The Communication Complexity of Private Simultaneous Messages, Revisited*

Benny Applebaum[1], Thomas Holenstein[2], Manoj Mishra[3], and Ofer Shayevitz[1]

[1] Tel Aviv University, Tel Aviv, Israel
{benny.applebaum,ofersha}@gmail.com
[2] Google, Zurich, Switzerland
thomas.holenstein@gmail.com
[3] N.I.S.E.R., Bhubaneswar, India manoj.mishra@niser.ac.in

Abstract. Private Simultaneous Message (PSM) protocols were introduced by Feige, Kilian and Naor (STOC '94) as a minimal non-interactive model for information-theoretic three-party secure computation. While it is known that every function $f : \{0,1\}^k \times \{0,1\}^k \rightarrow \{0,1\}$ admits a PSM protocol with exponential communication of $2^{k/2}$ (Beimel et al., TCC '14), the best known (non-explicit) lower-bound is $3k - O(1)$ bits. To prove this lower-bound, FKN identified a set of simple requirements, showed that any function that satisfies these requirements is subject to the $3k - O(1)$ lower-bound, and proved that a random function is likely to satisfy the requirements.

We revisit the FKN lower-bound and prove the following results:

(Counterexample) We construct a function that satisfies the FKN requirements but has a PSM protocol with communication of $2k + O(1)$ bits, revealing a gap in the FKN proof.

(PSM lower-bounds) We show that, by imposing additional requirements, the FKN argument can be fixed leading to a $3k - O(\log k)$ lower-bound for a random function. We also get a similar lower-bound for a function that can be computed by a polynomial-size circuit (or even polynomial-time Turing machine under standard complexity-theoretic assumptions). This yields the first non-trivial lower-bound for an explicit Boolean function partially resolving an open problem of Data, Prabhakaran and Prabhakaran (Crypto '14, IEEE Information Theory '16). We further extend these results to the setting of imperfect PSM protocols which may have small correctness or privacy error.

(CDS lower-bounds) We show that the original FKN argument applies (as is) to some weak form of PSM protocols which are strongly related to the setting of Conditional Disclosure of Secrets (CDS). This connection yields a simple combinatorial criterion for establishing linear $\Omega(k)$-bit CDS lower-bounds. As a corollary, we settle the complexity of the Inner Product predicate resolving an open problem of Gay, Kerenidis, and Wee (Crypto '15).

[*] This is the full version of a paper appearing in EUROCRYPT 2018. This work was done while the second author was at ETH Zurich and the third author was at Tel Aviv University.

1 Introduction

Information theoretic cryptography studies the problem of secure communication and computation in the presence of computationally unbounded adversaries. Unlike the case of computational cryptography whose full understanding is closely tied to basic open problems in computational complexity, information theoretic solutions depend "only" on non-computational (typically combinatorial or algebraic) objects. One may therefore hope to gain a full understanding of the power and limitations of information theoretic primitives. Indeed, Shannon's famous treatment of perfectly secure symmetric encryption [30] provides an archetypical example for such a study.

Unfortunately, for most primitives, the picture is far from being complete. This is especially true for the problem of *secure function evaluation* (SFE) [33], in which a set of parties $P_1, \ldots, P_m$ wish to jointly evaluate a function f over their inputs while keeping those inputs private. Seminal completeness results show that *any function* can be securely evaluated with information theoretic security [10, 13] (or computational security [33, 19]) under various adversarial settings. However, the *communication complexity* of these solutions is tied to the *computational complexity* of the function (i.e., its circuit size), and it is unknown whether this relation is inherent. For instance, as noted by Beaver, Micali, and Rogaway [8] three decades ago, we cannot even rule out the possibility that any function can be securely computed by a constant number of parties with communication that is polynomial in the input length, even in the simple setting where the adversary passively corrupts a single party. More generally, the communication complexity of securely computing a function (possibly via an inefficient protocol) is wide open, even in the most basic models.

1.1 A Minimal Model for Secure Computation

In light of the above, it makes sense to study the limitation of information theoretic secure computation in its simplest form. In [16] Feige, Kilian and Naor (hereinafter referred to as FKN) presented such a "Minimal Model for Secure Computation". In this model, Alice and Bob hold private inputs, x and y, and they wish to let Charlie learn the value of $f(x, y)$ without leaking any additional information. The communication pattern is minimal. Alice and Bob each send to Charlie a single message, a and b respectively, which depends on the party's input and on a random string r which is shared between Alice and Bob but is hidden from Charlie. Given (a, b) Charlie should be able to recover $f(x, y)$ without learning additional information. The parties are assumed to be computationally unbounded, and the goal is to minimize the communication complexity of the protocol (i.e., the total number of bits sent by Alice and Bob). Following [23], we refer to such a protocol as a *private simultaneous message* protocol (PSM).

Definition 1 (Private Simultaneous Messages). *A private simultaneous message (PSM) protocol $\Pi = (\Pi_A, \Pi_B, g)$ for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is a triple of functions $\Pi_A : \mathcal{X} \times \mathcal{R} \to \mathcal{A}$, $\Pi_B : \mathcal{Y} \times \mathcal{R} \to \mathcal{B}$, and $g : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ that satisfy the following two properties.*

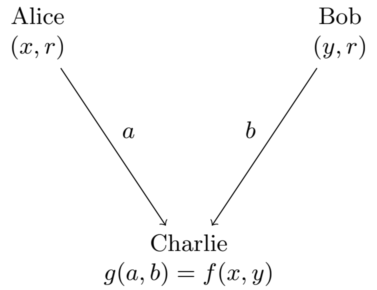

Fig. 1. Schematic of a PSM protocol.

- *(δ-Correctness) The protocol has correctness error of δ if for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ it holds that*

$$\Pr_{r \xleftarrow{\$} \mathcal{R}} [f(x, y) \neq g(\Pi_A(x, r), \Pi_B(y, r))] \leq \delta$$

- *(ϵ-Privacy) The protocol has privacy error of ϵ if for every pair of inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $(x', y') \in \mathcal{X} \times \mathcal{Y}$ for which $f(x, y) = f(x', y')$ the random variables*

$$(\Pi_A(x, r), \Pi_B(y, r)) \quad \textit{and} \quad (\Pi_A(x', r), \Pi_A(y', r), \tag{1}$$

induced by a uniform choice of $r \xleftarrow{\$} \mathcal{R}$, are ϵ-close in statistical distance.

We mainly consider perfect *protocols which enjoy both* perfect correctness *($\delta = 0$) and* perfect privacy *($\epsilon = 0$). We define the* communication complexity *of the protocol to be* $\log |\mathcal{A}| + \log |\mathcal{B}|$.

The correctness and privacy conditions assert that, for every pair of inputs (x, y) and (x', y'), the transcript distributions are either close to each other when $f(x, y) = f(x', y')$, or far apart when $f(x, y) \neq f(x', y')$. Hence, the joint computation of Alice and Bob, $C_r(x, y) = (\Pi_A(x, r), \Pi_B(y, r))$, can be also viewed as a "randomized encoding" [24, 5] (or "garbled version") of the function $f(x, y)$ that has the property of being 2-decomposable into an x-part and a y-part. Being essentially non-interactive, such protocols (and their multiparty variants [23]) have found various applications in cryptography (cf. [22, 2]). Moreover, it was shown in [9, 6] that PSM is the strongest model among several other non-interactive models for secret-sharing and zero-knowledge proofs.

FKN showed that any function $f : \{0, 1\}^k \times \{0, 1\}^k \to \{0, 1\}$ admits a PSM protocol [16]. The best known communication complexity is polynomial for log-space computable functions [16] and $O(2^{k/2})$ for general functions [9]. While it seems likely that some functions require super-polynomial communication, the best known lower-bound, due to the original FKN paper, only shows that a random function requires $3k - O(1)$ bits of communication. This lower-bound is somewhat weak but still non-trivial since an insecure solution (in which Alice and Bob just send their inputs to Charlie) costs $2k$ bits of communication. The question of improving this lower-bound is an intriguing open problem. In this paper, we aim for a more modest goal. Inspired by the general theory of communication complexity, we ask:

> How does the PSM complexity of a function f relate to its combinatorial properties? Is there a "simple" condition that guarantees a non-trivial lower-bound on the PSM complexity?

We believe that such a step is necessary towards proving stronger lower-bounds. Additionally, as we will see, this question leads to several interesting insights for related information-theoretic tasks.

1.2 Revisiting the FKN lower-bound

Our starting point is the original proof of the $3k$ lower-bound from [16]. In order to prove a lower-bound FKN relax the privacy condition by requiring that Charlie will not be able to recover the last bit of Alice's input. Formally, let us denote by $\bar{x}$ the string obtained by flipping the last bit of x. Then, the privacy condition (Eq. 1) is relaxed to hold only over *sibling* inputs (x, y) and $(\bar{x}, y)$ for which $f(x, y) = f(\bar{x}, y)$. We refer to this relaxation as *weak privacy.* Since (standard) privacy implies weak privacy, it suffices to lower-bound the communication complexity of weakly private PSM protocols.

To prove a lower-bound for random functions, FKN (implicitly) identify three conditions which hold for most functions and show that if a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ satisfies these conditions then any weak PSM for f has communication complexity of at least $3k - O(1)$. The FKN conditions are:

1. The function f is *non-degenerate*, namely, for every $x \neq x'$ there exists y for which $f(x, y) \neq f(x', y)$ and similarly, for every $y \neq y'$ there exists x for which $f(x, y) \neq f(x, y')$.
2. The function is *useful* in the sense that for at least $\frac{1}{2} - o(1)$ of the inputs (x, y) it holds that $f(x, y) = f(\bar{x}, y)$ where $\bar{x}$ denotes the string x with its last bit flipped. (An input (x, y) for which the equation holds is referred to as being *useful.*[4])
3. We say that $(x_1, \ldots, x_m) \times (y_1, \ldots, y_n)$ is a *complement similar* rectangle of f if $f(x_i, y_j) = f(\bar{x}_i, y_j)$ for every $1 \leq i \leq m$ and $1 \leq j \leq n$. Then, f has no complement similar rectangle of size mn larger than $M = 2^{k+1}$. Equivalently, the function $f'(x, y) = f(x, y) - f(\bar{x}, y)$, which can be viewed as a partial derivative of f with respect to its last coordinate, has no 0-monochromatic rectangle of size M.

We observe that the above conditions are, in fact, insufficient to prove a non-trivial lower-bound. As a starting point, we note that the inner-product function

[4] In the FKN terminology such an input (x, y) is referred to as being dangerous.

has low PSM complexity and has no large monochromatic rectangles. While the inner-product function cannot be used directly as a counterexample (since it has huge complement similar rectangles), we can construct a related function f such that: (1) the derivative f' is (a variant of) the inner product function and so f' has no large monochromatic rectangles; and (2) by applying some local preprocessing on Alice's input, the computation of $f(x,y)$ reduces to the computation of the inner product function. Altogether, we prove the following theorem (see Section 3).

Theorem 1 (FKN counterexample). *There exists a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ that satisfies the FKN conditions but has a (standard) PSM of communication complexity of $2k + O(1)$.*

Let us take a closer look at the proof of the FKN lower-bound to see where the gap is. The FKN proof boils down to showing that the set S_r of all possible transcripts (a,b) sent by Alice and Bob under a random string r, has relatively small intersection with the set $S_{r'}$ of all possible transcripts (a,b) sent by Alice and Bob under a different random string r'. Such a collision, $c = (a,b) \in S_r \cap S_{r'}$, is counted as a *trivial collision* if the inputs (x,y) that generate c under r are the same as the inputs (x',y') that generate c under r'. Otherwise, the collision is counted as *non-trivial.* The argument mistakenly assumes that all non-trivial collisions are due to sibling inputs, i.e., $(x',y') = (\bar{x},y)$. In other words, it is implicitly assumed that the transcript (a,b) *fully reveals* all the information about (x,y) except for the last input of x. (In addition to the value of $f(x,y)$ which is revealed due to the correctness property.) More formally, a *weakly private fully revealing PSM $\Pi = (\Pi_A, \Pi_B, g)$* for a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ is a perfect PSM for the function $f' : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}^{k-1} \times \{0,1\}^k \times \{0,1\}$ that takes (x,y) and outputs $(x[1:k-1], y, f(x,y))$, where $x[1:k-1]$ is the $(k-1)$-prefix of x. (See also Definition 4.)

We show that the original FKN argument holds if one considers fully-revealing PSM protocols. (See Theorem 8 for a slightly stronger version.)

Theorem 2 (LB's against weakly private fully revealing PSM). *Let $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ be a non-degenerate function. Let M be an upper-bound on size of the largest complement similar rectangle of f and let U be a lower-bound on the number of useful inputs of f. Then, any weakly-private fully-revealing PSM for f has communication complexity of at least $2 \log U - \log M - O(1)$. In particular, for all but $o(1)$ fraction of the functions $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$, we get a lower-bound of $3k - O(1)$.*

A lower-bound of c bits against fully-revealing weakly-private PSM easily yields a lower-bound of $c - 2k + 1$ bits for PSM. (Since a standard PSM can be turned into a fully-revealing weakly-private PSM by letting Alice/Bob append $x[1:k-1]$ and y to their messages.) Unfortunately, this loss (of $2k$ bits) makes the $3k$ bit lower-bound useless. Moreover, Theorem 1 shows that this loss is unavoidable. Put differently, fully-revealing weakly-private PSM may be more expensive than standard PSM. Nevertheless, as we will see in Section 1.4, lower-bounds for fully-revealing weakly-private PSM have useful implications for other models.

1.3 Fixing the PSM lower-bound

We show that the FKN argument can be fixed by posing stronger requirements on f. Roughly speaking, instead of limiting the size of complement similar rectangles, we limit the size of any pair of *similar* rectangles by a parameter M. That is, if the restriction of f to the ordered rectangle $R = (x_1, \dots, x_m) \times (y_1, \dots, y_\ell)$ is equal to the restriction of f to the ordered rectangle $R' = (x'_1, \dots, x'_m) \times (y'_1, \dots, y'_\ell)$ and the rectangles are disjoint in the sense that either $x_i \neq x'_i$ for every i, or $y_j \neq y'_j$ for every j, then the size $m\ell$ of R should be at most M. (See Section 2 for a formal definition.)

Theorem 3 (perfect-PSM LB's). *Let $\mathcal{X}, \mathcal{Y}$ be sets of size at least 3, and let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a non-degenerate function for which any pair of disjoint similar rectangles (R, R') satisfies $|R| \leq M$. Then, any perfect PSM for f has communication of at least $2(\log|\mathcal{X}| + \log|\mathcal{Y}|) - \log M - 2$.*

The theorem is proved by a distributional version of the FKN argument which also implies Theorem 2. (See Section 4.) As a corollary, we recover the original lower-bound claimed by FKN.

