Revealing Additional Information in Two-Party Computations*

Andreas Jakoby and Maciej Liśkiewicz **

Institut für Theoretische Informatik, Universität zu Lübeck, Germany
liskiewi/jakoby@tcs.uni-luebeck.de

Abstract. A two-argument function is computed privately by two parties if after the computation, no party should know anything about the other inputs except for what he is able to deduce from his own input and the function value. In [1] Bar-Yehuda, Chor, Kushilevitz, and Orlitsky give a complete characterisation of two-argument functions which can be computed privately (in the information-theoretical sense) in the Honest-But-Curious model and study protocols for "non-private" functions revealing as little information about the inputs as possible. The authors define a measure which determines for any function f the additional information $\mathcal{E}(f)$ required for computing f and claim that f is privately-computable if and only if $\mathcal{E}(f) = 0$. In our paper we show that the characterisation is false: we give a privately-computable function f with $\mathcal{E}(f) \neq 0$ and another function g with $\mathcal{E}(g) = 0$ that is *not* privately-computable. Moreover, we show some rather unexpected and strange properties of the measure for additional information given by Bar-Yehuda et al. and we introduce an alternative measure. We show that for this new measure the minimal leakage of information of randomized and deterministic protocols are equal. Finally, we present some general relations between the information gain of an optimal protocol and the communication complexity of a function.

1 Introduction

We investigate computations of functions of two n-bit inputs x and y by two players Alice holding x and Bob having y. For a given function f Alice (A) and Bob (B), both with unlimited computational power, communicate to determine $f(x, y)$ keeping as much of its input secret from the other party as possible. In this setting two models are considered in the literature. In the first one we assume that the players are honest but curious, that means they never deviate from the given protocol but try to acquire knowledge about the input bits of the other player only by observing the communication. In the second setting Alice or Bob can be malicious, i.e. they can cheat. In this paper we study privacy in the Honest-But-Curious setting.

* Supported by DFG research grant RE 672/5-1.

** On leave from Instytut Informatyki, Uniwersytet Wrocławski, Poland.

B. Roy (Ed.): ASIACRYPT 2005, LNCS 3788, pp. 121–135, 2005.

Private computation was introduced by Yao [8]. He considered the problem under cryptographic assumptions. Private computation in the information-theoretical secure setting has been introduced by Ben-Or et al. [3] and Chaum et al. [5]. Ben-Or et al. have presented a function that is not privately computable. A complete characterisation of such functions has been given independently by Kushilevitz [6] and Beaver [2]. This characterisation has been given by using so called forbidden submatrices. Let M be a matrix. We say that two row indices i and j are related ($i \sim j$) if there is a column k for which $M_{i,k} = M_{j,k}$. For example, the row indices of matrix T shown below are related while the rows of matrix T' are not related.

$$T = \begin{bmatrix} 0\,0 \\ 0\,1 \end{bmatrix}, \quad T' = \begin{bmatrix} 1\,0 \\ 0\,1 \end{bmatrix}. \tag{1}$$

We define the equivalence relation $\equiv$ to be the transitive closure of $\sim$. In a similar way, we define the relations $\sim$ and $\equiv$ on the columns of M. A matrix is forbidden if it is not monochromatic (i.e. not all elements of the matrix are the same), all its rows are equivalent with respect to $\equiv$ on rows, and all its columns are equivalent with respect to $\equiv$ on columns. Matrix T defined in (1) is a small example of a forbidden matrix and T' is an example for a not forbidden matrix. Privately-computable functions can be characterised as follows. Let M_f denote the communication matrix for the function f, i.e. an $2^n \times 2^n$ matrix such that rows and columns are indexed by n-bit inputs and for every $x, y \in \{0,1\}^n$ we have $(M_f)_{x,y} = f(x,y)$. For example T and T' in (1) are communication matrices of the two argument Boolean functions `AND` and `XOR`, respectively.

Theorem 1 ([6,2]). *In the Honest-But-Curious model a two-argument function f can be computed privately if and only if M_f does not contain any forbidden submatrix.*

Using this characterisation one can see that the majority of functions cannot be computed privately. For such functions it is natural to study the minimum amount of information about the individual inputs that must leak during their computation. There are several ways to quantify such a leakage. In [1] Bar-Yehuda et al. introduced three measures: a combinatorial measure $\mathcal{I}_{\mathrm{c}}$, an information-theoretic measure $\mathcal{I}_{\mathrm{i}}$, and a measure $\mathcal{I}_{\mathrm{c\text{-}i}}$ that includes both combinatorial and information-theoretic aspects. For the measures they proved general tight bounds on minimum amount of information about the inputs that must be revealed in a computation. Moreover, they showed that sacrificing some privacy can reduce the number of messages required during the computation.

In [1] the authors define for any function f the *additional* information $\mathcal{E}(f)$ required for computing f as a difference between $\mathcal{I}_{\mathrm{c}}(f)$ and $\log_2 |\operatorname{range}(f)|$, where $|\operatorname{range}(f)|$ denotes the cardinality of the range of function f. They claim that f is privately-computable if and only if $\mathcal{E}(f) = 0$. In our paper we show that the characterisation is false. We construct a privately-computable function f with $\mathcal{E}(f) \neq 0$. Moreover we show that for the function $f_{min}(x,y) = \min\{x,y\}$, where x and y are interpreted as integers from $\{0,1,\ldots,2^n-1\}$, it holds that

$\mathcal{E}(f_{min}) = 0$. On the other hand, f_{min} cannot be computed privately since the communication matrix of f_{min}:

$$M_{f_{min}} = \begin{bmatrix} 0\,0\,0\,0\ldots & 0 \\ 0\,1\,1\,1\ldots & 1 \\ 0\,1\,2\,2\ldots & 2 \\ \ldots & \\ 0\,1\,2\,3\ldots & 2^n-1 \end{bmatrix}$$

contains a forbidden submatrix. In fact, $M_{f_{min}}$ is not monochromatic and for every $x < 2^n - 1$ we have $f_{min}(x, x) = f_{min}(x + 1, x) = f_{min}(x, x + 1) = x$ and $f_{min}(x + 1, x + 1) = x + 1$ what implies that all its rows (columns, resp.) are in the same equivalence class.

