# Distributed Computing With Privacy

Himanshu Tyagi[†]

***Abstract*—A set of terminals that observe correlated data seek to compute functions of the data using interactive public communication. At the same time it is required that this communication, observed by an eavesdropper, does not reveal the value of a private function of the data. In general, the private function and the functions computed by the terminals can be all different. We show that a class of functions are securely computable if and only if the conditional entropy of data given the value of private function is greater than the least rate of interactive communication required for an appropriately chosen multiterminal source coding task. A single-letter formula is provided for this rate in special cases.**

## I. Introduction

We consider the following problem of distributed function computation under privacy constraints. The terminals in a set $\mathcal{M} = \{1, ..., m\}$ observe correlated data, and wish to compute functions $g_1, ..., g_m$, respectively, of their collective data. To this end, they communicate interactively over a noiseless channel of unlimited capacity. It is required that this communication must not reveal the value of a specified private function $g_0$ of the data. If such a communication protocol exists, the functions are said to be *securely computable*. We formulate a Shannon theoretic multiterminal source model that addresses the basic question: *When are the functions $g_0, ..., g_m$ securely computable?*

The answer to the general question above is not known. The case when the terminals in subset $A$ of $\mathcal{M}$ compute only the private function $g_0$ and those not in $A$ perform no computation was dealt with in [10]. Loosely speaking, denoting the collective data of the terminals by the random variable (rv) $X_{\mathcal{M}}$ and the random value of the function $g_0$ by the rv $G_0$, the maximum rate of randomness (in the data) that is independent of $G_0$ is $H(X_{\mathcal{M}}|G_0)$. It was shown in [10] that if $g_0$ is securely computable (by the terminals in $A$), then

$$H(X_{\mathcal{M}}|G_0) = H(X_{\mathcal{M}}) - H(G_0) \geq R^*, \tag{1}$$

and $g_0$ is securely computable if

$$H(X_{\mathcal{M}}|G_0) > R^*, \tag{2}$$

where $R^*$ has the operational significance of being the minimum overall rate of communication needed for a specific multiterminal source-coding task; this task does not involve any security constraint.

In this paper, we give necessary and sufficient conditions for secure computation of given functions. We extend the results of [10] and identify appropriate quantities $R^*$ to establish necessary and sufficient conditions of the same form as (1) and (2), respectively, for a broad class of settings involving the secure computation of multiple functions. The simpler case when the computed functions $g_1, ..., g_m$ correspond to the observations of subsets of terminals is studied separately as a problem of secure multiterminal source coding.

Under the sufficient condition (2), the secure computing scheme in [10] recovered the entire data at the (function seeking) terminals in $A$ using communication that is independent of $G_0$. Similarly, when one of the terminals computes the private function $g_0$, i.e., $g_i = g_0$ for some $i \in \mathcal{M}$, our secure computing scheme enables the recovery of entire data at the terminal $i$.

Unlike [10], we do not provide a single-letter formula for the quantity $R^*$, in general; nevertheless, conditions (1) and (2) provide a structural characterization of securely computable functions. Moreover, for special cases the conditions do take a single-letter form (see Example 1 and Corollary 4 below).

The problem of secure computing for multiple functions is formulated in the next section, followed by our main results in section III. Partial proof of our main result is sketched in the last section.

*Notation.* The set $\{1, ..., m\}$ is denoted by $\mathcal{M}$. For $i < j$, denote by $[i, j]$ the set $\{i, ..., j\}$. Let $X_1, ..., X_m, m \geq 2$, be rvs taking values in finite sets $\mathcal{X}_1, ..., \mathcal{X}_m$, respectively, and with a known probability mass function. Denote by $X_{\mathcal{M}}$ the collection of rvs $(X_1, ..., X_m)$, and by $X_{\mathcal{M}}^n = (X_1^n, ..., X_m^n)$ the $n$ independent and identically distributed (i.i.d). repetitions of rvs $X_{\mathcal{M}}$. For a subset $A$ of $\mathcal{M}$, denote by $X_A$ the rvs $(X_i, i \in A)$. Given $R_i \geq 0$, $1 \leq i \leq m$, let $R_A$ denote the sum $\sum_{i \in A} R_i$.

Finally, for $0 \leq \epsilon < 1$, we say an rv $U$ is $\epsilon$-recoverable from an rv $V$ if there exists a function $g$ of $V$ such that $\Pr(U = g(V)) \geq 1 - \epsilon$.