Corollary 1. *For a $1 - o(1)$ fraction of the functions $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ any perfect PSM protocol for f requires $3k - 2\log k - O(1)$ bits of total communication.*[5]

Proof. It is not hard to verify that $1 - o(1)$ fraction of all functions are non-degenerate. In Section 6 we further show that, for $1 - o(1)$ of the functions, any pair of disjoint similar rectangles (R, R') satisfies $|R| \leq k^2 \cdot 2^k$. The proof follows from Theorem 3. □

By partially de-randomizing the proof, we show that the above lower-bound applies to a function that is computable by a family of polynomial-size circuits, or, under standard complexity-theoretic assumptions, by a polynomial-time Turing machine. This resolves an open question of Data, Prabhakaran and Prabhakaran [15] who proved a similar lower-bound for an explicit *non-boolean* function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}^{k-1}$. Prior to our work, we could not even rule out the (absurd!) possibility that all efficiently computable functions admit a perfect PSM with communication of $2k + o(k)$.

Theorem 4. *There exists a sequence of polynomial-size circuits*

$$f = \left\{ f_k : \{0,1\}^k \times \{0,1\}^k \to \{0,1\} \right\}$$

[5] The constant 2 can be replaced by any constant larger than 1.

such that any perfect PSM for f_k has communication complexity of at least $3k - O(\log k)$ bits. Moreover, assuming the existence of a hitting-set generator against co-nondeterministic uniform algorithms, f is computable by a polynomial-time Turing machine.[6]

Remark 1 (On the hitting-set generator assumption). The exact definition of a hitting-set generator against co-nondeterministic uniform algorithms is postponed to Section 6. For now, let us just say that the existence of such a generator follows from standard Nissan-Wigderson type complexity-theoretic assumptions. In particular, it suffices to assume that the class E of functions computable in $2^{O(n)}$-deterministic time contains a function that has no sub-exponential nondeterministic circuits [28], or, more liberally, that some function in E has no sub-exponential time Arthur-Merlin protocol [21]. (See also the discussion in [7].)

Lower-bounds for imperfect PSM's. We extend Theorem 3 to handle imperfect PSM protocols by strengthening the non-degeneracy condition and the non self-similarity condition. This can be used to prove an imperfect version of Corollary 1 showing that, for almost all functions, an imperfect PSM with correctness error δ and privacy error ϵ must communicate at least

$$\min\{3k - 2\log(k), 2k + \log(1/\epsilon), 2k + \log(1/\delta)\} - O(1)$$

bits. An analogous extension of Theorem 4, yields a similar bound for an explicit function. (See Section 5.)

1.4 Applications to Conditional Disclosure of Secrets

We move on to the closely related model of *Conditional Disclosure of Secrets* (CDS) [18]. In the CDS model, Alice holds an input x and Bob holds an input y, and, in addition, Alice holds a secret bit s. The referee, Charlie, holds both x and y, but does not know the secret s. Similarly to the PSM case, Alice and Bob use shared randomness to compute the messages a and b that are sent to Charlie. The CDS requires that Charlie can recover s from (a, b) if and only if the predicate $f(x, y)$ evaluates to one.[7]

Definition 2 (Conditional Disclosure of Secrets). *A* conditional disclosure of secrets *(CDS) protocol $\Pi = (\Pi_A, \Pi_B, g)$ for a predicate $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ and domain $\mathcal{S}$ of secrets is a triple of functions $\Pi_A : \mathcal{X} \times \mathcal{S} \times \mathcal{R} \to \mathcal{A}$, $\Pi_B : \mathcal{Y} \times \mathcal{R} \to \mathcal{B}$ and $g : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \to \mathcal{S}$ that satisfy the following two properties:*

[6] It is worth mentioning that the proof of Theorem 4 strongly relies on the explicit combinatorial condition given in Theorem 3 (and we do not know how to obtain it directly from Corollary 1). This illustrates again the importance of relating PSM complexity to other more explicit properties of functions.

[7] Usually, it is assumed that both Alice and Bob hold the secret s. It is not hard to see that this variant and our variant (in which only Alice knows the secret) are equivalent up to at most 1-bit of additional communication.

1. *(Perfect Correctness) For every (x, y) that satisfies f and any secret $s \in \mathcal{S}$ we have that:*

$$\Pr_{r \overset{\$}{\leftarrow} \mathcal{R}} [g(x, y, \Pi_A(x, s, r), \Pi_B(y, r)) \neq s] = 0.$$

2. *(Perfect Privacy) For every input (x, y) that does not satisfy f and any pair of secrets $s, s' \in \mathcal{S}$ the distributions*

$$(x, y, \Pi_A(x, s, r), \Pi_B(y, r)) \quad \text{and} \quad (x, y, \Pi_A(x, s', r), \Pi_B(y, r)),$$

induced by $r \overset{\$}{\leftarrow} \mathcal{R}$ are identically distributed.

The communication complexity *of the CDS protocol is* $(\log |\mathcal{A}| + \log |\mathcal{B}|)$ *and its* randomness complexity *is* $\log |\mathcal{R}|$. *By default, we assume that the protocol supports single-bit secrets ($\mathcal{S} = \{0, 1\}$).*[8]

Intuitively, CDS is weaker than PSM since it either releases s or keeps it private but it cannot *manipulate the secret data.*[9] Still, this notion has found useful applications in various contexts such as information-theoretically private information retrieval (PIR) protocols [14], priced oblivious transfer protocols [1], secret sharing schemes for graph-based access structures (cf. [11, 12, 31]), and attribute-based encryption [20, 29].

The communication complexity of CDS. In light of the above, it is interesting to understand the communication complexity of CDS. Protocols with communication of $O(t)$ were constructed for t-size Boolean formula by [18] and were extended to t-size (arithmetic) branching programs by [25] and to t-size (arithmetic) span programs by [6]. Until recently, the CDS complexity of a general predicate $f : \{0, 1\}^k \times \{0, 1\}^k \to \{0, 1\}$ was no better than its PSM complexity, i.e., $O(2^{k/2})$ [9]. This was improved to $2^{O(\sqrt{k \log k})}$ by Liu, Vaikuntanathan and Wee [27]. Moreover, Applebaum et al. [4] showed that, for very long secrets, the amortized complexity of CDS can be reduced to $O(\log k)$ bits per bit of secret. Very recently, the amortized cost was further reduced to $O(1)$ establishing the existence of general CDS with constant rate [3].

Lower-bounds for the communication complexity of CDS were first established by Gay, Kerenidis, and Wee [17]. Their main result shows that the CDS communication of a predicate f is at least logarithmic in its randomized one-way communication complexity, and leads to an $\Omega(\log k)$ lower-bound for several explicit functions. Applebaum et al. [4] observed that weakly private PSM reduces to CDS. This observation together with the $3k$-bit FKN lower-bound for weakly

[8] One may consider imperfect variants of CDS. In this paper we restrict our attention to the (more common) setting of perfect CDS.

[9] This is analogous to the relation between Functional Encryption and Attribute Based Encryption. Indeed, CDS can be viewed as an information-theoretic one-time variant of Attribute Based Encryption.

private PSM has lead to a CDS lower-bound of $k-o(k)$ bits for some non-explicit predicate. (The reduction loses about $2k$ bits.)

In this paper, we further exploit the connection between CDS and PSM by observing that CDS protocols for a predicate $h(x,y)$ give rise to weakly private fully revealing PSM for the function $f((x \circ s),y) = h(x,y) \wedge s$, where $\circ$ denotes concatenation. By using our lower-bounds for weakly private fully revealing PSM's we get the following theorem. (See Section 7 for a proof.)

Theorem 5. *Let $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. Suppose that M upper-bounds the size of the largest 0-monochromatic rectangle of h and that for every $x \in \mathcal{X}$, the residual function $h(x,\cdot)$ is not the constant zero function. Then, the communication complexity of any perfect CDS for h is at least*

$$2\log|f^{-1}(0)| - \log M - \log|\mathcal{X}| - \log|\mathcal{Y}| - 1,$$

where $|f^{-1}(0)|$ denotes the number of inputs (x,y) that are mapped to zero.

Unlike the non-explicit lower-bound of [4], the above theorem provides a simple and clean sufficient condition for proving non-trivial CDS lower-bounds. For example, we can easily show that a random function has at least linear CDS complexity.

Corollary 2. *For all but a $o(1)$ fraction of the predicates $h : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$, any perfect CDS for h has communication of at least $k-4-o(1)$.*

Proof. Let $h : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ be a randomly chosen predicate. Let $K = 2^k$ and let $\epsilon = 1/\sqrt{K}$. There are exactly $2^K \cdot 2^K = 2^{2K}$ rectangles. Therefore, by a union-bound, the probability of having a 0-monochromatic rectangle of size $M = 2K(1+\epsilon)$ is at most

$$2^{2K} \cdot 2^{-M} = 2^{-2\epsilon K} = 2^{-\Omega(\sqrt{K})}.$$

Also, since h has K^2 inputs, the probability of having less than $(\frac{1}{2}-\epsilon)\cdot K^2$ unsatisfying inputs is, by a Chernoff bound, $2^{-\Omega(\epsilon^2 K^2)} = 2^{-\Omega(K)}$. Finally, by the union bound, the probability that there exists $x \in \mathcal{X}$ for which $h(x,\cdot)$ is the all-zero function is at most $K \cdot 2^{-K}$. It follows, by Theorem 5, that with probability of $1 - 2^{-\Omega(\sqrt{K})}$, the function h has a CDS complexity of at least $k-4-o(1)$. □

We can also get lower-bounds for explicit functions. For example, Gay et al. [17] studied the CDS complexity of the binary inner product function $h(x,y) = \langle x,y \rangle$. They proved an upper-bound of $k+1$ bits and a lower-bound of $\Omega(\log k)$ bits, and asked as an open question whether a lower-bound of $\Omega(k)$ can be established. (The question was open even for the special case of linear CDS for which [17] proved an $\Omega(\sqrt{k})$ lower-bound). By plugging the inner-product predicate into Theorem 5, we conclude:

Corollary 3. *Any perfect CDS for the inner product predicate $h_{ip} : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ requires at least $k-3-o(1)$ bits of communication.*

Proof. It suffices to prove the lower bound for the restriction of inner-product in which $x \neq 0^n$. It is well known (cf. [26]) that the largest monochromatic rectangle is of size $M = 2^k$, and the number of "zero" inputs is exactly $S = 2^{2k-1} - 2^{k-1}$. Hence, Theorem 5 yields a lower-bound of $k - 3 - o(1)$. □

This lower-bound matches the $k + 1$ upper-bound up to a constant additive difference (of 4 bits). It also implies that in any ABE scheme for the inner-product function which is based on the dual system methodology [32] either the ciphertext or the secret-key must be of length $\Omega(k)$. (See [17] for discussion.)

Organization. Following some preliminaries (Section 2), we present the counter example for the FKN lower-bound (Section 3). We then analyze the communication complexity of perfect PSM (Section 4) and imperfect PSM (Section 5). Based on these results, we obtain PSM lower-bounds for random and explicit functions (Section 6), as well as CDS lower-bounds (Section 7).

Acknowledgement. We are grateful to the reviewers of Eurocrypt 2018 and the Journal of Cryptology for the helpful and valuable comments on our manuscript. We especially thank the anonymous JOC referee for detecting a flaw in the original proof of Claims 9 and 13. BA and MM are supported by the European Union's Horizon 2020 Programme (ERC-StG-2014-2020) under grant agreement no. 639813 ERC-CLC, the Check Point Institute for Information Security, and the Blavatnik Interdisciplinary Cyber Research Center. OS and MM are supported by ERC grant no. 639573.

2 Preliminaries

For a string (or a vector) x of length n, and indices $1 \leq i \leq j \leq n$, we let $x[i]$ denote the i-th entry of x, and let $x[i : j]$ denote the string $(x[i], x[i+1] \ldots, x[j])$. By convention, all logarithms are taken base 2.

Rectangles. Let $\mathcal{X} \subset \{0,1\}^*$ and $\mathcal{Y} \subset \{0,1\}^*$ be some finite domains. (Typically, the set of all binary strings of some specified length.) An *(ordered) rectangle* of size $m \times n$ over $\mathcal{X} \times \mathcal{Y}$ is a pair $\rho = (\mathbf{x}, \mathbf{y})$, where $\mathbf{x} = (x_1, \ldots, x_m) \subseteq \mathcal{X}^m$ and $\mathbf{y} = (y_1, \ldots, y_n) \subseteq \mathcal{Y}^n$ satisfy $x_i \neq x_j$ and $y_i \neq y_j$ for all $i \neq j$. We say that (x, y) belongs to ρ if $x = x_i$ and $y = y_j$ for some i, j (or by abuse of notation we simply write $x \in \mathbf{x}$ and $y \in \mathbf{y}$). The size of an $(m \times n)$-rectangle ρ is mn, and its density with respect to some probability distribution μ over $\mathcal{X} \times \mathcal{Y}$, is $\sum_{x \in \mathbf{x}, y \in \mathbf{y}} \mu(x, y)$. Let $\rho = (\mathbf{x}, \mathbf{y})$ and $\rho' = (\mathbf{x}', \mathbf{y}')$ be a pair of $(m \times n)$-rectangles. We say that ρ and ρ' are *x-disjoint* (resp., *y-disjoint*) if $x_i \neq x'_i$ for all $i \in \{1, \ldots, m\}$ (resp., if $y_j \neq y'_j$ for all $j \in \{1, \ldots, n\}$). We say that ρ and ρ' are disjoint if they are either x-disjoint or y-disjoint.

As an example, consider the three (2×3)-rectangles $\rho_1 = \big((1,2),(5,6,7)\big)$, $\rho_2 = \big((2,1),(6,5,4)\big)$, and $\rho_3 = \big((1,3),(7,5,6)\big)$. Among those, ρ_1 and ρ_3 are y-disjoint but not x-disjoint, ρ_2 and ρ_3 are x-disjoint but not y-disjoint, and

ρ_1 and ρ_2 are both x-disjoint and y-disjoint. Therefore, each of these pairs is considered to be disjoint.

If $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is a function and ρ a rectangle of size $m \times n$, we let $f_{[\rho]}$ be the matrix M of size $m \times n$ whose entry M_{ij} is $f(x_i, y_j)$. A rectangle ρ is *0-monochromatic* (resp., 1-monochromatic) if $f_{[\rho]}$ is the all-zero matrix (resp., all-one matrix). A rectangle ρ is *similar* to a rectangle ρ' (with respect to f) if $f_{[\rho]} = f_{[\rho']}$. A rectangle $(\mathbf{x} = (x_1, \ldots, x_m), \mathbf{y})$ is *complement similar* if it is similar to the rectangle $((\bar{x}_1, \ldots, \bar{x}_m), \mathbf{y})$, where $\bar{x}$ denotes the string x with its last bit flipped.

Probabilistic notation. We will use calligraphic letters $\mathcal{A}$, $\mathcal{B}$, ..., to denote finite sets. Lower case letters denote values from these sets, i.e., $x \in \mathcal{X}$. Upper case letters usually denote random variables (unless the meaning is clear from the context).