We show also some rather strange properties of the measures for revealed information $\mathcal{I}_{\mathrm{c}}$, $\mathcal{I}_{\mathrm{i}}$, and $\mathcal{I}_{\mathrm{c\text{-}i}}$. For example, we show that $\mathcal{I}_{\mathrm{c}}(\texttt{AND}) = \mathcal{I}_{\mathrm{c}}(\texttt{XOR})$: the revealed information required for computing `AND` is the same as for `XOR` contradictory to the fact that `XOR` can be computed privately but `AND` cannot. The similar property holds for the remaining measures as well.

Furthermore, we introduce an alternative measure for the minimum revealed information, which is based on the information source defined in [4]. The revealed information of a protocol to a player is merely the logarithm of the number of different probability distributions on the communication strings a player can observe. For this measure we will show that f is a privately computable function if and only if the amount of the minimum revealed information is zero. We give some tight bounds of concrete functions and show a general lower bound for arbitrary two n-bit inputs functions.

We show that for our measure the minimal leakage of information for randomized and deterministic protocols are equal. Finally, we present some relations between the information gain of an optimal protocol and the communication complexity of a function. More precisely, we will give a lower bound for the leakage of information that is logarithmic on the communication complexity. We will show that for some specific functions this general bound is tight.

The paper is organized as follows. In the next section we give some preliminaries for communication complexity. In Section 3 we present the model of Bar-Yehuda et al. and we give there our analysis of their results. In Section 4 we discuss our measure for reviling additional information. The relation of the gain of additional information in randomized protocols and deterministic protocols is investigated in Section 5. Finally, in Section 6 we give a general relation between communication complexity and the additional information.

2 Communication Protocols

Let f be a function of two n-bit inputs x and y that are known to two parties A and B, respectively, each having unlimited computing power. The aim is to determine $f(x, y)$ by alternate transmitting messages over a noiseless binary channel according to a communication protocol. We consider two kinds

of protocols: deterministic and randomised. In deterministic case each message is determined by the input known to the party and by the previously received messages. We require that in every round of communication, the set of all possible messages is prefix-free. A protocol computes f if for every (x, y) each party deduces correctly the value $f(x, y)$. Let $\mathcal{P}(x, y)$ denote the concatenation of all communication messages of a protocol $\mathcal{P}$ exchanged between A and B during the computation on an input (x, y). Let communication complexity of protocol $\mathcal{P}$, denoted by $C_{\mathcal{P}}$, be the maximum length of $\mathcal{P}(x, y)$, and let the communication size $CS_{\mathcal{P}}$ be the number of different strings $\mathcal{P}(x, y)$, over all inputs (x, y). Define the deterministic communication complexity of f, denoted by $C_D(f)$, as the smallest $C_{\mathcal{P}}$ over all deterministic protocols $\mathcal{P}$ computing f and analogously let the communication size $CS_D(f)$ be the smallest $CS_D(f)$ over all $\mathcal{P}$.

For the randomised protocol $\mathcal{P}$ on an input (x, y), to determine communication messages A and B can use additionally random bit strings. In this paper we consider randomised protocols where each party A and B has access to a *private* random strings R_A and R_B, respectively. In this case the communication string $\mathcal{P}(x, y)$, defined again as the concatenation of all messages transmitted during an execution of $\mathcal{P}$ on (x, y), is a random string.

For a general survey of communication complexity see e.g. Kushilevitz and Nisan [7].

3 Additional Information - The Model of Bar-Yehuda et al.

In this section we will discuss the measuring of additional information defined in [1]. First we give the definitions and the results of [1] and we show next that some of the results are false, the measures are somehow inconsistent, and they have rather unexpected and strange properties.

3.1 The Results

Let us first present the definition of privacy cost in the combinatorial setting. Next the information-theoretic measure and the measure that includes both combinatorial and information-theoretic aspects will be considered.

To define the combinatorial measure $\mathcal{I}_{\mathrm{c}}(f)$ for a function f Bar-Yehuda et al. introduce a weak and a stronger definition of privacy cost. However, since the notions are equivalent to each other, we will recall the definition of $\mathcal{I}_{\mathrm{c}}$ using the notion of strong privacy only. To measure information leakage during computation of f we use an auxiliary function h, which like f, is a function of two n-bit strings. The ranges of both functions can be different. Intuitively speaking, a protocol $\mathcal{P}$ for f leaks at most h, or equivalently is h-private, if during the computation of $\mathcal{P}$ on (x, y) the information learned by a party about the input of the other party can be deduced from its own input and the value $h(x, y)$.

Definition 1 ([1]). *A protocol $\mathcal{P}$ for f is strongly h-private for A if*

1. *for every $x, y \in \{0,1\}^n$ $\mathcal{P}$ computes the value $f(x,y)$ correctly with probability 1 and*
2. *for every $x, y_1, y_2 \in \{0,1\}^n$, $h(x,y_1) = h(x,y_2)$ implies that for all random choices r of A, $\mathcal{P}(x,y_1)$ and $\mathcal{P}(x,y_2)$ have the same distribution, namely, for every communication string s,*

$$\Pr[s = \mathcal{P}(x,y_1)|r] \;=\; \Pr[s = \mathcal{P}(x,y_2)|r],$$

where the probability is taken over the random choices of B.

Strong h-privacy for B is defined analogously. To give more intuition let us consider the Boolean function f_{equ} defined on two n-bit strings:

$$f_{equ}(x,y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Furthermore, let us consider the (deterministic) protocol of [1] for computing f_{equ} on two n-bit strings $x = x_1 x_2 \ldots x_n$ and $y = y_1 y_2 \ldots y_n$:

Protocol 1. For all $i = 1, 2, \ldots n$ do:
1. A sends x_i to B;
2. If $x_i \neq y_i$ then B transmits 0 and exit; else if $x_i = y_i$ then B transmits 1.