## II. Problem formulation

We consider a multiterminal source model for function computation using public communication, under privacy constraints. This basic model was introduced in [4] in a separate context of SK generation with public transaction. Terminals $1, \ldots, m$ observe, respectively, the sequences $X_1^n, \ldots, X_m^n$ of length $n$. For $0 \leq i \leq m$, let $g_i : \mathcal{X}_{\mathcal{M}} \to \mathcal{Y}_i$ be given mappings, where the sets $\mathcal{Y}_i$ are finite. Further, for $0 \leq i \leq m$ and $n \geq 1$, the (single-letter) mapping $g_i^n : \mathcal{X}_{\mathcal{M}}^n \to \mathcal{Y}_i^n$ is

This work was supported by the U.S. National Science Foundation under Grants CCF0830697 and CCF1117546.

[†]Department of Electrical and Computer Engineering, and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA. Email: tyagi@umd.edu.

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

defined by

$$g_i^n(x_{\mathcal{M}}^n) = (g_i(x_{11}, \dots, x_{m1}), \dots, g_i(x_{1n}, \dots, x_{mn})),$$
$$x_{\mathcal{M}}^n = (x_1^n, \dots, x_m^n) \in \mathcal{X}_{\mathcal{M}}^n.$$

For convenience, we shall denote the rv $g_i^n\left(X_{\mathcal{M}}^n\right)$ by $G_i^n, n \geq 1$, and, in particular, $G_i^1 = g_i\left(X_{\mathcal{M}}\right)$ simply by $G_i$.

Each terminal $i \in \mathcal{M}$ wishes to compute the function $g_i^n(x_{\mathcal{M}}^n)$, without revealing $g_0^n(x_{\mathcal{M}}^n)$, $x_{\mathcal{M}}^n \in \mathcal{X}_{\mathcal{M}}^n$. To this end, the terminals are allowed to communicate over a noiseless public channel, possibly interactively in several rounds.

**Definition 1.** An *$r$-rounds interactive communication protocol* consists of mappings

$$f_{11}, ..., f_{1m}, ...., f_{r1}, ..., f_{rm},$$

where $f_{ij}$ denotes the communication sent by the $j$th node in the $i$th round of the protocol; specifically, $f_{ij}$ is a function of $X_j^n$ and the communication sent in the previous rounds $\{f_{kl} : 1 \leq k \leq i-1, l \in \mathcal{M}\}$. Denote the rv corresponding to the communication by

$$\mathbf{F} = F_{11}, ..., F_{1m}, ...., F_{r1}, ..., F_{rm},$$

noting that $\mathbf{F} = \mathbf{F}^{(n)}\left(X_{\mathcal{M}}^n\right)$. The rate[1] of $\mathbf{F}$ is $\frac{1}{n}\log\|\mathbf{F}\|$.

**Definition 2.** For $\epsilon_n > 0, n \geq 1$, we say that functions[2] $g_{\mathcal{M}} = (g_0, g_1, ..., g_m)$, with private function $g_0$, are *$\epsilon_n$-securely computable* ($\epsilon_n$- SC) from observations of length $n$, and public communication $\mathbf{F} = \mathbf{F}^{(n)}$, if

(i) $G_i^n$ is $\epsilon_n$- recoverable from $(X_i^n, \mathbf{F})$ for every $i \in \mathcal{M}$, and

(ii) $\mathbf{F}$ satisfies the secrecy condition

$$\frac{1}{n} I\left(G_0^n \wedge \mathbf{F}\right) \leq \epsilon_n.$$

*Remark.* The definition of secrecy here corresponds to "weak secrecy" [1], [7]. When our results have a single-letter form, our achievability schemes for secure computing attain "strong secrecy" in the sense of [8], [2], [4]. In fact, when we have a single-letter form, our proof can be modified to yield "strong secrecy."

By definition, for $\epsilon_n$-SC functions $g_{\mathcal{M}}$, the private function $G_0$ is effectively concealed from an eavesdropper with access to the public communication $\mathbf{F}$.

**Definition 3.** For private function $g_0$, we say that functions $g_{\mathcal{M}}$ are *securely computable* if $g_{\mathcal{M}}$ are $\epsilon_n$- SC from observations of length $n$ and public communication $\mathbf{F} = \mathbf{F}^{(n)}$, such that $\lim_n \epsilon_n = 0$.

Figure 1 shows the setup for secure computing.

