Optimality of the recursive data exchange protocol

Himanshu Tyagi[†] Shun Watanabe[‡]

***Abstract*—Multiple parties observe correlated data generated independent and identically (in time) from a known joint distribution. Parties communicate with each other interactively to enable each party to recover the data observed by all the other parties and attain *omniscience*. We characterize the asymptotic growth of the number of bits of interactive communication required by the parties to attain omniscience up to the second-order term. For the converse, we provide a single-shot lower bound for the required number of bits of communication, which yields the asymptotic result as a special case. It is shown that the knowledge of the distribution can be used to modify the recently proposed recursive universal data exchange protocol to render it optimal up to the second-order term. As a corollary, we provide a precise characterization of the reduction in communication for omniscience due to interaction.**

I. Introduction

In the multiparty data exchange or omniscience problem proposed in [4], parties observing correlated data seek to recover each other's data by communicating interactively. Each bit of communication is sent over a noiseless broadcast channel and is received by all the parties. For the specific case when the data of the parties is generated as an independent and identically distributed (IID) sequence (in time), [4] characterized the optimal rate of overall communication for omniscience and showed that it equals the value $R_{\mathsf{CO}}(\mathrm{P}_{\mathcal{M}})$ of a particular linear program (LP). The communication scheme proposed in [4], which attains the optimal rate, is based on the general scheme proposed in [3] and entails sending random hashes of appropriate rate by each party. The communication is *simple*, namely there is no interaction between the parties.

Recently, in [8], we considered the *individual sequence* version of the omniscience problem where the parties seek omniscience for every given sequence of observed data with large probability (and not only for the randomly generated IID sequence). We proposed an interactive communication protocol, termed the *recursive data exchange protocol* (RDE), which increases the number of bits communicated by each party in steps; a more detailed description of the protocol is provided in Section III-A below. An individual sequence performance bound for RDE was established in [8]. Specifically, it was shown that when the data sequence observed has a joint type Q, RDE attains omniscience with large probability and communicates no more than $nR_{\mathsf{CO}}(\mathrm{Q}) + \mathcal{O}(\sqrt{n})$ bits. As a corollary, it was shown for IID data that the protocol achieves the optimal rate without the knowledge of the distribution, with an excess rate of $\mathcal{O}(n^{-\frac{1}{2}}\sqrt{\log n})$ at a fixed n. In this paper, we show that a slight variant of RDE is, in fact, optimal up to the second-order asymptotic term when the IID distribution is known.

The utility of a protocol with type based individual sequence performance guarantees, such as the one summarized above, in attaining good second-order performance is clear: Once the joint type of the unknown data is known, we can easily communicate at appropriate rate to identify the exact sequence in that type set. Thus, heuristically, one purpose of interaction in the universal protocol is to reduce the uncertainty about the joint type of the data sequence. When the IID distribution generating the data is known, the joint type is specified to within a total variation distance of $\mathcal{O}(n^{-\frac{1}{2}})$ with large probability. But to get an optimal second-order performance, the uncertainty about the joint type must be reduced even further. If we directly use RDE, the $\mathcal{O}(n^{-\frac{1}{2}})$ excess rate resulting from the rounds of interaction will dominate the second-order term. To circumvent this bottleneck, we modify RDE by using the knowledge of the distribution to skip several rounds of interaction and reduce the overall communication.

To prove the second-order optimality of the protocol, we provide a single-shot lower bound for the number of bits that must be communicated to attain omniscience, which maybe of independent interest. The evaluation of this bound for the IID case yields the same tail-bound which appears in the performance of RDE, thereby establishing the optimality of the latter.

Note that RDE entails several rounds of interaction. We explicitly characterize the second order asymptotic performance of simple protocols such as the one used in [4] where the parties do not interact. As a corollary, we obtain a precise characterization of the reduction in communication for omniscience due to interaction.

The results of this paper extend those reported in [7], where the case of two parties was handled. But the multiparty protocol substantially generalizes the one used in [7]. Furthermore, the converse proof in [7] was based on a reduction argument relating data exchange to secret agreement. The converse proof here is new and is based on leveraging the structure of interactive protocols for data exchange.

The remainder of the paper is organized as follows. A formal description of the problem and a statement of our results is provided in the next section. Section III contains the proof of achievability including a brief description of the protocol and a sketch of its analysis. Section IV contains the aforementioned single-shot converse which yields the converse for the IID case as a corollary. The final section contains a

[†]Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. Email: htyagi@ece.iisc.ernet.in.
[‡]Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan. Email: shunwata@cc.tuat.ac.jp.

discussion on another notion of optimality for RDE.

