Electronic Colloquium on Computational Complexity, Report No. 156 (2012)

Limits of provable security for homomorphic encryption

Andrej Bogdanov[*] Chin Ho Lee[†]

Abstract

We show that public-key bit encryption schemes which support weak homomorphic evaluation of parity or majority cannot be proved message indistinguishable beyond $\mathrm{AM} \cap \mathrm{coAM}$ via general (adaptive) reductions, and beyond statistical zero-knowledge via reductions of constant query complexity.

Previous works on the limitation of reductions for proving security of encryption schemes make restrictive assumptions about the encryption algorithm (Brassard, Goldreich and Goldwasser, Akavia et al.) or about the reduction (Feigenbaum and Fortnow, Bogdanov and Trevisan, Akavia et al.) Our first result makes no assumptions of either sort.

Towards these results, we show that any homomorphic evaluator for parity or majority over sufficiently many inputs can be converted into a rerandomization algorithm. This is a procedure that converts a ciphertext into another ciphertext which is statistically close to being independent and identically distributed with the original one.

1 Introduction

In this work we revisit the question of basing cryptography on NP-hardness. If P equals NP then computationally secure encryption is impossible. Is the converse true?

Despite considerable efforts, there is no candidate encryption scheme whose security can be plausibly reduced to the worst-case hardness of some NP-complete problem. Neither is there conclusive evidence that rules out constructions of secure encryption schemes from NP-complete problems, although several obstacles have been pointed out over the years.

Restricting the encryption Brassard [Bra79] shows that no public-key encryption scheme can be proved secure beyond $\mathrm{NP} \cap \mathrm{coNP}$, but under the implicit assumption that every public key-ciphertext pair (queried by the reduction) can be decrypted uniquely. Goldreich and Goldwasser [GG98] argue that this assumption is unrealistic by giving examples of encryption schemes that do not satisfy it. They show that the conclusion holds under the relaxed assumption that invalid queries to the decryption oracle can be efficiently certified as such. (If the reduction is randomized, the limitation weakens to $\mathrm{AM} \cap \mathrm{coAM}$.)

Goldreich and Goldwasser warn that these assumptions are unrealistic as they do not apply to many known proofs of security. Bogdanov and Trevisan [BT06] point out the following example of Even

[*] `andrejb@cse.cuhk.edu.hk`. Dept. of Computer Science and Engineering and Institute for Theoretical Computer Science and Communications, Chinese University of Hong Kong. Work supported by grants RGC GRF CUHK410309 and CUHK410111.

[†] `chlee@cse.cuhk.edu.hk`. Dept. of Computer Science and Engineering, Chinese University of Hong Kong.

ISSN 1433-8092

and Yacobi [EY80]. They construct a public key encryption scheme and show how to solve an NP-hard problem using a distinguishing oracle. Their notion of security is unrealistic, as they require a perfect distinguishing oracle. However, their example illustrates that the restrictions imposed by Brassard and Goldreich and Goldwasser do not capture the difficulty of basing cryptography on NP hardness.

Akavia, Goldreich, Goldwasser, and Moshkovitz [AGGM06] rule out reductions from NP-complete problems to inverting one-way functions (the basis of private-key encryption) assuming that sizes of preimage sets are worst-case approximable in AM. The same considerations apply to their argument. There are natural examples of conjectured one-way functions (for example, Goldreich's function [Gol00]) not known to satisfy the aforementioned assumptions.

Restricting the reduction Another line of works makes restrictive assumptions about the type of reduction used to prove NP-hardness. Feigenbaum and Fortnow [FF93] show that a decision problem cannot be proven NP-hard on average (unless the polynomial hierarchy collapses) by a reduction that is non-adaptive and each of its queries is uniformly distributed. Bogdanov and Trevisan [BT06] obtain the same conclusion without restricting the distribution of queries, but still under non-adaptive reductions. More precisely, they show that if there is a non-adaptive reduction from a decision problem L to a problem in distributional NP, then L must be in AM/poly $\cap$ coAM/poly. In particular their result applies to the problem of inverting a one-way function. For this important case, Akavia et al. improve the limitation to AM $\cap$ coAM, also assuming the reduction is non-adaptive.

Lattice-based cryptography provides the only known examples of encryption schemes whose insecurity would imply worst-case solutions to conjectured hard problems, like finding short vectors in lattices. The reduction of Regev [Reg09], which gives the most efficient cryptosystems of this kind with a proof of security (against quantum algorithms), is adaptive. For certain settings of parameters, these cryptosystems support homomorphic evaluation of a bounded class of functionalities (and general functionalities under additional security assumptions).

Our results

We say a public-key encryption scheme supports weak homomorphic evaluation of $f\colon \{0,1\}^n \to \{0,1\}$ if there is an efficient algorithm that for every $x_1 \dots x_n \in \{0,1\}^n$ takes as inputs the public key and encryptions of the bits $x_1, \dots, x_n$ and outputs a ciphertext that decrypts to $f(x_1 \dots x_n)$. See Section 2 for a formal definition.

Our main theorem (Theorem 1) shows that any public key encryption scheme that supports weak homomorphic evaluation of parity (or majority) cannot be proved message indistinguishable beyond AM $\cap$ coAM, even under adaptive reductions. To the best of our knowledge this applies to all known schemes with homomorphic properties, including El Gamal encryption [Gam85], Paillier encryption [Pai99], as well as the more recent somewhat and fully homomorphic encryption schemes of Gentry [Gen09], Van Dijk et al. [vDGHV10], and Brakerski and Vaikuntanathan [BV11] (which build upon the lattice-based cryptosystems of Regev [Reg09] and Peikert [Pei09]).

In Theorem 2 we show that if the reduction has constant query complexity, then message indistinguishability cannot be proved beyond statistical zero knowledge (SZK), which is a subclass of AM $\cap$ coAM. We are not aware of any previous general results that show statistical zero knowledge as a limitation for proving security of encryptions.

The reductions we consider are randomized and meet the following definition: Given an input, the reduction makes arbitrary (adaptive) queries to a distinguishing oracle for bit encryptions. We require that for any (not necessarily efficient) distinguishing oracle, which may depend on the input to the reduction, the reduction outputs the correct answer. We do not know of any cryptographic reductions that treat the adversary as a black box which fall outside our definition.