Given two random variables A and B over the same set $\mathcal{A}$, we use $\|A - B\|$ to denote their *statistical distance* $\|A - B\| = \frac{1}{2} \sum_{a \in \mathcal{A}} |\Pr[A = a] - \Pr[B = a]|$. The min-entropy of A, denoted by $H_\infty(A)$, is minus the logarithm of the probability of the most likely value of A, i.e., $-\log \max_{a \in \mathcal{A}} \Pr[A = a]$.

3 A Counterexample to the FKN lower-bound

Let $\mathbf{T}_0, \mathbf{T}_1$ be a pair of $(k-1) \times (k-1)$ non-singular matrices (over the binary field $\mathbb{F} = \mathrm{GF}[2]$) with the property that $\mathbf{T} = \mathbf{T}_0 + \mathbf{T}_1$ is also non-singular. (The existence of such matrices is guaranteed via a simple probabilistic argument.[10]) Define the mapping $L : \mathbb{F}^k \to \mathbb{F}^k$ by

$$x \mapsto (\mathbf{T}_{x[k]} \cdot x[1 : k-1]) \circ x[k],$$

where $\circ$ denotes concatenation. That is, if the last entry of x is zero then L applies $\mathbf{T}_0$ to the $k-1$ prefix $x' = x[1 : k-1]$ and extends the resulting $k-1$ vector by an additional 0 entry, and if $x[k] = 1$ then the prefix x' is sent to $\mathbf{T}_1 x'$ and the vector is extended by an additional 1 entry. Note that L is a bijection (since $\mathbf{T}_0, \mathbf{T}_1$ are non-singular). The function $f : \mathbb{F}^k \times \mathbb{F}^k \to \mathbb{F}$ is defined by

$$(x, y) \mapsto \langle L(x), y \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner-product function over $\mathbb{F}$.

In Section 3.1, we will prove that f satisfies the FKN conditions (described in Section 1.2).

[10] When $k-1$ is even, there is a simple deterministic construction: Take $\mathbf{T}_0$ (resp., $\mathbf{T}_1$) to be the upper triangular matrix (resp., lower triangular matrix) whose entries on and above main diagonal (resp., on and below the diagonal) are ones and all other entries are zero. It is not hard to verify that both matrices are non-singular. Also $\mathbf{T} = \mathbf{T}_0 + \mathbf{T}_1$ has a zero diagonal and ones in all other entries and so $\mathbf{T}$ has full rank if $k-1$ is even. The same construction can be used when $k-1$ is odd, at the expense of obtaining a matrix $\mathbf{T}$ with an almost full rank that has only minor affect on the parameter M obtained in Lemma 1.

Lemma 1. *The function f is (1) non-degenerate, (2) useful, and (3) its largest complement similar rectangle is of size at most $M = 2^{k+1}$.*

Recall that f is non-degenerate if for every distinct $x \neq x'$ (resp., $y \neq y'$) the residual functions $f(x, \cdot)$ and $f(x', \cdot)$ (resp., $f(\cdot, y')$ and $f(\cdot, y')$) are distinct. It is useful if $\Pr_{x,y}[f(x, y) = f(\bar{x}, y)] \geq \frac{1}{2}$, where $\bar{x}$ denotes the string x with its last entry flipped. Also, a rectangle $R = (\mathbf{x}, \mathbf{y})$ is complement similar if $f(x, y) = f(\bar{x}, y)$ for every $x \in \mathbf{x}, y \in \mathbf{y}$.

In Section 3.2 we will show that f admits a PSM with communication complexity of $2k + O(1)$.

Lemma 2. *The function f has a PSM protocol with communication complexity of $2k + 2$.*

Theorem 1 follows from Lemma 1 and Lemma 2.

3.1 f satisfies the FKN properties (Proof of Lemma 1)

(1) f is non-degenerate. Fix $x_1 \neq x_2 \in \mathbb{F}^k$ and observe that $L(x_1) \neq L(x_2)$ (since L is a bijection). Therefore there exists y for which $f(x_1, y) = \langle L(x_1), y\rangle \neq \langle L(x_2), y\rangle = f(x_2, y)$. (In fact this holds for half of y's). Similarly, for every $y_1 \neq y_2$ there exists $v \in \mathbb{F}^k$ for which $\langle v, y_1\rangle \neq \langle v, y_2\rangle$, and since L is a bijection we can take $x = L^{-1}(v)$ and get that $f(x, y_1) = \langle v, y_1\rangle \neq \langle v, y_2\rangle = f(x, y_2)$.

(2) f is useful. Choose $x' \overset{\$}{\leftarrow} \mathbb{F}^{k-1}$ and $y \overset{\$}{\leftarrow} \mathbb{F}^k$ and observe that $f(x' \circ 0, y) = f(x' \circ 1, y)$ if and only if

$$\langle \mathbf{T}x', y[1 : k-1]\rangle + y_k = 0,$$

which happens with probability $\frac{1}{2}$.

(3) The largest complement similar rectangle is of size at most 2^{k+1}. Fix some rectangle $R = (\mathbf{x}, \mathbf{y})$, where $\mathbf{x} = (x_1, \ldots, x_m) \in (\mathbb{F}^k)^m$ and $\mathbf{y} = (y_1, \ldots, y_n) \in (\mathbb{F}^k)^n$. We show that if R is complement similar then $mn \leq 2 \cdot 2^k$. Since R is complement similar for every $x \in \mathbf{x}, y \in \mathbf{y}$ it holds

$$f(x, y) = f(\bar{x}, y),$$

which by definition of f implies that

$$\langle \mathbf{T}x' \circ 1, y\rangle = 0,$$

where x' is the $(k-1)$ prefix of x. Let d be the dimension of the linear subspace spanned by the vectors in $\mathbf{x}$, and so $m \leq 2^d$. Since $\mathbf{T}$ has full rank, the dimension of the subspace V spanned by $\{(\mathbf{T}x[1 : k-1] \circ 1) : x \in \mathbf{x}\}$ is at least $d-1$. (We may lose 1 in the dimension due to the removal of the last entry of the vectors $x \in \mathbf{x}$.) Noting that every $y \in \mathbf{y}$ is orthogonal to V, we conclude that the dimension of the subspace spanned by $\mathbf{y}$ is at most $k - (d-1)$. It follows that $n \leq 2^{k-(d-1)}$ and so $mn < 2 \cdot 2^k$. □

3.2 PSM for f (Proof of Lemma 2)

Note that f can be expressed as applying the inner product to v and y where v can be locally computed based on x. Hence it suffices to construct a PSM for the inner-product function and let Alice compute v and apply the inner-product protocol to v. (This reduction is a special instance of the so-called substitution lemma of randomize encoding, cf. [22, 2].) Lemma 2 now follows from the following lemma.

Lemma 3. *The inner product function* $h_{ip} : \mathbb{F}^k \times \mathbb{F}^k \to \mathbb{F}$ *has a PSM protocol with communication complexity of* $2k + 2$.

A proof of the lemma appears[11] in [27, Corollary 3]. For the sake of self-containment we describe here an alternative proof.

Proof. We show a PSM $\Pi = (\Pi_A, \Pi_B, g)$ with communication $2k$ under the promise that the inputs of Alice and Bob, x, y, are both not equal to the all zero vector. To get a PSM for the general case, let Alice and Bob locally extend their inputs x, y to $k+1$-long inputs $x' = x \circ 1$ and $y' = y \circ 1$. Then run the protocol Π and at the end let Charlie flip the outcome. It is easy to verify that the reduction preserves correctness and privacy. Since the inputs are longer by a single bit the communication becomes $2(k+1)$ as promised.

We move on to describe the protocol Π. The common randomness consists of a random invertible matrix $\mathbf{R} \in \mathbb{F}^{k \times k}$. Given non-zero $x \in \mathbb{F}^k$, Alice outputs $a = \mathbf{R}x$ where x is viewed as a column vector. Bob, who holds $y \in \mathbb{F}^k$, outputs $b = y^T \mathbf{R}^{-1}$. Charlie outputs ba.

Prefect correctness is immediate: $(y^T \mathbf{R}^{-1}) \cdot (\mathbf{R}x) = y^T x$, as required. To prove perfect privacy, we use the following claim.

Claim 6. *Let* $x, y \in \mathbb{F}^k$ *be non-zero vectors and denote their inner-product by* z. *Then, there exists an invertible matrix* $\mathbf{M} \in \mathbb{F}^{k \times k}$ *for which* $\mathbf{M}e_1 = x$ *and* $v_z^T \mathbf{M}^{-1} = y^T$ *where* e_i *is the i-th unit vector, and* v_z *is taken to be* e_1 *if* $z = 1$ *and* e_k *if* $z = 0$.

Proof. Let us first rewrite the condition $v_z^T \mathbf{M}^{-1} = y^T$ as $v_z^T = y^T \mathbf{M}$. Let $V \subset \mathbb{F}^k$ be the linear subspace of all vectors that are orthogonal to y. Note that the dimension of V is $k-1$. We distinguish between two cases based on the value of z.

Suppose that $z = 0$, that is, $x \in V$ and $v_z = e_k$. Then set the first column of $\mathbf{M}$ to be x and choose the next $k-2$ columns $\mathbf{M}_2, \ldots, \mathbf{M}_{k-1}$ so that together with x they form a basis for V. Let the last column $\mathbf{M}_k$ be some vector outside V. Observe that the columns are linearly independent and so $\mathbf{M}$ is invertible. Also, it is not hard to verify that $\mathbf{M}e_1 = x$ and that $y^T \mathbf{M} = e_k^T$.

Next, consider the case where $z = 1$, that is, $x \notin V$ and $v_z = e_1$. Then, take $\mathbf{M}_1 = x$ and let the other columns $\mathbf{M}_2, \ldots, \mathbf{M}_k$ to be some basis for V. Since x is non-zero the columns of $\mathbf{M}$ are linearly independent. Also, $\mathbf{M}e_1 = x$ and $y^T \mathbf{M} = e_1^T$. The claim follows. □

[11] We thank the anonymous reviewer for pointing this out.

We can now prove perfect privacy. Fix some non-zero $x, y \in \mathbb{F}^k$ and let $z = \langle x, y\rangle$. We show that the joint distribution of the messages (A, B) depends only on z. In particular, (A, B) is distributed identically to $(\mathbf{R}e_1, v_b^T\mathbf{R}^{-1})$ where $\mathbf{R}$ a random invertible matrix. Indeed, letting $\mathbf{M}$ be the matrix guaranteed in Claim 6 we can write

$$(\mathbf{R}x, y^T\mathbf{R}^{-1}) = (\mathbf{R}(\mathbf{M}e_1), (v_z^T\mathbf{M}^{-1})\mathbf{R}^{-1}).$$

Noting that $\mathbf{T} = \mathbf{RM}$ is also a random invertible matrix (since the the set of invertible matrices forms a group) we conclude that the RHS is identically distributed to $\mathbf{T}e_1, v_z^T\mathbf{T}^{-1}$, as claimed. □

Remark 2. Overall the PSM for f has the following form: Alice sends $a = \mathbf{R} \cdot (L(x) \circ 1)$ and Bob sends $b = (y \circ 1)^T\mathbf{R}^{-1}$ where $\mathbf{R} \in \mathbb{F}^{(k+1)\times(k+1)}$ is a random invertible matrix. The privacy proof shows that if the input (x, y) is mapped to (a, b) for some $\mathbf{R}$ then for every (x', y') for which $f(x, y) = f(x', y')$, there exists $\mathbf{R}'$ under which the input (x', y') is mapped to (a, b) as well. Hence, there are collisions between non-sibling inputs. As explained in the introduction, this makes the FKN lower-bound inapplicable.

4 Lower bound for perfect PSM protocols

In this Section we will prove a lower bound for perfect PSM protocols.

Definition 3. *For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and distribution μ over the domain $\mathcal{X} \times \mathcal{Y}$ with marginals μ_A and μ_B, define*

$$\alpha(\mu) = \max_{(R_1,R_2)} \min(\mu(R_1), \mu(R_2)),$$

where the maximum ranges over all pairs of similar disjoint rectangles (R_1, R_2). We also define

$$\beta(\mu) = \min_z \Pr[\,(X, Y) \neq (X', Y') \mid f(X, Y) = f(X', Y') = z\,],$$

where (X, Y) and (X', Y') represent two independent samples from μ, and minimum is taken over all z's in the support of $f(X, Y)$. Finally, we say that f is non-degenerate *with respect to μ if for every $x \neq x'$ in the support of μ_A there exists some $y \in \mathcal{Y}$ for which $f(x, y) \neq f(x', y)$, and similarly for every $y \neq y'$ in the support of μ_B there exists some $x \in \mathcal{X}$ for which $f(x, y) \neq f(x, y')$.*

In Section 4.1, we prove the following key lemma.

Lemma 4. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Then the communication complexity of any perfect PSM protocol is at least*

$$\max_\mu \log(1/\alpha(\mu)) + H_\infty(\mu) - \log(1/\beta(\mu)) - 1,$$

where the maximum is taken over all (not necessarily product) distribution μ under which f is non-degenerate.

The lower-bound is meaningful as long as β is not too small. Intuitively, this makes sure that the privacy requirement (which holds only over inputs on which the function agrees) is not trivial to achieve under μ.

For the special case of a Boolean function f, we can use the uniform distribution over $\mathcal{X} \times \mathcal{Y}$ and prove Theorem 3 from the introduction (restated here for the convenience of the reader).

Theorem 7 (Thm 3 restated). *Let $\mathcal{X}, \mathcal{Y}$ be sets of size at least 3. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a non-degenerate function for which any pair of disjoint similar rectangles (R, R') satisfies $|R| \leq M$. Then, any perfect PSM for f has communication of at least $2(\log|\mathcal{X}| + \log|\mathcal{Y}|) - \log M - 2$.*

Proof. By the key lemma (Lemma 4), it suffices to show that, for the uniform distribution μ, we have

$$\alpha(\mu) \leq M/(|\mathcal{X}||\mathcal{Y}|), \qquad H_\infty(\mu) = \log|\mathcal{X}| + \log|\mathcal{Y}|, \qquad \text{and } \beta(\mu) \geq 1/2.$$

The first two inequalities are immediate. The third inequality follows by noting that for every possible outcome $z \in \{0,1\}$,

$$\Pr[\,(X,Y) \neq (X',Y') \mid f(X,Y) = f(X',Y') = z\,] = 1 - \frac{1}{N_z} \geq 1 - \frac{1}{|\mathcal{X}| - 1} \geq 1/2,$$

where N_z is the number of preimages of z under f. The inequality $N_z \geq |\mathcal{X}| - 1$ follows by observing that if $N_z \leq |\mathcal{X}| - 2$, then there must be a pair $x \neq x'$ for which the restricted functions $f(x,\cdot)$ and $f(x',\cdot)$ are both equal to $1 - z$. This contradicts the non-degeneracy of f. □

We note that the additive constant 2 in the theorem statement can be replaced by $1 + o_k(1)$ when the size of the domain $\mathcal{X} \times \mathcal{Y}$ grows with k.