II. Second order asymptotics of data exchange

We begin by giving a description of the data exchange or the omniscience problem in a generative single-shot model (in contrast to the individual sequence setup of [8]). The parties in the set $\mathcal{M} = \{1, ..., m\}$ observe correlated data, with the ith party observing a discrete random variable X_i taking values in the finite set $\mathcal{X}_i$. The observations are generated from a known distribution $\mathrm{P}_{X_\mathcal{M}} = \mathrm{P}_{X_1,...,X_m}$. The m parties wish to know each other's data using interactive communication over a noiseless (error-free) channel. We restrict to tree-protocols with shared randomness[1]. In fact, it is easy to see that deterministic protocols suffice when the distribution is known. Nevertheless, the scheme that we present uses a randomized protocol, which may later be derandomized to obtain a deterministic protocol. During the protocol, the parties exchange transcript $\Pi = \pi(X_\mathcal{M})$. At the end of the protocol, party j outputs an estimate $\hat{X}^{(j)}_\mathcal{M} = \hat{X}^{(j)}_\mathcal{M}(X_j, \Pi)$ of $X_\mathcal{M}$. The length $|\pi|$ of the protocol π corresponds to the maximum number of bits communicated during an execution of the protocol.

Definition 1. For $0 \leq \epsilon < 1$, a protocol π constitutes an ϵ-data exchange (ϵ-DE) protocol if there exists, for each $j \in \mathcal{M}$, an estimate function $\hat{X}^{(j)}_\mathcal{M} = \hat{X}^{(j)}_\mathcal{M}(X_j, \Pi)$ of $X_\mathcal{M}$ such that

$$\mathrm{P}\left(\hat{X}^{(j)}_\mathcal{M} = X_\mathcal{M},\ \forall j \in \mathcal{M}\right) \geq 1 - \epsilon.$$

The minimum over lengths $|\pi|$ of ϵ-DE protocols $|\pi|$ is the minimum communication for ϵ-DE, denoted $L_\epsilon(X_\mathcal{M})$.

For IID observations $X^n_\mathcal{M}$ distributed according to $\mathrm{P}_{X_\mathcal{M}}$, the minimum rate of communication for omniscience, $R(\mathrm{P}_{X_\mathcal{M}})$, is defined as

$$R(\mathrm{P}_{X_\mathcal{M}}) := \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} L_\epsilon(X^n_\mathcal{M}).$$

The quantity $R(\mathrm{P}_{X_\mathcal{M}})$ was characterized in [4] as

$$R(\mathrm{P}_{X_\mathcal{M}}) = \min \left\{ \sum_{i=1}^m R_i : \sum_{i \in B} R_i \geq H(X_B | X_{B^c}), \quad \forall\, B \subsetneq \mathcal{M} \right\}. \tag{1}$$

The value of the LP on the right-side of (1) is denoted by $R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}})$. An alternative expression for $R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}})$ was obtained in [4] by looking at its dual form. In fact, [1], [2] showed that the optimization in the dual form can be restricted to the partitions of $\mathcal{M}$ and showed that[2]

$$R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}}) = \max_{\sigma \in \Sigma(\mathcal{M})} \mathbb{H}_\sigma(\mathcal{M} | \mathrm{P}_{X_\mathcal{M}}), \tag{2}$$

[1]For a formal description of multiparty tree protocol for data exchange, see [8].

[2]The fact that $R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}})$ is lower bounded by the right-side of (2) was already shown in [4].

where $\Sigma(\mathcal{M})$ denotes the set of partitions of $\mathcal{M}$, and, for each $\sigma \in \Sigma(\mathcal{M})$,

$$\mathbb{H}_\sigma(\mathcal{M} | \mathrm{P}_{X_\mathcal{M}}) := \frac{1}{|\sigma| - 1} \sum_{i=1}^{|\sigma|} H\left(X_\mathcal{M} | X_{\sigma_i}\right). \tag{3}$$

In this paper, we shall derive a more precise asymptotic characterization of $L_\epsilon(X^n_\mathcal{M})$ which is accurate up to the second-order asymptotic term. The densities of information quantities in (3) play an important role in our characterization. In particular, for every $\sigma \in \Sigma(\mathcal{M})$, denote

$$h_\sigma\left(x_\mathcal{M}\right) := \frac{1}{|\sigma| - 1} \sum_{i=1}^{|\sigma|} h(x_\mathcal{M} | x_{\sigma_i}), \tag{4}$$

where for every subset B of $\mathcal{M}$,

$$h(x_\mathcal{M} | x_B) = \log \frac{1}{\mathrm{P}_{X_\mathcal{M} | X_B}\left(x_\mathcal{M} | x_B\right)}.$$

Our main result in this paper is the following.