Lemma 5 which is used in the proofs of Theorems 1 and 2 gives a way to obtain rerandomization of ciphertexts from any homomorphic evaluator for parity or majority. While rerandomization has been used in constructions of homomorphic evaluators [Gen09, vDGHV10], it is not a priori clear that it is necessary for homomorphic evaluation. Homomorphic evaluation may be implemented deterministically while rerandomization requires randomness.

Examples Consider El Gamal encryption over a sufficiently large cyclic group G. We can view it as a bit encryption scheme like this. The public key is a pair of group elements (g, h) and the secret key is an integer s such that $g^s = h$. Let $v \neq 1$ be any group element. The encryption of a bit m has the form $\mathbf{Enc}_{g,h}(m) = (g^r, h^r v^m)$ where $r \sim [|G|]$ is random. The (homomorphic) decryption algorithm on input (a, b) finds the smallest t such that $b = a^s v^t$ and outputs t modulo 2. This scheme supports efficient homomorphic evaluation of parity by multiplying the corresponding ciphertexts, i.e. $\mathbf{Dec}_s(\mathbf{Enc}_{g,h}(m_1) \cdots \mathbf{Enc}_{g,h}(m_k)) = m_1 + \cdots + m_k \mod 2$.

Applying our main theorems, we conclude that El Gamal encryption cannot be proved secure beyond AM ∩ coAM using general reductions and beyond SZK using reductions of constant query complexity. In this example, one can give direct proofs of these statements by using specific properties of the discrete logarithm [GK93].

Other examples include Paillier encryption, which explicitly supports homomorphic evaluation of sums and therefore parities, as well as all the homomorphic encryption schemes of Gentry, van Dijk et al., and Brakerski and Vaikuntanathan, which support homomorphic evaluation of any functionality on a constant number of inputs (for sufficiently large values of the security parameter).

Our proof

The *symmetric increasing function* $sif_n(x)$ on n inputs, where n is odd, is a partial function over $\{0, 1\}^n$ that takes value 0 when x has Hamming weight $(n-1)/2$ and 1 when x has weight $(n+1)/2$. Its extensions over the boolean cube include parity and majority.

From homomorphic evaluation to rerandomization (Section 4) For simplicity let's assume that we have a *strong* homomorphic evaluator H for sif_n. This is an algorithm that takes as inputs independent encryptions of $x_1, \dots, x_n$ and outputs a ciphertext which is statistically close to an encryption of $sif_n(x_1, \dots, x_n)$. In Lemma 5 we show that H can be used to obtain an approximate *rerandomization* **Rer**: This is a procedure that takes an encryption as its input and produces an independent and identically distributed encryption as its output. Our rerandomization will be approximate in the sense that the input and output of **Rer** will be only statistically close to independent.

Our construction works as follows: Given a ciphertext C, generate $(n-1)/2$ independent encryptions of 0, $(n-1)/2$ independent encryptions of 1, randomly shuffle them together with C and feed the n resulting ciphertexts to the homomorphic evaluator for sif_n. By the strong homomorphic property, the output of the homomorphic evaluator will be identically distributed with C. But why should they be independent? From the point of view of the homomorphic evaluator, if C is an encryption of b, then it is indistinguishable from the other $(n-1)/2$ encryptions of b. Since the output of the homomorphic evaluator is bounded in length, the evaluator must "forget" most of the information about most of the ciphertexts it is given as inputs, including C as it is indistinguishable from the others. Therefore the output is forced to look almost statistically independent of C.

Lemma 5 also applies to weak homomorphic evaluators, in which case it achieves a weaker notion of rerandomization. While this weak rerandomization is sufficient to carry out the rest of the proof, to simplify this discussion we will assume the availability of strong rerandomization as described above.

From rerandomization to a distinguishing protocol (Section 5) To turn a reduction from distinguishing encryptions to L into a proof system for $\overline{L}$, we proceed as in previous works: The verifier plays the role of the reduction and the prover plays the role of the distinguishing oracle. The challenge is to force the prover to give answers that are consistent with a specific, fixed distinguishing oracle.

To illustrate the difficulties in the context of public key encryption, let us point out the deficiencies of some naive proof systems. Suppose the verifier submits a public key-ciphertext query (PK, C) to the prover, who is supposed to act as a distinguishing oracle. A natural attempt is to ask the prover to provide the message m and randomness R such that C is an encryption of m under public key PK with randomness R. This fails to account for the possibility that C may not be a valid ciphertext at all: Perhaps there is no pair (m, R) that encrypts to C under PK. It is not clear how a prover can certify such a statement. Another attempt would be to ask the prover for the secret key SK associated to PK. Again, it is not clear how to achieve completeness in case the public key is invalid and there is no corresponding secret key, or soundness in case the public key can be paired with several different secret keys (the choice of which may affect how different invalid ciphertexts decrypt).

Our protocol works as follows: Given a query (PK, C), the verifier asks the prover for the value b that encrypts to C, together with a proof that the rerandomization of C is statistically close to encryptions of b but statistically far from encryptions of $\overline{b}$. If the pair (PK, C) is properly distributed, this forces the prover to a unique correct answer. But since statistical closeness and statistical farness are both efficiently verifiable [BBM11, SV03], the prover can now also certify that a pair (PK, C) is *not* a valid public key-ciphertext pair. We call this protocol DP (the distinguishing protocol).

One important detail is that the protocols for statistical closeness and statistical farness are only guaranteed to solve promise versions of these problems: For a given gap $[\ell, r)$, they can distinguish distributions that are within statistical distance ℓ from those that are at distance at least r, but give no guarantee about the outcome for instances that fall inside the gap. Therefore DP is only complete and sound provided that none of the underlying instances fall inside the respective gaps.

The proof system (Section 7) Given a reduction R from a decision problem L to distinguishing encryptions, this suggests the following constant-round proof system for $\overline{L}$: On a given input, the verifier chooses randomness for the reduction and sends this randomness to the prover. The prover sends back a transcript of the reduction interacting with a distinguishing oracle, which includes a list of queries (PK_i, C_i) made by the reduction together with an answer a_i saying if C_i encrypts 0 or 1 under PK_i, or the pair (PK_i, C_i) is invalid $(\perp)$. The verifier and prover then apply the DP protocol to certify that all the answers a_i are correct.