The protocol is strongly h_{equ}-private for both A and B, where h_{equ} is defined as follows: $h_{equ}(x,y) = \min\{i : x_i \neq y_i\}$ if $x \neq y$ and $h_{equ}(x,y) = n+1$ otherwise. To see this, note that for the protocol $\mathcal{P}$ above and for every input (x,y) and (x,y') it holds that $\mathcal{P}(x,y) = \mathcal{P}(x,y')$ if and only if $h_{equ}(x,y) = h_{equ}(x,y')$. An analogous equivalence holds for every (x,y) and (x',y). Recall that $\mathcal{P}(x,y)$ for the deterministic protocol $\mathcal{P}$ denotes just the concatenation of all communication messages sent between A and B during the computation of $\mathcal{P}$ on (x,y).

Definition 2 ([1]). *Let h_1 and h_2 be functions of two n-bit inputs. A protocol $\mathcal{P}$ is strongly $(h_1;h_2)$-private if it is strongly h_1-private for A and strongly h_2-private for B. A protocol $\mathcal{P}$ is strongly h-private if it is strongly (h,h)-private. A function f is strongly h-private if it has a strongly h-private protocol.*

For example, f_{equ} is strongly h_{equ}-private. The revealed information $\mathcal{I}_c(f)$ and the additional information $\mathcal{E}(f)$ required for computing f are defined by

$$\begin{aligned} \mathcal{I}_c(f) &= \min\{\log_2 |\operatorname{range}(h)| : f \text{ is strongly } h\text{-private } \} \\ \mathcal{E}(f) &= \mathcal{I}_c(f) - \log_2 |\operatorname{range}(f)|. \end{aligned}$$

Hence, for the the function f_{equ} we have:

$$\mathcal{I}_c(f_{equ}) \leq \log_2(n+1) \quad \text{and} \quad \mathcal{E}(f) \leq \log_2(n+1) - 2. \tag{3}$$

In [1] Bar-Yehuda et al. observe the following claim which is false as we will see in the next section.

Claim 1 ([1], p. 1932). *A function f is privately-computable if and only if $\mathcal{I}_{\mathrm{c}}(f) = \log_2 |\operatorname{range}(f)|$, i.e., if and only if $\mathcal{E}(f) = 0$.*

For the min function:

$$f_{min}(x,y) = \begin{cases} x & \text{if } x \le y, \\ y & \text{otherwise,} \end{cases} \tag{4}$$

where x and y are interpreted as integers from $\{0, 1, \ldots, 2^n - 1\}$, the authors claim that

Claim 2 ([1], p. 1933]). $0 < \mathcal{E}(f_{min}) \le 1$.

This is not true, as we will see in the next section.

Now, we recall the definition of information-theoretic measure $\mathcal{I}_{\mathrm{i}}$ and a measure that includes both combinatorial and information-theoretic aspects $\mathcal{I}_{\text{c-i}}$. In this paper we will discuss only the deterministic counterpart of these measures (denoted by $\mathcal{I}_{\mathrm{i}}^{\mathrm{det}}$ and $\mathcal{I}_{\text{c-i}}^{\mathrm{det}}$) that refer to the leakage of information if the protocols are restricted to deterministic ones.

To define $\mathcal{I}_{\mathrm{i}}^{\mathrm{det}}$ and $\mathcal{I}_{\text{c-i}}^{\mathrm{det}}$ one has implicitly to assume a probability distribution for the input x and y. Let us consider the input strings as a pair (X, Y) of random variables drawn from some specified distribution which is known to both parties. For a deterministic protocol $\mathcal{P}$ define

$$I_{\mathcal{P}}(X,Y) = \max\{I(X; \mathcal{P}(X,Y)|Y), I(Y; \mathcal{P}(X,Y)|X)\}$$

to be the maximum of the information gained by A or B about the input of the other party that can be deduced from the complete communication strings $\mathcal{P}(X,Y)$ and its own input. Here $I(X;Y|Z)$ denotes the conditional mutual information. The information-theoretic measure $\mathcal{I}_{\mathrm{i}}^{\mathrm{det}}$ of additional information is defined as follows

$$I_f^{\mathrm{det}}(X,Y) = \min\{I_{\mathcal{P}}(X,Y) : \ \mathcal{P} \text{ is a deterministic protocol computing } f\}$$
$$\mathcal{I}_{\mathrm{i}}^{\mathrm{det}}(f) = \sup\{I_f^{\mathrm{det}}(X,Y) : \ (X,Y) \text{ is distributed over } \{0,1\}^n \times \{0,1\}^n\}.$$

Finally define the combinatorial-information-theoretic measure $\mathcal{I}_{\text{c-i}}^{\mathrm{det}}$ by

$$I_{\mathcal{P}} = \sup\{I_{\mathcal{P}}(X,Y) : \ (X,Y) \text{ is distributed over } \{0,1\}^n \times \{0,1\}^n\}$$
$$\mathcal{I}_{\text{c-i}}^{\mathrm{det}}(f) = \min\{I_{\mathcal{P}} : \mathcal{P} \text{ is a deterministic protocol computing } f\}.$$

3.2 Mistakes and Inconsistencies

In the following we show that some claims of [1] are false. We start our analysis showing the following useful lemma:

Lemma 1. *For every function f of two n-bit inputs the revealed information required for computing f is bounded by n, i.e. $\mathcal{I}_{\mathrm{c}}(f) \le n$.*

Note that the lemma does not follow from the simple relation between $\mathcal{I}_c$ and deterministic communication complexity that $\mathcal{I}_c(f) \leq C_D(f)$, since $C_D(f)$ can be equal to $n + |\operatorname{range}(f)|$. On the other hand the bound stated in the lemma seems to be quite natural: One party cannot gain more than n bit of information about the input of the other party in the sense of Shannon.