Theorem 1. *Given a distribution $\mathrm{P}_{X_\mathcal{M}}$, let $\Sigma^*(\mathcal{M}) \subseteq \Sigma(\mathcal{M})$ be the set of partitions σ of $\mathcal{M}$ which maximize $\mathbb{H}_\sigma(\mathcal{M} | \mathrm{P}_{X_\mathcal{M}})$. Further, let $\mathbf{Z}$ be a $d = |\Sigma^*(\mathcal{M})|$-dimensional Gaussian vector with mean $\mathbf{0}$ and covariance matrix given by*

$$\mathbf{V} = \mathrm{Cov}\big[\big(h_\sigma\left(X_\mathcal{M}\right) : \sigma \in \Sigma^*(\mathcal{M})\big)\big].$$

Then, we have

$$L_\epsilon(X^n_\mathcal{M}) = nR(\mathrm{P}_{X_\mathcal{M}}) + \sqrt{n} r^*(\epsilon) + o(\sqrt{n}),$$

where

$$r^*(\epsilon) := \inf \left\{ r : \mathbb{P}\left(\bigcup_{i=1}^d \{Z_i \geq r\} \right) \leq \epsilon \right\}.$$

Our proof of Theorem 1 relies on analysis of a modification of the universal protocol of [8] and a new single-shot converse bound. In particular, we show that the leading term in bounds for probability of error, in both achievability part given in Section III and converse part given in Section IV, is roughly

$$\mathbb{P}\left(\bigcup_{\sigma \in \Sigma(\mathcal{M})} \left\{ \frac{1}{n} \sum_{t=1}^n h_\sigma\left(X_{\mathcal{M},t}\right) \geq R_n \right\} \right) \tag{5}$$

for a fixed communication rate R_n. Note that, for each t, the expected value of the random variable $h_\sigma\left(X_{\mathcal{M},t}\right)$ equals $R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}})$ if $\sigma \in \Sigma^*(\mathcal{M})$ and is strictly less than $R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}})$ if $\sigma \notin \Sigma^*(\mathcal{M})$. Thus, for $R_n = R_{\texttt{CO}}(\mathrm{P}_{X_\mathcal{M}}) + \mathcal{O}(n^{-\frac{1}{2}})$, the probability of the each event $\left\{\frac{1}{n}\sum_{t=1}^n h_\sigma\left(X_{\mathcal{M},t}\right) \geq R_n\right\}$ for $\sigma \notin \Sigma^*(\mathcal{M})$ goes to 0 in the limit as n goes to infinity. Thus, by the union bound and the law of large numbers, we can omit the events corresponding to $\sigma \notin \Sigma^*(\mathcal{M})$ in bound (5). Finally, we complete the proof of Theorem 1 by using the multivariate Berry-Esséen theorem along the lines of [6].

The performance reported above is achieved by an interactive protocol. If we restrict to simple protocols with no interaction between the parties, using an information-spectrum analysis ($cf.$ [5, Sec. 7]) we can establish the following

characterization of the minimum length $L_\epsilon^{\mathrm{s}}(X_{\mathcal{M}}^n)$ of a simple ϵ-DE protocol.

Theorem 2. *Given a distribution $\mathrm{P}_{X_{\mathcal{M}}}$, let $(\tilde{Z}_{[B]} : \emptyset \neq B \subsetneq \mathcal{M})$ be a $(2^{|\mathcal{M}|}-2)$-dimensional Gaussian vector with mean $\mathbf{0}$ and the covariance matrix given by*

$$\tilde{\mathbf{V}} = \mathrm{Cov}\big[\big(h\left(X_{\mathcal{M}}|X_{B^c}\right) : B \subsetneq \mathcal{M}\big)\big].$$

Then, we have

$$L_\epsilon^{\mathrm{s}}(X_{\mathcal{M}}^n) = nR(\mathrm{P}_{X_{\mathcal{M}}}) + \sqrt{n}\tilde{r}^*(\epsilon) + o(\sqrt{n}),$$

where

$$\tilde{r}^*(\epsilon) := \inf\left\{\sum_{j=1}^m r_j : \mathbb{P}\left(\bigcup_{B\subsetneq\mathcal{M}}\left\{\tilde{Z}_{[B]} \geq \sum_{j\in B} r_j\right\}\right) \leq \epsilon\right\}.$$