This proof system is complete and sound, provided that all the inputs (PK_i, C_i, a_i) to the DP protocol satisfy its promise. But in general the verifier does not know in advance if the promise is satisfied or not. We resolve this issue by choosing the width of the gaps $[\ell, r)$ to be sufficiently small and by having the verifier randomize the location of the gaps. This should make it unlikely for any of the queries to fall inside the promise gap of DP.

This approach was also used by Bogdanov and Trevisan [BT06] in the context of non-adaptive reductions. An additional twist is required when the reduction is adaptive because the location of the gaps may affect the answers of the honest prover. For example, imagine an adaptive reduction that does a "binary search" for the gap $[\ell, r)$: If the first answer a is to the right of r, its next query will be $a/2$, and so on until it hits the gap. To handle such reductions, we want to make the location of the gaps in each round independent of the answers of the honest prover in the previous rounds. On the other hand, the locations of these gaps must be consistent with a specific, fixed distinguishing oracle that the prover is required to emulate.

To achieve both objectives we specify a randomized family of decryption oracles, where for each query to the oracle the gap location is random, and the gap locations among the various queries are q-wise independent, where q is an upper bound on the number of queries performed by the reduction. In the first round of the reduction the verifier chooses a random oracle from this family and sends its (polynomial length) description to the prover. The honest prover is then expected to give answers that are consistent with this instantiation of the decryption oracle. By independence, the probability that any of the queries made by the honest prover falls inside the gap will be small. In Section 6.1 we develop the relevant complexity-theoretic framework and we prove Theorem 1 in Section 7.1.

To prevent any of the queries from falling into the gaps $[\ell, r)$, the size of the gaps needs to be inverse proportional to the number of queries made by the reduction. Unless the reduction makes a bounded number of queries, this requires protocols for statistical closeness and statistical farness where the verifier runs in time inverse polynomial to the size of the gap and the gap can be at an arbitrary location. Such protocols were developed by Bhatnagar, Bogdanov, and Mossel [BBM11][1] and we use them in the proof of Theorem 1.

For reductions that make a constant number of queries, it is sufficient to have statistical closeness/farness protocols over a constant number of disjoint gaps $[\ell, r)$. Sahai and Vadhan [SV03] give implementations of such protocols in SZK. Using their protocols and the closure properties of SZK which we recall in Section 6.2, we prove Theorem 2 in Section 7.2.

[1] Technically their statement is not as strong as the one we need here, but their proof can be easily adapted. We provide the details in Appendix A.

2 Definitions

In this section we give definitions of homomorphic evaluation and rerandomization. Although for the proofs of Theorems 1 and 2 we only require weak homomorphic evaluation and weak rerandomization, we also give the corresponding strong notions. Lemma 5, which shows how to convert homomorphic evaluation into rerandomization, applies to both the weak and the strong notions.

Homomorphic evaluation and rerandomization Let $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ be a bit encryption scheme. Fix a security parameter s and let $(PK, SK) \sim \mathbf{Gen}(1^s)$ the distribution on key pairs. (We will assume that s is implicit in the public and secret keys.)

- We say H is a *strong homomorphic evaluator* for $f \colon K_n \to \{0,1\}$ with error ε if for all $m \in K_n$, the random variables
$$(PK, H_{PK}(\mathbf{Enc}_{PK}(m_1), \dots, \mathbf{Enc}_{PK}(m_n))) \quad \text{and} \quad (PK, \mathbf{Enc}_{PK}(f(m)))$$
(where all encryptions are independent) are within statistical distance ε.
- We say H is a *weak homomorphic evaluator* for f with error ε if for all $m \in K_n$,
$$\Pr[\mathbf{Dec}_{SK}(PK, H_{PK}(\mathbf{Enc}_{PK}(m_1), \dots, \mathbf{Enc}_{PK}(m_n))) = f(m)] \geq 1 - \varepsilon,$$
where all encryptions are independent.

A bit encryption scheme is *efficient* if $\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec}$ all run in time polynomial in the security parameter s. We say a partial function $f \colon \{0,1\}^* \to \{0,1\}$ has an *efficient homomorphic evaluator* if there exists a polynomial p in the security parameter such that $f_{p(s)}$ can be evaluated homomorphically in polynomial time uniformly over the security parameter s.

Let $\mathbf{Rer}$ be a randomized function that maps public keys and ciphertexts into ciphertexts. In the following definitions R and R' are independent choices of randomness for $\mathbf{Rer}$.

- We say $\mathbf{Rer}$ is a *strong rerandomization* with error ε if for every $b \in \{0,1\}$, the random variables
$$(PK, E, \mathbf{Rer}_{PK}(E, R)) \quad \text{and} \quad (PK, E, E')$$
where $E, E' \sim \mathbf{Enc}_{PK}(b)$ are independent are within statistical distance ε.
- We say $\mathbf{Rer}$ is a *weak rerandomization* with decryption error ε and rerandomization error ρ if for every $b \in \{0,1\}$, $\Pr[\mathbf{Dec}_{SK}(\mathbf{Rer}_{PK}(\mathbf{Enc}_{PK}(b))) = b] \geq 1 - \varepsilon$ and the random variables
$$(PK, \mathbf{Rer}_{PK}(E, R), \mathbf{Rer}_{PK}(E, R')) \quad \text{and} \quad (PK, \mathbf{Rer}_{PK}(E, R), \mathbf{Rer}_{PK}(E', R'))$$
where $E, E' \sim \mathbf{Enc}_{PK}(b)$ are independent are within statistical distance ρ.

The symmetric increasing function For simplicity we will assume n is odd. Let $K_n \subseteq \{0,1\}^n$ denote the strings of hamming weight $(n-1)/2$ and $(n+1)/2$. The *symmetric increasing function* $sif_n \colon K_n \to \{0,1\}$ takes value 0 on the strings of weight $(n-1)/2$ and value 1 on the strings of hamming weight $(n+1)/2$. Let $K_* = \cup_{n \text{ odd}} K_n$ and $sif \colon K_* \to \{0,1\}$ be the function that equals sif_n on input length n.

The functions sif can be viewed as partial functions on the boolean cube. Their extensions include parity (on input lengths of the form $4k+1$) and majority.

3 The main theorems

We say $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ supports weak homomorphic evaluation of $f\colon K_* \to \{0,1\}$ with error ε if it has an efficient homomorphic evaluator for f with error ε.

