Universal Multiparty Data Exchange and Secret Key Agreement

Himanshu Tyagi[†] Shun Watanabe[‡]

Abstract

Multiple parties observing correlated data seek to recover each other's data and attain omniscience. To that end, they communicate interactively over a noiseless broadcast channel – each bit transmitted over this channel is received by all the parties. We give a universal interactive communication protocol, termed the *recursive data exchange protocol* (RDE), which attains omniscience for any sequence of data observed by the parties and provide an individual sequence guarantee of performance. As a by-product, for observations of length n, we show the universal rate optimality of RDE up to an $\mathcal{O}(n^{-1/2}\sqrt{\log n})$ term in a generative setting where the data sequence is independent and identically distributed (in time). Furthermore, drawing on the duality between omniscience and secret key agreement due to Csiszár and Narayan, we obtain a universal protocol for generating a multiparty secret key of rate at most $\mathcal{O}(n^{-1/2}\sqrt{\log n})$ less than the maximum rate possible. A key feature of RDE is its *recursive structure* whereby when a subset A of parties recover each-other's data, the rates appear as if the parties have been executing the protocol in an alternative model where the parties in A are collocated.

I. Introduction

An m party omniscience protocol is an interactive communication protocol that enables m parties to recover each other's data. The communication is error-free and is in a broadcast mode wherein the transmission of each party is received by all the other parties. Such protocols were first considered in [14] in a two-party setup, where bounds for the number of bits communicated on average and in the worst-case were derived for the case when no error is allowed. The m party version, and the omniscience terminology, was proposed in [12] where the collective observations of the parties was assumed to be an independent and identically distributed (IID) sequence generated from a known distribution[1] $\mathrm{P}_{X_1\cdots X_m}$. It

[†]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.

[1]Throughout we shall restrict to finite random variables and use the phrase probability distribution interchangeably with probability mass function (pmf).

was shown in [12] that a simultaneous communication protocol based on sending random hash bits of appropriate rates attains the optimal sum-rate $R\left(\mathrm{P}_{X_1\cdots X_m}\right)$. A common feature of these prior works is that the protocol relies on the knowledge of the underlying distribution $\mathrm{P}_{X_1\cdots X_m}$. Note that the protocol proposed in [12] relies on the classic multiterminal source coding scheme given in [10]. Thus, it inherits the following universality feature from that scheme: If for $1 \leq i \leq m$ the ith party communicates rate R_i, the protocol attains omniscience for any source distribution $\mathrm{P}_{X_1\cdots X_m}$ for which the rate vector $(R_1, \ldots, R_m)$ lies in the omniscience rate-region corresponding to $\mathrm{P}_{X_1\cdots X_m}$. Nevertheless, this provides no guarantee of universal optimality for the sum-rate $(R_1 + \cdots + R_m)$ for an arbitrary source $\mathrm{P}_{X_1\cdots X_m}$.

A naive protocol entails using the first n' samples to estimate the entropies involved and then applying the optimal protocol of [12] with rates satisying the entropy constraints. Specifically, by using the estimator for entropy proposed in [29], we can estimate the entropy to within an approximation error of $\mathcal{O}(1/\sqrt{n'})$ using n' samples, where the constants implied by $\mathcal{O}$ depend on the support size of the distribution. This results in an universally sum-rate optimality protocol, but for observations of length n, the overall excess rate of communication over the optimal rate is $\mathcal{O}(n'/n{+}1/\sqrt{n'})$, which is at best $\mathcal{O}(n^{-1/3})$. Furthermore, there is no guarantee of performance for this protocol for a fixed sequence $(\mathbf{x}_1, ..., \mathbf{x}_m)$ observed by the parties.

In this paper, we present a protocol for omniscience, termed the *recursive data exchange protocol* (RDE), that is universal and works for individual sequences of data in the spirit of [31], namely it attains omniscience with probability close to 1 for every specific data sequence. For a given sequence $(\mathbf{x}_1, ..., \mathbf{x}_m)$ of data consisting of n length observations, RDE attains an excess communication rate of $\mathcal{O}(n^{-1/2})$ over $R\left(\mathrm{P}_{\mathbf{x}_1\cdots \mathbf{x}_m}\right)$ where $\mathrm{P}_{\mathbf{x}_1\cdots \mathbf{x}_m}$ denotes the joint type of the observations. As a consequence, we show that for the generative model where the data of the parties is IID, RDE is universally sum-rate optimal with an excess rate of $\mathcal{O}(n^{-1/2}\sqrt{\log n})$. Note that even for the case when the underlying distribution is known, the optimal rate can only be achieved asymptotically and an excess rate is often needed. In particular, for[2] $m = 2$, the precise leading asymptotic term in excess worst-case rate was established in [25] and was shown to be $\mathcal{O}(n^{-1/2})$.

An interesting application of RDE appears in *secret key* (SK) agreement [17], [1], [12]. Specifically, Csiszár and Narayan showed in [12] that an optimum-rate SK can be generated by first attaining omniscience and then extracting secure bits from the recovered data. We follow the same procedure here with RDE in place of the omniscience protocol of [12] and obtain a universal SK of rate at most $\mathcal{O}(n^{-1/2}\sqrt{\log n})$ less than the optimal average and the worst-case rate. Note that for the case $m = 2$

[2]For $m > 2$, a variant of RDE is shown in [26] to attain the optimal second-order asymptotic term, which is $O(n^{-1/2})$, for worst-case rates when the distribution is known.

with known distribution, the precise leading asymptotic term in the gap to optimal worst-case rate was established in [15] and was shown to be $\mathcal{O}(n^{-1/2})$. Therefore, for multiparty data exchange as well as SK agreement RDE can roughly attain the worst-case performance for the case of known distributions, without requiring the knowledge of the distribution. Also, for average rate, the universal $O(n^{-1/2}\sqrt{\log n})$ gap to optimal rates attained by RDE is to our knowledge the best-known.

It was shown in [30] that interaction enables an $\mathtt{ACK} - \mathtt{NACK}$ based universal variable-length coding scheme for the Slepian-Wolf problem, where only party 1 needs to send its data to party 2. Our protocol, too, is interactive in a similar spirit, but it relies on carefully increasing the rate of communication for each party. Note that while for $m = 2$ a simple extension of the protocol in [30] works for the data exchange problem as well, this is not the case when $m > 2$. For $m > 2$, the order in which the parties communicate must be carefully chosen. We give a very simple criterion for choosing this order of communication and show that the resulting protocol is universally rate-optimal. Specifically, the encoders in RDE send random hash bits corresponding their inputs, while the decoders, which use a variant of minimum entropy decoding, try to decode the observations of any subset of communicating parties. A key feature of RDE is its *recursive structure* whereby when a subset A of parties recover each-other's data, the rates appear as if the parties have been executing the protocol in an alternative model where the parties in A are collocated from the start. To enable this, the parties communicate in the order of the entropies of their empirical types, with the highest entropy party communicating first, followed by the next highest entropy party, and so on. The delay in communication between the parties is chosen to ensure that for every pair of communicating parties, the difference of their rates of communication, at any instance, is equal to the difference of the entropies of their marginal types. We follow this policy and increase the rate in steps until a subset of parties can attain *local omniscience*, $i.e.$, recover each other's data.

Our encoders are easy to implement, but the decoders are theoretical constructs which use type classes to form a list of guesses for the data of other parties. Furthermore, since we try to decode the data of every possible subset of communicating parties, the complexity of our decoder is exponential in m. Nevertheless, we believe that RDE is a stepping-stone towards a practical protocol for the multiparty data exchange problem.

There is a rich literature relating to the problems considered here. Following the seminal work of Slepian and Wolf [23], which introduced fixed-length distributed source coding for two parties, universal error-exponents for the multiparty extension of this problem were considered in [11], [9], [20]. For the case of two parties, universal variable length protocols with optimal average rate were proposed in [13], [30].

In particular, the protocol used in [30] has excess rate less than[3] $\mathcal{O}(n^{-1/2})$, which is the best-known. A related protocol was used in [25] in a single-shot setup which, when applied to IID observations with a known distribution, was shown to be of optimal worst-case length even up to the second-order asymptotic term. A slight variant of the data exchange or omniscience problem, which assumes the data of the parties to be elements of a finite field and requires exact recovery using linear communication, has been considered in [22], [24], [7], [18], [19]. While RDE doesn't directly relate to these works, we propose it as an alternative approach for ensuring data exchange in these settings.

The remainder of this paper is organized as follows: The next section contains the formal description of the omniscience problem. We first describe an idealized version of RDE, $\mathrm{RDE}_{\mathtt{id}}$, in Section III where we assume that the rates can be continuously increased and an ideal decoder is available. We also illustrate the working of $\mathrm{RDE}_{\mathtt{id}}$ with examples. Ideal assumptions are removed in the subsequent section which contains a complete description of RDE and our main results about its performance. The SK agreement problem and our universal SK agreement protocol based on RDE are described in Section V. All the proofs are given in Section VI. Our proofs rely on technical properties of the formula for minimum communication for omniscience. Some of these properties are new and maybe of independent interest.

Notations. We start by recalling the standard notations: We consider discrete random variables X taking values in a finite set $\mathcal{X}$ and with pmf P_X. Denote the set $\{1, ..., m\}$ of all parties by $\mathcal{M}$. For random variables $(X_i : i \in \mathcal{M})$ and $A \subseteq \mathcal{M}$, X_A denotes the collection $(X_i : i \in A)$. Also, X_A^n denotes the sequence of IID random variables $\{X_{A,t}\}_{t=1}^n$, where $X_{A,t} = (X_{i,t} : i \in A)$. Similarly, $\mathcal{X}_A$ denotes the product set $\prod_{i\in A} \mathcal{X}_i$ and $\mathcal{X}^n = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. For given distributions P and Q, their variational distance is denoted by $\|\mathrm{P} - \mathrm{Q}\| = \frac{1}{2}\sum_x |\mathrm{P}(x) - \mathrm{Q}(x)|$. While our protocols are conceptually simple, the analysis is notationally heavy and relies on some bespoke notations. For easy reference, we summarize all nonstandard notations used in this paper in Table I. We often need to think of a subset of parties as a single party and use natural extensions of our notations to indicate such cases. For instance, for a partition σ of $A \subseteq \mathcal{M}$ or of $\mathcal{M}$, the notation $\left(R^*_{\sigma_i}(A_\sigma) : 1 \le i \le |\sigma|\right)$ extends $R^*_i(A)$ given in Table I and denotes the solution $\left(R_1, \ldots, R_{|\sigma|}\right)$ for equations

$$\sum_{j \ne i} R_j = H\left(X_A | X_{\sigma_i}\right), \quad \forall\, 1 \le i \le |\sigma|.$$

Note that we have abused the subscript notation, with different connotations in different contexts. For instance, we use the notation A_σ for a partition σ of A, which represents the set A as a collection of

[3]For $m = 2$ even RDE has excess rate less than $\mathcal{O}(n^{-1/2})$. The extra $\mathcal{O}(\sqrt{\log n})$ factor for a general m appears since the optimal sum-rate may not be a concave function of $\mathrm{P}_{X_1 \cdots X_m}$ for $m > 2$, and we take recourse to a Taylor approximation of the sum-rate function.

elements $\sigma_i \in \sigma$. However, the specific connotation should be clear from the context.

Notation	Description
$\Sigma(A)$	Set of all nontrivial partitions of A
$\lvert\sigma\rvert$	Number of parts in the partition σ
$\sigma_f(A)$	The finest partition $\{\{i\} : i \in A\}$ of A
$\sigma_B(A)$, $B \subsetneq A$	The partition $\{\{A \setminus B\}, \{i\} : i \in B\}$ of A
R_A	Sum rate $\sum_{i \in A} R_i$
$\mathcal{R}_{\mathtt{CO}}(A)$	Set of all vectors $(R_i : i \in A)$ s.t. $R_B \geq H(X_B \mid X_{A \setminus B})$, $\forall\, B \subsetneq A$
$\mathcal{R}^{\Delta}_{\mathtt{CO}}(A)$	Set of all vectors $(R_i : i \in A)$ s.t. $R_B \geq H(X_B \mid X_{A \setminus B}) + \lvert B\rvert\Delta$, $\forall\, B \subsetneq A$
$R_{\mathtt{CO}}(A)$	Minimum of R_A over all $\mathbf{R} \in \mathcal{R}_{\mathtt{CO}}(A)$
$\mathbb{H}_\sigma(A)$, $\sigma \in \Sigma(A)$	$\frac{1}{\lvert\sigma\rvert - 1} \sum_{i=1}^{\lvert\sigma\rvert} H\left(X_A \mid X_{\sigma_i}\right)$
$R_i^*(A), i \in A$	Solution of $\sum_{j \neq i} R_j = H(X_A \mid X_i)$, $\forall\, i \in A$
A_σ, $\sigma \in \Sigma(A)$	$\{A_{\sigma_i} : 1 \leq i \leq \lvert\sigma\rvert\}$, where $A_{\sigma_i} = \sigma_i$ treated as a single party
$\lvert\pi\rvert$	Maximum number of bits communicated in any execution of the protocol π
$\lvert\pi\rvert_{\mathtt{av}}$	Expected value of the number of bits communicated in an execution of the protocol π

TABLE I: Summary of notations used in the paper.

II. OMNISCIENCE

We begin with the description of the problem for IID observations. Specifically, parties in a set $\mathcal{M} = \{1, \ldots, m\}$ observe an IID sequence $X_{\mathcal{M}}^n = (X_{\mathcal{M}1}, \ldots, X_{\mathcal{M}n})$, with the ith party observing $\{X_{it}\}_{t=1}^n$ and $X_{\mathcal{M}t} = (X_{it} : i \in \mathcal{M}) \sim \mathrm{P}_{X_{\mathcal{M}}}$ denoting the collective data at the tth time instance. The parties have access to shared public randomness (public coins) U such that U is independent jointly of $X_{\mathcal{M}}^n$. Furthermore, the ith party, $i \in \mathcal{M}$, has access to private randomness (private coins) U_i such that $U_{\mathcal{M}}$, U, and $X_{\mathcal{M}}^n$ are mutually independent. Thus, the ith party observes (X_i^n, U_i, U).

For simplicity, we restrict our exposition to *tree-protocols* ($cf.$ [16]) described below. A tree-protocol π for $\mathcal{M}$ consists of a binary tree, termed the *protocol-tree*, with the vertices labeled by the elements of $\mathcal{M}$. The protocol starts at the root and proceeds towards the leaves. When the protocol is at vertex v with label i_v, party i_v communicates a bit b_v based on its local observations $(X_{i_v}^n, U_{i_v}, U)$. The protocol proceeds to the left- or the right-child of v, respectively, if b_v is 0 or 1. The protocol terminates when it reaches a leaf, at which point each party produces an output based on its local observations and the bits communicated during the protocol, namely the transcript $\Pi = \pi(X_{\mathcal{M}}^n, U_{\mathcal{M}}, U)$. Note that for tree-protocols the set of possible transcripts is prefix-free. Also, note that the output is not included in the transcript of the protocol, but is computed locally at each party. The literature on distributed function computation often focuses on Boolean functions and includes the 1-bit output as a part of the protocol transcript (cf. [16]). This results in a negligible 1-bit loss in communication. However, including the output in the transcript in our setup makes the data exchange problem trivial since the optimal protocol shall entail each party declaring its observation.

Figure 1 shows an example of a protocol tree for $m = 3$. The label of each node represents the party which determines the communicated bit at that node; the final boxes represent the termination of the protocol, at which point an output is produced by each party.

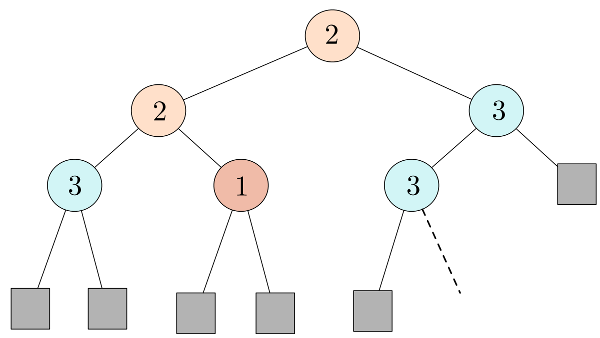

Fig. 1: A multiparty protocol tree.

