# Conditional Disclosure of Secrets: Amplification, Closure, Amortization, Lower-bounds, and Separations*

Benny Applebaum† Barak Arkis† Pavel Raykov† Prashant Nalini Vasudevan‡

February 20, 2017

**Abstract**

In the *conditional disclosure of secrets* problem (Gertner et al., J. Comput. Syst. Sci., 2000) Alice and Bob, who hold inputs $x$ and $y$ respectively, wish to release a common secret $s$ to Carol (who knows both $x$ and $y$) if only if the input $(x, y)$ satisfies some predefined predicate $f$. Alice and Bob are allowed to send a single message to Carol which may depend on their inputs and some joint randomness and the goal is to minimize the communication complexity while providing information-theoretic security.

Following Gay, Kerenidis, and Wee (Crypto 2015), we study the communication complexity of CDS protocols and derive the following positive and negative results.

- (**Closure**) A CDS for $f$ can be turned into a CDS for its complement $\bar{f}$ with only a minor blow-up in complexity. More generally, for a (possibly non-monotone) predicate $h$, we obtain a CDS for $h(f_1, \ldots, f_m)$ whose cost is essentially linear in the formula size of $h$ and polynomial in the CDS complexity of $f_i$.
- (**Amplification**) It is possible to reduce the privacy and correctness error of a CDS from constant to $2^{-k}$ with a multiplicative overhead of $O(k)$. Moreover, this overhead can be amortized over $k$-bit secrets.
- (**Amortization**) Every predicate $f$ over $n$-bit inputs admits a CDS for multi-bit secrets whose amortized communication complexity per secret bit grows linearly with the input length $n$ for sufficiently long secrets. In contrast, the best known upper-bound for single-bit secrets is exponential in $n$.
- (**Lower-bounds**) There exists a (non-explicit) predicate $f$ over $n$-bit inputs for which any perfect (single-bit) CDS requires communication of at least $\Omega(n)$. This is an exponential improvement over the previously known $\Omega(\log n)$ lower-bound.
- (**Separations**) There exists an (explicit) predicate whose CDS complexity is exponentially smaller than its randomized communication complexity. This matches a lower-bound of Gay et. al., and, combined with another result of theirs, yields an exponential separation between the communication complexity of linear CDS and non-linear CDS. This is the first provable gap between the communication complexity of linear CDS (which captures most known protocols) and non-linear CDS.

Our results solve several open problems posed by Gay et al., and have applications to secret-sharing schemes for forbidden-graph access structures.

*Research supported by the European Union's Horizon 2020 Programme (ERC-StG-2014-2020) under grant agreement no. 639813 ERC-CLC, and the Check Point Institute for Information Security.

†Tel-Aviv University, bennyap@post.tau.ac.il, barakark@mail.tau.ac.il, raykov.pavel@gmail.com.

‡MIT, prashantv91@gmail.com. This research was done while visiting Tel Aviv University.

# 1 Introduction

Consider a pair of computationally-unbounded parties, Alice and Bob, each holding an $n$-bit input, $x$ and $y$ respectively, to some public predicate $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Alice and Bob also hold a joint secret $s \in \{0,1\}$ and have access to a joint source of randomness $r \xleftarrow{R} \{0,1\}^\rho$. The parties wish to disclose the secret $s$ to a third party, Carol, if and only if the predicate $f(x,y)$ evaluates to 1. To this end, Alice (resp., Bob) should send to Carol a single message $a = a(x,s;r)$ (resp., $b = b(y,s;r)$). Based on the transcript $(a,b)$ and the inputs $(x,y)$, Carol should be able to recover the secret $s$ if and only if $f(x,y) = 1$. (Note that Carol is assumed to know $x$ and $y$.) That is, we require two properties:

- *Correctness*: There exists a decoder algorithm $\mathsf{Dec}$ that recovers $s$ from $(x,y,a,b)$ with high probability whenever $(x,y)$ is a 1-input (i.e., $f(x,y) = 1$);
- *Privacy*: There exists a simulator $\mathsf{Sim}$ that, given a 0-input $(x,y)$ (for which the predicate evaluates to 0), samples the joint distribution of the transcript $(x,y,a,b)$ up to some small deviation error.

The main goal is to minimize the communication complexity of the protocol which is taken to be the total bit-length of the messages $a$ and $b$. (See Section 3 for formal definitions.)

This form of *Conditional Disclosure of Secrets* (CDS) was introduced by Gertner, Ishai, Kushilevitz and Malkin [GIKM00] as a tool for adding data privacy to information-theoretically private information retrieval (PIR) protocols [CKGS98] and was later used in the computational setting as a light-weight alternative to zero-knowledge proofs (cf. [AIR01]). Apart from these applications, CDS plays a central role in the design of secret sharing schemes for graph-based access structures (cf. [BD91, CSGV93, SS97]) and in the context of attribute-based encryption [GPSW06, SW05]. In fact, CDS can be *equivalently formulated* under any of these frameworks as discussed below.

**Secret sharing for forbidden graphs.** CDS can be viewed as a special form of secret sharing for graph-based access structure (cf. [BD91, CSGV93, SS97]). Specifically, consider a secret-sharing scheme whose parties are the nodes of a bipartite graph $G = (X \cup Y, E)$ and a pair of parties $(x,y) \in X \times Y$ should be able to recover the secret $s$ if and only if they are connected by an edge (we do not require any privacy/correctness condition for other subsets of parties). Then, we can represent the secret-sharing problem as the problem of realizing a CDS for the predicate $f_G(x,y) = 1 \Leftrightarrow (x,y) \in E$ and vice-versa by setting the share of the $x$-th node (resp., $y$-th node) to be the message $a(x,s;r)$ (resp., $b(y,s;r)$). The communication complexity of the CDS protocol therefore corresponds to the size of shares.

**Attribute-based encryption.** CDS can be further viewed as a limited form of private-key attribute-based encryption [GPSW06, SW05] which offers one-time information-theoretic security. In such an encryption scheme both the decryption key $a_x$ of a receiver and the ciphertext $b_y$ of a sender are associated with some public attributes $x$ and $y$, respectively. The receiver should be able to decrypt the plaintext $m$ from the ciphertext $b_y$ using the key $a_x$ only if the attributes $x$ and $y$ "match" according to some predefined policy, i.e., satisfy some predicate $f(x,y)$. Using CDS for $f$, we can derive such a one-time secure scheme by letting the decryption key be Alice's message, $a_x = a(x,s;r)$, for a random secret $s$, and taking the ciphertext to be Bob's message

$b_y = b(y, s; r)$ together with a padded-version of the message $m \oplus s$. (Here we can think of $r, s$ as the sender's private-key.) In fact, it was shown by Attrapadung [Att14] and Wee [Wee14] that even in the computational setting of public-key (multi-user) attribute-based encryption (ABE), *linear CDS schemes* (in which the computation of Alice and Bob can be written as a linear function in the secret an the randomness) form a central ingredient. As a result, the ciphertext size and secret key of the ABE directly depend on the communication complexity of the underlying CDS.

**The communication complexity of CDS.** In light of the above, it is interesting to understand the communication complexity of CDS. Unfortunately, not much is known. Gertner et al. [GIKM00] showed that any predicate $f$ that can be computed by a $t$-size Boolean formula admits a perfect linear CDS (with zero correctness/privacy error) with communication complexity of $O(t)$. This result was extended by Ishai and Wee [IW14] to $s$-size (arithmetic) branching programs and by Applebaum and Raykov [AR16] to $s$-size (arithmetic) span programs (though in the latter case correctness is imperfect). Beimel, Ishai, Kumaresan and Kushilevitz [BIKK14] proved that the CDS complexity of the *worst* predicate $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ over $n$-bit inputs is at most $O(2^{n/2})$. A similar upper-bound was later established by Gay, Kerenidis, and Wee [GKW15] for the case of linear CDS, where a matching (non-explicit) lower-bound follows from the work of Mintz [Min12]. Gay et al. [GKW15] also initiated a systematic treatment of the communication complexity of CDS and established the first lower-bounds on the communication complexity of general CDS. Their main result relates the CDS communication of a predicate $f$ to its randomized communication complexity. Roughly speaking, it is shown that a general CDS for $f$ must communicate at least $\Omega(\log(\mathsf{R}(f)))$ bits, and a linear CDS must communicate at least $\Omega(\sqrt{\mathsf{R}(f)})$, where $\mathsf{R}(f)$ denotes the number of bits communicated in a randomized protocol that need to be exchanged between Alice and Bob in order to compute $f$ with constant error probability.[1] This yields (explicit) lower-bounds of $\Omega(\log(n))$ and $\Omega(\sqrt{n})$ for concrete $n$-bit predicates. Overall, for general CDS, there is a double-exponential gap between the best known (logarithmic) lower-bound and the best known (exponential) upper bound.

# 2 Our Results

Following Gay et al.[GKW15], we conduct a systematic study of the complexity of CDS. Unlike previous works, we focus on *manipulations* and *transformations* of various forms of CDS. Our approach yields several positive and negative results regarding the complexity of CDS, and answers several open problems posed in previous works. We proceed with a statement of our results.

## 2.1 Closure properties

We begin by asking whether one can generally combine CDS for basic predicates $f_1, \ldots, f_m$ into a CDS for a more complicated predicate $h(f_1, \ldots, f_m)$. Using standard secret sharing techniques, one can derive such a transformation when $h$ is a monotone function (with overhead proportional to the monotone formula size of $h$). However, these techniques fail to support non-monotone operations. Our first observation asserts that linear CDS for $f$ can be easily transformed into a linear CDS for its complement $\overline{f} \equiv 1 - f$.

[1] More precisely, $\mathsf{R}(f)$ can be replaced with the communication complexity of one-message protocol from Alice to Bob plus the communication complexity of one-message protocol from Bob to Alice.

**Theorem 2.1** (Linear CDS is closed under complement)**.** *Suppose that $f$ has a linear* CDS *with randomness complexity of $\rho$ and communication complexity of $t$, then $\overline{f}$ has a linear* CDS *scheme with randomness complexity of $2t + \rho + 1$ and communication complexity of $2(\rho + 1)$.*

The theorem generalizes to arbitrary finite field $\mathbb{F}$. (See Section 4.1.) Roughly speaking, we rely on the following observation. It can be shown that, for a fixed input $(x, y)$, the parties jointly compute some linear operator $T_{x,y}$ that has a high rank whenever $f(x, y) = 0$, and low rank when $f(x, y) = 1$. We "reverse" the CDS by essentially moving to the dual $T^*_{x,y}$ of $T_{x,y}$ whose rank is high when $f(x, y) = 1$, and low when $f(x, y) = 0$. One still has to find a way to distributively compute the mapping $T^*_{x,y}$. We solve this technicality by using a private simultaneous message protocol (PSM) [FKN94] that allows Alice and Bob to securely release an image of $T^*_{x,y}$ to Carol without leaking any additional information.

Next, we show that a similar "reversing transformation" exists for general (non-linear and imperfect) CDS protocols.

**Theorem 2.2** (CDS is closed under complement)**.** *Suppose that $f$ has a* CDS *with randomness complexity of $\rho$ and communication complexity of $t$ and privacy/correctness errors of $2^{-k}$. Then $\overline{f} \equiv 1 - f$ has a* CDS *scheme with similar privacy/correctness errors and randomness/communication complexity of $O(k^3\rho^2 t + k^3\rho^3)$.*

Imitating the argument used for the case of linear CDS, we consider, for an input $(x, y)$ and secret $s$, the probability distribution $D^s_{x,y}$ of the messages $(a, b)$ induced by the choice of the common random string. Observe that the distributions $D^0_{x,y}$ and $D^1_{x,y}$ are statistically far when $f(x) = 1$ (due to correctness), and are statistically close when $f(x, y) = 0$ (due to privacy). Therefore, to prove Theorem 2.2 we should somehow *reverse* statistical distance, i.e., construct a CDS whose corresponding distributions $E^0_{x,y}$ and $E^1_{x,y}$ are close when $D^0_{x,y}$ and $D^1_{x,y}$ are far, and vice versa. A classical result of Sahai and Vadhan [SV03] (building on Okamoto [Oka96]) provides such a reversing transformation for efficiently-samplable distributions (represented by their sampling circuits). As in the case of linear CDS, this transformation cannot be used directly since the resulting distributions do not "decompose" into an $x$-part and a $y$-part. Nevertheless, we can derive a decomposable version of the reversing transformation by employing a suitable PSM protocol. (See Section 4.2 for details.)

Theorems 2.1 and 2.2 have several interesting applications. Exploiting the ability to combine CDS's under AND/OR operations, we can further show that CDS is "closed" under (non-monotone) formulas, i.e., one can obtain a CDS for $h(f_1, \ldots, f_m)$ whose cost is essentially linear in the formula size of $h$ and polynomial in the CDS complexity of $f_i$. (See Section 4.3 for details.)

**Application to secret-sharing for forbidden graph.** Theorem 2.1 shows that the design of linear CDS for "very dense" functions that output 1 on all but $\ell$ inputs, reduces to the design of linear CDS for "very sparse" functions that output 1 on few, at most $\ell$, inputs. Since the latter case can be easily implemented via a linear CDS with $O(\ell)$ communication and randomness complexity, we get a linear CDS for $\ell$-dense functions with a similar complexity. Moving to the terminology of forbidden graph access structure, it follows that any highly dense bipartite $G$, obtained by removing at most $\ell$ edges from the complete bipartite graph over $N$ nodes, admits a secret sharing scheme whose total share size is at most $O(N + \ell)$. Moreover, if $G$ is obtained by removing $\ell$ edges from an arbitrary bipartite graph $F$, we get a complexity of $O(M + N + \ell)$ where $M$ is the total share size of the access structure associated with $F$. Indeed, this follows by the closure of CDS

under formulas; Letting $f, g$ denote the edge-indicator function of $F, G$, we can write $g(x, y)$ by the formula $f(x, y) \wedge p_1(x, y) \wedge \cdots \wedge p_\ell(x, y)$ where $p_i$ is a "point function" that evaluates to zero on the $i$-th edge that was excluded from $F$. Similar questions regarding the communication complexity of very dense forbidden graph access structure were recently studied by Beimel et al. [BFP16] for general (non-bipartite) graphs (see also [BFM16]). Our computational approach (as opposed to the combinatorial approach of [BFP16]) provides better bounds for the special case of bipartite graphs. (See Section 4.4 for details.)

## 2.2 Amplification

We move on to the study the robustness of CDS with respect to privacy and correctness errors. Borrowing tools from Sahai and Vadhan [SV03], it can be shown that CDS with constant correctness and privacy error of, say $1/3$, can be boosted into a CDS with an error of $2^{-k}$ at the expense of increasing the communication by a factor of $O(k^5)$. We show that in the context of CDS one can reduce the overhead to $O(k)$ and amortize it over long secrets.