A *γ-distinguishing oracle* for $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ is a function D such that

$$\Pr[D(PK, \mathbf{Enc}_{PK}(0)) \text{ accepts}] - \Pr[D(PK, \mathbf{Enc}_{PK}(1)) \text{ accepts}] > \gamma.$$

A *reduction* from a decision problem L to γ-distinguishing encryptions in $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ is an efficient randomized oracle algorithm $R^?$ such that for every valid input x there exists a γ-distinguishing oracle D such that $R^D(x) = L(x)$ with probability at least $8/9$. (For our results the exact constant won't matter, as long as it is strictly greater than $1/2$.)

Theorem 1. *Let $\varepsilon \in (0, 1/18)$ be any constant and $\delta \geq 2\sqrt{\varepsilon}$. Let $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ be a public key encryption scheme that supports homomorphic evaluation of sif with error at most ε. If there is a reduction from L to $(1-\delta)$-distinguishing $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$, then L is in $\mathrm{AM} \cap \mathrm{coAM}$.*

We will assume that the reduction is *query length regular*: On input x, the reduction first computes a query length $m \geq |x|$ and only makes queries of length m. The theorem can be proved without this assumption, but we make it for notational convenience.

In the case when the reduction has constant query complexity, a stronger conclusion can be obtained.

Theorem 2. *Let q be any constant, $\delta > 0$, and $\varepsilon = \varepsilon(q, \delta)$ a sufficiently small constant. Let $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ be a public key encryption scheme that supports homomorphic evaluation of sif with error at most ε. If there is a reduction from L to $(1-\delta)$-distinguishing $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ that makes at most q queries, then L is in statistical zero-knowledge.*

4 Ciphertext rerandomization from homomorphic evaluation

In this section we show how to convert a homomorphic evaluation algorithm for sif into a rerandomization. Let H denote entropy and I denote mutual information.

Claim 3. *Let $X_1, \dots, X_n$ be i.i.d. random variables and $I \sim \{1, \dots, n\}$ a uniformly random index. Let F, G, G' be random variables such that (1) F and G are independent conditioned on X_I, (2) F is independent of I, (3) G and G' are identically distributed and (4) F and G' are independent. Then the random variables (F, G) and (F, G') are within statistical distance $\sqrt{2\,\mathrm{H}(F)/n}$.*

Proof.

$$\begin{aligned}\mathrm{H}(X_I \mid F) &\geq \mathrm{H}(X_I \mid F, I) \quad \text{(conditioning reduces entropy)}\\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{H}(X_i \mid F) \geq \frac{1}{n}\mathrm{H}(X_1, \dots, X_n \mid F) \geq \frac{1}{n}(\mathrm{H}(X_1, \dots, X_n) - \mathrm{H}(F)) = \mathrm{H}(X_I) - \frac{\mathrm{H}(F)}{n}.\end{aligned}$$

Since F and G are conditionally independent of X_I, $\mathrm{I}(F; G) \leq \mathrm{I}(F; X_I)$. Therefore

$$\mathrm{I}(F; G) \leq \mathrm{I}(F; X_I) = \mathrm{H}(X_I) - \mathrm{H}(X_I \mid F) \leq \frac{\mathrm{H}(F)}{n}$$

and the conclusion follows by Pinsker's inequality [Pin64]. □

In the special case when $G = X_I$ we get the following corollary.

Corollary 4. *Let $X_1, \dots, X_n$ be i.i.d and $I \sim \{1, \dots, n\}$ a uniformly random index and F be independent of I. Then (F, X_I) and (F, X) are within statistical distance $\sqrt{2\,\mathrm{H}(F)/n}$, where X is i.d. with $X_1, \dots, X_n$ and independent of F.*

The next lemma shows how to obtain rerandomization from homomorphic evaluation of the sif function.

Lemma 5. *If $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ has a strong (resp. weak) efficient homomorphic evaluator for sif_{2k+1} with error ε, then it has a strong (resp. weak) efficient rerandomization* **Rer** *with error at most $\varepsilon + \sqrt{2c/(k+1)}$ (resp. decryption error ε and rerandomization error $\sqrt{2c/(k+1)}$), where c is the length of ciphertexts.*

Proof. We prove the strong version of the lemma and then describe the changes needed to obtain the weak version. Define **Rer** to be the following procedure. Given public key PK and ciphertext E:

1. Let
$$X_i = \begin{cases} \mathbf{Enc}_{PK}(0, R_i) & \text{for } 1 \le i \le k, \\ E & \text{for } i = k+1, \textit{and} \\ \mathbf{Enc}_{PK}(1, R_i) & \text{for } k+2 \le i \le 2k+1. \end{cases}$$
2. Choose a random permutation π on the set $\{1, \dots, 2k+1\}$.
3. Output $F = H_{PK}(X_{\pi(1)}, \dots, X_{\pi(2k+1)})$.

We will now assume that $E \sim \mathbf{Enc}_{PK}(0)$; the case $E \sim \mathbf{Enc}_{PK}(1)$ is similar. We first condition on the choice of the public key PK, letting ε_{PK} denote the statistical distance between the two distributions in the definition of strong homomorphic evaluator conditioned on PK.

The random variables $X_1, \dots, X_{k+1}$ and F satisfy the assumptions of the Corollary 4, so (X_I, F) (where $I \sim \{1, \dots, k+1\}$) is within statistical distance $\sqrt{2c/(k+1)}$ from (E', F), where $E' \sim \mathbf{Enc}_{PK}(0)$ is independent of F. On the one hand, by the randomness of π, (X_I, F) is identically distributed to $(E, F) = (E, \mathbf{Rer}_{PK}(E))$. On the other hand, by the strong homomorphic property, (E', F) is within distance $\varepsilon_{\mathrm{PK}}$ of a pair of independent random encryptions of 0 under PK. So conditioned on PK, the statistical distance in rerandomization is at most $\varepsilon_{\mathrm{PK}} + \sqrt{2c/(k+1)}$. Averaging over $\varepsilon_{\mathrm{PK}}$ we prove the strong version of the lemma.

