# On the Communication Complexity of Secure Computation

Deepesh Data and Vinod M. Prabhakaran School of Technology and Computer Science Tata Institute of Fundamental Research Mumbai 400 005

Manoj M. Prabhakaran Department of Computer Science University of Illinois, Urbana-Champaign Urbana, IL 61801

July 11, 2018

#### Abstract

Information theoretically secure multi-party computation (MPC) has been a central primitive of modern cryptography. However, relatively little is known about the communication complexity of this primitive.

In this work, we develop powerful information theoretic tools to prove lower bounds on the communication complexity of MPC. We restrict ourselves to a concrete setting involving 3-parties, in order to bring out the power of these tools without introducing too many complications. Our techniques include the use of a data processing inequality for residual information — i.e., the gap between mutual information and Gács-Körner common information, a new information inequality for 3-party protocols, and the idea of distribution switching by which lower bounds computed under certain worst-case scenarios can be shown to apply for the general case.

Using these techniques we obtain tight bounds on communication complexity by MPC protocols for various interesting functions. In particular, we show concrete functions which have "communication-ideal" protocols, which achieve the minimum communication simultaneously on all links in the network. Also, we obtain the first *explicit* example of a function that incurs a higher communication cost than the input length, in the secure computation model of Feige, Kilian and Naor [FKN94], who had shown that such functions exist. We also show that our communication bounds imply tight lower bounds on the amount of randomness required by MPC protocols for many interesting functions.

We identify a *multi-secret sharing* primitive that is interesting on its own right, but also has the property that lower bounds on its share sizes serve as lower bounds for communication complexity of MPC protocols. While often the resulting bounds are tight, we can use our results to give a concrete example where there is a gap between the share sizes and the communication complexity.

# Contents

T	Introduction	1
	1.1 Results and Techniques	2
	1.2 Related Work	4
<b>2</b>	Preliminaries	4
3	Lower Bounds on Communication Complexity         3.1 Preliminary Lower Bounds	7 7 8 10
4	Lower Bounds on Randomness	13
5	Application to Specific Functions	14
$\mathbf{A}$	Preliminaries: More Details	18
В	Entropy Lower Bounds as Communication Lower Bounds	20
$\mathbf{C}$	Connections to Secure Sampling and Correlated Multi-Secret Sharing	20
D	Proofs of the Main Theorems	24
${f E}$	Proofs Omitted from Section 3	<b>25</b>
$\mathbf{F}$	Details omitted from Section 5	28
	F.1 Secure Computation of GROUP-ADD	28
	F.2 Secure Computation of SUM	29
	F.3 Secure Computation of CONTROLLED ERASURE	30
	F.4 Secure Computation of REMOTE $\binom{m}{1}$ -OT $\frac{n}{2}$	32
	F.5 Secure Computation of AND	33
$\mathbf{G}$	Lower Bounds for Dependent Inputs	34
Н	Dependence on Input Distributions	35

## 1 Introduction

Information theoretically secure multi-party computation has been a central primitive of modern cryptography. The seminal results of Ben-Or, Goldwasser, and Wigderson [BGW88] and Chaum, Crépeau, and Damgård [CCD88] showed that information theoretically secure function computation is possible between parties connected by pairwise, private links as long as only a strict minority may collude in the honest-but-curious model (and a strictly less than one-third minority may collude in the malicious model). Since then, several protocols have improved the efficiency of these protocols.

However, relatively less is known about lower bounds on the amount of communication required by a secure multi-party computation protocol, with a few notable exceptions [Kus89, FY92, CK93, FKN94]. In fact, [IK04] shows that establishing strong communication lower bounds (even with restrictions on the number of rounds) would imply breakthrough lower bound results for other well-studied problems like private-information retrieval and locally decodable codes. Further, due to known upper bounds on the communication needed in a secure multi-party computation protocol [DI06], such lower bounds would imply non-trivial circuit complexity lower bounds — a notoriously hard problem in theoretical computer science. The goal of this work is to develop tools to tackle the difficult problem of lower bounds for communication in secure multi-party computation, even if they do not immediately have direct implications to circuit complexity or locally decodable codes.

It is instructive to compare the problem of communication complexity lower bounds for secure multiparty computation with that when there is no security requirement involved. This latter problem has been extensively studied — over the last three and a half decades, starting with [Yao79] — resulting in a rich collection of results and techniques. Unfortunately, many of the techniques in the communication complexity setting are not relevant in the setting of secure computation: for instance, for communication complexity Yao's minimax theorem allows one to consider only deterministic protocols with public randomness, but in the secure computation setting, one must allow private randomness, and hence it is not sufficient to consider only deterministic protocols. This rules out several powerful combinatorial approaches from the communication complexity literature. But over the last decade or so (see for example, [KLL+12] and references therein), a slew of information theoretic tools have been developed, which in many cases subsume more complicated combinatorial approaches.

Following this lead, the approach we take in this work is to develop novel information-theoretic tools to obtain lower bounds on the communication complexity of secure computation. Indeed, the tools we develop and use have connections with similar tools developed in the context of communication complexity and related problems. In particular, all these tools are related to notions of "common information" introduced by Gács-Körner [GK73] and Wyner [Wyn75].<sup>2</sup>

In this work we restrict our study to a concrete setting that brings out the power of these tools without introducing too many additional complications. Our setting involves 3 parties (with security against corruption of any single party) of which only two parties have inputs, X and Y, and only the third party receives an output Z as a (possibly randomized) function of the inputs. This class of functions is similar to that studied in [FKN94], but our protocol model is more general (since it allows fully interactive communication), making it harder to establish lower bounds.

<sup>&</sup>lt;sup>1</sup>Of course, communication complexity lower bounds continue to hold for secure computation as well, but these bounds as such are (apparently) very loose (since there is a trivial upper bound for communication complexity, which is at most the size of all inputs and outputs).

<sup>&</sup>lt;sup>2</sup>In communication complexity and related problems, the lower bound techniques relate to Wyner common information [PP, BP13], whereas the tools in this work are more directly related to Gács-Körner common information. Wyner common information and Gács-Körner common information have been generalized to a measure of correlation represented as the "tension region" in [PP12].

#### 1.1 Results and Techniques

We study the setting shown in Figure 1. We obtain lower bounds on the expected number of bits that need to be exchanged between each pair of parties when securely evaluating a (possibly randomized) function of two inputs so that Alice and Bob have one input each, and Charlie receives the output. In fact, our bounds are on the entropy of the transcript between each pair,<sup>3</sup> and hence hold even when the protocol is amortized over several instances with independent inputs. Further, these bounds do not depend on the input distribution (as long as the distribution has full support) and hold even if the protocol is allowed to depend on the input distribution.

At a high-level, the ingredients in deriving of our lower-bounds are the following:

• Firstly, we observe that, since Alice and Bob do not obtain any outputs, they are both forced to reveal their inputs fully (upto equivalent inputs) to the rest of the system. This implies that the transcripts of a secure computation form the shares of the inputs and outputs according to an appropriately defined "correlated multiple secret sharing scheme" (CMSS).<sup>4</sup> Hence, lower bounds on the entropies of the shares in a CMSS imply lower bounds on the entropies of the messages in a secure computation protocol. One can immediately obtain a naïve lower bound on the entropies of the shares in a correlated multiple secret sharing scheme: specifically, if X, Y, Z are the secrets, and  $M_{23}$  denotes the part of shares that is not available to a party who should learn only X, then we can see that  $H(M_{23}) \geq H(Y, Z|X)$ .<sup>5</sup>

We strengthen the naïve lower bounds by relying on a "data-processing inequality" for residual information — i.e., the gap between mutual-information and (Gács-Körner) common information — which lets us relate the residual information between the shares to the residual information between the secrets. This bound is given in Theorem 1.

- We can further improve the above lower bounds using a new tool, called distribution switching. The key idea is that the security requirement forces the distribution of the transcript on certain links to be independent of the inputs. Hence, we can optimize our bounds over all input distributions having full support. Further, this shows that even if the protocol is allowed to depend on the input distribution, our bounds (which depend only on the function being evaluated) hold for every input distribution that has full support over the input domain. The resulting bound is summarised in Theorem 2.
- As it turns out, CMSS lower bounds are in general weak, because a CMSS can in fact be strictly more efficient than a secure computation protocol that the CMSS problem is derived from (see Appendix C). To go beyond the CMSS bounds, we need to exploit the fact that in a protocol, the transcripts have to be generated by the parties interactively, rather than be created by an omniscient "dealer". An important technical contribution of this work is to provide a new tool towards this, in the form of a new information inequality for 3-party interactive protocols (Lemma 4). We use this to derive a bound (Theorem 3) that serves as an intermediate result for us.

<sup>&</sup>lt;sup>3</sup>The entropy bounds translate to bounds on the expected number of bits communicated, when we require that the messages on the individual links are encoded using (possibly adaptively chosen) prefix-free codes. See Appendix B.

 $<sup>^4</sup>$ We remark that our notion of multiple secret sharing schemes is different from that of [BSC<sup>+</sup>94], which (implicitly) required that secrets with different access structures be independent of each other. In our case, Z is typically strongly correlated with X, Y, often via a deterministic function.

<sup>&</sup>lt;sup>5</sup> We point out a simple example for which one can obtain a tight bound from this naïve bound for CMSS: addition (in any group) requires one group element to be communicated between every pair of players, even with amortization over several independent instances. Previous lower bounds for secure evaluation of addition (in any group) [FY92, CK93], while considering an arbitrary number of parties, either restricted themselves to bounding the *number of messages* required, or relied on non-standard security requirements. (For the 3-party case, for semi-honest security, results of [FY92, CK93] only imply that all three links should be used. [FY92] did give a lower bound on the number of bits communicated as well, but this was shown only under a non-standard security requirement called *unstoppability*.)

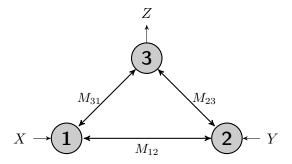


Figure 1 A three-party secure computation problem. Alice (party-1) has input X and Bob (party-2) has Y. We require that (i) Charlie (party-3) obtains as output a randomized function of the other two parties' inputs, distributed as  $p_{Z|XY}$ , (ii) Alice and Bob learn no additional information about each other's inputs, and (iii) Charlie learns nothing more about X, Y than what is revealed by Z. All parties can talk to each other, over multiple rounds over bidirectional pairwise private links.

Using the idea of distribution switching, we can significantly improve the above lower bounds by optimizing them using appropriate distributions of inputs. In fact, we can take the different terms in our bounds and optimize each of them separately using different distributions over the inputs. The resulting bounds (Theorem 4 and Theorem 5) are often stronger than what can be obtained by considering a single input distribution for the entire expression.

The resulting bounds are summarized in Theorem 1, Theorem 2, Theorem 4 and Theorem 5. While we restrict our attention to a 3 party setting, to the best of our knowledge, these are the first *generic* lower bounds which apply to any function. To illustrate their use, we apply them to several interesting example functions. In particular, we show the following:

- We analyze secure protocols for two functions GROUP-ADD, CONTROLLED-ERASURE and REMOTE-OT and, applying our lower bounds, show that these protocols achieve *optimal communication complexity simultaneously on each link*. We call such a protocol a *communication-ideal* protocol. We leave it open to characterize which functions have communication-ideal protocols.
- We use our lower bounds to establish a separation between secret sharing and secure computation: we show that there exists a function (in fact, the AND function) which has a CMSS scheme with a share strictly smaller than the number of bits in the transcript on the corresponding link in any secure computation protocol for that function. While such a separation is natural to expect, we note that proving it requires exploiting the properties of an interactive protocol.
- We show an *explicit* deterministic function  $f: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{n-1}$  which has a communication-ideal protocol in which Charlie's total communication cost is (and must be at least) 3n-1 bits. In contrast, [FKN94] showed that there exist functions  $f: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ , for which Charlie must receive at least 3n-4 bits, if the protocol is required to be in their non-interactive model. (Note that our bound is incomparable to that of [FKN94], since we require the output of our function to be longer; on the other hand, our bound uses an explicit function, and continues to hold even if we allow unrestricted interaction.)
- Our lower bounds for communication complexity also yield lower bounds on the amount of randomness needed in secure computation protocols. We analyze secure protocols for GROUP-ADD, CONTROLLED-ERASURE, REMOTE-OT and SUM, and prove that these protocols are *randomness-optimal*, i.e., they use the least amount of randomness.

#### 1.2 Related Work

Communication complexity of multi-party computation without security requirements has been widely studied since [Yao79] (see [KN97]), and more recently has seen the use of information-theoretic tools as well, in [CSWY01] and subsequent works. Independently, in the information theory literature communication requirements of interactive function computation have been studied (e.g. [OR01]).

In secure multi-party computation, there has been a vast literature on information-theoretic security, focusing on building efficient protocols, as well as characterizing various aspects like corruption models that admit secure protocols (e.g. [BGW88, CCD88, Cha89, HM97, FHM99, HLMR12]) and the number of rounds of interaction needed (e.g. [FL82, GIKR01, FGG+06, PCRR09, KKK09]), In the computational security setting, [NN01] gave upper bounds on the communication complexity of 2-party secure computation in terms of the communication complexity without security requirements. In the information-theoretic security setting, [DI06] upper bounded the communication complexity of multi-party secure computation in terms of the circuit complexity of the computation.

But lower-bounding communication complexity has received much less attention. For 2-party secure computation with security against passive corruption of one party (when the function admits such a protocol), communication complexity was combinatorially characterized in [Kus89]. Franklin and Yung [FY92] showed that the number of messages used in a protocol must be quadratic in n, the number of parties (if security against corrupting  $t = \Omega(n)$  parties is required). Further [FY92, CK93] gave tight lower and upper bounds on the number of messages needed for secure computation of the "modulosum" function by n parties; relying on a stronger corruption model (fail-stop corruption), [FY92] also argued a lower bound for the amortized communication complexity of secure summation. [FKN94] obtained a lower bound on the communication complexity for a restricted class of 3-party protocols; along with positive results, they gave a modest lower bound for communication needed for evaluating random functions in this model. The difficulty of obtaining general lower bounds was pointed out in [IK04], who related such lower bounds to lower bounds for locally decodable codes and private information retrieval protocols. The connection to private information retrieval protocols was recently used in [BIKK14] to, among other things, derive the best known general upper bound on communication for Boolean functions in the model of [FKN94]. Note that this upper bound is exponential in the number of input bits compared to the lower bound of [FKN94] which is only linear. The question of how much randomness is required for secure computation seems to have received even less attention; we are aware of [KM97, BSPV99, GR05, LA14].

Information-theoretic tools have been successfully used in deriving bounds in various cryptographic problems like key agreement (e.g. [MW03]), secure 2-party computation (e.g. [DM99]) and secret-sharing and its variants (e.g. [BO11] and [BSC<sup>+</sup>94]). In this work, we rely on information-theoretic tools developed in [WW08, PP12], which also considered cryptographic problems. Some preliminary observations leading to this work appeared in [DP13] (as referenced at the appropriate points, below).

## 2 Preliminaries

**Notation.** We write  $p_X$  to denote the distribution of a discrete random variable X;  $p_X(x)$  denotes  $\Pr[X=x]$ . When clear from the context, the subscript of  $p_X$  will be omitted. The conditional distribution denoted by  $p_{Z|U}$  specifies  $\Pr[Z=z|U=u]$ , for each value z that Z can take and each value u that U can take. A randomized function of two variables, is specified by a probability distribution  $p_{Z|XY}$ , where X, Y denote the two input variables, and Z denotes the output variable. For a sequence of random variables  $X_1, X_2, \ldots$ , we denote by  $X^n$  the vector  $(X_1, \ldots, X_n)$ .

For random variables T, U, V, we write the Markov chain T - U - V to indicate that T and V are conditionally independent conditioned on U: i.e., I(T; V|U) = 0. All logarithms are to the base 2. The binary entropy function is denoted by  $H_2(p) = -p \log p - (1-p) \log(1-p)$ ,  $p \in (0,1)$ .

**Protocols.** A 3-party protocol  $\Pi$  is specified by a collection of "next message functions"  $(\Pi_1, \Pi_2, \Pi_3)$  which probabilistically map a *state* of the protocol to the next state (in a restricted manner), and output functions  $(\Pi_1^{\text{out}}, \Pi_2^{\text{out}}, \Pi_3^{\text{out}})$  used to define the outputs of the parties as probabilistic functions of their views. We shall also allow the protocol to depend on the distribution of the inputs to the parties. (This would allow one to tune a protocol to be efficient for a suitable input distribution. Allowing this makes our lower bounds stronger; on the other hand, none of the protocols we give for our examples require this flexibility. See discussion in Appendix H.)

In Appendix A we formalize a well-formed 3-party protocol. Without loss of generality, the state of the protocol consists only of the inputs received by each party and the transcript of the messages exchanged so far.<sup>6</sup> We denote the final transcripts on the three links, after executing protocol  $\Pi$  on its specified input distribution by  $M_{12}^{\Pi}$ ,  $M_{23}^{\Pi}$  and  $M_{31}^{\Pi}$ . When  $\Pi$  is clear from the context, we simply write  $M_{12}$  etc. We define  $M_1 = (M_{12}, M_{31})$  as the transcripts that party 1 can see;  $M_2$  and  $M_3$  are defined similarly. We define the view of the  $i^{th}$  party,  $V_i$  to consist of  $M_i$  and that party's inputs and outputs (if any).

It is easy to see that a (well-formed) protocol, along with an input distribution, fully defines a joint distribution over all the inputs, outputs and the joint transcripts on all the links.