**Theorem 2.3** (Amplification)**.** *A CDS $F$ for $f$ which supports a single-bit secret with privacy and correctness error of $1/3$, can be transformed into a CDS $G$ for $k$-bit secrets with privacy and correctness error of $2^{-\Omega(k)}$ and communication/randomness complexity which are larger than those of $F$ by a multiplicative factor of $O(k)$.*

The proof relies on constant-rate rump secret sharing schemes. (See Section 5.)

## 2.3 Amortizing CDS over long secrets

The above theorem suggests that there may be non-trivial savings when the secrets are long. We show that this is indeed the case, partially resolving an open question of Gay, Kerenidis, and Wee [GKW15].

**Theorem 2.4** (Amortization over long secrets)**.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a predicate. Then, for sufficiently large $m$, there exists a perfect linear CDS which supports $m$-bit secrets with total communication complexity of $O(nm)$.*

Recall that for a single-bit secret, the best known upper-bound for a general predicate is $O(2^{n/2})$ [BIKK14, GKW15]. In contrast, Theorem 2.4 yields an amortized complexity of $O(n)$ per each bit of the secret. The constant in the big-O notation is not too large (can be taken to be 12). Unfortunately, amortization kicks only when the value of $m$ is huge (double exponential in $n$). Achieving non-trivial savings for shorter secrets is left as an interesting open problem.

The proof of Theorem 2.4 is inspired by a recent result of Potechin [Pot16] regarding amortized space complexity.[2] Our proof consists of two main steps.

We begin with a *batch-CDS* scheme in which Alice holds a single input $x$, Bob holds a single input $y$, and both parties hold $2^{2^{2n}}$ secrets, one for each predicate in $\mathcal{F}_n = \{f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}\}$. The scheme releases the secret $s_f$ if and only if $f$ evaluates to 1 on $(x, y)$. Using a recursive construction, it is not hard to realize such a CDS with communication complexity of $O(n|\mathcal{F}_n|)$.

[2] In fact, Theorem 2.4 can be derived from Potechin's theorem by extending the connection between space-limited computation and CDS to the setting of multiple secrets. Instead, we present a self-contained proof which directly manipulates CDS and does not go through other computational models. This proof is arguably simpler, more instructive and yields (slightly) better amortized complexity.

Next, we use batch-CDS to get a CDS for a (single) predicate $f$ and a vector $s$ of $m = |\mathcal{F}_n|$ secrets, which is indexed by predicates $p \in \mathcal{F}_n$. We secret-share each bit $s_p$ into two parts $\alpha_p, \beta_p$ and collectively release all $\alpha_p$'s via batch-CDS (where $\alpha_p$ is associated with the predicate $p$). Finally, we collectively release all $\beta_p$'s via batch-CDS where $\beta_p$ is associated with the predicate $h_p$ that outputs 1 on $(x, y)$ if and only if $h$ and the target function $f$ agree on $(x, y)$. The key-observation is that $\alpha_p$ and $\beta_p$ are released if only if $f$ and $p$ evaluates to 1. As a result we get perfect privacy and *semi-correctness*: For 1-inputs $(x, y)$, exactly half of the secrets $s_p$ are released (the ones for which $p$ evaluates to 1). The latter property can be upgraded to perfect correctness by adding redundancy to the secrets (via a simple pre-encoding). See Section 6 for full details.

### 2.4 Linear lower-bound

We change gears and move from upper-bounds to lower-bounds. Specifically, we derive the first linear lower-bound on the communication complexity of general CDS.

**Theorem 2.5** (Lower-bound)**.** *There exists a predicate $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ for which any perfect (single-bit) CDS requires communication of at least $0.99n$.*

Previously the best known lower-bound for general CDS (due to [GKW15]) was logarithmic in $n$. As noted by [GKW15], an "insecure" realization of CDS requires a single bit, and so Theorem 2.5 provides a rare example of a provable linear gap in communication complexity between secure and insecure implementation of a natural task. (As argued in [GKW15], even super-constant gaps are typically out of reach.)

The proof of the lower-bound (given in Section 7) relies, again, on CDS manipulations. Consider a generalized version of CDS where the parties wish to release some Boolean function $f(x, y, s)$ defined over $x, y$ and the secret $s$. We show that one can construct such a "generalized CDS" for a function $f$ based on a standard CDS for a related predicate $g : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. In particular, we use a standard CDS to release the value of $s$ only if the residual function $f(x, y, \cdot)$ depends on $s$ (i.e., $g(x, y) = f(x, y, 0) \oplus f(x, y, 1)$). This way the output $f(x, y, s)$ can be always computed, either trivially, based on $x, y$ alone, or based on the additional knowledge of $s$, which is leaked when its value matters. Moreover, privacy is preserved since $s$ is leaked only when its value matters, which means that it can be derived anyway from $f(x, y, s)$ and $(x, y)$. We conclude that a lower-bound on CDS follows from a lower-bound on generalized-CDS. We then note that such a lower-bound essentially appears in the work of Feige, Kilian and Naor [FKN94]. Indeed, "generalized-CDS" can be equivalently viewed as a weakened version of private simultaneous message protocols for which the lower-bound of [FKN94] applies.[3]

### 2.5 CDS vs. linear CDS vs. communication complexity

Let us denote by $\mathsf{CDS}(f)$ the minimal communication complexity of CDS for $f$ with a single bit of secret and constant privacy/correctness error (say 0.1). We define $\mathsf{linCDS}(f)$ similarly with respect to linear CDS protocols.

We re-visit the connection between CDS-complexity and randomized communication complexity, and show that the former can be exponentially smaller than the latter. Since linear CDS complexity is at least polynomial in the communication complexity ($\mathsf{linCDS}(f) \geq \Omega(\sqrt{\mathsf{R}(f)})$), as

[3] CDS, generalized CDS, and PSM, can be all captured under the frameowrk of partial garbling studied by Ishai and Wee [IW14].

shown by [GKW15], we also conclude that general CDS can have exponentially-smaller communication than linear CDS.

**Theorem 2.6** (Separation)**.** *There exists an (explicit) partial function $f$ for which (1)* $\mathsf{CDS}(f) \leq O(\log \mathsf{R}(f))$ *and (2)* $\mathsf{CDS}(f) \leq O(\log \mathsf{linCDS}(f))$*.*

The first part of the theorem matches the lower-bound $\mathsf{CDS}(f) \geq \Omega(\log \mathsf{R}(f))$ established by [GKW15].[4] The second part provides the first separation between linear CDS and general (non-linear) CDS, resolving an open question of [GKW15].

The proof of Theorem 2.6 can be viewed as the communication complexity analog of Aaronson's [Aar12] oracle separation between the complexity class **SZK** of problems admitting *statistical-zero knowledge proofs* [GMR88], and the class **QMA** of problems admitting Quantum Merlin Arthur proofs. (See Section 8 for details.)

### 2.6 Discussion: The big picture

**CDS vs. SZK.** Our results highlight an important relation between conditional disclosure of secrets to statistical-zero knowledge protocols. A CDS protocol reduces the computation of $f(x, y)$, to an estimation of the statistical distance between a pair of "2-decomposable" distributions $D^0 = (a(x, 0; r), b(y, 0; r))$ and $D^1 = (a(x, 1; r), b(y, 1; r))$, similarly to the way that languages that admit a statistical zero-knowledge proofs are reduced to the analogous problem of estimating the statistical distance between a pair of efficiently-samplable distributions [SV03]. This simple insight has turned to be extremely useful for importing techniques from the domain of SZK to the CDS world.

**CDS: The low-end of information-theoretic protocols.** Determining the communication complexity of information-theoretic secure protocols is a fundamental research problem. Despite much efforts, we have very little understanding of the communication complexity of simple cryptographic tasks, and for most models, there are exponentially-large gaps between the best known upper-bounds to the best known lower-bounds. In an attempt to simplify the problem, one may try to focus on the most basic settings with a minimal non-trivial number of players (namely, 3) and the simplest possible communication pattern (e.g., single message protocols). Indeed, in this minimal communication model, conditional disclosure of secrets captures the notion of secret-sharing, just like private simultaneous message protocols (PSM) capture the notion of secure computation, and zero-information Arthur-Merlin games (ZAM) [GPW15] capture the notion of (non-interactive) zero-knowledge. Of all three variants, CDS is the simplest one: For any given predicate $f$ the CDS communication of $f$ is essentially upper-bounded by its ZAM complexity which is upper-bounded by its PSM complexity [AR16]. Hence, CDS should be the easiest model for obtaining upper-bounds (protocols) whereas PSM should be the easiest model for proving lower-bounds.

Our results, however, demonstrate that the current techniques for proving PSM lower-bounds [FKN94] also apply to the CDS model. The situation is even worse, since, by Theorem 2.4, the amortized communication complexity of CDS is indeed linear (per bit). We therefore conclude that proving a super-linear lower-bound in the PSM model requires a method that fails to lower-bound the amortized communication of CDS. Put differently, lower-bounds techniques which do not distinguish between PSM complexity and amortized CDS complexity cannot prove super-linear lower-bounds.

[4] The original lower-bound, which is stated for perfect CDS and for total functions, readily generalizes to partial functions and imperfect CDS. See Appendix A.

This "barrier" provides a partial explanation for the lack of strong (super-linear) lower-bounds for PSM. It will be interesting to further formalize this argument and present some syntactic criteria that determines whether a lower-bound technique is subject to the CDS barrier.

**Acknowledgement**

We would like to thank Amos Beimel and Hoteck Wee for useful discussions. We also thank Amos Beimel for pointing out the relevance of our results to secret-sharing schemes for dense forbidden graphs.

# 3 Preliminaries

Through the paper, real numbers are assumed to be rounded up when being typecast into integers ($\log n$ always becomes $\lceil \log n \rceil$, for instance). The *statistical distance* between two discrete random variables, $X$ and $Y$, denoted by $\Delta(X;Y)$ is defined by $\Delta(X;Y) := \frac{1}{2}\sum_z |\Pr[X=z] - \Pr[Y=z]|$. We will also use statistical distance for probability distributions, where for a probability distribution $D$ the value $\Pr[D=z]$ is defined to be $D(z)$.

## 3.1 Conditional disclosure of secrets

We define the notion of Conditional Disclosure of Secrets [GIKM00].

**Definition 3.1** (CDS)**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. Let $F_1 : \mathcal{X} \times \mathcal{S} \times \mathcal{R} \to \mathcal{T}_1$ and $F_2 : \mathcal{Y} \times \mathcal{S} \times \mathcal{R} \to \mathcal{T}_2$ be deterministic encoding algorithms, where $\mathcal{S}$ is the secret domain. Then, the pair $(F_1, F_2)$ is a* CDS *scheme for $f$ if the function $F(x,y,s,r) = (F_1(x,s,r), F_2(y,s,r))$ that corresponds to the joint computation of $F_1$ and $F_2$ on a common $s$ and $r$, satisfies the following properties:*

1. *($\delta$-Correctness) There exists a deterministic algorithm* Dec*, called a decoder, such that for every 1-input $(x,y)$ of $f$ and any secret $s \in \mathcal{S}$ we have that:*

$$\Pr_{r \xleftarrow{R} \mathcal{R}} [\mathsf{Dec}(x, y, F(x,y,s,r)) \neq s] \leq \delta$$

2. *($\varepsilon$-Privacy) There exists a simulator* Sim *such that for every 0-input $(x,y)$ of $f$ and any secret $s \in \mathcal{S}$: it holds that*

$$\Delta_{r \xleftarrow{R} \mathcal{R}} (\mathsf{Sim}(x,y) \ ; \ F(x,y,s,r)) \leq \varepsilon$$

*The* communication complexity *of the* CDS *protocol is $(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|)$ and its* randomness complexity *is $\log|\mathcal{R}|$. If $\delta$ and $\varepsilon$ are zeros, such a* CDS *scheme is called* perfect*.*

By default, we let $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$, $\mathcal{Z} = \{0,1\}$, $\mathcal{S} = \{0,1\}^s$, $\mathcal{R} = \{0,1\}^\rho$, $\mathcal{T}_1 = \{0,1\}^{t_1}$, and $\mathcal{T}_2 = \{0,1\}^{t_2}$ for positive integers $n, s, \rho, t_1$, and $t_2$.

**Linear CDS.** We say that a CDS scheme $(F_1, F_2)$ is *linear* over a finite field $\mathbb{F}$ (or simply linear) if, for any fixed input $(x, y)$, the functions $F_1(x, s, r)$ and $F_2(y, s, r)$ are linear over $\mathbb{F}$ in the secret $s$ and in the randomness $r$, where the secret, randomness, and messages are all taken to be vectors over $\mathbb{F}$ , i.e., $\mathcal{R} = \mathbb{F}^\rho$, $\mathcal{S} = \mathbb{F}^s$, $\mathcal{T}_1 = \mathbb{F}^{t_1}$ and $\mathcal{T}_2 = \mathbb{F}^{t_2}$. (By default, we think of $\mathbb{F}$ as the binary field, though our results hold over general fields.) Such a linear CDS can be canonically represented by a sequence of matrices $(M_x)_{x\in\mathcal{X}}$ and $(M_y)_{y\in\mathcal{Y}}$ where $M_x \in \mathbb{F}^{t_1\times(1+\rho)}$ and $M_y \in \mathbb{F}^{t_2\times(1+\rho)}$ and $F_1(x, s, r) = M_x \cdot \begin{pmatrix} s \\ \boldsymbol{r} \end{pmatrix}$ and $F_2(x, s, r) = M_y \cdot \begin{pmatrix} s \\ \boldsymbol{r} \end{pmatrix}$. It is not hard to show that any linear CDS with non-trivial privacy and correctness errors (smaller than 1) is actually perfect. Moreover, the linearity of the senders also implies that the decoding function is linear in the messages (cf. [GKW15]).[5]

**Definition 3.2.** *We denote by* $\mathsf{CDS}(f)$ *the least communication complexity of a* CDS *protocol for* $f$ *with* $\frac{1}{10}$*-correctness and* $\frac{1}{10}$*-privacy.* $\mathsf{linCDS}(f)$ *is defined analogously for linear* CDS *protocols.*

**Remark 3.3** (Input-dependent communication)**.** *We assume that the messages sent by the senders are of fixed length regardless of their inputs. That is, for every input* $x, y$ *and* $r$*, the output of* $F_1$ *and* $F_2$ *is always a string of fixed length. However, the CDS definition naturally generalizes to the case where these functions have different output-length per different inputs,* $x$ *and* $y$*, and so the protocol can have different communication per-input. Such a view is needed when one measures the total communication over all possible* $(x, y)$ *inputs. (As typically done in the context of graph-forbidden access structures). We note that most of our transformations remain meaningful in this model since they blow-up the communication on a per-input basis.*

## 3.2 Private simultaneous message protocols

We will also need the following model of information-theoretic non-interactive secure computation that was introduced by [FKN94], and was later named as *Private Simultaneous Message* (PSM) protocols by [IK97].