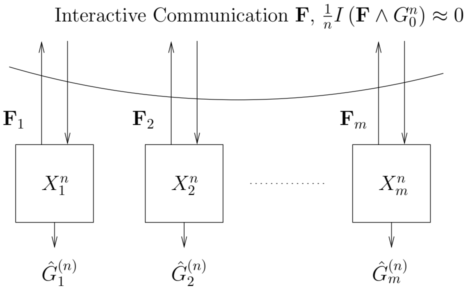

Fig. 1. Secure computation of $g_1, ..., g_m$ with private function $g_0$

In this paper, we give necessary and sufficient conditions for the secure computability of certain classes of functions $g_{\mathcal{M}} = (g_0, g_1, ..., g_m)$. There are three classes of problems studied. In the first two classes, we require at least one of the terminals to compute the private function $g_0$, i.e. $g_i = g_0$ for some $1 \leq i \leq m$. The subclass with the additional restriction $g_i = g_i\left(g_0\right)$ for all $1 \leq i \leq m$ (along with $g_i = g_0$ for some $i$) allows for a more structured secure computing protocol, and is studied first. We note that the formulation in [10], in which the terminals in a given subset $A$ of $\mathcal{M}$ are required to compute (only) $g_0$ securely, is a further particularization with

$$g_i = \begin{cases} g_0, & i \in A, \\ \text{constant}, & \text{otherwise}. \end{cases} \tag{3}$$

It was shown in [10] that (1) and (2) constitute, respectively, necessary and sufficient conditions for the functions above to be securely computable, with $R^*$ being the minimum rate of interactive communication $\mathbf{F}$ that enables all the terminals in $\mathcal{M}$ to attain *omniscience* (see [4]), i.e., recover *all* the data $X_{\mathcal{M}}^n$, using $\mathbf{F}$ and the *decoder side information* $G_0^n$ given to the terminals in $\mathcal{M} \setminus A$. In fact, it was shown that when condition (2) holds, it is possible to recover $X_{\mathcal{M}}^n$ using communication that is independent of $G_0^n$.

The last class of problems we study is a generalization of the previous instance of *secure multiterminal source coding*. Specifically, we consider the situation where each terminal wishes to recover some subset $X_{\mathcal{M}_i}^n$ of the sources where $\mathcal{M}_i \subseteq \mathcal{M} \setminus \{i\}$, i.e.,

$$g_i\left(X_{\mathcal{M}}\right) = X_{\mathcal{M}_i}, \quad i \in \mathcal{M}. \tag{4}$$

While a characterization of securely computable functions in the general sense of Definition 3 is unresolved, for the specific classes above we provide matching necessary and sufficient conditions for the secure computability of $g_{\mathcal{M}}$. The guiding heuristic in this work is the following generalized interpretation of the results of [10]: Conditions (1) and (2) constitute, respectively, the necessary and sufficient conditions for functions $g_{\mathcal{M}} = (g_0, g_1, ..., g_m)$ to be securely computable, where $R^*$ is the infimum of the rates of interactive communication $\mathbf{F}'$ in a multiterminal source coding problem described below:

For each $1 \leq i \leq m$, the following must hold simultaneously:

(P1) $G_i^n$ is $\epsilon_n$-recoverable from $(X_i^n, \mathbf{F}')$, and

(P2) $X_{\mathcal{M}}^n$ is $\epsilon_n$-recoverable from $(X_i^n, G_0^n, \mathbf{F}')$, i.e., the terminals attain omniscience, with $G_0^n$ as side information that is used only for decoding (but is not used for the communication $\mathbf{F}'$),

[1] All logarithms are with respect to the base 2.

[2] The abuse of notation $g_{\mathcal{M}} = (g_0, g_1, ..., g_m)$ simplifies our presentation.

where $\epsilon_n \to 0$ as $n \to \infty$[3]. For the specific case in (3), $R^*$ above has a single-letter formula. In general, a single-letter expression for $R^*$ is not known.

Our results, described in section III, are obtained by simple adaptations of this principle. Unlike [10], our conditions, in general, are not of a single-letter form. Nevertheless, they provide a structural characterization of secure computability. As an application, our results provide simple conditions for secure computability in the following illustrative example.