To prove the weak version, it is clear by construction that the decryption error is upper bounded by the homomorphic evaluation error. Let F and G be two independent instantiations of **Rer** on the same input E. Conditioned on PK, the random variables $X_1, \dots, X_{k+1}, F, G$ satisfy the assumptions of Claim 3, so (F, G) and (F, G'), where G' is i.i.d with G and therefore with F, are within statistical distance $\sqrt{2c/(k+1)}$. □

5 The distinguishing protocol

In this section we describe a constant-round interactive proof system DP that, given input (PK, C, b), certifies that C is an encryption of b under PK when $b \in \{0, 1\}$ and that (PK, C) is an invalid pair when $b = \perp$. The proof system is parametrized by two gaps $[\ell, r)$ and $[\ell', r')$, which describe a promise on the inputs.

We will assume we have the following constant-round protocols for statistical closeness ($SC_{[\ell,r)}$) and statistical farness ($SF_{[\ell,r)}$), where $0 \leq \ell < r \leq 1$. The protocols take as inputs a pair of sampler circuits D, D' producing distributions over the same set $\{0,1\}^m$ with the following completeness / soundness properties:

- If D, D' are at statistical distance less than ℓ, $SC_{[\ell,r)}(D, D')$ accepts with probability $1 - \sigma$.
- If D, D' are at statistical distance at least r, $SC_{[\ell,r)}(D, D')$ rejects with probability $1 - \sigma$.
- If D, D' are at statistical distance at least r, $SF_{[\ell,r)}(D, D')$ accepts with probability $1 - \sigma$.
- If D, D' are at statistical distance less than ℓ, $SF_{[\ell,r)}(D, D')$ rejects with probability $1 - \sigma$.

Here σ can be any inverse polynomial in the size of the input. The following two theorems state the existence of these protocols. The second one is stronger as provides statistical zero-knowledge implementation, but makes a stronger assumption about the gaps.

Formally we will view SC and SF as promise problems that take ℓ, r, D, D' as their inputs. Theorem 6 essentially follows from work of Bhatnagar, Bogdanov, and Mossel [BBM11]. We provide the missing details in Appendix A.

Theorem 6. *For $r > \ell$, the problems SC and SF are in* AM *where the running time of the verifier is polynomial in the size of D, the size of D', and $1/(r - \ell)$.*

Theorem 7 is proved by Sahai and Vadhan [SV03].

Theorem 7. *For $r^2 > \ell$, the problems SC and SF are in* SZK *where the running time of the verifier is polynomial in the size of D, the size of D', and $1/\ell^{1/\log(r^2/\ell)}$.*

The protocol DP will certify that the rerandomization of C is close to an rerandomized encryption of b but far from a rerandomized encryption of $\bar{b}$ when $b \in \{0,1\}$. When $b = \perp$, it certifies that either the rerandomized encryptions of 0 and 1 are close, or the rerandomized encryption of C is far from both.

Let $Z_{PK,b}$ be the circuit that on input R, R' outputs $\mathbf{Rer}_{PK}(\mathbf{Enc}_{PK}(b, R), R')$, i.e. a rerandomized encryption of b.

The distinguishing protocol $DP_{[\ell,r),[\ell',r')}$

```
On input $(PK, C, b)$, where $b \in \{0, 1, \perp\}$:
If $b = 0$ or $b = 1$:
    Verifier and Prover execute $SF_{[\ell,r)}(Z_{PK,0}, Z_{PK,1})$.
    If the protocol rejects, reject. Otherwise:
        Verifier and Prover execute $SC_{[\ell',r')}(Z_{PK,b}, \mathbf{Rer}_{PK}(C))$.
        If the protocol accepts, accept, else reject.
If $b = \perp$:
    Verifier and Prover execute $SC_{[\ell,r)}(Z_{PK,0}, Z_{PK,1})$.
    If the protocol accepts, accept. Otherwise:
```

```
9.        Verifier and Prover execute $SF_{[\ell',r')}(Z_{PK,0}, \mathbf{Rer}_{PK}(C))$.
10.       Verifier and Prover execute $SF_{[\ell',r')}(Z_{PK,1}, \mathbf{Rer}_{PK}(C))$.
11.       If both accept, accept, else reject.
```

The distinguishing oracle We now define an oracle π that distinguishes between encryptions of 0 and encryptions of 1. This oracle answers $\perp$ on all queries (PK, C) that do not represent valid key-ciphertext pairs and answers bad on all queries that fall inside the gaps of the underlying protocols SC and SF. We then show that for every pair (PK, C) that falls outside the gaps, $b = \pi(PK, C)$ is the unique answer that makes $DP(PK, C, b)$ accept.

Assume $\ell' < r/2$ and consider the following oracle

$$\pi_{[\ell,r),[\ell',r')}(PK, C) = \begin{cases} \perp, & \text{if } d < \ell \text{ or } (d \geq r \text{ and } d_0 \geq r' \text{ and } d_1 \geq r') \\ 0, & \text{if } d \geq r \text{ and } d_0 < \ell' \text{ (and so } d_1 \geq \ell') \\ 1, & \text{if } d \geq r \text{ and } d_1 < \ell' \text{ (and so } d_0 \geq \ell') \\ \text{bad}, & \text{if } d \in [\ell, r) \text{ or } d_0 \in [\ell', r') \text{ or } d_1 \in [\ell', r') \end{cases}$$

where $d = \mathrm{sd}(Z_{PK,0}, Z_{PK,1})$ and $d_b = \mathrm{sd}(Z_{PK,b}, \mathbf{Rer}_{PK}(C))$ (for $b \in \{0, 1\}$).

Let $\pi = \pi_{[\ell,r),[\ell',r')}$ and $DP = DP_{[\ell,r),[\ell',r')}$. The following claim shows that π is a distinguishing oracle.

Claim 8. *Assume* $\mathbf{Rer}$ *is a rerandomization with decryption error* $\varepsilon < (1 - r)^2/2$ *and rerandomization error* $\rho < \ell'^2$. *Then* $\Pr[\pi(PK, \mathbf{Enc}_{PK}(b)) = b] \geq 1 - \sqrt{2\varepsilon} - \sqrt{\rho}$ *for every* $b \in \{0, 1\}$.