The (worst-case) length $|\pi|$ of a protocol π is the maximum number of bits that are transmitted in any execution of the protocol and equals the depth of the protocol-tree. Also, the average length $|\pi|_{\texttt{av}}$ is given by the expected value of the number of bits transmitted in an execution of the protocol π.

In the omniscience problem, the parties engage in interactive communication to recover each other's data. A protocol π constitutes an ϵ-omniscience protocol if, at the end of the protocol, the ith party can output an estimate $\widehat{\mathbf{X}}_i = \widehat{\mathbf{X}}_i(X_i^n, U_i, U, \Pi) \in \mathcal{X}_{\mathcal{M}}^n$ such that

$$\mathbb{P}\left(\widehat{\mathbf{X}}_i = X_{\mathcal{M}}^n : i \in \mathcal{M}\right) \geq 1 - \epsilon.$$

Definition 1 (Communication for omniscience)**.** Given IID observations with a common distribution $\mathrm{P}_{X_{\mathcal{M}}}$ as above, for $0 \leq \epsilon < 1$, a rate $R \geq 0$ is an ϵ-achievable omniscience rate if there exists an ϵ-omniscience protocol π with length $|\pi|$ less than nR, for all n sufficiently large. The infimum over all ϵ-achievable omniscience rates is denoted by $R_\epsilon(\mathrm{P}_{X_{\mathcal{M}}})$. The *minimum rate of communication for omniscience* $R(\mathrm{P}_{X_{\mathcal{M}}})$ is given by

$$R(\mathrm{P}_{X_{\mathcal{M}}}) = \lim_{\epsilon \to 0} R_\epsilon(\mathrm{P}_{X_{\mathcal{M}}}).$$

The *minimum average rate of communication for omniscience* $R^{\mathtt{av}}(\mathrm{P}_{X_{\mathcal{M}}})$ is defined similarly by replacing length $|\pi|$ with average length $|\pi|_{\mathtt{av}}$.

The fundamental quantity $R(\mathrm{P}_{X_{\mathcal{M}}})$ was characterized in [12] as

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

Following [12], the collection of all rate vectors $\mathbf{R} = (R_1, \ldots, R_m)$ satisfying the constraints in (1), termed the CO region, will be denoted by $\mathcal{R}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right)$, and the minimum sum-rate by $R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right)$. When the distribution $\mathrm{P}_{X_{\mathcal{M}}}$ is clear from the context, we shall omit it from the notation and simply use $\mathcal{R}_{\mathtt{CO}}\left(\mathcal{M}\right)$ and $R_{\mathtt{CO}}\left(\mathcal{M}\right)$.

While the result in [12] was shown to hold only for $R(\mathrm{P}_{X_{\mathcal{M}}})$, the same characterization holds for $R^{\mathtt{av}}(\mathrm{P}_{X_{\mathcal{M}}})$ as well. Indeed, note that the set of distinct transcripts of a tree protocol π is prefix-free. Therefore, the lengths of these transcripts satisfy Kraft's inequality, and so, $H(\Pi) \leq |\pi|_{\mathtt{av}}$. By proceeding exactly as in [12], we can see that $R^{\mathtt{av}}(\mathrm{P}_{X_{\mathcal{M}}}) \geq R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$. On the other hand, clearly $R^{\mathtt{av}}(\mathrm{P}_{X_{\mathcal{M}}}) \leq R(\mathrm{P}_{X_{\mathcal{M}}}) = R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$, whereby for every distribution $\mathrm{P}_{X_{\mathcal{M}}}$, we have

$$R^{\mathtt{av}}(\mathrm{P}_{X_{\mathcal{M}}}) = R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}).$$

An alternative expression for $R_{\mathtt{CO}}(\mathrm{P}_{X_{\mathcal{M}}})$ was obtained in [12] by looking at its dual form. In fact, by leveraging on the complementary slackness property, [3], [5] showed that the optimization in the dual form can be restricted to the partitions of $\mathcal{M}$ and showed that[4]

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

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

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

Note that the fact that $R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$ is lower bounded by the right-side of (2) was shown earlier in [12]. RDE directly achieves the right-side of (2), thereby providing an alternative, "operational" proof for the tightness of this lower bound for $R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$ from [12].

While there can be several maximizers of $\mathbb{H}_\sigma$, there exists a maximizing partition which is a further partition of any other maximizing partition [4, Theorem 5.2], the finest maximizing partition; we shall call this finest maximizer of $\mathbb{H}_\sigma$ in (2) the *finest dominant partition* (FDP), which was called *fundamental*

[4]An alternative proof of (2) was provided in [4] by using techniques from submodular optimization.

partiion in [4]. The *finest partition* $\sigma_f(\mathcal{M}) := \{\{i\}, i \in \mathcal{M}\}$ plays a particularly important role in RDE. Note that when the finest partition is FDP, the optimal rate assignment is uniquely given by the solution $\mathbf{R}^* = (R_1^*, \ldots, R_m^*)$ of

$$\sum_{i \in \mathcal{M}\setminus\{j\}} R_i = H(X_{\mathcal{M}}|X_j), \quad j = 1, \ldots, m. \tag{4}$$

III. Universal protocol for omniscience under ideal assumptions

We give a universal protocol for omniscience, which, when a sequence $\mathbf{x}_{\mathcal{M}}$ is observed, will transmit communication of rate no more than $R_{\texttt{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}})$. To present the main idea behind RDE, we first describe it assuming the following ideal assumptions.

Specifically, we make two assumptions:

(a) *Continuous rate assumption:* Communication-rate, defined as the total number of bits of communication up to a certain time divided by n, can be increased continuously in time[5]; and

(b) *Ideal decoder assumption:* We assume the availability of an error-free, ideal decoder $\mathrm{DEC}_{\texttt{id}}$ which correctly decodes a sequence once sufficient communication has been sent and declares a `NACK` otherwise.[6]

A standard universal decoder used in source coding is the minimum entropy decoder which, given side-information $\mathbf{y}$ and an nR-bit[7] random hash[8] of X^n, searches for the unique sequence $\mathbf{x}$ such that the joint type $\mathrm{P}_{\overline{X}\,\overline{Y}} = \mathrm{P}_{\mathbf{xy}}$ satisfies $H\left(\overline{X}\middle|\overline{Y}\right) \leq R$ and the hash of $\mathbf{x}$ matches the received hash bits. The decoder that we prescribe in the next section works on a similar principle except that it searches for any possible subset of sequences it can decode with the current rate. To avoid the additional complications due to decoding error, we first assume the availability of an ideal decoder $\mathrm{DEC}_{\texttt{id}}$ which enables omniscience for all parties $j \in A$ as soon as the rate received from the parties in A is sufficient. That is, the ideal decoder guarantees that each party $i \in A$ can recover the correct sequence $\mathbf{x}_A$ if the rates of communication $\mathbf{R} = (R_i : i \in \mathcal{A})$ satisfy $\mathbf{R} \in \mathcal{R}_{\texttt{CO}}(A|\mathrm{P}_{\mathbf{x}_A})$. Furthermore, if $\mathbf{R} \notin \mathcal{R}_{\texttt{CO}}(A|\mathrm{P}_{\mathbf{x}_A})$, the ideal decoder does

[5]Clearly, this does not hold in practice since the number of bits of communication can be increased only in steps of discrete sizes. The continuous rate assumption allows us to examine, loosely speaking, the "fluid limit" behavior of RDE.

[6]In analysis of the ideal protocol, we do not account for the rate needed to send `NACK`s. In practice, each `NACK` symbol counts for a bit of communication and the size Δ of discrete increments must be chosen carefully to render the rate needed to send `NACK`s negligible.

[7]nR is required to be an integer. When this is not the case, we simply use $\lceil nR \rceil$ bits in place of nR. This convention will be used throughout this paper and will be accounted for in our analysis.

[8]A "random hash" of X^n is a bit sequence produced by a function $f : \mathcal{X}^n \to \{0,1\}^{nR}$ which is chosen randomly (using public randomness) from a class of functions satisfying the 2-universal property [2]. For instance, the class of all functions satisfies the 2-universal property and, therefore, standard "random binning" (cf. [8]) produces a random hash.

not mistakenly output a wrong sequence $\mathbf{x}'_A$, but declares a `NACK` instead. Protocol 1 summarizes our assumed ideal decoder $\mathrm{DEC}_{\mathtt{id}}$.

Protocol 1: Ideal decoder $\mathrm{DEC}_{\mathtt{id}}(j, \sigma, \mathbf{R})$

Input: An index $1 \le j \le m$, a partition $\sigma \in \Sigma(\mathcal{M})$, a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$.
Output: An `ACK` message $(\mathtt{ACK}, A)$ or a `NACK` message

1) For σ_i such that $j \in \sigma_i$, search for the maximal set $A \subseteq \mathcal{M}$ such that $\sigma_i \subsetneq A$ and $(R_l : l \in A) \in \mathcal{R}_{\mathtt{CO}}\left(A \mid \mathrm{P}_{\mathbf{x}_A}\right)$, and reveal $\mathbf{x}_A$ to party j.
2) **if** *If such an A was found in Step 1* **then**
   return $(\mathtt{ACK}, A)$.
   else
   return `NACK`.

With this ideal decoder at our disposal, under the continuous rates assumption, finding a universal protocol is tantamount to finding a policy for increasing the rates $(R_1, \ldots, R_m)$ such that when the rate vector enters $\mathcal{R}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{\mathbf{x}_\mathcal{M}}\right)$ for the first time, the sum-rate is $R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{\mathbf{x}_\mathcal{M}}\right)$. Note that initially the marginal types $\mathrm{P}_{\mathbf{x}_i}$ are available to each party and can be transmitted using $\mathcal{O}(\log n)$ bits, since there are only polynomially many types. Also, if a subset A attains local omniscience in the middle of the protocol, any $j \in A$ upon recovering $\mathbf{x}_A$ can transmit $\mathrm{P}_{\mathbf{x}_A}$ in $\mathcal{O}(\log n)$ bits to all the parties, who in turn can use it to compute $H(\mathrm{P}_{\mathbf{x}_A})$.

As an illustration, consider the simple case when $m = 2$. Parties first share $\mathrm{P}_{\mathbf{x}_1}$ and $\mathrm{P}_{\mathbf{x}_2}$; suppose $H(\mathrm{P}_{\mathbf{x}_1}) \ge H(\mathrm{P}_{\mathbf{x}_2})$. Then, party 1 starts communicating and increases its rate R_1 at slope[9] 1. When the rate R_1 reaches $H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_2})$, party 2 starts communicating at slope 1 as well. Throughout the protocol, each party is trying to decode the other using the ideal decoder $\mathrm{DEC}_{\mathtt{id}}$ and they keep on communicating as long as the ideal decoders output `NACK`s. The parties will decode each other as soon as (R_1, R_2) enters $\mathcal{R}_{\mathtt{CO}}\left(\{1,2\}|\mathrm{P}_{\mathbf{x}_1,\mathbf{x}_2}\right)$, *i.e.*, when

$$R_1 \ge H(\overline{X}_1|\overline{X}_2) \text{ and } R_2 \ge H(\overline{X}_2|\overline{X}_1),$$

where $(\overline{X}_1, \overline{X}_2) \sim \mathrm{P}_{\mathbf{x}_1,\mathbf{x}_2}$. Note that once both parties start communicating, the difference $R_1 - R_2$ is maintained as $H(\overline{X}_1) - H(\overline{X}_2)$. Thus, when (R_1, R_2) enters $\mathcal{R}_{\mathtt{CO}}\left(\{1,2\}\right)$, it holds that

$$R_1 = H(\overline{X}_1|\overline{X}_2) \text{ and } R_2 = H(\overline{X}_2|\overline{X}_1);$$

the red line in Figure 2 illustrates[10] this evolution of rates.

[9]The slope is defined as the derivative of rate with respect to the time under the continuous rate assumption.

[10]It is also possible to proceed along the blue line for the $m = 2$ case. However, its extension to a general m is not clear.

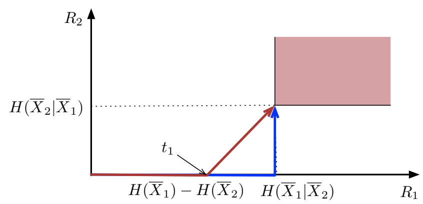

Fig. 2: Illustration of protocol for $m = 2$. The transition point t_1 depends only on the marginal types $\mathrm{P}_{\mathbf{x}_1}$ and $\mathrm{P}_{\mathbf{x}_2}$.

RDE extends the idea above to a general m. We design RDE so that the first subset A which attains local omniscience does so by using communication only from the parties in A and of sum rate

$$R_A = \mathbb{H}_{\sigma_f(A)}(A|\mathrm{P}_{\mathbf{x}_A}) = \sum_{i \in A} R_i^*(A); \tag{5}$$

see (13) in Lemma 7 given in Section VI below for the second equality. To that end, we note (see Lemma 7 for a proof) that for every A

$$R_i^*(A) - R_j^*(A) = H(\overline{X}_i) - H(\overline{X}_j). \tag{6}$$

A key point here is that for $\mathrm{P}_{\mathbf{x}_\mathcal{M}}$ this difference can be computed using only the marginal types $\mathrm{P}_{\mathbf{x}_i}$ and $\mathrm{P}_{\mathbf{x}_j}$. RDE ensures that for every pair (i, j) of communicating parties, the rate of communication

$$R_i - R_i^*(A) = R_j - R_j^*(A),$$

which by (6) in turn can be ensured if the *constant difference* property, namely

$$R_i - R_j = H(\overline{X}_i) - H(\overline{X}_j), \tag{7}$$

is maintained throughout the protocol for every pair of communicating parties. Thus, all communicating parties i reach the rate $R_i^*(A)$ at the same time. Specifically, we first arrange parties in decreasing order of the entropy of the empirical distribution of their local observations, which are shared in $\mathcal{O}(\log n)$-bits. Assuming $H(\mathrm{P}_{\mathbf{x}_1}) \geq H(\mathrm{P}_{\mathbf{x}_2}) \geq \cdots \geq H(\mathrm{P}_{\mathbf{x}_m})$, party 1 starts communicating, and the ith party starts communicating when $R_1 \geq H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_i})$. This ensures the constant difference property (7) for every pair (i, j) of communicating parties. For notational convenience, we assign -1 to R_i when the ith party has not started communicating; the rate vector $(0, -1, -1, \ldots, -1)$ indicates that party 1 starts communicating and every one else remains quiet. When a subset A attains local omniscience, we decrease the rate-slope for each party $i \in A$ to $1/|A|$, thereby ensuring that collectively parties in A increase the rate of communication R_A at slope 1. Note that since parties in A have recovered $\mathbf{x}_A$, any one party $i \in A$ can compute the type $\mathrm{P}_{\mathbf{x}_A}$ and transmit it using $\mathcal{O}(\log n)$ bits. Our main observation is that at this point the rates appear as if the parties in A were collocated to begin with and have been executing the protocol as a single party. In particular, $R_A - R_j = H(\overline{X}_A) - H(\overline{X}_j)$ for any communicating party j outside A. The second crucial observation is that for the first subset A which attains local omniscience, $(R_i^*(A) : i \in A) \in \mathcal{R}_{\texttt{CO}}(A)$. Since by (5) $\sum_{i\in A} R_i^*(A)$ is a lower bound for $\mathcal{R}_{\texttt{CO}}(A)$, the parties in A cannot attain local omniscience before they communicate at sum-rate $\sum_{i\in A} R_i^*(A)$. Further, RDE ensures that all parties in A reach the rate $R_i^*(A)$ at the same time. Thus, the parties in A must have communicated at sum-rate

$$R_A = \sum_{i\in A} R_i^*(A) = \mathbb{H}_{\sigma_f(A)}(A|\mathrm{P}_{\mathbf{X}_A}) \tag{8}$$

when they attain local omniscience. As the protocol proceeds, subsets of parties keep attaining local omniscience and start behaving as a single party. Proceeding recursively, it follows that when all parties attain omniscience, the rate of communication must equal $\mathbb{H}_\sigma(\mathcal{M}|\mathrm{P}_{\mathbf{X}_\mathcal{M}})$ for some $\sigma \in \Sigma(\mathcal{M})$, which in view of (2) is no more than $R_{\texttt{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{X}_\mathcal{M}})$ and must be optimal in the limit as $n \to \infty$.