*Example* 1. We consider the case of $m = 2$ terminals that observe binary symmetric sources (BSS) with underlying rvs $X_1, X_2$ with joint pmf given by

$$\Pr(X_1 = 0, X_2 = 0) = \Pr(X_1 = 1, X_2 = 1) = \frac{1-\delta}{2},$$
$$\Pr(X_1 = 0, X_2 = 1) = \Pr(X_1 = 1, X_2 = 0) = \frac{\delta}{2},$$

where $0 < \delta < 1/2$. The results of this paper will allow us to provide conditions for the secure computability of the four choices of $g_0, g_1, g_2$ below; it will follow by Theorem 1 that functions $g_0, g_1, g_2$ are securely computable if

$$h(\delta) < \tau,$$

and conversely, if the functions above are securely computable, then

$$h(\delta) \leq \tau,$$

where $h(\tau) = -\tau \log \tau - (1-\tau)\log(1-\tau)$, and the constant $\tau = \tau(\delta)$ depends on the choice of the function. These characterizations are summarized in the next table.

| $g_0$ | $g_1$ | $g_2$ | $\tau$ |
|---|---|---|---|
| $X_1 \oplus X_2$ | $X_1 \oplus X_2$ | $X_1 \oplus X_2$ | $1/2$ |
| $X_1 \oplus X_2$ | $X_1 \oplus X_2$ | $\phi$ | $1$ |
| $X_1 \oplus X_2,\ X_1.X_2$ | $X_1 \oplus X_2,\ X_1.X_2$ | $X_1.X_2$ | $2\delta/3$ |
| $X_1 \oplus X_2$ | $X_1 \oplus X_2$ | $X_1.X_2$ | $2/3$ |

The results for the first two settings follow from [10]. The third and fourth results are new. In these settings, terminal 1 is required to recover the private function; our results below show that the conditions for the secure computability in these cases remain unchanged even if this terminal is required to attain omniscience. Note that since $h(\delta) < 1$ for all $0 < \delta < 1/2$, there exists a communication protocol for securely computing the functions in the second setting. By contrast, a secure computing protocol for the functions in the third setting does not exist for any $0 < \delta < 1/2$, since $h(\delta) > 2\delta/3$. □

## III. Characterization of securely computable functions

In this section, we characterize securely computable functions for three settings. Our necessary and sufficient conditions entail the comparison of $H(X_{\mathcal{M}}|G_0)$ with a rate $R^*$; the specific choice of $R^*$ depends on the functions $g_{\mathcal{M}}$.

**(1)** For $0 < m_0 < m$, and for private function $g_0$, let

$$g_i = \begin{cases} g_0, & i \in [1, m_0], \\ g_i(g_0), & i \in [m_0 + 1, m]. \end{cases} \tag{5}$$

Denote by $\mathcal{R}_1^*(g_{\mathcal{M}})$ the closure of the (nonempty) set of pairs

$$\left(R_{\mathbf{F}}^{(1)}, \frac{1}{n} I(G_0^n \wedge \mathbf{F})\right),$$

for all $n \geq 1$ and interactive communication $\mathbf{F}$, where

$$R_{\mathbf{F}}^{(1)} = \frac{1}{n} H(\mathbf{F}) + \frac{1}{n} \sum_{i=m_0+1}^{m} H(G_i^n | X_i^n, \mathbf{F}) + \inf R_{\mathcal{M}}, \tag{6}$$

with the infimum taken over rates $R_1, ..., R_m$ satisfying the following constraints:

**(1a)** $\forall L \subsetneq \mathcal{M}$, $[1, m_0] \nsubseteq L$,
$$R_L \geq \frac{1}{n} H\left(X_L^n | X_{\mathcal{M}\setminus L}^n, \mathbf{F}\right);$$

**(1b)** $\forall L \subsetneq \mathcal{M}$, $[1, m_0] \subseteq L$,
$$R_L \geq \frac{1}{n} H\left(X_L^n | X_{\mathcal{M}\setminus L}^n, G_0^n, \mathbf{F}\right).$$

The quantity $\inf_{n,\mathbf{F}} R_{\mathbf{F}}^{(1)}$ corresponds to the solution of a multiterminal source coding problem. Specifically, it is the infimum of the rates of interactive communication that satisfy (P1) and (P2) above (see [3, Theorem 13.5], [4]).

Note that the SK-based scheme for secure computing given in section IV relies critically on the functional relation $g_i = g_i(g_0)$.