Proof. First we show that the statistical distance between the distributions

$$(PK, \mathbf{Rer}_{PK}(\mathbf{Enc}_{PK}(0, R), R') = Z_{PK,0}) \quad \text{and} \quad (PK, \mathbf{Rer}_{PK}(\mathbf{Enc}_{PK}(1, R), R') = Z_{PK,1})$$

is at least $1 - 2\varepsilon$. Consider the test T that on input (PK, C), samples SK conditioned on PK, and outputs $\mathbf{Dec}_{SK}(C)$. Since $\mathbf{Rer}$ is a rerandomization with decryption error ε, we have for every $b \in \{0, 1\}$

$$\Pr[\mathbf{Dec}_{SK}(\mathbf{Rer}_{PK}(\mathbf{Enc}_{PK}(b))) = b] \geq 1 - \varepsilon.$$

Therefore T distinguishes the two distributions with advantage $1 - 2\varepsilon$. By Markov's inequality, for at least a $1 - \sqrt{2\varepsilon}$ fraction of the PK, the statistical distance between $Z_{PK,0}$ and $Z_{PK,1}$ is at least $1 - \sqrt{2\varepsilon}$. Since $\mathbf{Rer}$ has a rerandomization error ρ, the statistical distance between

$$(PK, \mathbf{Rer}_{PK}(C, R), \mathbf{Rer}_{PK}(C, R')) \quad \text{and} \quad (PK, \mathbf{Rer}_{PK}(C, R), \mathbf{Rer}_{PK}(C', R'))$$

(where $C, C' \sim \mathbf{Enc}_{PK}(b)$ are independent) is at most ρ. By Markov's inequality, for at least a $1 - \sqrt{\rho}$ fraction of the pairs (PK, C), the statistical distance between $\mathbf{Rer}_{PK}(C, R')$ and $\mathbf{Rer}_{PK}(C', R') = Z_{PK,b}$ is at most $\sqrt{\rho} < \ell'$. The claim follows by taking a union bound. □

The following claims are immediate from the definitions and the completeness and soundness assumptions on SC and SF.

Claim 9. *(Completeness) Assume* $\ell' < r/2$ *and* $\pi(PK, C) \neq \text{bad}$. *Then* $DP(PK, C, \pi(PK, C))$ *accepts with probability at least* $1 - 3\sigma$.

Claim 10. *(Soundness) Assume* $\ell' < r/2$. *If* $DP(PK, C, b)$ *accepts with probability more than* 3σ, *then* $\pi(PK, C) \in \{b, \text{bad}\}$.

6 Complexity theoretic setup

In this section we cover the complexity-theoretic framework for the proofs of Theorems 1 and 2.

6.1 Promise oracles for adaptive reductions

Let Ξ be any finite set of values that includes the special symbol bad. An *oracle family* over input length m with size d is a multiset Π of functions $\pi \colon \{0,1\}^m \to \Xi$, $1 \leq i \leq d$. We say Π is *ε-bad* if for every input x, $\Pr_{\pi \sim \Pi}[\pi(x) = \text{bad}] \leq \varepsilon$.

Let $F \colon \{0,1\}^m \to [d]$ be a function. The oracle $\Pi_F \colon \{0,1\}^m \to \Xi$ is given by $\Pi_F(z) = \pi_{F(z)}(z)$. In the lemma below F will be a randomized function of the same form.

Lemma 11. *Let $R^?$ be a reduction that on an input of length n, makes at most q queries of length m. Let Π be an oracle family of size d. Assume d is a power of two. There exists a randomized function $F \colon \{0,1\}^m \to [d]$ such that:*

- *F is computable in time (and hence uses randomness) polynomial in m, q, and d.*
- *For every input x of length n, the probability that $R^{\Pi_F}(x)$ never receives* bad *as an answer to any of its queries is at least $(1-\varepsilon)^q$.*

Proof. Fix m and let $F \colon \{0,1\}^m \to [d]$ be a q-wise independent function family. Using standard constructions, F can be described by $O(mq)$ random bits and is computable in time polynomial in m, q, and d.

Let $(Q_1, a_1), \ldots, (Q_q, a_q)$ denote the query-answer pairs of the reduction when interacting with the oracle Π_F. We may assume all queries are distinct. We write the probability that any of the a_i's equals bad as a product of conditional probabilities. Let p_i be the probability that $a_i \neq \text{bad}$ conditioned on $a_1, \ldots, a_{i-1} \neq \text{bad}$.

Let's look at p_1 first. Since Π is ε-bad, the probability that a_1 is bad is at most ε and $p_1 \geq 1-\varepsilon$. Now we consider p_i. Since F is q-wise independent it follows that conditioned on every possible collection of values of $F(Q_1), \ldots, F(Q_{i-1})$ (which in particular determine the event $a_1, \ldots, a_{i-1} \neq \text{bad}$), $F(Q_i)$ is uniformly distributed in $[d]$. Since Π is ε-bad, the conditional probability that $a_i = \text{bad}$ can be at most ε, and so $p_i \geq 1 - \varepsilon$. Multiplying the conditional probabilities gives the second part of the lemma. □

6.2 Statistical zero-knowledge

We recall some results about the complexity of statistical zero-knowledge SZK. Sahai and Vadhan [SV03] show that the statistical distance problem $SD = SF_{[1/9,8/9)}$ is complete for SZK under many-one reductions.

We also need the following result of Sahai and Vadhan [SV03] that states the closure of SZK under truth-table reductions.

Theorem 12. *There is a deterministic algorithm that takes as input instances $x_1, \ldots, x_k$ of SD and a boolean predicate $P \colon \{0,1\}^k \to \{0,1\}$ and outputs an instance x of SD such that $SD(x) =$*

$P(SD(x_1), \ldots, SD(x_k))$. The running time of the algorithm is polynomial in 2^k and the sizes of $x_1, \ldots, x_k$.

We also need the following fact, which says that reductions within SZK can without loss of generality be assumed deterministic.

Claim 13. *If there is a randomized many-one reduction R from L to SD such that $\Pr[SD(R(x)) = L(x)] \geq p$, where p is any constant above $1/2$, then L is in* SZK.

Proof. The reduction takes input x and randomness r and produces a pair of circuits D, D'. Let $E_x(r,s)$ (resp. $E'_x(r,s)$) be the circuits that on input r, s runs the reduction on input x and randomness r and outputs $D(s)$ (resp., $D'(s)$).

Assume $L(x) = SD(R(x))$ with probability at least $8/9$ over the randomness of the reduction. For $x \in L$, the statistical distance between E_x and E'_x is at least $(8/9)^2 \geq 2/3$ because at least $8/9$ choices of r contribute at least $8/9$ to the statistical distance. If $x \notin L$, then the statistical distance is at most $8/9 \cdot 1/9 + 1/9 \cdot 1 \leq 1/3$, because for at least $8/9$ choices of r the statistical distance over s is at most $1/9$, and for the other choices it is at most 1. Therefore L reduces deterministically to $SF_{[1/3,2/3)}$, so L is in SZK by Theorem 7.