Weakly private fully revealing PSM. We can also derive a lower-bound on the communication complexity of weakly private fully revealing PSM. We begin with a formal definition.

Definition 4 (Weakly Private Fully Revealing PSM). *A* weakly private fully revealing PSM *$\Pi = (\Pi_A, \Pi_B, g)$ for a function $f : \{0,1\}^{k_1} \times \{0,1\}^{k_2} \to \mathcal{Z}$ is a perfect PSM for the function $f' : \{0,1\}^{k_1} \times \{0,1\}^{k_2} \to \{0,1\}^{k_1-1} \times \{0,1\}^{k_2} \times \mathcal{Z}$ that takes (x,y) and outputs $(x[1:k_1-1], y, f(x,y))$, where $x[1:k_1-1]$ is the (k_1-1)-prefix of x.*

In the following, we say that f is *weakly non-degenerate* if for every x there exists y such that $f(x,y) \neq f(\bar{x},y)$. Recall that an input (x,y) is useful if $f(x,y) = f(\bar{x},y)$. We prove the following (stronger) version of Theorem 2 from the introduction.

Theorem 8. *Let $f : \{0,1\}^{k_1} \times \{0,1\}^{k_2} \to \{0,1\}$ be a weakly non-degenerate function. Let M be an upper-bound on the size of the largest complement similar* *rectangle of f and let U be a lower-bound on the number of useful inputs of f. Then, any weakly-private fully-revealing PSM for f has communication complexity of at least $2\log U - \log M - 2$. In particular, for all but an $o(1)$ fraction of the predicates $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ we get a lower-bound of $3k - 4 - o(1)$.*

Proof. Let f' be the function defined in Definition 4 based on f. We will prove a lower-bound on the communication complexity of any perfect PSM for f'. Let μ be the uniform distribution over the set of useful inputs. Since f is weakly non-degenerate the function f' is non-degenerate under μ. The first part of the theorem follows from Lemma 4 by noting that

$$\alpha(\mu) \leq M/U, \qquad \beta(\mu) = 1/2, \qquad \text{and} \qquad H_\infty(\mu) \geq \log U.$$

Indeed, the bound on α follows by noting that the complement similar rectangles are the only similar disjoint rectangles, and the bound on β follows by noting that every z in the support of $f'(\mu)$ has exactly two preimages for which μ assigns equal weights.

To prove the second ("in particular") part observe that, for a random function f, each pair of inputs (x, y) and $(\bar{x}, y)$ gets the same f-value with probability $\frac{1}{2}$ independently of other inputs. Hence, with all but $o(1)$ probability, a fraction of $\frac{1}{2} - o(1)$ of all 2^{2k-1} of the pairs is mapped to the same value, and so there will be $2^{2k-1}(1 - o(1))$ useful inputs. (Since each successful pair contributes two useful inputs.) Also, each M-size rectangle R is complement similar with probability 2^{-M}. By taking a union bound over all $2^{2^{k+1}}$ rectangles, we conclude that f has an $M = 2^{k+1}(1 + o(1))$-size complement similar rectangle with probability at most $2^{2^{k+1}-M} = o(1)$. We conclude that, all but an $o(1)$ fraction of the functions, do not have weakly-private fully-revealing PSM with complexity smaller than $3k - 4 - o(1)$. □

A direct combinatorial proof of this theorem (based on the FKN argument) appears in Appendix A.

4.1 Proof of the Key Lemma (Lemma 4)

Fix some function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and let $\Pi = (\Pi_A, \Pi_B, g)$ be a perfect PSM protocol for f. Let μ denote some distribution over the domain $\mathcal{X} \times \mathcal{Y}$ and assume that f is non-degenerate with respect to μ.

We will use a probabilistic version of the FKN proof. In particular, consider two independent executions of Π on inputs that are sampled independently from μ. We let X, Y and R (resp., X', Y' and R') denote the random variables that represent the inputs of Alice and Bob and their shared randomness in the first execution (resp., second execution), and let A and B (resp., A' and B') denote the random variables that represent the messages sent by Alice and Bob in the first execution (resp., second execution). Thus, we can for example write $\Pr[(A, B) = (A', B') \wedge X \neq X']$ to denote the probability that the messages in the two executions match while the two inputs for Alice are different.

To simplify notation somewhat, we define the following events:

$$\begin{aligned}
\mathcal{P}^{(=)} &:\equiv (A = A') \wedge (B = B') \\
\mathcal{I}^{(=)} &:\equiv (X = X') \wedge (Y = Y') \\
\mathcal{I}^{(\neq)} &:\equiv (X \neq X') \vee (Y \neq Y') \equiv \neg \mathcal{I}^{(=)} \\
\mathcal{F}^{(=)} &:\equiv f(X,Y) = f(X',Y')
\end{aligned}$$

(The notation $\mathcal{P}$ is chosen to indicate equivalence/inequivalence of *Protocol* message and $\mathcal{I}$ to indicate equivalence/inequivalence of the *Inputs*.) Our lower-bound follows from the following claims.

Claim 9. *The communication complexity of Π is at least* $\log(1/\Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)}]) - \log(1/\beta)$.

Proof. We upper-bound the collision probability $\Pr[(A,B) = (A',B')]$ of two random executions by showing that

$$\Pr[\mathcal{P}^{(=)}] \leq \frac{\Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)}]}{\beta}. \tag{2}$$

Because the collision probability of two independent instances of a random variable is at least the inverse of the alphabet size, the alphabet of A and B must have size at least $\beta / \Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)}]$. Thus, in total the protocol requires

$$\log(1/\Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)}]) - \log(1/\beta)$$

bits of communication.

We move on to prove (2). That is, we show that $\Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)}]$ is at least $\beta \Pr[\mathcal{P}^{(=)}]$. First, by perfect correctness $\mathcal{P}^{(=)}$ happens only when $\mathcal{F}^{(=)}$ happens, hence

$$\Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)}] = \Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)} \wedge \mathcal{F}^{(=)}] = \Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)} \mid \mathcal{F}^{=}] \Pr[\mathcal{F}^{=}].$$

Denoting by $\mathcal{F}^{=z}$ the event $f(X,Y) = f(X',Y') = z$, the RHS equals to

$$\sum_z \Pr[\mathcal{I}^{(\neq)} \wedge \mathcal{P}^{(=)} \mid \mathcal{F}^{=z}] \Pr[\mathcal{F}^{=z}], \tag{3}$$

where the sum ranges over all z's in the support Z of $f(X,Y)$. Below we argue that, by perfect privacy, for every $z \in Z$, the conditional random variables $[\mathcal{I}^{(\neq)} \mid \mathcal{F}^{=z}]$ and $[\mathcal{P}^{(=)} \mid \mathcal{F}^{=z}]$ are statistically independent. Hence, (3) equals to

$$\begin{aligned}
\sum_z \Pr[\mathcal{I}^{(\neq)} \mid \mathcal{F}^{=z}] \Pr[\mathcal{P}^{(=)} \mid \mathcal{F}^{=z}] \Pr[\mathcal{F}^{=z}] &\geq \sum_z \beta \Pr[\mathcal{P}^{(=)} \mid \mathcal{F}^{=z}] \Pr[\mathcal{F}^{=z}] \\
&= \beta \Pr[\mathcal{P}^{(=)}],
\end{aligned}$$

(recall that $\beta = \min_z \Pr[\mathcal{I}^{(\neq)} \mid \mathcal{F}^{=z}]$), and we established (2).

It remains to show that, for every $z \in Z$, the conditional random variables $[\mathcal{I}^{(\neq)} \mid \mathcal{F}^{=z}]$ and $[\mathcal{P}^{(=)} \mid \mathcal{F}^{=z}]$ are statistically independent. To see this, note that by privacy, for every fixing of (X, Y, X', Y') to (x, y, x', y') for which $f(x, y) = f(x', y') = z$, the distribution of the random variables $[(A, B, A', B') \mid X = x, Y = y, X' = x', Y' = y']$ remains the same regardless of the values of (x, y, x', y'). Hence, conditioning on $\mathcal{F}^{(=z)}$, the random variable (A, B, A', B') is independent of (X, Y, X', Y'), which implies that $[\mathcal{I}^{(\neq)} \mid \mathcal{F}^{=z}]$ and $[\mathcal{P}^{(=)} \mid \mathcal{F}^{=z}]$ are statistically independent. This completes the proof of Claim 9. □

Claim 10. *For any pair of strings r and r',*

$$\Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} | R = r, R' = r'] \leq 2\alpha(\mu) 2^{-H_\infty(\mu)}.$$

Proof. We see that

$$\begin{aligned}\Pr\big[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} | R = r \wedge R' = r'\big] \leq \Pr\big[\mathcal{P}^{(=)} \wedge (X \neq X') | R = r \wedge R' = r'\big] \\ + \Pr\big[\mathcal{P}^{(=)} \wedge (Y \neq Y') | R = r \wedge R' = r'\big] .\end{aligned}$$

Due to symmetry it suffices to bound the first summand by $\alpha(\mu)2^{-H_\infty(\mu)}$.

Say that x *collides* with x' if $\mathit{\Pi}_A(x, r) = \mathit{\Pi}_A(x', r')$. Restricting our attention to x's in the support of μ_A, we claim that every x can collide with at most a single x'. Indeed, if this is not the case, then $\mathit{\Pi}_A(x, r) = \mathit{\Pi}_A(x', r') = \mathit{\Pi}_A(x'', r')$. The second equality implies that when the randomness is r', for every y, the messages (a, b) communicated under (x', y) are equal to the ones communicated under (x'', y). By perfect correctness, this implies that $f(x', y) = f(x'', y)$ for every y, contradicting the non-degeneracy of f under μ. Analogously, let us say that y collides with y' if $\mathit{\Pi}_B(y, r) = \mathit{\Pi}_B(y', r')$. The same reasoning shows that every y in the support of μ_B can collide with at most a single y' in the support of μ_B.

Let $\mathbf{x} = (x_1, \ldots, x_m)$ and $\mathbf{x}' = (x'_1, \ldots, x'_m)$ be a complete list of entries for which x_i collides with x'_i and $x_i \neq x'_i$ and $\mu_A(x_i), \mu_A(x'_i) > 0$. Analogously let $\mathbf{y} = (y_1, \ldots, y_n)$ and $\mathbf{y}' = (y'_1, \ldots, y'_n)$ be a complete list for which y_i collides with y'_i and $\mu_B(y_i), \mu_B(y'_i) > 0$. (Note that we do not require $y_i \neq y'_i$.) Since collisions are unique (as explained above), the tuples $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}'$ are uniquely determined up to permutation.

By definition, the tuples (x, y, x', y') with $x \neq x'$, and $(a, b) = (a', b')$ are exactly those of the form (x_i, y_j, x'_i, y'_j) for some i and j.

Now, consider the two x-disjoint rectangles $\rho = (\mathbf{x}, \mathbf{y})$ and $\rho' = (\mathbf{x}', \mathbf{y}')$ and assume, without loss of generality, that $\mu(\rho) \leq \mu(\rho')$. Since Alice and Bob both send the same messages with randomness r on inputs (x_i, y_j) as they send with randomness r' on inputs x'_i, y'_j, we see that it must be that $f(x_i, y_j) = f(x'_i, y'_j)$ if the protocol is correct. Therefore, $f_{[\rho]} = f_{[\rho']}$, and so $\mu(\rho) \leq \alpha(\mu)$.

To complete the argument, note that $\mathcal{P}^{(=)} \wedge (X \neq X')$ can only happen if we pick $(X, Y) = (x_i, y_j)$ and $(X', Y') = (x'_i, y'_j)$ for some i, j. The event that there exists i, j for which $(X, Y) = (x_i, y_j)$ has probability at most $\alpha(\mu)$. The event that $(X', Y') = (x'_i, y'_j)$ for the same (i, j) has probability at most $\max_{x,y} \mu(x, y) = 2^{-H_\infty(\mu)}$. □

Claim 10 implies that

$$\begin{aligned}\Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)}] &= \sum_{r,r'} \Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} | R = r, R' = r'] \Pr[R = r, R' = r'] \\ &\leq \sum_{r,r'} 2\alpha(\mu) 2^{-H_\infty(\mu)} \Pr[R = r, R' = r'] \\ &\leq 2\alpha(\mu) 2^{-H_\infty(\mu)}.\end{aligned}$$

Combined with Claim 9, this yields Lemma 4. □

5 Lower bounds for imperfect PSM protocols

In this section we prove a lower-bound on the communication complexity of imperfect PSM protocols. For this, we will have to strengthen the requirements from the function f.

Definition 5. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$.*

- *We call f* somewhat regular *if for every possible image $z \in \mathcal{Z}$ the number of preimages $f^{-1}(z)$ is at least $0.9\frac{|\mathcal{X}|\cdot|\mathcal{Y}|}{|\mathcal{Z}|}$ and at most $1.1\frac{|\mathcal{X}|\cdot|\mathcal{Y}|}{|\mathcal{Z}|}$.*
- *We call f* strongly non-degenerate *if for any $x \neq x'$ we have $|\{y | f(x, y) = f(x', y)\}| \leq 0.9|\mathcal{Y}|$ and for any $y \neq y'$ we have $|\{x | f(x, y) = f(x, y')\}| \leq 0.9|\mathcal{X}|$.*
- *A pair of ordered $m \times n$ rectangles $\rho = (\mathbf{x}, \mathbf{y})$ and $\rho' = (\mathbf{x}', \mathbf{y}')$ in which either $x_i \neq x'_i$ for all $i \in [m]$, or $y_i \neq y'_i$ for all $i \in [n]$ are called* approximately similar *if for 0.99 of the pairs (i, j) we have $f(x_i, y_j) = f(x'_i, y'_j)$.*

(All the constants in the above definition are somewhat arbitrary and other constants may be chosen.)

In Section 5.1 we prove the following theorem:

Theorem 11. *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a somewhat-regular, strongly non-degenerate function whose largest approximately similar pair of rectangles is of size at most M. Then, any PSM for f with privacy error of ϵ and correctness error of $\delta < \frac{1}{100}$, requires at least*

$$\log|\mathcal{X}| + \log|\mathcal{Y}| + \min \left\{ \begin{array}{l} \log|\mathcal{X}| + \log|\mathcal{Y}| - \log\left(\frac{1}{\Pr[\mathcal{F}^{(=)}]}\right), \\ \log|\mathcal{X}| + \log|\mathcal{Y}| - \log M, \\ \log(1/\epsilon), \\ \log(1/\delta) - \log\left(\frac{1}{\Pr[\mathcal{F}^{(=)}]}\right) \end{array} \right\} - c \qquad (4)$$

bits of communication, where c is some universal constant (that does not depend on f) and $\Pr[\mathcal{F}^{(=)}] = \Pr[f(X, Y) = f(X', Y')]$ when (X, Y) and (X', Y') are picked independently and uniformly at random from $\mathcal{X} \times \mathcal{Y}$.