Secure Computation. We consider three party computation functionalities, in which Alice and Bob (parties 1 and 2) receive as inputs the random variables  $X \in \mathcal{X}$  and  $Y \in \mathcal{Y}$ , respectively, and Charlie (party 3) produces an output  $Z \in \mathcal{Z}$  distributed according to a specified distribution  $p_{Z|XY}$ . In particular, we can consider a deterministic function evaluation functionality where Z = f(X,Y) with probability 1, for some function  $f: \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ . The set  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  are always finite. In secure computation, we shall consider the inputs to the computation to come from a distribution  $p_{XY}$  over  $\mathcal{X} \times \mathcal{V}$ .

A (perfectly) secure computation protocol  $\Pi(p_{XY}, p_{Z|XY}) = (\Pi_1, \Pi_2, \Pi_3, \Pi_3^{\text{out}})$  for  $(p_{XY}, p_{Z|XY})$  is a protocol which satisfies the following conditions:

- Correctness: Output of Charlie, is distributed according to  $p_{Z|X=x,Y=y}$ , where x,y are the inputs to Alice and Bob
- Privacy: The privacy condition corresponds to "1-privacy", wherein at most one party is passively corrupt. Corresponding to security against Alice, Bob and Charlie, respectively, we have the following three Markov chains.  $M_1 X (Y, Z)$ ,  $M_2 Y (X, Z)$  and  $M_3 Z (X, Y)$ . Equivalently, in terms of the views,  $I(V_1; (Y, Z)|X) = I(V_2; (X, Z)|Y) = I(V_3; (X, Y)|Z) = 0$ .

Intuitively, the privacy condition guarantees that even if one party (say Alice) is curious, and retains its view from the protocol (i.e.,  $M_1$ ), this view reveals nothing more to it about the inputs and outputs of the other parties (namely, Y, Z), than what its own inputs and outputs reveal (as long as the other parties erase their own views). In other words, a curious party may as well simulate a view for itself based on just its inputs and outputs, rather than retain the actual view it obtained from the protocol execution.

For simplicity, we prove all our results for *perfect security* as defined above; this is also the setting for classical positive results like that of [BGW88]. But all our bounds do extend to the setting of statistical security, as we shall show in the full version of this paper (following [WW10, PP12] who extend similar results to the statistical security case).<sup>7</sup> Also, the above security requirements are for an

<sup>&</sup>lt;sup>6</sup>Since the parties are computationally unbounded, there is no need to allow private randomness as part of the state; randomness for a party can always be resampled at every round conditioned on the inputs, outputs and messages in that party's view.

<sup>&</sup>lt;sup>7</sup>We remark that we do not know if our bounds extend to a relaxed security setting sometimes considered in the information theory literature: there the error in security is only required to go to 0 as the size of the input grows to infinity. Instead, we use the standard cryptographic security requirement that for any fixed input length, the error can be driven arbitrarily close to zero by choosing a large enough security parameter.

honest execution of the protocol (corresponding to honest-but-curious or passive corruption of at most one party). The lower bounds derived in this model typically continue to hold for active corruption as well (since for many functionalities, every protocol secure against active corruption is a protocol secure against passive corruption), but in fact, in our setting (where 1 out of 3 parties is corrupted), the functions we consider simply do not have secure protocols against active corruption.

Communication Complexity and Entropy. A standard approach to lowerbounding the number of bits in a string is to lowerbound its entropy. However, in an interactive setting, a party sees the messages in each round, rather than just a concatenation of all the bits sent over the entire protocol. In a setting where we allow variable length messages, this would seem to allow communicating more bits of information than the length of the transcript itself. But this allows the parties to learn when the message transmitted in a round ends, implicitly inserting an end-of-message marker into the bit stream. To account for this, one can require that the message sent at every round is a codeword in a prefix-free code. (The code itself can be dynamically determined based on previous messages exchanged over the link.) As shown in Appendix B, with this requirement, the number of bits communicated in each link is indeed lowerbounded by the entropy of the transcript in that link.

**Normal Forms.** In Appendix A, we describe a normal form for a randomized function  $p_{Z|XY}$  as well as for the pair  $(p_{XY}, p_{Z|XY})$ , where  $p_{XY}$  and  $p_{Z|XY}$  are the input distribution and the function respectively. Essentially, these normal forms merge "equivalent" inputs and outputs. As argued there, it suffices to study the communication complexity of secure computation for functions  $p_{Z|XY}$  and for pairs  $(p_{XY}, p_{Z|XY})$  in normal form.

Communication-Ideal Protocol. We say that a protocol  $\Pi(p_{XY}, p_{Z|XY})$  for securely computing a randomized function  $p_{Z|XY}$ , for a distribution  $p_{XY}$  is communication-ideal if for each  $ij \in \{12, 23, 31\}$ ,

$$H(M_{ij}^{\Pi}) = \inf_{\Pi'(p_{XY}, p_{Z|XY})} H(M_{ij}^{\Pi'}),$$

where the infimum is over all secure protocols for  $p_{Z|XY}$  with the same distribution  $p_{XY}$ . That is, a communication-ideal protocol achieves the least entropy possible for every link, simultaneously. We remark that it is not clear, a priori, how to determine if a given function  $p_{Z|XY}$  has a communication-ideal protocol for a given distribution  $p_{XY}$ .

#### Common Information and Residual Information

Gács and Körner [GK73] introduced the notion of common information to measure a certain aspect of correlation between two random variables. The Gács-Körner common information of a pair of correlated random variables (U, V) can be defined as  $H(U \sqcap V)$ , where  $U \sqcap V$  is a random variable with maximum entropy among all random variables Q that are determined both by U and by V (i.e., there are functions f and g such that Q = f(U) = g(V)). In [PP12], the gap between mutual information and common information was termed residual information:  $RI(U; V) := I(U; V) - H(U \sqcap V)$ .

In [WW08], Wolf and Wullschleger identified (among other things) the following important data processing inequality for residual information.

**Lemma 1** ([WW08]). If T, U, V, W are jointly distributed random variables such that the following two Markov chains hold: (i) U - T - W, and (ii) T - W - V, then

$$RI(T; W) \le RI((U, T); (V, W)).$$

The Markov chain conditions above correspond to the requirement that it is secure (against honestbut-curious adversaries) to require a pair of parties holding the views (U, T) and (V, W), to produce outputs T, W, respectively, because for the first party, the rest of its view, U, can be simulated based on the output T, independent of the output W (and similarly, for the second party). The lemma states that under such a secure transformation from views to outputs, the residual information can only decrease.

In [PP12], the following alternate definition of residual information was given, which will be useful in lowerbounding conditional mutual information terms.

$$RI(U;V) = \min_{\substack{Q: \exists f, g \text{ s.t.} \\ Q = f(U) = g(V)}} I(U;V|Q). \tag{1}$$

The random variable Q which achieves the minimum is, in fact,  $U \sqcap V$ . Note that the residual information is always non-negative.

# 3 Lower Bounds on Communication Complexity

This section is divided into three parts. In Subsection 3.1, we derive preliminary lower bounds for secure computation. In each of the subsequent subsections, we give different improvements of the lower bounds derived in Subsection 3.1.

#### 3.1 Preliminary Lower Bounds

We first state the following basic lemma for any protocol for secure computation. Similar results have appeared in the literature earlier (for instance, special cases of Lemma 2 appear in [DM00, WW10, DP13]).

Lemma 2 states the simple fact that, for  $(p_{XY}, p_{Z|XY})$  in normal form, the information about a party's input must flow out through the links she/he is part of, and the information about Charlie's output must flow in through the links he is part of. This crucially relies on the fact that Alice and Bob obtain no output, and Charlie has no input in our model.

**Lemma 2.** Suppose  $(p_{XY}, p_{Z|XY})$  is in normal form. Then, in any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , the cut isolating Alice from Bob and Charlie must reveal Alice's input X, i.e.,  $H(X|M_{12}, M_{31}) = 0$ . Similarly,  $H(Y|M_{12}, M_{23}) = 0$  and  $H(Z|M_{23}, M_{31}) = 0$ .

A proof is given in Appendix E. We obtain a preliminary lower bound in Theorem 1 below by using the above lemma and the data-processing inequality for residual information in Lemma 1. Note that the assumption of  $(p_{XY}, p_{Z|XY})$  being in normal form below is without loss of generality (Appendix A).

**Theorem 1.** Any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where  $(p_{XY}, p_{Z|XY})$  is in normal form, should satisfy the following lower bounds on the entropy of the transcripts on each link.

$$H(M_{23}) \ge \max\{RI(X;Z), RI(X;Y)\} + H(Y,Z|X),$$
 (2)

$$H(M_{31}) \ge \max\{RI(Y;Z), RI(X;Y)\} + H(X,Z|Y),$$
 (3)

$$H(M_{12}) \ge \max\{RI(X;Z), RI(Y;Z)\} + H(X,Y|Z).$$
 (4)

*Proof.* We shall prove (2). The other two inequalities follow symmetrically.

$$H(M_{23}) \ge \max\{H(M_{23}|M_{31}), H(M_{23}|M_{12})\}$$

$$= \max\{I(M_{23}; M_{12}|M_{31}), I(M_{23}; M_{31}|M_{12})\} + H(M_{23}|M_{12}, M_{31}).$$
(5)

Firstly, we can bound the last term of (5) as follows (to already get a naïve bound):

$$H(M_{23}|M_{12}, M_{31}) \stackrel{\text{(a)}}{=} H(M_{23}, Y, Z|M_{12}, M_{31}, X)$$
  
  $\geq H(Y, Z|M_{12}, M_{31}, X) \stackrel{\text{(b)}}{=} H(Y, Z|X),$ 

where (a) follows from Lemma 2 and (b) follows from the privacy against Alice. Next, we lower bound the first term inside max of (5) by RI(X; Z) as follows. Firstly,

$$I(M_{23}; M_{12}|M_{31}) = I(M_{23}, M_{31}; M_{12}, M_{31}|M_{31}) \ge RI(M_{23}, M_{31}; M_{12}, M_{31}), \tag{6}$$

where the inequality follows from (1), the alternate definition of residual information, by taking  $Q = M_{31}$ . Now, by privacy against Charlie, we have  $(M_{23}, M_{31}) - Z - X$  and privacy against Alice, we have  $(M_{12}, M_{31}) - X - Z$ . Applying Lemma 1 with the above markov chains, together with Lemma 2, we get

$$RI(M_{23}, M_{31}; M_{12}, M_{31}) \ge RI(Z; X) = RI(X; Z).$$

Similarly, we can lower bound the second term inside max of (5) by RI(X;Y), completing the proof.  $\square$ 

A consequence of Lemma 2 is that the transcripts in a secure computation protocol form shares in a CMSS scheme for the same distribution  $p_{XYZ} = p_{XY}p_{Z|XY}$ ; see Appendix C. There we derive bounds on the sizes of these shares which, in fact, imply Theorem 1 (and Theorem 2). In the rest of the paper we will restrict our attention to  $p_{XY}$  which have full support (and, without loss of generality,  $p_{Z|XY}$  expressed in normal form). This will allow us to strengthen the preliminary bounds in Theorem 1. Notice that such  $(p_{XY}, p_{Z|XY})$  are in normal form and hence Lemma 2 holds.

#### 3.2 Distribution Switching and Improved Lower Bounds - I

To improve the bounds in Theorem 1, we give a technique, distribution switching, which significantly improves the above bounds and leads to one of our main theorems.

The following lemma states that privacy requirements imply that the transcript  $M_{12}$  generated by a secure protocol computing  $p_{Z|XY}$  is independent of both the inputs. Moreover, if the function  $p_{Z|XY}$  satisfies some additional constraints, then the other two transcripts also become independent of the inputs. For a distribution  $p_{XY}$ , a bipartite graph on vertex set  $\mathcal{X} \cup \mathcal{Y}$  is said to be the *characteristic bipartite graph of*  $p_{XY}$ , if  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$  are connected whenever  $p_{XY}(x,y) > 0$ . The proof of the following lemma is along the lines of a similar lemma in [DP13] and we prove it in Appendix E for completeness.

**Lemma 3.** Consider a function  $p_{Z|XY}$  not necessarily in normal form.

- 1. Suppose that  $p_{XY}$  is such that the characteristic bipartite graph of  $p_{XY}$  is connected. Then, for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , we have  $I(X, Y, Z; M_{12}) = 0$ .
- 2. Suppose  $p_{XY}$  has full support and  $p_{Z|XY}$  satisfies the following condition:

**Condition 1.** There is no non-trivial partition  $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$  (i.e.,  $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$  and neither  $\mathcal{X}_1$  nor  $\mathcal{X}_2$  is empty), such that if  $\mathcal{Z}_k = \{z \in \mathcal{Z} : x \in \mathcal{X}_k, y \in \mathcal{Y}, p(z|x,y) > 0\}, k = 1, 2$ , their intersection  $\mathcal{Z}_1 \cap \mathcal{Z}_2$  is empty.

Then, for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , we have  $I(X, Y, Z; M_{31}) = 0$ .

3. Suppose  $p_{XY}$  has full support and  $p_{Z|XY}$  satisfies the following condition:

Condition 2. There is no non-trivial partition  $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$  such that if  $\mathcal{Z}_k = \{z \in \mathcal{Z} : x \in \mathcal{X}, y \in \mathcal{Y}_k, p(z|x,y) > 0\}, k = 1, 2$ , their intersection  $\mathcal{Z}_1 \cap \mathcal{Z}_2$  is empty.

Then, for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , we have  $I(X, Y, Z; M_{23}) = 0$ .

We point out that  $p_{XY}$  having a connected characteristic bipartite graph is a weaker condition than  $p_{XY}$  having full support.

#### Distribution Switching

We will now strengthen the lower bounds from Theorem 1. Specifically, we will argue that even if the protocol is allowed to depend on the input distribution (as we do here), privacy requirements will require that the lower bounds derived for when the distributions of the inputs are changed continue to hold for the original setting.

We note that any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where distribution  $p_{XY}$  has full support, continues to be a secure protocol even if we switch the input distribution to a different one  $p_{X'Y'}$ . This follows directly from examining the correctness and privacy conditions required for a protocol to be secure.

**Theorem 2.** For any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where  $p_{Z|XY}$  is in normal form and  $p_{XY}$  has full support, we have the following strengthening of (4):

$$H(M_{12}) \ge \max \{ \sup_{p_{X'Y'}} \left( RI(X'; Z') + H(X', Y'|Z') \right), \sup_{p_{X'Y'}} \left( RI(Y'; Z') + H(X', Y'|Z') \right) \}, \tag{7}$$

where the supremizations are over  $p_{X'Y'}$  having full support and the objective functions are evaluated using  $p_{X'Y'Z'}(x,y,z) = p_{X'Y'}(x,y)p_{Z|XY}(z|x,y)$ .

If  $p_{Z|XY}$  (in normal form) satisfies Condition 1 of Lemma 3, then for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where  $p_{XY}$  has full support, we have the following strengthening of (3):

$$H(M_{31}) \ge \max \{ \sup_{p_{X'Y'}} \left( RI(Y'; Z') + H(X', Z'|Y') \right), \sup_{p_{X'Y'}} \left( RI(X'; Y') + H(X', Z'|Y') \right) \}, \tag{8}$$

where the supremizations are over  $p_{X'Y'}$  having full support and the objective functions are evaluated using  $p_{X'Y'Z'}(x,y,z) = p_{X'Y'}(x,y)p_{Z|XY}(z|x,y)$ .

If  $p_{Z|XY}$  (in normal form) satisfies Condition 2 of Lemma 3, then for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where  $p_{XY}$  has full support, we have the following strengthening of (2):

$$H(M_{23}) \ge \max\{\sup_{p_{X'Y'}} \left( RI(X'; Z') + H(Y', Z'|X') \right), \sup_{p_{X'Y'}} \left( RI(X'; Y') + H(Y', Z'|X') \right) \}, \tag{9}$$

where the supremizations are over  $p_{X'Y'}$  having full support and the objective functions are evaluated using  $p_{X'Y'Z'}(x,y,z) = p_{X'Y'}(x,y)p_{Z|XY}(z|x,y)$ .

*Proof.* By Lemma 3, it follows that the transcript  $M_{12}$  of the protocol (under both the original and the switched input distributions) must remain independent of the input data X, Y. This allows us to argue using Theorem 1, that

$$H(M_{12}) \ge \max \{ \sup_{p_{X'Y'}} \left( RI(X'; Z') + H(X', Y'|Z') \right), \sup_{p_{X'Y'}} \left( RI(Y'; Z') + H(X', Y'|Z') \right) \},$$

where the supremizations are over  $p_{X'Y'}$  having full support and the objective functions are evaluated using  $p_{X'Y'Z'}(x, y, z) = p_{X'Y'}(x, y)p_{Z|XY}(z|x, y)$ .

Similarly, if the function  $p_{Z|XY}$  satisfies the condition 1 and 2 of Lemma 3, we can show the other two bounds on  $H(M_{31})$  and  $H(M_{23})$  as well.

In Appendix C, we derive similar bounds for the size of the shares of a CMSS scheme.

#### 3.3 An Information Inequality for Protocols and Improved Lower Bounds - II

We can give a different improvement to Theorem 1 by exploiting the fact that in a protocol, transcripts are generated by the parties interactively, rather than by an omniscient dealer. Towards this, we derive an information inequality relating the transcripts on different links in general 3-party protocols, in which parties do not share any common or correlated randomness or correlated inputs at the beginning of the protocol. Note that our model for protocols does indeed satisfy these conditions when the inputs are independent of each other.