If $\Pr[L(x) = SD(R(x))]$ is any constant above $1/2$, the probability can be first amplified to $8/9$ via Theorem 12 with the majority predicate. □

Combining Theorem 12 and Claim 13 we get the following corollary.

Corollary 14. *Suppose there is a randomized algorithm A that on input x of length n and randomness r computes inputs $x_1, \ldots, x_k$ and a predicate $P\colon \{0,1\}^k \to \{0,1\}$, where $k = O(\log n)$ and accepts if $P(SD(x_1), \ldots, SD(x_k))$ is true. If $\Pr[A(x) = L(x)] \geq p$, where p is any constant greater than $1/2$, then L is in* SZK.

7 Proofs of the main theorems

7.1 Proof of Theorem 1

Let $F_\omega\colon \{0,1\}^m \to [d]$ be the randomized function from Lemma 11, with ω describing the randomness. We set $I_j = \left[\frac{1}{3} + \frac{j-1}{3d}, \frac{1}{3} + \frac{j}{3d}\right)$ and $I'_j = \frac{1}{3} I_j$, where $1 \leq j \leq d$.

The decision protocol DL: On input x:

V: Compute the oracle query length m. Let d be the smallest power of two above $90q$ where q is an upper bound on the number of queries $R^?(x)$ makes. Choose randomness r for the reduction and randomness ω for F_ω. Send r, d, ω to the prover.

P: Send a sequence $((PK_i, C_i), b_i)$, $1 \leq i \leq q$ of oracle query-answer pairs.

V: Verify that the received query-answer pairs determine an accepting computation of $R^?(x, r)$. If not, reject. For every query i, compute $j = F_\omega(PK_i, C_i)$ and let $[\ell_i, r_i) = I_j$ and $[\ell'_i, r'_i) = I'_j$.

V, P: Execute in parallel the protocols $DP_{[\ell_i,r_i),[\ell'_i,r'_i)}(PK_i, C_i, b_i)$ for $1 \leq i \leq q$ with completeness/soundness gap $\sigma = 1/9q$. If any of them result in rejection, reject. Otherwise, accept.

Let $\pi_j = \pi_{I_j,I'_j}$ and Π_F be the oracle from Lemma 11.

Claim 15. *The oracle family $\{\pi_j\}_{1\leq j\leq d}$ is at most $3/d$-bad.*

Proof. Query (PK, C) is bad for π_j if $\mathrm{sd}(Z_{PK,0}, Z_{PK,1}) \in I_j$ or $\mathrm{sd}(Z_{PK,0}, \mathbf{Rer}_{PK}(C)) \in I'_j$ or $\mathrm{sd}(Z_{PK,1}, \mathbf{Rer}_{PK}(C)) \in I'_j$. Since the intervals I_j are disjoint, and so are the intervals I'_j, each of the three events occurs with probability at most $1/d$, so their union occurs with probability at most $3/d$. □

Proof of Theorem 1 It is sufficient to prove that $L \in \mathrm{AM}$. By applying the same argument to its complement $\overline{L}$ we also get $L \in \mathrm{coAM}$.

Assume $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$ supports homomorphic evaluation of sif with error at most ε and there is a reduction $R^?$ from L to $(1-\delta)$-distinguishing encryptions.

We instantiate the constructions with the following parameters. Let ε be the homomorphic evaluation error and c an upper bound on the length of ciphertexts queried by the reduction on input x. Set the number of inputs $2k+1$ to sif to $4c/\varepsilon^4$. Let $\mathbf{Rer}$ be the rerandomization from Lemma 5 with this value of k. By Lemma 5 its decryption error is ε and its rerandomization error is at most $\rho \leq \varepsilon^2$. The protocol DP is instantiated with this rerandomization.

Claim 16. *For an appropriate choice of parameters and for every F, Π_F is a $(1-\delta)$-distinguishing oracle.*

Proof. Notice that all the intervals I_j are within $[1/3, 2/3)$ and I'_j are all within $[1/9, 2/9)$. Since $\varepsilon < 1/18$ and $\rho < 1/81$, we have for every j, π_j satisfies the assumptions of Claim 8, which shows that

$$\Pr[\pi_j(PK, \mathbf{Enc}_{PK}(b)) = b] \geq 1 - \sqrt{2\varepsilon} - \sqrt{\rho} \geq 1 - 2\sqrt{\varepsilon} \geq 1 - \delta.$$

Since Π_F equals some π_j on every query, it follows that the same formula is true for Π_F, so Π_F is a $(1-\delta)$-distinguishing oracle. □

By Theorem 6, the verifier for DL can be implemented in polynomial time. Theorem 1 the follows by the next two claims:

Claim 17. *(Completeness) If $x \in L$, there exists a prover that makes $DL(x)$ accept with probability at least $2/3$.*

Proof. In the second step, the prover will emulate $R^{\Pi_F}(x, r)$. If the oracle returns bad on any of the queries in this emulation, the prover aborts (causing the verifier to reject). In the fourth step, the prover emulates the honest prover for $DP_{[\ell_i,r_i),[\ell'_i,r'_i)}$.

Let B be the event that $R^{\Pi_F}(x, r)$ rejects or Π_F returns bad on any of the queries in $R^{\Pi_F}(x, r)$ or any of DP protocols rejects. If B does not happen, then the verifier accepts. We upper bound the rejecting probability of the verifier by taking a union bound. Since Π_F is a distinguishing oracle, R^{Π_F} rejects with probability at most $1/9$. By Claim 15 and Lemma 11, R^{Π_F} returns bad with probability at most $1 - (1 - 3/d)^q \leq 1/9$; by Claim 9, each of the step 4 protocols rejects with probability at most $1/9q$, so by a union bound B happens with probability at most $1/3$. □

Claim 18. *(Soundness) If $x \notin L$ then no prover makes $DL(x)$ accept with probability at least $1/3$.*

Proof. Assume $x \notin L$. If $DL(x)$ accepts, then at least one of the following must be true:

1. $R^{\Pi_F}(x, r)$ accepts, or
2. Π_F returns bad on at least one query in $R^{\Pi_F}(x, r)$, or
3. $DL(x)$ accepts, $R^{\Pi_F}(x, r)$ rejects, and Π_F never returns bad.