In the special case of a Boolean function f, it holds that $\Pr[\mathcal{F}^{(=)}] = \Pr[f(X,Y) = f(X',Y')] \geq 1/2$, and the communication lower-bound simplifies to

$$\log|\mathcal{X}| + \log|\mathcal{Y}| + \min\{\log|\mathcal{X}| + \log|\mathcal{Y}| - \log M, \log(1/\epsilon), \log(1/\delta)\} - c$$

where c is some universal constant. In Section 6, we will use Theorem 11 to prove imperfect PSM lower-bounds for random functions and for efficiently computable functions.

5.1 Proof of Theorem 11

In the following, all random variables are defined as in Section 4 where μ is simply the uniform distribution over $\mathcal{X} \times \mathcal{Y}$. In order to analyze protocols that are not perfectly private or perfectly correct, we first study the mistakes the protocol does in a few more details. Fix a PSM protocol $\Pi = (\Pi_A, \Pi_B, g)$ for f with correctness error of $\delta < 1/100$ and privacy error of ϵ. We let $\beta(x,r)$ be the probability that $f(X,Y) \neq g(A,B)$ conditioned on $(X,R) = (x,r)$, and (by abuse of notation) define $\beta(y,r)$ analogously. That is,

$$\begin{aligned}\beta(x,r) &:= \Pr[f(X,Y) \neq g(A,B) | (X,R) = (x,r)],\\ \beta(y,r) &:= \Pr[f(X,Y) \neq g(A,B) | (Y,R) = (y,r)].\end{aligned}$$

We then consider the event

$$\mathcal{H}_\delta :\equiv \Big(\beta(X,R) \leq 5\delta\Big) \wedge \Big(\beta(X',R') \leq 5\delta\Big) \wedge \Big(\beta(Y,R) \leq 5\delta\Big) \wedge \Big(\beta(Y',R') \leq 5\delta\Big)\ .$$

Intuitively, $\mathcal{H}_\delta$ occurs if what the players see is unlikely to produce an error.

Claim 12. $\Pr[\mathcal{H}_\delta] \geq \frac{1}{5}$.

Proof. We compute

$$\begin{aligned}\Pr[\neg\mathcal{H}_\delta] &\leq 2\Pr[\beta(X,R) > 5\delta \vee \beta(Y,R) > 5\delta]\\ &\leq 2\Pr[\beta(X,R) > 5\delta] + 2\Pr[\beta(Y,R) > 5\delta],\end{aligned}$$

where we used the union bound. Because

$$\begin{aligned}\Pr[f(X,Y) \neq g(A,B)] &\geq \Pr[\beta(X,R) > 5\delta] \cdot \Pr[f(X,Y) \neq g(A,B) | \beta(X,R) > 5\delta]\\ &> \Pr[\beta(X,R) > 5\delta] \cdot 5\delta,\end{aligned}$$

and similarly for $\Pr[\beta(Y,R) > 5\delta]$, we get $\Pr[\neg\mathcal{H}_\delta] \leq 4\frac{\Pr[f(X,Y) \neq g(A,B)]}{5\delta} \leq \frac{4}{5}$. □

Our main goal is to derive an upper-bound on the conditional probability $\Pr[\mathcal{P}^{(=)} | \mathcal{H}_\delta]$. Since the event $\mathcal{H}_\delta$ represents a restriction of the alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{R}$, the conditional collision probability $\Pr[\mathcal{P}^{(=)} | \mathcal{H}_\delta]$ is indicative of only a subset of the alphabet $\mathcal{A} \times \mathcal{B}$. That is,

$$\Pr[\mathcal{P}^{(=)} | \mathcal{H}_\delta] \geq \frac{1}{|\mathcal{A} \times \mathcal{B}|}.$$

Therefore, an upper-bound on the conditional probability $\Pr[\mathcal{P}^{(=)}|\mathcal{H}_\delta]$ translates into a lower-bound on the communication complexity.

Now,

$$\Pr[\mathcal{P}^{(=)} \wedge \mathcal{H}_\delta] = \Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(=)} \wedge \mathcal{H}_\delta] + \Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} \wedge \mathcal{H}_\delta] \, .$$

We upper-bound each of the two summands. In Section 5.3 we show that

$$\Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(=)} \wedge \mathcal{H}_\delta] \le \frac{1}{|\mathcal{X}||\mathcal{Y}|}\left(\frac{1.5}{\Pr[\mathcal{F}^{(=)}]} \cdot \left(\frac{1}{|\mathcal{X}||\mathcal{Y}|} + 60\delta\right) + \epsilon\right). \tag{5}$$

In Section 5.4 we prove that

$$\Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} \wedge \mathcal{H}_\delta] \le \max\left\{\tilde{c}\frac{\delta^2}{|\mathcal{X}||\mathcal{Y}|}, \frac{M}{|\mathcal{X}|^2|\mathcal{Y}|^2}\right\}, \tag{6}$$

where $\tilde{c}$ is a universal constant. We conclude that

$$\Pr[\mathcal{P}^{(=)} \wedge \mathcal{H}_\delta] \le \frac{O(1)}{|\mathcal{X}||\mathcal{Y}|} \cdot \max\left\{\frac{1}{|\mathcal{X}||\mathcal{Y}|\Pr[\mathcal{F}^{(=)}]}, \frac{\delta}{\Pr[\mathcal{F}^{(=)}]}, \epsilon, \frac{M}{|\mathcal{X}||\mathcal{Y}|}\right\}.$$

Since $\Pr[\mathcal{H}_\delta] \ge \frac{1}{5}$, we get:

$$\Pr[\mathcal{P}^{(=)}|\mathcal{H}_\delta] \le \frac{O(1)}{|\mathcal{X}||\mathcal{Y}|} \cdot \max\left\{\frac{1}{|\mathcal{X}||\mathcal{Y}|\Pr[\mathcal{F}^{(=)}]}, \frac{\delta}{\Pr[\mathcal{F}^{(=)}]}, \epsilon, \frac{M}{|\mathcal{X}||\mathcal{Y}|}\right\}$$

Thus,

$$\begin{aligned}|\mathcal{A} \times \mathcal{B}| &\ge \frac{1}{\Pr[\mathcal{P}^{(=)}|\mathcal{H}_\delta]} \\ &\ge |\mathcal{X}||\mathcal{Y}| \cdot \Omega\left(\min\left\{|\mathcal{X}||\mathcal{Y}|\Pr[\mathcal{F}^{(=)}], \frac{|\mathcal{X}||\mathcal{Y}|}{M}, \frac{\Pr[\mathcal{F}^{(=)}]}{\delta}, \frac{1}{\epsilon}\right\}\right),\end{aligned}$$

which proves Theorem 11. □

5.2 A useful fact

The proofs of Equations (5) and (6) both exploit the non-degeneracy of f via the following simple fact.

Fact 13. *Suppose that f is strongly non-degenerate and that (Π_A, Π_B, g) is a PSM for f. Suppose that there is a set $\mathbf{x} \subset \mathcal{X}$ of at least k distinct inputs whose A-message collide under some string r, that is, $\Pi_A(x,r) = \Pi_A(x',r)$ for any $x, x' \in \mathbf{x}$. Then, for at least $k-1$ of the inputs $x \in \mathbf{x}$'s, the protocol conditioned on r errs for at least $1/20$ fraction of the y's, i.e.,*

$$\Pr_{y \xleftarrow{\$} \mathcal{Y}}[g(\Pi_A(x,r), \Pi_B(y,r)) \neq f(x,y)] \ge 1/20.$$

An analogues fact holds for collisions over the B-messages.

Proof. Assume, towards a contradiction, that there is a pair of distinct inputs $x \neq x' \in \mathbf{x}$, for which

$$\Pr_{y \overset{\$}{\leftarrow} \mathcal{Y}}[g(\Pi_A(x,r), \Pi_B(y,r)) \neq f(x,y)] < 1/20,$$

and

$$\Pr_{y \overset{\$}{\leftarrow} \mathcal{Y}}[g(\Pi_A(x',r), \Pi_B(y,r)) \neq f(x',y)] < 1/20,$$

since $\Pi_A(x,r) = \Pi_A(x',r)$, we conclude that $f(x,y)$ and $f(x',y)$ disagree on less than 0.1-fraction of the y's, which contradicts the fact that f is strongly non-degenerate. □

5.3 Proof of Eq. (5)

Eq. (5) follows by combining the following two claims.

Claim 14. *Suppose that f is somewhat regular and that (Π_A, Π_B, g) is a PSM for f with privacy error of ϵ. Then,*

$$\Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(=)} \wedge \mathcal{H}_\delta] \leq \frac{1}{|\mathcal{X}||\mathcal{Y}|}\left(1.5\Pr[\mathcal{P}^{(=)}|\mathcal{F}^{(=)}] + \epsilon\right).$$

Proof. First,

$$\begin{aligned}\Pr[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(=)} \wedge \mathcal{H}_\delta] &\leq \Pr[\mathcal{I}^{(=)} \wedge \mathcal{P}^{(=)}] \\ &= \Pr[\mathcal{I}^{(=)}]\Pr[\mathcal{P}^{(=)} \mid \mathcal{I}^{(=)}] \\ &= \frac{1}{|\mathcal{X}||\mathcal{Y}|}\Pr[\mathcal{P}^{(=)} \mid \mathcal{I}^{(=)}]. \end{aligned} \tag{7}$$

To establish the claim it suffices to show that

$$\Pr[\mathcal{P}^{(=)} \mid \mathcal{I}^{(=)}] \leq \epsilon + 1.5\left(\Pr[\mathcal{P}^{(=)} \mid \mathcal{F}^{(=)}]\right). \tag{8}$$

By Bayes' rule, we can rewrite (8) as

$$\sum_{x,y} p(x,y)q(x,y) \leq \epsilon + 1.5\left(\sum_{x,y} p'(x,y)q'(x,y)\right). \tag{9}$$

where

$$p(x,y) := \Pr[\mathcal{P}^{(=)} \mid \mathcal{I}^{(=)}, (X,Y) = (x,y)], \quad q(x,y) := \Pr[(X,Y) = (x,y) \mid \mathcal{I}^{(=)}],$$

and

$$p'(x,y) := \Pr[\mathcal{P}^{(=)} \mid \mathcal{F}^{(=)}, (X,Y) = (x,y)], \quad q'(x,y) := \Pr[(X,Y) = (x,y) \mid \mathcal{F}^{(=)}].$$

In the remainder of the proof, we establish (9), by relating, for every (x, y), the value of $p(x, y)$ to $p'(x.y)$ and the value of $q(x, y)$ to $q'(x, y)$.

First, observe that conditioning on $\mathcal{I}^{(=)}$ does not affect the distribution of (X, Y), hence

$$\forall (x, y): \quad q(x, y) = \Pr[(X, Y) = (x, y)] = 1/N, \tag{10}$$

where $N = |\mathcal{X}| \cdot |\mathcal{Y}|$. Also, letting N_z denote the number of preimages of $z \in \mathcal{Z}$ under f, it holds that

$$\forall (x, y): \quad q'(x, y) = \frac{N_{f(x,y)}}{\sum_z N_z^2} \geq \frac{0.9(N/|\mathcal{Z}|)}{|\mathcal{Z}|(1.1)^2 N^2/|\mathcal{Z}|^2} = \frac{0.9}{(1.1)^2 N}, \tag{11}$$

where the inequality follows from the fact that f is somewhat regular. (Recall that "somewhat regularity" just means that, for every z, $N_z \in [0.9\frac{N}{|\mathcal{Z}|}, 1.1\frac{N}{|\mathcal{Z}|}]$.) By combining Equations (10),(11), we get that

$$\forall (x, y): \quad q(x, y) \leq \frac{(1.1)^2}{0.9} q'(x, y) < 1.5 q'(x, y). \tag{12}$$

We will later use privacy to show that

$$\forall (x, y): \quad p(x, y) \leq p'(x, y) + \epsilon. \tag{13}$$

By combining Equations (12), and (13), we get

$$\begin{aligned}
\sum_{x,y} p(x, y) q(x, y) &\leq \sum_{x,y} (p'(x, y) + \epsilon) q(x, y) \\
&= \sum_{x,y} p'(x, y) q(x, y) + \epsilon \sum_{x,y} p'(x, y) \\
&\leq 1.5 \left(\sum_{x,y} p'(x, y) q'(x, y) \right) + \epsilon,
\end{aligned}$$

and (9) follows.

It remains to prove (13). Since the event $\mathcal{P}^{(=)}$ is a function of (A, B, A', B'), the difference $|p(x, y) - p'(x, y)|$ is upper-bounded by the statistical distance

$$\begin{aligned}
&\Big\| \Big((A, B, A', B') \mid (X, Y) = (x, y), \mathcal{I}^{(=)} \Big) \\
&\qquad - \Big((A, B, A', B') \mid (X, Y) = (x, y), \mathcal{F}^{(=)} \Big) \Big\|.
\end{aligned} \tag{14}$$

We show that, for every (x, y), (14) is at most ϵ. In fact, we prove a stronger statement: for every (x, y) and (x', y') for which $f(x, y) = f(x', y')$, the random variable

$$[(A, B, A', B') \mid (X, Y) = (x, y), (X', Y') = (x, y)]$$

is ϵ-close in statistical distance to

$$[(A, B, A', B') \mid (X, Y) = (x, y), (X', Y') = (x', y')].$$

Indeed, in both cases the (A, B)-part is statistically independent of (A', B'), and therefore the statistical distance is just the statistical distance between

$$[(A', B') \mid (X', Y') = (x, y)] \quad \text{and} \quad [(A', B') \mid (X', Y') = (x', y')]$$

which is at most ϵ by ϵ-privacy. This completes the proof of the claim. □

By Claim 14, in order to prove Eq. (5) it suffices to prove the following claim.

Claim 15. *Suppose f is strongly non-degenerate and (Π_A, Π_B, g) is a PSM with correctness error of δ. Then,*

$$\Pr[\mathcal{P}^{(=)}|\mathcal{F}^{(=)}] \leq \frac{1}{\Pr[\mathcal{F}^{(=)}]} \cdot \left(\frac{1}{|\mathcal{X}||\mathcal{Y}|} + 60\delta\right).$$

Proof. We define $k(r, a) := |\{x|\Pi_A(x, r) = a\}|$ and $k(r, b) := |\{y|\Pi_B(y, r) = b\}|$, the number of inputs to the respective part of a protocol that map to a, respectively to b, for a fixed randomness r.