**Definition 3.4** (PSM)**.** *Let* $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ *be a function. We say that a pair of deterministic encoding algorithms* $F_1 : \mathcal{X} \times \mathcal{R} \to \mathcal{T}_1$ *and* $F_2 : \mathcal{Y} \times \mathcal{R} \to \mathcal{T}_2$ *are* PSM *for* $f$ *if the function* $F(x, y, r) = (F_1(x, r), F_2(y, r))$ *that corresponds to the joint computation of* $F_1$ *and* $F_2$ *on a common* $r$*, satisfies the following properties:*

1. *(*$\delta$*-Correctness) There exists a deterministic algorithm* Dec*, called decoder, such that for every input* $(x, y)$ *we have that:*

$$\Pr_{r\xleftarrow{R}\mathcal{R}}[\mathsf{Dec}(F(x, y, r)) \neq f(x, y)] \leq \delta.$$

2. *(*$\varepsilon$*-Privacy) There exists a randomized algorithm (simulator)* Sim *such that for any input* $(x, y)$ *it holds that:*

$$\Delta_{r\xleftarrow{R}\mathcal{R}}(\mathsf{Sim}(f(x, y)); F(x, y, r)) \leq \varepsilon.$$

[5] One can further consider a seemingly weaker form of linearity in which only the decoder is linear [GKW15]. Indeed, our separation between linear CDS and standard CDS applies to this setting as well.

*The* communication complexity *of the* PSM *protocol is defined as the total encoding length* $(\log|\mathcal{T}_1|+\log|\mathcal{T}_2|)$*, and the* randomness complexity *of the protocol is defined as the length* $\log|\mathcal{R}|$ *of the common randomness. If* $\delta$ *and* $\varepsilon$ *are zeros, such a* PSM *scheme is called* perfect*. The scheme is* balanced *[AIK04] if the simulator maps the uniform distribution over* $\mathcal{Z}$ *to the uniform distribution over* $\mathcal{T}=\mathcal{T}_1\times\mathcal{T}_2$ *and the decoder maps the uniform distribution over* $\mathcal{T}$ *to the uniform distribution over* $\mathcal{Z}$*.*

### 3.3 Randomized encoding and CDS encoding

When talking about PSM protocols, we will use $F(x,y,r)$ as abbreviation for $(F_1(x,r),F_2(y,r))$, and analogously for CDS. When we do not need to explicitly argue about the common randomness, we will suppress it as an argument to $F$ – that is, we will use $F(x,y)$ to denote the random variable produced by $F(x,y,r)$ for uniformly random $r$. Moreover, observe that the correctness and privacy conditions of both PSM and CDS are phrased as properties of the joint mapping $F$. One can therefore consider a *non-decomposable* CDS/PSM $F$ which respects privacy and correctness, but (possibly) fails to decompose into an $x$-part and a $y$-part (i.e., some of its outputs depend both on $x$ and $y$). In this case, we can ignore the partition of the input into $x,y$ and parse them as a single argument $w=(x,y)$. Following [IK00, AIK04] we refer to this generalization of PSM as *randomized encoding* of $f$, and to the generalized version of CDS as a CDS-encoding of $f$. The notion of perfect and balanced PSM and perfect and linear CDS carry naturally to this setting as well. These non-decomposable variants can be trivially realized (for PSM set $F(x,y)=f(x,y)$ and for CDS take $F(x,y,s)=f(x,y)\wedge s$). Nevertheless, they offer a useful abstraction. In particular, we will use these non-decomposable notions as a useful stepping stone towards obtaining a decomposable realization.

## 4 Closure properties

In this section, we establish several closure properties of CDS. We begin with closure under complement for linear CDS, then, extend the result to general CDS, and finally, prove that general and linear CDS are closed under $\mathbf{NC}^1$ circuits (or equivalently under Boolean formulas).

### 4.1 Reversing Linear CDS

We begin by proving Theorem 2.1 (restated here for the convenience of the reader).

**Theorem 4.1** (Linear CDS is closed under complement)**.** *Let* $f$ *be a function that has a linear* CDS *scheme* $F$ *with randomness complexity of* $\rho$ *field elements and communication complexity of* $t$ *field elements. Then, the complement function* $\overline{f}\equiv 1-f$ *has a linear* CDS *scheme with randomness complexity of* $(2t+\rho+1)$ *field elements and communication complexity of* $2(\rho+1)$ *field elements.*

*Proof.* Let $F_1,F_2$ be a linear CDS scheme for $f$ with randomness complexity $\rho$ and communication complexity $t$, where $F_1(x,s,\boldsymbol{c})$ and $F_2(y,s,\boldsymbol{c})$ are computed by applying matrices $M_x$ and $M_y$ to the vector $\begin{pmatrix}s\\ \boldsymbol{c}\end{pmatrix}$, respectively. We parse $M_x=(\boldsymbol{v}_x|T_x)$ and $M_y=(\boldsymbol{v}_y|T_y)$, i.e., $\boldsymbol{v}_x$ (resp., $\boldsymbol{v}_y$) denotes the first column of $M_x$ (resp., $M_y$), and $T_x$ (resp., $T_y$) denotes the remaining columns. In the following we fix $x,y$ to be some inputs and let $\boldsymbol{v}=\begin{pmatrix}\boldsymbol{v}_x\\ \boldsymbol{v}_y\end{pmatrix}$, $T=\begin{pmatrix}T_x\\ T_y\end{pmatrix}$, $M=(\boldsymbol{v}|T)$.

One can observe that due to the linearity of CDS, it holds that $f(x,y) = 0$ if and only if $\boldsymbol{v} \in \mathsf{colspan}(T)$. Indeed, the joint distribution of the messages, $M\boldsymbol{c}$, is uniform over the subspace $\mathcal{U}_s = \mathsf{colspan}(T) + s\boldsymbol{v}$. If $\boldsymbol{v} \in \mathsf{colspan}(T)$ the subspace $\mathcal{U}_s$ collapses to $\mathsf{colspan}(T)$ regardless of the value of $s$ (and so we get perfect privacy), whereas for $\boldsymbol{v} \notin \mathsf{colspan}(T)$, different secrets $s \neq s'$ induce disjoint subspaces $\mathcal{U}_s$ and $\mathcal{U}_{s'}$, and so the secret can be perfectly recovered.

Based on this observation, one can construct a *non-decomposable* CDS for $\overline{f}$ (in which Alice and Bob are viewed as a single party) as follows. Compute a random mask $\boldsymbol{\alpha}^T\boldsymbol{v}$ (where $\boldsymbol{\alpha}$ is a random vector), and output the masked secret bit $d = s + \boldsymbol{\alpha}^T\boldsymbol{v}$ together with the row vector $\boldsymbol{\gamma} = \boldsymbol{\alpha}^T T$. The decoding procedure starts by finding a vector $\boldsymbol{z}$ such that $\boldsymbol{v} = T\boldsymbol{z}$ (such a vector always exists since $\boldsymbol{v} \in \mathsf{colspan}(T)$ if $f(x,y) = 0$), and then outputs $d - \boldsymbol{\gamma}\boldsymbol{z} = s + \boldsymbol{\alpha}^T\boldsymbol{v} - (\boldsymbol{\alpha}^T T)\boldsymbol{z} = s$. Of course, the resulting scheme is not decomposable, however, we can fix this problem by letting Alice and Bob compute a PSM of it. We proceed with a formal description.

We construct CDS scheme $G = (G_1, G_2)$ for $\overline{f}$ as follows: Alice and Bob get shared randomness $q = (u, \boldsymbol{w}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$, where $u \in \mathbb{F}$, $\boldsymbol{w} \in \mathbb{F}^\rho$, and $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{F}^t$. Then they compute

$$G_1(x,s,q) = (\boldsymbol{\alpha}_1^T T_x + \boldsymbol{w}, \boldsymbol{\alpha}_1^T \cdot \boldsymbol{v}_x + u + s) \quad \text{and} \quad G_2(y,s,q) = (\boldsymbol{\alpha}_2^T T_y - \boldsymbol{w}, \boldsymbol{\alpha}_2^T \cdot \boldsymbol{v}_y - u).$$

The decoder on input $(\boldsymbol{m}_1, b_1)$ from Alice and $(\boldsymbol{m}_2, b_2)$ from Bob does the following: it finds a vector $\boldsymbol{z}$ such that $\boldsymbol{v} = T\boldsymbol{z}$ and outputs $b_1 + b_2 - (\boldsymbol{m}_1 + \boldsymbol{m}_2) \cdot \boldsymbol{z}$.

We now prove that the pair $(G_1, G_2)$ is a CDS for $\overline{f}$ starting with correctness. Fix an input $(x,y)$ for which $f(x,y) = 0$. Recall that in this case $\boldsymbol{v} \in \mathsf{colspan}(T)$, and so the decoder can find $\boldsymbol{z}$ as required. It is not hard to verify that in this case the decoding formula recovers the secret. Indeed, letting $\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}$, we have

$$b_1 + b_2 - (\boldsymbol{m}_1 + \boldsymbol{m}_2) \cdot \boldsymbol{z} = s + \boldsymbol{\alpha}^T \cdot \boldsymbol{v} - (\boldsymbol{\alpha}^T \cdot T) \cdot \boldsymbol{z} = s + \boldsymbol{\alpha}^T \cdot \boldsymbol{v} - \boldsymbol{\alpha}^T \cdot \boldsymbol{v} = s.$$

We now turn to proving the perfect privacy of the protocol. Fix some input $(x,y)$ such that $f(x,y) = 1$ and let $M = (\boldsymbol{v}|T)$ be the joint linear mapping. To prove privacy, it suffices to show that, for random $\boldsymbol{\alpha} \xleftarrow{R} \mathbb{F}^{2t}$, the first entry of the vector $\boldsymbol{\alpha}^T M$ is uniform conditioned on the other entries of the vector. To see this, first observe that $\boldsymbol{\alpha}^T M$ is distributed uniformly subject to the linear constraints $\boldsymbol{\alpha}^T M \cdot \boldsymbol{r} = 0$ induced by all vectors $\boldsymbol{r}$ in the Kernel of $M$. Therefore, $\boldsymbol{\alpha}^T\boldsymbol{v}$ is uniform conditioned on $\boldsymbol{\alpha}^T T$ if and only if all $\boldsymbol{r}$'s in the Kernel of $M$ have 0 as their first entry. Indeed, if this is not the case, then $\boldsymbol{v} \in \mathsf{colspan}(T)$, and so $(x,y)$ cannot be 1-input of $f$.

Finally, observe that the protocol consumes $(2t + \rho + 1)$ field elements for the joint randomness, and communicates a total number of $2\rho + 2$ field elements. □

**Remark 4.2** (Preserving input-dependent complexity)**.** *Theorem 4.1 extends to the input-dependent setting mentioned in Remark 3.3. Let $F$ be a linear CDS for $f$ and let $(M_x)_{x\in\mathcal{X}}$ and $(M_y)_{y\in\mathcal{Y}}$ denote its canonical representation. We measure the* effective randomness *of $F$ per input $x$ (resp., $y$) as the number of non-zero columns in $M_x$ (resp., $M_y$), excluding the first column $\boldsymbol{v}_x$ (resp., $\boldsymbol{v}_y$). Let $\rho_1(x)$ and $\rho_2(y)$ denote the effective randomness of the CDS $F$ for $f$ for an input $x$ and $y$, respectively. Then, Theorem 4.1 can be used to obtain a linear CDS $G'$ for $\overline{f}$ in which Alice communicates $\rho_1(x) + 1$ field elements for an input $x$, and Bob communicates $\rho_2(y) + 1$ field elements for an input $y$. The idea is to note that some of the bits sent in the CDS $G$ (defined in the proof of Theorem 4.1) can be omitted. Indeed, suppose that the $i$-th columns of the matrix $T_x$ equals to zero (recall that $T_x$ is obtained from $M_x$ by excluding the first column as in the proof of Theorem 4.1).*

*Then, in the CDS $G$, the $i$-th bit of Alice's message $\boldsymbol{\alpha}_1^T T_x + \boldsymbol{w}$ equals to $\boldsymbol{w}_i$ and so the receiver learns $\boldsymbol{w}_i$. As a result, we can just fix it to zero without losing privacy. However, now we can reduce the communication of Alice by omitting the $i$-th bit (i.e., $\boldsymbol{w}_i$) from the message. Applying this shrinkage to every redundant column in $T_x$ and every redundant column in $T_y$, yields the result.*

## 4.2 Reversing general CDS

We continue by proving Theorem 2.2 (restated below).

**Theorem 4.3** (CDS is closed under complement)**.** *Suppose that $f$ has a CDS with randomness complexity of $\rho$ and communication complexity of $t$ and privacy/correctness errors of $2^{-k}$. Then $\overline{f} \equiv 1 - f$ has a CDS scheme with similar privacy/correctness errors and randomness/communication complexity of $O(k^3\rho^2 t + k^3\rho^3)$.*

We begin with the following reversing transformation of Sahai and Vadhan [SV03, Corollary 4.18].

**Construction 4.4** (Statistical Distance Reversal)**.** *Let $D^0, D^1 : Q \to L$ be a pair of functions where $Q = \{0,1\}^\rho$ and $L = \{0,1\}^t$. For a parameter $k$, let $m = k^3\rho^2$, and let $\mathcal{H} = \{h : \{0,1\}^m \times Q^m \times L^m \to S\}$ be a family of 2-universal hash functions where $S = \{0,1\}^{(\rho+1)m-2(m/k)-k}$. The functions $C^0$ and $C^1$ take an input $(\vec{b}, \vec{r}, \vec{b'}, \vec{r'}, h, u) \in \left(\{0,1\}^m \times Q^m\right)^2 \times \mathcal{H} \times U$, and output the tuple*

$$(D^{\vec{b}}(\vec{r}), \vec{b}, h, z)$$

*where $D^{\vec{b}}(\vec{r}) =: (D^{b_1}(r_1), \ldots, D^{b_m}(r_m))$, and*

$$z = \begin{cases} h(\vec{b}, \vec{r}, D^{\vec{b'}}(\vec{r'})) & \text{for } C^0 \\ u & \text{for } C^1 \end{cases}.$$

In the following, we denote by $D^0$ (resp., $D^1, C^0, C^1$) the probability distributions induced by applying the function $D^0$ (resp., $D^1, C^0, C^1$) to a uniformly chosen input.