It can be shown that when $\tilde{\mathbf{V}}$ is a non-singular covariance matrix, $r^*(\epsilon) < \tilde{r}^*(\epsilon)$ holds. Therefore, interactive protocols are strictly better than simple protocols in this case. Furthermore, Theorems 1 and 2 together establish that the precise gain due to interaction is given by

$$\sqrt{n}\left(\tilde{r}^*(\epsilon) - r^*(\epsilon)\right) + o(\sqrt{n}).$$

III. Achievability

A. A modified recursive data exchange protocol

We use a slight modification of RDE to attain the optimal performance. RDE, as described in [8] has three components:

1) A subroutine termed OMN which iteratively increments the rate of communication of each party until a subset of parties recover each other, *i.e.*, attain *local omniscience*;
2) a decoder subroutine which is a variant of the minimum entropy decoder and returns the data of a subset of parties whenever possible; else it outputs a `NACK` and requests for transmission of more hash bits;
3) an outer routine which repeatedly calls OMN with updated input parameters.

The subroutine OMN increments the rates of each party in steps of size Δ. Parties start communicating in order of the entropies of the marginal types of their local data, with the ith party starting communication when the rate R_1 of the highest entropy party crosses, roughly, $H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_i})$ where $\mathrm{P}_{\mathbf{x}_i}$ denotes the type of the sequence $\mathbf{x}_i$. The following property ensues: The rates of communicating parties differ by as much as the difference of the entropies of their marginal types. If the rates are incremented in this fashion, we showed in [8] that when the decoder of party i recovers the data of party j, then the decoder of party j must also recover the data of party i. Thus, when OMN terminates, a subset of parties must have attained local omniscience (if an error has not occurred). The key observation in [8], which also motivates the name of the protocol, is that when OMN terminates, the parties in the subsets that have attained local omniscience have rates which appear as if the parties in the subset are collocated and have been executing the protocol as a single party. Therefore, the outer routine continues by making another call to OMN, but for a modified model where the parties in the omniscience attaining subsets are collocated. Finally, if no error occurs, the outer routine terminates when all parties attain local omniscience. The main result in [8] showed that for each sequence $\mathbf{x}_{\mathcal{M}}$, the parties attain omniscience after communicating roughly $nR_{\mathtt{CO}}(\mathrm{P}_{\mathbf{x}_{\mathcal{M}}})$ bits. Formally, for a given sequence $\mathbf{x}_{\mathcal{M}}$, the probability of error of the protocol is bounded above by

$$C_1\left(\frac{\log|\mathcal{X}_{\mathcal{M}}|}{\Delta_n} + m\right)p(n)2^{-n\Delta},$$

where C_1 is a constant depending only on m and $p(n)$ is a polynomial in n. Furthermore, if an error does not occur, the number of bits communicated by the protocol for input $\mathbf{x}_{\mathcal{M}}$ is bounded above by

$$nR_{\mathtt{CO}}(\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}) + nC_2\Delta + C_3\frac{\log|\mathcal{X}_{\mathcal{M}}|}{\Delta} + C_4\log n,$$

where C_2, C_3, C_4 are constants depending only on m. The term $\frac{\log|\mathcal{X}_{\mathcal{M}}|}{\Delta}$ in the expression above corresponds to 1 bit of communication required for each `ACK-NACK`, namely 1 bit for each round of interactive communication. The choice of Δ which optimizes the expression above is roughly $\Delta = 1/\sqrt{n}$. However, this results in an excess rate of $\mathcal{O}(1/\sqrt{n})$, which may asymptotically lead to second-order suboptimality of the protocol. To circumvent this, we modify the protocol above by skipping some rounds of interaction using the knowledge of the joint distribution $\mathrm{P}_{X_{\mathcal{M}}}$.