**(2)** The next case is a relaxation of the previous model in that the restriction $g_i = g_i(g_0)$ for $i \in [m_0 + 1, m]$ is dropped. Let $\mathcal{R}_2^*(g_{\mathcal{M}})$ denote the closure of the set of pairs

$$\left(R_{\mathbf{F}}^{(2)}, \frac{1}{n} I(G_0^n \wedge \mathbf{F})\right),$$

for all $n \geq 1$ and interactive communication $\mathbf{F}$, where

$$R_{\mathbf{F}}^{(2)} = \frac{1}{n} H(\mathbf{F}) + \inf\left[R'_{[m_0+1,m]} + R_{\mathcal{M}}\right], \tag{7}$$

with the infimum taken over rates $R_1, ..., R_m$ and $R'_{m_0+1}, ..., R'_m$ satisfying the following constraints:

**(2a)** $\forall L \subsetneq \mathcal{M}$, $[1, m_0] \nsubseteq L$,
$$R_L \geq \frac{1}{n} H\left(X_L^n | X_{\mathcal{M}\setminus L}^n, \mathbf{F}\right);$$

**(2b)** for $m_0 < j \leq m$,
$$R'_j \geq \frac{1}{n} H\left(G_j^n | X_j^n, \mathbf{F}\right);$$

**(2c)** $\forall L \subseteq \mathcal{M}, [1, m_0] \subseteq L$, and $L' \subseteq [m_0 + 1, m]$ with either $L \neq \mathcal{M}$ or $L' \neq [m_0 + 1, m]$,
$$R'_{L'} + R_L$$
$$\geq \frac{1}{n} H\left(G_{L'}^n, X_L^n | G_{[m_0+1,m]\setminus L'}^n, X_{\mathcal{M}\setminus L}^n, G_0^n, \mathbf{F}\right).$$

[3] Thus, (P1) and (P2) require any terminal computing $g_0$ to become omniscient.

The quantity $\inf_{n,\mathbf{F}} R_{\mathbf{F}}^{(2)}$ corresponds to the solution of a multiterminal source coding problem, and is the infimum of the rates of interactive communication $\mathbf{F}'$ that satisfy (P1) and (P2) above, and additionally satisfies:

(P3) $X_{\mathcal{M}}^n$ is $\epsilon_n$-recoverable from $\left(G_j^n, G_0^n, \mathbf{F}'\right)$, $m_0 < j \leq m$.

This modification corresponds to the introduction of $m - m_0$ dummy terminals, with the $j$th dummy terminal observing $G_j^n$, $m_0 < j \leq m$; the dummy terminals can be realized by a terminal $i$ in $[1, ..., m_0]$ that recovers $X_{\mathcal{M}}^n$ from $(X_i^n, \mathbf{F})$. As remarked before, the SK-based secure computing scheme in section IV does not work for this case, and a secure computing scheme based on the dummy terminals above is used instead. The conditions (P2) and (P3) above correspond to omniscience at the terminals in the extended model, with $G_0^n$ provided as side information only for decoding.

**(3)** The last case concerns multiterminal source coding without revealing the private data (see (4)). Denote by $\mathcal{R}_3^*(g_{\mathcal{M}})$ the closure of the set of pairs

$$\left(R_{\mathbf{F}}^{(3)}, \frac{1}{n} I\left(G_0^n \wedge \mathbf{F}\right)\right),$$

for all interactive communication $\mathbf{F}$, where

$$R_{\mathbf{F}}^{(3)} = \frac{1}{n} H(\mathbf{F}) + \inf R_{\mathcal{M}}, \tag{8}$$

with rates $R_1, ..., R_m$ satisfying the following constraints:

**(3a)** For $1 \leq i \leq m$, $\forall\, L \subseteq \mathcal{M}_i \subseteq \mathcal{M} \setminus \{i\}$,

$$R_L \geq \frac{1}{n} H\left(X_L^n | X_{\mathcal{M}_i \setminus L}^n, X_i^n, \mathbf{F}\right);$$

**(3b)** $\forall L \subsetneq \mathcal{M}$,

$$R_L \geq \frac{1}{n} H\left(X_L^n | X_{\mathcal{M} \setminus L}^n, G_0^n, \mathbf{F}\right).$$

As before, the quantity $\inf_{n,\mathbf{F}} R_{\mathbf{F}}^{(3)}$ corresponds to the infimum of the rates of interactive communication that satisfy (P1) and (P2) above.

Our main result below characterizes securely computable functions for the three settings above.