**Lemma 4.** In any well-formed 3-party protocol, if the inputs to the parties are independent of each other, then, for  $\{\alpha, \beta, \gamma\} = \{1, 2, 3\}$ ,

$$I(M_{\gamma\alpha}; M_{\beta\gamma}) \ge I(M_{\gamma\alpha}; M_{\beta\gamma}|M_{\alpha\beta}).$$

We prove the lemma in Appendix E. We further note that, as in (6),  $I(M_{\gamma\alpha}; M_{\beta\gamma}|M_{\alpha\beta}) \ge RI(M_{\gamma\alpha}M_{\alpha\beta}; M_{\beta\gamma}M_{\alpha\beta})$ . Hence, if the inputs are independent of each other,

$$I(M_{\gamma\alpha}; M_{\beta\gamma}) \ge I(M_{\gamma\alpha}; M_{\beta\gamma} | M_{\alpha\beta}) \ge RI(M_{\gamma\alpha} M_{\alpha\beta}; M_{\beta\gamma} M_{\alpha\beta}). \tag{10}$$

This inequality provides us with a means to exploit the protocol structure behind transcripts. Below, Theorem 3 (specifically, (13)) shows that the term  $\max\{RI(X;Z),RI(Y;Z)\}$  in (4) can be replaced by RI(X;Z) + RI(Y;Z). Note that Theorem 3 is stated and proven for independent inputs, that is,  $p_{XY} = p_X p_Y$ . In Appendix G we show that (using ideas of distribution switching from Subsection 3.2) it implies lower bounds for dependent inputs  $p_{XY}$  with full support as well. However, this extension is not required in the sequel where we derive our main theorems.

**Theorem 3.** Any secure protocol  $\Pi(p_X p_Y, p_{Z|XY})$ , where  $p_{Z|XY}$  is in normal form and  $p_X$ ,  $p_Y$  have full support, should satisfy the following lower bounds on the entropy of the transcripts on each link.

$$H(M_{23}) \ge RI(X; Z) + H(Y, Z|X)$$
 (11)

$$H(M_{31}) \ge RI(Y; Z) + H(X, Z|Y)$$
 (12)

$$H(M_{12}) > RI(X;Z) + RI(Y;Z) + H(X,Y|Z)$$
 (13)

*Proof.* Firstly, note that (11) and (12) follow from (2) and (3). To prove (13), we have

$$H(M_{12}) = I(M_{12}; M_{23}) + H(M_{12}|M_{23})$$
  
=  $I(M_{12}; M_{23}) + I(M_{12}; M_{31}|M_{23}) + H(M_{12}|M_{23}, M_{31})$   
 $\geq RI(X; Z) + RI(Y; Z) + H(X, Y|Z),$ 

where the last inequality used  $H(M_{12}|M_{23},M_{31}) \geq H(X,Y|Z)$  and  $I(M_{12};M_{31}|M_{23}) \geq RI(Y;Z)$  (both as in the proof of Theorem 1) as well as  $I(M_{12};M_{23}) \geq RI(X;Z)$  (by (10), which applies since we assume independent inputs).

In Appendix C we show that the above proof can be extended to derive lower bounds for secure sampling. We can improve the bounds in Theorem 3 using distribution switching which leads to our main theorems. The following lemma states that if the inputs X and Y are independent, then privacy requirements imply that certain transcripts generated by a secure protocol computing  $p_{Z|XY}$ , are independent of certain data. More precisely, we show the following.

**Lemma 5.** For any secure protocol  $\Pi(p_X p_Y, p_{Z|XY})$ , where  $p_{Z|XY}$  may not be in normal form, should satisfy  $I(X; M_{23}) = I(Y; M_{31}) = 0$ .

Proof.  $I(X; M_{23}) = 0$  follows from  $I(X; M_{23}) \leq I(X; M_{23}, Y) = I(X; Y) + I(X; M_{23}|Y) = 0$ , where the last equality follows from independence of X and Y and privacy against Bob. Similarly we can show  $I(Y; M_{31}) = 0$ .

If the inputs X and Y are independent, i.e.,  $p_{XY} = p_X p_Y$ , then the transcripts  $M_{23}$  and  $M_{31}$  of the protocol (under both the original and the switched input distributions) must remain independent of X and Y respectively (by Lemma 5). Note that for the independence of  $M_{12}$  and (X,Y), we do not need the independence of inputs. Rather, since we used Lemma 4 (which requires independence of inputs) to get (13), we are forced to consider independent inputs if we use the bound on  $H(M_{12})$  in (13). This allows us to argue using Theorem 3, that

$$H(M_{12}) \ge \sup_{p_{X'}p_{Y'}} RI(X'; Z') + RI(Y'; Z') + H(X', Y'|Z'),$$

where the supremum is over  $p_{X'}, p_{Y'}$  which have full support and the terms in the right hand side are evaluated under the joint distribution

$$p_{X',Y',Z'}(x,y,z) = p_{X'}(x)p_{Y'}(y)p_{Z|X,Y}(z|x,y).$$

Similarly, from Lemma 5, we know that  $M_{31}$  is independent of Y, and  $M_{23}$  is independent of X. Hence,

$$H(M_{31}) \ge \sup_{p_{Y'}} RI(Y'; Z') + H(X, Z'|Y'),$$
 (14)

$$H(M_{23}) \ge \sup_{p_{X'}} RI(X'; Z') + H(Y, Z'|X'),$$
 (15)

where the right hand side of (14) is evaluated under  $p_{X,Y',Z'}(x,y,z) = p_X(x)p_{Y'}(y)p_{Z|X,Y}(z|x,y)$ . Similarly, for the bound on  $H(M_{23})$ .

In fact, we can show an even stronger bound than above by a more careful application of distribution switching. This leads us to second of our three main lower bound theorems, which is proved in Appendix D.

Remark: The above discussion of distribution switching used Lemma 5, which holds for independent inputs, that is,  $p_{XY} = p_X p_Y$ . However, as we argue in Appendix G, the resulting lower bounds also hold for dependent inputs  $p_{XY}$  with full support; we state the theorem for this general case.

**Theorem 4.** The following communication complexity bounds hold for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where  $p_{Z|X,Y}$  is in normal form and  $p_{XY}$  has full support:

$$H(M_{23}) \ge \left(\sup_{p_{X'}} RI(X'; Z')\right) + \left(\sup_{p_{X''}} H(Y, Z''|X'')\right),$$
 (16)

$$H(M_{31}) \ge \left(\sup_{p_{Y'}} RI(Y'; Z')\right) + \left(\sup_{p_{Y''}} H(X, Z''|Y'')\right),$$
 (17)

$$H(M_{12}) \ge \max \left\{ \begin{array}{l} \sup_{p_{X'}} \left( \sup_{p_{Y'}} RI(Y'; Z') \right) + \left( \sup_{p_{Y''}} RI(X'; Z'') + H(X', Y''|Z'') \right), \\ \sup_{p_{Y'}} \left( \sup_{p_{X'}} RI(X'; Z') \right) + \left( \sup_{p_{X''}} RI(Y'; Z'') + H(X'', Y'|Z'') \right) \end{array} \right\}, \quad (18)$$

where the supremizations are over distributions  $p_{X'}, p_{X''}, p_{Y''}, p_{Y''}$  having full support. The terms in the right hand side of (16) are evaluated using the distribution  $p_Y$  of the data Y of Bob, i.e.,

$$p_{X',Y,Z'}(x,y,z) = p_{X'}(x)p_Y(y)p_{Z|X,Y}(z|x,y),$$
  
$$p_{X'',Y,Z''}(x,y,z) = p_{X''}(x)p_Y(y)p_{Z|X,Y}(z|x,y).$$

Similarly, the terms in (17) are evaluated using the distribution  $p_X$  of the data X of Alice. The lower bound in (18) does not depend on the distributions  $p_X$  and  $p_Y$  of the data. The terms on the top row of (18), for instance, are evaluated using

$$p_{X',Y',Z'}(x,y,z) = p_{X'}(x)p_{Y'}(y)p_{Z|X,Y}(z|x,y),$$
  
$$p_{X',Y'',Z''}(x,y,z) = p_{X'}(x)p_{Y''}(y)p_{Z|X,Y}(z|x,y).$$

When the function satisfies certain additional constraints, we can strengthen the lower bounds on the  $H(M_{23})$  and  $H(M_{31})$  as shown in the following theorem which is proved in Appendix D.

**Theorem 5.** Consider any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where  $p_{XY}$  has full support and  $p_{Z|XY}$  is in normal form.

1. Suppose the function  $p_{Z|XY}$  satisfies Condition 1 of Lemma 3, that is, there is no non-trivial partition  $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$  (i.e.,  $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$  and neither  $\mathcal{X}_1$  nor  $\mathcal{X}_2$  is empty), such that if  $\mathcal{Z}_k = \{z \in \mathcal{Z} : x \in \mathcal{X}_k, y \in \mathcal{Y}, p(z|x,y) > 0\}, k = 1, 2$ , their intersection  $\mathcal{Z}_1 \cap \mathcal{Z}_2$  is empty. Then, we have the following strengthening of (17).

$$H(M_{31}) \ge \sup_{p_{X'}} \left( \left( \sup_{p_{Y'}} RI(Y'; Z') \right) + \left( \sup_{p_{Y''}} H(X', Z''|Y'') \right) \right),$$
 (19)

where the suprimizations are over distributions  $p_{X'}, p_{Y'}, p_{Y''}$  having full support and the terms in the right hand side are evaluated using the distribution

$$p_{X'Y'Z'Y''Z''}(x',y',z',y'',z'') = p_{X'}(x')p_{Y'}(y')p_{Z|XY}(z'|x',y')p_{Y''}(y'')p_{Z|XY}(z''|x',y'').$$

2. Suppose the function  $p_{Z|XY}$  satisfies Condition 2 of Lemma 3, that is, there is no non-trivial partition  $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$  such that if  $\mathcal{Z}_k = \{z \in \mathcal{Z} : x \in \mathcal{X}, y \in \mathcal{Y}_k, p(z|x,y) > 0\}, k = 1, 2$ , their intersection  $\mathcal{Z}_1 \cap \mathcal{Z}_2$  is empty. Then, we have the following strengthening of (16).

$$H(M_{23}) \ge \sup_{p_{X'}} \left( \left( \sup_{p_{X'}} RI(X'; Z') \right) + \left( \sup_{p_{X''}} H(Y', Z''|X'') \right) \right),$$
 (20)

where the supremizations are over distributions  $p_{X'}, p_{X''}, p_{Y'}$  having full support and the terms in the right hand side are evaluated using the distribution

$$p_{X'Y'Z'X''Z''}(x',y',z',x'',z'') = p_{X'}(x')p_{Y'}(y')p_{Z|XY}(z'|x',y')p_{X''}(x'')p_{Z|XY}(z''|x'',y').$$

Note that in Theorem 2, Theorem 4 and Theorem 5, any choice of  $p_{X'Y'}$ ,  $p_{X'}$ ,  $p_{X''}$ ,  $p_{Y''}$ ,  $p_{Y''}$ ,  $p_{Y''}$  (with full support) will yield a lower bound. For a given function, while all choices do yield valid lower bounds, one is often able to obtain the *best* lower bound analytically (as in Theorem 7, where it is seen to be the best as it matches an upper bound) or numerically (as in Theorem 8).

To summarize, for any secure computation problem  $(p_{XY}, p_{Z|XY})$ , expressed in the normal form, Theorem 1 gives lower bounds on entropies of transcripts on all three links. If  $p_{XY}$  has full support and  $p_{Z|XY}$  is in normal form, then for  $H(M_{31})$ , our best lower bound is the larger of (3) and (17); for  $H(M_{23})$ , it is the larger of (2) and (16) and for  $H(M_{12})$ , it is the larger of (7) and (18). In addition, if  $p_{Z|XY}$  satisfies condition 1 of Lemma 3, then for  $H(M_{31})$ , our best lower bound is the larger of (8) and (19); if  $p_{Z|XY}$  satisfies condition 2 of Lemma 3, then for  $H(M_{23})$ , our best lower bound is the larger of (9) and (20).

## 4 Lower Bounds on Randomness

In this section, we provide lower bounds on the amount of randomness required in secure computation protocols. Although our focus in this paper is not to prove lower bounds on the amount of randomness, it turns out that we may apply the above lower bounds on communication to derive bounds on the amount of randomness required. We show in Section 5 that they give tight bounds on randomness required for the specific functions we analyse.

**Definition 1.** The randomness required to securely compute a function  $p_{Z|XY}$  for input distribution  $p_{XY}$  is defined as

$$\rho(p_{XY}, p_{Z|XY}) = \inf_{\Pi(p_{XY}, p_{Z|XY})} H(V_1, V_2, V_3 | X, Y),$$

where  $V_i$  is the view of party-i at the end of the protocol and infimum is over all secure protocols for  $(p_{XY}, p_{Z|XY})$ .

We can simplify the entropy term in the above definition as follows.

$$H(V_1, V_2, V_3 | X, Y) \ge H(M_{31}, M_{12}, M_{23} | X, Y)$$
  
  $\ge \max\{H(M_{31} | X, Y), H(M_{12} | X, Y), H(M_{23} | X, Y)\}.$ 

Since the above inequalities are true for any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , we have

$$\rho(p_{XY}, p_{Z|XY}) \ge \max\{H(M_{31}|X, Y), H(M_{12}|X, Y), H(M_{23}|X, Y)\}. \tag{21}$$

**Theorem 6.** Consider any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ .

- 1. If the characteristic bipartite graph of  $p_{XY}$  is connected, then  $\rho(p_{XY}, p_{Z|XY}) \geq H(M_{12})$ .
- 2. If  $p_{XY}$  has full support and  $p_{Z|XY}$  satisfies Condition 1 of Lemma 3, then  $\rho(p_{XY}, p_{Z|XY}) \geq H(M_{31})$ .
- 3. If  $p_{XY}$  has full support and  $p_{Z|XY}$  satisfies Condition 2 of Lemma 3, then  $\rho(p_{XY}, p_{Z|XY}) \geq H(M_{23})$ .

*Proof.* By Lemma 3, we have the following:

- 1. If the characteristic bipartite graph of  $p_{XY}$  is connected, then  $H(M_{12}|X,Y)=H(M_{12})$ . This, together with (21) implies  $\rho(p_{XY},p_{Z|XY}) \geq H(M_{12})$ .
- 2. If  $p_{XY}$  has full support and  $p_{Z|XY}$  satisfies Condition 1, then  $H(M_{31}|X,Y) = H(M_{31})$ . This, together with (21) implies  $\rho(p_{XY}, p_{Z|XY}) \ge H(M_{31})$ .
- 3. If  $p_{XY}$  has full support and  $p_{Z|XY}$  satisfies Condition 2, then  $H(M_{23}|X,Y)=H(M_{23})$ . This, together with (21) implies  $\rho(p_{XY},p_{Z|XY}) \geq H(M_{23})$ .

Hence, we can apply the lower bounds developed in Section 3 to obtain lower bounds on randomness. If  $p_{XY}$  has full support and  $p_{Z|XY}$  is in normal form, then for  $H(M_{12})$ , our best lower bound is the larger of (7) and (18). In addition, if  $p_{Z|XY}$  satisfies condition 1 of Lemma 3, then for  $H(M_{31})$ , our best lower bound is the larger of (8) and (19); if  $p_{Z|XY}$  satisfies condition 2 of Lemma 3, then for  $H(M_{23})$ , our best lower bound is the larger of (9) and (20).

We call a protocol randomness-optimal, if the total number of random bits used by the protocol is optimal. In all the examples we consider in the next section, the amount of randomness does not depend on the input distribution  $p_{XY}$  as long as they have full support, so, instead of writing  $\rho(p_{XY}, p_{Z|XY})$ , we simply write  $\rho(p_{Z|XY})$ .

13

# 5 Application to Specific Functions

In this section we consider a few important examples, and apply our generic lower bounds from above to these examples, to obtain interesting results. While many of these results are natural to conjecture, they are not easy to prove (see, for instance, Footnote 5).

Optimality of the FKN Protocol. Feige et al. [FKN94] provided a generic (non-interactive) secure computation protocol for all 3-party functions in our model. This protocol uses a straight-forward (but "inefficient") reduction from an arbitrary function to a variant of the oblivious transfer problem, which we shall call the remote OT function (defined below), and then gives a simple protocol for this new function. While the resulting protocol is inefficient for most functions, one could ask whether the protocol that [FKN94] used for REMOTE OT itself is optimal. We use our lower bounds from above to show that indeed, this is the case.

The REMOTE  $\binom{m}{1}$ -OT $^n_2$  function, is defined as follows: Alice's input  $X = (X_0, X_1, \ldots, X_{m-1})$  is made up of m bit-strings each of length n, and Bob has an input  $Y \in \{0, 1, \ldots, m-1\}$ . Charlie wants to compute  $Z = f(X,Y) = X_Y$ . Figure 5 in Appendix F.4 gives the simple protocol for this function from [FKN94] (rephrased as a protocol in our model). It requires nm bits to be exchanged over the Alice-Charlie (31) link,  $n + \log m$  bits over the Bob-Charlie (23) link and  $nm + \log m$  bits over the Alice-Bob (12) link. The total number of random bits used in the protocol is  $nm + \log m$ . In Appendix F.4, we prove the following theorem, which shows that this protocol is optimal and in fact, a communication-ideal protocol. We also prove that this protocol is randomness-optimal.

**Theorem 7.** Any secure protocol  $\Pi(p_{XY}, \text{REMOTE-OT})$  for computing REMOTE  $\binom{m}{1}$ -OT<sup>n</sup> for inputs X and Y where  $p_{XY}$  has full support, must satisfy

$$H(M_{31}) \ge nm$$
,  $H(M_{23}) \ge n + \log m$ , and  $H(M_{12}) \ge nm + \log m$ ,  $\rho(\text{REMOTE-OT}) \ge nm + \log m$ .

