{rfr}}^* &\triangleq \{(R_1, \ldots, R_m) : \forall \mathcal{S} \in \mathfrak{H} : R_{\mathcal{S}} \in \text{SW}(X_{\mathcal{S}})\}\\ \mathcal{R}_{\text{dfr}}^* &\triangleq \{(R_1, \ldots, R_m) \in \mathcal{R}_{\text{rfr}}^* : \forall \mathcal{S}_1, \mathcal{S}_2 \in \mathfrak{H} :\\ &\qquad \text{if } \exists r \in \mathcal{R}(\mathcal{S}_2) : H_r(X_{\mathcal{S}_1 \cap \mathcal{S}_2} | W) = 0\\ &\qquad \text{then } R_{\mathcal{S}_1 \cap \mathcal{S}_2} \in \text{SW}(X_{\mathcal{S}_1 \cap \mathcal{S}_2})\}.\end{aligned}$$

The following theorem gives the rate regions explicitly.

Theorem 2: The fixed-rate achievable regions are given by

$$\mathcal{R}_{\text{dfr}} = \mathcal{R}_{\text{dfr}}^* \qquad \text{and} \qquad \mathcal{R}_{\text{rfr}} = \mathcal{R}_{\text{rfr}}^*.$$

VII. Proof of Theorem 2

A. Converse for Randomized Coding

Assume $(R_1, \ldots, R_m)$ is randomized-fixed-rate achievable. Fix $\mathcal{S} \in \mathfrak{H}$. Suppose $\mathcal{S}^c$ are the traitors and perform a black hole attack. Thus $\hat{X}_{\mathcal{S}}^n$ must be based entirely on $\{f_i(X_i^n)\}_{i \in \mathcal{S}}$, and since $\Pr(X_{\mathcal{S}} \neq \hat{X}_{\mathcal{S}})$ can be made arbitrarily small, by the converse of the Slepian–Wolf theorem, which holds even if the encoders may use randomness, $R_{\mathcal{S}} \in \text{SW}(X_{\mathcal{S}})$.

B. Converse for Deterministic Coding

Assume $(R_1, \ldots, R_m)$ is deterministic-fixed-rate achievable. The converse for randomized coding holds equally well here, so $(R_1, \ldots, R_m) \in \mathcal{R}_{\text{rfr}}^*$. We prove by contradiction that $(R_1, \ldots, R_m) \in \mathcal{R}_{\text{dfr}}^*$ as well. Suppose $(R_1, \ldots, R_m) \in \mathcal{R}_{\text{rfr}}^* \setminus \mathcal{R}_{\text{dfr}}^*$, meaning that for some $\mathcal{S}_1$, $\mathcal{S}_2 \in \mathfrak{H}$, there exists $r \in \mathcal{R}(\mathcal{S}_2)$ such that $H_r(X_{\mathcal{S}_1 \cap \mathcal{S}_2} | W) = 0$ but $R_{\mathcal{S}_1 \cap \mathcal{S}_2} \notin \text{SW}(X_{\mathcal{S}_1 \cap \mathcal{S}_2})$. Consider the case that $\mathcal{H} = \mathcal{S}_1$ and r is such that $H_r(\mathcal{S}_1 \cap \mathcal{H} | W) = 0$. Thus the traitors always have access to $X_{\mathcal{S}_1 \cap \mathcal{H}}^n$.