To help the reader build heuristics for the complete protocol and its analysis, we provide a sketch of the analysis for the ideal situation and consider the ideal version $\mathrm{RDE}_{\texttt{id}}$. The formal proofs for the ideal case closely follow those for the results for the actual protocol and have been omitted. As mentioned, $\mathrm{RDE}_{\texttt{id}}$ proceeds recursively by increasing the rates with fixed slopes until a subset attains omniscience, at which point the slopes are changed so that the parties in an omniscience attaining subset behave as if they are collocated. We describe the one-step omniscience protocol $\mathrm{OMN}_{\texttt{id}}$ in Protocol 2. The protocol takes as input a partition σ such that parties in any one part are behaving as collocated parties, a vector $\mathbf{H} = (H_{\sigma_i}, 1 \le i \le |\sigma|)$ consisting of estimates of entropy for marginal distribution of parties in any part of σ, and a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$ of rates of communication sent by all the parties up to this point.

Note that a "valid" rate vector should reflect that parties in any one part have communicated enough to attain local omniscience. Also, since we shall recursively call $\mathrm{OMN}_{\texttt{id}}$, the only rate vectors $\mathrm{OMN}_{\texttt{id}}$ encounters are those which can arise by increasing the rates in the manner of RDE. We call the collection of rate vectors satisfying the two conditions above $(\sigma, \mathbf{H})$-valid. Formally,

Definition 2. For $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$ and $\mathbf{H} = (H_{\sigma_1}, \ldots, H_{\sigma_k})$ with $H_{\sigma_1} \ge H_{\sigma_2} \ge \cdots \ge H_{\sigma_k}$, a rate vector $(R_1, \ldots, R_m)$ is $(\sigma, \mathbf{H})$-valid if

$$(R_j, j \in \sigma_i) \in \mathcal{R}_{\texttt{CO}}(\sigma_i), \quad \forall\, i \text{ s.t. } |\sigma_i| \ge 2,$$

and $(R_{\sigma_i}, 1 \le i \le k)$ can be obtained by starting with $(0, -1, -1, \ldots, -1)$ and incrementing the rates as

Protocol 2: $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R})$

Input: A partition $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$, an entropy estimate vector $\mathbf{H} = (H_{\sigma_i} : 1 \le i \le k)$, a rate vector $\mathbf{R} = (R_1, \dots, R_m)$; we assume that $\mathbf{H}$ is sorted, *i.e.*, $H_{\sigma_1} \ge H_{\sigma_2} \ge \cdots \ge H_{\sigma_k}$.

Output: A rate vector $\mathbf{R}^{\mathtt{out}}$, a family of subsets $\mathcal{O}$ that have attained omniscience.

1) Initialize $s := \max\{i : R_{\sigma_i} \ge 0\}$.
2) All parties j such that $j \in \sigma_i$ for some $1 \le i \le s$ increase their rates R_j at slope $1/|\sigma_i|$.
3) **if** *There exists $i > s$ such that $R_{\sigma_1} \ge H_{\sigma_1} - H_{\sigma_i}$* **then**
   set $R_j = 0$ for all $j \in \sigma_i$, and set $s = \max\{i : R_{\sigma_i} \ge 0\}$.
4) For all j such that $j \in \sigma_i$ for some $1 \le i \le s$, execute $\mathrm{DEC}_{\mathtt{id}}(j, \sigma, \mathbf{R})$, which outputs `NACK` or $(\mathtt{ACK}, A_j)$.
5) **if** *All parties send a* `NACK` **then**
   return to Step 2.
   else
   Identify the omniscience family

$$\mathcal{O} = \{B \subset \mathcal{M} : \text{ all } j \in B \text{ returned } (\mathtt{ACK}, B)\}.$$

   Set $\mathbf{R}^{\mathtt{out}} = \mathbf{R}$ and return $(\mathbf{R}, \mathcal{O})$.

in Protocol 2 when the parties in each part σ_i are collocated, *i.e.*, each part σ_i starts increasing its rate at slope 1 once $R_{\sigma_1} \ge H_{\sigma_1} - H_{\sigma_i}$.

As mentioned earlier, instead of initializing all rates with 0 in RDE, and in the definition of a valid rate vector, we distinguish between rate 0 and rate -1 for a technical reason. A rate of -1 indicates that the party is not participating in the protocol yet and will not even attempt to decode. In contrast, a 0 rate indicates that the party has not yet communicated any bits, but has started decoding and will increment its communication rate in each step from here on.

The result below shows a recursive property of $\mathrm{OMN}_{\mathtt{id}}$ that renders RDE universally rate-optimal. Specifically, it shows that if $\mathbf{R}$ is $(\sigma, \mathbf{H})$-valid then, when $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R})$ terminates, the output rate vector is $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}})$-valid where $\sigma^{\mathtt{out}}$ is a sub-partition of σ which is obtained by combining the parts that have achieved local omniscience; $\mathbf{H}^{\mathtt{out}}$ is the corresponding estimate for entropies of the marginals of parts of $\sigma^{\mathtt{out}}$. Furthermore, for every set A that attains local omniscience, the sum-rate R_A at the end of $\mathrm{OMN}_{\mathtt{id}}$ is exactly $\mathbb{H}_{\sigma_f(A_\sigma)}(A_\sigma)$.[11]

Theorem 1. *For $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$ and $\mathbf{H} = (H_{\sigma_1}, \dots, H_{\sigma_k})$ with $H_{\sigma_1} \ge H_{\sigma_2} \ge \cdots H_{\sigma_k}$, let $\mathbf{R}^{\mathtt{in}} = (R_1^{\mathtt{in}}, \dots, R_m^{\mathtt{in}})$ be $(\sigma, \mathbf{H})$-valid. Then, if $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R}^{\mathtt{in}})$ is executed, the final rates $\mathbf{R}^{\mathtt{out}}$ and the omniscience family $\mathcal{O}$ satisfy the following:*

[11]When $A = \cup_{l=1}^{c} \sigma_{i_l}$, by our convention $\mathbb{H}_{\sigma_f(A_\sigma)}(A_\sigma) = \mathbb{H}_{\{\sigma_{i_1}|\cdots|\sigma_{i_c}\}}(A|\mathrm{P}_{\mathtt{X}_A})$.

1) Every $A \in \mathcal{O}$ consists of parts of σ, i.e.,

$$A = \bigcup_{l=1}^{c} \sigma_{i_l}$$

for some $\{i_1, \dots, i_c\} \subseteq \{1, \dots, |\sigma|\}$, and the sum-rate $R_A^{\mathtt{out}}$ satisfies

$$R_A^{\mathtt{out}} = \mathbb{H}_{\{\sigma_{i_1}|\cdots|\sigma_{i_c}\}}\left(A|\mathrm{P}_{\mathbf{x}_A}\right).$$

2) Let $\sigma^{\mathrm{out}} \in \Sigma(\mathcal{M})$ be the partition obtained by combining the parts in σ that belong to the same A in $\mathcal{O}$. Let $H_{\sigma_i^{\mathrm{out}}}$ denote the entropy of the type of $\mathbf{x}_{\sigma_i^{\mathrm{out}}}$. Then, with $\mathbf{H}^{\mathrm{out}} = \left(H_{\sigma_i^{\mathrm{out}}}, 1 \leq i \leq |\sigma^{\mathrm{out}}|\right)$, $\mathbf{R}^{\mathrm{out}}$ is $(\sigma^{\mathrm{out}}, \mathbf{H}^{\mathrm{out}})$-valid.

In fact, Theorem 1 is a special case of Theorem 2, and the proof of the former follows from that of the latter given below. However, we provide a brief sketch of the proof of Theorem 1 here to highlight the key ideas and, also, to clarify the technical proof of Theorem 2.

Proof sketch. For simplicity, assume that σ consists of singletons, $i.e.$, $\sigma = \sigma_f(\mathcal{M})$. The main component of our proof is the following claim:

Claim: *The parties in a subset A attain local omniscience exactly when each R_i, $i \in A$, reaches $R_i^*(A)$.*

As mentioned before, all communicating parties $i \in A$ reach $R_i^*(A)$ simultaneously, and the parties in A cannot attain local omniscience before this happens. The proof of the claim follows from Lemma 11 given in Section VI, since no subset of A has attained local omniscience before A.

The theorem follows. Indeed, the first assertion holds by (8). For the second assertion, we need to show that for two subsets A and B in $\mathcal{O}$, $R_A^{\mathtt{out}} - R_B^{\mathtt{out}} = H(X_A) - H(X_B)$. The complete proof considers various cases depending on if A (or B) contains a party i with nonnegative $R_i^{\mathtt{in}}$. We illustrate the proof for a case when there exist $i \in A$ and $j \in B$ with $R_i^{\mathtt{in}}, R_j^{\mathtt{in}} \geq 0$. Since $\mathbf{R}^{\mathtt{in}}$ is valid for $\sigma = \sigma_f(\mathcal{M})$ and the communicating parties maintain the difference of their rates, it follows from the claim above that

$$\begin{aligned}
R_A^{\mathtt{out}} - R_B^{\mathtt{out}} &= R_{A\setminus\{i\}}^{\mathtt{out}} - R_{B\setminus\{j\}}^{\mathtt{out}} + R_i^{\mathtt{out}} - R_j^{\mathtt{out}} \\
&= R_{A\setminus\{i\}}^{\mathtt{out}} - R_{B\setminus\{j\}}^{\mathtt{out}} + R_i^{\mathtt{in}} - R_j^{\mathtt{in}} \\
&= R_{A\setminus\{i\}}^{\mathtt{out}} - R_{B\setminus\{j\}}^{\mathtt{out}} + H(X_i) - H(X_j) \\
&= \sum_{l \in A\setminus\{i\}} R_l^*(A) - \sum_{k \in B\setminus\{j\}} R_k^*(B) + H(X_i) - H(X_j) \\
&= H(X_A|X_i) - H(X_B|X_j) + H(X_i) - H(X_j)
\end{aligned}$$

$$= H(X_A) - H(X_B).$$

Other cases can be handled similarly. Therefore, $\mathbf{R}^{\mathtt{out}}$ is valid for $\sigma^{\mathtt{out}}$. □

Thus, if we proceed by recursively calling $\mathrm{OMN}_{\mathtt{id}}$, each time with $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}}, \mathbf{R}^{\mathtt{out}})$ obtained from the previous call, we shall ultimately attain omniscience using the sum rate $\mathbb{H}_\sigma(\mathcal{M})$ for some partition σ. Since $\mathbb{H}_\sigma(\mathcal{M})$ is a lower bound for $\mathcal{R}_{\mathtt{CO}}(\mathcal{M})$ by (2), this rate must be optimal. We summarize the overall ideal protocol in Protocol 3.

Protocol 3: $\mathrm{RDE}_{\mathtt{id}}$: The recursive data exchange protocol under ideal conditions

1) Initialize $\sigma = \sigma_f(\mathcal{M})$, $\mathbf{R} = (0, -1, -1, \ldots, -1)$, $k = |\sigma|$.
2) **while** $k > 1$ **do**
   (i) For $1 \leq i \leq k$, a party $j \in \sigma_i$ computes $\mathrm{P}_{\mathbf{x}_{\sigma_i}}$ and broadcasts it.
   Each party computes $H_{\sigma_i} = H\left(\mathrm{P}_{\mathbf{x}_{\sigma_i}}\right)$, $1 \leq i \leq k$.
   (ii) Let $\mathbf{H}$ be the sorted version of $(H_{\sigma_i} : 1 \leq i \leq k)$, *i.e.*, assume $H_{\sigma_1} \geq H_{\sigma_2} \geq \cdots \geq H_{\sigma_k}$.
   Call $\mathrm{OMN}_{\mathtt{id}}(\sigma, \mathbf{H}, \mathbf{R})$.
   Let $(\mathbf{R}^{\mathtt{out}}, \mathcal{O})$ be its output.
   (iii) Let $\sigma^{\mathtt{out}} = \{\sigma_i : \sigma_i \in \sigma \text{ s.t. } \sigma_i \not\subset A \ \forall A \in \mathcal{O}\} \bigcup \{A : A \in \mathcal{O}\}$.
   Update $\mathbf{R} = \mathbf{R}^{\mathtt{out}}$, $\sigma = \sigma^{\mathtt{out}}$, and $k = |\sigma^{\mathtt{out}}|$.

Remark 1. Recently, it was shown in [6] that if a set A corresponds to a part in the partition that attains the maximum in (2), then omniscience can be attained in such a manner that the parties in A can attain omniscience along the way from the communication of the parties in A. RDE explicitly has this feature and attains omniscience for each part of the maximizing partition along the way.

We conclude this section with a few illustrative examples to demonstrate the working of the ideal version $\mathrm{RDE}_{\mathtt{id}}$. The first example is for $m = 3$ and exhibits a case where $\sigma_f(\mathcal{M})$ is the FDP.

Example 1. Let $X_1 \sim \mathsf{Ber}(1/2)$, $X_3 \sim \mathsf{Ber}(q)$, and $X_2 = X_1 \oplus X_3$. In this case, $\mathcal{R}_{\mathtt{CO}}(\{1, 2, 3\})$ is given by rate vectors satisfying the following linear constraints:

$$R_1 + R_2 \geq 1,$$

$$R_2 + R_3 \geq h(q),$$

$$R_1 + R_3 \geq h(q).$$

When $\frac{1}{2} < h(q) \leq 1$, the finest partition is the FDP, and

$$R_{\mathtt{CO}}(\{1, 2, 3\}) = \mathbb{H}_{\{1|2|3\}} = \frac{1 + 2h(q)}{2}.$$

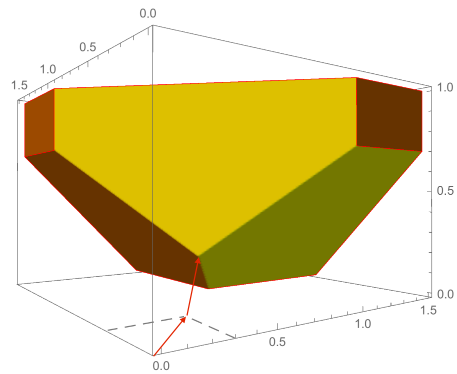

Fig. 3: Illustration of $\mathcal{R}_{\texttt{CO}}(\{1,2,3\})$ for Example 1.

The CO region is depicted in Figure 3. As can be seen from the figure, $R_{\texttt{CO}}(\{1,2,3\})$ is achieved by the unique rate assignment $\mathbf{R}^* = (1/2, 1/2, (2h(q)-1)/2)$. In $\mathrm{RDE}_{\texttt{id}}$, parties 1 and 2 communicate first and increase their rates at slope 1 until $R_1 = R_2 = H(X_1) - H(X_3) = H(X_2) - H(X_3) = 1 - h(q)$. At this point, party 3 starts communicating and all the parties increase their rates at slope 1. Owing to the initial lead of R_1 and R_2 over R_3, all the parties reach $\mathbf{R}^*$ simultaneously.

When $\mathbb{H}_\sigma$ is maximized by a partition σ other than the finest partition $\sigma_f(\mathcal{M})$, as $\mathrm{RDE}_{\texttt{id}}$ proceeds, the parties in parts of σ attain local omniscience, along the way, before all the parties attain omniscience. Consider the following example, again for $m = 3$.

Example 2. Let $W_1, W_2 \sim \mathsf{Ber}(1/2)$ and $V_1, V_2 \sim \mathsf{Ber}(q)$ for some $0 < q < \frac{1}{2}$, and let $X_1 = (W_1, W_2)$, $X_2 = (W_1 \oplus V_1, W_2)$, and $X_3 = W_2 \oplus V_2$. In this case, the partition $\{12|3\}$ is the FDP, $\mathbb{H}_{\{12|3\}} = 1 + 3h(q)$, and $\mathrm{RDE}_{\texttt{id}}$ proceeds as follows: Parties 1 and 2 start increase their rates at slope 1. When their rates reach $h(q)$, they attain local omniscience. At this point they start increasing their rates at slope $1/2$ and continue doing so until $R_1 + R_2$ reaches $H(X_1, X_2) - H(X_3) = 1 + h(q)$. Now, party 3 starts communicating at slope 1. When all the parties reach $((1+2h(q))/2, (1+2h(q))/2, h(q))$, they attain omniscience.