**Theorem 1.** *For $i = 1, 2, 3$, with functions $g_0, g_1, ..., g_m$ as in the case $(i)$ above, the functions $g_{\mathcal{M}}$ are securely computable if the following condition holds:*

$$H\left(X_{\mathcal{M}} | G_0\right) > R_i^*\left(g_{\mathcal{M}}\right). \tag{9}$$

*Conversely, if the functions above are securely computable, then*

$$H\left(X_{\mathcal{M}} | G_0\right) \geq R_i^*\left(g_{\mathcal{M}}\right), \tag{10}$$

*where*

$$R_i^*\left(g_{\mathcal{M}}\right) = \inf_{(x,0) \in \mathcal{R}_i^*(g_{\mathcal{M}})} x, \quad i = 1, 2, 3. \tag{11}$$

Theorem 1 affords the following heuristic interpretation. The quantity $H\left(X_{\mathcal{M}} | G_0\right)$ represents the maximum rate of randomness in $X_{\mathcal{M}}^n$ that is (nearly) independent of $G_0^n$. On the other hand, $R_i^*\left(g_{\mathcal{M}}\right)$ is an appropriate rate of communication for the computation of $g_{\mathcal{M}}$; we show that latter being less than $H\left(X_{\mathcal{M}} | G_0\right)$ guarantees the secure computability of $g_{\mathcal{M}}$.

Although the conditions for secure computability above are not of a single-letter form in general, they do reduce to such a form for specific instances. The following result provides a sufficient condition for obtaining single-letter conditions for characterizing securely computable functions.

**Lemma 2.** *For case $(i)$, $i = 1, 2, 3$, if for all $n \geq 1$ and interactive communication* $\mathbf{F}$

$$R_{\mathbf{F}}^{(i)} \geq R_{\mathbf{F}}^{(i)} \,|_{\mathbf{F}=\textit{constant}} =: \; R_{\textit{constant}}^{(i)}, \tag{12}$$

*then $R_i^*\left(g_{\mathcal{M}}\right) = R_{\textit{constant}}^{(i)} = \inf_{n,\mathbf{F}} R_{\mathbf{F}}^{(i)}$.*

The proof is a simple consequence of the definition of $R_i^*\left(g_{\mathcal{M}}\right)$ in (11). Note that $R_{\text{constant}}^{(i)}$ has a single-letter form.

*Remark.* As mentioned before, the quantity $\inf_{n,\mathbf{F}} R_{\mathbf{F}}^{(i)}$ is the infimum of the rates of interactive communication that satisfies (P1),(P2) for $i = 1, 3$, and satisfies (P1)-(P3) for $i = 2$. Thus, when the conditions of Lemma 2 hold, we have from Theorem 1 that $g_{\mathcal{M}}$ are securely computable if

$$H\left(X_{\mathcal{M}} | G_0\right) > R_{\text{constant}}^{(i)},$$

and if $g_{\mathcal{M}}$ are securely computable then

$$H\left(X_{\mathcal{M}} | G_0\right) \geq R_{\text{constant}}^{(i)},$$

where $R_{\text{constant}}^{(i)}$ is the minimum rate of communication that satisfies (P1), (P2) for $i = 1, 3$, and satisfies (P1)-(P3) for $i = 2$.

As a consequence of Lemma 2, we obtain below a single-letter characterization of securely computable functions, with $m = 2$, in a special case; the following lemma, which is a special case of [5, Lemma B.1] (see also [6, Theorem 1]), is instrumental to our proof.

**Lemma 3.** *Let $m = 2$. For an interactive communication* $\mathbf{F}$*, we have*

$$H(\mathbf{F}) \geq H\left(\mathbf{F} | X_1^n\right) + H\left(\mathbf{F} | X_2^n\right).$$

We next consider case (1) for two terminals.

**Corollary 4.** *For $m = 2$, for functions $g_0, g_1, g_2$ with $g_1 = g_0$ and $g_2 = g_2\left(g_0\right)$, we have*

$$R_1^*\left(g_{\mathcal{M}}\right) = H\left(X_2 | X_1\right) + H\left(G_2 | X_2\right) + H\left(X_1 | X_2, G_0\right). \tag{13}$$