More Functions with Communication-Ideal and Randomness-Optimal Protocols. GROUP-ADD, addition in any group has a communication-ideal and randomness-optimal protocol, for any input distribution with full support (see Appendix F.1). As mentioned in Footnote 5, this is easy to see for the uniform distribution, and using distribution switching, we can see that the same holds as long as the input distribution has full support. A more interesting example, is a function called CONTROLLED-ERASURE that was studied in [DP13]. We resolve the communication complexity of this secure computation problem fully, by showing that the protocol for this function from [DP13] is in fact communication-ideal as well as randomness-optimal, again, for every input distribution with full support.

Separating Secure and Insecure Computation. A basic question of secure computation is whether it needs more bits to be communicated than the input-size itself (which suffices for insecure computation). While natural to expect, it is not easy to prove this. In their restricted model, [FKN94] showed a non-explicit result, that for securely computing most Boolean functions on the domain  $\{0,1\}^n \times \{0,1\}^n$ , Charlie is required to receive at least 3n-4 bits, which is significantly more than the 2n bits sufficient for insecure computation.

REMOTE  $\binom{2}{1}$ -OT $^n_2$  from above already gives us an explicit example of a function where this is true: the total input size is 2n+1, but the communication is at least  $H(M_{31})+H(M_{23})\geq 3n+1$ . To present an easy comparison to the lower bound of [FKN94], we can consider a symmetrized variant of REMOTE  $\binom{2}{1}$ -OT $^n_2$ , in which two instances of REMOTE  $\binom{2}{1}$ -OT $^n_2$  are combined, one in each direction. More specifically,  $X=(A_0,A_1,a)$  where  $A_0,A_1$  are of length (n-1)/2 (for an odd n) and a is a single bit; similarly  $Y=(B_0,B_1,b)$ ; the output of the function is defined as an (n-1) bit string  $f(X,Y)=(A_b,B_a)$ . Considering (say) the uniform input distribution over X,Y, the bounds for

REMOTE  $\binom{2}{1}$ -OT<sup>n</sup> add up to give us  $H(M_{31}) \ge 3(n-1)/2 + 1$  and  $H(M_{23}) \ge 3(n-1)/2 + 1$ , so that the communication with Charlie is lowerbounded by  $H(M_{31}) + H(M_{23}) \ge 3n - 1$ .

This compares favourably with the bound of [FKN94] in many ways: our lower bound holds even in a model that allows interaction; in particular, this makes the gap between insecure computation (n-1) bits in our case, 2n bits for [FKN94]) and secure computation (about 3n bits for both) somewhat larger. More importantly, our lower bound is explicit (and tight for the specific function we use), whereas that of [FKN94] is existential. However, our bound does not subsume that of [FKN94], who considered Boolean functions. Our results do not yield a bound significantly larger than the input size, when the output is a single bit. It appears that this regime is more amenable to combinatorial arguments, as pursued in [FKN94], rather than information theoretic arguments. Finally, for the case of random Boolean functions, it is plausible that the actual communication cost is exponential in the input length, but none of the current techniques are capable of delivering such a result. We leave it as a fascinating open problem to obtain tight bounds in this regime, possibly by combining combinatorial and information-theoretic approaches.

Separating Secure Computation and Secret Sharing. Another natural separation one expects is between the amount of communication needed when the views (or transcripts) are generated by a secure computation protocol, versus when they are generated by an omniscient "dealer" so that the security requirements are met. As mentioned before, the latter setting corresponds to the share sizes in a CMSS scheme. Again, while such a separation is expected, it is not very easy to establish this, especially for explicit examples. It requires us to establish a strong lower bound for the secure computation problem as well as provide a CMSS scheme that is better. None of the examples considered above yield this separation.

We establish the separation using the 3-party AND function, defined as follows: Alice has an input bit X, Bob has an input bit Y and Charlie should obtain  $Z = f(X, Y) = X \wedge Y$ . There is a CMSS scheme that achieves  $\log(3) \leq 1.6$  bits of entropy for all three shares  $M_{12}, M_{23}$  and  $M_{31}$  (see Theorem 12). However, the following lower bounds, proven in Appendix F.5 using numerical optimization, shows that in a secure computation protocol,  $H(M_{12})$  should be strictly larger than this.

**Theorem 8.** Any secure protocol  $\Pi(p_{XY}, AND)$  for computing AND for inputs X and Y where  $p_{XY}$  has full support over  $\{0,1\}^n \times \{0,1\}^n$ , must satisfy

$$H(M_{31}) \ge n \log(3)$$
,  $H(M_{23}) \ge n \log(3)$ , and  $H(M_{12}) \ge n(1.826)$ ,  $\rho(\text{AND}) \ge n(1.826)$ .

The best known protocol for AND (which resembles the CMSS scheme above, and first appeared in [FKN94]) achieves  $H(M_{12}) = 1 + \log(3)$ ,  $H(M_{23}) = H(M_{31}) = \log(3)$  (see Appendix F.5). Our lower bounds on  $H(M_{31})$  and  $H(M_{23})$  match with the protocol requirements on these links. The bound on  $H(M_{12})$  is not known to be tight. The protocol given in Appendix F.5 requires  $1 + \log(3)$  random bits and we prove a lower bound of 1.826.

Open Problems. We close with a brief list of concrete open problems from this work. For secure computation of AND and SUM (see Appendix F.2 for SUM) there is a gap between the best known upper bound and our lower bound for  $M_{12}$  link. These specific examples point to challenges in obtaining good lower bounds. Another important problem is to find an explicit example for a *Boolean* function in which the communication to Charlie must be significantly larger than the total input size. Note that [FKN94] gave an existential result (in their restricted model) and the explicit example in this work does not have Boolean output. The case of random Boolean functions, where communication being exponential in the input length is a plausible, but unproven result, was already mentioned.

# References

- [BGW88] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Proc.* 20th STOC, pages 1–10. ACM, 1988. 1, 4, 5
- [BIKK14] Amos Beimel, Yuval Ishai, Ranjit Kumaresan, and Eyal Kushilevitz. On the cryptographic complexity of the worst functions. In *Theory of Cryptography*, volume 8349, pages 317–342. Springer Berlin Heidelberg, 2014. 4
- [BO11] Amos Beimel and Ilan Orlov. Secret sharing and non-shannon information inequalities. IEEE Transactions on Information Theory, 57(9):5634–5649, 2011. 4
- [BP13] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. In FOCS, pages 688–697, 2013. 1
- [BSC<sup>+</sup>94] Carlo Blundo, Alfredo De Santis, Giovanni Di Crescenzo, Antonio Giorgio Gaggia, and Ugo Vaccaro. Multi-secret sharing schemes. In Yvo Desmedt, editor, *CRYPTO*, volume 839 of *Lecture Notes in Computer Science*, pages 150–163. Springer, 1994. 2, 4
- [BSPV99] Carlo Blundo, Alfredo De Santis, Giuseppe Persiano, and Ugo Vaccaro. Randomness complexity of private computation. *Computational Complexity*, 8(2):145–168, 1999. 4
- [CCD88] David Chaum, Claude Crépeau, and Ivan Damgård. Multiparty unconditionally secure protocols. In *Proc.* 20th STOC, pages 11–19. ACM, 1988. 1, 4
- [Cha89] David Chaum. The spymasters double-agent problem: Multiparty computations secure unconditionally from minorities and cryptographically from majorities. In Gilles Brassard, editor, *CRYPTO*, volume 435 of *Lecture Notes in Computer Science*, pages 591–602. Springer, 1989. 4
- [CK93] Benny Chor and Eyal Kushilevitz. A communication-privacy tradeoff for modular addition. Inf. Process. Lett., 45(4):205–210, 1993. 1, 2, 4
- [CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In FOCS, pages 270–278. IEEE, 2001. 4
- [DI06] Ivan Damgård and Yuval Ishai. Scalable secure multiparty computation. In Cynthia Dwork, editor, *CRYPTO*, volume 4117 of *Lecture Notes in Computer Science*, pages 501–520. Springer, 2006. 1, 4
- [DM99] Yevgeniy Dodis and Silvio Micali. Lower bounds for oblivious transfer reductions. In Jacques Stern, editor, *EUROCRYPT*, volume 1592 of *Lecture Notes in Computer Science*, pages 42–55. Springer, 1999. 4
- [DM00] Yevgeniy Dodis and Silvio Micali. Parallel reducibility for information-theoretically secure computation. In Mihir Bellare, editor, *CRYPTO*, volume 1880 of *Lecture Notes in Computer Science*, pages 74–92. Springer, 2000. 7
- [DP13] Deepesh Data and Vinod M. Prabhakaran. Communication requirements for secure computation. In *Proc. 51st Annual Allerton Conference on Communication, Control, and Computing*, 2013. 4, 7, 8, 14, 30

- [FGG+06] Matthias Fitzi, Juan Garay, Shyamnath Gollakota, C Pandu Rangan, and Kannan Srinathan. Round-optimal and efficient verifiable secret sharing. In *Theory of Cryptography*, pages 329–342. Springer, 2006. 4
- [FHM99] Matthias Fitzi, Martin Hirt, and Ueli M. Maurer. General adversaries in unconditional multi-party computation. In Kwok-Yan Lam, Eiji Okamoto, and Chaoping Xing, editors, ASIACRYPT, volume 1716 of Lecture Notes in Computer Science, pages 232–246. Springer, 1999. 4
- [FKN94] Uriel Feige, Joe Kilian, and Moni Naor. A minimal model for secure computation (extended abstract). In *STOC*, pages 554–563. ACM, 1994. 1, 3, 4, 14, 15, 32, 33
- [FL82] Michael J. Fischer and Nancy A. Lynch. A lower bound for the time to assure interactive consistency. *Inf. Process. Lett.*, 14(4):183–186, 1982. 4
- [FY92] Matthew K. Franklin and Moti Yung. Communication complexity of secure computation (extended abstract). In *STOC*, pages 699–710. ACM, 1992. 1, 2, 4
- [GIKR01] Rosario Gennaro, Yuval Ishai, Eyal Kushilevitz, and Tal Rabin. The round complexity of verifiable secret sharing and secure multicast. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 580–589. ACM, 2001. 4
- [GK73] Peter Gács and János Körner. Common information is far less than mutual information. Problems of Control and Information Theory, 2(2):149–162, 1973. 1, 6
- [GR05] Anna Gál and Adi Rosén. Omega(log n) lower bounds on the amount of randomness in 2-private computation. SIAM J. Comput., 34(4):946–959, 2005. 4
- [HLMR12] Martin Hirt, Christoph Lucas, Ueli Maurer, and Dominik Raub. Passive corruption in statistical multi-party computation. In Adam Smith, editor, *The 6th International Conference on Information Theoretic Security ICITS 2012*, Lecture Notes in Computer Science. Springer-Verlag, 2012. Full Version available from http://eprint.iacr.org/2012/272. 4
- [HM97] Martin Hirt and Ueli M. Maurer. Complete characterization of adversaries tolerable in secure multi-party computation (extended abstract). In *PODC*, pages 25–34, 1997. 4
- [IK04] Yuval Ishai and Eyal Kushilevitz. On the hardness of information-theoretic multiparty computation. In *EUROCRYPT*, pages 439–455, 2004. 1, 4
- [KKK09] Jonathan Katz, Chiu-Yuen Koo, and Ranjit Kumaresan. Improving the round complexity of vss in point-to-point networks. *Inf. Comput.*, 207(8):889–899, 2009. 4
- [KLL<sup>+</sup>12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *FOCS*, pages 500–509, 2012. 1
- [KM97] Eyal Kushilevitz and Yishay Mansour. Randomness in private computations. SIAM J. Discrete Math., 10(4):647–661, 1997. 4
- [KN97] Eyal Kushilevitz and Noam Nisan. Communication complexity. Cambridge University Press, New York, 1997. 4
- [Kus89] Eyal Kushilevitz. Privacy and communication complexity. In *FOCS*, pages 416–421. IEEE, 1989. 1, 4

- [LA14] Eun Jee Lee and Emmanuel Abbe. A shannon approach to secure multi-party computations. CoRR, abs/1401.7360, 2014. 4, 29
- [MW03] Ueli M. Maurer and Stefan Wolf. Secret-key agreement over unauthenticated public channels iii: Privacy amplification. *IEEE Transactions on Information Theory*, 49(4):839–851, 2003.
- [NN01] Moni Naor and Kobbi Nissim. Communication preserving protocols for secure function evaluation. In *STOC*, pages 590–599, 2001. 4
- [OR01] Alon Orlitsky and James R Roche. Coding for computing. *Information Theory*, *IEEE Transactions on*, 47(3):903–917, 2001. 4
- [PCRR09] Arpita Patra, Ashish Choudhary, Tal Rabin, and C Pandu Rangan. The round complexity of verifiable secret sharing revisited. In Advances in Cryptology-CRYPTO 2009, pages 487– 504. Springer, 2009. 4
- [PP] Manoj M. Prabhakaran and Vinod M. Prabhakaran. Communication complexity lower bounds from assisted common information. Under Preparation. 1
- [PP12] Vinod M. Prabhakaran and Manoj M. Prabhakaran. Assisted common information with an application to secure two-party sampling. arXiv, abs/1206.1282, 2012. Preliminary versions appeared at ISIT 2010 and ISIT 2011. 1, 4, 5, 6, 7
- [WW08] Stefan Wolf and Jürg Wullschleger. New monotones and lower bounds in unconditional two-party computation. *IEEE Transactions on Information Theory*, 54(6):2792–2797, 2008. 4, 6
- [WW10] Severin Winkler and Jürg Wullschleger. On the efficiency of classical and quantum oblivious transfer reductions. In *CRYPTO*, pages 707–723, 2010. Full version available at http://arxiv.org/abs/1205.5136. 5, 7
- [Wyn75] A. D. Wyner. The wire-tap channel. The Bell system Technical Journal, 54(8):1355–1387, 1975.
- [Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *STOC*, pages 209–213. ACM, 1979. 1, 4

# A Preliminaries: More Details

**Protocol.** In an execution of the protocol, (a subset of) the parties receive inputs from the specified distribution, and exchange messages over private, point-to-point links. Without loss of generality, the state of the protocol consists only of the inputs received by each party and the bits transmitted on each link, so far. This is because, as the parties are computationally unbounded, private randomness for a party can always be resampled at every round conditioned on the inputs, outputs and messages in that party's view. Each  $\Pi_i$  specifies a distribution over  $\{0,1\}^* \times \{0,1\}^*$  (corresponding to the messages to be transmitted on the two links party i is connected to) conditioned on the inputs of party i and the messages in the i-th party's links so far. For a protocol to be well-formed, we require that the message sent by one party to another at any round is a codeword in a prefix-free code such that the code itself is determined by the messages exchanged between the two parties in previous rounds; also, we require

<sup>&</sup>lt;sup>8</sup>Without such a restriction, we implicitly allow "end-of-message" markers, and the rest of the communication could have fewer bits than the entropy of the transcript annotated with rounds.

that with probability 1, the protocol terminates — i.e., it reaches a state when all  $\Pi_i$  output empty strings (it is not important for our lower bounds that the parties realize when this happens).

The final transcript in each of the three links consists of two strings, obtained by simply concatenating all the bits sent in each direction on that link, by the time the protocol has terminated.

A Normal Form for Functionality  $p_{Z|XY}$ . For a randomized function  $p_{Z|XY}$ , define the relation  $x \equiv x'$  for  $x, x' \in \mathcal{X}$  to hold if  $\forall y \in \mathcal{Y}, z \in \mathcal{Z}, p(z|x,y) = p(z|x',y)$ ; similarly  $y \equiv y'$  is defined. We also define  $z \equiv z'$  if there exists a constant c such that  $\forall x \in \mathcal{X}, y \in \mathcal{Y}, p(z|x,y) = c \cdot p(z'|x,y)$ . We say that  $p_{Z|XY}$  is in the normal form if  $x \equiv x' \Rightarrow x = x', y \equiv y' \Rightarrow y = y'$  and  $z \equiv z' \Rightarrow z = z'$ .

It is easy to see that one can transform any randomized function  $p_{Z|XY}$  to one in normal form  $p_{Z'|X'Y'}$ , with possibly smaller alphabets, so that any secure computation protocol for the former can be transformed to one for the latter with the same communication costs, and vice versa. (To define X',  $\mathcal{X}$  is modified by replacing all x in an equivalence class of  $\equiv$  with a single representative; Y' and Z' are defined similarly. The modification to the protocol, in either direction, is for each party to locally map X to X' etc., or vice versa; notice that the Z' to Z map is potentially randomized.) Hence it is enough to study the communication complexity of securely computing functions in the normal form.

**A Normal Form for**  $(p_{XY}, p_{Z|XY})$ . We define a normal form for the pair  $(p_{XY}, p_{Z|XY})$ , where  $p_{XY}$  is the input distribution and the randomized function is  $p_{Z|XY}$  as follows:

**Definition 2.** For a pair  $(p_{XY}, p_{Z|XY})$ , define the relations  $x \cong x'$ ,  $y \cong y'$  and  $z \cong z'$  as follows.

- 1. Take any  $x, x' \in \mathcal{X}$  and define  $\mathcal{S}_{x,x'} = \{y \in \mathcal{Y} : p_{XY}(x,y) > 0, p_{XY}(x',y) > 0\}$ . We say that  $x \cong x'$ , if  $\forall y \in \mathcal{S}_{x,x'}$  and  $z \in \mathcal{Z}$ , we have  $p_{Z|XY}(z|x,y) = p_{Z|XY}(z|x',y)$ .
- 2. Take any  $y, y' \in \mathcal{Y}$  and define  $\mathcal{S}_{y,y'} = \{x \in \mathcal{X} : p_{XY}(x,y) > 0, p_{XY}(x,y') > 0\}$ . We say that  $y \cong y'$ , if  $\forall x \in \mathcal{S}_{y,y'}$  and  $z \in \mathcal{Z}$ , we have  $p_{Z|XY}(z|x,y) = p_{Z|XY}(z|x,y')$ .
- 3. Take  $z, z' \in \mathcal{Z}$  and define  $\mathcal{S} = \{(x, y) : p_{XY}(x, y) > 0\}$ . We say that  $z \cong z'$ , if for some constant  $c \geq 0$  and  $\forall (x, y) \in \mathcal{S}$ , we have  $p_{Z|XY}(z|x, y) = c \cdot p_{Z|XY}(z'|x, y)$ .