We first bound $\Pr[\mathcal{P}^{(=)}]$. Fix randomness r, r', and x, y, and let $a = \Pi_A(x, r)$, $b = \Pi_B(y, r)$. We see that

$$\Pr[\mathcal{P}^{(=)}|R = r, R' = r', X = x, Y = y] = \frac{k(r', a)k(r', b)}{|\mathcal{X}||\mathcal{Y}|}.$$

We can write the RHS as

$$\frac{1}{|\mathcal{X}||\mathcal{Y}|} + \frac{(k(r', a) - 1)|\mathcal{Y}| \cdot (k(r', b) - 1)|\mathcal{X}|}{|\mathcal{X}|^2|\mathcal{Y}|^2} + \frac{(k(r', a) - 1)|\mathcal{Y}|}{|\mathcal{X}||\mathcal{Y}|^2} + \frac{(k(r', b) - 1)|\mathcal{X}|}{|\mathcal{X}|^2|\mathcal{Y}|}.$$

Thus, by letting $k(r, a, b) = \max\{(k(r, a) - 1)|\mathcal{Y}|, (k(r, b) - 1)|\mathcal{X}|\}$, we get

$$\begin{aligned}\Pr[\mathcal{P}^{(=)}|R = r, R' = r', X = x, Y = y]& \\ \leq \frac{1}{|\mathcal{X}||\mathcal{Y}|} + \frac{(k(r', a, b))^2}{|\mathcal{X}|^2|\mathcal{Y}|^2} + \frac{k(r', a, b)}{|\mathcal{X}||\mathcal{Y}|^2} + \frac{k(r', a, b)}{|\mathcal{X}|^2|\mathcal{Y}|}.& \qquad (15)\end{aligned}$$

Because f is strongly non-degenerate, we see that for any a, b, we have

$$\begin{aligned}\Pr[f(X', Y') \neq g(A', B')|R' = r']& \\ &\geq \max\left\{\frac{1}{20} \cdot \frac{(k(r', a) - 1)}{|\mathcal{X}|}, \frac{1}{20} \cdot \frac{(k(r', b) - 1)}{|\mathcal{Y}|}\right\} \\ &= \max\left\{\frac{1}{20} \cdot \frac{(k(r', a) - 1) \cdot |\mathcal{Y}|}{|\mathcal{X}||\mathcal{Y}|}, \frac{1}{20} \cdot \frac{(k(r', b) - 1)|\mathcal{X}|}{|\mathcal{X}||\mathcal{Y}|}\right\} \\ &= \frac{1}{20|\mathcal{X}||\mathcal{Y}|}k(r', a, b).\end{aligned}$$

Indeed, if $k(r', a) > 1$, then by Fact 13, for at least $k(r', a) - 1$ of the choices of x' that map to a, the protocol makes an error for at least $\frac{|\mathcal{Y}|}{20}$ choices of y' (and similarly, for the symmetric case where $k(r', b) > 1$). Multiplying both sides by $\Pr[R' = r']$ and summing up the inequalities over all r', we get:

$$\Pr[f(X', Y') \neq g(A', B')] \geq \frac{1}{20|\mathcal{X}||\mathcal{Y}|} \mathsf{E}[k(R', a, b)].$$

Since the PSM is δ-correct, the LHS is at most δ, and we get:

$$20\delta \geq \frac{\mathsf{E}(k(R', a, b))}{|\mathcal{X}||\mathcal{Y}|}.$$

Plugging this into (15), and taking expectations over r, r', x, y, we conclude that

$$\begin{aligned} \Pr[\mathcal{P}^{(=)}] &\leq \frac{1}{|\mathcal{X}||\mathcal{Y}|} + 20\delta + \frac{20\delta}{|\mathcal{Y}|} + \frac{20\delta}{|\mathcal{X}|} \\ &\leq \frac{1}{|\mathcal{X}||\mathcal{Y}|} + 60\delta \end{aligned}$$

Since $\Pr[\mathcal{P}^{(=)}|\mathcal{F}^{(=)}] \leq \Pr[\mathcal{P}^{(=)}]/\Pr[\mathcal{F}^{(=)}]$, we get the claim. □

5.4 Proof of Eq. (6)

Claim 16. *Suppose that f is a strongly non-degenerate function whose largest approximately similar pair of rectangles is of size at most M, and assume that (Π_A, Π_B, g) is a PSM protocol for f with correctness error of $\delta < \frac{1}{100}$. Then, for any fixed choice r and r' of the randomness, we have*

$$\Pr\left[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} \wedge \mathcal{H}_\delta \mid R = r \wedge R' = r'\right] \leq \max\left\{\tilde{c}\frac{\delta^2}{|\mathcal{X}||\mathcal{Y}|}, \frac{M}{|\mathcal{X}|^2|\mathcal{Y}|^2}\right\},$$

where $\tilde{c}$ is some universal constant.

Proof. We see that $\Pr\left[\mathcal{P}^{(=)} \wedge \mathcal{I}^{(\neq)} \wedge \mathcal{H}_\delta | R = r \wedge R' = r'\right]$ is at most

$$\begin{aligned} &\Pr\left[\mathcal{P}^{(=)} \wedge (X \neq X') \wedge \mathcal{H}_\delta | R = r \wedge R' = r'\right] \\ &\quad + \Pr\left[\mathcal{P}^{(=)} \wedge (Y \neq Y') \wedge \mathcal{H}_\delta | R = r \wedge R' = r'\right]. \end{aligned}$$

Due to symmetry it suffices to bound the first summand.

As in the proof of the perfect case, we say that x and x' are *colliding* (under r and r') if $\Pi_A(x, r) = \Pi_A(x', r')$. We restrict our attention to inputs x and x' for which $\beta(x, r) \leq 5\delta$ and $\beta(x', r') \leq 5\delta$.

Claim. *Suppose that $\beta(x, r) \leq 5\delta$. Then x can collide with at most a single x' for which $\beta(x', r') \leq 5\delta$.*

Proof. If this is not the case, then there exist $x' \neq x''$ for which $\Pi_A(x, r) = \Pi_A(x', r') = \Pi_A(x'', r')$ and $\beta(x', r'), \beta(x'', r') \leq 5\delta$. By Fact 13, either $\beta(x', r')$ or $\beta(x'', r')$ must be at least $\frac{1}{20}$ which is larger than 5δ for $\delta < 1/100$. Thus we have reached a contradiction. □

A similar conclusion for $y, y' \in \mathcal{Y}$, holds as well. That is, for every $y \in \mathcal{Y}$ with $\beta(y, r) \leq 5\delta$ there exists at most a single y' with $\beta(y', r') \leq 5\delta$ for which a collision $\varPi_B(y, r) = \varPi_B(y', r')$ occurs.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}' = (x'_1, \ldots, x'_n)$ be a complete list of colliding elements for which all of the following hold for all i:

$$\begin{aligned} \varPi_A(x_i, r) &= \varPi_A(x'_i, r') & x_i &\neq x'_i \\ \beta(x_i, r) &\leq 5\delta, & \beta(x'_i, r') &\leq 5\delta \ . \end{aligned}$$

Analogously, let $\mathbf{y} = (y_1, \ldots, y_m)$ and $\mathbf{y}' = (y'_1, \ldots, y'_m)$ be a complete list of colliding elements for which

$$\varPi_B(y_i, r) = \varPi_B(y'_i, r'), \quad \beta(y_i, r) \leq 5\delta, \quad \text{and } \beta(y'_i, r') \leq 5\delta \ .$$

(In the latter case, pairs (y_i, y'_i) are allowed to be equal.) By definition, the resulting rectangles $\rho = (\mathbf{x}, \mathbf{y})$ and $\rho' = (\mathbf{x}', \mathbf{y}')$ are x-disjoint. Note that Alice and Bob both send the same messages with randomness r on inputs (x_i, y_j) as they send with randomness r' on inputs x'_i, y'_j, for any (i, j). Also, by the above claim, the tuples $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}'$ are uniquely determined up to permutation. We conclude that the event $\mathcal{P}^{(=)} \wedge (X \neq X') \wedge \mathcal{H}_\delta$ happens if and only if we pick $(X, Y) = (x_i, y_j)$ from ρ and $(X', Y') = (x'_i, y'_j)$ from ρ', for some (i, j). The probability of this is $\frac{mn}{|\mathcal{X}||\mathcal{Y}|} \cdot \frac{1}{|\mathcal{X}||\mathcal{Y}|}$.

If $n \leq 200 \cdot 5\delta \cdot |\mathcal{X}|$ and $m \leq 200 \cdot 5\delta \cdot |\mathcal{Y}|$, then $\frac{mn}{|\mathcal{X}|^2|\mathcal{Y}|^2}$ is at most $\frac{40\,000 \cdot 25\delta^2}{|\mathcal{X}||\mathcal{Y}|}$, and we are done. Otherwise, either $n > 200 \cdot 5\delta \cdot |\mathcal{X}|$ or $m > 200 \cdot 5\delta \cdot |\mathcal{Y}|$ holds. We will show that in this case ρ and ρ' are approximately similar rectangles and conclude that $mn \leq M$, giving us $\frac{mn}{|\mathcal{X}|^2|\mathcal{Y}|^2} \leq \frac{M}{|\mathcal{X}|^2|\mathcal{Y}|^2}$.

Claim. *Suppose that $n > 200 \cdot 5\delta \cdot |\mathcal{X}|$ or $m > 200 \cdot 5\delta \cdot |\mathcal{Y}|$. Then the rectangles ρ and ρ' are approximately similar.*

Proof. Let us assume that $m > 200 \cdot 5\delta \cdot |\mathcal{Y}|$. (The other case is handled symmetrically). For every $(i, j) \in [n] \times [m]$ it holds that

$$g_r(x_i, y_j) = g_{r'}(x'_i, y'_j),$$

where $g_r = g(\varPi_A(x, r), \varPi_B(y, r))$. Hence, to prove that $f_{[\rho]}$ agrees with $f_{[\rho']}$ on all but 1/100 of the inputs, it suffices to show that g_r (resp., $g_{r'}$) disagrees with f over at most 1/200-fraction of the entries of ρ (resp., ρ'). Indeed, for every $x_i \in \mathbf{x}$ (resp., $x'_i \in \mathbf{x}'$) the number of y's for which the PSM errs over r (resp., over r') is at most $5\delta|\mathcal{Y}|$. Overall, for each of the rectangles ρ and ρ', the number of decoding errors produced by the PSM (over r and r' respectively) is bounded above by $n \times 5\delta|\mathcal{Y}|$. As a result, f disagrees with g_r (resp., $g_{r'}$) on at most $\frac{n \times 5\delta|\mathcal{Y}|}{mn} \leq 1/200$ fraction of the inputs in ρ (resp., ρ'). The claim follows.

This completes the proof of Claim 16. □

6 Imperfect PSM lower-bounds for random and explicit functions

In this section we will show that most functions have non-trivial imperfect PSM complexity, and establish the existence of an explicit function that admits a non-trivial imperfect PSM lower-bound. Formally, in Section 6.1 we will prove the following theorem (which strengthens Corollary 1 from the introduction).

Theorem 17. *For a $1 - o(1)$ fraction of the functions $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ any PSM protocol for f with privacy error of ϵ and correctness error of δ, $\delta < \frac{1}{100}$, requires at least*

$$\ell(k, \epsilon, \delta) = \min\{3k - 2\log(k), 2k + \log(1/\epsilon), 2k + \log(1/\delta)\} - c \tag{16}$$

bits of communication, where c is some universal constant.

By de-randomizing the proof, we derive (in Section 6.2) the following theorem (which strengthens Theorem 4 from the introduction).

Theorem 18. *There exists a sequence of polynomial-size circuits*

$$f = \left\{f_k : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}\right\}$$

such that any δ-correct ϵ-private PSM for f_k has communication complexity of at least $\ell(k, \epsilon, \delta)$ bits (as defined in (16)). Moreover, assuming the existence of a hitting-set generator against co-nondeterministic uniform algorithms, there exists an explicit family f which is computable by a polynomial-time Turing machine whose imperfect PSM communication complexity is at least $\ell(k, \epsilon, \delta) - O(\log k)$.

The reader is advised to read the following subsections sequentially since the proof of Theorem 18 builds over the proof of Theorem 17.

6.1 Lower bounds for random functions (Proof of Thm. 17)

We will need the following definition.

Definition 6 (good function). *We say that a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ is* good *if it satisfies the following conditions:*

1. *For every set $S \subset \{0,1\}^k \times \{0,1\}^k$ of k^2 consecutive pairs of strings (according to some predefined order over $\{0,1\}^k \times \{0,1\}^k$), the function f, restricted to the inputs in S, is somewhat regular (as per Definition 5). That is,*

$$0.9\frac{k^2}{2} \leq \sum_{(x,y)\in S} f(x, y) \leq 1.1\frac{k^2}{2}.$$

2. *For every $x \neq x'$ and every set $\mathbf{y}$ of k^2 consecutive strings (according to some predefined order over $\{0,1\}^k$), it holds that $f(x,y) = f(x',y)$ for at most 0.9-fraction of the elements $y \in \mathbf{y}$.*
3. *Similarly, for every $y \neq y'$ and set $\mathbf{x}$ of k^2 consecutive strings (according to some predefined order over $\{0,1\}^k$), it holds that $f(x,y) = f(x,y')$ for at most 0.9-fraction of $x \in \mathbf{x}$.*
4. *For every pair of $k^2 \times k^2$ x-disjoint or y-disjoint rectangles R, R', it holds that $f_{[R]}$ disagrees with $f_{[R']}$ on at least 0.01 fraction of the entries.*

Claim 19. *Any good $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ satisfies the conditions of Theorem 11 with $M = 2^k \cdot k^2$, and therefore any δ-correct ϵ-private PSM for f, $\delta < \frac{1}{100}$, requires communication of*

$$\ell(k, \epsilon, \delta) = \min\{3k - 2\log(k), 2k + \log(1/\epsilon), 2k + \log(1/\delta)\} - c,$$

for some universal constant c.

Proof. Fix some good f. First, since f is somewhat regular on every block of consecutive inputs (Condition (1)), it is also somewhat regular over the whole domain. Next, Condition (2) guarantees that $f(x,\cdot)$ and $f(x',\cdot)$ differ on 0.1 fraction of each k^2 block of consecutive y's, and therefore, overall, they must differ on a 0.1 fraction of all possible y's. Applying the same argument on the y-axis (using condition (3)), we conclude that a good f must be strongly non-degenerate.

Finally, we claim that a good f cannot have a pair of x-disjoint approximately similar $m \times n$ rectangles R, R' of size $mn \geq 2^k \cdot k^2$. To see this, observe that the latter condition implies that m, n are both larger than k^2, and therefore, again by an averaging argument, there must exists a pair of $k^2 \times k^2$ x-disjoint sub-rectangles $R'_0 \subseteq R_0, R'_1 \subseteq R_1$ which are also approximately similar. Applying the same argument to y-disjoint rectangles we conclude that any good f satisfies the conditions of Theorem 11. □

We say that a family of functions $\{f_z : \mathcal{A} \to \mathcal{B}\}_{z \in \mathcal{Z}}$ is t-wise independent functions if for any t-tuple of distinct inputs $(a_1, \ldots, a_t)$ and for a uniformly chosen $z \overset{\$}{\leftarrow} \mathcal{Z}$, the joint distribution of $(f_z(a_1), \ldots, f_z(a_t))$ is uniform over $\mathcal{B}^t$.