To wit, it follows from the description of the decoder that for a sequence $\mathbf{x}_{\mathcal{M}}$, when the parties in a subset A attain local omniscience, the parties in A must be communicating and the rates $(R_i, i \in A)$ must belong to the region $\mathcal{R}_{\mathtt{CO}}^{\Delta}(A|\mathrm{P}_{\mathbf{x}_A})$, which is the set of all vectors $(R_i, i \in A)$ such that

$$\sum_{i\in B} R_i \geq H(X_B|X_{A\setminus B}) + |B|\Delta, \quad \forall\, B \subsetneq A.$$

At this point OMN terminates and the outer protocol calls another instance of OMN with modified parameters. The rates are incremented in steps of size Δ so that such *transition points* are not missed. If we were assured that for a certain number of rounds the rate vector $(R_i, i \in A)$ will remain outside $\mathcal{R}_{\mathtt{CO}}^{\Delta}(A|\mathrm{P}_{\mathbf{x}_A})$, where $\mathrm{P}_{\mathbf{x}_A}$ denotes the joint type of the sequence $\mathbf{x}_A = (\mathbf{x}_i, i \in A)$, for every subset A that has not yet attained local omniscience, we need not wait for a `NACK` message from the decoders to increment the rates and can directly increase the rate to the final rate at the end of these rounds. The next observation allows us to use the knowledge of $\mathrm{P}_{X_{\mathcal{M}}}$ to identify such regions.

Lemma 3 (Continuity)**.** *For every $\Delta > 0$, $n \in \mathbb{N}$, and the type $\hat{\mathrm{P}}_{X_{\mathcal{M}}^n}$ of $X_{\mathcal{M}}^n$, with probability greater than $1 - 2|\mathcal{X}_{\mathcal{M}}|/n^2$ it holds for every $A \subseteq \mathcal{M}$ that*

$$\mathcal{R}_{\mathtt{CO}}^{\Delta}\left(A|\hat{\mathrm{P}}_{X_A^n}\right) \subseteq \mathcal{R}_{\mathtt{CO}}^{\Delta-\nu_n}\left(A|\mathrm{P}_{X_A}\right) \subseteq \mathcal{R}_{\mathtt{CO}}^{\Delta-2\nu_n}\left(A|\hat{\mathrm{P}}_{X_A^n}\right),$$

where

$$\nu_n = |\mathcal{X}_{\mathcal{M}}| \cdot \sqrt{\frac{\log n}{n}} \cdot \log\left(\frac{n}{\log n}\right).$$

The proof is omitted due to lack of space.

Therefore, if we are executing the protocol for data generated by $\mathrm{P}_{X_{\mathcal{M}}}$, with large probability we will not encounter a transition point as long as for every subset A that has not attained local omniscience, the rate vector $(R_i, i \in A)$ remains outside $\mathcal{R}_{\mathsf{CO}}^{\Delta-\nu_n}(A|\mathrm{P}_{X_A})$. Based on this observation, we can simply modify RDE as follows to reduce the number of rounds of interaction and thereby the number of bits communicated:

1) In the outer routine, instead of calling OMN with the current rates, the parties keep on incrementing their rates until for one of the subsets A of communicating parties that has not attained local omniscience, $(R_i, i \in A)$ enters $\mathcal{R}_{\mathsf{CO}}^{\Delta-\nu_n}(A|\mathrm{P}_{X_A})$. At this point, the outer routine calls OMN with the current rates.
2) The OMN subroutine proceeds as before, but terminates if the overall number of bits communicated exceeds a fixed number $nR_{\max}$. At this point the outer routine terminates as well and the protocol concludes.

B. Analysis of the protocol

We now provide a bound for the probability of error of the proposed protocol and show that it achieves the second-order performance claimed in Theorem 1. Suppose that for a subset A of communicating parties that has not attained local omniscience, the rate vector $(R_i, i \in A)$ belongs to $\mathcal{R}_{\mathsf{CO}}^{\Delta-\nu_n}(A|\mathrm{P}_{X_A})$. At this point, OMN is called and the rates of each party in A are incremented by Δ per round. By Lemma 3, the modified protocol engages in at most $2\nu_n/\Delta$ rounds before the parties in A attain local omniscience. Since there are at most 2^m transition points (one for each subset of $\mathcal{M}$), the total number of rounds of interaction in execution of the modified protocol is at most $2^{m+1}\nu_n/\Delta$. The next result follows by modifying the proof of [8, Theorem 4] to reflect the reduced number of rounds of interaction, accounting for the probability of the event excluded in Lemma 3, and setting $\Delta = \sqrt{\nu_n/n}$.