A pair  $(p_{XY}, p_{Z|XY})$  is said to be in normal form if  $x \cong x' \Rightarrow x = x'$ ,  $y \cong y' \Rightarrow y = y'$ , and  $z \cong z' \Rightarrow z = z'$ .

We can assume without loss of generality that  $(p_{XY}, p_{Z|XY})$  is in normal form. Otherwise, suppose  $x, x' \in \mathcal{X}$ , where  $x \neq x'$  and  $x \cong x'$ . In this case we can safely merge x and x' into a single  $x^*$  without affecting anything. Now,  $\forall (y, z) \in \mathcal{Y} \times \mathcal{Z}$ , define  $p_{XY}(x^*, y) = p_{XY}(x, y) + p_{XY}(x', y)$  and

$$p_{Z|XY}(z|x^*, y) = \begin{cases} p_{Z|XY}(z|x, y) & \text{if } y \in \mathcal{S}_{x, x'}, \\ p_{Z|XY}(z|x, y) & \text{if } p_{XY}(x, y) \ge 0 \text{ and } p_{XY}(x', y) = 0, \\ p_{Z|XY}(z|x', y) & \text{if } p_{XY}(x, y) = 0 \text{ and } p_{XY}(x', y) > 0, \end{cases}$$

which gives  $p_{XYZ}(x^*,y,z) = p_{XYZ}(x,y,z) + p_{XYZ}(x',y,z)$ . Similarly we can merge equivalent y's. For  $z \neq z'$  with  $z \cong z'$ , we can merge z and z' into a single  $z^*$  by defining  $p_{Z|XY}(z^*|x,y) = p_{Z|XY}(z|x,y) + p_{Z|XY}(z'|x,y)$ , which gives  $p_{XYZ}(x,y,z^*) = p_{XYZ}(x,y,z) + p_{XYZ}(x,y,z')$ .

It is easy to see that one can transform any pair  $(p_{XY}, p_{Z|XY})$  defined by the given  $p_{XY}$  and  $p_{Z|XY}$  to one in normal form  $(p_{X'Y'}, p_{Z'|X'Y'})$  using the above described modification, with possibly smaller alphabets, so that any secure computation protocol for the former can be transformed to one for the latter with the same communication costs, and vice versa. The modification to the protocol, in either direction, is for each party to locally map X to X' etc., or vice versa; notice that the Z' to Z map is potentially randomized. Hence, it is enough to study secure computation problems  $(p_{XY}, p_{Z|XY})$  in normal form.

# B Entropy Lower Bounds as Communication Lower Bounds

Our interest is in providing bounds on the amount of communication needed. In this paper, we derive lower bounds on the entropies,  $H(M_{12})$ ,  $H(M_{23})$ ,  $H(M_{31})$ , of the transcripts on the links. As we argue below, these bounds will lowerbound the expected number of bits exchanged over the link when we require that the message (bit string) sent by one party to another in any round is a codeword in a prefix-free code such that the code itself is determined by the messages exchanged between the two parties in previous rounds. This is a natural requirement when variable length bit strings are allowed as messages. Notice that without this restriction, since a party sees the messages in each round rather than just a concatenation of all the bits sent over the entire protocol, implicit end-of-message markers are present which may convey implicit information (without being accounted for in communication complexity which is measured by the length of the transcript). The prefix-free requirement eliminates the possibility of information being implicitly conveyed through end-of-message markers.

Let us denote by  $M_{i\vec{j},t}$  the message sent by party-i to party-j in round-t. Let  $L_{i\vec{j},t}$  be the (potentially random) length in bits of this message, and let  $L_{ij} = \sum_{t=1}^{N} L_{i\vec{j},t} + L_{j\vec{i},t}$  be the number of bits exchanged over the link ij in either direction. We are interested in lower bounds for  $\mathbb{E}[L_{ij}]$ . We have

$$H(M_{ij}) = \sum_{t=1}^{\infty} H(M_{i\vec{j},t}, M_{j\vec{i},t} | M_{i\vec{j}}^{t-1}, M_{j\vec{i}}^{t-1})$$

$$\leq \sum_{t=1}^{\infty} H(M_{i\vec{j},t} | M_{i\vec{j}}^{t-1}, M_{j\vec{i}}^{t-1}) + H(M_{j\vec{i},t} | M_{i\vec{j}}^{t-1}, M_{j\vec{i}}^{t-1})$$

$$\stackrel{\text{(a)}}{\leq} \sum_{t=1}^{\infty} \mathbb{E}[L_{i\vec{j},t}] + \mathbb{E}[L_{j\vec{i},t}]$$

$$= \mathbb{E}[L_{ij}],$$

where (a) follows from the fact that the prefix-code of which  $M_{i\vec{j},t}$  is a codeword, is a function of the conditioning random variables  $(M_{i\vec{j}}^{t-1}, M_{j\vec{i}}^{t-1})$ , and hence the conditional entropy  $H(M_{i\vec{j},t}|M_{i\vec{j}}^{t-1}, M_{j\vec{i}}^{t-1})$  is no larger than  $\mathbb{E}[L_{i\vec{j},t}]$  (by Kraft's inequality and non-negativity of Kullback-Leibler divergence); similarly for the second term.

# C Connections to Secure Sampling and Correlated Multi-Secret Sharing

**Secure Sampling.** In secure sampling functionalities, none of the parties receives any input, but all three parties produce outputs. The functionality is specified by a joint distribution  $p_{XYZ}$  and the protocol for sampling  $p_{XYZ}$  is specified by  $\Pi(p_{XYZ})$ . The correctness condition in this case is that the outputs of Alice, Bob and Charlie are distributed according to  $p_{XYZ}$ . The security conditions remain the same as in the case of secure computation, that is, none of the parties can infer anything about the other parties' outputs other than what they can from their own outputs.

**A Normal Form for**  $p_{XYZ}$ . For a joint distribution  $p_{XYZ}$ , define the relation  $x \sim x'$  for  $x, x' \in \mathcal{X}$  to hold if  $\exists c \geq 0$  such that  $\forall y \in \mathcal{Y}, z \in \mathcal{Z}, p(x, y, z) = c \cdot p(x', y, z)$ . Similarly, we define  $y \sim y'$  for  $y, y' \in \mathcal{Y}$  and  $z \sim z'$  for  $z, z' \in \mathcal{Z}$ . We say that  $p_{XYZ}$  is in the normal form if  $x \sim x' \Rightarrow x = x', y \sim y' \Rightarrow y = y'$  and  $z \sim z' \Rightarrow z = z'$ .

It is easy to see that one can transform any distribution  $p_{XYZ}$  to one in normal form  $p_{X'Y'Z'}$ , with possibly smaller alphabets, so that any secure sampling protocol for the former can be transformed to one for the latter with the same communication costs, and vice versa. (To define X', X is modified by

removing all x such that p(x) = 0 and then replacing all x in an equivalence class of  $\sim$  with a single representative; Y' and Z' are defined similarly. The modification to the protocol, in either direction, is for each party to locally map X to X' etc., or vice versa.) Hence it is enough to study the communication complexity of securely sampling distributions in the normal form.

Now, we show an analog of Lemma 2 for secure sampling protocols.

**Lemma 6.** Suppose  $p_{XYZ}$  is in normal form. Then, in any secure sampling protocol  $\Pi(p_{XYZ})$ , the cut isolating Alice from Bob and Charlie must determine Alice's output X, i.e.,  $H(X|M_{12},M_{31})=0$ . Similarly,  $H(Y|M_{12},M_{23})=0$  and  $H(Z|M_{23},M_{31})=0$ .

Proof. We only prove  $H(X|M_{12},M_{31})=0$ ; the other ones, i.e.,  $H(Y|M_{12},M_{23})=0$  and  $H(Z|M_{23},M_{31})=0$  are similarly proved. We need to show that for every  $m_{12},m_{31}$  with  $p(m_{12},m_{31})>0$ , there is a (necessarily unique)  $x\in\mathcal{X}$  such that  $p(x|m_{12},m_{31})=1$ . Suppose, to the contrary, that we have a secure sampling protocol resulting in a p.m.f.  $p(x,y,z,m_{12},m_{31})$  such that there exists  $x,x'\in\mathcal{X}, x\neq x'$  and  $m_{12},m_{31}$  satisfying  $p(m_{12},m_{31})>0$ ,  $p(x|m_{12},m_{31})>0$  and  $p(x'|m_{12},m_{31})>0$ . Since  $p(m_{12},m_{31})>0$  and  $p(x|m_{12},m_{31})>0$  imply  $p_X(x)>0$ , there exists (y,z) s.t.  $p_{XYZ}(x,y,z)>0$ .

- (i) The definition of a protocol implies that  $p(x, y, z, m_{12}, m_{31})$  can be written as  $p_{YZ}(y, z)p(m_{12}, m_{31}|y, z)p(x|m_{12}, m_{31})$ .
- (ii) Privacy against Alice implies that  $p(x, y, z, m_{12}, m_{31})$  can be written as  $p_{XYZ}(x, y, z)p(m_{12}, m_{31}|x)$ .
- (iii) (i) and (ii) gives  $p_{YZ}(y,z)p(m_{12},m_{31}|y,z)p(x|m_{12},m_{31}) = p_{XYZ}(x,y,z)p(m_{12},m_{31}|x).$

By assumption,  $p(m_{12}, m_{31}) > 0$  and  $p(x|m_{12}, m_{31}) > 0$ , which imply that  $p(m_{12}, m_{31}|x) > 0$ . And since  $p_{XYZ}(x, y, z) > 0$ , we have from (iii) that  $p(m_{12}, m_{31}|y, z) > 0$ . Now consider (x', y, z). By assumption,  $p(m_{12}, m_{31}) > 0$  and  $p(x'|m_{12}, m_{31}) > 0$ , which imply  $p(m_{12}, m_{31}|x') > 0$ . Since  $p(m_{12}, m_{31}|y, z) > 0$  from above, (iii) implies that  $p_{XYZ}(x', y, z) > 0$ . Define  $\alpha \triangleq \frac{p(x, y, z)}{p(x', y, z)}$ . Since  $p_{XYZ}(x, y', z')$  or p(x', y', z') is non-zero. Assume that any one of these is non-zero, then applying the above arguments will give us that the other one should also be non-zero.

- (iv) Dividing the expression in (iii) by the one we obtain when we apply the above arguments to (x', y, z) gives  $\frac{p(x|m_{12}, m_{31})}{p(x'|m_{12}, m_{31})} = \alpha \cdot \frac{p(m_{12}, m_{31}|x)}{p(m_{12}, m_{31}|x')}$ .
- (v) Repeating (i)-(iv) for (x, y', z') and (x', y', z'), we get  $\frac{p(x|m_{12}, m_{31})}{p(x'|m_{12}, m_{31})} \neq \alpha \cdot \frac{p(m_{12}, m_{31}|x)}{p(m_{12}, m_{31}|x')}$ , which contradicts (iv).

**Theorem 9.** Any secure sampling protocol  $\Pi(p_{XYZ})$ , where  $p_{XYZ}$  is in normal form, should satisfy the following lower bounds on the entropy of the transcripts on each link.

$$H(M_{23}) \ge RI(X; Z) + RI(X; Y) + H(Y, Z|X),$$
  
 $H(M_{31}) \ge RI(Y; Z) + RI(X; Y) + H(X, Z|Y),$   
 $H(M_{12}) \ge RI(X; Z) + RI(Y; Z) + H(X, Y|Z).$ 

*Proof.* From Lemma 6, we have  $H(X|M_{12}, M_{31}) = 0$ ,  $H(Y|M_{12}, M_{23}) = 0$  and  $H(Z|M_{23}, M_{31}) = 0$ . Note that we can apply Lemma 4 for secure sampling of dependent X, Y and Z, because, in the beginning, parties only have independent randomness, but no inputs. In the end, they output from a

joint distribution  $p_{XYZ}$ , where X, Y and Z may be dependent, but this does not affect the requirements of Lemma 4 in any way. The proof for  $H(M_{23})$  is given below; the other two bounds follows similarly.

$$H(M_{23}) = I(M_{12}; M_{23}) + H(M_{23}|M_{12})$$

$$= I(M_{12}; M_{23}) + I(M_{31}; M_{23}|M_{12}) + H(M_{23}|M_{12}, M_{31})$$

$$\stackrel{\text{(a)}}{\geq} I(M_{12}; M_{23}|M_{31}) + I(M_{31}; M_{23}|M_{12}) + H(M_{23}|M_{12}, M_{31})$$

$$\stackrel{\text{(b)}}{\geq} RI(X; Z) + RI(X; Y) + H(Y, Z|X),$$

where (a) used  $I(M_{12}; M_{23}) \ge I(M_{12}; M_{23}|M_{31})$ , which follows from Lemma 4; (b) used  $I(M_{12}; M_{23}|M_{31}) \ge RI(X; Z)$ ,  $I(M_{31}; M_{23}|M_{12}) \ge RI(X; Y)$  and  $H(M_{23}|M_{12}, M_{31}) \ge H(Y, Z|X)$ , which we have shown in the proof of Theorem 1.

We remark that if the marginal distributions satisfy  $p_{XY} = p_X p_Y$  (i.e., X and Y are independent), then a secure computation protocol for  $p_{Z|XY}$  can be turned into a secure sampling protocol (with the same communication costs), by having Alice and Bob locally sample inputs X and Y according to  $p_X$  and  $p_Y$  and then run the computation protocol. So, whenever X and Y are independent, the lower bounds on communication for secure sampling imply lower bounds for secure computation.

Correlated Multi-Secret Sharing Schemes. We define a notion of secret-sharing, called Correlated Multi-Secret Sharing (CMSS) that is closely related to secure sampling/computation problem. We will show that lower bounds on the entropy of shares of such secret-sharing schemes will also be lower bounds on entropy of transcripts for the corresponding secure computation protocols. However, we shall show a separation between the efficiency of secret-sharing (where there is an omniscient dealer) and a protocol, using the stronger lower bounds we have established in Section 3.

**Definition 3.** Given a graph G = (V, E), an adversary structure  $A \subseteq 2^V$ , and a joint distribution  $p_{(X_v)_{v \in V}}$  over random variables  $X_v$  indexed by  $v \in V$ , a correlated multiple secret sharing scheme for  $(G, p_{(X_v)_{v \in V}})$  defines a distribution  $p_{(M_e)_{e \in E}|(X_v)_{v \in V}}$  of shares  $M_e$  for each edge  $e \in E$ , such that the following hold. Below, for  $S \subseteq E$ ,  $M_S$  stands for the collection of all  $M_e$  for  $e \in S$ ; similarly  $X_T$  is defined for  $T \subseteq V$ ;  $E_v \subseteq E$  denotes the set of edges incident on a vertex V.

- Correctness: For all  $v \in V$ ,  $H(X_v|M_{E_v}) = 0$ .
- Privacy: For every set  $T \in \mathcal{A}$ , let  $E_T = \bigcup_{v \in T} E_v$ ; then,  $I(X_{\overline{T}}; M_{E_T} | X_T) = 0$ .

Below we give a specialised version of the above general definition which is suitable to our setting, where G is the clique over the vertex set  $V = \{1, 2, 3\}$ , and  $\mathcal{A} = \{\{1\}, \{2\}, \{3\}\}$  (corresponding to 1-privacy).

We define  $\Sigma$  to be a correlated multi-secret sharing scheme for a joint distribution  $p_{XYZ}$  (with respect to our fixed adversary structures) if it probabilistically maps secrets (X, Y, Z) to shares  $M_{12}, M_{23}, M_{31}$  such that the following conditions hold:

- Correctness:  $H(X|M_{12}, M_{31}) = H(Y|M_{12}, M_{23}) = H(Z|M_{23}, M_{31}) = 0.$
- Privacy:

$$\begin{split} &I((M_{12},M_{31});(Y,Z)|X)=0 & \quad & \text{(privacy against Alice)}, \\ &I((M_{12},M_{23});(X,Z)|Y)=0 & \quad & \text{(privacy against Bob)}, \\ &I((M_{23},M_{31});(X,Y)|Z)=0 & \quad & \text{(privacy against Charlie)}. \end{split}$$

We point out that while the correctness condition relates only to the supports of X, Y and Z individually, the privacy condition is crucially influenced by the joint distribution.

**Theorem 10.** Any CMSS scheme for any joint distribution  $p_{XYZ}$  satisfies

$$H(M_{12}) \ge \max\{RI(X;Z), RI(Y;Z)\} + H(X,Y|Z),$$
  
 $H(M_{23}) \ge \max\{RI(X;Z), RI(X;Y)\} + H(Y,Z|X),$   
 $H(M_{31}) \ge \max\{RI(Y;Z), RI(X;Y)\} + H(X,Z|Y).$ 

*Proof.* We proceed along the lines of the proof of Theorem 1, except that here we do not need Lemma 2 to argue that  $H(X|M_{12}, M_{31}) = H(Y|M_{12}, M_{23}) = H(Z|M_{23}, M_{31}) = 0$ , instead, these follow from the correctness of CMSS.