Claim 20. *Pick $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ uniformly at random among all such functions. Then, with probability $1 - o(1)$, the resulting function is good. Moreover, this holds even if f is chosen from a family of k^4-wise independent functions.*

Proof. Choose f randomly from a family of k^4-wise independent hash functions. By a multiplicative Chernoff bound, the probability that f fails to be somewhat regular over some fixed set $S \subset \{0,1\}^k \times \{0,1\}^k$ of size k^2, is $2^{-\Omega(k^2)}$. (Indeed, $\{f(x,y)\}_{(x,y)\in S}$ is just a uniform sequence of k^2 random bits.) Since we consider at most $2^{2k}/k^2$ (consecutive) sets of inputs we can apply a union-bound and conclude that condition (1) fails with probability at most $2^{2k-\Omega(k^2)} = 2^{-\Omega(k^2)}$.

Next, we establish properties (2) and (3). Fix a pair of $x \neq x'$ and a k^2-subset $\mathbf{y} \subset \{0,1\}^k$ of consecutive y's. By a Chernoff bound, the probability that $f(x,y) = f(x',y)$ for more than 0.9 of $y \in \mathbf{y}$ is at most $2^{-\Omega(k^2)}$. There are at most 2^{2k} pairs of x, x', and at most 2^k different sets $\mathbf{y}$ of consecutive y's, therefore by a union bound the probability that condition (2) does not hold is $2^{3k} 2^{-\Omega(k^2)} = 2^{-\Omega(k^2)}$. A similar argument, shows that (3) fails with a similar probability.

We move on to prove there is no pair of approximately similar x-disjoint rectangles of size exactly $k^2 \times k^2$. (Again, the case of y-disjoint rectangles is treated similarly.)

Let $m = k^2$. Fix two x-disjoint $m \times m$-rectangles $R = (\mathbf{x}, \mathbf{y})$ and $R' = (\mathbf{x}', \mathbf{y}')$. We want to give an upper bound on the probability that $f_{[R]}$ agrees with $f_{[R']}$ on 0.99 of their entries. This event happens only if the entries of f satisfy all but 0.01 of the the m^2 equations $f(x_i, y_j) = f(x'_i, y'_j)$ for $(i,j) \in \{1, \ldots, m\} \times \{1, \ldots, m\}$. The probability that any such equation is satisfied is $\frac{1}{2}$: since the rectangles are x-disjoint the equation is non-trivial. We can further find a subset T of at least $m^2/2$ such equations such that each equation in the subset uses an entry $f(x,y)$ that is not used in any other equation. Let us fix some $0.01m^2$ subset S of equations that are allowed to be unsatisfied. After removing S from T, we still have at least $0.49m^2$ equations that are simultaneously satisfied with probability of at most $2^{-0.49m^2}$. There are at most $2^{H_2(0.01)m^2}$ sets S (where H_2 is the binary entropy function), and at most 2^{2mk} choices for R and 2^{2mk} choices for R'. Hence, by a union bound, the probability that (3) fails is at most

$$2^{-0.49m^2 + 0.081m^2 + 4m^{3/2}} < 2^{-\Omega(m^2)},$$

the claim follows. □

Theorem 17 follows from Claims 19 and 20. □

6.2 Explicit lower-bound (Proof of Thm. 18)

Our next goal is to obtain an explicit lower-bound. We begin by noting that good functions (as per definition 6) can be identified by efficient co-nondeterministic algorithms.

Definition 7. *A co-nondeterministic algorithm $M(z,v)$ is a Turing machine that takes z as its primary input and v as a witness. For each $z \in \{0,1\}^*$ we say that M* rejects *z if there exist a witness v such that $M(z,v) = 1$, and say that M* accepts *z otherwise.*

Claim 21. *There exists a co-nondeterministic algorithm M that given some s-bit representation of a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ accepts f if and only if f is good. The complexity of M is $O(k^4 t)$ where $t \geq k$ is an upper-bound on the time complexity of evaluating f on a given point.*

Proof. It suffices to describe a polynomial-time verifiable witness for the failure of each of the goodness conditions.

A failure of (1) is witnessed by a set $S \subset \{0,1\}^k \times \{0,1\}^k$ of k^2 consecutive inputs. The verification process computes $s = \sum_{(x,y)\in S} f(x,y)$ in time $O(k^2 t + k^2 \log k)$, and outputs 1 if s is not in $[0.9k^2/2, 1.1k^2/2]$.

If f is not good due to (2), then the witness is a pair $x \neq x'$ and a k^2-set $\mathbf{y}$ of consecutive y's. Since f can be efficiently evaluated we can verify that $f(x,y) = f(x',y)$ for more than 0.9-fraction of the y's in $\mathbf{y}$ in times $O(k^2 t + k^2 \log k)$. A violation of (3) is treated similarly.

If f is not good due to (4), then the witness is a pair of x-disjoint or y-disjoint $k^2 \times k^2$ rectangles R, R' that are approximately similar. Again, we can verify the validity of this witness in time $O(k^4 t + k^4 \log k)$, which simplifies to $O(k^4 t)$ since $t \geq k$. □

Let $s(k) = \text{poly}(k)$ and let $\left\{f_z : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}\right\}_{z\in\{0,1\}^s}$ be a family of k^4-wise independent functions with an *evaluator* algorithm F which takes an index $z \in \{0,1\}^s$ and input $(x,y) \in \{0,1\}^k \times \{0,1\}^k$ and outputs in time $t(k)$ the value of $f_z(x,y)$. (Such an F can be based on k^4-degree polynomials over a field of size $\Theta(k^4)$). Claims 19 and 20 imply that for most choices of z, the function f_z has an imperfect PSM complexity of at least $\ell(k,\epsilon,\delta)$. Since F is efficiently computable, for every z there is a polynomial-size circuit that computes f_z. Hence, there exists a polynomial-size computable function for which the $\ell(k,\epsilon,\delta)$ lower-bound holds, and the first part of Theorem 18 follows.

To prove the second part, we use a properly chosen pseudorandom generator (PRG) $G : \{0,1\}^{O(\log k)} \to \{0,1\}^s$ to "derandomize" the family $\{f_z\}$. That is, we define the function $g : \{0,1\}^{O(\log k)} \times \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ which takes (w,x,y) and outputs $f_z(x,y)$ where $z = G(w) \in \{0,1\}^s$. Concretely, we require G to "hit" the image of any co-nondeterministic algorithms of complexity $T = O(k^4 t)$. Formally, this means that for every T-time co-nondeterministic algorithm M it holds that if $\Pr_z[M(z) = 1] \geq \frac{1}{2}$ then there exists a "seed" r for which $M(G(r)) = 1$.

Taking M to be the algorithm from Claims 21, we conclude, by Claims 20 and 19, that for some seed w, the function $f_{G(w)}$ has an imperfect PSM complexity of at least $\ell(k,\epsilon,\delta)$. Let us parse g as a two-party function, say by partitioning w to two halves w_A, w_B and giving (x, w_A) to Alice, and y, w_B to Bob. We conclude that g must have an imperfect PSM complexity of at least $\ell(k,\epsilon,\delta)$. Since the input length k' of Alice and Bob becomes longer by an additional $O(\log k)$ bits, the lower-bound becomes at least $\ell(k',\epsilon,\delta) - O(\log k')$, as claimed. The second part of Theorem 18 follows. □

7 Lower-bounds for Conditional Disclosure of Secrets

In this section we derive CDS lower bounds. We begin with a reduction from fully revealing weakly hiding PSM (Definition 4) to CDS.

Claim 22. *Let $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. Define the function $f : \mathcal{X}' \times \mathcal{Y} \to \{0,1\}$ where $\mathcal{X}' = \mathcal{X} \times \{0,1\}$ by $f((x,s),y) = s \wedge h(x,y)$. If h has a perfect CDS with communication complexity of c then f has a weakly-private fully-revealing PSM with complexity of $c + \log|\mathcal{X}| + \log|\mathcal{Y}|$.*

Proof. Given a CDS protocol $\mathit{\Pi} = (\mathit{\Pi}_A, \mathit{\Pi}_B, g)$ for h we construct a weakly-private fully-revealing PSM for f as follows. Given an input (x,s), Alice sends $(x, a = \mathit{\Pi}_A(x,s,r))$ where x plays the role of the Alice's input in the CDS, s plays the role of the secret, and r is a shared string uniformly sampled from $\mathcal{R}$. Bob takes his input y, and sends $(y, b = \mathit{\Pi}_B(y,r))$. Charlie outputs $h(x,y) \wedge g(x,y,a,b)$.

It is not hard to verify that the protocol is perfectly correct and fully revealing. Indeed, a PSM decoding error happens only if $g(x,y,a,b)$ fails to decode the secret s (which happens with probability zero). To prove weak privacy observe that if f agrees on a pair of inputs, $((x,0),y)$ and $((x,1),y)$, then $h(x,y)$ must be zero. By CDS privacy, for $R \xleftarrow{\$} \mathcal{R}$ the distribution $(x,y,\mathit{\Pi}_A(x,0,R),\mathit{\Pi}_B(y,R))$ is identical to the distribution $(x,y,\mathit{\Pi}_A(x,1,R),\mathit{\Pi}_B(y,R))$, as required. □

Next, we show that the properties of f needed for applying Theorem 8, follow from simple requirements on h. In the following, we say that $x \in \mathcal{X}$ is a *null input* if the residual function $h(x,\cdot)$ is the constant zero function.

Claim 23. *Let h and f be as in Claim 22. Then*

1. *The size of the largest complement similar rectangle of f equals to the size of the largest 0-monochromatic rectangle of h.*
2. *The number U of useful inputs of f is exactly two times larger than the number of inputs that are mapped by h to zero.*
3. *If h has no input x for which the residual function $h(x,\cdot)$ is the constant zero function, then f is weakly non-degenerate.*

Proof. The claim follows immediately by noting that for every (x,y) it holds that $f((x,1),y) = f((x,0),y)$ if and only if $h(x,y) = 0$. We proceed with a formal argument.

1. Consider some complement similar rectangle $R = (\mathbf{x}' \times \mathbf{y})$ of f. For every $(x,b) \in \mathbf{x}'$ and $y \in \mathbf{y}$, it holds that
$$f((x,b),y) = f((x,1-b),y),$$
and therefore $h(x,y) = 0$ and R is a 0-monochromatic rectangle of h.
2. Every input (x,y) that does not satisfy h induces an unordered pair, $((x,1),y)$ and $((x,0),y)$, of useful inputs for f. Therefore, the number of (ordered) useful inputs of f is exactly $2|h^{-1}(0)|$.
3. Fix some $(x,s) \in \mathcal{X}'$ and assume, towards a contradiction, that for every y it holds that $f((x,s),y) = f((x,1-s),y)$. By the definition of f this means that $h(x,y) = 0$ for every y, contradicting our assumption on h.

□

Theorem 5 (restated here for convenience) now follows immediately from the lower-bound on weakly-private fully revealing PSM (Theorem 8).

Theorem 24 (Theorem 5 restated). *Let $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. Suppose that M upper-bounds the size of the largest 0-monochromatic rectangle of h and that for every $x \in \mathcal{X}$, the residual function $h(x,\cdot)$ is not the constant zero function. Then, the communication complexity of any perfect CDS for h is at least*

$$2\log|f^{-1}(0)| - \log M - \log|\mathcal{X}| - \log|\mathcal{Y}| - 1,$$

where $|f^{-1}(0)|$ denotes the number of inputs (x,y) that are mapped to zero.

Proof. Let $h : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate that satisfies the theorem requirement. That is, M upper-bounds the size of the largest 0-monochromatic rectangle of h, there at least S inputs that are mapped to zero, and for every $x \in \mathcal{X}$, the residual function $h(x,\cdot)$ is not the constant zero function.

Suppose that h has a perfect CDS with communication complexity of c. By Claim 22, the function f (defined in the claim) has a weakly-private fully-revealing PSM with complexity of at most

$$c + \log|\mathcal{X}| + \log|\mathcal{Y}|,$$

which, by Claim 23 and Theorem 8, is at least

$$2\log U - \log M - 2 = 2\log S - \log M - 1.$$

It follows that

$$c \geq 2\log S - \log M - 1 - (\log|\mathcal{X}| + \log|\mathcal{Y}|),$$

as required. □

Example 1 (The index predicate). As a sanity check, consider the index predicate $f_{ind} : [k] \times \{0,1\}^k \to \{0,1\}$ which given an index $i \in [k]$ and a string $y \in \{0,1\}^k$ outputs $y[i]$, the i-th bit of y. Clearly exactly half of all inputs are mapped to 0. Also, for every i the residual function $f(i,\cdot)$ is not the constant zero. Finally, every zero rectangle is of the form $I \times \{y : y[i] = 0, \forall i \in I\}$ where $I \subseteq [k]$. This implies that the size of any such rectangle is exactly $|I| \cdot 2^{k-|I|} \leq 2^{k-1}$. Plugging this into Theorem 24, we get a lower-bound of

$$2(k + \log k - 1) - (k-1) - k - \log k - 1 \geq \log k - 2.$$

References

1. William Aiello, Yuval Ishai, and Omer Reingold. Priced oblivious transfer: How to sell digital goods. In *EUROCRYPT*, pages 119–135, 2001.
2. Benny Applebaum. Garbled circuits as randomized encodings of functions: a primer. In Yehuda Lindell, editor, *Tutorials on the Foundations of Cryptography.*, pages 1–44. Springer International Publishing, 2017.

3. Benny Applebaum and Barak Arkis. Conditional disclosure of secrets and \$d\$-uniform secret sharing with constant information rate. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:189, 2017.
4. Benny Applebaum, Barak Arkis, Pavel Raykov, and Prashant Nalini Vasudevan. Conditional disclosure of secrets: Amplification, closure, amortization, lower-bounds, and separations. In *CRYPTO*, pages 727–757, 2017.
5. Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC^0. In *FOCS*, pages 166–175, 2004.
6. Benny Applebaum and Pavel Raykov. From private simultaneous messages to zero-information arthur-merlin protocols and back. In *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part II*, pages 65–82, 2016.
7. Boaz Barak, Shien Jin Ong, and Salil P. Vadhan. Derandomization in cryptography. *SIAM J. Comput.*, 37(2):380–400, 2007.
8. Donald Beaver, Silvio Micali, and Phillip Rogaway. The round complexity of secure protocols (extended abstract). In *STOC*, pages 503–513, 1990.
9. Amos Beimel, Yuval Ishai, Ranjit Kumaresan, and Eyal Kushilevitz. On the cryptographic complexity of the worst functions. In *TCC*, pages 317–342, 2014.
10. Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In *STOC*, pages 1–10, 1988.
11. Ernest F. Brickell and Daniel M. Davenport. On the classification of ideal secret sharing schemes. *J. Cryptology*, 4(2):123–134, 1991.
12. Renato M. Capocelli, Alfredo De Santis, Luisa Gargano, and Ugo Vaccaro. On the size of shares for secret sharing schemes. *J. Cryptology*, 6(3):157–167, 1993.
13. David Chaum, Claude Crépeau, and Ivan Damgård. Multiparty unconditionally secure protocols (extended abstract). In *STOC*, pages 11–19, 1988.
14. Benny Chor, Eyal Kushilevitz, Oded Goldreich, and Madhu Sudan. Private information retrieval. *J. ACM*, 45(6):965–981, 1998.
15. Deepesh Data, Vinod M. Prabhakaran, and Manoj M. Prabhakaran. Communication and randomness lower bounds for secure computation. *IEEE Trans. Information Theory*, 62(7):3901–3929, 2016.
16. Uriel Feige, Joe Kilian, and Moni Naor. A minimal model for secure computation (extended abstract). In *STOC*, pages 554–563, 1994.
17. Romain Gay, Iordanis Kerenidis, and Hoeteck Wee. Communication complexity of conditional disclosure of secrets and attribute-based encryption. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part II*, volume 9216 of *Lecture Notes in Computer Science*, pages 485–502. Springer, 2015.
18. Yael Gertner, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin. Protecting data privacy in private information retrieval schemes. *J. Comput. Syst. Sci.*, 60(3):592–629, 2000.
19. Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game or a completeness theorem for protocols with honest majority. In *STOC*, 1987.
20. Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In Ari Juels, Rebecca N. Wright, and Sabrina De Capitani di Vimercati, editors, *Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006*, pages 89–98. ACM, 2006.