Note that $\{1,2\}$ attain local omniscience even before 3 starts communicating, illustrating the recursive structure of $\mathrm{RDE}_{\texttt{id}}$ wherein a subset attaining local omniscience start behaving as if the parties in it were collocated to begin with. In fact, this recursive property holds even when only a subset of communicating parties attains omniscience, as our final example with $m = 4$ illustrates. The situation for $m = 4$ captures the typical case for our general analysis – establishing the recursive nature of the protocol at situations similar to that illustrated by the point t_3 in Figure 4 constitutes the main step in our analysis.

Example 3. Let $W_1, W_2, W_3 \sim \mathsf{Ber}(1/2)$ and $V_1, V_2 \sim \mathsf{Ber}(q)$ for some $0 < q < \frac{1}{2}$, and let $X_1 =$

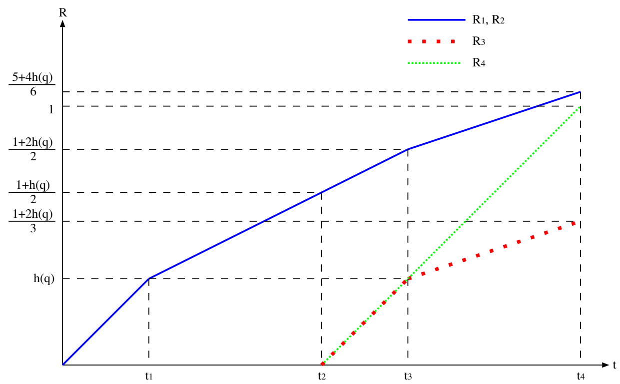

Fig. 4: The evolution of rates for Example 3.

(W_1, W_2), $X_2 = (W_1 \oplus V_1, W_2)$, $X_3 = W_2 \oplus V_2$, and $X_4 = W_3$. Note that the observations of subset $\{1,2,3\}$ are exactly as in Example 2. In this case, the partition $\{123|4\}$ is the FDP, $\mathbb{H}_{\{123|4\}} = 3 + 2h(q)$, and $\mathrm{RDE}_{\mathtt{id}}$ proceeds as in Figure 4. At t_1, parties 1 and 2 attain local omniscience and change the slopes of R_1 and R_2 to $1/2$. At t_2, parties 3 and 4 start communicating. At t_3, parties in $\{1,2,3\}$ attain local omniscience and change their slope to $1/3$. Note that up to t_3 the evolution of (R_1, R_2, R_3) is exactly the same as that in Example 2. Also, at t_3 the rate difference $(R_1 + R_2 + R_3 - R_4)$ equals $H(X_1, X_2, X_3) - H(X_4) = 1 + 2h(q)$. Thus, after t_3 the rate pair $(R_1 + R_2 + R_3, R_4)$ behaves as if the parties in $\{1,2,3\}$ were collocated to begin with. Finally, all parties attain omniscience at t_4.

IV. Universal protocol for omniscience: Full description

Moving now to the real world, rates must be increased in discrete increments and a positive decoding error probability must be tolerated. To that end, the parties incrementally transmit independent hash bits, $n\Delta$ at a time. The ideal decoder of the previous section is replaced with a typical decoder $\mathrm{DEC}(j, \sigma, \mathbf{R})$ which searches for the maximal set A such that there exists a unique sequence $\mathbf{x}_A$ that contains the current rate vector in its CO region and is consistent with the local observation and the received hash values. In fact, instead of working with the original CO region $\mathcal{R}_{\mathtt{CO}}(A)$, we use the more restrictive region $\mathcal{R}^{\Delta}_{\mathtt{CO}}(A)$ consisting of vectors $(R_i, i \in \mathcal{A})$ such that

$$R_B \geq H(X_B \mid X_{A \setminus B}) + |B|\Delta, \quad \forall\, B \subsetneq A.$$

The complete decoder is described in Protocol 4.

Note that the decoder declares $(\mathtt{ACK}, A)$ if it can find a unique maximal set A and a unique sequence $\mathbf{x}_A$, declares $\mathtt{NACK}$ if it finds no such set, or an $\mathtt{ERR}$ otherwise. In fact, an error may occur even when it is not detected, *i.e.*, when $\mathtt{ERR}$ is not transmitted. However, we can identify an event $\mathcal{E}$ (described formally

Protocol 4: $\mathrm{DEC}(j, \sigma, \mathbf{R})$

Input: An index $1 \leq j \leq m$, a partition $\sigma \in \Sigma(\mathcal{M})$, a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$

Output: A NACK message, an ACK message $(\texttt{ACK}, A)$, or an error message ERR.

1) For σ_i such that $j \in \sigma_i$, find the maximal set $A \subseteq \mathcal{M}$ such that $\sigma_i \subsetneq A$ and there exists a unique sequence $\hat{\mathbf{x}}_A$ such that the hashes of $\hat{\mathbf{x}}_A$ match all the previously received hashes from parties in $A \setminus \{j\}$ and the joint type $\mathrm{P}_{\overline{X}_A}$ of $\hat{\mathbf{x}}_A$ satisfies the following:
   (i) $\mathrm{P}_{\overline{X}_j} = \mathrm{P}_{\mathbf{x}_j}$, and
   (ii) $(R_i : i \in A) \in \mathcal{R}^{\Delta}_{\texttt{CO}}\left(A \mid \mathrm{P}_{\overline{X}_A}\right)$.
2) **if** *there is a unique maximal A found in Step 1* **then**
   return $(\texttt{ACK}, A)$.
   else if *there is no sequence found in Step 1 for any set A* **then**
   return NACK.
   else if *there are multiple As found or multiple sequences $\hat{\mathbf{x}}_A$ are found for any A in Step 1* **then**
   return ERR.

in Section VI-B) of small probability such that under $\mathcal{E}^c$ the real decoder DEC behaves exactly like $\mathrm{DEC}_{\texttt{id}}$, but with $\mathcal{R}_{\texttt{CO}}(\mathcal{A})$ replaced with $\mathcal{R}^{\Delta}_{\texttt{CO}}(\mathcal{A})$. Therefore, omniscience can be achieved in a similar manner as the ideal protocol of the previous section.

The main component of RDE is the one step omniscience protocol OMN described in Protocol 5, which uses DEC for decoding. Protocol OMN proceeds very much like the ideal protocol except that a new party i starts communicating when $R_1 \geq H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_i}) + \alpha\Delta$, where $\alpha \in \mathbb{N}$ is an increasing threshold parameter which is updated as the protocol proceeds. Throughout the protocol, a rate $R_i = -1$ indicates that the ith party is not yet transmitting and only parties with $R_i \geq 0$ communicate. The decoder tries to attain omniscience only among the communicating parties.

The ideal protocol of the previous section works due to its recursive structure whereby when a subset A attains local omniscience, the rate vector appears as if the parties in A have been collocated from the start. Moreover, the first subset to attain local omniscience does so by using a communication of rate $\mathbb{H}_{\sigma_f(A)}$. Both these properties were captured by Theorem 1. The result below establishes a similar recursive property of OMN. However, the definition of "validity" needs to be modified from Definition 2 – in place of the operational definition in the ideal case, we use the more technical definition below which captures all the key features that we need.

Definition 3. For $\alpha \in \mathbb{N}$, $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$ and $\mathbf{H} = (H_{\sigma_1}, \ldots, H_{\sigma_k})$, a rate vector $(R_1, \ldots, R_m)$ is $(\sigma, \mathbf{H}, \alpha)$-valid if, for $s = \max\{i : R_{\sigma_i} \geq 0\}$, the following conditions hold:

(i) *(Approximate constant difference)* For $1 \leq i, j \leq s$,

$$R_{\sigma_i} - R_{\sigma_j} \leq H_{\sigma_i} - H_{\sigma_j} + \alpha\Delta;$$

Protocol 5: $\mathrm{OMN}(\sigma, \alpha, \mathbf{H}, \mathbf{R})$

Input: A partition $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$, an $\alpha \in \mathbb{N}$, an entropy estimate vector $\mathbf{H} = (H_{\sigma_i} : 1 \le i \le k)$, a rate vector $\mathbf{R} = (R_1, \ldots, R_m)$; we assume that $\mathbf{H}$ is sorted, *i.e.*, $H_{\sigma_1} \ge H_{\sigma_2} \ge \cdots \ge H_{\sigma_k}$

Output: A rate vector $\mathbf{R}^{\mathtt{out}}$, a family of subsets $\mathcal{O}$ that have attained omniscience.

1) Initialize $s := \max\{i : R_{\sigma_i} \ge 0\}$.
2) All parties j such that $j \in \sigma_i$ for some $1 \le i \le s$ send $\lceil n\Delta/|\sigma_i| \rceil$ random hash bits. Update $R_j \to R_j + \Delta/|\sigma_i|$.
3) **if** *There exists* $i > s$ *such that* $R_{\sigma_1} \ge H_{\sigma_1} - H_{\sigma_i} + \alpha\Delta$ **then**
   set $R_j = 0$ for all $j \in \sigma_i$, and set $s = \max\{i : R_{\sigma_i} \ge 0\}$.
4) For all j such that $j \in \sigma_i$ for some $1 \le i \le s$, execute $\mathrm{DEC}(j, \sigma, \mathbf{R})$, which outputs `NACK`, $(\mathtt{ACK}, A_j)$, or `ERR`.
5) **if** *All parties send a* `NACK` **then**
   return to Step 2.
   else if *No party declares an* `ERR` *and some parties declare an* `ACK`*,* **then**
   Identify the omniscience family

$$\mathcal{O} = \{B \subset \mathcal{M} : \text{ all } j \in B \text{ returned } (\mathtt{ACK}, B)\}.$$

   if $\mathcal{O}$ *is nonempty* **then**
      Set $\mathbf{R}^{\mathtt{out}} = \mathbf{R}$, and return $(\mathbf{R}, \mathcal{O})$.
   else
      declare an error.
   else
   declare an error.

(ii) *(Noncommunicating parties)*

$$R_{\sigma_1} < H_{\sigma_1} - H_{\sigma_{s+1}} + \alpha\Delta; \tag{9}$$

(iii) *(Combined parties)* $\forall\, 1 \le i \le k$ such that $|\sigma_i| \ge 2$,

$$(R_j : j \in \sigma_i) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_i); \tag{10}$$

(iv) *(Separate parts)* for all $A \subseteq \{1, \ldots, k\}$ with $|A| \ge 2$,

$$(R_j : j \in \sigma_i, i \in A) \notin \mathcal{R}^{\Delta}_{\mathtt{CO}}\left(\bigcup_{i \in A} \sigma_i\right).$$

The constant difference condition is crucial for ensuring the recursive nature of RDE under ideal conditions. In general, since the rates must be incremented in discrete steps, the approximate version in Condition (i) has been introduced in the place of the original constant difference condition. For noncommunicating parties, Condition (ii) must be satisfied so that Condition (i) is maintained for those parties in future rounds when they start communicating. Condition (iii) ensures that the current rates are enough for parties in each part to attain local omniscience, while Condition (iv) ensures that σ is the maximal partition such that the parties in each part can attain local omniscience at current rates.

The following theorem captures our key observation about OMN; its proof is given in Section VI-B.

Theorem 2. *For $\alpha \in \mathbb{N}$, $\sigma \in \Sigma(\mathcal{M})$ with $|\sigma| = k$ and $\mathbf{H} = (H_{\sigma_1}, \ldots, H_{\sigma_k})$ with $H_{\sigma_1} \geq H_{\sigma_2} \geq \cdots H_{\sigma_k}$, let $\mathbf{R}^{\mathtt{in}} = (R_1^{\mathtt{in}}, \ldots, R_m^{\mathtt{in}})$ be $(\sigma, \mathbf{H}, \alpha)$-valid. Then, if $\mathrm{OMN}(\sigma, \alpha, \mathbf{H}, \mathbf{R}^{\mathtt{in}})$ is executed and error $\mathcal{E}$ (defined in Section VI-B) does not occur, the final rates $\mathbf{R}^{\mathtt{out}}$ and the omniscience family $\mathcal{O}$ satisfy the following:*

(I) For every $A \in \mathcal{O}$, it holds that

a) A consists of parts of σ, i.e.,

$$A = \bigcup_{l=1}^{c} \sigma_{i_l}$$

for some $\{i_1, \ldots, i_c\}$, and

b) denoting by A_σ the set $\{\sigma_{i_1}, \ldots, \sigma_{i_c}\}$, we have

$$R^*_{\sigma_{i_l}}(A_\sigma) - 2\alpha\Delta \leq R^{\mathtt{out}}_{\sigma_{i_l}} \leq R^*_{\sigma_{i_l}}(A_\sigma) + (m + 2\alpha)\Delta, \ 1 \leq l \leq c.$$

(II) Let $\sigma^{\mathtt{out}} \in \Sigma(\mathcal{M})$ be the partition obtained by combining the parts in σ that belong to the same A in $\mathcal{O}$. Let $H_{\sigma_i^{\mathtt{out}}}$ denote the entropy of the type of $\mathbf{x}_{\sigma_i^{\mathtt{out}}}$. Then, with $\mathbf{H}^{\mathtt{out}} = \left(H_{\sigma_i^{\mathtt{out}}}, 1 \leq i \leq |\sigma^{\mathtt{out}}|\right)$, $\mathbf{R}^{\mathtt{out}}$ is $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}}, c'_m\alpha)$-valid, where c'_m is a constant depending only on m.

We are now in a position to describe RDE. We begin by calling OMN with $\sigma = \sigma_f(\mathcal{M})$, $\alpha = 1$, the sorted entropy estimates $\mathbf{H}$ computed from marginal empirical distributions $\mathrm{P}_{\mathbf{x}_i}$, and the rate vector $\mathbf{R} = (0, -1, \ldots, -1)$ indicating that party 1 starts communicating and every one else remains quiet. Note that $\mathbf{R}$ is $(\sigma, \mathbf{H}, 1)$-valid. A new party i starts communicating when $R_1 \geq H_1 - H_i + \Delta$. If no error occurs, OMN will terminate when a subset A attains omniscience. In view of Theorem 2, at this point R_A should be close to $\mathbb{H}_{\sigma_f(A)}(A|\mathrm{P}_{\mathbf{x}_A})$ and the rates will be $(\sigma^{\mathtt{out}}, \mathbf{H}^{\mathtt{out}}, c'_m\alpha)$-valid. Thus, we are in a similar situation as the first call to OMN except that α must be replaced by $c'_m\alpha$ and the parties in a single part of $\sigma^{\mathtt{out}}$ are behaving as if they are collocated. The protocol proceeds by calling OMN again with these updated parameters. Note that under $\mathcal{E}^c$, any party $j \in A$ for $A \in \mathcal{O}$ can correctly compute $\mathrm{P}_{\mathbf{x}_A}$ and transmit it using $O(\log n)$ bits. Proceeding recursively in this manner, the protocol stops when parties in $\mathcal{M}$ attain omniscience, which by Theorem 2 can only happen when the sum-rate $R_\mathcal{M}$ is close to $\mathbb{H}_\sigma(\mathcal{M}|\mathrm{P}_{\mathbf{x}_\mathcal{M}})$ for some partition σ of $\mathcal{M}$. Thus, omniscience will be attained in communication of rate roughly less than $R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{x}_\mathcal{M}})$. We formally describe RDE in Protocol 6 and summarise its performance in Theorem 3.

We close with the following result claiming the universal rate optimality of RDE for every IID

Protocol 6: RDE: The recursive data exchange protocol

1) Initialize $\sigma = \sigma_f(\mathcal{M})$, $\mathbf{R} = (0, -1, -1, \ldots, -1)$, $k = |\sigma|$, $\alpha = 1$.
2) **while** $k > 1$ **do**
   (i) For $1 \leq i \leq k$, a party $j \in \sigma_i$ computes $\mathrm{P}_{\mathbf{x}_{\sigma_i}}$ and broadcasts it.
   Each party computes $H_{\sigma_i} = H\left(\mathrm{P}_{\mathbf{x}_{\sigma_i}}\right)$, $1 \leq i \leq k$.
   (ii) Let $\mathbf{H}$ be the sorted version of $(H_{\sigma_i} : 1 \leq i \leq k)$, *i.e.*, assume $H_{\sigma_1} \geq H_{\sigma_2} \geq \cdots \geq H_{\sigma_k}$.
   Call $\mathrm{OMN}(\sigma, \alpha, \mathbf{H}, \mathbf{R})$.
   if *There is no error declared* **then**
   let $(\mathbf{R}^{\mathtt{out}}, \mathcal{O})$ be its output.
   else
   Terminate.
   (iii) Let $\sigma^{\mathtt{out}} = \{\sigma_i : \sigma_i \in \sigma \text{ s.t. } \sigma_i \not\subset A \ \forall A \in \mathcal{O}\} \bigcup \{A : A \in \mathcal{O}\}$.
   Update $\mathbf{R} = \mathbf{R}^{\mathtt{out}}$, $\sigma = \sigma^{\mathtt{out}}$, $k = |\sigma^{\mathtt{out}}|$, and $\alpha \to c'_m \alpha$.

distribution. Proof is a simple consequence of Theorem 2 and is given in Section VI. Note that while Protocol 6 is a variable length protocol, its fixed length variant can be obtained simply by aborting the protocol once the total number of bits communicated crosses nR.