If  $p_{XYZ} = p_{XY}p_{Z|XY}$ , where  $(p_{XY}, p_{Z|XY})$  is in normal form, using Lemma 2, the bounds in Theorem 10 imply bounds in Theorem 1. If  $p_{XYZ}$  has full support, then we can further strengthen the bounds in Theorem 10 by applying distribution switching.

**Theorem 11.** Consider any CMSS scheme for a joint distribution  $p_{XYZ}$ , where  $p_{XYZ}$  has full support.

- 1.  $H(M_{12}) \ge \max\{\sup_{p_{X'Y'Z'}} (RI(X';Z') + H(X',Y'|Z')), \sup_{p_{X'Y'Z'}} (RI(Y';Z') + H(X',Y'|Z'))\},$ where  $p_{X'Y'Z'}$  is any distribution for which the characteristic bipartite graph of  $p_{X'Y'}$  is connected.
- 2.  $H(M_{23}) \ge \max\{\sup_{p_{X'Y'Z'}} (RI(X';Z') + H(Y',Z'|X')), \sup_{p_{X'Y'Z'}} (RI(X';Y') + H(Y',Z'|X'))\},$ where  $p_{X'Y'Z'}$  is any distribution for which the characteristic bipartite graph of  $p_{Y'Z'}$  is connected.
- 3.  $H(M_{31}) \ge \max\{\sup_{p_{X'Y'Z'}} (RI(Y';Z') + H(X',Z'|Y')), \sup_{p_{X'Y'Z'}} (RI(X';Y') + H(X',Z'|Y'))\},$ where  $p_{X'Y'Z'}$  is any distribution for which the characteristic bipartite graph of  $p_{X'Z'}$  is connected.

Proof. First we observe that we can apply distribution switching to CMSS schemes also, i.e., if we have a CMSS  $\Sigma(p_{XYZ})$ , where  $p_{XYZ}$  has full support, it will remain a CMSS if we change the distribution to a different one  $p_{X'Y'Z'}$ . This follows from the correctness and privacy conditions of a CMSS. Proceeding as in the proof of Lemma 3, we can show that for any CMSS  $\Sigma(p_{XYZ})$ , connectedness of the characteristic bipartite graph of  $p_{XY}$  implies  $I(X,Y,Z;M_{12})=0$ . The other two, i.e., connectedness of the characteristic bipartite graph of  $p_{XZ}$  implies  $I(X,Y,Z;M_{23})=0$  and connectedness of the characteristic bipartite graph of  $p_{YZ}$  implies  $I(X,Y,Z;M_{23})=0$ , follow similarly. Now, we can apply the distribution switching to the bounds in Theorem 10.

It is easy to see that any secure sampling protocol  $\Pi(p_{XYZ})$ , where  $p_{XYZ}$  is in normal form, yields a CMSS scheme for the same joint distribution  $p_{XYZ}$ : An omniscient dealer can always produce the shares  $M_{12}, M_{23}, M_{31}$  which are precisely the transcripts produced by the secure sampling protocol. Now, correctness for this CMSS follows from Lemma 6, and privacy of CMSS scheme follows from the privacy of the secure sampling protocol. Thus the lower bounds on the transcripts produced by a CMSS scheme for a given  $p_{XYZ}$  in normal form, gives lower bounds on the corresponding links for any secure sampling protocol for this  $p_{XYZ}$ . If  $p_{XYZ} = p_{XY}p_{Y|XY}$ , where  $(p_{XY}, p_{Z|XY})$  is in normal form, then lower bounds for CMSS schemes provide lower bounds for secure computation problems. As we discuss in page 15, this lower bound is not tight in general, i.e., there is a function (in fact the AND function) for which there is a CMSS scheme which requires less communication than what our lower bound for secure computation for that function provides. Towards this, here we give upper bounds on the share sizes of a 3-party CMSS for AND which is defined as X and Y independent and uniformly distributed bits, and  $Z = X \wedge Y$ .

**Theorem 12.** For  $p_{XYZ}$  such that X and Y independent and uniformly distributed bits, and  $Z = X \wedge Y$ , there is a CMSS  $\Sigma(p_{XYZ})$  which has  $H(M_{12}) = H(M_{23}) = H(M_{31}) = \log(3)$ .

*Proof.* Consider a CMSS scheme  $\Sigma$  defined as follows. Let  $(\alpha, \beta, \gamma)$  be a random permutation of the set  $\{0, 1, 2\}$ . Let  $M_{12} = \alpha$ , and

$$M_{31} = \begin{cases} \alpha & \text{if } X = 1 \\ \beta & \text{if } X = 0 \end{cases}$$
  $M_{23} = \begin{cases} \alpha & \text{if } Y = 1 \\ \gamma & \text{if } Y = 0 \end{cases}$ 

It can be seen that this scheme satisfies the correctness and privacy requirements (in particular,  $(M_{12}, M_{31})$  is uniformly random, conditioned on  $M_{12} = M_{31}$  when X = 1 and conditioned on  $M_{12} \neq M_{31}$  when X = 0).  $H(M_{12}^{\Sigma}) = H(M_{23}^{\Sigma}) = H(M_{31}^{\Sigma}) = \log 3 < 1.585$ .

Theorem 11 implies that this scheme is optimal.

## D Proofs of the Main Theorems

*Proof of Theorem 4.* Here, we prove our lower bounds only for *independent* inputs, i.e.,  $p_{XY} = p_X p_Y$ , but as we show in Appendix G, they also hold for dependent inputs  $p_{XY}$  with full support.

Suppose we have a secure protocol for computing  $p_{Z|XY}$  in the normal form under  $p_X, p_Y$  which have full support. Consider  $H(M_{23})$ ,

$$H(M_{23}) = I(M_{23}; M_{12}) + I(M_{23}; M_{31}|M_{12}) + H(M_{23}|M_{12}, M_{31}).$$

By Lemma 5,  $M_{23}$  is independent of X. So, by distribution switching, we know that we may switch the distribution of X to, say,  $p_{X''}$  which also has full support and the resulting  $M_{23}$  has the same distribution as under  $p_X$ , i.e.,

$$H(M_{23}) = \sup_{p_{X''}} I(M_{23}; M_{12}) + I(M_{23}; M_{31}|M_{12}) + H(M_{23}|M_{12}, M_{31}).$$

Under this switched distribution, let us consider the first term  $I(M_{23}; M_{12})$ . Let us notice that, by privacy against Bob,  $(M_{23}, M_{12})$  must again be independent of X''. Hence, even if we switch the distribution of X to, say  $p_{X'}$ , the joint distribution of  $(M_{23}, M_{12})$  must remain unchanged. Hence, we have that  $I(M_{23}; M_{12})$  under the distribution  $p_{X''}$  is the same as that under  $p_{X'}$ . Therefore,

$$H(M_{23}) = \left(\sup_{p_{X'}} I(M_{23}; M_{12})\right) + \left(\sup_{p_{X''}} I(M_{23}; M_{31}|M_{12}) + H(M_{23}|M_{12}, M_{31})\right).$$

Now proceeding as in the proof of Theorem 3, we have

$$H(M_{23}) \ge \left(\sup_{p_{X'}} RI(X'; Z')\right) + \left(\sup_{p_{X''}} H(Y, Z''|X'')\right).$$

The bound on  $H(M_{31})$  follows in an identical fashion. To see the bounds on  $H(M_{12})$ , let us recall that  $M_{12}$  is independent of X, Y (by Lemma 3) and hence we may switch the distributions of both X and Y. Furthermore, let us note that we may write  $H(M_{12})$  in two different ways.

$$H(M_{12}) = [I(M_{12}; M_{31})] + [I(M_{12}; M_{23}|M_{31}) + H(M_{12}|M_{23}, M_{31})]$$
(22)

$$H(M_{12}) = [I(M_{12}; M_{23})] + [I(M_{12}; M_{31}|M_{23}) + H(M_{12}|M_{23}, M_{31})].$$
(23)

Using (22) and proceeding as we did for  $H(M_{23})$  leads to the top row of the right hand side of (18), and (23) leads to the bottom row.

Proof of Theorem 5. Here, we prove our lower bounds only for independent inputs, i.e.,  $p_{XY} = p_X p_Y$ , but as we show in Appendix G, they also hold for dependent inputs  $p_{XY}$  with full support.

In the proof of Lemma 3, we show that under condition 1,  $M_{31}$  is independent of both X, Y. This allows us to switch the distribution of both X and Y as we did for bounding  $H(M_{12})$  in the proof of Theorem 4. Proceeding in an identical fashion as there leads us to (19). Similarly, under condition 2,  $M_{23}$  is independent of X, Y which leads to (20).

## E Proofs Omitted from Section 3

Proof of Lemma 2. First we will show  $H(X|M_{12}, M_{13}) = 0$ ; the other one, i.e.  $H(Y|M_{12}, M_{23}) = 0$ , is similarly proved. We apply a cut-set argument. Consider the cut isolating Alice from Bob and Charlie.

We need to show that for every  $m_{12}, m_{31}$  with  $p(m_{12}, m_{31}) > 0$ , there is a (necessarily unique)  $x \in \mathcal{X}$  such that  $p(x|m_{12}, m_{31}) = 1$ . Suppose, to the contrary, that we have a secure protocol resulting in a p.m.f.  $p(x, y, z, m_{12}, m_{31})$  such that there exists  $x, x' \in \mathcal{X}$ ,  $x \neq x'$  and  $m_{12}, m_{31}$  satisfying  $p(m_{12}, m_{31}) > 0$ ,  $p(x|m_{12}, m_{31}) > 0$  and  $p(x'|m_{12}, m_{31}) > 0$ . For these x, x', since  $(p_{XY}, p_{Z|XY})$  is in the normal form,  $\exists (y, z) \in \mathcal{Y} \times \mathcal{Z}$  such that  $p_{XY}(x, y) > 0$ ,  $p_{XY}(x', y) > 0$  and  $p_{Z|X,Y}(z|x, y) \neq p_{Z|X,Y}(z|x', y)$ .

- (i) The definition of a protocol implies that  $p(x, y, z, m_{12}, m_{31})$  can be written as  $p_{X,Y}(x,y)p(m_{12}, m_{31}|x,y)p(z|m_{12}, m_{31},y)$ .
- (ii) Privacy against Alice implies that  $p(m_{12}, m_{31}|x, y, z) = p(m_{12}, m_{31}|x)$ .
- (iii) Using (ii) in (i), we get  $p(x, y, z, m_{12}, m_{31}) = p_{X,Y}(x, y)p(m_{12}, m_{31}|x)p(z|m_{12}, m_{31}, y)$ .
- (iv) Correctness and (ii) imply that we can also write  $p(x, y, z, m_{12}, m_{31}) = p_{X,Y}(x, y)p_{Z|X|Y}(z|x, y)p(m_{12}, m_{31}|x)$ .
- (v) Since  $p_{X,Y}(x,y)p(m_{12},m_{31}|x) > 0$ , from (iii) and (iv), we get  $p(z|m_{12},m_{31},y) = p_{Z|X,Y}(z|x,y)$ .

Applying the above arguments to  $(x', y, z, m_{12}, m_{31})$ , we get  $p(z|m_{12}, m_{31}, y) = p_{Z|X,Y}(z|x', y)$ , leading to the contradiction  $p(z|m_{12}, m_{31}, y) \neq p(z|m_{12}, m_{31}, y)$ , since by assumption  $p_{Z|X,Y}(z|x, y) \neq p_{Z|X,Y}(z|x', y)$ .

For  $H(Z|M_{23}, M_{31}) = 0$ , we need to show that for every  $m_{23}, m_{31}$  with  $p(m_{23}, m_{31}) > 0$ , there is a (necessarily unique)  $z \in \mathcal{Z}$  such that  $p(z|m_{23}, m_{31}) = 1$ . Suppose, to the contrary, that we have a secure protocol resulting in a p.m.f.  $p(x, y, z, m_{23}, m_{31})$  such that there exists  $z, z' \in \mathcal{Z}$ ,  $z \neq z'$  and  $m_{23}, m_{31}$  satisfying  $p(m_{23}, m_{31}) > 0$ ,  $p(z|m_{23}, m_{31})$  and  $p(z'|m_{23}, m_{31}) > 0$ . Since  $(p_{XY}, p_{Z|XY})$  is in normal form, there exists (x, y) s.t.  $p_{XY}(x, y) > 0$  and  $p_{Z|X,Y}(z|x, y) > 0$ .

- (i) The definition of a protocol implies that  $p(x, y, z, m_{23}, m_{31})$  can be written as  $p_{X,Y}(x,y)p(m_{23}, m_{31}|x,y)p(z|m_{23}, m_{31})$ .
- (ii) Privacy against Charlie implies that  $p(x, y, z, m_{23}, m_{31})$  can be written as  $p_{X,Y}(x,y)p(z|x,y)p(m_{23},m_{31}|z)$ .
- (iii) (i) and (ii) gives  $p(m_{23}, m_{31}|x, y)p(z|m_{23}, m_{31}) = p_{Z|X,Y}(z|x, y)p(m_{23}, m_{31}|z).$

By assumption,  $p(m_{23}, m_{31}) > 0$  and  $p(z|m_{23}, m_{31}) > 0$ , which imply that  $p(m_{23}, m_{31}|z) > 0$ . And since  $p_{Z|X,Y}(z|x,y) > 0$ , we have from (iii) that  $p(m_{23}, m_{31}|x,y) > 0$ . Now consider (x,y,z'). By assumption,  $p(m_{23}, m_{31}) > 0$  and  $p(z'|m_{23}, m_{31}) > 0$ , which imply  $p(m_{23}, m_{31}|z') > 0$ . Since  $p(m_{23}, m_{31}|x,y) > 0$  from above, (iii) implies that  $p_{Z|X,Y}(z'|x,y) > 0$ . Define  $\alpha \triangleq \frac{p(z|x,y)}{p(z'|x,y)}$ . Since  $(p_{XY}, p_{Z|XY})$  is in normal form,  $\exists (x',y') \in (\mathcal{X},\mathcal{Y})$  s.t.  $p_{XY}(x',y') > 0$  and  $p_{Z|X,Y}(z|x',y') \neq \alpha \cdot p_{Z|X,Y}(z'|x',y')$ . Since  $\alpha \neq 0$ ,

at least one of p(z|x',y') or p(z'|x',y') is non-zero. Assume that any one of these is non-zero, then applying the above arguments will give us that the other one should also be non-zero.

- (iv) Dividing the expression in (iii) by the one we obtain when we apply the above arguments to (x, y, z') gives  $\frac{p(z|m_{23}, m_{31})}{p(z'|m_{23}, m_{31})} = \alpha \cdot \frac{p(m_{23}, m_{31}|z)}{p(m_{23}, m_{31}|z')}$ .
- (v) Repeating (i)-(iv) for (x', y', z) and (x', y', z'), we get  $\frac{p(z|m_{23}, m_{31})}{p(z'|m_{23}, m_{31})} \neq \alpha \cdot \frac{p(m_{23}, m_{31}|z)}{p(m_{23}, m_{31}|z')}$ , which contradicts (iv).

#### Proof of Lemma 3.

1. To show  $I(X, Y, Z; M_{12}) = 0$ , we need only show that  $I(X; M_{12}) = 0$ , since  $I(X, Y, Z; M_{12}) = I(X; M_{12}) + I(Y, Z; M_{12}|X)$  and the second term is equal to zero by the privacy against Alice.

For  $I(X; M_{12}) = 0$ , we need to show that  $p(m_{12}|x) = p(m_{12}|x')$  for all  $x, x' \in \mathcal{X}$ . Take some  $x, x' \in \mathcal{X}, x \neq x'$ . Suppose there is a  $y \in \mathcal{Y}$  s.t.  $p_{XY}(x,y) > 0, p_{XY}(x',y) > 0$ . Then, by privacy against Alice  $p(m_{12}, x, y) = p_{X,Y}(x,y)p(m_{12}|x)$  and by privacy against Bob  $p(m_{12}, x, y) = p_{X,Y}(x,y)p(m_{12}|y)$ . By comparing these two, we get  $p(m_{12}|x) = p(m_{12}|y)$ . Applying the above arguments to  $p(m_{12}, x', y)$  gives  $p(m_{12}|x') = p(m_{12}|y)$ . Hence,  $p(m_{12}|x) = p(m_{12}|x')$ .

Connectedness of the characteristic bipartite graph of  $p_{XY}$  implies that for every  $x, x' \in \mathcal{X}$ , there is a sequence  $x_0 = x, x_1, x_2, \dots, x_{L-1}, x_L = x' \in \mathcal{X}$  such that for every pair  $(x_{l-1}, x_l), l = 1, 2, \dots, L$ , there is a  $y_l \in \mathcal{Y}$  s.t.  $p_{X,Y}(x_{l-1}, y_l) > 0$  and  $p_{X,Y}(x_l, y_l) > 0$ . Hence,  $p(m_{12}|x) = p(m_{12}|x_1) = p(m_{12}|x_2) = \dots = p(m_{12}|x')$ .

2. To show  $I(X, Y, Z; M_{31}) = 0$  under condition 1, we need only show that  $I(X; M_{31}) = 0$ , since  $I(X, Y, Z; M_{31}) = I(X; M_{31}) + I(Y, Z; M_{31}|X)$  and the second term is equal to zero by the privacy against Alice.