Theorem 4. *Denoting by $\hat{\mathrm{P}}_{X_{\mathcal{M}}^n}$ the type of the sequence $X_{\mathcal{M}}^n$, the probability of error P_e for the protocol described above is bounded as*

$$\mathrm{P}_e \le \mathbb{P}\left(R_{\mathsf{CO}}(\hat{\mathrm{P}}_{X_{\mathcal{M}}^n}) + C_5\sqrt{|\mathcal{X}_{\mathcal{M}}|}\cdot\left(\frac{\log n}{n}\right)^{\frac{3}{4}} \ge R_{\max}\right) + C_1\left(\frac{\log|\mathcal{X}_{\mathcal{M}}|}{n}+m\right)p(n)2^{-\sqrt{n\nu_n}} + \frac{2|\mathcal{X}_{\mathcal{M}}|}{n^2},$$

where C_5 depends only on m.

Next, we weaken the individual-sequence-based performance bound above to obtain a form amenable to second-order asymptotic analysis. To that end, we use the following observation.

Lemma 5. *For every $R_n \ge 0$,*

$$\mathbb{P}\left(R_{\mathsf{CO}}(\hat{\mathrm{P}}_{X_{\mathcal{M}}^n}) \ge R_n\right) \le \mathbb{P}\left(\bigcup_{\sigma\in\Sigma(\mathcal{M})}\left\{\frac{1}{n}\sum_{t=1}^{n} h_\sigma\left(X_{\mathcal{M},t}\right)\right\} \ge R_n\right),$$

where $h_\sigma(\cdot)$ is defined in (4)*.*

The proof is omitted due to lack of space and is based on the observation that for $\overline{X}_{\mathcal{M}}$ distributed according to $\hat{\mathrm{P}}_{X_{\mathcal{M}}^n}$

$$H\left(\overline{X}_{\mathcal{M}}|\overline{X}_B\right) \le \mathbb{E}\left[\log\frac{1}{\mathrm{P}_{X_{\mathcal{M}}|X_B}\left(\overline{X}_{\mathcal{M}}|\overline{X}_B\right)}\right].$$

Combining the bounds above, we obtain our final bound for P_e.

Corollary 6. *The probability of error P_e for the protocol described above is bounded as*

$$\mathrm{P}_e \le \mathbb{P}\left(\bigcup_{\sigma\in\Sigma(\mathcal{M})}\left\{\frac{1}{n}\sum_{t=1}^{n} h_\sigma\left(X_{\mathcal{M},t}\right)\right\} \ge R_{\max}-\delta_n\right) + \epsilon_n,$$

where $\delta_n = o(\sqrt{n})$ and $\epsilon_n = o(1)$.

IV. Converse

We establish the following single-shot converse bound for multiparty data exchange.

Theorem 7. *For every ϵ-DE protocol π and $\lambda > 0$, we have*

$$|\pi| \ge \lambda + \frac{m}{m-1}\cdot\log(P_\lambda - \epsilon) - \frac{m^2}{m-1}$$

where the tail-probability P_λ is given by

$$P_\lambda = \mathbb{P}\left(\bigcup_{\sigma\in\Sigma(\mathcal{M})}\{h_\sigma(X_{\mathcal{M}}) > \lambda\}\right).$$

In fact, we prove the following alternative form of Theorem 7 which shows that for any protocol π, the probability of error P_e is bounded below as

$$\mathrm{P}_e \ge \mathbb{P}\left(\bigcup_{\sigma\in\Sigma(\mathcal{M})}\{h_\sigma(X_{\mathcal{M}}) > \lambda\}\right) - \sum_{\sigma\in\Sigma(\mathcal{M})} 2^{\frac{|\sigma|-1}{|\sigma|}(|\pi|-\lambda)}. \tag{6}$$

This alternative form leads to the bound in (5) when the observations are IID.

Proof. Let $\hat{X}_{\mathcal{M}}^{(j)} = \hat{X}_{\mathcal{M}}^{(j)}(X_j, \pi(X_{\mathcal{M}}))$, $j \in \mathcal{M}$, denote the output function of the jth party for π. Further, for a partition $\sigma \in \Sigma(\mathcal{M})$, let

$$\mathcal{T}_\sigma(\lambda) = \left\{x_{\mathcal{M}} \in \mathcal{X}_{\mathcal{M}} : h_\sigma(x_{\mathcal{M}}) \le \lambda\right\},$$

and

$$\mathcal{T}(\lambda) = \bigcap_{\sigma\in\Sigma(\mathcal{M})}\mathcal{T}_\sigma(\lambda).$$

We show that

$$\mathbb{P}\left(\hat{X}_{\mathcal{M}}^{(j)} = X_{\mathcal{M}},\ \forall j \in \mathcal{M}\right)$$

$$\leq \mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{T}(\lambda)\right) + \sum_{\sigma \in \Sigma(\mathcal{M})} 2^{\frac{|\sigma|-1}{|\sigma|}(|\pi|-\lambda)}.$$

which is equivalent to (6) and Theorem 7.