**Fact 4.5** (Corollary 4.18 in [SV03])**.** *In the set-up of Construction 4.4, the following holds for every parameter $k$.*

1. *If $\Delta(D^0, D^1) < 2^{-k}$ then $\Delta(C^0, C^1) > 1 - 2^{-k}$.*
2. *If $\Delta(D^0, D^1) > 1 - 2^{-k}$, then $\Delta(C^0, C^1) < 2^{-k}$.*

Fact 4.5 allows to transform a CDS $F(x, y, s, r) = (F_1(x, s; r), F_2(y, s; r))$ for the function $f$, into a *CDS encoding* $C$ for $\overline{f}$. For inputs $x, y$ and secret $s$, the CDS encoding $C$ samples a message from the distribution $C_{xy}^s$ obtained by applying Construction 4.4 to the distributions $D_{xy}^0 = F(x, y, 0, r)$ and $D_{xy}^1 = F(x, y, 1)$.

Unfortunately, the resulting CDS encoding is not decomposable since the hash function is applied jointly to the $x$-th and $y$-th components of the distributions $D_{xy}^0$ and $D_{xy}^1$. We fix the problem by using a PSM of $h$. Let us begin with the following more general observation that shows that $h$ can be safely replaced with its randomized encoding.

**Lemma 4.6.** *Under the set-up of Construction 4.4, for every $h \in \mathcal{H}$ let $\hat{h}$ be a perfect balanced randomized encoding of $h$ with randomness space $V$ and output space $\hat{S}$. The function $E^0$ (resp., $E^1$) is defined similarly to $C^0$ (resp., $C^1$) except that the input is $(\vec{b}, \vec{r}, \vec{b'}, \vec{r'}, h, v, \hat{s}) \in \left(\{0,1\}^m \times Q^m\right)^2 \times \mathcal{H} \times V \times \hat{S}$ and the output is identical except for the z-part which is replaced by*

$$\hat{z} = \begin{cases} \hat{h}(\vec{b}, \vec{r}, D^{\vec{b'}}(\vec{r'}); v) & \text{for } E^0 \\ \hat{s} & \text{for } E^1 \end{cases}.$$

*Then, the conclusion of Fact 4.5 holds for $E^0$ and $E^1$ as well. Namely, for every parameter $k$,*

1. *if $\Delta(D^0, D^1) < 2^{-k}$ then $\Delta(E^0, E^1) > 1 - 2^{-k}$;*
2. *if $\Delta(D^0, D^1) > 1 - 2^{-k}$, then $\Delta(E^0, E^1) < 2^{-k}$.*

*Proof.* Fix $D^0$ and $D^1$. We prove that $\Delta(E^0, E^1) = \Delta(C^0, C^1)$ and conclude the lemma from Fact 4.5. Indeed, consider the randomized mapping $T$ which maps a tuple $(a, b, h, z)$ to $(a, b, h, \mathsf{Sim}(z))$ where $\mathsf{Sim}$ is the simulator of the encoding $\hat{h}$. Then, by the perfect privacy and the balanced property, $T$ takes $C^0$ to $E^0$ and $C^1$ to $E^1$. Since statistical distance can only decrease when the same probabilistic process is applied to two random variables, it follows that $\Delta(C^0, C^1) \leq \Delta(T(C^0), T(C^1)) = \Delta(E^0, E^1)$. For the other direction, consider the mapping $T'$ which maps a tuple $(a, b, h, \hat{z})$ to $(a, b, h, \mathsf{Dec}(\hat{z}))$ where $\mathsf{Dec}$ is the decoder of the encoding. Then, by the perfect correctness and by the balanced property, $T$ takes $E^0$ to $C^0$ and $E^1$ to $C^1$. It follows that $\Delta(E^0, E^1) \leq \Delta(T'(E^0), T'(E^1)) = \Delta(C^0, C^1)$, and the lemma follows. □

We can now prove Theorem 4.3.

*Proof of Theorem 4.3.* Let $F = (F_1, F_2)$ be a CDS for the function $f$ with randomness complexity $\rho$, communication $t$ and privacy/correctness error of $2^{-k}$. For inputs $x, y$ and secret $\sigma$, the CDS for $\overline{f}$ will be based on the functions $E^\sigma$ defined in Lemma 4.6 where $D^0(r) = F(x, y, 0; r)$ and $D^1(r) = F(x, y, 1; r)$. In particular, we will instantiate Lemma 4.6 as follows.

Let

$$\alpha = (\vec{b}, \vec{r}, D^{\vec{b}}_{xy}(\vec{r})), \quad \text{where } D^{\vec{b}}_{xy}(\vec{r}) := (F(x, y, b_1; r_1), \ldots, F(x, y, b_m; r_m))$$

be the input to the hash function $h$. Let $n_0 = m(1+\rho+t)$ denote the length of $\alpha$. Observe that each bit of $\alpha$ depends either on $x$ or on $y$ but not in both (since $F$ is a CDS). Let $A \subset [n_0]$ denote the set of entries which depend on $x$ and let $B = [n_0] \setminus A$ be its complement. Let $n_1 = (\rho+1)m - 2m/k - k$ denote the output length of the hash function family $\mathcal{H}$. We implement $\mathcal{H} = \{h\}$ by using Toeplitz matrices. That is, each function is defined by a binary Toeplitz matrix $M \in \mathbb{F}_2^{n_1 \times n_0}$ (in which each descending diagonal from left to right is constant) and a vector $w \in \mathbb{F}_2^{n_1}$, and $h(\alpha) = M\alpha + w$. Let us further view the hash function $h(\alpha)$ as a two-argument function $h(\alpha_A, \alpha_B)$ and let

$$\hat{h}(\alpha_A, \alpha_B; v) = (M_A\alpha_A + w + v, M_B\alpha_B - v),$$

where $v \in \mathbb{F}_2^{n_1}$ and $M_A$ (resp. $M_B$) is the restriction of $M$ to the columns in $A$ (resp., columns in $B$). It is not hard to verify that $\hat{h}$ is a perfect balanced PSM for $h$. (Indeed decoding is performed by adding Alice's output to Bob's output, and simulation is done by splitting an output $\beta$ of $h$ into two random shares $c_1, c_2 \in \mathbb{F}_2^{n_1}$ which satisfy $c_1 + c_2 = \beta$.)

Consider the randomized mapping $E^{\sigma}_{xy}$ obtained from Lemma 4.6 instantiated with $D^s(r) = D^s_{xy}(r) = F(x, y, s; r)$ and the above choices of $\hat{h}$. We claim that $E^{\sigma}_{xy}$ is a CDS for $\overline{f}$ with privacy and correctness error of $2^{-k}$. To see this first observe that, by construction, the output of $E^{\sigma}_{xy}$ can be decomposed into an $x$-component $E_1(x, \sigma)$ and a $y$-component $E_2(y, \sigma)$. (All the randomness that is used as part of the input to $E$ is consumed as part of the joint randomness of the CDS.)

To prove privacy, fix some 0-input $(x, y)$ of $\overline{f}$ and note that $f(x, y) = 1$ and therefore, by the correctness of the CDS $F$, it holds that $\Delta(D^0_{xy}, D^1_{xy}) > 1 - 2^{-k}$. We conclude, by Lemma 4.6, that $\Delta(E^0_{xy}, E^1_{xy}) < 2^{-k}$ and privacy holds. For correctness, fix some 1-input $(x, y)$ of $\overline{f}$ and note that $f(x, y) = 0$ and therefore, by the privacy of the CDS $F$, it holds that $\Delta(D^0_{xy}, D^1_{xy}) < 2^{-k}$. We conclude, by Lemma 4.6, that $\Delta(E^0_{xy}, E^1_{xy}) > 1-2^{-k}$ and so correctness holds (by using the optimal distinguisher as a decoder). Finally, since the description length of $h$ is $n_0 + 2n_1$ the randomness complexity of $\hat{h}$ is $n_1$ and the communication complexity of $\hat{h}$ is $2n_1$, the overall communication and randomness complexity of the resulting CDS is $O(k^3\rho^2 t + k^3\rho^3)$. □

### 4.3 Closure under formulas

Closure under formulas can be easily deduced from Theorems 4.1 and 4.3.

**Theorem 4.7.** *Let $g$ be a boolean function over $m$ binary inputs that can be computed by a $\sigma$-size formula. Let $f_1, \ldots, f_m$ be $m$ boolean functions over $\mathcal{X} \times \mathcal{Y}$ each having a $\mathsf{CDS}$ with $t$ communication and randomness complexity, and $2^{-k}$ privacy and correctness errors. Then, the function $h : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ defined by $g(f_1(x, y), \ldots, f_m(x, y))$ has a $\mathsf{CDS}$ scheme with $O(\sigma k^3 t^3)$ randomness and communication complexity, and $\sigma 2^{-k}$ privacy and correctness errors. Moreover, in the case of linear CDS, the communication and randomness complexity are only $O(\sigma t)$ and the resulting CDS is also linear.*

*Proof.* Without loss of generality, assume that the formula $g$ is composed of AND and OR gates and all the negations are at the bottom (this can be achieved by applying De Morgan's laws) and are not counted towards the formula size. We prove the theorem with an upper-bound of $\sigma \cdot Ck^3t^3$ where $C$ is the constant hidden in the big-O notation in Theorem 4.3 (the upper-bound on the communication/ranodmness complexity of the complement of a CDS).

The proof is by induction on $\sigma$. For $\sigma = 1$, the formula $g$ is either $f_i(x, y)$ or $\overline{f}_i(x, y)$ for some $i \in [m]$, in which case the claim follows either from our assumption on the CDS for $f_i$ or from Theorem 4.3. To prove the induction step, consider a $\sigma$-size formula $g(f_1, \ldots, f_m)$ of the form $g_1(f_1, \ldots, f_m) \diamond g_2(f_1, \ldots, f_m)$ where $\diamond$ is either AND or OR, $g_1$ and $g_2$ are formulas of size $\sigma_1$ and $\sigma_2$, respectively, and $\sigma = \sigma_1 + \sigma_2 + 1$. For the case of an AND gate, we additively secret share the secret $s$ into random $s_1$ and $s_2$ subject to $s_1 + s_2 = s$ and use a CDS for $g_1$ with secret $s_1$ and for $g_2$ for the secret $s_2$. For the case of OR gate, use a CDS for $g_1$ with secret $s$ and for $g_2$ for the secret $s$. By the induction hypothesis, the communication and randomness complexity are at most $\sigma_1 \cdot Ck^3t^3 + \sigma_2 \cdot Ck^3t^3 + 1 \leq \sigma Ck^3t^3$, and the privacy/correctness error grow to $\sigma_1 2^{-k} + \sigma_2 2^{-k} \leq \sigma 2^{-k}$, as required.

The extension to the linear case follows by plugging the upper-bound from Theorem 4.1 to the basis of the induction, and by noting that the construction preserves linearity. □

### 4.4 Application: secret sharing for dense forbidden graphs

As explained in the introduction CDS schemes naturally correspond to secret sharing schemes for forbidden bipartite graphs. Recently, Beimel et al. [BFP16] studied the complexity of such schemes for the case of dense graphs in which all but few edges exist. In the setting of CDS, this corresponds to functions which are almost always 1.

**Corollary 4.8.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a predicate which takes the value 1 on all but $\ell$ inputs. Then, there exists a linear CDS for $f$ for which the total communication complexity over all inputs $x$ and $y$ is $2(N+\ell)$ where $N = 2^n$. More generally, for any predicate $h : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with CDS of total complexity of $M$, there exists a CDS for $h \wedge f$ with total complexity of $M + 2(N+\ell)$.*

*Proof.* Consider the following simple CDS for the complement $g$ of $f$. Let $(x_i, y_i)_{i\in[\ell]}$ be the points accepted by $g$. The shared randomness consists of $\ell$ random bits $r = (r_i)_{i\in[\ell]}$. Given a secret $s$ and input $x$ and $y$, Alice outputs $(s + r_i)_{i:x_i=x}$ and Bob outputs $(r_i)_{i:y_i=y}$. It is not hard to verify that this is a linear CDS for $g$. Moreover, observe that the effective randomness complexity per $x$ is exactly $D_1(x)$, the number of tuples $(x_i, y_i)$ in which $x$ appears. Similarly, the effective randomness complexity per $y$ is exactly $D_2(y)$, the number of tuples $(x_i, y_i)$ in which $y$ appears. Applying the input-dependent version of Theorem 4.1 (as outlined in Remark 4.2), yields a CDS for $f$, in which the communication complexity per $x$ is $D_1(x)+1$ and per $y$ is $D_2(y)+1$. Summing-up over all $x$ and $y$, yields a total communication complexity of $\sum_{x\in\{0,1\}^n} D_1(x)+1+\sum_{y\in\{0,1\}^n} D_2(y)+1 = 2(2^n+\ell)$, as required.

For the "Moreover" part, we secret share $s$ to random $s_1$ and $s_2$ subject to $s_1 + s_2 = s$ and use the $h$-CDS for releasing $s_1$ and the $f$-CDS for releasing $s_2$. (This can be viewed as a special case of Theorem 4.7.) □

Our results correspond to secret sharing schemes for dense forbidden *bipartite graphs* with $N = 2^n$ nodes on each side and where $\ell$ edges are removed, either from the complete graph, or from the graph that represents the function $h$. For general (possibly non-bipartite) dense graphs with $\binom{N}{2} - \ell$ edges, Beimel et al. [BFP16] provide an upper-bound of $O(N + \ell^{7/6})$ for $\ell < N$ and $O(N^{7/6+2\beta/2})$ for $\ell = N^{1+\beta}, \beta \leq 1/2$. The latter result generalizes to the case where $\ell = N^{1+\beta}$ edges are removed from an arbitrary graph $H$ at the expense of an additive overhead of $M$ – the total share-size of a secret-sharing realizing $H$. The bounds in Corollary 4.8 are better, though they apply to the more restrictive setting of bipartite graphs.

## 5 Amplifying correctness and privacy of CDS

In this section we show how to simultaneously reduce the correctness and privacy error of a CDS scheme $F$. Moreover, the transformation has only minor cost when applied to long secrets.