Theorem 3. *There exist constants $C_i > 0, i = 1, \ldots, 4$ depending only on m and a polynomial $p(n)$ depending on $\mathcal{X}_i$, $i \in \mathcal{M}$, such that for every $\Delta > 0$ and every sequence $\mathbf{x}_{\mathcal{M}}$, the probability of error for Protocol 6 is bounded above by*

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

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

$$nR_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}) + nC_2\Delta + C_3 \left(\frac{\log |\mathcal{X}_{\mathcal{M}}|}{\Delta} + m \right) + C_4 \log n. \tag{11}$$

Corollary 4. *For $\Delta = \frac{1}{\sqrt{n}}$ and every distribution $\mathrm{P}_{X_{\mathcal{M}}}$, Protocol 6 has a probability of error ϵ_n vanishing to 0 as $n \to \infty$ and average length $|\pi|_{\mathtt{av}}$ less than*[12]

$$nR_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) + \mathcal{O}(\sqrt{n \log n}).$$

Furthermore, for a fixed $R > 0$, the fixed-length variant of Protocol 6 has probability of error ϵ_n vanishing to 0 as $n \to \infty$ for all distributions $\mathrm{P}_{X_{\mathcal{M}}}$ that satisfy

$$R > R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) + \mathcal{O}\left(\sqrt{n^{-1} \log n}\right).$$

[12]The constant implied by $\mathcal{O}(\sqrt{n \log n})$ depends on $\mathrm{P}_{X_{\mathcal{M}}}$; see (43) below.

V. Universal secret key agreement

Closely related to the omniscience problem is the SK agreement problem where the parties seek to generate shared random bits which are almost independent of the communication used to generate them. Specifically, an (ϵ, δ)-SK agreement protocol consists of an interactive communication protocol π with public randomness U, private randomness U_i at Party i, and with the output of the ith party $K_i = K_i(X_i^n, U_i, U, \Pi)$ such that there exists a $\mathcal{K}$-valued random variable K satisfying the *recoverability condition*

$$\mathbb{P}\left(K_i = K, \forall\, i \in \mathcal{M}\right) \geq 1 - \epsilon,$$

and the *secrecy condition*[13]

$$\|\mathrm{P}_{K\Pi U} - \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{\Pi U}\| \leq \delta,$$

where $\mathrm{P}_{\mathtt{unif}}$ denotes the uniform distribution on $\mathcal{K}$.

Definition 4 (Secret key capacity)**.** For $\epsilon, \delta \in [0, 1)$, a rate $R \geq 0$ is an (ϵ, δ)-achievable SK rate if there exists a $\mathcal{K}^{(n)}$-valued (ϵ, δ)-SK with $\log |\mathcal{K}^{(n)}| \geq nR$ for all n sufficiently large. The supremum over all (ϵ, δ)-achievable SK rates is called the (ϵ, δ)-SK capacity, denoted $C_{\epsilon,\delta}(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}})$. The SK capacity for $\mathrm{P}_{X_\mathcal{M}}$ is given by

$$C(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}) = \lim_{\epsilon+\delta \to 0} C_{\epsilon,\delta}(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}).$$

Theorem 5 ([12])**.** *Given a distribution* $\mathrm{P}_{X_\mathcal{M}}$*,*

$$C\left(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}\right) = H\left(X_\mathcal{M}\right) - R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}\right).$$

In fact, it was shown in [27], [28] that a strong converse holds and $C_{\epsilon,\delta}(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}) = C(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}})$ for all $\epsilon + \delta < 1$.

The achievability of rate $H\left(X_\mathcal{M}\right) - R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}\right)$ was shown in [12] by establishing a connection between SK agreement and omniscience. In particular, a SK achieving capacity was generated by first communicating at rate $R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}}\right)$ to attain omniscience, and then extracting a SK from $X_\mathcal{M}^n$ which is almost independent of the communication used for omniscience. Following the same methodology, we provide a universal SK agreement protocol which builds upon the universal omniscience protocol of the previous section.

[13] We assume that the public randomness U is available to the eavesdropper.

We consider a slight generalization of the definition of SK above, which admits variable length SKs. An (ϵ, δ)-SK K and its estimates $K_1, \ldots, K_m$ now take values in $\mathcal{K} = \{0,1\}^*$, the set of finite length binary sequences. The recoverability condition remains as before. However, the secrecy condition needs to be modified. Specifically, denoting by T the random length of K, which we assume to be available to the eavesdropper, the secrecy condition now requires

$$\sum_t \mathrm{P}_T(t) \left\| \mathrm{P}_{K\Pi U|T=t} - \mathrm{P}_{\mathtt{unif},t} \times \mathrm{P}_{\Pi U|T=t} \right\| \leq \delta,$$

where $\mathrm{P}_{\mathtt{unif},t}$ denotes the uniform distribution on $\{0,1\}^t$. The average achievable rate and average SK capacity are defined as above with the worst-case length $\log|\mathcal{K}|$ replaced by the average length $\mathbb{E}[T]$. Instead of introducing a new notation for average SK capacity, we note that it equals $C(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$ and, with an abuse of notation, use $C(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$ to denote both the SK capacity and the average SK capacity. Indeed, the achievability is the same as above since a fixed length SK constitutes a variable length SK. For the converse, denoting

$$\epsilon_t := 1 - \mathbb{P}(K = K_i, i \in \mathcal{M}|T=t) \ \text{ and } \delta_t := \left\| \mathrm{P}_{K\Pi U|T=t} - \mathrm{P}_{\mathtt{unif},t} \times \mathrm{P}_{\Pi U|T=t} \right\|,$$

it follows by applying the converse proof of [12] for each fixed value $T = t$ that

$$\frac{t}{n} \leq C(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) + g_1(\epsilon_t) + g_2(\delta_t),$$

where g_1 and g_2 are concave, increasing functions satisfying $g_i(x) \to 0$ as $x \to 0$. Thus,

$$\begin{aligned}
\frac{\mathbb{E}[T]}{n} &\leq C(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) + \mathbb{E}[g_1(\epsilon_T) + g_2(\delta_T)] \\
&\leq C(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) + g_1(\mathbb{E}[\epsilon_T]) + g_2(\mathbb{E}[\delta_T]) \\
&\leq C(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) + g_1(\epsilon) + g_2(\delta),
\end{aligned}$$

where the last two inequalities hold since g_i, $i = 1, 2$, are concave and increasing.

We present a universal SK agreement protocol that generates a SK of average length $nC(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) - \mathcal{O}(\sqrt{n \log n})$ without the knowledge of the underlying distribution $\mathrm{P}_{X_{\mathcal{M}}}$. Specifically, first the parties use Protocol 6 with $\Delta = 1/\sqrt{n}$ to recover $X_{\mathcal{M}}^n$. If no error occurs and the recovered sequence is $\mathbf{x}_{\mathcal{M}}$, by Theorem 3 the number of bits communicated is no more than

$$l(\mathbf{x}_{\mathcal{M}}) = nR_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}) + \mathcal{O}(\sqrt{n}).$$

We extract a SK from recovered $\mathbf{x}_{\mathcal{M}}^n$ by randomly hashing[14] it to roughly $nH(\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}) - l(\mathbf{x}_{\mathcal{M}})$ values. Formal description of the protocol is given in Protocol 7; the length of the SK is tuned to the secrecy parameter δ.

Protocol 7: A universal SK agreement protocol

Input: Step size parameter Δ and secrecy parameter δ

1) Parties execute Protocol 6 with step-size Δ.
2) **if** *Protocol 6 completes without declaring an error* **then**

   Protocol 6.

   Each party $i \in \mathcal{M}$ forms an estimate K_i of the SK as follows:

   (i) Denoting by $\mathrm{P}^{(i)}$ the type of the estimate $\mathbf{x}_{\mathcal{M}}^{(i)}$ of $X_{\mathcal{M}}^n$ at Party i and by ϵ_n the maximum error probability of Protocol 6, set $l\left(\mathrm{P}^{(i)}\right)$ to be the quantity in (11) for $\mathrm{P}^{(i)}$ and

$$k\big(\mathrm{P}^{(i)}\big) = nH\big(\mathrm{P}^{(i)}\big) - l\big(\mathrm{P}^{(i)}\big) - |\mathcal{X}_{\mathcal{M}}|\log(n+1) - 2\log\frac{1}{\delta - 2\epsilon_n} + 2;$$

   (ii) generate K_i by randomly hashing $\mathbf{x}_{\mathcal{M}}^{(i)}$ to $k\left(\mathrm{P}^{(i)}\right)$ bits.

   else

   Declare an error.

Theorem 6. *For $\Delta = \frac{1}{\sqrt{n}}$, $0 < \delta < 1$, and every distribution $\mathrm{P}_{X_{\mathcal{M}}}$, Protocol 7 generates a variable length (ϵ_n, δ)-SK with ϵ_n vanishing to 0 as $n \to \infty$ and average length greater than*

$$nC(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) - \mathcal{O}(\sqrt{n\log n}). \tag{12}$$

VI. Technical results and proofs

This section contains the proofs of our results. We begin by noting a few properties of the mathematical quantities involved in our proofs.

A. Properties of CO region and related quantities

With a general subset $A \subseteq \mathcal{M}$ in the role of $\mathcal{M}$, we define the notations $R_i^*(A)$ and $\mathbb{H}_\sigma(A)$, $\sigma \in \Sigma(A)$ in a similar manner as in (4) and (3), respectively. Our first lemma notes some simple properties of $\mathbb{H}_\sigma(A)$ and $R_i^*(A)$.

Lemma 7. *For $A \subseteq \mathcal{M}$ and $\sigma \in \Sigma(A)$, the following relations hold between $R_i^*(A)$ and $\mathbb{H}_\sigma(A)$:*

$$\sum_{i \in A} R_i^*(A) = \mathbb{H}_{\sigma_f}(A); \tag{13}$$

[14]The random hash can be replaced by a randomly selected member of a 2-universal hash family.

$$R_i^*(A) = \mathbb{H}_{\sigma_f}(A) - H(X_A|X_i), \quad \forall i \in A; \tag{14}$$

$$\sum_{i \in B} R_i^*(A) - H(X_A|X_{A\setminus B}) = |B|\Big[\mathbb{H}_{\sigma_f}(A) - \mathbb{H}_{\sigma_B}(A)\Big], \tag{15}$$

where the final equality holds for every $B \subsetneq A$, with the shorthand σ_B for the partition $\sigma_B(A) \in \Sigma(A)$ given by $\{\{A \setminus B\}, \{i\} : i \in B\}$.

Furthermore, $R_i^*(A)$ satisfies the following properties:

$$\sum_{j \in A} R_j^*(A) - R_i^*(A) = H(X_A|X_i), \quad \forall\, i \in A; \tag{16}$$

$$R_i^*(A) - R_j^*(A) = H(X_i) - H(X_j), \quad \forall\, i, j \in A. \tag{17}$$

Finally, for $A \subseteq \mathcal{M}$ and $\sigma \in \Sigma(A)$, similar results holds for $R_{\sigma_i}^*(A_\sigma)$, $1 \le i \le |\sigma|$, with A_σ in place of A.

Proof. Since $(R_i^*(A) : i \in A)$ is the solution of

$$\sum_{\substack{j \in A \\ j \neq i}} R_j = H(X_A|X_i), \ i \in A,$$

by taking the summation of all the constraints and by dividing by $|A| - 1$, we have (13). Then, by subtracting the constraint for i from (13), we have (14). From (14), for every $B \subsetneq A$ it holds that

$$\sum_{i \in B} R_i^*(A) = |B|\mathbb{H}_{\sigma_f}(A) - \sum_{i \in B} H(X_A|X_i). \tag{18}$$

Also,

$$\mathbb{H}_{\sigma_B}(A) = \frac{1}{|B|}\left[\sum_{i \in B} H(X_A|X_i) + H(X_A|X_{A\setminus B})\right],$$

which is equivalent to

$$\sum_{i \in B} H(X_A|X_i) = |B|\mathbb{H}_{\sigma_B}(A) - H(X_A|X_{A\setminus B}).$$

Combining this with (18), we have (15).

By taking the difference of (13) and (14), we have (16); (17) also follows from (14). The final statement is proved exactly in the same manner by regarding X_{σ_i} as a single random variable. □

Next, we prove another useful relation between $\mathbb{H}$ and R_i^* showing that the difference $\sum_{i \in B} R_i^*(A) - \mathbb{H}_{\sigma_f}(B)$ must have the same sign as $\mathbb{H}_{\sigma_{\overline{B}}}(A) - \mathbb{H}_{\sigma_f}(A)$, where $\overline{B}$ denotes $A \setminus B$ and, as before, we have used the shorthand σ_B for the partition $\sigma_B(A)$ of A.

Lemma 8. *For every $B \subsetneq A \subseteq \mathcal{M}$ with $\overline{B} = A \backslash B$,*

$$\sum_{i \in B} R_i^*(A) = \mathbb{H}_{\sigma_f}(B) + \frac{|B||\overline{B}|}{|B|-1}\Big[\mathbb{H}_{\sigma_{\overline{B}}}(A) - \mathbb{H}_{\sigma_f}(A)\Big]. \tag{19}$$

For $A \subseteq \mathcal{M}$ and $\sigma \in \Sigma(A)$, similar results holds for $R_{\sigma_i}^*(A_\sigma)$, $1 \leq i \leq |\sigma|$, with $B \subsetneq A_\sigma$ in place of $B \subsetneq A$.

Proof. First we have

$$\begin{aligned}
&(|B|-1)\left[R_B^*(A) - \mathbb{H}_{\sigma_f}(B)\right] \\
&= (|B|-1)R_B^*(A) - \sum_{i \in B} H(X_B|X_i) \\
&= (|B|-1)\left[|B|\mathbb{H}_{\sigma_f}(A) - \sum_{i \in B} H(X_A|X_i)\right] - \sum_{i \in B} H(X_B|X_i) \\
&= \frac{|B|-1}{|A|-1}\left[|B|\sum_{i \in A} H(X_A|X_i) - (|A|-1)\sum_{i \in B} H(X_A|X_i)\right] - \sum_{i \in B} H(X_B|X_i) \\
&= \frac{(|B|-1)|B|}{|A|-1}\sum_{i \in \overline{B}} H(X_A|X_i) + \left(\frac{(|B|-1)|B|}{|A|-1} - |B|\right)\sum_{i \in B} H(X_B|X_i) \\
&\quad + \left(\frac{(|B|-1)|B|}{|A|-1} - (|B|-1)\right)|B|H(X_A|X_B),
\end{aligned} \tag{20}$$

where we used (14) in the second equality. On the other hand, we have

$$\begin{aligned}
&|B||\overline{B}|\left[\mathbb{H}_{\sigma_{\overline{B}}}(A) - \mathbb{H}_{\sigma_f}(A)\right] \\
&= \frac{|B|}{|A|-1}\left[(|A|-1)\sum_{i \in \overline{B}} H(X_A|X_i) + (|A|-1)H(X_A|X_B) - |\overline{B}|\sum_{i \in A} H(X_A|X_i)\right] \\
&= \frac{|B|}{|A|-1}\left[(|B|-1)\sum_{i \in \overline{B}} H(X_A|X_i) + (|A|-1)H(X_A|X_B) - |\overline{B}|\sum_{i \in B}\left(H(X_B|X_i) + H(X_A|X_B)\right)\right] \\
&= \frac{(|B|-1)|B|}{|A|-1}\sum_{i \in \overline{B}} H(X_A|X_i) - \frac{|B||\overline{B}|}{|A|-1}\sum_{i \in B} H(X_B|X_i) \\
&\quad + \frac{|B|}{|A|-1}\left((|A|-1) - |B||\overline{B}|\right)H(X_A|X_B),
\end{aligned} \tag{21}$$

where we used $|\overline{B}| = |A| - |B|$ in the second equality. We can verify that the coefficient of each term in (20) and (21) coincides. Thus, we have (19).