Proof. Let f be a two-argument function f over $\{0,1\}^n \times \{0,1\}^n$ and let $\mathcal{P}$ be an arbitrary protocol which computes f correctly with probability 1. Define the function $g(x,y) = (x+y) \mod 2^n$ considering x and y as integer in $\{0,\ldots,2^n-1\}$. It is easy to verify that $\mathcal{P}$ is strongly g-private. In fact, for every $x_1, x_2, y_1, y_2 \in \{0,1\}^n$ with $x_1 \neq x_2$ and $y_1 \neq y_2$ we have $g(x_1,y_1) \neq g(x_1,y_2)$ and $g(x_1,y_1) \neq g(x_2,y_1)$. Hence, Condition (2) of Definition 1 is fulfilled. Because $|\operatorname{range}(g)| = 2^n$, we get

$$\mathcal{I}_c(f) = \min\{\log_2 |\operatorname{range}(h)| : f \text{ is strongly } h\text{-private } \} \leq \log_2 |\operatorname{range}(g)| \;=\; n.$$

□

As a counterexample of the characterisation given in Claim 1 consider the function $\varphi : \{0,1\}^n \times \{0,1\}^n \;\to\; \{0\} \cup \{0,1\}^n$ defined for any $n \geq 2$:

$$\varphi(x,y) = \begin{cases} y & \text{if } x = 0^n, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Proposition 1. *Function φ can be computed privately but $\mathcal{E}(\varphi) \neq 0$.*

Proof. Note that $0 \neq 0^n$, hence the communication matrix M_φ does not contain a forbidden submatrix: M_φ is not monochromatic and the first row of M_φ is not equivalent with any other row of the matrix. Hence by the characterisation by Kushilevitz and Beaver (Theorem 1) we know that φ can be computed privately. On the other hand according to the definition of the additional information required for computing φ and by Lemma 1 we can conclude that

$$\mathcal{E}(\varphi) \;=\; \mathcal{I}_c(\varphi) - \log_2 |\operatorname{range}(\varphi)| \;\leq\; n - \log_2(2^n+1) \;<\; 0.$$

□

Therefore Claim 1 is false: For privately-computable function φ we have both $\mathcal{I}_c(\varphi) < \log_2 |\operatorname{range}(\varphi)|$ and $\mathcal{E}(\varphi) \neq 0$. This example shows a strange property of the definition of $\mathcal{E}(\varphi)$: *The additional information required for computing a function can be negative.*

Using again Lemma 1 one can show that Claim 2 is false:

Proposition 2. *For the function f_{min} defined in (4) it holds that*

$$\mathcal{I}_c(f_{min}) = n \quad \textit{and} \quad \mathcal{E}(f_{min}) = 0.$$

Proof. By Lemma 1 we get

$$\mathcal{E}(f_{min}) \;=\; \mathcal{I}_{\mathrm{c}}(f_{min}) - \log_2 |\operatorname{range}(f_{min})| \;\leq\; n - \log_2(2^n) \;=\; 0.$$

It is not difficult to show that $\mathcal{E}(f_{min}) \geq 0$. In fact, if $\mathcal{I}_{\mathrm{c}}(f_{min}) < n$ then there exists a function h such that f_{min} is strongly h-private and $\log_2 |\operatorname{range}(h)| < n$. Consider $x = 2^n - 1$, then for any pair $y_1, y_2 \in \{0, 1, \ldots, 2^n - 1\}$ with $y_1 \neq y_2$ we have $f_{min}(x, y_1) \neq f_{min}(x, y_2)$. This implies the inequality $h(x, y_1) \neq h(x, y_2)$, contradicting the assumption that $\log_2 |\operatorname{range}(h)| < n$. □

Note that the communication matrix $M_{f_{min}}$ of f_{min} contains a forbidden submatrix (see a discussion in Section 1). Hence by Theorem 1, f_{min} is not privately-computable. By Propositions 1 and 2 one can conclude

Theorem 2. *There exists a privately-computable function φ, with $\mathcal{E}(\varphi) \neq 0$ and another function f, with $\mathcal{E}(f) = 0$ that is* not *privately-computable.*

Now we will discuss some inconsistencies of the definitions for additional information. We will show that in fact none of these definitions suits well for measuring additional information properly. In Section 4 we will give a new definition for additional information.

For the function φ, defined in (5), let us consider two (deterministic) protocols $\mathcal{P}_1$ and $\mathcal{P}_2$ that compute φ. The protocol $\mathcal{P}_1$ works on x, y as follows: A sends 0 if $x = 0^n$ and 1 otherwise. If B receives 0 then he sends y to A and otherwise B stops the computation. In protocol $\mathcal{P}_2$, A sends 0 if $x = 0^n$ and 1 otherwise and then B sends y to A. Obviously in both cases each party can determine correctly the value of the function at the end of the communication. Note that $\mathcal{P}_1$ is private protocol in a common sense (more precisely 1-private, see e.g. [6] for the definition) while $\mathcal{P}_2$ is not private. We can say even more: Using $\mathcal{P}_2$ A gains full information about the input of B. On the other hand, both $\mathcal{P}_1$ and $\mathcal{P}_2$ are optimal with respect to $\mathcal{I}_{\mathrm{c}}$. To see this, consider the function $g(x, y) = (x + y) \bmod 2^n$ used in the proof of Lemma 1. We get that both $\mathcal{P}_1$ and $\mathcal{P}_2$ are strongly g-private and the optimality follows from the obvious fact that

$$\mathcal{I}_{\mathrm{c}}(\varphi) = n = |\operatorname{range}(g)|.$$

$\mathcal{I}_{\mathrm{i}}^{\mathrm{det}}$ and $\mathcal{I}_{\mathrm{c\text{-}i}}^{\mathrm{det}}$ measure the additional information wrong, as well. According to the definition of $I_{\mathcal{P}}$ we have for both protocols $\mathcal{P}_1, \mathcal{P}_2$

$$\begin{aligned} I_{\mathcal{P}_i} &= \sup_{(X,Y)} I_{\mathcal{P}_i}(X, Y) \\ &= \sup_{(X,Y)} \max\{\, H(X|Y) - H(X|\mathcal{P}_i(X,Y), Y), \\ &\qquad\qquad H(Y|X) - H(Y|\mathcal{P}_i(X,Y), X)\} \\ &= H(Y) \end{aligned}$$

and therefore $I_{\mathcal{P}_1} = I_{\mathcal{P}_2}$. Hence neither $\mathcal{I}_{\mathrm{i}}^{\mathrm{det}}$ nor $\mathcal{I}_{\mathrm{c\text{-}i}}^{\mathrm{det}}$ measures *the additional* information which can be gain by a party during the computation.