**Theorem 5.1.** *Let $f : X \times Y \to \{0,1\}$ be a predicate and let $F$ be a CDS for $f$ which supports 1-bit secrets with correctness error $\delta_0 = 0.1$ and privacy error $\varepsilon_0 = 0.1$. Then, for every integer $k$ there exists a CDS $G$ for $f$ with $k$-bit secrets, privacy and correctness errors of $2^{-\Omega(k)}$. The communication (resp., randomness) of $G$ larger than those of $F$ by a multiplicative factor of $O(k)$.*

*Proof.* Let $\varepsilon$ be some constant larger than $\varepsilon_0$. Let $\mathsf{E}$ be a randomized mapping that takes $k$-bit message $s$ and $O(k)$-random string into an encoding $c$ of length $m = \Theta(k)$ with the following properties:

1. If $\mathsf{E}(s)$ is passes through a Binary Symmetric Channel with crossover probability $\delta_0$ then $s$ can be recovered with probability $1 - \exp(-\Omega(k))$.

2. For any pair of secrets $s$ and $s'$ and any set $T \subset [m]$ of size at most $\varepsilon m$, the $T$-restricted encoding of $s$ is distributed identically to the $T$-restricted encoding of $s'$, i.e., $(\mathsf{E}(s)_i)_{i \in T} \equiv (\mathsf{E}(s')_i)_{i \in T}$.

That is, $\mathsf{E}$ can be viewed as a ramp secret-sharing scheme with 1-bit shares which supports robust reconstruction. Such a scheme can be based on any linear error-correcting code with good dual distance [CCG+07]. In particular, by using a random linear code, we can support $\varepsilon_0 = \delta_0 = 0.1$ or any other constants which satisfy the inequality $1 - H_2(\delta) > H_2(\varepsilon)$.

Given the CDS $F = (F_1, F_2)$ we construct a new CDS $G = (G_1, G_2)$ as follows. Alice and Bob jointly map the secret $s \in \{0,1\}^k$ to $c = \mathsf{E}(s; r_0)$ (using joint randomness $r_0$). Then, for every $i \in [m]$, Alice outputs $F_1(x, c_i; r_i)$ and Bob outputs $F_2(y, c_i; r_i)$, where $r_1, \ldots, r_m$ are given as part of the shared randomness.

Let us analyze the correctness of the protocol. Fix some $x, y$ for which $f(x, y) = 1$. Consider the decoder which given $(v_1, \ldots, v_m)$ and $x, y$ applies the original decoder of $F$ to each coordinate separately (with the same $x, y$), and passes the result $\hat{c} \in \{0,1\}^m$ to the decoding procedure of $\mathsf{E}$, promised by Property (1) above. By the correctness of $F$, each bit $\hat{c}_i$ equals to $c_i$ with probability of at least $1 - \delta_0$. Therefore, the decoder of $\mathsf{E}$ recovers $c$ with all but $1 - \exp(-\Omega(k))$ probability.

Consider the simulator which simply applies $G$ to the secret $s' = 0^k$. Fix $x$ and $y$ and a secret $s$. To upper-bound the statistical distance between $G(x, y, s')$ and $G(x, y, s')$, we need the following standard "coupling fact" (cf. [MPR07, Lemma 5] for a similar statement).

**Fact 5.2.** *Any pair of distributions, $(D_0, D_1)$ whose statistical distance is $\varepsilon$ can be coupled into a joint distribution $(E_0, E_1, b)$ with the following properties:*

1. *The marginal distribution of $E_0$ (resp., $E_1$) is identical to $D_0$ (resp., $D_1$).*

2. *$b$ is an indicator random variable which takes the value 1 with probability $\varepsilon$.*

3. *Conditioned on $b = 0$, the outcome of $E_0$ equals to the outcome of $E_1$.*

Define the distributions $D_0 := F(x, y, 0)$ and $D_1 := F(x, y, 1)$, and let $(E_0, E_1, b)$ be the coupled version of $D_0, D_1$ derived from Fact 5.2. Let $c = \mathsf{E}(s)$ and $c' = \mathsf{E}(s')$. Then,

$$G(x, y, s) = (E^1_{c_1}, \ldots, E^m_{c_m}),$$

and

$$G(x, y, s') = (E^1_{c'_1}, \ldots, E^m_{c'_m}),$$

where for each $i \in [m]$ the tuple $(E^i_0, E^i_1, b^i)$ is sampled jointly and independently from all other tuples. Let $T = \{i \in [m] : b_i \neq 0\}$. Then, it holds that

$$\begin{aligned} \Delta(G(x, y, s); G(x, y, s')) &\leq \Delta((T, (E^i_{c_i})_{i \in T}); (T, (E^i_{c'_i})_{i \in T})) \\ &\leq \Pr[|T| > \varepsilon m] \leq \exp(-\Omega(k)), \end{aligned}$$

where the first inequality follows from Fact 5.2, the second inequality follows from the second property of $\mathsf{E}$ and the last inequality follows from a Chernoff bound (recalling that $\varepsilon - \varepsilon' > 0$ is a constant and $m = \Theta(k)$). The theorem follows. □

**Remark 5.3** (Optimization)**.** *The polarization lemma of Sahai and Vadhan [SV03] provides an amplification procedure which works for a wider range of parameters. Specifically, their transformation can be applied as long as the initial correctness and privacy errors satisfy the relation $\delta_0^2 > \varepsilon_0$. (Some evidence suggest that this condition is, in fact, necessary for any amplification procedure [HR05].) Unfortunately, the communication overhead in their reduction is polynomially larger than ours and does not amortize over long secrets. It is not hard to combine the two approaches and get the best of both worlds. In particular, given a CDS with constant correctness and privacy errors which satisfy $\delta_0^2 > \varepsilon_0$, use the polarization lemma with constant security parameter $k_0$ to reduce the errors below the threshold needed for Theorem 5.1, and then use the theorem to efficiently reduce the errors below $2^{-k}$. The resulting transformation has the same asymptotic tradeoff between communication, error, and secret length, and can be used for a wider range of parameters. (This, in particular, yields the statement of Theorem 2.3 in the introduction in which $\delta_0$ and $\varepsilon_0$ are taken to be $1/3$).*

**Remark 5.4** (Preserving efficiency)**.** *Theorem 5.1 preserves efficiency (of the CDS senders and decoder) as long as the encoding $\mathsf{E}$, and its decoding algorithm are efficient. This can be guaranteed by replacing the random linear codes (for which decoding is not know to be efficient) with an Algebraic Geometric Codes (as suggested in [CCG+07]; see also Claim 4.1 in [IKOS09] and [GS96, CC06]). This modification requires to start with smaller (yet constant) error probabilities $\delta_0, \varepsilon_0$. As in Remark 5.3, this limitation can be easily waived. First use the inefficient transformation (based on random binary codes) with constant amplification $k_0 = O(1)$ to reduce the privacy/correctness error below the required threshold, and then use the efficient amplification procedure (based on Algebraic Geometric Codes).*

# 6 Amortizing the communication for long secrets

In this section we show that, for sufficiently long secrets, the amortized communication cost of CDS for $n$-bit predicates is $O(n)$ bits per each bit of the secret. As explained in the introduction, in order to prove this result we first amortize CDS over many different predicates (applied to the same input $(x, y)$). We refer to this version of CDS as *batch-CDS*, formally defined below.

**Definition 6.1** (batch-$\mathsf{CDS}$)**.** *Let $\mathcal{F} = (f_1, \ldots, f_m)$ be an $m$-tuple of predicates over the domain $\mathcal{X} \times \mathcal{Y}$. Let $F_1 : \mathcal{X} \times \mathcal{S}^m \times \mathcal{R} \to \mathcal{T}_1$ and $F_2 : \mathcal{Y} \times \mathcal{S}^m \times \mathcal{R} \to \mathcal{T}_2$ be deterministic encoding algorithms, where $\mathcal{S}$ is the secret domain (by default $\{0,1\}$). Then, the pair $(F_1, F_2)$ is a batch-$\mathsf{CDS}$ scheme for $\mathcal{F}$ if the function $F(x, y, s, r) = (F_1(x, s, r), F_2(y, s, r))$, that corresponds to the joint computation of $F_1$ and $F_2$ on a common $s$ and $r$, satisfies the following properties:*

1. *(Perfect correctness)*[6] *There exists a deterministic algorithm $\mathsf{Dec}$, called a decoder, such that for every every $i \in [m]$, every 1-input $(x, y)$ of $f_i$ and every vector of secrets $s \in \mathcal{S}^m$, we have that:*

$$\Pr_{r \xleftarrow{R} \mathcal{R}} [\mathsf{Dec}(i, x, y, F(x, y, s, r)) = s_i] = 1.$$

[6] For simplicity, we consider only perfectly correct and perfectly private batch-CDS, though the definition can be generalized to the imperfect case as well.

2. *(Perfect privacy) There exists a simulator* $\mathsf{Sim}$ *such that for every input* $(x, y)$ *and every vector of secrets* $s \in \mathcal{S}^m$*, the following distributions are identical*

$$\mathsf{Sim}(x, y, \hat{s}) \qquad \textit{and} \qquad F(x, y, s, r),$$

*where* $r \stackrel{R}{\leftarrow} \mathcal{R}$ *and* $\hat{s}$ *is an* $m$*-long vector whose* $i$*-th component equals to* $s_i$ *if* $f_i(x, y) = 1$*, and* $\perp$ *otherwise.*

*The* communication complexity *of the* $\mathsf{CDS}$ *protocol is* $(\log |\mathcal{T}_1| + \log |\mathcal{T}_2|)$.

In the following, we let $\mathcal{F}_n$ denote the $2^{2^{2n}}$-tuple which contains all predicates $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined over pairs of $n$-bit inputs (sorted according to some arbitrary order).

**Lemma 6.2.** *$\mathcal{F}_n$-batch CDS can be implemented with communication complexity of $3|\mathcal{F}_n|$. Moreover the protocol is linear.*

*Proof.* The proof is by induction on $n$. For $n = 1$, it is not hard to verify that any predicate $f : \{0,1\} \times \{0,1\} \to \{0,1\}$ admits a CDS with a total communication complexity of at most 2 bits. Indeed, there are 16 such predicates, out of which, six are trivial in the sense that the value of $f$ depends only in the inputs of one of the parties (and so they admit a CDS with 1 bit of communication), and the other ten predicates correspond, up to local renaming of the inputs, to AND, OR, and XOR, which admit simple linear 2-bit CDS as follows. For AND, Alice and Bob send $s \cdot x + r$ and $r \cdot y$; for OR, they send $x \cdot s$ and $y \cdot s$; and, for XOR, they send $s + x \cdot r_1 + (1 - x)r_2$ and $y \cdot r_2 + (1 - y)r_1$ (where $r$ and $(r_1, r_2)$ are shared random bits and addition/multiplication are over the binary field). It follows, that $\mathcal{F}_1$-batch CDS can be implemented with total communication of at most $2|\mathcal{F}_1|$. (In fact, this bound can be improved by exploiting the batch mode.)

Before proving the induction step. Let us make few observations. For $(\alpha, \beta) \in \{0,1\}^2$, consider the mapping $\phi_{\alpha,\beta} : \mathcal{F}_{n+1} \to \mathcal{F}_n$ which maps a function $f \in \mathcal{F}_{n+1}$ to the function $g \in \mathcal{F}_n$ obtained by restricting $f$ to $x_{n+1} = \alpha$ and $y_{n+1} = \beta$. The mapping $\phi_{\alpha,\beta}$ is onto, and is $D$-to-1 where $D = |\mathcal{F}_{n+1}|/|\mathcal{F}_n|$. We can therefore define a mapping $T_{\alpha,\beta}(f)$ which maps $f \in \mathcal{F}_{n+1}$ to $(g, i) \in \mathcal{F}_n \times [D]$ such that $f$ is the $i$-th preimage of $g$ under $\phi_{\alpha,\beta}$ with respect to some fixed order on $\mathcal{F}_{n+1}$. By construction, for every fixed $(\alpha, \beta)$, the mapping $T_{\alpha,\beta}$ is one-to-one.

We can now prove the induction step; That is, we construct $\mathcal{F}_{n+1}$-batch CDS based on $D$ copies of $\mathcal{F}_n$-batch CDS. Given input $x \in \{0,1\}^{n+1}$ for Alice, $y \in \{0,1\}^{n+1}$ for Bob, and joint secrets $(s_f)_{f \in \mathcal{F}_{n+1}}$, the parties proceed as follows.

1. Alice and Bob use $D$ copies of $\mathcal{F}_n$-batch CDS with inputs $x' = (x_1, \ldots, x_n)$ and $y' = (y_1, \ldots, y_n)$. In the $i$-th copy, for every predicate $g \in \mathcal{F}_n$, a random secret $r_{g,i} \in \{0,1\}$ is being used. (The $r_{g,i}$'s are taken from the joint randomness of Alice and Bob.)

2. For every $f \in \mathcal{F}_{n+1}$ and $(\alpha, \beta) \in \{0,1\}^2$, Alice and Bob release the value $\sigma_{f,\alpha,\beta} = s_f + r_{g,i}$ where $(g, i) = T_{\alpha,\beta}(f)$ iff the last bits of their inputs, $x_{n+1}$ and $y_{n+1}$, are equal to $\alpha$ and $\beta$, respectively. This step is implemented as follows. For each $f$, Alice sends a pair of bits

$$c_{f,0} = \sigma_{f,x_{n+1},0} + r'_{f,0}, \quad \text{and} \quad c_{f,1} = \sigma_{f,x_{n+1},1} + r'_{f,1},$$

and Bob sends $r'_{f,y_{n+1}}$ where $r'_{f,0}, r'_{f,1}$ are taken from the joint randomness.

The decoding procedure is simple. If the input $(x, y) \in \{0,1\}^{n+1} \times \{0,1\}^{n+1}$ satisfies $f \in \mathcal{F}_{n+1}$, the decoder does the following: (1) Computes $(g, i) = T_{x_{n+1}, y_{n+1}}(f)$ and retrieves the value of $r_{g,i}$ which is released by the batch-CDS since $g(x', y') = f(x, y) = 1$; (2) Collects the values $c_{f,x_{n+1}}$ and $r'_{f,y_{n+1}}$ sent during the second step, and recovers the value of $s_f$ by computing $c_{f,x_{n+1}} - r'_{f,y_{n+1}} - r_{g,i}$.