The second statement is proved exactly in the same manner by regarding X_{σ_i} as a single random variable. $\square$

As RDE proceeds, subsets of parties that have attained local omniscience start behaving as one. In the next recursive step of the protocol such sets of parties behaving as one attain omniscience. The next lemma ensures that when the rate is sufficient for these sets of parties to attain omniscience, it is sufficient also for the individual members of these sets to attain omniscience.

Lemma 9. *For a subset $A \subseteq \mathcal{M}$ and a partition $\sigma \in \Sigma(A)$ with $|\sigma| = k$, suppose that for every $1 \leq i \leq k$*

$$(R_j : j \in \sigma_i) \in \mathcal{R}_{\mathsf{CO}}^{\Delta}(\sigma_i), \tag{22}$$

and

$$(R_{\sigma_i} : 1 \leq i \leq k) \in \mathcal{R}_{\mathsf{CO}}^{\Delta}(A_\sigma), \tag{23}$$

where the elements of the set A_σ consist of parts of the partition σ (each part treated as a single element). Then, it holds that

$$(R_i : i \in A) \in \mathcal{R}_{\mathsf{CO}}^{\Delta}(A).$$

Proof. We prove that for any $B \subsetneq A$,

$$R_B \geq H(X_B|X_{A\setminus B}) + |B|\Delta.$$

Without loss of generality, we can assume

$$B = \left(\bigcup_{i=1}^{k'} B_i\right) \cup \left(\bigcup_{i=k'+1}^{k} B_i\right)$$

for some $1 \leq k' \leq k$, where $B_i \subsetneq \sigma_i$ for $1 \leq i \leq k'$ and $B_i = \sigma_i$ for $k'+1 \leq i \leq k$ (B_i may be empty set for $1 \leq i \leq k'$). Then, from (22) and (23), we have

$$\begin{aligned}
R_B &= \sum_{i=1}^{k'} R_{B_i} + \sum_{i=k'+1}^{k} R_{B_i} \\
&\geq \sum_{i=1}^{k'} H(X_{B_i}|X_{\sigma_i\setminus B_i}) + H(X_{\sigma_{k'+1}}, \ldots, X_{\sigma_k}|X_{A_\sigma\setminus\{\sigma_{k'+1},\ldots,\sigma_k\}}) + |B|\Delta \\
&= \sum_{i=1}^{k'} H(X_{B_i}|X_{\sigma_i\setminus B_i}) + H(X_{B_{k'+1}}, \ldots, X_{B_k}|X_{A\setminus\cup_{i=k'+1}^{k} B_i}) + |B|\Delta \\
&\geq \sum_{i=1}^{k'} H(X_{B_i}|X_{A\setminus\cup_{j=i}^{k} B_j}) + H(X_{B_{k'+1}}, \ldots, X_{B_k}|X_{A\setminus\cup_{i=k'+1}^{k} B_i}) + |B|\Delta \\
&= H(X_B|X_{A\setminus B}) + |B|\Delta.
\end{aligned}$$

□

The next observation helps us to relax the assumptions of the previous lemma by showing that the collocated parts of a partition will attain local omniscience even if a collection of (nonempty) subsets of each part attains local omniscience.

Lemma 10. *For a subset $A \subseteq \mathcal{M}$ and a partition $\sigma \in \Sigma(A)$ with $|\sigma| = k$, let $B_i \subseteq \sigma_i$ be nonempty for $1 \leq i \leq k$. Suppose that for every $1 \leq i \leq k$*

$$(R_j : j \in \sigma_i) \in \mathcal{R}^{\Delta}_{\mathsf{CO}}(\sigma_i), \tag{24}$$

and

$$(R_j : j \in B_i, 1 \leq i \leq k) \in \mathcal{R}^{\Delta}_{\mathsf{CO}}\left(\bigcup_{i=1}^{k} B_i\right). \tag{25}$$

Then, it holds that

$$(R_{\sigma_i} : 1 \leq i \leq k) \in \mathcal{R}^{\Delta}_{\mathsf{CO}}(A_\sigma).$$

Proof. It suffices to show that for any $C = \cup_{i=1}^{c} \sigma_i$ with $1 \leq c \leq k$

$$R_C \geq H(X_C | X_{A_\sigma \setminus C}) + |C|\Delta. \tag{26}$$

To that end, we have from (24) and (25) that

$$\begin{aligned}
R_C &= \sum_{i=1}^{c} R_{\sigma_i} \\
&= \sum_{i=1}^{c} R_{\sigma_i \setminus B_i} + \sum_{i=1}^{c} R_{B_i} \\
&\geq \sum_{i=1}^{c} H(X_{\sigma_i \setminus B_i} | X_{B_i}) + H(X_{B_1}, \ldots, X_{B_c} | X_{B_{c+1}}, \ldots, X_{B_k}) + |C|\Delta \\
&\geq \sum_{i=1}^{c} H(X_{\sigma_i \setminus B_i} | X_{\cup_{j=1}^{i-1}(\sigma_j \setminus B_j)}, X_{B_1}, \ldots, X_{B_c}, X_{\sigma_{c+1}}, \ldots, X_{\sigma_k}) \\
&\quad + H(X_{B_1}, \ldots, X_{B_c} | X_{\sigma_{c+1}}, \ldots, X_{\sigma_k}) + |C|\Delta \\
&= H(X_C | X_{A_\sigma \setminus C}) + |C|\Delta.
\end{aligned}$$

□

We need to show that when each call to OMN terminates, which happens when a subset A attains local omniscience, the rate of communication used for each party $i \in A$ is $R_i^*(A)$ (or if OMN was called with a partition σ then the same property holds with i and A, respectively, replaced by σ_l and A_σ, where A_σ is the set of parts that comprise A). Recall that OMN ensures that for each communicating party (or a set consisting of collocated parties) the difference $R_i - R_i^*(A)$ is maintained for every $A \subseteq \mathcal{M}$. Therefore, all the parties in A will reach $R_i^*(A)$ at the same time, and, since before reaching this rate their sum-rate will not be sufficient for omniscience, it suffices to show that the rate vector $(R_i^*(A) : i \in A)$ lies the omniscience region for A. The next technical lemma shows that there must be some subset A for which this holds and constitutes the main step in our proof. We show a slight generalization which holds when the parties in parts of $\sigma \in \Sigma(\mathcal{M})$ are collocated.

Lemma 11. *For a partition $\sigma \in \Sigma(\mathcal{M})$ and $A \subseteq \mathcal{M}$ such that*

$$A = \bigcup_{l=1}^{c} \sigma_{i_l},$$

there exists $B \subseteq \{1, \ldots, c\}$ with $|B| \geq 2$ such that

$$(R^*_{\sigma_{i_l}}(A_\sigma) + |\sigma_{i_l}|\Delta : l \in B) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\{\sigma_{i_l} : l \in B\}), \tag{27}$$

where A_σ is the set of parts σ_i that comprise A, with each part treated as a single element.

Proof. Since (27) is equivalent to

$$(R^*_{\sigma_{i_l}}(A_\sigma) : l \in B) \in \mathcal{R}_{\mathtt{CO}}(\{\sigma_{i_l} : l \in B\}),$$

we prove the claim for $\Delta = 0$. We proceed by induction on c. For $c = 2$, since

$$R^*_{\sigma_{i_1}}(A_\sigma) = H(X_{\sigma_{i_1}} | X_{\sigma_{i_2}}),$$
$$R^*_{\sigma_{i_2}}(A_\sigma) = H(X_{\sigma_{i_2}} | X_{\sigma_{i_1}}),$$

$B = \{1, 2\}$ satisfies the claim. Suppose that the claim holds for all $c \leq b$. For $c = b + 1$, if

$$(R^*_{\sigma_{i_l}}(A_\sigma) : 1 \leq l \leq c) \in \mathcal{R}_{\mathtt{CO}}(A_\sigma),$$

then $B = \{1, \ldots, c\}$ satisfies the claim. Otherwise, there exists $\overline{C} \subsetneq \{1, \ldots, c\}$ with $\overline{C} \neq \emptyset$ such that, with $C = \{1, \ldots, c\} \backslash \overline{C}$,

$$\sum_{l \in \overline{C}} R^*_{\sigma_{i_l}}(A_\sigma) < H(X_{\cup_{l \in \overline{C}} \sigma_{i_l}} | X_{\cup_{l \in C} \sigma_{i_l}}).$$

Then, by Lemma 8 and (15) of Lemma 7, it holds that

$$\sum_{l\in C} R^*_{\sigma_{i_l}}(A_\sigma) > \mathbb{H}_{\sigma_f}(\{\sigma_{i_j} : j \in C\}) \\ = \sum_{l\in C} R^*_{\sigma_{i_l}}(\{\sigma_{i_j} : j \in C\}). \tag{28}$$

Since

$$R^*_{\sigma_{i_l}}(A_\sigma) - R^*_{\sigma_{i_{l'}}}(A_\sigma) = H(X_{\sigma_{i_l}}) - H(X_{\sigma_{i_{l'}}}) \\ = R^*_{\sigma_{i_l}}(\{\sigma_{i_j} : j \in C\}) - R^*_{\sigma_{i_{l'}}}(\{\sigma_{i_j} : j \in C\})$$

for every $l \neq l'$ ($cf.$ (17) of Lemma 7), (28) implies

$$R^*_{\sigma_{i_l}}(A_\sigma) > R^*_{\sigma_{i_l}}(\{\sigma_{i_j} : j \in C\}), \ \forall\, l \in C. \tag{29}$$

Since $|C| \leq b$, by the induction hypothesis, there exists $B \subseteq C$ such that

$$(R^*_{\sigma_{i_l}}(\{\sigma_{i_j} : j \in C\}) : l \in B) \in \mathcal{R}_{\texttt{CO}}(\{\sigma_{i_l} : l \in B\}),$$

which together with (29) implies that B satisfies the claim for $c = b + 1$. □

We are now in a position to prove the main results.

B. Proof of Theorem 2

Before going to the proof of Theorem 2, let us formally define the error event $\mathcal{E}$. We shall show that this event will happen with vanishing probability.

Let L denote the maximum number of rounds of communication for party 1 over all possible values $\mathbf{x}_\mathcal{M}$ of the data sequence[15]. Since the protocol terminates either correctly or erroneously once the rate vector enters the omniscience region, and since

$$\big(\log|\mathcal{X}_{\sigma_i}| + |\sigma_i|\Delta : 1 \leq i \leq |\sigma|\big) \in \mathcal{R}^{\Delta}_{\texttt{CO}}(\mathcal{M}_\sigma|\mathrm{P}_\mathbf{x})$$

holds for any $\sigma \in \Sigma(\mathcal{M})$ and any sequence $\mathbf{x}$, L is bounded above as

$$L \leq \max_{\sigma\in\Sigma(\mathcal{M})} \max_{1\leq i\leq|\sigma|} \frac{\log|\mathcal{X}_{\sigma_i}|}{\Delta} + |\sigma_i| \\ \leq \frac{\log|\mathcal{X}_\mathcal{M}|}{\Delta} + m. \tag{30}$$

[15]Since party 1 is the first one to communicate and continues to communicate till the last round in RDE, the number of times other parties communicate does not exceed L.

For a fixed $\mathbf{x} \in \mathcal{X}_{\mathcal{M}}^n$, let $L(\mathbf{x})$ be the maximum number of rounds of communication when $\mathbf{x}$ is observed by the parties. With a slight abuse of notation, denote by $\mathbf{R}(l)$ the rate of communication after l rounds, $1 \leq l \leq L$, if the protocol does not declare an error till then. Also, let $h_l(\mathbf{x}_i)$ denote the random hash bits sent by the ith party (observing $\mathbf{x}_i$) in the lth round.

For $B \subsetneq A \subseteq \mathcal{M}$ and $1 \leq l \leq L(\mathbf{x})$, let

$$\mathcal{T}_l^B(\mathbf{x}) = \left\{\mathbf{x}'_A : (R_i(l) : i \in A) \in \mathcal{R}_{\mathtt{CO}}^{\Delta}(A|\mathrm{P}_{\mathbf{x}'_A}) \text{ and } \{i \in A : \mathbf{x}'_i \neq \mathbf{x}_i\} = B\right\}.$$

Note that

$$\begin{aligned} |\mathcal{T}_l^B(\mathbf{x})| &\leq p(n) \max_{\substack{\mathrm{P}_{\overline{X}_A} \in \mathcal{P}_n(\mathcal{X}_A): \\ (R_i(l): i \in A) \in \mathcal{R}_{\mathtt{CO}}^{\Delta}\left(A|\mathrm{P}_{\overline{X}_A}\right)}} |\{\mathbf{x}'_A : \mathrm{P}_{\mathbf{x}'_A} = \mathrm{P}_{\overline{X}_A}, \{i \in A : \mathbf{x}'_i \neq \mathbf{x}_i\} = B\}| \\ &\leq p(n) \max_{\substack{\mathrm{P}_{\overline{X}_A} \in \mathcal{P}_n(\mathcal{X}_A): \\ (R_i(l): i \in A) \in \mathcal{R}_{\mathtt{CO}}^{\Delta}\left(A|\mathrm{P}_{\overline{X}_A}\right)}} 2^{nH\left(\overline{X}_A|\overline{X}_{A\setminus B}\right)} \\ &\leq p(n) 2^{nR_B(l) - n|B|\Delta}, \end{aligned} \tag{31}$$

where $\mathcal{P}_n(\mathcal{X}_A)$ is the set of all types on $\mathcal{X}_A$ and $p(n)$ is the number of types and is polynomial in n.

For $B \subsetneq A$, denote by $\mathcal{E}_A(l, B)$ the error event

$$\mathcal{E}_A(l, B) = \left\{\exists \mathbf{x}'_A \in \mathcal{T}_l^B(\mathbf{x}) \text{ s.t. } h_k(\mathbf{x}_j) = h_k(\mathbf{x}'_j)\ \forall j \in B, \forall 1 \leq k \leq l\right\}.$$

Finally, let

$$\begin{aligned} \mathcal{E}_A(l) &= \bigcup \left\{\mathcal{E}_A(l, B) : B \neq \emptyset, B \subsetneq A\right\}, \\ \mathcal{E} &= \bigcup \left\{\mathcal{E}_A(l) : 1 \leq l \leq L(\mathbf{x}), A \subseteq \mathcal{M}\right\}. \end{aligned}$$

Lemma 12. *There exists a constant C depending only on m such that, for every sequence $\mathbf{x} \in \mathcal{X}_{\mathcal{M}}^n$, the probability of the error event $\mathcal{E} = \mathcal{E}(\mathbf{x})$ defined above is bounded by $CLp(n)2^{-n\Delta}$.*

Remark 2. Suppose for a sequence $\mathbf{x} \in \mathcal{X}_{\mathcal{M}}^n$, error $\mathcal{E}$ does not occur. By definitions of $\mathcal{E}$ and DEC, when OMN terminates, a set A belongs to $\mathcal{O}$ if and only if the final rates of communication $\mathbf{R}$ satisfy $(R_i : i \in A) \in \mathcal{R}_{\mathtt{CO}}^{\Delta}(A|\mathrm{P}_{\mathbf{x}})$.

Proof. By noting the bound in (31), we have

$$\begin{aligned} \Pr(\mathcal{E}_A(l, B) \mid X_{\mathcal{M}}^n = \mathbf{x}) &\leq \frac{1}{2^{nR_B(l)}} |\mathcal{T}_l^B(\mathbf{x})| \\ &\leq p(n) 2^{-n\Delta}, \end{aligned}$$

where the first inequality uses the bound for probability of collision event $h_k(\mathbf{x}_j) = h_k(\mathbf{x}'_j)$ which holds for a random hash. (In fact, the same bound holds for a randomly selected member of a 2-universal hash family.) Thus,

$$\begin{aligned}
\Pr(\mathcal{E} \mid X^n_{\mathcal{M}} = \mathbf{x}) &\leq 2^m \cdot L \cdot \max_{A \subseteq \mathcal{M}, 1 \leq l \leq L(\mathbf{x})} \Pr(\mathcal{E}_A(l) \mid X^n_{\mathcal{M}} = \mathbf{x}) \\
&\leq 4^m \cdot L \cdot \max_{A \subseteq \mathcal{M}, 1 \leq l \leq L(\mathbf{x}), B \subsetneq A} \Pr(\mathcal{E}_A(l, B) \mid X^n_{\mathcal{M}} = \mathbf{x}) \\
&\leq 4^m \cdot L \cdot p(n) \cdot 2^{-n\Delta}.
\end{aligned}$$

□

Proof of Theorem 2. We prove each statement of Theorem 2 separately.