For all $\mathcal{S} \in \mathfrak{H}$, let $D(X_{\mathcal{S}})$ be the subset of $T_\epsilon^n(X_{\mathcal{S}})$ such that all sequences in D are decoded correctly if $\mathcal{S}^c$ are the traitors and no matter what messages they send. Thus the probability that $X_{\mathcal{S}}^n \in D(X_{\mathcal{S}})$ is large. Let $D(X_{\mathcal{S}_1 \cap \mathcal{H}})$ be the marginal intersection of $D(X_{\mathcal{S}_1})$ and $D(X_{\mathcal{H}})$. That is, it is the set of sequences $x_{\mathcal{S}_1 \cap \mathcal{H}}^n$ such that there exists $x_{\mathcal{S}_1 \setminus \mathcal{H}}^n$ and $x_{\mathcal{H} \setminus \mathcal{S}_1}^n$ with $(x_{\mathcal{S}_1 \cap \mathcal{H}}^n x_{\mathcal{S}_1 \setminus \mathcal{H}}^n) \in D(X_{\mathcal{S}_1})$ and $(x_{\mathcal{S}_1 \cap \mathcal{H}}^n x_{\mathcal{H}_n \setminus \mathcal{S}_1}) \in D(X_{\mathcal{H}})$. Note that with high probability $X_{\mathcal{S}_1 \cap \mathcal{H}}^n \in D(X_{\mathcal{S}_1 \cap \mathcal{H}})$. Suppose $X_{\mathcal{S}_1 \cap \mathcal{H}}^n \in D(X_{\mathcal{S}_1 \cap \mathcal{H}})$ and $(X_{\mathcal{S}_1 \cap \mathcal{H}}^n X_{\mathcal{H} \setminus \mathcal{S}_1}^n) \in D(X_{\mathcal{H}})$, so by the definition of D, $\hat{X}_{\mathcal{S}_1 \cap \mathcal{H}}^n = X_{\mathcal{S}_1 \cap \mathcal{H}}^n$. Since $R_{\mathcal{S}_1 \cap \mathcal{H}} \notin \text{SW}(X_{\mathcal{S}_1 \cap \mathcal{H}})$, there is some $x'^n_{\mathcal{S}_1 \cap \mathcal{H}} \in D(X_{\mathcal{S}_1 \cap \mathcal{H}})$ mapping to the same codewords as $X_{\mathcal{S}_1 \cap \mathcal{H}}$ such that $x'^n_{\mathcal{S}_1 \cap \mathcal{H}} \neq X_{\mathcal{S}_1 \cap \mathcal{H}}^n$. Because the traitors have access to $X_{\mathcal{S}_1 \cap \mathcal{H}}$, they can construct $x'^n_{\mathcal{S}_1 \cap \mathcal{H}}$, and also find $x'^n_{\mathcal{S}_1 \setminus \mathcal{H}}$ such that $(x'^n_{\mathcal{S}_1 \cap \mathcal{H}} x'^n_{\mathcal{S}_1 \setminus \mathcal{H}}) \in D(X_{\mathcal{S}_1})$. If the traitors report $x'^n_{\mathcal{S}_1 \setminus \mathcal{H}}$, then we have a contradiction, since this situation is identical to that of the traitors being $\mathcal{S}_1^c$, in which case, by the definition of D, $\hat{X}_{\mathcal{S}_1 \cap \mathcal{H}}^n = x'^n_{\mathcal{S}_1 \cap \mathcal{H}}$.

C. Achievability for Deterministic Coding

Fix $(R_1, \ldots, R_m) \in \mathcal{R}_{\text{dfr}}^*$. Our achievability scheme will be a simple extension of the random binning proof of the Slepian–Wolf theorem given in [14]. Each encoding function $f_i : \mathcal{X}_i^n \to \{1, \ldots, 2^{nR_i}\}$ is constructed by means of a random binning procedure. Decoding is then performed as follows. For each $\mathcal{S} \in \mathfrak{H}$, if there is at least one $x_{\mathcal{S}}^n \in T_\epsilon^n(X_{\mathcal{S}})$ matching all received codewords from $\mathcal{S}$, let $\hat{x}_{i,\mathcal{S}}^n$ be one such sequence for all $i \in \mathcal{S}$. If there is no such sequence, leave $\hat{x}_{i,\mathcal{S}}^n$ null. Note that we produce a separate estimate $\hat{x}_{i,\mathcal{S}}^n$ of X_i^n for all $\mathcal{S} \ni i$. Let $\hat{x}_i^n$ equal one non-null $\hat{x}_{i,\mathcal{S}}^n$.

We now consider the probability of error. With high probability, $\hat{x}_{i,\mathcal{H}}^n = X_i^n$ for honest i. Thus all we need to show is that for all other $\mathcal{S} \in \mathfrak{H}$ with $i \in \mathcal{S}$, $\hat{x}_{i,\mathcal{S}}$ is null or also equal to X_i^n. Fix $\mathcal{S} \in \mathfrak{H}$. If there is some $r \in \mathcal{R}(\mathcal{S})$ with $H_r(X_{\mathcal{H} \cap \mathcal{S}} | W) = 0$, then by the definition of $\mathcal{R}_{\text{dfr}}^*$, $R_{\mathcal{H} \cap \mathcal{S}} \in \text{SW}(X_{\mathcal{H} \cap \mathcal{S}})$. Thus with high probability the only sequence $x_{\mathcal{H} \cap \mathcal{S}}^n \in T_\epsilon^n(X_{\mathcal{H} \cap \mathcal{S}})$ matching all received codewords will be $X_{\mathcal{H} \cap \mathcal{S}}^n$, so $\hat{x}_{i,\mathcal{S}}^n = X_i^n$ for all $i \in \mathcal{H} \cap \mathcal{S}$.