*Proof:* The constraints (1a) and (1b) satisfied by rates $R_1, R_2$ in the definition of $R_{\mathbf{F}}^{(1)}$ are

$$R_2 \geq \frac{1}{n} H\left(X_2^n | X_1^n, \mathbf{F}\right),$$
$$R_1 \geq \frac{1}{n} H\left(X_1^n | X_2^n, G_0^n, \mathbf{F}\right),$$

which further yields

$$R_{\mathbf{F}}^{(1)} = \frac{1}{n}\left[H\left(\mathbf{F}\right) + H\left(G_2^n | X_2^n, \mathbf{F}\right)\right.$$
$$\left. + H\left(X_2^n | X_1^n, \mathbf{F}\right) + H\left(X_1^n | X_2^n, G_0^n, \mathbf{F}\right)\right]. \tag{14}$$

Thus, $R^{(1)}_{\text{constant}}$ equals the term on the right side of (13). From $H\left(G_2|G_0\right)=0$, and the expression for $R^{(1)}_{\mathbf{F}}$ above, we have

$$\begin{aligned} R^{(1)}_{\mathbf{F}} &\geq \frac{1}{n}\left[H(\mathbf{F}) - H\left(\mathbf{F}|X_1^n\right) - H\left(\mathbf{F}|X_2^n\right)\right] + R^{(1)}_{constant} \\ &\geq R^{(1)}_{constant}, \end{aligned}$$

where the last inequality follows from Lemma 3. The result then follows from Lemma 2. □

We next derive simple conditions for secure computability for the BSS in Example 1

*Example* 2. Consider the setup of Example 1, with $g_0 = g_1 = X_1 \oplus X_2, X_1.X_2$ and $g_2 = X_1.X_2$. By Corollary 4 and the observation $H\left(G_2|X_2\right) = h(\delta)/2$, we get $R_1^*\left(g_{\mathcal{M}}\right) = 3h(\delta)/2$. Since $H\left(X_1, X_2 \mid G_0\right) = H\left(X_1, X_2 \mid X_1 \oplus X_2\right) - H\left(X_1.X_2 \mid X_1 \oplus X_2\right) = \delta$, the characterization of secure computability claimed in Example 1 follows from Theorem 1. □

*Example* 3. In the setup of Example 1, consider $g_0 = g_1 = X_1 \oplus X_2$ and $g_2 = X_1.X_2$. This choice of $g_0, g_1, g_2$ is an instance of case (2) above. For an interactive communication $\mathbf{F}$, a manipulation of constraints (2a), (2b), (2c) in the definition of $R^{(2)}_{\mathbf{F}}$, yields

$$\begin{aligned} R^{(2)}_{\mathbf{F}} = \frac{1}{n}&\left[H(\mathbf{F}) + H\left(X_1^n|X_2^n, G_0^n, G_2^n, \mathbf{F}\right)\right. \\ &+ \max\left\{H\left(X_2^n|G_0^n, G_2^n, \mathbf{F}\right), H\left(X_2^n|X_1^n, \mathbf{F}\right)\right\} \\ &\left.+ H\left(G_2^n|X_2^n, \mathbf{F}\right)\right]. \end{aligned} \tag{15}$$

It follows from $H\left(X_1^n|X_2^n, G_0^n, G_2^n, \mathbf{F}\right) = 0$ that

$$\begin{aligned} R^{(2)}_{constant} &= H\left(G_2|X_2\right) + \max\left\{H\left(X_2|G_0, G_2\right), H\left(X_2|X_1\right)\right\} \\ &= \frac{h(\delta)}{2} + \max\left\{\delta, h(\delta)\right\} = \frac{3}{2}h(\delta), \end{aligned} \tag{16}$$

as $h(\delta) > \delta$ for $0 < \delta < 1/2$.

Next, note from (15) that for any interactive communication $\mathbf{F}$,

$$\begin{aligned} R^{(2)}_{\mathbf{F}} \geq \frac{1}{n}&\left[H(\mathbf{F}) - H\left(\mathbf{F}|X_1^n\right) - H\left(\mathbf{F}|X_2^n\right)\right] \\ &+ H\left(G_2|X_2\right) + H\left(X_2|X_1\right) \\ &\geq H\left(G_2|X_2\right) + H\left(X_2|X_1\right) = \frac{3}{2}h(\delta), \end{aligned} \tag{17}$$

where the last inequality above follows from Lemma 3. The characterization in Example 1 follows from (16), (17), and $H\left(X_1, X_2|G_0\right) = 1$, using Lemma 2 and Theorem 1. □

## IV. Outline of proof of Theorem 1

In this section we present the key ideas in the proof of sufficiency part of Theorem 1 for case (1). The sufficiency proof for other cases, and the proof of necessity, are omitted due to space constraints.