To that end, note that $x_{\mathcal{M}} \in \mathcal{T}_\sigma(\lambda)^c$ implies

$$\mathrm{P}_{X_{\mathcal{M}}}\left(x_{\mathcal{M}}\right) \leq 2^{-\lambda \frac{|\sigma|-1}{|\sigma|}} \left(\prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\left(x_{\sigma_i}\right)\right)^{\frac{1}{|\sigma|}}. \tag{7}$$

Denote by $\mathcal{S}$ the support $\mathtt{supp}(\mathrm{P}_{X_{\mathcal{M}}})$ of $\mathrm{P}_{X_{\mathcal{M}}}$ and by $\mathcal{C}$ the set of sequences for which an error does not occur, *i.e.*,

$$\mathcal{C} := \left\{ x_{\mathcal{M}} \in \mathcal{S} : \hat{X}^{(j)}_{\mathcal{M}}(x_j, \pi(x_{\mathcal{M}})) = x_{\mathcal{M}},\ \forall j \in \mathcal{M} \right\}.$$

Further, for each transcript τ of π, let

$$\mathcal{S}(\tau) = \{x_{\mathcal{M}} \in \mathcal{S} : \pi(x_{\mathcal{M}}) = \tau\}$$

and

$$\mathcal{C}(\tau) = \left\{ x_{\mathcal{M}} \in \mathcal{S} : \pi(x_{\mathcal{M}}) = \tau,\ \hat{X}^{(j)}_{\mathcal{M}}(x_j, \tau) = x_{\mathcal{M}},\ \forall j \in \mathcal{M} \right\};$$

clearly, $\mathcal{C}(\tau) \subseteq \mathcal{S}(\tau)$. With these notations, we have

$$\begin{aligned}\mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C}\right) &= \mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}(\lambda)\right) + \mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}(\lambda)^c\right) \\ &\leq \mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{T}(\lambda)\right) + \sum_{\sigma \in \Sigma(\mathcal{M})} \mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}_\sigma(\lambda)^c\right),\end{aligned}$$

where the inequality follows from the union bound. To complete the proof, we show that

$$\mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}_\sigma(\lambda)^c\right) \leq 2^{\frac{|\sigma|-1}{|\sigma|}(|\pi|-\lambda)}.$$

Indeed, for each $\sigma \in \Sigma(\mathcal{M})$, we have

$$\begin{aligned}&\mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}_\sigma(\lambda)^c\right) \\ &= \sum_{\tau} \mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C}(\tau) \cap \mathcal{T}_\sigma(\lambda)^c\right) \\ &= \sum_{\tau} \sum_{x_{\mathcal{M}} \in \mathcal{C}(\tau) \cap \mathcal{T}_\sigma(\lambda)^c} \mathrm{P}_{X_{\mathcal{M}}}\left(x_{\mathcal{M}}\right) \\ &\leq 2^{-\lambda \frac{|\sigma|-1}{|\sigma|}} \sum_{\tau} \sum_{x_{\mathcal{M}} \in \mathcal{C}(\tau)} \left(\prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\left(x_{\sigma_i}\right)\right)^{\frac{1}{|\sigma|}},\end{aligned}$$

where the inequality is by (7). To complete the proof, we note two properties of the set $\mathcal{C}(\tau)$; the proofs are omitted due to lack of space.

1) The set $\mathcal{C}(\tau)$ consists of sequences which are disjoint along each coordinate $j \in \mathcal{M}$, *i.e.*, the set is of the form $\mathcal{C}(\tau) = \{x^{(\tau)}_{\mathcal{M},\ell} = (x^{(\tau)}_{1,\ell}, \ldots, x^{(\tau)}_{m,\ell}) : \ell = 1, \ldots, L_\tau\}$ for some integer L_τ, where $x^{(\tau)}_{j,\ell} \neq x^{(\tau)}_{j,\tilde{\ell}}$ for every $\ell \neq \tilde{\ell}$ and $j \in \mathcal{M}$.
2) Let $\mathcal{S}_j(\tau)$ be the projection of the set $\mathcal{S}(\tau)$ along the jth coordinate, i.e.,

$$\mathcal{S}_j(\tau) = \{x_j \in \mathcal{X}_j : \exists\, x_{\mathcal{M}\setminus\{j\}} \text{ s.t. } (x_1, \ldots, x_m) \in \mathcal{S}(\tau)\}.$$

Then, the rectangles $\mathcal{S}_1(\tau) \times \cdots \times \mathcal{S}_m(\tau)$ and $\mathcal{S}_1(\tilde{\tau}) \times \cdots \times \mathcal{S}_m(\tilde{\tau})$ are disjoint for every $\tau \neq \tilde{\tau}$.