Finally, let us consider the two argument functions `AND` and `XOR`. We have:

$$\mathcal{I}_{\mathrm{c\text{-}i}}^{\mathrm{det}}(\texttt{AND}) = \mathcal{I}_{\mathrm{c\text{-}i}}^{\mathrm{det}}(\texttt{XOR}) = 1.$$

But XOR can be computed privately and therefore *no* additional information can be gained during a computation of XOR. On the other hand, AND cannot be computed privately.

4 Additional Information - New Measure

In the following we will present an alternative measure for additional information, that is based on the information source defined in [4].

Definition 3. *Let $\mathcal{P}$ be a protocol for a function f which for every $x, y \in \{0,1\}^n$ computes the value $f(x,y)$ correctly with probability 1. Let $x \in \{0,1\}^n$, $z \in$ range(f) and let r be a random string provided to A. Define the* information source *of A on x, z, and r as the set of different probability distributions on the communication strings A, holding x and r, can observe during all computations of $\mathcal{P}$ that give the result z:*

$$\mathcal{S}_{\mathcal{P},A}(x,z,r) = \{(\mu_{x,y}(s_1), \mu_{x,y}(s_2), \ldots) : \ y \in \{0,1\}^n, f(x,y) = z\}$$

where $\mu_{x,y}(s_k) = \Pr[\mathcal{P}(x,y) = s_k | r]$. Define the size of the information source as

$$s_{\mathcal{P},A}(x,z) = \max_r |\mathcal{S}_{\mathcal{P},A}(x,z,r)|.$$

Analogously we define $\mathcal{S}_{\mathcal{P},B}(y,z,r)$ - the information source *of B on y, z, and r and the size $s_{\mathcal{P},B}(y,z)$.*

If $\mathcal{P}$ is a deterministic protocol then we will omit r in $\mathcal{S}_{\mathcal{P},A}(x,z,r)$ and write just $\mathcal{S}_{\mathcal{P},A}(x,z)$. Now we are ready to define a new *combinatorial* measure for the additional information, analogy of $\mathcal{I}_c$, that we will denote by $\mathcal{J}_c$.

Definition 4. *The additional information of $\mathcal{P}$ revealed to A is defined as*

$$J_{\mathcal{P},A} = \max\{\log_2 s_{\mathcal{P},A}(x,z) : \ x \in \{0,1\}^n, \ z \in \text{range}(f) \ \} \ .$$

Analogously we define $J_{\mathcal{P},B}$. The additional information that can be deduced running a protocol $\mathcal{P}$ is $J_{\mathcal{P}} = \max\{J_{\mathcal{P},A}, J_{\mathcal{P},B}\}$. The additional information required for computing f is

$$\mathcal{J}_c(f) = \min\{J_{\mathcal{P}} : \ \mathcal{P} \text{ is a protocol computing } f\}.$$

We have the following characterisation of privately computable functions:

Theorem 3. *A function f is privately computable if and only if $\mathcal{J}_c(f) = 0$.*

The proof of the theorem is straightforward and we skip it here.

We can redefine the measure $\mathcal{J}_c$ in term of h-privacy used by Bar-Yehuda et al. (see Definition 2).

Definition 5. *Let h be a function of two n-bit inputs and let protocol $\mathcal{P}$ for a function f be strongly h-private. Analogously to Definition 3 and 4 let*

$$s^h_{\mathcal{P},A}(x,z) = |\{h(x,y) : \ y \in \{0,1\}^n, f(x,y) = z\}|$$

and

$$J^h_{\mathcal{P},A} = \max\{\log_2 s^h_{\mathcal{P},A}(x,z) : \ x \in \{0,1\}^n,\ z \in \mathrm{range}(f)\ \}\ .$$

Analogously define $s^h_{\mathcal{P},B}$ *and* $J^h_{\mathcal{P},B}$ *for* B. *Then* $J^h_{\mathcal{P}} = \max\{J^h_{\mathcal{P},A}, J^h_{\mathcal{P},B}\}$.

Theorem 4. *For every function* f *it holds*

$$\mathcal{J}_{\mathrm{c}}(f) = \min\{J^h_{\mathcal{P}} : \ \mathcal{P} \textit{ is strongly } h\textit{-private protocol for } f\}.$$

Our measure modifies the definition of Bar Yehuda et al. [1] by considering the result of the function. The proof Theorem 4 uses some facts that we get from the derandomisation of an optimal protocol. We will present such a derandomisation in the next section.