In addition, it is not hard to verify that perfect privacy holds. Indeed, suppose that $(x, y) \in \{0,1\}^{n+1} \times \{0,1\}^{n+1}$ does not satisfy $f$. Then, the only $s_f$-dependent value which is released is $s_f \oplus r_{g,i}$ where $g$ is the restriction of $f$ to $(x_{n+1}, y_{n+1})$. However, since $(x, y)$ fails to satisfy $f$, its prefix does not satisfy $g$ and therefore $r_{g,i}$ remains hidden from the receiver. Formally, we can perfectly simulate the view of the receiver as follows. First simulate the first step using $D$ calls to the simulator of $\mathcal{F}_n$-batch CDS with random secrets $r_{g,i}$. Then simulate the second step by sampling, for each $f$, three values $c_{f,0}, c_{f,1}$ and $r'$ which are uniform if $f(x, y) = 0$, and, if $f(x, y) = 1$, satisfy the linear constraint $s_f = c_{f,x_{n+1}} - r'_{f,y_{n+1}} - r_{g,i}$ where $(g, i) = T_{\alpha,\beta}(f)$.

Finally the communication complexity equals to the complexity of $D$ copies of batch CDS for $\mathcal{F}_n$ (communicated in the first step) plus $3|\mathcal{F}_{n+1}|$ bits (communicated at the second step). Therefore, by the induction hypothesis, the overall communication, is $3|\mathcal{F}_{n+1}| + 3Dn|\mathcal{F}_n|$. Recalling that $D = |\mathcal{F}_{n+1}|/|\mathcal{F}_n|$, we derive an upper-bound of $3(n+1)|\mathcal{F}_{n+1}|$, as required. □

We use Lemma 6.2 to amortize the complexity of CDS over long secrets.

**Theorem 6.3.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a predicate. Then, for $m = |\mathcal{F}_n|/2 = 2^{2^{2n}}/2$, there exists a perfect linear CDS which supports $m$-bit secrets with total communication complexity of $12nm$.*

The case of longer secrets of length $m > |\mathcal{F}_n|/2$ (as in Theorem 2.4) can be treated by partitioning the secret to $|\mathcal{F}_n|/2$-size blocks and applying the CDS for each block separately. The overall communication complexity is upper-bounded by $13nm$.

*Proof.* Given a vector $S$ of $m = |\mathcal{F}_n|/2$ secrets, we duplicate each secret twice and index the secrets by predicates $p \in \mathcal{F}_n$ such that $s_p = s_{\bar{p}}$ (i.e., a predicate and its complement index the same secret). On inputs $x, y$, Alice and Bob make two calls to $\mathcal{F}_n$-batch CDS (with the same inputs $x, y$). In the first call the secret associated with a predicate $p \in \mathcal{F}_n$ is a random values $r_p$. In the second call, for every predicate $h \in \mathcal{F}_n$, we release the secret $s_p \oplus r_p$ where $p$ is the unique predicate for which $p = f + h + 1$ (where addition is over the binary field).

**Correctness**. Suppose that $f(x, y) = 1$. Recall that each of the original secrets $S_i$ appears in two copies $(s_p, s_{\bar{p}})$ for some predicate $p$. Since one of these copies is satisfied by $(x, y)$, it suffices to show that, whenever $p(x, y) = 1$, the secret $s_p$ can be recovered. Indeed, for such a predicate $p$, the value $r_p$ is released by the first batch-CDS, and the value $s_p \oplus r_p$ is released by the second batch-CDS. The latter follows by noting that the predicate $h$ which satisfies $p = f + h + 1$ is also satisfied, since $h(x, y) = p(x, y) + f(x, y) + 1 = 1$. It follows that $s_p$ can be recovered for every $p$ which is satisfied by $(x, y)$, as required.

**Privacy**. Suppose that $f(x, y) = 0$. We show that all the "virtual secrets" $s_p$ remain perfectly hidden in this case. Indeed, for $h$ and $p$ which satisfy $p = f + h + 1$, it holds that, whenever $f(x, y) = 0$, either $h(x, y) = 0$ or $p(x, y) = 0$, and therefore, for any $p$, either $r_p$ or $s_p \oplus r_p$ are released, but never both.

Finally, using Lemma 6.2, the total communication complexity of the protocol is $2 \cdot 3 \cdot n \cdot |\mathcal{F}_n| = 12nm$, as claimed. □

# 7 A Linear Lower Bound on CDS

Here we show that the lower bound on the communication complexity of PSM protocols proven in [FKN94] can be extended to apply for CDS as well. We do this by showing how to use CDS protocols to construct PSM protocols that are only required to hide a certain small pre-specified set of input bits (as opposed to the whole input). We define this notion of PSM below.

**Definition 7.1** ($b$-bit PSM)**.** *Consider a function $f : (\mathcal{W} \times \mathcal{X}) \times \mathcal{Y} \to \mathcal{Z}$, with $\log|\mathcal{W}| \geq b$ for some $b > 0$. We say that a pair of deterministic encoding algorithms $F_1 : \mathcal{W} \times \mathcal{X} \times \mathcal{R} \to \mathcal{T}_1$ and $F_2 : \mathcal{Y} \times \mathcal{R} \to \mathcal{T}_2$ constitute a $b$-bit PSM for $f$ if the function $F((w,x),y,r) = (F_1(w,x,r), F_2(y,r))$ satisfies the following properties:*

1. *($\delta$-Correctness) There exists a deterministic algorithm Dec, called the* decoder*, such that for every input $((w,x),y)$ we have that:*

$$\Pr_{r \xleftarrow{R} \mathcal{R}}[\mathsf{Dec}(F((w,x),y,r)) \neq f((w,x),y)] \leq \delta.$$

2. *($b$-bit $\varepsilon$-Privacy) There exists a randomized algorithm Sim such that for any input $((w,x),y)$ it holds that:*

$$\underset{r \xleftarrow{R} \mathcal{R}}{\Delta}(\mathsf{Sim}(f((w,x),y),x,y); F((w,x),y,r)) \leq \varepsilon.$$

*The communication complexity of the protocol is defined as the total encoding length $(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|)$, and the randomness complexity of the protocol is defined as $\log|\mathcal{R}|$.*

*By default, the above sets are to be taken to be $\mathcal{W} = \{0,1\}^b$, $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$, $\mathcal{Z} = \{0,1\}$, $\mathcal{R} = \{0,1\}^\rho$, $\mathcal{T}_1 = \{0,1\}^{t_1}$, and $\mathcal{T}_2 = \{0,1\}^{t_2}$ for some positive integers $b, n, \rho, t_1$, and $t_2$.*

**Lemma 7.2** (CDS to 1-bit PSM)**.** *If every Boolean function on $\mathcal{X} \times \mathcal{Y}$ has a CDS protocol with communication complexity $t$, then every Boolean function on $(\{0,1\} \times \mathcal{X}) \times \mathcal{Y}$ has a 1-bit PSM protocol with communication complexity $(t + 1 + \log|\mathcal{X}| + \log|\mathcal{Y}|)$, with the same correctness and privacy guarantees.*

*Proof.* Suppose we want to construct a 1-bit PSM protocol for a function $f : (\{0,1\} \times \mathcal{X}) \times \mathcal{Y} \to \{0,1\}$. Let $(G_1, G_2, \mathsf{Dec}_{\mathsf{CDS}})$ be a CDS protocol for the function $g(x,y) = f((0,x),y) \oplus f((1,x),y)$ with communication complexity $t$.

We use this to construct our 1-bit PSM protocol $(F_1, F_2, \mathsf{Dec})$ for $f$. Let $s$ be a bit from the common randomness. $F_1$ is now defined as $F_1((w,x),(s,r)) = (G_1(x,s,r), w \oplus s, x)$, and $F_2$ is defined as $F_2(y,(s,r)) = (G_2(y,s,r), y)$.

Dec, on input $((g_1, w \oplus s, x), (g_2, y))$, works by first checking whether given $x$ and $y$, the value of $f$ still depends on $w$. If not, it simply computes $f$ using $x$ and $y$. If it does depend on $w$, this implies that $f((0,x),y) \neq f((1,x),y)$, and $g(x,y) = 1$, and so $\mathsf{Dec}_{\mathsf{CDS}}(x,y,g_1,g_2)$ outputs $s$, which can be used to retrieve $w$ from $(w \oplus s)$, and now the whole input is known and $f$ can be computed. This argues correctness, and the error here is at most that in the CDS protocol. The communication is also seen to be at most $(t + \log|\mathcal{X}| + \log|\mathcal{Y}| + 1)$.

Let $\mathsf{Sim}_{\mathsf{CDS}}$ be the simulator for the CDS protocol for $g$. The simulator $\mathsf{Sim}(f((w,x),y),x,y)$ works by first checking whether $f((0,x),y) = f((1,x),y)$. If it isn't, then the value of $w$ is determined by $x$, $y$, and the value of $f$ and, knowing $w$, Sim can compute $F_1$ and $F_2$ by itself, thus

simulating them perfectly. If not, this implies that $g(x,y) = 0$. In this case, $\mathsf{Sim}$ first computes $(g_1^*, g_2^*) \leftarrow \mathsf{Sim}_{\mathsf{CDS}}(x,y)$, picks a random bit $s^*$, and outputs $((g_1^*, s^*, x), (g_2^*, y))$. The simulation error is:

$$\begin{aligned}\Delta(\mathsf{Sim}(f((w,x),y),x,y); F(x,y,c)) \\ &= \Delta((\mathsf{Sim}_{\mathsf{CDS}}(x,y), s^*, x, y); (G(x,y,s), w \oplus s, x, y)) \\ &= \Delta((\mathsf{Sim}_{\mathsf{CDS}}(x,y), s^*); (G(x,y,s), w \oplus s))\end{aligned}$$

Note that here $s^*$ and $(w \oplus s)$ have the same marginal distribution, which is the uniform distribution over $\{0,1\}$. Also, $\mathsf{Sim}_{\mathsf{CDS}}(x,y)$ is independent of $s^*$. Writing out the expansion of $\Delta$ in terms of differences in probabilities and using Bayes' Theorem along with the above observation gives us the following:

$$\begin{aligned}&\Delta((\mathsf{Sim}_{\mathsf{CDS}}(x,y), s^*); (G(x,y,s), w \oplus s)) \\ &= \frac{1}{2} \sum_{m \in \mathcal{T}_1 \times \mathcal{T}_2, b \in \{0,1\}} \Big| \Pr[(\mathsf{Sim}_{\mathsf{CDS}}(x,y), s^*) = (m,b)] \\ &\qquad\qquad - \Pr[(G(x,y,s), w \oplus s) = (m,b)] \Big| \\ &= \frac{1}{2} \sum_{m \in \mathcal{T}_1 \times \mathcal{T}_2, b \in \{0,1\}} \Big| \Pr[s^* = b] \Pr[\mathsf{Sim}_{\mathsf{CDS}}(x,y) = m] \\ &\qquad\qquad - \Pr[w \oplus s = b] \Pr[G(x,y,b \oplus w) = m] \Big| \\ &= \frac{1}{2} \sum_{b \in \{0,1\}} \frac{1}{2} \sum_{m \in \mathcal{T}_1 \times \mathcal{T}_2} \Big| \Pr[\mathsf{Sim}_{\mathsf{CDS}}(x,y) = m] - \Pr[G(x,y,b \oplus w) = m] \Big| \\ &= \frac{1}{2} \left[ \Delta(\mathsf{Sim}_{\mathsf{CDS}}(x,y); G(x,y,0)) + \Delta(\mathsf{Sim}_{\mathsf{CDS}}(x,y); G(x,y,1)) \right]\end{aligned}$$

By the $\varepsilon$-privacy of the $\mathsf{CDS}$ scheme (since the value of $g(x,y)$ is 0), each summand in the right-hand side above is at most $\varepsilon$. Hence the total simulation error is at most $\varepsilon$. □

In [FKN94] it was shown that there exists a Boolean function on $\{0,1\}^n \times \{0,1\}^n$ such that any perfect 1-bit $\mathsf{PSM}$ for it requires at least $2.99n$ bits of communication. Using Lemma 7.2 along with this lower bound, we have the following theorem.

**Theorem 7.3.** *There is a Boolean function on $\{0,1\}^n \times \{0,1\}^n$ such that any perfect $\mathsf{CDS}$ protocol for it has communication complexity at least $0.99n$.*

We can generalise the above approach to construct $b$-bit $\mathsf{PSM}$ protocols for larger values of $b$ as follows.

**Lemma 7.4** ($\mathsf{CDS}$ to $b$-bit $\mathsf{PSM}$)**.** *If every Boolean function on $\mathcal{X} \times \mathcal{Y}$ has a $\mathsf{CDS}$ protocol with communication complexity $t$ then, for any $b > 0$, every Boolean function on $(\{0,1\}^b \times \mathcal{X}) \times \mathcal{Y}$ has a $b$-bit $\mathsf{PSM}$ protocol with communication complexity $(2^{2^b}(t+1) + \log|\mathcal{X}| + \log|\mathcal{Y}|)$, with the same correctness guarantee and with privacy that is degraded by a factor of $2^{2^b}$.*

*Proof sketch.* The idea behind the construction is that the function $f_{x,y}(w) = f((w,x),y)$, where $w$ is $b$ bits long, can be one of only $2^{2^b}$ functions – call this set of functions $\mathcal{H} = \{h_i\}$. For each of these $h_i$'s, we define a function $g_i(x,y)$ that indicates whether $f_{x,y} \equiv h_i$. Note that once the $\mathsf{PSM}$ decoder knows $x$ and $y$, the information that the value of $f$ reveals to it about $w$ is exactly $f_{x,y}(w)$, which is the same as $h_i(w)$ if $g_i(x,y) = 1$.

In our construction, first we have $F_1$ reveal $x$ and $F_2$ reveal $y$. Now we wish to, for each $i$, reveal $h_i(w)$ if and only if $g_i(x,y) = 1$. To do this, for each $i$, we choose a random bit $s_i$ from the common randomness, reveal $h_i(w) \oplus s_i$, and run the $\mathsf{CDS}$ protocol for $g_i$ with $s_i$ as the secret.

The correctness is preserved because whenever the $\mathsf{CDS}$ for the "correct" value of $i$ is correct, our protocol is correct.

The simulator, given $x$, $y$ and $f((w,x),y)$, first outputs $x$ and $y$. It then finds the $i'$ such that $g_{i'}(x,y) = 1$. For every other $i$, it publishes a random $s_i^*$ and the output of the $\mathsf{CDS}$ simulator for the function $g_i$ with inputs $x, y$ and secret 0. For $i'$, it publishes $(f((w,x),y) \oplus s_{i'}^*)$ for a random $s_{i'}^*$ and the messages of the $\mathsf{CDS}$ protocol for $g_{i'}$ with inputs $x, y$ and secret $s_{i'}^*$. Privacy error goes from $\varepsilon$ to $2^{2^b}\varepsilon$ because of arguments similar to those in the proof of Lemma 7.2 being applied to each invocation of the $\mathsf{CDS}$ protocol, all of which are mutually independent. □

# 8 Separating CDS and Insecure Communication

Here we show an explicit function whose randomised communication complexity is much higher than its $\mathsf{CDS}$ communication complexity. For simplicity, assume that $n$ below is a power of 2; the statements made here can be shown to be true for a general $n$ along the same lines.