The obvious extensions of these properties for any partition σ of coordinates $\mathcal{M}$ also hold. Using these properties and the fore-mentioned bound for $\mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}_\sigma(\lambda)^c\right)$, we have

$$\begin{aligned}&\mathrm{P}_{X_{\mathcal{M}}}\left(\mathcal{C} \cap \mathcal{T}_\sigma(\lambda)^c\right) \\ &\leq 2^{-\lambda \frac{|\sigma|-1}{|\sigma|}} \sum_{\tau} \sum_{x_{\mathcal{M}} \in \mathcal{C}(\tau)} \left(\prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\left(x_{\sigma_i}\right)\right)^{\frac{1}{|\sigma|}}, \\ &= 2^{-\lambda \frac{|\sigma|-1}{|\sigma|}} \sum_{\tau} \sum_{\ell=1}^{L_\tau} \left(\prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\left(x^{(\tau)}_{\sigma_i,\ell}\right)\right)^{\frac{1}{|\sigma|}} \\ &\leq 2^{-\lambda \frac{|\sigma|-1}{|\sigma|}} \sum_{\tau} \prod_{i=1}^{|\sigma|} \left(\sum_{\ell=1}^{L_\tau} \mathrm{P}_{X_{\sigma_i}}\left(x^{(\tau)}_{\sigma_i,\ell}\right)\right)^{\frac{1}{|\sigma|}} \\ &\leq 2^{-\lambda \frac{|\sigma|-1}{|\sigma|}} \sum_{\tau} \left(\prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\left(\mathcal{S}_{\sigma_i}(\tau)\right)\right)^{\frac{1}{|\sigma|}} \\ &\leq \left(2^{-\lambda+|\pi|}\right)^{\frac{|\sigma|-1}{|\sigma|}} \left(\sum_{\tau} \prod_{i=1}^{|\sigma|} \mathrm{P}_{X_{\sigma_i}}\left(\mathcal{S}_{\sigma_i}(\tau)\right)\right)^{\frac{1}{|\sigma|}} \\ &\leq \left(2^{-\lambda+|\pi|}\right)^{\frac{|\sigma|-1}{|\sigma|}},\end{aligned}$$

where the second and the fourth inequalities use Hölder's inequality, the third inequality uses property 1 and the final inequality uses property 2.

□

V. Discussion

An application of the single-shot converse bound in Section IV to the uniform distribution on a given joint type class yields a converse in an individual sequence setup. It can be shown that if the parties observe a sequence $\mathbf{x}_{\mathcal{M}}$ of joint type $\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}$ which is known to the parties, they still require communication roughly of rate $R_{\mathtt{CO}}(\mathrm{P}_{\mathbf{x}_{\mathcal{M}}})$ to attain omniscience. This, when combined with the performance bound for RDE in [8], shows that RDE attains a worst-case redundancy of $\mathcal{O}(\sqrt{n})$ bits when compared with a class of data exchange protocols which know the joint type. We have omitted this result due to lack of space.

References

[1] C. Chan, "On tightness of mutual dependence upperbound for secret-key capacity of multiple terminals," *arXiv:0805.3200*, 2008.

[2] C. Chan and L. Zheng, "Mutual dependence for secret key agreement," *Proc. Annual Conference on Information Sciences and Systems (CISS)*, 2010.

[3] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 155–165, March 1980.

[4] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, December 2004.

[5] T. S. Han, *Information-Spectrum Methods in Information Theory [English Translation]*. Series: Stochastic Modelling and Applied Probability, Vol. 50, Springer, 2003.

[6] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, 2014.

[7] H. Tyagi, P. Viswanath, and S. Watanabe, "Interactive communication for data exchange," *Proc. IEEE International Symposium on Information Theory*, pp. 1806 – 1810, 2015.

[8] H. Tyagi and S. Watanabe, "Universal multiparty data exchange and secret key agreement," *CoRR*, vol. abs/1605.01033, 2016.