Proof (of Theorem 4). Let f be a function. We show first that

$$\mathcal{J}_{\mathrm{c}}(f) \le \min\{J^h_{\mathcal{P}} : \ \mathcal{P} \text{ is strongly } h\text{-private protocol for } f\}. \tag{6}$$

Assume h is function and $\mathcal{P}$ is a strongly h-private protocol for computing f such that $J^h_{\mathcal{P}}$ is minimum among all such functions h and protocols $\mathcal{P}$. By the definition of h-privacy we have that for every $x, y_1, y_2 \in \{0,1\}^n$, $h(x,y_1) = h(x,y_2)$ implies that for all random choices r of A, $\mathcal{P}(x,y_1)$ and $\mathcal{P}(x,y_2)$ have the same distribution. Hence, for every $x \in \{0,1\}^n$ and $z \in \mathrm{range}(f)$ we have

$$s_{\mathcal{P},A}(x,z) \ \le\ |\{h(x,y) : \ y \in \{0,1\}^n, f(x,y) = z\}| \ =\ s^h_{\mathcal{P},A}(x,z).$$

Similarly we have: $s_{\mathcal{P},B}(y,z) \le s^h_{\mathcal{P},B}(y,z)$. Hence both $J_{\mathcal{P},A} \le J^h_{\mathcal{P},A}$ and $J_{\mathcal{P},B} \le J^h_{\mathcal{P},B}$ are true and therefore we get $J_{\mathcal{P}} \le J^h_{\mathcal{P}}$. This implies that Inequality (6) is true.

To see that the inverse inequality to (6) is also true, we apply Theorem 6. Let $\mathcal{P}$ be a protocol for f such that $J_{\mathcal{P}}$ is minimal among all protocols computing f. By Theorem 6 there exists a deterministic protocol $\mathcal{P}'$ for f such that

$$J_{\mathcal{P}'} \le J_{\mathcal{P}} = \mathcal{J}_{\mathrm{c}}(f).$$

Since $\mathcal{P}'$ is deterministic, we can define a function h for every $x, y \in \{0,1\}^n$ as follows: $h(x,y) = \mathcal{P}'(x,y)$. Obviously, $\mathcal{P}'$ is strongly h-private and it is true that $J^h_{\mathcal{P}'} = J_{\mathcal{P}'}$. Hence, by the inequality above one can conclude:

$$\min\{J^h_{\mathcal{P}} : \ \mathcal{P} \text{ is strongly } h\text{-private protocol for } f\} \ \le\ J^h_{\mathcal{P}'} \ =\ J_{\mathcal{P}'} \ \le \mathcal{J}_{\mathrm{c}}(f).$$

This completes the proof. □

Using Theorem 4 we get that $\mathcal{J}_{\mathrm{c}}(f) \le \mathcal{I}_{\mathrm{c}}(f)$ for every function f. However the difference can be very big: e.g. for f_{min} we have by Proposition 2 that $\mathcal{I}_{\mathrm{c}}(f_{min}) = n$. On the other hand using the protocol given in [1]:

Protocol 2. For all $i = 0, 1, \ldots 2^n - 1$ or until the first 1 is transmitted do:
1. A transmits bit 1 if $x = i$ and 0 otherwise;
2. B transmits bit 1 if $y = i$ and 0 otherwise.

we get that $\mathcal{J}_c(f_{min}) \leq 1$. Since f_{min} cannot be computed privately, we obtain the equality $\mathcal{J}_c(f_{min}) = 1$.

For the equality function f_{equ} (see (2)), we get $\mathcal{I}_c(f_{equ}) \leq \log_2(n+1)$ (compare the inequalities (3)). By the fact shown in [1] that for any deterministic protocol $\mathcal{P}$ which computes f_{equ} there is $v \in \{0,1\}^n$ such that the size of the set

$$\{\mathcal{P}(x,v) : x \in \{0,1\}^n\} \cup \{\mathcal{P}(v,y) : y \in \{0,1\}^n\}$$

has at lest $n+2$ elements, we obtain for $z = 0$: $s_{\mathcal{P},A}(v,z) + s_{\mathcal{P},B}(v,z) \geq n+2$ and finally that $\mathcal{J}_c(f_{equ}) \geq \log_2(n+2)/2 > \log_2 n - 1$. Hence we get the following bounds: $\log_2 n - 1 < \mathcal{J}_c(f_{equ}) \leq \mathcal{I}_c(f_{equ}) \leq \log_2(n+1)$.

We close the section by giving a general lower bound for $\mathcal{J}_c$. Recall that a rectangle in $\{0,1\}^n \times \{0,1\}^n$ is a Cartesian product $R = V \times H$ with $V, H \subseteq \{0,1\}^n$. The rectangle R is f-constant if f is constant over R. Obviously, every protocol for $\mathcal{P}$ partitions the communication matrix M_f into f-constant rectangles. Let r_f be the largest width of an f-constant rectangle.

Theorem 5. *For every Boolean function f of two n-bit inputs*

$$\mathcal{J}_c(f) \geq n - \log_2 r_f - 2.$$

The proof of Theorem 3 of [1] works for our Theorem.

Using the general bound given in the Theorem above one can find lower bounds for Boolean functions f communication matrix of which is of the Hadamard type (see [1]). From this characterisation we get e.g. that for the n-variable inner product mod 2 function defined as

$$f_{in}(x,y) = \sum_{i=1}^{n} x_i \cdot y_i \mod 2 \tag{7}$$

it holds that $\mathcal{J}_c(f_{in}) \geq n/2 - 2$.

5 Derandomisation

In this section we will show that every randomized protocol $\mathcal{P}$ that computes the function f correctly with probability 1 can be simulated by a deterministic protocol $\mathcal{P}'$ such that the additional information that can be deduced running protocol $\mathcal{P}'$ is bounded by the additional information that can be deduced running protocol $\mathcal{P}$, i.e. $J_{\mathcal{P}'} \leq J_{\mathcal{P}}$. We will start by the derandomisation of the part of A.

Let us assume that A performing $\mathcal{P}$ starts the communication and let ℓ be an upper bound for the number of random bits used by A. In the algorithm below A simulates the t-th round of the computation of $\mathcal{P}$, with $t = 1, 2, 3, \ldots$ as follows: On a given input x A computes iteratively string c_t and a subset $\mathcal{R}_t \subseteq \{0,1\}^{\leq \ell}$ of all binary strings of lengths less or equal to ℓ, such that c_t is a complete communication string of a computation during the first t rounds and

$\mathcal{R}_t$ is a subset of all possible random strings that can be used by A. A string r is in $\mathcal{R}_t$ if there exists a computation of $\mathcal{P}$ such that the first t rounds of the computation are consistent with c_t when A on x and r. Define $\mathcal{R}_0 = \{0,1\}^{\leq \ell}$ and let c_0 be the empty string.