If $H\left(X_{\mathcal{M}}|G_0\right) > R_1^*\left(g_{\mathcal{M}}\right)$ holds, then from the definition of $R_1^*\left(g_{\mathcal{M}}\right)$, for all sufficiently small $\epsilon > 0$ there exist $n \geq 1$ and interactive communication $\mathbf{F} = \mathbf{F}\left(X_{\mathcal{M}}^n\right)$ such that

$$\frac{1}{n} I\left(G_0^n \wedge \mathbf{F}\right) < \epsilon,$$

and

$$\frac{1}{n} H\left(X_{\mathcal{M}}^n|G_0^n, \mathbf{F}\right) > \frac{1}{n}\sum_{j=m_0+1}^{m} H\left(G_j^n|X_j^n, \mathbf{F}\right) + R_{\mathcal{M}}, \tag{18}$$

where $R_1, ..., R_m$ satisfy the constraints (1a), (1b). From (18), using the approach of the proof of sufficiency in [10, Theorem 5], we first show the existence of an interactive communication $\mathbf{F}' = \mathbf{F}'\left(X_{\mathcal{M}}^N\right)$ that is almost independent of $G_0^N$, and attains omniscience at the terminals in $\mathcal{M}$, with side information $G_0^N$ given for decoding to the terminals in $[m_0+1, m]$, for $N = nk$ sufficiently large; the interactive communication $\mathbf{F}'$ includes $\mathbf{F}$. Next, for $m_0 < j \leq m$, denoting by $\hat{F}_j$ the Slepian-Wolf codeword for $G_j^N$ given $X_j^N$ and $\mathbf{F}'$, we show the existence of rvs $K_j = K_j\left(X_j^N\right)$ of approximate rate

$$\frac{1}{N} H\left(G_j^N|X_j^N, \mathbf{F}'\right),$$

that are almost independent of

$$\left(G_0^N, \mathbf{F}', K_l \oplus \hat{F}_l, 1 \leq l \leq j-1\right);$$

hence, the communication $\left(\mathbf{F}', K_l \oplus \hat{F}_l, m_0 < l \leq m\right)$ is almost independent of $G_0^N$.

For $m_0 < l \leq m$, $K_l \oplus \hat{F}_l$ is used as a one-time-pad to send $\hat{F}_l$, and so $G_j^N$, to terminal $l$ observing $\left(X_l^N, \mathbf{F}'\right)$. The existence of $K_{m_0+1}, ..., K_m$ follows by extending the proof of [9, Theorem 4], using (18), and the observation

$$\frac{1}{N} H\left(X_j^N|G_0^N, \mathbf{F}'\right) \approx \frac{1}{N} H\left(X_{\mathcal{M}}^N|G_0^N, \mathbf{F}'\right).$$

Therefore, $g_{\mathcal{M}}$ is $\epsilon$-SC for all $\epsilon$ sufficiently small.

## Acknowledgement

This work benefited greatly from the discussions with Prof. Prakash Narayan. Further, his comments on the manuscript helped to improve the presentation.

## References

[1] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography–part i: Secret sharing,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, 1993.

[2] I. Csiszár, “Almost independence and secrecy capacity,” *Prob. Pered. Inform.*, vol. 32, no. 1, pp. 48–57, 1996.

[3] I. Csiszár and J. Körner, *Information theory: Coding Theorems for Discrete Memoryless Channels*. 2nd Edition. Cambridge University Press, 2011.

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

[5] I. Csiszár and P. Narayan, “Secrecy capacities for multiterminal channel models,” *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2437–2452, 2008.

[6] M. Madiman and P. Tetali, “Information inequalities for joint distributions, with interpretations and applications,” *IEEE Trans. Inform. Theory*, vol. 56, pp. 2699–2713, 2010.

[7] U. M. Maurer, “Secret key agreement by public discussion from common information,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, May 1993.

[8] ——, *Communications and Cryptography: Two sides of One Tapestry*, R.E. Blahut et al., Eds. ed. Norwell, MA: Kluwer, 1994, ch. 26, pp. 271–285.

[9] H. Tyagi, P. Narayan, and P. Gupta, “Secure computing,” *Proc. Int. Symp. Inform. Theory*, pp. 2612 – 2616, June 2010.

[10] H. Tyagi, P. Narayan, and P. Gupta, “When is a function securely computable?” *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6337–6350, Oct 2011.