Now consider the case that $H_r(X_{\mathcal{H} \cap \mathcal{S}} | W) > 0$ for all $r \in \mathcal{R}(\mathcal{S})$. For convenience, let $Y = X_{\mathcal{H} \cap \mathcal{S}}$ and $Z = X_{\mathcal{T}}$. Let $R_Y = \sum_{i \in \mathcal{H} \cap \mathcal{S}} R_i$ and $R_Z = \sum_{i \in \mathcal{T}} R_i$. Since $R_{\mathcal{S}} \in \text{SW}(X_{\mathcal{S}})$, $R_Y + R_Z \geq H(YZ) + \eta$ for some η. Let $b_Y(y^n)$ be the set of sequences in $\mathcal{Y}^n$ that map to the same codewords as y^n, and let $b_Z \subset \mathcal{Z}^n$ be the set of sequences mapping to the codewords sent by the traitors. Then Y may be decoded incorrectly only if there is some $y'^n \in b_Y(Y^n)$ and some $z^n \in b_Z$ such that $y'^n \neq Y^n$ and $(y'^n z^n) \in T_\epsilon^n(YZ)$. For some $w^n \in \mathcal{W}^n$

$$\begin{aligned}&\Pr(\exists y'^n \in b_Y(Y^n) \setminus \{Y^n\}, z^n \in b_Z :\\ &\quad (y'^n z^n) \in T_\epsilon^n(YZ) | W^n = w^n)\\ &\leq \Pr\left(Y^n \notin T_\epsilon^n(Y|w^n) | W^n = w^n\right) + \sum_{y^n \in T_\epsilon^n(Y|w^n)} p(y^n|w^n)\\ &\quad \cdot \mathbf{1}\{\exists y'^n \in b_Y(y^n) \setminus \{y^n\}, z^n \in b_Z : (y'^n z^n) \in T_\epsilon^n(YZ)\}\\ &\leq \epsilon + 2^{-n(H(Y|W)-\epsilon)} \sum_{z^n \in b_Z \cap T_\epsilon^n(Z)} k_3(z^n, w^n) \qquad (36)\end{aligned}$$

where

$$\begin{aligned}k_3(z^n, w^n) \triangleq |\{&y^n \in T_\epsilon^n(Y|w^n) :\\ &\exists y'^n \in b_Y(y^n) \cap T_\epsilon^n(Y|z^n) \setminus \{y^n\}\}|.\end{aligned}$$

On average, the number of typical y^n put into a bin is at most $2^{n(H(Y)-R_Y+\epsilon)}$, so we can use (23) to assume with high probability than no more than $2^{n(H(Y)-R_Y+2\epsilon)}$ are put into any bin. Note that

$$\begin{aligned}\sum_{z^n\in T_\epsilon^n(Z)} & k_3(z^n,w^n)\\ &\le \sum_{z^n\in T_\epsilon^n(Z)}\sum_{y^n\in T_\epsilon^n(Y|w^n)} |b_Y(y^n)\cap T_\epsilon^n(Y|z^n)\backslash\{y^n\}|\\ &= \sum_{y^n\in T_\epsilon^n(Y|w^n)}\sum_{y'^n\in b_Y(y^n)\cap T_\epsilon^n(Y|z^n)\backslash\{y^n\}} |T_\epsilon^n(Z|y'^n)|\\ &\le 2^{n(H(Y|W)+\epsilon)}2^{n(H(Y)-R_Y+2\epsilon)}2^{n(H(Z|Y)+\epsilon)}\\ &= 2^{n(H(YZ)+H(Y|W)-R_Y+4\epsilon)}.\end{aligned}$$

The average k_3 sum over typical z^n in a given bin is thus

$$2^{n(H(YZ)+H(Y|W)-R_Y-R_Z+4\epsilon)} \le 2^{n(H(Y|W)+4\epsilon-\eta)}.$$

We can use an argument similar to that in Section V-F, partitioning $T_\epsilon^n(Z)$ into different l values, to show that with high probability, since $H(Y|W) > 0$, for all bins b_Z,

$$\sum_{z^n\in T_\epsilon^n(Z)\cap b_Z} k_3(z^n,w^n) \le 2^{n(H(Y|W)+5\epsilon-\eta)}.$$