We need to show that  $p(m_{31}|x) = p(m_{31}|x')$  for all  $x, x' \in \mathcal{X}$ . Take some  $x, x' \in \mathcal{X}, x \neq x'$ . Suppose there is a  $z \in \mathcal{Z}$  s.t.  $p_{Z|X,Y}(z|x,y), p_{Z|X,Y}(z|x',y') > 0$  for some  $y, y' \in \mathcal{Y}$ . Then, by privacy against Alice  $p(m_{31},x,z) = p_{X,Z}(x,z)p(m_{31}|x)$  and by privacy against Charlie  $p(m_{31},x,z) = p_{X,Z}(x,z)p(m_{31}|z)$ . By comparing these two, and since  $p_{X,Z}(x,z) > 0$  (which follows from the assumption that  $p_{X,Y}$  has full support), we get  $p(m_{31}|x) = p(m_{31}|z)$ . Applying the above arguments to  $p_{X,Z}(x,z)$  gives  $p(m_{31}|x') = p(m_{31}|z)$ . Hence,  $p(m_{31}|x) = p(m_{31}|x')$ .

Condition 1 implies that for every  $x, x' \in \mathcal{X}$ , there is a sequence  $x_0 = x, x_1, x_2, \dots, x_{L-1}, x_L = x' \in \mathcal{X}$  such that for every pair  $(x_{l-1}, x_l), l = 1, 2, \dots, L$ , there is a  $z_l \in \mathcal{Z}$  s.t.  $p_{Z|X,Y}(z_l|x_{l-1}, y_l), p_{Z|X,Y}(z_l|x_l, y'_l) > 0$  for some  $y_l, y'_l \in \mathcal{Y}$ . Hence,  $p(m_{31}|x) = p(m_{31}|x_1) = p(m_{31}|x_2) = \dots = p(m_{31}|x')$ .

3. The other case under condition 2 follows similarly.

*Proof of Lemma 4.* We will apply induction on the number of rounds of the protocol.

Base case: At the beginning of the protocol, all the transcripts  $M_{\gamma\alpha}$ ,  $M_{\beta\gamma}$  and  $M_{\alpha\beta}$  are empty. So, the inequality is trivially true.

Inductive step: Assume that the inequality is true at the end of round t, and we prove it for t+1. For simplicity, let us denote the transcript  $M_{\gamma\alpha}$  (similarly others) at the end of round t by  $M_{\gamma\alpha}$  itself and at the end of round t+1 by  $\widetilde{M}_{\gamma\alpha}$ . We denote by  $\Delta M$ , the new message sent in round t+1 and if

that message is sent from party  $\gamma$  to party  $\alpha$ , we denote it by  $\Delta M_{\gamma\dot{\alpha}}$  and so  $\widetilde{M}_{\gamma\alpha}$  becomes  $(M_{\gamma\alpha}, \Delta M_{\gamma\dot{\alpha}})$ .

Observe that we need to consider only three kinds of messages exchanged in round t+1, which are  $\Delta M_{\vec{\beta}\alpha}$ ,  $\Delta M_{\vec{\beta}\gamma}$  and  $\Delta M_{\gamma\vec{\beta}}$ . The inequality for other three kinds of messages is similarly proved. Since the parties do not share any common or correlated randomness, the new message that one party (say,  $\beta$ ) sends to another (say,  $\alpha$ ) is conditionally independent of the transcript  $(M_{\gamma\alpha})$  between the other two parties ( $\gamma$  and  $\alpha$ ) conditioned on the transcripts  $(M_{\alpha\beta}, M_{\beta\gamma})$  on both of the links to which that party (namely,  $\beta$ ) is associated with. So we have the following:

$$I(M_{\gamma\alpha}; \Delta M_{\vec{\beta\alpha}} | M_{\alpha\beta}, M_{\beta\gamma}) = 0, \tag{24}$$

$$I(M_{\gamma\alpha}; \Delta M_{\vec{\beta\gamma}} | M_{\alpha\beta}, M_{\beta\gamma}) = 0, \tag{25}$$

$$I(M_{\alpha\beta}; \Delta M_{\gamma\beta}|M_{\beta\gamma}, M_{\gamma\alpha}) = 0.$$
 (26)

# 1. If $\Delta M = \Delta M_{\vec{\beta \alpha}}$ , then

$$\begin{split} I(\widetilde{M}_{\gamma\alpha};\widetilde{M}_{\beta\gamma}) &\overset{\text{(a)}}{=} I(M_{\gamma\alpha};M_{\beta\gamma}) \\ &\overset{\text{(b)}}{\geq} I(M_{\gamma\alpha};M_{\beta\gamma}|M_{\alpha\beta}) \\ &\overset{\text{(c)}}{=} I(M_{\gamma\alpha};M_{\beta\gamma},\Delta M_{\vec{\beta\alpha}}|M_{\alpha\beta}) \\ &\overset{\geq}{=} I(\underbrace{M_{\gamma\alpha}}_{\widetilde{M}_{\gamma\alpha}};\underbrace{M_{\beta\gamma}}_{\widetilde{M}_{\beta\gamma}}|\underbrace{M_{\alpha\beta},\Delta M_{\vec{\beta\alpha}}}_{\widetilde{M}_{\alpha\beta}}) \\ &= I(\widetilde{M}_{\gamma\alpha};\widetilde{M}_{\beta\gamma}|\widetilde{M}_{\alpha\beta}), \end{split}$$

where (a) follows because  $\widetilde{M}_{\gamma\alpha} = M_{\gamma\alpha}$  and  $\widetilde{M}_{\beta\gamma} = M_{\beta\gamma}$ , (b) follows from the induction hypothesis and (c) follows from (24).

# 2. If $\Delta M = \Delta M_{\vec{\beta\gamma}}$ , then

$$\begin{split} I(\widetilde{M}_{\gamma\alpha};\widetilde{M}_{\beta\gamma}) &\overset{\text{(d)}}{=} I(M_{\gamma\alpha};M_{\beta\gamma},\Delta M_{\vec{\beta\gamma}}) \\ &\geq I(M_{\gamma\alpha};M_{\beta\gamma}) \\ &\overset{\text{(e)}}{\geq} I(M_{\gamma\alpha};M_{\beta\gamma}|M_{\alpha\beta}) \\ &\overset{\text{(f)}}{\geq} I(\underbrace{M_{\gamma\alpha};M_{\beta\gamma},\Delta M_{\vec{\beta\gamma}}}_{\widetilde{M}_{\beta\gamma}}|\underbrace{M_{\alpha\beta}}_{\widetilde{M}_{\alpha\beta}}) \\ &= I(\widetilde{M}_{\gamma\alpha};\widetilde{M}_{\beta\gamma}|\widetilde{M}_{\alpha\beta}), \end{split}$$

where (d) follows because  $\widetilde{M}_{\gamma\alpha} = M_{\gamma\alpha}$  and  $\widetilde{M}_{\beta\gamma} = (M_{\beta\gamma}, \Delta M_{\vec{\beta\gamma}})$ , (e) follows from the induction hypothesis and (f) follows from (25).

3. If  $\Delta M = \Delta M_{\vec{\gamma}\beta}$ , then

$$\begin{split} I(\widetilde{M}_{\gamma\alpha};\widetilde{M}_{\beta\gamma}) &\overset{\text{(g)}}{=} I(M_{\gamma\alpha};M_{\beta\gamma},\Delta M_{\gamma\vec{\beta}}) \\ &= I(M_{\gamma\alpha};M_{\beta\gamma}) + I(M_{\gamma\alpha};\Delta M_{\gamma\vec{\beta}}|M_{\beta\gamma}) \\ &\overset{\text{(h)}}{=} I(M_{\gamma\alpha};M_{\beta\gamma}) + I(M_{\gamma\alpha};\Delta M_{\gamma\vec{\beta}}|M_{\beta\gamma}) + I(M_{\alpha\beta};\Delta M_{\gamma\vec{\beta}}|M_{\beta\gamma},M_{\gamma\alpha}) \\ &= I(M_{\gamma\alpha};M_{\beta\gamma}) + I(M_{\gamma\alpha},M_{\alpha\beta};\Delta M_{\gamma\vec{\beta}}|M_{\beta\gamma}) \\ &\overset{\text{(i)}}{\geq} I(M_{\gamma\alpha};M_{\beta\gamma}|M_{\alpha\beta}) + I(M_{\gamma\alpha};\Delta M_{\gamma\vec{\beta}}|M_{\beta\gamma},M_{\alpha\beta}) \\ &= I(\underbrace{M_{\gamma\alpha};M_{\beta\gamma},\Delta M_{\gamma\vec{\beta}}}_{\widetilde{M}_{\beta\gamma}}|\underbrace{M_{\alpha\beta}}_{\widetilde{M}_{\alpha\beta}}) \\ &= I(\widetilde{M}_{\gamma\alpha};\widetilde{M}_{\beta\gamma}|\widetilde{M}_{\alpha\beta}), \end{split}$$

where (g) follows because  $\widetilde{M}_{\gamma\alpha}=M_{\gamma\alpha}$  and  $\widetilde{M}_{\beta\gamma}=(M_{\beta\gamma},\Delta M_{\gamma\beta})$ , (h) follows from (26) and (i) follows from the induction hypothesis.

# F Details omitted from Section 5

#### F.1 Secure Computation of GROUP-ADD

Let  $\mathbb{G}$  be a (possibly non-abelian) group with binary operation +. The function GROUP-ADD is defined as follows: Alice has an input  $X \in \mathbb{G}$ , Bob has an input  $Y \in \mathbb{G}$  and Charlie should get Z = f(X,Y) = X + Y.

In Figure 2, we recapitulate a well-known simple protocol for securely computing the above function. The protocol requires a  $|\mathbb{G}|$ -ary symbol to be exchanged per computation over each link. As we show below, this protocol is easily seen to be optimal in terms of expected number of bits on each link as well as the amount of randomness. For vectors  $X, Y \in \mathbb{G}^n$ , we write X + Y to denote the component-wise computation.

#### Algorithm 1: Secure Computation of GROUP-ADD

**Require:** Alice & Bob have input vectors  $X, Y \in \mathbb{G}^n$ .

**Ensure:** Charlie securely computes the component-wise

$$Z = X + Y$$
.

- 1: Charlie samples n i.i.d. uniformly distributed elements  $K = (K_1, K_2, ..., K_n)$  from  $\mathbb{G}$  using his private randomness; sends it to Bob as  $M_{\vec{32}} = K$ .
- 2: Bob sends  $M_{\vec{21}} = Y + M_{\vec{32}}$  to Alice.
- 3: Alice sends  $M_{\vec{13}} = X + M_{\vec{21}}$  to Charlie.
- 4: Charlie outputs  $Z = M_{\vec{13}} K$ .

**Figure 2** An optimal protocol for secure computation in any group  $\mathbb{G}$ . The protocol requires a  $|\mathbb{G}|$ -ary symbol to be exchanged per computation over each link.

**Theorem 13.** Any secure protocol for computing in a Group  $\mathbb{G}$ , where  $p_{XY}$  has full support over  $\mathbb{G}^n \times \mathbb{G}^n$ , must satisfy

$$H(M_{12}), H(M_{23}), H(M_{31}) \ge n \log |\mathbb{G}|,$$
  
 $\rho(\text{GROUP-ADD}) \ge n \log |\mathbb{G}|.$ 

*Proof.* It is easy to see that the above function satisfies Condition 1 and Condition 2 of Lemma 3. We will only need the last terms (corresponding to the naïve bounds H(X', Y'|Z') etc., but with distribution switching) of (7), (8) and (9) for  $H(M_{12})$ ,  $H(M_{31})$  and  $H(M_{23})$  respectively. Since we are computing a deterministic function, and Y can be determined from (X, Z), the last terms in each of the these bounds will reduce to the following:

$$H(M_{12}) \ge \sup_{p_{X'Y'}} H(X'|Z'),$$
  

$$H(M_{31}) \ge \sup_{p_{X'Y'}} H(X'|Y'),$$
  

$$H(M_{23}) \ge \sup_{p_{X'Y'}} H(Y'|X').$$

The optimum bounds for  $M_{12}$ ,  $M_{31}$  and  $M_{23}$  are obtained by taking X' and Y' to be independent and uniform over  $\mathbb{G}^n$ , which gives  $H(M_{12}), H(M_{31}), H(M_{23}) \geq n \log |\mathbb{G}|$ .

From Theorem 6 and the above bound on  $H(M_{12})$ , we have  $\rho(\text{GROUP-ADD}) \geq n \log |\mathbb{G}|$ , which implies that the above protocol is randomness-optimal.

#### F.2 Secure Computation of SUM

The SUM function is defined as follows: Alice and Bob have one bit input  $X \in \{0,1\}$  and  $Y \in \{0,1\}$  respectively. Charlie wants to compute the arithmetic sum Z = f(X,Y) = X+Y. Figure 3 recapitulates a simple protocol for this function. This protocol requires a ternary symbol to be exchanged per computation over each link. We show in below that our bounds give  $H(M_{31}), H(M_{23}) \ge \log(3)$  and  $H(M_{12}) \ge 1.5$ . Thus, while the protocol matches the lower bound on  $H(M_{31})$  and  $H(M_{23})$ , there is a gap for  $H(M_{12})$ . While the protocol requires  $H(M_{12}) = \log(3)$ , the lower bound is only  $H(M_{12}) \ge 1.5$ . We also show that this protocol is randomness-optimal, which proves a recent conjecture of [LA14] for three users.

For vectors  $U, V \in \{0, 1, 2\}^n$ , we write U + V to denote the component-wise addition modulo-3.

#### **Algorithm 4:** Secure Computation of SUM

**Require:** Alice and Bob have input vectors  $X, Y \in \{0, 1\}^n$ .

**Ensure:** Charlie securely computes the component-wise SUM Z = X + Y.

- 1: Charlie samples n i.i.d. uniformly distributed elements  $K = (K_1, K_2, ..., K_n)$  from  $\{0, 1, 2\}$  using his private randomness; sends it to Alice as  $M_{\vec{i}\vec{1}} = K$ .
- 2: Alice sends  $M_{1\overline{2}} = M_{3\overline{1}} + X$  to Bob.
- 3: Bob sends  $M_{\vec{23}} = M_{\vec{12}} + Y$  to Charlie.
- 4: Charlie outputs  $Z = M_{\vec{23}} K$ .

Figure 3 A protocol to compute SUM. The protocol requires a ternary symbol to be exchanged over all the three links per computation. We show a lower bound of log(3) both on Alice-Charlie and Bob-Charlie links and a lower bound of 1.5 on Alice-Bob link.

**Theorem 14.** Any secure protocol for computing SUM, where  $p_{XY}$  has full support over  $\{0,1\}^n \times \{0,1\}^n$  must satisfy

$$H(M_{31}), H(M_{23}) \ge n \log(3)$$
 and  $H(M_{12}) \ge 1.5n$ ,  
 $\rho(\text{SUM}) \ge n \log(3)$ .

*Proof.* It is easy to see that SUM satisfies Condition 1 and Condition 2 of Lemma 3. Also, RI(Y; Z) = I(Y; Z) and RI(Z; X) = I(Z; X). It turns out that for  $H(M_{31})$  and  $H(M_{23})$ , the bounds in (8) and (9) are better than (19) and (20) respectively. Since X can be determined from (Y, Z) and Y can be determined from (X, Z), we can simplify the bounds in (8) and (9) to the following:

$$H(M_{31}) \ge \sup_{p_{X'Y'}} (H(Z')),$$
  
$$H(M_{23}) \ge \sup_{p_{X'Y'}} (H(Z')).$$

For  $H(M_{31})$ , taking  $p_{X'Y'}(0,0) = p_{X'Y'}(1,1) = 1/3$  and  $p_{X'Y'}(0,1) = p_{X'Y'}(1,0) = 1/6$  gives  $H(M_{31})$ ,  $H(M_{23}) \ge n \log(3)$ . For  $H(M_{12})$ , the bound in (18) is better than (7) and (18) simplifies to

$$H(M_{12}) \ge \sup_{p_{X'}} \left\{ \sup_{p_{Y'}} I(Y'; Z') + \sup_{p_{Y''}} \left\{ I(X'; Z'') + H(X', Y''|Z'') \right\} \right\}.$$

The second term simplifies to H(X'). Taking  $X', Y' \sim \text{Bern}(1/2)$  gives  $H(M_{12}) \geq 1.5n$ .

Since SUM satisfies condition 1 of Lemma 3. So, from Theorem 6, we have  $\rho(SUM) \ge H(M_{31})$ , which from the above calculation is lower bounded by  $n \log(3)$ , implying the randomness-optimality of the above protocol.

#### F.3 Secure Computation of Controlled Erasure

The controlled erasure function from [DP13] is shown below. Alice's input X acts as the "control" which decides whether Charlie receives an erasure  $(\Delta)$  or Bob's input Y.

$$\begin{array}{c|cc} & & & y \\ x & \hline 0 & 1 \\ \hline 0 & \Delta & \Delta \\ 1 & 0 & 1 \\ \end{array}$$

Notice that Charlie always find out Alice's control bit, but does not learn Bob's bit when it is erased. This function does not satisfy Condition 1 of Lemma 3.

Figure 4 gives a protocol (repeated from [DP13]) for securely computing this function on each location of strings of length n. Bob sends his input string to Charlie under the cover of a one-time pad and reveals the key used to Alice. Alice sends his input to Charlie compressed using a Huffman code (replaced by Lempel-Ziv if we want the protocol to be distribution independent). He also sends to Charlie those key bits he received corresponding to the locations where there is no erasure (i.e., where his input bit is 1). When  $X \sim \text{Bernoulli}(p)$  and  $Y \sim \text{Bernoulli}(q)$ , i.i.d., where  $p, q \in (0, 1)$ , the expected message length for Alice-Charlie link is  $\mathbb{E}[L_{31}] < nH_2(p) + 1 + np$ , the messages lengths on the other two links are determinisitically n each,  $L_{12} = L_{23} = n$ . Here we prove the optimality of this protocol for  $X \sim \text{Bernoulli}(p)$  and  $Y \sim \text{Bernoulli}(q)$ , where  $p, q \in (0, 1)$ ; [DP13] only considered the case where  $X, Y \sim \text{Bernoulli}(1/2)$ . We also prove that this protocol is randomness-optimal.