**Definition 8.1** (Communication Complexity)**.** *Consider a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. A protocol between two parties (with shared randomness) who are given inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively, is said to compute $f$ if for every $(x,y) \in \mathcal{X} \times \mathcal{Y}$, the parties arrive at the correct value of $f$ at the end of it with probability at least $2/3$.*

*The communication cost of a protocol is the most number of bits exchanged by the parties over all possible inputs and all values of shared randomness. The* Randomised Communication Complexity *of $f$, denoted $\mathsf{R}(f)$, is the least communication cost of any protocol computing $f$.*

Gay et al. [GKW15] showed that if a function has a $\mathsf{CDS}$ protocol with communication complexity $t$ then, roughly, $\log \mathsf{R}(f) \leq 2t$ (in fact, they show this bound for one-way communication complexity). We show that this bound is optimal (up to constant factors) by exhibiting a function that has a $\mathsf{CDS}$ protocol with low communication, but has high randomised communication complexity (in fact, also high quantum communication complexity – see [She11] for relevant definitions and explanations). Towards this, we first introduce the following problem.

**Definition 8.2** (The Collision Problem)**.** *The* Collision Problem ($\mathsf{Col}_n$) *is a promise problem defined over a subset of $\{0,1\}^{n \log n}$ as follows. For an input $x \in \{0,1\}^{n \log n}$, divide $x$ into $n$ blocks of $\log n$ bits each. Each such $x$ can now be used to define a function $f_x : \{0,1\}^{\log n} \to \{0,1\}^{\log n}$, where $f_x(i)$ is the $i^{th}$ block of $x$ (when $i$ is interpreted as an integer in $[n]$). $\mathsf{Col}_n(x)$ is defined to be 1 if $f_x$ is a permutation, 0 if $f_x$ is 2-to-1, and is undefined otherwise.*

We use the above problem in conjunction with Sherstov's Pattern Matrix method [She11] for proving communication complexity lower bounds. We define the following function that corresponds to what would be called a "pattern matrix" of $\mathsf{Col}_n$.

**Definition 8.3.** *The promise problem* $\mathsf{PCol}_n : \{0,1\}^{4n\log n} \times [4]^{n\log n} \to \{0,1\}$ *is defined as follows. On an input* $(x, y)$*, first divide* $x$ *into* $n \log n$ *blocks of size* 4 *bits each. From the* $i^{th}$ *block, select the bit* $x_{i,y_i}$ *that is specified by the* $i^{th}$ *coordinate of* $y$ *(which is an element of* $\{1,2,3,4\}$*) to get the string* $x_y$ *of length* $n \log n$*. The output of the function is* $\mathsf{Col}_n(x_y)$*.*

The pattern matrix method gives us a way to lower bound the randomised communication complexity of a function constructed in this manner using lower bounds on the approximate degree (denoted by $\mathsf{deg}$ and which we do not define here) of the underlying function. We use known results to derive the following Corollary 8.4.

**Corollary 8.4.** $\mathsf{R}(\overline{\mathsf{PCol}}_n) = \mathsf{R}(\mathsf{PCol}_n) \geq \Omega(n^{1/3})$

*Proof.* It follows from [She11] that $\mathsf{R}(\mathsf{PCol}_n) \geq \Omega(\mathsf{deg}(\mathsf{Col}_n))$. Combined with the fact that $\mathsf{deg}(\mathsf{Col}_n) \geq \Omega(n^{1/3})$ (which follows from [Amb05, Kut05]), we derive the corollary. □

Next we show that $\overline{\mathsf{PCol}}_n$ has a very efficient $\mathsf{CDS}$ protocol.

**Lemma 8.5.** *There is a* $\mathsf{CDS}$ *protocol for* $\overline{\mathsf{PCol}}_n$ *with* $\frac{1}{3}$*-completeness, perfect privacy, and communication complexity* $O(\log n)$*.*

In order to prove this lemma, we will need the following simple lemma which shows how to simulate messages generated by applying a $\mathsf{PSM}$ protocol to a set of inputs that are distributed jointly. It says that these can be simulated by sampling the corresponding distribution over the function outputs and running the $\mathsf{PSM}$ simulator on these sampled outputs.

**Lemma 8.6.** *Consider any function* $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$*, and a* $\mathsf{PSM}$ *protocol* $(F_1, F_2)$ *for it with* $\epsilon$*-privacy realized by a simulator* $\mathsf{Sim}$*. For any integer* $k > 0$ *and any joint distribution* $\overline{(X,Y)}$ *over* $(\mathcal{X} \times \mathcal{Y})^k$*, let* $\overline{Z}$ *be the distribution over* $\mathcal{Z}^k$ *obtained by sampling* $\overline{(x,y)} = ((x_1,y_1),\ldots,(x_k,y_k))$ *from* $\overline{(X,Y)}$ *and then computing* $(f(x_1,y_1),\ldots,f(x_k,y_k))$*. Then,*

$$\Delta\left((\mathsf{Sim}(z_1),\ldots,\mathsf{Sim}(z_k)) ; (F(x_1,y_1),\ldots,F(x_k,y_k))\right) \leq k\epsilon,$$

*where* $\overline{(x,y)} \leftarrow \overline{(X,Y)}, \overline{z} \leftarrow \overline{Z}$*. In particular, if the* $\mathsf{PSM}$ *is perfect, the above statistical distance is zero.*

The proof (which is standard) appears in Appendix B. We can now prove Lemma 8.5.

*Proof of Lemma 8.5.* Given input $(x, y) \in \{0,1\}^{4n\log n} \times [4]^{n\log n}$ and secret bit $s$, the idea behind the $\mathsf{CDS}$ protocol is to convey through the messages a uniformly random element from the range of $f_{x_y}$ if $s = 1$, and a uniformly random element from $\{0,1\}^{\log n}$ if $s = 0$. If $\overline{\mathsf{PCol}}_n(x, y) = 0$, $f_{x_y}$ is a permutation, and hence the distributions in the two cases are identical. If $\overline{\mathsf{PCol}}_n(x, y) = 1$, $f_{x_y}$'s range covers only half the co-domain and so the two cases can be distinguished.

We now construct a $\mathsf{CDS}$ protocol $(F_1, F_2)$ that functions as above. Let $G = (G_1, G_2)$ be the perfect $\mathsf{PSM}$ protocol for the finite function $\mathsf{ind} : \{0,1\}^4 \times [4] \to \{0,1\}$ that takes $(a, b)$ as input and outputs the bit in $a$ that is pointed to by $b$. Let $\mathsf{Dec}_{\mathsf{PSM}}$ be a perfect decoder for $G$. The $\mathsf{CDS}$ protocol $(F_1, F_2)$ works as follows.

- First an index $i \in [n]$ is sampled from the common randomness. (In the case of $s = 1$, $f_{x_y}(i)$ is the information that will be output jointly by $F_1$ and $F_2$.)

- Note that the value $f_{x_y}(i)$ consists of $\log n$ bits, each of which is encoded by 4 bits in $x$ and a value in [4] in $y$ – let the relevant parts of $x$ and $y$ be $(x_i^1, \ldots, x_i^{\log n})$ and $(y_i^1, \ldots, y_i^{\log n})$ respectively, where $x_i^j \in \{0,1\}^4$ and $y_i^j \in [4]$.
- If $s = 1$, for each $j \in [\log n]$, $F_1$ outputs $g_1^j = G_1(x_i^j, r_j)$, and $F_2$ outputs $g_2^j = G_2(y_i^j, r_j)$, where $r_j$ is from the common randomness.
- If $s = 0$, for each $j \in [\log n]$, $F_1$ outputs $g_1^j = G_1(w^j, r_j)$, and $F_2$ outputs $g_2^j = G_2(y_i^j, r_j)$, where each $w^j$ is chosen at random from $\{0,1\}^4$.

The $\mathsf{CDS}$ decoding procedure $\mathsf{Dec}$ works as follows.

- Input: $(x, y, (g_1^1, \ldots, g_1^{\log n}), (g_2^1, \ldots, g_2^{\log n}))$.
- For each $j \in [\log n]$, compute $z_j \leftarrow \mathsf{Dec}_{\mathsf{PSM}}(g_1^j, g_2^j)$ to get the string $z$.
- If there exists an $i$ such that $f_{x_y}(i) = z$, output 1, else output 0.

If $\overline{\mathsf{PCol}}_n(x, y) = 1$, $f_{x_y}$ is 2-to-1. By the perfect correctness of the $\mathsf{PSM}$ protocol, if $s = 1$, $z = f_{x_y}(i)$ for the $i$ chosen by $(F_1, F_2)$, and so $\mathsf{Dec}$ always outputs 1. If $s = 0$, $z$ is a random string in $\{0,1\}^{\log n}$, and $\mathsf{Dec}$ outputs 0 exactly when $z$ falls outside the range of $f_{x_y}$; this happens with probability 1/2 as $f_{x_y}$ is 2-to-1, and $\mathsf{Dec}$ outputs 1 otherwise.

This gives only $\frac{1}{2}$-correctness but this error is only one-sided, and so by repeating the protocol once more and checking whether $z$ lies in the range of $f_{x_y}$ both times, this error can be reduced, giving $\frac{1}{4}$-correctness. This repetition does not degrade privacy which, as shown below, is perfect.

If $\overline{\mathsf{PCol}}_n(x, y) = 0$, $f_{x_y}$ is a permutation. The output of $F_1$ and $F_2$ is simulated as follows using $\mathsf{Sim}_{\mathsf{PSM}}$, the simulator for the perfect $\mathsf{PSM}$ protocol used above. Our simulator $\mathsf{Sim}$, given $(x, y)$ as input, first picks random bits $z_1^*, \ldots, z_{\log n}^*$. It then outputs $(\mathsf{Sim}_{\mathsf{PSM}}(z_1^*), \ldots, \mathsf{Sim}_{\mathsf{PSM}}(z_{\log n}^*))$.

The simulation error is:

$$\Delta((\mathsf{Sim}_{\mathsf{PSM}}(z_1^*), \ldots, \mathsf{Sim}_{\mathsf{PSM}}(z_{\log n}^*)); ((g_1^1, g_2^1), \ldots, (g_1^{\log n}, g_2^{\log n})))$$

where the $(g_1^j, g_2^j)$'s are the $\mathsf{PSM}$ messages in the protocol description.

First we consider the case $s = 1$. Recall that the $(g_1^j, g_2^j)$'s are computed by first selecting $i \in [n]$ at random and computing the $\mathsf{PSM}$ messages for $\mathsf{ind}(x_i^j, y_i^j)$, which is the $j^{\text{th}}$ bit of $f_{x_y}(i)$. As the range of $f_{x_y}$ is uniform over $\{0,1\}^{\log n}$, over the randomness of $i$ each $\mathsf{ind}(x_i^j, y_i^j)$ is a uniformly random bit independent of all the other $\mathsf{ind}(x_i^{j'}, y_i^{j'})$'s. Thus, $(\mathsf{ind}(x_i^1, y_i^1), \ldots, \mathsf{ind}(x_i^{\log n}, y_i^{\log n}))$ is distributed the same as $(z_1^*, \ldots, z_{\log n}^*)$, and so by Lemma 8.6, the above simulation error is zero as we are using a perfect $\mathsf{PSM}$ protocol.

Similarly, when $s = 0$, the $(g_1^j, g_2^j)$'s are computed by first selecting $i \in [n]$ at random and computing the $\mathsf{PSM}$ messages for $\mathsf{ind}(w^j, y_i^j)$ for uniformly random $w^1, \ldots, w^j \in \{0,1\}^4$. So again each $\mathsf{ind}(w^j, y_i^j)$ is a uniformly random bit independent of all the other $\mathsf{ind}(w^{j'}, y_i^{j'})$'s, and by Lemma 8.6, the simulation error is again zero.

The $\mathsf{PSM}$ for each bit of $z$ is for a finite-sized function and its communication complexity is some constant, so the total communication is $\Theta(\log n)$. □

Gay et al. [GKW15] showed the following relationships between the randomized communication complexity of a Boolean function and the complexity of general and linear $\mathsf{CDS}$ protocols for it with single-bit secrets. While they originally showed these for perfect protocols, we extend their proof to work for imperfect ones in Appendix A.

**Theorem 8.7** ([GKW15])**.** *For any (partial or total) Boolean function $f$,*

$$\mathsf{CDS}(f) \geq \frac{1}{2}\log \mathsf{R}(f) \qquad \textit{and} \qquad \mathsf{linCDS}(f) \geq \frac{1}{10}\sqrt{\mathsf{R}(f)}$$

The following corollary of Lemma 8.5 and Corollary 8.4 shows that the above bound on $\mathsf{CDS}$ in general is tight up to constant factors.

**Corollary 8.8.** *There exists a partial Boolean function $f$ such that:*

$$\mathsf{CDS}(f) \leq O(\log \mathsf{R}(f))$$

Following from Corollary 8.8 and Theorem 8.7, the next corollary says that there are functions for which general $\mathsf{CDS}$ protocols can do much better than linear $\mathsf{CDS}$ protocols.

**Corollary 8.9.** *There exists a partial Boolean function $f$ such that:*

$$\mathsf{CDS}(f) \leq O(\log \mathsf{linCDS}(f))$$

**Remark 8.10.** *In fact, [GKW15] showed that Theorem 8.7 holds even for "weakly-linear" $\mathsf{CDS}$ protocols in which only the decoding process is assumed to be linear (and the senders are allowed to be non-linear). Corollary 8.9 therefore generalizes to this case as well.*

# References


[Aar12] S. Aaronson. Impossibility of succinct quantum proofs for collision-freeness. *Quantum Information & Computation*, 12(1-2):21–28, 2012.

[AIK04] B. Applebaum, Y. Ishai, and E. Kushilevitz. Cryptography in $\mathrm{NC}^0$. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 166–175. IEEE Computer Society, 2004.

[AIR01] W. Aiello, Y. Ishai, and O. Reingold. Priced oblivious transfer: How to sell digital goods. In B. Pfitzmann, editor, *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*, volume 2045 of *Lecture Notes in Computer Science*, pages 119–135. Springer, 2001.

[Amb05] A. Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[AR16] B. Applebaum and P. 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.

[Att14] N. Attrapadung. Dual system encryption via doubly selective security: Framework, fully secure functional encryption for regular languages, and more. In P. Q. Nguyen and E. Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 557–577. Springer, 2014.