Proof of (Ia): Denoting $B_i = A \cap \sigma_i$, let $\{i_1, \ldots, i_c\}$ be those indices $i \in \{1, \ldots, k\}$ for which $B_i \neq \emptyset$. Since $\mathbf{R}^{\mathtt{in}}$ is $(\sigma, \mathbf{H}, \alpha)$-valid, it satisfies

$$(R^{\mathtt{in}}_j : j \in \sigma_i) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_i), \ \forall\, 1 \leq i \leq k \text{ s.t. } |\sigma_i| \geq 2,$$

and therefore, so does $\mathbf{R}^{\mathtt{out}}$, i.e.,

$$(R^{\mathtt{out}}_j : j \in \sigma_i) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_i), \ \forall\, 1 \leq i \leq k \text{ s.t. } |\sigma_i| \geq 2.$$

Furthermore, since an error does not occur and the parties in A attain omniscience, by Remark 2

$$(R^{\mathtt{out}}_j : j \in A) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(A).$$

Thus, by Lemma 10 and Lemma 9,

$$(R^{\mathtt{out}}_j : j \in \sigma_{i_l}, 1 \leq l \leq c) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}\left(\bigcup_{l=1}^{c} \sigma_{i_l}\right).$$

Therefore, since no error has occurred, by Remark 2 the parties in $\cup_{l=1}^{c} \sigma_{i_l}$ must attain omniscience. But by the definition of $\mathcal{O}$ the set A is a maximal set attaining omniscience and $A \subseteq \cup_{l=1}^{c} \sigma_{i_l}$. Hence, A must be $\cup_{l=1}^{c} \sigma_{i_l}$.

Proof of (Ib): As a preparation of our proof, we first show that for $1 \leq j, l \leq c$, the difference between $(R^{\mathtt{out}}_{\sigma_{i_j}} - R^*_{\sigma_{i_j}}(A_\sigma))$ and $(R^{\mathtt{out}}_{\sigma_{i_l}} - R^*_{\sigma_{i_l}}(A_\sigma))$ is bounded above by $(2\alpha + 1)\Delta$. Indeed, for $1 \leq j, l \leq c$, the parties in σ_{i_j} and σ_{i_l} are communicating when OMN terminates. In fact, defining

$$i_l \leq s^{\mathtt{in}} := \max\{i : R^{\mathtt{in}}_{\sigma_i} \geq 0\},$$

the parties in σ_{i_l} were communicating even when OMN was initiated, and therefore,

$$R^{\mathtt{out}}_{\sigma_1} - R^{\mathtt{out}}_{\sigma_{i_l}} = R^{\mathtt{in}}_{\sigma_1} - R^{\mathtt{in}}_{\sigma_{i_l}},$$

which by the assumption that $\mathbf{R}^{\mathtt{in}}$ is $(\sigma, \mathbf{H}, \alpha)$-valid yields

$$H_{\sigma_1} - H_{\sigma_{i_l}} - \alpha\Delta \leq R^{\mathtt{out}}_{\sigma_1} - R^{\mathtt{out}}_{\sigma_{i_l}} \leq H_{\sigma_1} - H_{\sigma_{i_l}} + \alpha\Delta. \tag{32}$$

On the other hand, if $i_l > s^{\mathtt{in}}$, then the parties in σ_{i_l} start communicating when

$$R_{\sigma_1} = \lceil H_{\sigma_1} - H_{\sigma_{i_l}} + \alpha\Delta \rceil_{\Delta},$$

where $\lceil a \rceil_{\Delta} := \min\{i\Delta : i \in \mathbb{N}, i\Delta \geq a\}$. Thereafter, the parties in σ_1 as well as σ_{i_l} communicate at sum-rate Δ per round. Thus, in this case,

$$R^{\mathtt{out}}_{\sigma_1} - R^{\mathtt{out}}_{\sigma_{i_l}} = \lceil H_{\sigma_1} - H_{\sigma_{i_l}} + \alpha\Delta \rceil_{\Delta}. \tag{33}$$

Upon combining (32) and (33), we get that for every $1 \leq j, l \leq c$,

$$\begin{aligned} R^{\mathtt{out}}_{\sigma_{i_j}} - R^{\mathtt{out}}_{\sigma_{i_l}} &\leq H_{\sigma_{i_j}} - H_{\sigma_{i_l}} + (2\alpha + 1)\Delta \\ &= R^*_{\sigma_{i_j}}(A_\sigma) - R^*_{\sigma_{i_l}}(A_\sigma) + (2\alpha + 1)\Delta, \end{aligned} \tag{34}$$

where the previous equation is by (17).

Now, we prove the lower bound in (Ib). Suppose that there exists a $j \in \{1, \ldots, c\}$ such that

$$R^{\mathtt{out}}_{\sigma_{i_j}} < R^*_{\sigma_{i_j}}(A_\sigma) + (|\sigma_{i_j}| - 2\alpha - 1)\Delta.$$

It follows from (34) that

$$\begin{aligned} \sum_{l=1}^{c} R^{\mathtt{out}}_{\sigma_{i_l}} &< \sum_{l=1}^{c} R^*_{\sigma_{i_l}}(A_\sigma) + \sum_{l=1}^{c} |\sigma_{i_l}|\Delta \\ &= \sum_{l=1}^{c} R^*_{\sigma_{i_l}}(A_\sigma) + |A|\Delta \\ &= \mathbb{H}_{\sigma_f}(A_\sigma) + |A|\Delta, \end{aligned} \tag{35}$$

where the previous equation is by (13). Also, since no error occurs and parties in A attain omniscience, by Remark 2,

$$(R^{\mathtt{out}}_j : j \in A) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(A),$$

which in turn implies that

$$\begin{aligned}\sum_{l=1}^{c} R^{\mathtt{out}}_{\sigma_{i_l}} &= \frac{1}{c-1}\sum_{l=1}^{c}\sum_{\substack{j=1\\ j\neq l}}^{c} R^{\mathtt{out}}_{\sigma_{i_j}} \\ &\geq \frac{1}{c-1}\sum_{l=1}^{c}\left[H(X_A|X_{\sigma_{i_l}}) + (|A|-|\sigma_{i_l}|)\Delta\right] \\ &= \mathbb{H}_{\sigma_f}(A_\sigma) + |A|\Delta,\end{aligned}$$

which contradicts (35). Thus, for every $1 \leq l \leq c$,

$$\begin{aligned}R^{\mathtt{out}}_{\sigma_{i_l}} &\geq R^*_{\sigma_{i_l}}(A_\sigma) + (|\sigma_{i_l}| - 2\alpha - 1)\Delta \\ &\geq R^*_{\sigma_{i_l}}(A_\sigma) - 2\alpha\Delta.\end{aligned}$$

Moving to the proof of the upper bound in (Ib), suppose that there exists an l such that

$$R^{\mathtt{out}}_{\sigma_{i_l}} > R^*_{\sigma_{i_l}}(A_\sigma) + (m + 2\alpha + 1)\Delta. \tag{36}$$

From Lemma 11, there exists $B \subseteq \{1, \ldots, c\}$ with $|B| \geq 2$ such that

$$(R^*_{\sigma_{i_l}}(A_\sigma) + |\sigma_{i_l}|\Delta : l \in B) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\{\sigma_{i_l} : l \in B\}). \tag{37}$$

Then, (34), (36) and (37) imply (by noting $|\sigma_{i_l}| < m$) that

$$(R^{\mathtt{out}}_{\sigma_{i_l}} - \Delta : l \in B) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\{\sigma_{i_l} : l \in B\}). \tag{38}$$

Also, note that for every $j \in \sigma_{i_l}$, $1 \leq l \leq c$ with $|\sigma_{i_l}| \geq 2$,

$$R^{\mathtt{out}}_j - R^{\mathtt{in}}_j \geq \frac{\Delta}{|\sigma_{i_l}|}, \tag{39}$$

since otherwise there is no communication in the execution of OMN, which in turn by Remark 2 contradicts the assumption that $\mathbf{R}^{\mathtt{in}}$ is $(\sigma, \mathbf{H}, \alpha)$-valid. Upon combining (39) with (10), we get

$$\left(R^{\mathtt{out}}_j - \frac{\Delta}{|\sigma_{i_l}|} : j \in \sigma_{i_l}\right) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_{i_l})$$

for every $1 \leq l \leq c$ with $|\sigma_{i_l}| \geq 2$, which together with (38) and Lemma 9 yields

$$\left(R^{\mathtt{out}}_j - \frac{\Delta}{|\sigma_{i_l}|} : j \in \sigma_{i_l}, l \in B\right) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}\left(\bigcup_{l\in B}\sigma_{i_l}\right).$$

But then by Remark 2 the parties in $\cup_{l\in B}\sigma_{i_l}$ attain omniscience one round before OMN terminates, which is a contradiction since no error has occurred and OMN must terminate as soon as a subset in $\mathcal{O}$ is recognized.

Proof of (II): For each $\sigma_i^{\texttt{out}} \in \sigma^{\texttt{out}}$ either $\sigma_i^{\texttt{out}} \in \sigma$ or $\sigma_i^{\texttt{out}} \in \mathcal{O}$; in the latter case, by (Ia), $\sigma_i^{\texttt{out}}$ must equal a union of parts of σ. Note that, by the argument leading to (34), for every $\sigma_i, \sigma_j \in \sigma$ such that $R_{\sigma_i}^{\texttt{out}} \geq 0$ and $R_{\sigma_j}^{\texttt{out}} \geq 0$

$$R_{\sigma_i}^{\texttt{out}} - R_{\sigma_j}^{\texttt{out}} \leq H_{\sigma_i} - H_{\sigma_j} + (2\alpha + 1)\Delta. \tag{40}$$

Also, for $\sigma_i^{\texttt{out}} = \cup_{l=1}^{c} \sigma_{i_l} \in \mathcal{O}$, note that[16] by (Ib)

$$\begin{aligned}\sum_{l=2}^{c} R_{\sigma_{i_l}}^{*}(\{\sigma_{i_1}, \ldots, \sigma_{i_c}\}) - 2\alpha c\Delta &\leq \sum_{l=2}^{c} R_{\sigma_{i_l}}^{\texttt{out}} \\ &\leq \sum_{l=2}^{c} R_{\sigma_{i_l}}^{*}(\{\sigma_{i_1}, \ldots, \sigma_{i_c}\}) + (mc + 2\alpha c)\Delta,\end{aligned}$$

which by (16) is the same as

$$\begin{aligned}H(X_{\sigma_i^{\texttt{out}}} | X_{\sigma_{i_1}}) - 2\alpha c\Delta &\leq \sum_{l=2}^{c} R_{\sigma_{i_l}}^{\texttt{out}} \\ &\leq H(X_{\sigma_i^{\texttt{out}}} | X_{\sigma_{i_1}}) + (mc + 2\alpha c)\Delta.\end{aligned} \tag{41}$$

To prove condition (i) in the definition of a valid rate vector ($cf.$ Definition 3), consider $\sigma_i^{\texttt{out}}$ and $\sigma_j^{\texttt{out}}$ such that $R_{\sigma_i^{\texttt{out}}}^{\texttt{out}} \geq 0$ and $R_{\sigma_j^{\texttt{out}}}^{\texttt{out}} \geq 0$. The following three cases are possible:

- Case $\sigma_i^{\texttt{out}}, \sigma_j^{\texttt{out}} \in \sigma$: In this case, the claim follows from (40).
- Case $\sigma_i^{\texttt{out}} \in \mathcal{O}$, $\sigma_j^{\texttt{out}} \in \sigma$: Let $\sigma_i^{\texttt{out}} = \cup_{l=1}^{c} \sigma_{i_l}$. Then, by (40) and (41)

$$\begin{aligned}R_{\sigma_i^{\texttt{out}}}^{\texttt{out}} - R_{\sigma_j^{\texttt{out}}}^{\texttt{out}} &= \sum_{l=2}^{c} R_{\sigma_{i_l}}^{\texttt{out}} + R_{\sigma_{i_1}}^{\texttt{out}} - R_{\sigma_j^{\texttt{out}}}^{\texttt{out}} \\ &\leq H(X_{\sigma_i^{\texttt{out}}} | X_{\sigma_{i_1}}) + H(X_{\sigma_{i_1}}) - H(X_{\sigma_j^{\texttt{out}}}) + (mc + 2\alpha c + 2\alpha + 1)\Delta \\ &= H(X_{\sigma_i^{\texttt{out}}}) - H(X_{\sigma_j^{\texttt{out}}}) + (mc + 2\alpha c + 2\alpha + 1)\Delta,\end{aligned}$$

  and similarly,

$$R_{\sigma_j^{\texttt{out}}}^{\texttt{out}} - R_{\sigma_i^{\texttt{out}}}^{\texttt{out}} \leq H(X_{\sigma_j^{\texttt{out}}}) - H(X_{\sigma_i^{\texttt{out}}}) + (2\alpha c + 2\alpha + 1)\Delta.$$

[16] We show the argument for $\sum_{l=2}^{c} R_{\sigma_{i_l}}^{\texttt{out}}$; the same argument extends to $\sum_{l=1, l\neq i}^{c} R_{\sigma_{i_l}}^{\texttt{out}}$ for every $i \in \{2, \ldots, c\}$.

- Case $\sigma_i^{\mathtt{out}}, \sigma_j^{\mathtt{out}} \in \mathcal{O}$: Using argument similar to the previous case, we can show

$$R^{\mathtt{out}}_{\sigma_i^{\mathtt{out}}} - R^{\mathtt{out}}_{\sigma_j^{\mathtt{out}}} \leq H(X_{\sigma_i^{\mathtt{out}}}) - H(X_{\sigma_j^{\mathtt{out}}}) + (mc + 4\alpha c + 2\alpha + 1)\Delta.$$

Condition (ii) can be proved similarly by considering two cases: $\sigma_1^{\mathtt{out}} \in \sigma$ and $\sigma_1^{\mathtt{out}} \in \mathcal{O}$. Specifically, let $s' = \max\{i : R^{\mathtt{out}}_{\sigma_i} \geq 0\}$. If $\sigma_1^{\mathtt{out}} \in \sigma$, then $\sigma_1^{\mathtt{out}} = \sigma_1$ and condition (ii) holds since the party[17] $\sigma^{\mathtt{out}}_{s'+1}$ did not start communicating. On the other hand, if $\sigma_1^{\mathtt{out}} = \cup_{l=1}^{c} \sigma_{i_l} \in \mathcal{O}$, then

$$\begin{aligned} R^{\mathtt{out}}_{\sigma_1^{\mathtt{out}}} &= R^{\mathtt{out}}_{\sigma_1} + R^{\mathtt{out}}_{\sigma_{i_1}} - R^{\mathtt{out}}_{\sigma_1} + \sum_{l=2}^{c} R^{\mathtt{out}}_{\sigma_{i_l}} \\ &< H_{\sigma_1} - H_{\sigma^{\mathtt{out}}_{s'+1}} + H_{\sigma_{i_1}} - H_{\sigma_1} + H(X_{\sigma_1^{\mathtt{out}}} | X_{\sigma_{i_1}}) + (mc + 2\alpha c + 3\alpha + 1)\Delta \\ &= H_{\sigma_1^{\mathtt{out}}} - H_{\sigma^{\mathtt{out}}_{s'+1}} + (mc + 2\alpha c + 3\alpha + 1)\Delta, \end{aligned}$$

where the strict inequality is by (40) and (41), since the party $\sigma^{\mathtt{out}}_{s'+1}$ did not start communicating.