#### Algorithm 2: Secure Computation of Controlled Erasure

**Require:** Alice & Bob have input bits  $X^n, Y^n \in \{0, 1\}^n$ .

Ensure: Charlie securely computes the Controlled erasure function

$$Z_i = f(X_i, Y_i), \qquad i = 1, \dots, n.$$

- 1: Bob samples n i.i.d. uniformly distributed bits  $K^n$  from his private randomness; sends it to Alice as  $M_{2\vec{1},1}=K^n$ . Bob sends to Charlie his input  $Y^n$  masked (bit-wise) with  $K^n$  as  $M_{2\vec{3},1}=Y^n\oplus K^n$ .
- 2: Alice sends his input  $X^n$  to Charlie compressed using a Huffman code (or Lempel-Živ if we want the protocol to not depend on the input distribution of  $X^n$ ); let  $c(X^n)$  be the codeword. Alice also sends to Charlie the sequence of key bits  $K_i$  corresponding to the locations where his input  $X_i$  is 1.

$$M_{\vec{12}.2} = c(X^n), (K_i)_{i:X_i=1}.$$

3: Charlie outputs

$$Z_i = \begin{cases} \Delta, & \text{if } X_i = 0\\ (Y_i \oplus K_i) \oplus K_i, & \text{if } X_i = 1. \end{cases}$$

Figure 4 A protocol to compute CONTROLLED ERASURE function. For  $X \sim \text{Bernoulli}(p)$  and  $Y \sim \text{Bernoulli}(q)$ , both i.i.d and  $p, q \in (0, 1)$ , the expected message lengths per bit are  $\mathbb{E}[L_{31}] < n(H_2(p) + p) + 1$ ,  $L_{12} = n$ , and  $L_{23} = n$ . We show that these are asymptotically optimal by showing the following lower bounds:  $H(M_{31}) \geq n(H_2(p) + p)$ ,  $H(M_{12}) \geq n$  and  $H(M_{23}) \geq n$ .

**Theorem 15.** Any secure protocol for computing CONTROLLED ERASURE for  $X \sim Bernoulli(p)$  and  $Y \sim Bernoulli(q)$ , both i.i.d., with  $p, q \in (0, 1)$  over block length n must satisfy

$$H(M_{31}) \ge n(H_2(p) + p), \quad H(M_{12}) \ge n, \quad and \quad H(M_{23}) \ge n,$$
  
 $\rho(\text{CONTROLLED-ERASURE}) \ge n.$ 

*Proof.* It is easy to see that this function satisfies only Condition 2 of Lemma 3. We also have RI(X;Z) = 0 and RI(Y;Z) = I(Y;Z) for this function. Since Condition 1 of Lemma 3 is not satisfied, our best bound for  $H(M_{31})$  is given by (17). Since X is independent of Y" in (17) and we are computing a deterministic function, the bound in (17) simplifies to the following:

$$H(M_{31}) \ge \sup_{p_{Y'}} \{I(Y'^n; Z'^n)\} + H(X^n).$$

The optimum bound for  $H(M_{31})$  is obtained by taking  $Y' \sim \text{Bernoulli}(1/2)$ , which gives  $H(M_{31}) \geq n(p + H_2(p))$ . For  $H(M_{23})$ , we can apply the bound in (9), which simplifies to the following:

$$H(M_{23}) \ge \sup_{p_{X'Y'}} \{ H(Y'^n|X'^n) \}.$$

Taking Y' to be independent of X' and Y' ~ Bernoulli(1/2) gives  $H(M_{31}) \ge n(p+H_2(p))$ . For  $H(M_{12})$ , we can apply the bound in (7), which simplifies to the following:

$$H(M_{12}) \ge \sup_{p_{X'Y'}} \left( I(Y'; Z') + H(X', Y'|Z') \right)$$

$$= \sup_{p_{X'Y'}} \left( -H(Z'|Y') + H(X', Y') \right)$$

$$\stackrel{\text{(a)}}{=} \sup_{p_{X'Y'}} \left( -H(X'|Y') + H(X', Y') \right)$$

$$= \sup_{p_{X'Y'}} \left( H(X') \right),$$

where (a) follows because, we can determine X from Z and Z is a deterministic function of (X, Y). Now, taking  $X' \sim \text{Bernoulli}(1/2)$  gives  $H(M_{12}) \geq n$ .

From Theorem 6 and the above bound on  $H(M_{12})$ , we have  $\rho(\text{CONTROLLED-ERASURE}) \geq n$ , which implies that the above protocol is randomness-optimal.

# F.4 Secure Computation of REMOTE $\binom{m}{1}$ -OT $\binom{n}{2}$

**Algorithm 3:** Secure Computation of REMOTE  $\binom{m}{1}$ -OT<sub>2</sub>

**Require:** Alice has m input bit strings  $X_0, X_1, \ldots, X_{n-1}$  each of length n & Bob has an input  $Y \in \{0, 1, \ldots, m-1\}$ .

**Ensure:** Charlie securely computes the REMOTE  $\binom{m}{1}$ -OT<sub>2</sub><sup>n</sup>:  $Z = X_Y$ .

- 1: Alice samples  $nm + \log m$  indep., uniformly distributed bits from her private randomness. Denote the first m blocks each of length n of this random string by  $K_0, K_1, \ldots, K_{m-1}$  and the last  $\log m$  bits by  $\pi$ . Alice sends it to Bob as  $M_{12.1} = (K_0, K_1, \ldots, K_{m-1}, \pi)$ .
- 2: Alice computes  $M^{(i)} = X_{\pi+i \pmod{m}} \oplus K_{\pi+i \pmod{m}}$ ,  $i \in \{0, 1, ..., m-1\}$  and sends to Charlie  $M_{\vec{13},2} = (M^{(0)}, M^{(1)}, ..., M^{(m-1)})$ . Bob computes  $C = Y \pi \pmod{m}$ ,  $K = K_Y$  and sends to Charlie  $M_{\vec{23},2} = (C, K)$ .
- 3: Charlie outputs  $Z = M^{(C)} \oplus K$ .

Figure 5 A protocol to securely compute REMOTE  $\binom{m}{1}$ -OT $\binom{n}{2}$ , which is a special case of the general protocol given in [FKN94]. The protocol requires nm bits to be exchanged over the Alice-Charlie (31) link,  $n + \log m$  bits over the Bob-Charlie (23) link and  $nm + \log m$  bits over the Alice-Bob (12) link. We show optimality of our protocol by showing that any protocol must exchange an expected nm bits over the Alice-Charlie (31) link,  $n + \log m$  bite over the Bob-Charlie (23) link and  $nm + \log m$  bits over the Alice-Bob (12) link.

Proof of Theorem 7. REMOTE  $\binom{m}{1}$ -OT<sub>2</sub><sup>n</sup> satisfies Condition 1 and Condition 2 of Lemma 3. We also have, RI(Y;Z) = I(Y;Z) and RI(Z;X) = I(Z;X). It turns out that for  $H(M_{31})$  and  $H(M_{23})$ , (8) and (9) give the same bounds as (19) and (20) respectively. We will consider the bounds in (19) and (20) in the following. Since X' is independent of Y'' in (19), X'' is independent of Y' in (20) and we are computing a deterministic function, the bounds in (19) and (20) simplify to the following:

$$H(M_{31}) \ge \sup_{p_{X'}} \left\{ \left( \sup_{p_{Y'}} I(Y'; Z') \right) + H(X') \right\},$$
  
$$H(M_{23}) \ge \sup_{p_{Y'}} \left\{ \left( \sup_{p_{X'}} I(X'; Z') \right) + H(Y') \right\}.$$

Taking X' and Y' to be uniform, we get  $H(M_{31}) \ge nm$ . To derive a lower bound on  $H(M_{23})$ , take  $Y' \sim \text{unif}\{0,1\}$  and X' distributed as below

$$p_{X'_0, X'_1, \dots, X'_{m-1}}(x_0, x_1, \dots, x_{m-1}) = \begin{cases} \frac{1}{2^n} - \epsilon, & x_0 = x_1 = \dots = x_{m-1} \\ \epsilon/(2^{n(m-1)} - 1), & \text{otherwise,} \end{cases}$$

where  $\epsilon > 0$  can be made arbitrarily small to make I(Z'; X') as close to n as desired. This gives a bound of  $H(M_{23}) \ge n + \log m$ . For  $H(M_{12})$ , the bottom row of (18) simplifies to

$$H(M_{12}) \ge \sup_{p_{X'}} \left\{ \sup_{p_{X'}} I(X'; Z') + \sup_{p_{X''}} \left\{ I(Y'; Z'') + H(X'', Y'|Z'') \right\} \right\}.$$

Taking Y' and X'' to be uniform and X' to be as below

$$p_{X_0',X_1',\dots,X_{m-1}'}(x_0,x_1,\dots,x_{m-1}) = \begin{cases} \frac{1}{2^n} - \epsilon, & x_0 = x_1 = \dots = x_{m-1} \\ \epsilon/(2^{n(m-1)} - 1), & \text{otherwise,} \end{cases}$$

where  $\epsilon > 0$  can be made arbitrarily small to make I(X'; Z') as close to n as desired. This gives a bound of  $H(M_{12}) \ge nm + \log m$ .

From Theorem 6 and the above bound on  $H(M_{12})$ , we have  $\rho(\text{REMOTE-OT}) \geq nm + \log m$ , which implies that the above protocol is randomness-optimal.

#### F.5 Secure Computation of AND

Algorithm 5: Secure Computation of AND

Require: Alice has an input bit X & Bob has a bit Y.

**Ensure:** Charlie securely computes the AND  $Z = X \wedge Y$ .

1: Alice samples a uniform random permutation  $(\alpha, \beta, \gamma)$  of (0, 1, 2) from her private randomness; sends it to Bob  $M_{\vec{12}} = (\alpha, \beta, \gamma)$  (using a symbol from an alphabet of size 6).

2: Alice sends  $\alpha$  to Charlie if X=1, and  $\beta$  if X=0. Bob sends  $\alpha$  to Charlie if Y=1, and  $\gamma$  if Y=1.

$$M_{31} = \begin{cases} \alpha & \text{if } X = 1 \\ \beta & \text{if } X = 0 \end{cases}$$
  $M_{23} = \begin{cases} \alpha & \text{if } Y = 1 \\ \gamma & \text{if } Y = 0 \end{cases}$ 

3: Charlie outputs Z = 1 if  $M_{31} = M_{23}$ , and 0 otherwise.

Figure 6 A protocol to compute AND [FKN94]. The protocol requires a ternary symbol to be exchanged over the Alice-Charlie (31) and Bob-Charlie (23) links and symbols from an alphabet of size 6 over the Alice-Bob (12) link per AND computation.

Proof of Theorem 8. We will prove the result only for n = 1, i.e., when input consists of only one bit. The result for general n follows by taking in the following proof  $X_i$ 's,  $Y_i$ 's and  $Y_i$ "s to be i.i.d.

It is easy to see that AND satisfies Condition 1 and Condition 2 of Lemma 3. Also, RI(Y;Z) = I(Y;Z) and RI(Z;X) = I(Z;X). It turns out that for  $H(M_{31})$  and  $H(M_{23})$ , the bounds in (8) and (9) are better than (19) and (20) respectively. The simplified bounds in (8) and (9) are as follows:

$$H(M_{31}) \ge \sup_{p_{X'Y'}} \left( I(Y'; Z') + H(X', Z'|Y') \right),$$
  
$$H(M_{23}) \ge \sup_{p_{X'Y'}} \left( I(X'; Z') + H(Y', Z'|X') \right).$$

For  $H(M_{31})$ , take  $p_{X'Y'}(0,0) = p_{X'Y'}(1,0) = p_{X'Y'}(1,1) = (1-\epsilon)/3$  and  $p_{X'Y'}(0,1) = \epsilon$ , where  $\epsilon > 0$  can be made arbitrarily small to make  $H(M_{31})$  as close to log(3) as we desire.

For  $H(M_{23})$ , take  $p_{X'Y'}(0,0) = p_{X'Y'}(0,1) = p_{X'Y'}(1,1) = (1-\epsilon)/3$  and  $p_{X'Y'}(1,0) = \epsilon$ , where  $\epsilon > 0$  can be made arbitrarily small to make  $H(M_{23})$  as close to  $\log(3)$  as we desire.

For  $H(M_{12})$ , (18) simplifies to

$$H(M_{12}) \ge \sup_{p_{X'}} \left\{ \sup_{p_{Y'}} I(Y'; Z') + \sup_{p_{Y''}} \left\{ I(X'; Z'') + H(X', Y''|Z'') \right\} \right\}.$$

The second term simplifies to  $H(X') + p_{X'}(0)$  by taking Y'' to be uniform. Taking  $p_{X'}(1) = 0.456$  and  $p_{Y'}(1) = 0.397$  gives  $H(M_{12}) \ge 1.826$ .

From Theorem 6 and the above bound on  $H(M_{12})$ , we have  $\rho(AND) \geq n(1.826)$ , whereas the protocol requires  $1 + \log 3$  random bits.

Note: We need the use of Lemma 4 (information inequality) only to improve the bound on  $H(M_{12})$  in REMOTE-OT, SUM and AND. All other bounds in all other functions do not need the use information inequality.

# G Lower Bounds for Dependent Inputs

We will show that all our lower bounds proven for independent inputs hold for dependent inputs as well provided the distribution has full support. In Subsection 3.2, we observed that any secure protocol  $\Pi(p_{XY}, p_{Z|XY})$ , where distribution  $p_{XY}$  has full support, continues to be a secure protocol even if we switch the input distribution to a different one  $p_{\widetilde{X}\widetilde{Y}}$ .

Since we can switch to any distribution  $p_{\widetilde{X}\widetilde{Y}}$ , in particular, we can switch to  $p_{\widetilde{X}\widetilde{Y}}$ , where  $\widetilde{X}$  and  $\widetilde{Y}$  have the same marginals as X and Y respectively, i.e.,  $p_{\widetilde{X}}(x) = p_X(x), \forall x \in \mathcal{X}$  and  $p_{\widetilde{Y}}(y) = p_Y(y), \forall y \in \mathcal{Y}$ . This allows us to argue that the communication lower bounds for  $\Pi(p_{\widetilde{X}\widetilde{Y}}, p_{Z|XY})$  also hold for  $\Pi(p_{XY}, p_{Z|XY})$ . To prove this, we show below that the resulting marginal distributions on the transcripts remain the same as the original ones, implying the same entropies.

Let denote the resulting distribution on the transcript on 12 link by  $M_{12}$  and similarly on the other two links.

$$\begin{split} p_{\widetilde{M}_{12}}(m_{12}) &= \sum_{x,y} p_{\widetilde{M}_{12}|\widetilde{X}\widetilde{Y}}(m_{12}|x,y) p_{\widetilde{X}\widetilde{Y}}(x,y) \\ &\stackrel{\text{(a)}}{=} \sum_{x,y} p_{M_{12}|XY}(m_{12}|x,y) p_{\widetilde{X}\widetilde{Y}}(x,y) \\ &\stackrel{\text{(b)}}{=} \sum_{x,y} p_{M_{12}|X}(m_{12}|x) p_{\widetilde{X}\widetilde{Y}}(x,y) \\ &= \sum_{x} p_{M_{12}|X}(m_{12}|x) \sum_{y} p_{\widetilde{X}\widetilde{Y}}(x,y) \\ &= \sum_{x} p_{M_{12}|X}(m_{12}|x) p_{\widetilde{X}}(x) \\ &= \sum_{x} p_{M_{12}|X}(m_{12}|x) p_{X}(x) \\ &= p_{M_{12}}(m_{12}), \end{split}$$

where (a) follows from the fact that in a secure computation protocol, once Alice and Bob are given inputs X=x and Y=y respectively, the protocol produces  $(m_{12},m_{23},m_{31},z)$  according to the conditional distribution  $p_{M_{12}M_{23}M_{31}Z|XY}(m_{12},m_{23},m_{31},z|x,y)$  and this conditional distribution does not depend on the distribution  $p_{XY}$ , hence,  $p_{\widetilde{M}_{12}|\widetilde{X}\widetilde{Y}}(m_{12}|x,y)=p_{M_{12}|XY}(m_{12}|x,y)$ ; and (b) follows from privacy against Alice. This implies that  $H(\widetilde{M}_{12})=H(M_{12})$ . Similarly we can prove  $H(\widetilde{M}_{23})=H(M_{23})$  and  $H(\widetilde{M}_{31})=H(M_{31})$ .

Proofs of our lower bounds for secure computation in Theorem 4 and Theorem 5 assumed independent inputs. For them to hold for dependent ones, we can take  $p_{\widetilde{X}\widetilde{Y}}$  in above to be a product distribution  $p_{\widetilde{X}\widetilde{Y}} = p_{\widetilde{X}}p_{\widetilde{Y}}$  with  $\widetilde{X}$  and  $\widetilde{Y}$  having the same marginals as X and Y.

# H Dependence on Input Distributions

Our communication lower bounds were developed for protocols whose designs may take into account the distributions of X and Y. Specifically, the right hand sides of (7) and (18) do not depend on the distributions  $p_X p_Y$  of the inputs. Thus, even though we allow the protocol to depend on the distributions, our lower bound on  $H(M_{12})$  does not. The same is true for (8) and (19) for  $H(M_{31})$  and (9) and (20) for  $H(M_{23})$ , which apply when the function satsifies certain conditions. As the CONTROLLED-ERASURE example (Appendix F.3) demonstrates, when these conditions are not satisfied, the communication complexity of the optimal protocol may indeed depend on the distribution of the input. Notice that the specific protocols we have given in this paper do not need the knowledge of the input distributions.