1. If t is odd then for every $r \in \mathcal{R}_{t-1}$ A simulates (deterministically) the t-th round of the computation of $\mathcal{P}$ on input x with the random string r that is consistent with the communication string c_{t-1} and computes a communication string for the tth round. Let w_t be lexicographically smallest among all such strings. Then A computes $\mathcal{R}_t := \{r \in \mathcal{R}_{t-1} \mid A \text{ sends } w_t \text{ on } x, r, c_{t-1}\}$ and $c_t := c_{t-1} \circ w_t$ and sends w_t to B. For two strings v and v', by $v \circ v'$ we denote concatenation of v and v'.
2. If t is even and u_t is a message received by A from B in tth round, then $c_t := c_{t-1} \circ u_t$.

Assume that the protocol stops in round T, then it is easy to see that for every input y, every possible result z, and every random string of B, A chooses for every pair of inputs x, x' the communication string s such that it is the lexicographically smallest string with $\Pr[\mathcal{P}(x,y) = s|r], \Pr[\mathcal{P}(x',y) = s|r] > 0$. Hence, inputs x, x' that gives the same distribution on y, z, r when running $\mathcal{P}$ gives also the same distribution when running the deterministic protocol $\mathcal{P}'$.

Note that we can derandomize the part of B's protocol analogously. Hence, we can conclude:

Lemma 2. *For every protocol $\mathcal{P}$ there exists a deterministic protocol $\mathcal{P}'$ computing the same function, such that for every choice of x, y, z $s_{\mathcal{P}',A}(x,z) \leq s_{\mathcal{P},A}(x,z)$ and $s_{\mathcal{P}',B}(y,z) \leq s_{\mathcal{P},B}(y,z)$.*

Theorem 6. *For every protocol $\mathcal{P}$ there exists an deterministic protocol $\mathcal{P}'$ computing the same function, such that $J_{\mathcal{P}'} \leq J_{\mathcal{P}}$.*

This result generalises the result of Kushilevitz [6] that a protocol can be computed privately in the two party scenario iff it can be computed privately by a deterministic protocol.

Using our simulation result, we can directly deduce some bounds for the size of a minimal information source. Let s_f be the minimum size of the information source of a protocol computing f, i.e. let

$$s_f = \min_{\mathcal{P}} \max_{x,y,z} \{s_{\mathcal{P},A}(x,z), s_{\mathcal{P},B}(y,z)\}$$

(note that $\mathcal{J}_c(f) = \log_2 s_f$).

Corollary 1. $s_f \leq CS_D(f)$.

Proof. Assume that $s_f > CS_D(f)$ and let $\mathcal{P}$ be a deterministic protocol that achieve s_f and $\mathcal{P}'$ be a deterministic protocol that achieve $CS_D(f)$. Assume that $s_f = s_{\mathcal{P},A}(x,z)$ for appropriate chosen values x, z. Then the number of communication strings seen by A on input x and result z when running $\mathcal{P}$ is

even higher then the number of communication strings seen by both parties when running $\mathcal{P}'$ on arbitrary inputs. Hence, the size of the information source when running $\mathcal{P}'$ is smaller than the size of the information source when running $\mathcal{P}$ – contradicting the assumption that $\mathcal{P}$ achieves the minimum size of the information source. □

Corollary 2. $CS_D(f) = \min_{deterministic\ \mathcal{P}\ computes\ f} |\bigcup_{x,z} S_{\mathcal{P},A}(x,z)|$.

Proof. Let $\mathcal{P}$ be a deterministic protocol for f such that

$$\left|\bigcup_{x,z} S_{\mathcal{P},A}(x,z)\right| \;=\; \min_{\text{deterministic } \mathcal{P}' \text{ computes } f} \left|\bigcup_{x,z} S_{\mathcal{P}',A}(x,z)\right| \;.$$

Since $\mathcal{P}$ is deterministic every distribution in the set $\bigcup_{x,z} S_{\mathcal{P},A}(x,z)$ rates exactly one communication string with a strictly positive probability. Furthermore, the set determines all communication strings used when running $\mathcal{P}$. The claim follows from the observation, that $\mathcal{P}$ is chosen such that the number of used communication strings is minimal. □

6 Lower Bounds on Size of the Information Source

Corollary 1 gives a general upper bound on the minimum size of the information source s_f. This bound is not tight. In fact, it is well known (see e.g [7]) that for the equality function f_{equ} it holds that $CS_D(f_{equ}) \geq 2^n$ and $C_D(f_{equ}) = n$. On the other hand from the Protocol 1 it follows that for any optimal protocol $\mathcal{P}$ we get $s_{\mathcal{P},A}(x,z)$, $s_{\mathcal{P},B}(y,z) \leq n$ for every x, y, z. Hence $s_{f_{equ}} \leq n < 2^n \leq CS_D(f_{equ})$. In this section we will prove a linear lower bound for the size of the information source with respect to the communication complexity, i.e. we show that for any f $s_f \in \Omega(C_D(f)/|\operatorname{range}(f)|)$. In particular for f_{equ} we get $C_D(f_{equ})/4 - 1 \leq s_{f_{equ}}$.