Applying this to (36) gives

$$\begin{aligned}\Pr(\exists y'^n \in b_Y(Y^n)\backslash\{y^n\}, z^n\in b_Z :&\\ (y'^n z^n)\in T_\epsilon^n(YZ)|W^n = w^n) &\le \epsilon + 2^{n(6\epsilon-\eta)}.\end{aligned}$$

Letting $\eta > 6\epsilon$ ensures that the probability of error is always small no matter what bin b_Z the traitors choose.

D. Achievability for Randomized Coding

We perform essentially the same coding procedure as with deterministic coding, expect we also apply randomness in a similar fashion as with variable-rate coding. The only difference from the deterministic coding scheme is that each sensor has a set of C identically created subcodebooks, from which it randomly chooses one, then sends the chosen subcodebook index along with the codeword. Decoding is the same as for deterministic coding. An argument similar to that in Section V-D can be used to show small probability of error.

VIII. Conclusion

We gave an explicit characterization of the region of achievable rates for a Byzantine attack on distributed source coding with variable-rate codes, deterministic fixed-rate codes, and randomized fixed-rate codes. We saw that a different set of rates were achievable for the three cases, and gave converse proofs and rate achieving coding schemes for each. Variable-rate achievability was shown using an algorithm in which sensors use randomness to make it unlikely that the traitors can fool the coding process.

Much more work could be done in the area of Byzantine network source coding. Multiterminal rate distortion [15], [16] could be studied, or other topologies, such as side information. However, perhaps the biggest drawback in this paper is that, as we discussed in the introduction, because the traitors cannot in general be identified, it is difficult to imagine applications that do not require some post processing of the source estimates, for example to estimate some underlying process. Thus it would make sense to solve the coding and estimation problems simultaneously, such as in the CEO problem [17].

References

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

[2] S. Marano, V. Matta, and L. Tong, "Distributed inference in the presence of Byzantine sensors," in *Proc. 40th Annu. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 29-Nov. 1 2006.

[3] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, pp. 382–401, Jul. 1982.

[4] D. Dolev, "The Byzantine generals strike again," *J. Algor.*, vol. 3, no. 1, pp. 14–30, 1982.

[5] R. Perlman, "Network Layer Protocols with Byzantine Robustness," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1988.

[6] L. Zhou and Z. J. Haas, "Securing ad hoc networks," *IEEE Netw. Mag.*, vol. 13, pp. 24–30, Nov./Dec. 1999.

[7] Y. Hu and A. Perrig, "Security and privacy in sensor networks," *IEEE Security Privacy Mag*, vol. 2, pp. 28–39, 2004.

[8] T. Ho, B. Leong, R. Koetter, M. Médard, M. Effrons, and D. Karger, "Byzantine modification detection in multicast networks using randomized network coding," in *IEEE Proc. Intl. Symp. Inf. Theory*, Jun.-Jul. 27–2, 2004, p. 143.

[9] B. Awerbuch, D. Holmer, C. Nita-Rotaru, and H. Rubens, "An on-demand secure routing protocol resilient to Byzantine failures," in *ACM Workshop Wireless Security (WiSe)*, Sep. 2002.

[10] O. Kosut and L. Tong, "Capacity of cooperative fusion in the presence of Byzantine sensors," in *Proc. 44th Annu. Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 27–29, 2006.

[11] T. H. S. Jaggi, M. Langberg, and M. Effros, "Correction of adversarial errors in networks," in *Proc. Int. Symp. Inf. Theory Applicat.*, Adelaide, Australia, 2005.

[12] A. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.

[13] A. Wyner, "The common information of two dependent random wariables," *IEEE Trans. Inf. Theory*, vol. 21, pp. 163–179, Mar. 1975.

[14] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. 21, pp. 226–228, Mar. 1975.

[15] S. Y. Tung, "Multiterminal Source Coding," PhD, Cornell University, Ithaca, NY, 1978.

[16] T. Berger, *The Information Theory Approach to Communications*, G. Longo, Ed. Berlin, Germany: Springer-Verlag, 1978, Chapter Multiterminal source coding.

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