21. Dan Gutfreund, Ronen Shaltiel, and Amnon Ta-Shma. Uniform hardness versus randomness tradeoffs for arthur-merlin games. *Computational Complexity*, 12(3-4):85–130, 2003.
22. Yuval Ishai. Randomization techniques for secure computation. In Manoj Prabhakaran and Amit Sahai, editors, *Secure Multi-Party Computation*, volume 10 of *Cryptology and Information Security Series*, pages 222–248. IOS Press, 2013.
23. Yuval Ishai and Eyal Kushilevitz. Private simultaneous messages protocols with applications. In *ISTCS (Israel Symposium on Theory of Computing and Systems)*, pages 174–184, 1997.
24. Yuval Ishai and Eyal Kushilevitz. Randomizing polynomials: A new representation with applications to round-efficient secure computation. In *FOCS*, pages 294–304, 2000.
25. Yuval Ishai and Hoeteck Wee. Partial garbling schemes and their applications. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, volume 8572 of *Lecture Notes in Computer Science*, pages 650–662. Springer, 2014.
26. Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, 1997.
27. Tianren Liu, Vinod Vaikuntanathan, and Hoeteck Wee. Conditional disclosure of secrets via non-linear reconstruction. In *CRYPTO*, pages 758–790, 2017.
28. Peter Bro Miltersen and N. V. Vinodchandran. Derandomizing arthur-merlin games using hitting sets. In *FOCS*, pages 71–80, 1999.
29. Amit Sahai and Brent Waters. Fuzzy identity-based encryption. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings*, volume 3494 of *Lecture Notes in Computer Science*, pages 457–473. Springer, 2005.
30. Claude E. Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28:656–715, 1949.
31. Hung-Min Sun and Shiuh-Pyng Shieh. Secret sharing in graph-based prohibited structures. In *Proceedings IEEE INFOCOM '97, The Conference on Computer Communications, Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Driving the Information Revolution, Kobe, Japan, April 7-12, 1997*, pages 718–724. IEEE, 1997.
32. Brent Waters. Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In Dario Catalano, Nelly Fazio, Rosario Gennaro, and Antonio Nicolosi, editors, *Public Key Cryptography - PKC 2011 - 14th International Conference on Practice and Theory in Public Key Cryptography, Taormina, Italy, March 6-9, 2011. Proceedings*, volume 6571 of *Lecture Notes in Computer Science*, pages 53–70. Springer, 2011.
33. Andrew Chi-Chih Yao. Protocols for secure computations (extended abstract). In *FOCS*, pages 160–164, 1982.

A Lower-bound for weakly-private fully-revealing protocols

In this section we provide an alternative combinatorial proof to Theorem 8 by following the outline of the original FKN argument. Recall that f is *weakly non-degenerate* if for every x there exists y such that $f(x,y) \neq f(\bar{x},y)$.

Let $\Pi = (\Pi_A, \Pi_B, g)$ be a perfect PSM protocol for f whose shared randomness is sampled uniformly from the set $\mathcal{R}$. In the following we let $C_r(x, y) = (\Pi_A(x, r), \Pi_B(x, r))$ denote the function that for randomness r, maps the inputs of Alice and Bob, x, y, to the messages (a, b). To prove Theorem 8 we show that at least $U^2/(2M)$ different messages (a, b) are being sent in Π. In fact, we will show that this happens even if we restrict our attention to some well chosen (ordered) subset of random strings $\mathbf{r} = (r_1, \ldots, r_L)$. We will also restrict our attention only to useful inputs (x, y) (for which $f(x, y) = f(\bar{x}, y)$). That is, we will prove that

$$|\bigcup_{r \in \mathbf{r}} \{C_r(x, y) : (x, y) \text{ is useful}\}| \geq U^2/(4M) \tag{17}$$

For a given $\mathbf{r} = (r_1, \ldots, r_L)$ consider the following counting scheme.

```
1. Initialize $S$ to be an empty set.
2. For $i = 1, \ldots, L$ do:
    (a) For every useful $(x, y)$: If $c = C_{r_i}(x, y)$ does not appear in $S$ add it to $S$.
        Else, discard it and call it a collision.
3. Output $|S|$.
```

Clearly, the output equals to the LHS of (17). The lower-bound will be established by showing that there are not too many *collisions*. We begin by noting that C_r is injective for any r.

Claim 25. *For every r, the function $C_r(\cdot)$ is injective.*

Proof. Suppose that $C_r(x, y) = C_r(x', y')$ for some r. Since the protocol is fully revealing it must hold that $y = y'$ and $x' \in \{x, \bar{x}\}$. Assume, towards a contradiction, that $x' = \bar{x}$, and let $a = \Pi_A(x, r) = \Pi_A(\bar{x}, r)$. For any y, let $b(y) = \Pi_B(y; r)$. Then, by perfect correctness, it holds that

$$f(x, y) = g(a, b(y)) = f(\bar{x}, y)$$

for every y. Contradicting the fact that f is weakly non-degenerate. □

Suppose that (x, y) and (x', y') collide over r and r'. That is, $c = C_r(x, y) = C_{r'}(x', y')$. Then either the collision is *trivial*, i.e., $(x, y) = (x', y')$, or $(x, y) \neq (x', y')$. In the latter case, it must hold that $x' = \bar{x}$ and $y' = y$ since the protocol is fully revealing. Indeed, otherwise, the value of $(x[1 : k - 1], y)$ cannot be recovered from c. We refer to this type of collision as *non-trivial.*[12]

While it easy to get some upper-bound on the amount of non-trivial collisions, it is somewhat tricky to upper-bound the number of trivial collisions. For this

[12] This is the point where the additional property of fully revealing is being used.

reason, we treat the two cases asymmetrically. In particular, getting back to our counting algorithm, consider a useful (x, y) which is excluded in the i-th step. We refer to this exclusion as *non-trivial* if there exists some $r_j, j < i$ for which $C_{r_i}(x, y) = C_{r_j}(\bar{x}, y)$. Otherwise, we refer to the exclusion as *trivial*. In the latter case, $c = C_{r_i}(x, y)$ appears in the image of C_{r_j} for (one or more) $r_j, j < i$ but only as an image of (x, y).

The following claim relates the number of non-trivial collisions to the size of the largest complement similar rectangle of f. (Recall that the latter is upper-bounded by M.)

Claim 26. *Every pair of distinct random string $r \neq r'$ has at most M non-trivial collisions. Consequently, for every choice of $\mathbf{r} = (r_1, \dots, r_L)$, the total number of non-trivial exclusions in the above process is at most $ML^2/2$.*

Proof. Fix a pair of random strings $r \neq r'$ and consider all their non-trivial collisions $(x_1, y_1), \dots, (x_t, y_t)$. For every i, it holds that $C_r(x_i, y_i) = C_{r'}(\bar{x}_i, y_i)$ and therefore $\Pi_A(x_i, r) = \Pi_A(\bar{x}_i, r')$ and $\Pi_B(y_i, r) = \Pi_B(y_i, r')$. It follows that for every i, j, $C_r(x_i, y_j) = C_{r'}(\bar{x}_i, y_j)$, and so by perfect correctness the rectangle $R = \{(x_i, y_j) : 1 \leq i, j \leq t\}$ is complement similar. We conclude that r and r' can have at most M non-trivial collisions, which implies that the total number of non-trivial collisions over $\mathbf{r}$ is at most $\sum_{i=1}^{L-1} iM \leq ML^2/2$. □

The next claim handles trivial collisions by relating their number to the number of non-trivial collisions.

Claim 27. *For every $1 \leq L \leq |\mathcal{R}|$, there exists a sequence of L distinct strings $\mathbf{r} = (r_1, \dots, r_L)$ for which the total number of trivial exclusions in the above process is upper-bounded by the number of non-trivial exclusions.*

Proof. By induction on L. The case of $L = 1$ trivially holds. Fix some sequence $\mathbf{r} = (r_1, \dots, r_{L-1})$ of length $L - 1$ for which the claim holds. We show that we can always extend $\mathbf{r}$ by an additional string $r \notin \mathbf{r}$ such that $T(r)$, the number of trivial exclusions contributed by r, is upper-bounded by $N(r)$, the number of non-trivial exclusions contributed by r. For this aim, we will show that the expectation, over a random $r \notin \mathbf{r}$, of $\delta(r) = N(r) - T(r)$ is positive.

For every useful (x, y) and transcript c, let $N_{x,y,c}(r)$ be an indicator that takes the value 1 if $C_r(x, y) = c$ and there exists $r' \in \mathbf{r}$ for which $C_{r'}(\bar{x}, y) = c$. Similarly, let $T_{x,y,c}(r)$ be an indicator that takes the value 1 if $C_r(x, y) = c$ and there exists $r' \in \mathbf{r}$ for which $C_{r'}(x, y) = c$ but there is no $r'' \in \mathbf{r}$ for which $C_{r''}(\bar{x}, y) = c$. Note that $N(r) = \sum_{x,y,c} N_{x,y,c}(r)$ and $T(r) = \sum_{x,y,c} T_{x,y,c}(r)$. It will be useful to "symmetrize" $N(r) - T(r)$ around x and $\bar{x}$ and write it as

$$\frac{1}{2} \sum_{x,y,c} (N_{x,y,c}(r) - T_{x,y,c}(r)) + (N_{\bar{x},y,c}(r) - T_{\bar{x},y,c}(r)).$$

Hence, by the linearity of expectation, it suffices to show that for every x, y, c,

$$\mathsf{E}_{r:r\notin\mathbf{r}}[(N_{x,y,c}(r) - T_{x,y,c}(r)) + (N_{\bar{x},y,c}(r) - T_{\bar{x},y,c}(r))] \geq 0. \tag{18}$$

We establish this via case analysis.

1. c is new. That is, for every $r' \in \mathbf{r}$, neither $C_{r'}(x,y) = c$ nor $C_{r'}(\bar{x},y) = c$. Then, for every r, $N_{x,y,c}(r) = T_{x,y,c}(r) = N_{\bar{x},y,c}(r) = T_{\bar{x},y,c}(r) = 0$ and (18) holds.
2. c already appears both under (x,y) and $(\bar{x},y)$. That is, there exists $r', r'' \in \mathbf{r}$ for which $C_{r'}(x,y) = C_{r''}(\bar{x},y) = c$. Then, for every $r \notin \mathbf{r}$,
$$N_{x,y,c}(r) = N_{\bar{x},y,c}(r) \geq 0, \quad \text{and} \quad T_{x,y,c}(r) = T_{\bar{x},y,c}(r) = 0,$$
where the inequality is strict when $C_r(x,y) = c$. Eq. (18) follows.
3. c already appears under (x,y) (i.e., $\exists r' \in \mathbf{r}$ s.t. $C_{r'}(x,y) = c$) but never appeared under $(\bar{x},y)$ (i.e., $\nexists r' \in \mathbf{r}$ s.t. $C_{r'}(\bar{x},y) = c$). In this case, we can partition all the $r \notin \mathbf{r}$ to three types.
   - (Bad) $C_r(x,y) = c$. In this case, $N_{x,y,c}(r) - T_{x,y,c}(r) = -1$ and $N_{\bar{x},y,c}(r) = T_{\bar{x},y,c}(r) = 0$ (since $C_r(\bar{x},y) \neq c$ due to injectivity).
   - (Good) $C_r(\bar{x},y) = c$. In this case $N_{\bar{x},y,c}(r) - T_{\bar{x},y,c}(r) = 1$ and $N_{x,y,c}(r) = T_{x,y,c}(r) = 0$ (since $C_r(x,y) \neq c$ due to injectivity).
   - (Neutral) $C_r(x,y), C_r(\bar{x},y) \neq c$. In this case, $N_{x,y,c}(r) = T_{x,y,c}(r) = N_{\bar{x},y,c}(r) = T_{\bar{x},y,c}(r) = 0$.

   We argue that, outside $\mathbf{r}$, there are more Good r's than Bad r's. Indeed, by perfect privacy, the total number of Good r's among all $r \in \mathcal{R}$ is equal to the total number of Bad r's among all $r \in \mathcal{R}$. Since $\mathbf{r}$ contains at least one Bad string but not a single Good string, the set $\mathcal{R} \setminus \mathbf{r}$ contains more Good strings than bad strings. Hence, (18) follows.
4. The last case is symmetric to the previous case (i.e., c already appears under $(\bar{x},y)$ but never appeared under (x,y)). This case is handled similarly to the previous case.

This completes the proof of Claim 27. □

In order to complete the argument, we prove a lower-bound on the number of random strings. Recall that U denotes the number of useful inputs.

Claim 28. *The size of $\mathcal{R}$ is at least U/M.*

Proof. Let $\mathcal{U}$ denote the set of all useful inputs. Fix some random string $r \in \mathcal{R}$. We count the number of non-trivial collisions ℓ with r. That is,
$$\ell = |\{(x,y,r') : (x,y) \in \mathcal{U} \text{ and } C_r(x,y) = C_{r'}(\bar{x},y)\}|.$$
Since C_r is injective (Claim 25), the set $C_r(\mathcal{U})$ consists of at least U transcripts. By perfect privacy, each such transcript $c = C_r(x,y), (x,y) \in \mathcal{U}$ must be an image of $(\bar{x},y)$ under some (different) $C_{r'}$. Hence, $\ell \geq U$. On the other hand, by Claim 26, each pair of distinct $r \neq r'$ can have at most M non-trivial collisions, and therefore $\ell \leq |\mathcal{R}|M$. The claim follows. □

We can now complete the proof. Let $L = U/(2M) < |\mathcal{R}|$. Apply the counting algorithm to a sequence $\mathbf{r} = (r_1, \ldots, r_L)$ of L distinct strings which satisfies Claim 27. Since C_r is injective for every $r \in \mathbf{r}$, Step 2a is performed LU times, out of which there are at most $ML^2/2$ non-trivial exclusions (Claim 27) and $ML^2/2$ trivial exclusions (Claim 27). Hence, S contains at least $LU - ML^2 = U^2/(4M)$ strings. The theorem follows. □