For a node v of the communication tree let X_v and Y_v denote the sets of input strings of A and B, respectively, such that on the input pairs $(x,y) \in X_v \times Y_v$ the protocol reaches v. Let $s_{\mathcal{P},A,v}(x,z)$ denote the size of the information source of the subprotocol of $\mathcal{P}$ starting in v and restricting the inputs to $X_v \times Y_v$. Let $s_{\mathcal{P},B,v}(y,z)$ be defined analogously. Finally, define

$$\operatorname{range}(v) \;=\; \{\; f(x,y) \mid (x,y) \in X_v \times Y_v \;\} \;.$$

Without loss of generality let us restrict ourselves only to the protocols $\mathcal{P}$ sending no unnecessary bits for computing the function. Formally assume that all internal nodes of a communication tree of $\mathcal{P}$ have degree at least 2. We start with the following observation:

Lemma 3. *Let $\mathcal{P}$ be a deterministic protocol computing a function f and let $v_1, \ldots, v_t$ be a leaf-to-root path in the communication tree of $\mathcal{P}$. Then for all $i \in \{1, \ldots, t\}$ there exists $x \in X_{v_i}, y' \in Y_{v_i}$, and $z, z' \in \operatorname{range}(f)$ such that*

$$\max\{s_{\mathcal{P},A,v_i}(x,z), s_{\mathcal{P},B,v_i}(y',z')\} \;\geq\; \left\lceil \frac{i}{2 \cdot |\operatorname{range}(v_i)|} \right\rceil \;-\; 1 \;.$$

Proof. The proof follows for $i = 1$ since for every leaf v_1 of the communication tree we have $s_{\mathcal{P},A,v_1}(x,z) = s_{\mathcal{P},B,v_1}(y,z) = 0$.

Consider now an internal node v_i, with $i > 1$. Let $u_1, \ldots, u_d$ be all successors of v_i in the communication tree. Obviously, v_{i-1} is one of the nodes u_j. Let us assume, that A has to send some message in v_i, then for all $x \in X_{v_{i-1}} \subset X_{v_i}$, $y \in Y_{v_{i-1}} = Y_{v_i}$, and $z = f(x,y)$:

$$s_{\mathcal{P},A,v_i}(x,z) = \max\{1,\ s_{\mathcal{P},A,v_{i-1}}(x,z)\}.$$

On the other hand one can prove that for the information source of B we have

$$s_{\mathcal{P},B,v_i}(y,z) = \sum_{j\in\{1,\ldots,d\} \text{ with } z\in\text{range}(u_j)} \max\{1,\ s_{\mathcal{P},B,u_j}(y,z)\}.$$

Therefore we can bound the quantity as follows

$$s_{\mathcal{P},B,v_i}(y,z) \geq \begin{cases} 1 + \max\{1, s_{\mathcal{P},B,v_{i-1}}(y,z)\} & \text{if } z \in \text{range}(u_j) \text{ for some } u_j \neq v_{i-1} \\ \max\{1, s_{\mathcal{P},B,v_{i-1}}(y,z)\} & \text{else.} \end{cases}$$

Assume that there are k nodes on the sub-path $v_1, \ldots, v_i$ where A sends a message to B. Then there exists $z' \in \text{range}(v_i)$ such that for at least

$$\left\lceil \frac{k}{|\text{range}(v_i)|} \right\rceil - 1$$

of these nodes v_j it holds that $z' \in \text{range}(v_{j-1}) \cap \text{range}(u)$ for some direct successor $u \neq v_{j-1}$ of v_j . Note that we can show simular bounds for $s_{\mathcal{P},A,v_i}(x,z)$ and $s_{\mathcal{P},B,v_i}(y,z)$ if Bob sends a message. The claim follows immediately since either A or B has to send some message in at least $\lceil i/2 \rceil$ of the nodes $v_1, \ldots, v_i$. □

As a corollary we obtain:

Corollary 3. *For every function f of two n-bit inputs it is true*

$$\frac{C_D(f)}{2 \cdot |range(f)|} - 1 \quad \leq \quad s_f.$$

Combining the corollary above with Theorem 6 we can conclude the following lower bound on the additional information:

Theorem 7. *For every function f of two n-bit inputs we have*

$$\mathcal{J}_c(f) \geq \log_2 C_D(f) - \log_2 |\,\text{range}(f)| - O(1).$$

7 Conclusions

In this paper measures for revealed information required for computing f have been considered. We have analysed the measures given by Bar-Yehuda et al.

and have showed that some results presented in [1] are wrong. Moreover we have observed some unnatural properties of the measures. We have introduced a new definition for the additional information for two party protocols and have given some bounds for concrete functions for the additional information. We get e.g. that for the n-variable inner product mod 2 function it is true that $\mathcal{J}_c(f_{in}) \geq n/2 - 2$. An interesting open problem is to show lower and upper bounds on $\mathcal{J}_c$ for another specific functions. A further task to do is to investigate a tradeoff between the additional information and the number of rounds for communication protocols.

References

1. Reuven Bar-Yehuda, Benny Chor, Eyal Kushilevitz, and Alon Orlitsky. Privacy, additional information, and communication. *IEEE Transactions on Information Theory*, 39(6):1930–1943, 1993. An early version of this paper appear in *Proc. of 5th IEEE Structure in Complexity Theory*, 1990, pp. 55–65.
2. Donald Beaver. Perfect Privacy for Two Party Protocols. Technical Report TR-11-89, Harvard University, 1989.
3. Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Proc. of the 20th Ann. ACM Symp. on Theory of Computing (STOC)*, pages 1–10. ACM Press, 1988.
4. Markus Bläser, Anderas Jakoby, Maciej Liśkiewicz, and Bodo Manthey, Privacy in Non-Private Environments. Proceedings of the 10th Annual International Cryptology Conference on the Theory and Application of Cryptology and Information Security (Asiacrypt), 2004, 137-151.
5. David Chaum, Claude Crépeau, and Ivan Damgård. Multiparty unconditionally secure protocols. In *Proc. of the 20th Ann. ACM Symp. on Theory of Computing (STOC)*, pages 11–19. ACM Press, 1988.
6. Eyal Kushilevitz. Privacy and communication complexity. *SIAM Journal on Discrete Mathematics*, 5(2):273–284, 1992.
7. Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1997.
8. Andrew Chi-Chih Yao. Protocols for secure computations. In *Proc. of the 23rd Ann. IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 160–164. IEEE Computer Society, 1982.