For condition (iii), if $\sigma_i^{\mathtt{out}}$ equals to a part of σ, then $(R_j^{\mathtt{out}} : j \in \sigma_i^{\mathtt{out}}) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_i^{\mathtt{out}})$ since $(R_j^{\mathtt{in}} : j \in \sigma_i^{\mathtt{out}}) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_i^{\mathtt{out}})$ and $R_j^{\mathtt{out}} \geq R_j^{\mathtt{in}}$ for every $1 \leq j \leq m$. On the other hand, if $\sigma_i^{\mathtt{out}}$ belongs to $\mathcal{O}$, then $(R_j^{\mathtt{out}} : j \in \sigma_i^{\mathtt{out}}) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}(\sigma_i^{\mathtt{out}})$ by Remark 2.

Finally, for condition (iv), if there exists $A \subseteq \{1, \dots, |\sigma^{\mathtt{out}}|\}$, $|A| \geq 2$, such that

$$(R_j^{\mathtt{out}} : j \in \sigma_i^{\mathtt{out}}, i \in A) \in \mathcal{R}^{\Delta}_{\mathtt{CO}}\left(\bigcup_{i \in A} \sigma_i^{\mathtt{out}}\right),$$

then by Remark 2 $\cup_{i \in A} \sigma_i^{\mathtt{out}} \in \mathcal{O}$, which further implies that $\cup_{i \in A} \sigma_i^{\mathtt{out}} \in \mathcal{O}$ is a part of $\sigma^{\mathtt{out}}$, a contradiction. Thus, condition (iv) must hold for $\mathbf{R}^{\mathtt{out}}$. $\square$

C. *Proofs of Theorem 3 and Corollary 4*

For Theorem 3, by Lemma 12 the probability of the error event $\mathcal{E} = \mathcal{E}(\mathbf{x})$ is bounded above by $C_1 L p(n) 2^{-n\Delta}$ for some constant C_1, where L is the maximum number of rounds and is bounded above by $\frac{\log |\mathcal{X}_{\mathcal{M}}|}{\Delta} + m$ using (30). Under the assumption that the error event $\mathcal{E}$ did not occur, at the end of the jth call to OMN with input partition σ, Theorem 2 guarantees that the total number of bits sent by each subset $A \in \mathcal{O}$ is bounded above by[18]

$$n\mathbb{H}_{\sigma_f(A_\sigma)}(A_\sigma | \mathrm{P}_{\mathbf{x}_A}) + nc(m + 2\alpha_j)\Delta + C_3 L + C_4 \log n, \tag{42}$$

[17] It can be seen that the parts of σ which did not start communicating must be singleton.

[18] The $\log n$ term corresponds to the bits communicated to share types of the locally recovered observations. Additional $C_3 L$ bits are added to account for the overhead arising from rounding-off the required number of bits to an integer and ACK/NACK bits for each round.

for some constants $C_3, C_4 > 0$, where α_j is recursively defined by setting $\alpha_1 = 1$ and $\alpha_{j+1} = c'_m \alpha_j$ with c'_m given in Theorem 2-(II). Since the size of partition σ strictly decreases in each execution of OMN, the number of calls to OMN is at most m and α_j remains bounded above by a constant that depends only on m. Theorem 3 follows upon using (42) for $A = \mathcal{M}$, and noting that

$$\begin{aligned}\mathbb{H}_{\sigma_f(\mathcal{M}_\sigma)}(\mathcal{M}_\sigma | \mathrm{P}_{\mathbf{x}_\mathcal{M}}) &= \mathbb{H}_\sigma(\mathcal{M} | \mathrm{P}_{\mathbf{x}_\mathcal{M}}) \\ &\leq R_{\mathtt{CO}}(\mathcal{M} | \mathrm{P}_{\mathbf{x}_\mathcal{M}}),\end{aligned}$$

where the inequality is by (2).

Corollary 4 is obtained as a consequence of Theorem 3 as follows. First, note that under the error event $\mathcal{E}$, which occur with probability less than $C_1 p(n) L 2^{-n\Delta}$, the number of communicated bits is bounded above by $C_5 n$ for some constant $C_5 > 0$. Next, by the Taylor approximation of the entropy function around $\mathrm{P}_{X_\mathcal{M}}$, for $\mathrm{Q}_{X_\mathcal{M}}$ satisfying $\|\mathrm{P}_{X_\mathcal{M}} - \mathrm{Q}_{X_\mathcal{M}}\| \leq \delta$ and $\mathtt{supp}(\mathrm{Q}_{X_\mathcal{M}}) \subset \mathtt{supp}(\mathrm{P}_{X_\mathcal{M}})$, we have

$$\left| R_{\mathtt{CO}}(\mathcal{M} | \mathrm{P}_{X_\mathcal{M}}) - R_{\mathtt{CO}}(\mathcal{M} | \mathrm{Q}_{X_\mathcal{M}}) \right| \leq C_6 \delta \tag{43}$$

for a sufficiently small δ, where $C_6 > 0$ is a constant that depends[19] on $\mathrm{P}_{X_\mathcal{M}}$. Denoting

$$\mathcal{B}_\delta(\mathrm{P}_{X_\mathcal{M}}) := \{\mathrm{Q}_{X_\mathcal{M}} : \|\mathrm{P}_{X_\mathcal{M}} - \mathrm{Q}_{X_\mathcal{M}}\| \leq \delta, \ \mathtt{supp}(\mathrm{Q}_{X_\mathcal{M}}) \subset \mathtt{supp}(\mathrm{P}_{X_\mathcal{M}})\},$$

Theorem 3 implies that, when $\mathcal{E}$ does not occur, the number of bits communicated is no more than

$$\begin{aligned}&nR_{\mathtt{CO}}(\mathcal{M} | \mathrm{P}_{X_\mathcal{M}}) + C_6 n\delta + \Pr\left(\mathtt{type}(X^n_\mathcal{M}) \notin \mathcal{B}_\delta(\mathrm{P}_{X_\mathcal{M}})\right) n \log|\mathcal{X}_\mathcal{M}| + C_2 n\Delta + C_3 L + C_4 \log n \\ &\leq nR_{\mathtt{CO}}(\mathcal{M} | \mathrm{P}_{X_\mathcal{M}}) + C_6 n\delta + 2|\mathcal{X}_\mathcal{M}| \exp(-2n\delta^2) n \log|\mathcal{X}_\mathcal{M}| + C_2 n\Delta + C_3 L + C_4 \log n,\end{aligned}$$

where the inequality uses the Hoeffding bound

$$\Pr\left(\mathtt{type}(X^n_\mathcal{M}) \notin \mathcal{B}_\delta(\mathrm{P}_{X_\mathcal{M}})\right) \leq 2|\mathcal{X}_\mathcal{M}| \exp(-2n\delta^2).$$

The claimed upper bound for the expected number of bits communicated follows by combining the bounds under $\mathcal{E}$ and $\mathcal{E}^c$ and setting $\delta = \sqrt{\frac{\log n}{n}}$, $\Delta = \frac{1}{\sqrt{n}}$. □

D. Proof of Theorem 6

We first recall the leftover hash lemma ($cf.$ [21]); a proof of the version stated below is given in, for instance, [15, Appendix B].

[19] The dependence of C_6 on $\mathrm{P}_{X_\mathcal{M}}$ can be omitted by replacing δ with $\delta \log \frac{|\mathcal{X}_\mathcal{M}|}{\delta}$.

Lemma 13 (Leftover Hash). *Consider random variables X and V taking values in finite sets $\mathcal{X}$ and $\mathcal{V}$, respectively. Let S be a random seed such that f_S is uniformly distributed over a 2-universal hash family. Then, for $K = f_S(X)$, we have*

$$\|\mathrm{P}_{KVS} - \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_V \times \mathrm{P}_S\|_1 \leq \frac{1}{2}\sqrt{|\mathcal{V}|2^{-H_{\min}(\mathrm{P}_X)}},$$

where $\mathrm{P}_{\mathtt{unif}}$ is the uniform distribution on $\mathcal{K}$ and

$$H_{\min}(\mathrm{P}_X) = -\log \max_x \mathrm{P}_X\left(x\right).$$

We assume that the public randomness U used in Protocol 6 is available to the eavesdropper. Denote by $\mathcal{E}$ the error event of Protocol 6, which is determined by $(X_{\mathcal{M}}^n, U)$, and by Π' an expurgated transcript defined as

$$\Pi' = \begin{cases} \Pi, & \text{if } (X_{\mathcal{M}}^n, U) \notin \mathcal{E}, \\ \mathtt{constant}, & \text{otherwise.} \end{cases}$$

Our security analysis will show that Π' reveals negligible information about the SK and then use the large probability of agreement between Π and Π' to claim the security of the SK. Note that when the joint type of $X_{\mathcal{M}}^n$ is $\mathrm{P}_{\overline{X}_{\mathcal{M}}}$ and an error did not occur in Protocol 6, the length of the transcript Π is bounded by $l(\mathrm{P}_{\overline{X}_{\mathcal{M}}})$; thereby the length of Π' is bounded by $l(\mathrm{P}_{\overline{X}_{\mathcal{M}}})$ as well.

For each realization $X_{\mathcal{M}}^n = \mathbf{x}$, we generate a SK of length $k(\mathrm{P}_{\mathbf{x}})$ by randomly hashing $\mathbf{x}$ to $k(\mathrm{P}_{\mathbf{x}})$ bits[20]. Clearly, the recoverability condition is satisfied with $1 - \epsilon_n$, where ϵ_n is the error probability of Protocol 6. Without loss of generality, we assume that the eavesdropper has access to the joint type $\mathrm{P}_{\mathbf{x}}$ of $\mathbf{x}$. Note that such an eavesdropper has potentially more information than that available to the actual eavesdropper in our protocol. Thus, security against this stronger eavesdropper implies security against the actual eavesdropper. Denoting by $T = t$ a fixed realization of the random type, triangular inequality yields

$$\begin{aligned}
&\sum_{t \in \mathcal{P}_n(\mathcal{X}_{\mathcal{M}})} P_T(t) \|\mathrm{P}_{K\Pi U|T=t} - \mathrm{P}_{\mathtt{unif},t} \times \mathrm{P}_{\Pi U|T=t}\| \\
&\quad \leq \sum_{t \in \mathcal{P}_n(\mathcal{X}_{\mathcal{M}})} P_T(t) \Big[\|\mathrm{P}_{K\Pi U|T=t} - \mathrm{P}_{K\Pi' U|T=t}\| + \|\mathrm{P}_{\Pi' U|T=t} - \mathrm{P}_{\Pi U|T=t}\| \\
&\qquad + \|\mathrm{P}_{K\Pi' U|T=t} - \mathrm{P}_{\mathtt{unif}} \times \mathrm{P}_{\Pi' U|T=t}\| \Big],
\end{aligned}$$

[20]Specifically, we use a seeded extractor for each fixed joint type $\mathrm{P}_{\mathbf{x}}$. For ease of presentation, we omit the dependence on seed from our notation.

where $\mathrm{P}_{\mathtt{unif},t}$ is the uniform distribution on $\{0,1\}^{k(t)}$. The first two terms on the right-side above are each bounded above by $\Pr\left(\Pi \neq \Pi'\right)$. Also, by Lemma 13 applied for each fixed (t,u), the third term is bounded above by

$$\sum_{t\in\mathcal{P}_n(\mathcal{X}_\mathcal{M})} P_T(t)\frac{1}{2}\sqrt{2^{l(t)+k(t)}(n+1)^{|\mathcal{X}_\mathcal{M}|}2^{-H(t)}},$$

where we have used the independence of U and $X^n_\mathcal{M}$ and the observation that

$$\begin{aligned}H_{\min}(\mathrm{P}_{X^n_\mathcal{M}|T=t,U=u}) &= H_{\min}(\mathrm{P}_{X^n_\mathcal{M}|T=t})\\ &\geq nH(t) - |\mathcal{X}_\mathcal{M}|\log(n+1).\end{aligned}$$

Thus, by combining the bounds above, we get

$$\begin{aligned}&\sum_{t\in\mathcal{P}_n(\mathcal{X}_\mathcal{M})} P_T(t)\|\mathrm{P}_{K\Pi U|T=t} - \mathrm{P}_{\mathtt{unif},t}\times\mathrm{P}_{\Pi U|T=t}\|\\ &\leq 2\Pr\left(\Pi\neq\Pi'\right) + \sum_{t\in\mathcal{P}_n(\mathcal{X}_\mathcal{M})} P_T(t)\frac{1}{2}\sqrt{2^{l(t)+k(t)}(n+1)^{|\mathcal{X}_\mathcal{M}|}2^{-H(t)}}\\ &\leq \delta,\end{aligned}$$

where the previous inequality uses $\mathbb{P}\left(\Pi\neq\Pi'\right)\leq\epsilon_n$ and the definitions of $l(t)$, $k(t)$, and δ. The average length $\sum_{t\in\mathcal{P}_n(\mathcal{X}_\mathcal{M})} P_T(t)k(t)$ is lower bounded by (12) using Theorem 5, in a similar manner as the proof of Corollary 4. □

ACKNOWLEDGMENT

SW is supported in part by the JSPS KAKENHI under grant 16H06091. HT is supported in part by the Defense Research and Development Organisation (DRDO), India under grant DRDO0649. Authors thank Navin Kashyap for pointing to [4, Theorem 5.2] and [6].

REFERENCES

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography–part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[2] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143 – 154, 1979.

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

[4] C. Chan, A. Al-Bashabsheh, J. B. Ebrahimi, T. Kaced, and T. Liu, "Multivariate mutual information inspired by secret-key agreement," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1883–1913, Oct 2015.

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

[6] C. Chan, A. Al-Bashabsheh, J. Ebrahimi, T. Kaced, S. Kadhe, T. Liu, A. Sprintson, M. Yan, and Q. Zhou, "Successive omniscience," in *2015 International Symposium on Network Coding (NetCod)*, June 2015, pp. 21–25.

[7] T. A. Courtade, B. Xie, and R. D. Wesel, "Optimal exchange of packet for universal recovery in broadcast networks," *Proc. Military Communication Conference*, pp. 2250–2255, 2010.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory. 2nd edition*. John Wiley & Sons Inc., 2006.

[9] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Trans. Inf. Theory*, vol. 28, no. 4, pp. 585–592, July 1982.

[10] 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.

[11] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, January 1981.

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

[13] S. C. Draper, "Universal incremental slepian-wolf coding," in *in 42nd Annual Allerton Conf*, 2004.

[14] A. El Gamal and A. Orlitsky, "Interactive data compression," *FOCS*, pp. 100–108, 1984.

[15] M. Hayashi, H. Tyagi, and S. Watanabe, "Secret key agreement: General capacity and second-order asymptotics," *CoRR*, vol. abs/1411.0735, 2014.

[16] E. Kushilevitz and N. Nisan, *Communication Complexity*. New York, NY, USA: Cambridge University Press, 1997.

[17] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[18] N. Milosavljevic, S. Pawar, S. E. Rouayheb, M. Gastpar, and K. Ramchandran, "Deterministic algorithm for the cooperative data exchange problem," *Proc. IEEE International Symposium on Information Theory*, pp. 474–478, 2011.

[19] ——, "Efficient algorithm for the data exchange problem," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1878–1896, April 2016.

[20] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, November 1994.

[21] R. Renner, "Security of quantum key distribution," *Ph. D. Dissertation, ETH Zurich*, 2005.

[22] S. E. Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," *Proc. IEEE Information Theory Workshop*, pp. 1–5, 2010.

[23] D. Slepian and J. Wolf, "Noiseless coding of correlated information source," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[24] A. Sprintson, P. Sadeghi, G. Booker, and S. E. Rouayheb, "A randomized algorithm and performance bounds for coded cooperative data exchange," *Proc. IEEE International Symposium on Information Theory*, pp. 1888–1892, 2010.

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

[26] H. Tyagi and S. Watanabe, "Optimality of the recursive data exchange protocol," *In Preparation*.

[27] ——, "A bound for multiparty secret key agreement and implications for a problem of secure computing," in *EUROCRYPT*, 2014, pp. 369–386.

[28] ——, "Converses for secret key agreement and secure computing," *CoRR*, vol. abs/1404.5715, 2014.

[29] Y. Wu and P. Yang, “Minimax rates of entropy estimation on large alphabets via best polynomial approximation,” *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3702–3720, June 2016.

[30] E.-H. Yang and D.-K. He, “Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder,” *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1808–1824, April 2010.

[31] J. Ziv and A. Lempel, “Compression of individual sequences via variable-rate coding,” *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530–536, September 1978.