[BD91] E. F. Brickell and D. M. Davenport. On the classification of ideal secret sharing schemes. *J. Cryptology*, 4(2):123–134, 1991.

[BFM16] A. Beimel, O. Farràs, and Y. Mintz. Secret-sharing schemes for very dense graphs. *J. Cryptology*, 29(2):336–362, 2016.

[BFP16] A. Beimel, O. Farràs, and N. Peter. Secret sharing schemes for dense forbidden graphs. In V. Zikas and R. D. Prisco, editors, *Security and Cryptography for Networks - 10th International Conference, SCN 2016, Amalfi, Italy, August 31 - September 2, 2016, Proceedings*, volume 9841 of *Lecture Notes in Computer Science*, pages 509–528. Springer, 2016.

[BIKK14] A. Beimel, Y. Ishai, R. Kumaresan, and E. Kushilevitz. On the cryptographic complexity of the worst functions. In Y. Lindell, editor, *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings*, volume 8349 of *Lecture Notes in Computer Science*, pages 317–342. Springer, 2014.

[CC06] H. Chen and R. Cramer. Algebraic geometric secret sharing schemes and secure multi-party computations over small fields. In C. Dwork, editor, *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings*, volume 4117 of *Lecture Notes in Computer Science*, pages 521–536. Springer, 2006.

[CCG+07] H. Chen, R. Cramer, S. Goldwasser, R. de Haan, and V. Vaikuntanathan. Secure computation from random error correcting codes. In M. Naor, editor, *Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings*, volume 4515 of *Lecture Notes in Computer Science*, pages 291–310. Springer, 2007.

[CKGS98] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *J. ACM*, 45(6):965–981, 1998.

[CSGV93] R. M. Capocelli, A. D. Santis, L. Gargano, and U. Vaccaro. On the size of shares for secret sharing schemes. *J. Cryptology*, 6(3):157–167, 1993.

[FKN94] U. Feige, J. Kilian, and M. Naor. A minimal model for secure computation (extended abstract). In F. T. Leighton and M. T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 554–563. ACM, 1994.

[GIKM00] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin. Protecting data privacy in private information retrieval schemes. *J. Comput. Syst. Sci.*, 60(3):592–629, 2000.

[GKW15] R. Gay, I. Kerenidis, and H. Wee. Communication complexity of conditional disclosure of secrets and attribute-based encryption. In R. Gennaro and M. 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.

[GMR88] S. Goldwasser, S. Micali, and R. L. Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM J. Comput.*, 17(2):281–308, 1988.

[GPSW06] V. Goyal, O. Pandey, A. Sahai, and B. Waters. Attribute-based encryption for fine-grained access control of encrypted data. In A. Juels, R. N. Wright, and S. D. C. 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.

[GPW15] M. Göös, T. Pitassi, and T. Watson. Zero-information protocols and unambiguity in arthur-merlin communication. In T. Roughgarden, editor, *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 113–122. ACM, 2015.

[GS96] A. Garcia and H. Stichtenoth. On the asymptotic behavior of some towers of function fields over finite fields. *Journal of Number Theory*, 61(2):248–273, 1996.

[HR05] T. Holenstein and R. Renner. One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption. In V. Shoup, editor, *Advances in Cryptology - CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, volume 3621 of *Lecture Notes in Computer Science*, pages 478–493. Springer, 2005.

[IK97] Y. Ishai and E. Kushilevitz. Private simultaneous messages protocols with applications. In *ISTCS*, pages 174–184, 1997.

[IK00] Y. Ishai and E. Kushilevitz. Randomizing polynomials: A new representation with applications to round-efficient secure computation. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 294–304. IEEE Computer Society, 2000.

[IKOS09] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai. Extracting correlations. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 261–270. IEEE Computer Society, 2009.

[IW14] Y. Ishai and H. Wee. Partial garbling schemes and their applications. In J. Esparza, P. Fraigniaud, T. Husfeldt, and E. 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.

[Kut05] S. Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005.

[Min12] Y. Mintz. Information ratios of graph secret-sharing schemes. Master's thesis, Dept. of Computer Science, Ben Gurion University, 2012.

[MPR07] U. M. Maurer, K. Pietrzak, and R. Renner. Indistinguishability amplification. In A. Menezes, editor, *Advances in Cryptology - CRYPTO 2007, 27th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2007, Proceedings*, volume 4622 of *Lecture Notes in Computer Science*, pages 130–149. Springer, 2007.

[Oka96] T. Okamoto. On relationships between statistical zero-knowledge proofs. In G. L. Miller, editor, *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996*, pages 649–658. ACM, 1996.

[Pot16] A. Potechin. A note on amortized space complexity. *CoRR*, abs/1611.06632, 2016.

[She11] A. A. Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[SS97] H. Sun and S. 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.

[SV03] A. Sahai and S. P. Vadhan. A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003.

[SW05] A. Sahai and B. Waters. Fuzzy identity-based encryption. In R. 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.

[Wee14] H. Wee. Dual system encryption via predicate encodings. In Y. Lindell, editor, *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings*, volume 8349 of *Lecture Notes in Computer Science*, pages 616–637. Springer, 2014.

# A Communication Complexity and Imperfect CDS Protocols

In this section, we extend the relationships between CDS and randomised communication complexity shown by Gay et al. [GKW15] to include imperfect CDS protocols. We prove the following theorem. (The terms involved are defined in Section 3 and Section 8.)

**Theorem A.1.** *For any (partial or total) Boolean function $f$,*

$$\mathsf{CDS}(f) \geq \frac{1}{2}\log \mathsf{R}(f)$$
$$\mathsf{linCDS}(f) \geq \frac{1}{10}\sqrt{\mathsf{R}(f)}$$

Recall that $\mathsf{CDS}(f)$ is the least communication complexity of any CDS protocol for $f$ with $\{0,1\}$ as the secret domain that has $\frac{1}{10}$ correctness and privacy. And that $\mathsf{linCDS}(f)$ is the same, but for linear protocols. We will prove Theorem A.1 using the following more general lemma that we prove afterward.

**Lemma A.2.** *Consider any function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Suppose $f$ has a CDS protocol $(F_1, F_2, \mathsf{Dec})$ with $\frac{1}{10}$-correctness and $\frac{1}{10}$-privacy, with domains as follows: $F_1 : \mathcal{X} \times \{0,1\} \times \mathcal{R} \to \mathcal{T}_1$, $F_2 : \mathcal{X} \times \{0,1\} \times \mathcal{R} \to \mathcal{T}_2$, and $\mathsf{Dec} : \mathcal{X} \times \mathcal{Y} \times \mathcal{T}_1 \times \mathcal{T}_2 \to \{0,1\}$. Let $\mathcal{H}$ be any superset of all possible functions $\{h : \mathcal{T}_1 \times \mathcal{T}_2 \to \{0,1\}\}$ that $\mathsf{Dec}(x, y, \cdot, \cdot)$ could possibly be for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Then,*

$$\mathsf{R}(f) \leq 100 \log|\mathcal{H}| \left(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|\right)$$

*Proof of Theorem A.1.* A lower bound for a CDS protocol for $f$ with $\frac{1}{10}$ correctness and privacy can be obtained by taking $\mathcal{H}$ to be the set of all possible functions from $\mathcal{T}_1 \times \mathcal{T}_2 \to \{0,1\}$. There are $2^{|\mathcal{T}_1||\mathcal{T}_2|}$ of these. We then have from Lemma A.2:

$$\begin{aligned} & \mathsf{R}(f) \leq 100\,|\mathcal{T}_1|\,|\mathcal{T}_2|\,(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|) \\ \implies & \log \mathsf{R}(f) \leq \log 100 + (\log|\mathcal{T}_1| + \log|\mathcal{T}_2|) + \log(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|) \\ & \leq 2(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|) \end{aligned}$$

This is true for any such CDS protocol. Note that $(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|)$ is the communication complexity of the CDS protocol in question. So this implies that $\log \mathsf{R}(f) \leq 2\mathsf{CDS}(f)$.

The lower bound on $\mathsf{linCDS}(f)$ is similarly obtained by taking $\mathcal{H}$ to be the set of all linear functions over vectors spaces that may be contained in $\mathcal{T}_1 \times \mathcal{T}_2$, as linear CDS protocols always have linear reconstruction. In this case, $\mathcal{T}_1$ and $\mathcal{T}_2$ would have to be of the form $\mathbb{F}^{t_1}$ and $\mathbb{F}^{t_2}$ for some $t_1, t_2$, and $\mathcal{H}$ would then contain $\mathbb{F}^{t_1+t_2} = |\mathcal{T}_1|\,|\mathcal{T}_2|$ functions. Lemma A.2 now immediately gives us the following:

$$\begin{aligned} & \mathsf{R}(f) \leq 100(\log|\mathcal{T}_1| + \log|\mathcal{T}_2|)^2 \\ \implies & \sqrt{\mathsf{R}(f)} \leq 10 \cdot \mathsf{linCDS}(f) \end{aligned}$$

□

*Proof of Lemma A.2.* Given a CDS protocol $(F_1, F_2, \mathsf{Dec})$ as in the hypothesis, we construct a single message protocol (with shared randomness) for parties $A$, who is given an $x \in \mathcal{X}$, and $B$, who is given a $y \in \mathcal{Y}$, to compute $f(x, y)$ as follows.

- For an integer $N$ that shall be determined later, the shared randomness is used to sample $N$ random bits $s_1, \ldots, s_N$, and also $r_1, \ldots, r_N \in \mathcal{R}$.
- For each $i \in [N]$, $A$ computes and sends $a_i = F_1(x, s_i, r_i)$ to $B$.
- For each $i \in [N]$, $B$ computes, in order, $b_i = F_2(y, s_i, r_i)$ and, for each $h \in H$, $s_i^h = h(a_i, b_i)$.
- If there is an $h \in H$ such that for more than $3/4$ values of $i \in [N]$, $s_i^h = s_i$, then $B$ decides that $f(x, y) = 1$, else 0.

If $f(x, y) = 1$, by the $\frac{1}{10}$-correctness of the $\mathsf{CDS}$ protocol, we know that there exists an $h^* \in H$, namely $\mathsf{Dec}(x, y, \cdot, \cdot)$, such that $\Pr[h^*(a_i, b_i) = s_i] \geq 9/10$. By the Chernoff bound, the probability that the communication protocol is wrong in this case can be bounded as follows:

$$\Pr\left[\left|\left\{i : s_i^{h^*} = s_i\right\}\right| \leq \frac{3}{4}N\right] \leq e^{-N/80}$$

If $f(x, y) = 0$, by the $\frac{1}{10}$-privacy of the $\mathsf{CDS}$ protocol and the triangle inequality, the statistical distance between the distributions $F(x, y, 0)$ and $F(x, y, 1)$ is at most $2/10$. This implies that for any function $h$, if $s_i$ is chosen at random, $\Pr[h(a_i, b_i) = s_i] \leq 6/10$. Using the union bound and the Chernoff bound, in order, the probability that the communication protocol is wrong in this case can be bounded as follows:

$$\begin{aligned}\Pr\left[\exists h \in \mathcal{H} : \left|\left\{i : s_i^h = s_i\right\}\right| \geq \frac{3}{4}N\right] &\leq |\mathcal{H}| \Pr\left[\left|\left\{i : s_i^h = s_i\right\}\right| \geq \frac{3}{4}N\right] \\ &\leq |\mathcal{H}|\, e^{-N/80}\end{aligned}$$

So if $N$ is chosen to be, say, $(100 \log |\mathcal{H}|)$, the error probability in both cases would be much less than $1/3$, and this would be a valid communication protocol computing $f$.

The total communcation involved is $N \log |\mathcal{T}_1| \leq 100 \log |\mathcal{H}| \, (\log |\mathcal{T}_1| + \log |\mathcal{T}_2|)$, as required. □

# B Proof of Lemma 8.6

We need the following standard facts about statistical distance.

**Fact B.1.** *For any random variable $A$ over $\mathcal{A}$ and variables $B, B'$ over the same domain that are jointly distributed with $A$,*

$$\Delta(B; B') \leq \sum_{a \in \mathcal{A}} \Pr[A = a] \cdot \Delta(B_{|A=a}; B'_{|A=a})$$

*where $B_{|A=a}$ is the variable $B$ conditioned on $A = a$.*

**Fact B.2.** *For any two tuples of mutually independent random variables $(A_1, \ldots, A_k)$ and $(B_1, \ldots, B_k)$ over the same domain,*

$$\Delta((A_1, \ldots, A_k); (B_1, \ldots, B_k)) \leq \sum_{i \in [k]} \Delta(A_i; B_i)$$

We can now prove Lemma 8.6.

*Proof of Lemma 8.6.* Expanding out what the $z_i$'s are defined to be in the lemma statement and then using Fact B.1, with $\overline{(X,Y)}$ as the distribution $A$ there, we conclude that the statistical distance

$$\Delta\left((\mathsf{Sim}(z_i))_{i\in[k]}\,;(F(x_i,y_i))_{i\in[k]}\right) \quad \text{where } \overline{(x,y)} \leftarrow \overline{(X,Y)}, \overline{z} \leftarrow \overline{Z},$$

equals the statistical distance

$$\Delta\left(\left(\mathsf{Sim}(f(x'_i,y'_i))\right)_{i\in[k]}\,;(F(x_i,y_i))_{i\in[k]}\right) \quad \text{where } \overline{(x,y)}, \overline{(x',y')} \leftarrow \overline{(X,Y)},$$

which is upper-bounded by

$$\sum_{\overline{(x,y)}\in(\mathcal{X}\times\mathcal{Y})^k} \Pr\left[\overline{(X,Y)} = \overline{(x,y)}\right] \cdot \Delta\left((\mathsf{Sim}(f(x_i,y_i)))_{i\in[k]}\,;(F(x_i,y_i))_{i\in[k]}\right).$$

Note that once the $x_i$'s and $y_i$'s are fixed, the $F(x_i,y_i)$'s are all independently distributed, and so are the $\mathsf{Sim}(f(x_i,y_i))$'s. Hence, using Fact B.2, we have for any $\overline{(x,y)} \in (\mathcal{X}\times\mathcal{Y})^k$:

$$\begin{aligned}\Delta\left((\mathsf{Sim}(f(x_i,y_i)))_{i\in[k]}\,;(F(x_i,y_i))_{i\in[k]}\right) &\le \sum_{i\in[k]} \Delta(\mathsf{Sim}(f(x_i,y_i)); F(x_i,y_i)) \\ &\le k\epsilon\end{aligned}$$

where the second inequality is from the $\epsilon$-privacy of the $\mathsf{PSM}$ protocol. This proves the lemma. □