We upper bound the probabilities of each of these events. Since Π_F is a distinguishing oracle, the first one occurs with probability at most $1/9$. By Claim 15 and Lemma 11, the second one occurs with probability at most $1 - (1 - 3/d)^q \leq 1/9$. If the third event is satisfied, then b_i must differ from $\Pi_F(PK_i, C_i) = \pi_{p_i,q_i}(PK_i, C_i)$ for at least one i. By Claim 10, the i'th instantiation of the DP protocol in then accepts with probability at most $1/9$. By a union bound, $DL(x)$ accepts with probability at most $1/3$. □

7.2 Proof of Theorem 2

Let $I_j, 1 \leq j \leq d$ be the following collection of intervals: $I_j = [\ell_j, r_j)$ where $r_1 = 1/2$, $\ell_j = r_j^2/4$, and $r_{j+1} = \ell_j$. Let $I'_j = \frac{1}{3} I_j$. Assume the reduction makes at most q queries on every input and let $d = 27q \cdot 3^q$.

By Theorem 7, for every j the problems $SC_{I_j}, SC_{I'_j}, SF_{I_j}, SF_{I'_j}$ are all in SZK so by Theorem 12 and the completeness of SD, DP_{I_j,I'_j} is also in SZK for every j.

Consider the following algorithm A. On input x, choose randomness r for R and a random $j \sim [d]$ and accept if there exists a sequence of answers $(a_1, \ldots, a_q) \in \{0, 1, \perp\}^q$ such that $R(x, r)$ accepts given these oracle answers and $DP_{I_j,I'_j}(Q_i, a_i)$ accepts for all $1 \leq i \leq q$. Since DP_{I_j,I'_j} is in SZK and SD is complete for SZK, A satisfies the assumption of Corollary 14, so if we can prove that $\Pr[A(x) = L(x)] \geq 2/3$, it will follow that L is in SZK.

Say j is bad if $\pi_j = \pi_{I_j,I'_j}$ answers bad on any pair (Q, a) queried by A. Since A makes at most $q3^q$ queries, by Claim 15 and a union bound the probability that A answers bad on any of its queries is at most $1/9$.

Fix an input x. By our choice of parameters, when ε is sufficiently small and $\rho = \varepsilon^2$, Claim 8 guarantees that π_j is a $(1 - 4\varepsilon)$-decryption oracle for every $1 \leq j \leq d$. So for at least $8/9$ fraction of r, $R^{\pi_j}(x, r) = L(x)$. Therefore with probability at least $7/9$, both $R^{\pi_j}(x, r) = L(x)$ and π_j never answers bad on any of A's queries. By Claims 9 and Claim 10, it must then hold that $a = \pi_j(Q)$ for all query-answer pairs (Q, a) made by A, and so $A(x) = L(x)$.

Acknowledgment

We thank Alon Rosen for pointing out that our results apply to homomorphic schemes that support evaluations of parity and not only majority.

References

[AGGM06] Adi Akavia, Oded Goldreich, Shafi Goldwasser, and Dana Moshkovitz. On basing one-way functions on NP-hardness. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, 2006.

[BBM11] Nayantara Bhatnagar, Andrej Bogdanov, and Elchanan Mossel. The computational complexity of estimating convergence time. In *Proceedings of the 15th International Workshop on Randomization and Computation (RANDOM)*, 2011.

[Bra79] Gilles Brassard. Relativized cryptography. In *Proceedings of the 20th IEEE Symposium on Foundations of Computer Science*, pages 383–391, 1979.

[BT06] Andrej Bogdanov and Luca Trevisan. On wost-case to average-case reductions for NP problems. *SIAM J. Comp.*, 36(4), 2006.

[BV11] Z. Brakerski and V. Vaikuntanathan. Efficient Fully Homomorphic Encryption from (Standard) LWE. In *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science*, 2011.

[EY80] Shimon Even and Yacob Yacobi. Cryptography and NP-completeness. In *Proceedings of the 7th ICALP*, volume 85 of *LNCS*, pages 195–207. Springer-Verlag, 1980.

[FF93] Joan Feigenbaum and Lance Fortnow. Random-self-reducibility of complete sets. *SIAM Journal on Computing*, 22:994–1005, 1993.

[Gam85] T. El Gamal. A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans. Info. Theory*, 31(4):469–472, 1985.

[Gen09] Craig Gentry. Fully Homomorphic Encryption Using Ideal Lattices. In *STOC*, pages 169–178, 2009.

[GG98] Oded Goldreich and Shafi Goldwasser. On the possibility of basing cryptography on the assumption that $P \neq NP$. Unpublished manuscript, 1998.

[GK93] Oded Goldreich and Eyal Kushilevitz. A perfect zero-knowledge proof system for a problem equivalent to the discrete logarithm. *J. Cryptology*, 6(2):97–116, 1993.

[Gol00] Oded Goldreich. Candidate one-way functions based on expander graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 7(090), 2000.

[Pai99] P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptology – Eurocrypt '99*, pages 223–238, 1999.

[Pei09] Chris Peikert. Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In *Proceedings of the 41th ACM Symposium on Theory of Computing*, pages 333–342, New York, NY, USA, 2009. ACM.

[Pin64] M. S. Pinsker. *Information and information stability of random variables and processes.* Holden-Day, 1964.

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6), 2009.

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

[vDGHV10] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan. Fully Homomorphic Encryption from Integers. In *Eurocrypt*, 2010.

A An AM protocol for statistical closeness

Bogdanov, Bhatnagar and Mossel [BBM11] show the existence of a protocol for statistical farness (SF) meeting the specifications of Theorem 6. They also give a protocol for statistical closeness (SC), but they only provide a soundness proof for gaps $[\ell, r)$ satisfying $r/\ell \geq 4$. We show how to extend their protocol and analysis to general gaps.

Theorem 19. *For $r > \ell$, the problem $SC_{[\ell,r)}(D, D')$ is in* AM *where the running time of the verifier is polynomial in the size of D, the size of D', and $1/(r - l)$.*

Let $N(t) = |\{\omega : |D^{-1}(\omega)| \geq t \text{ and } |D'^{-1}(\omega)| \geq t\}|$. From [BBM11], there is a lower bound protocol for $N(t)$ with completeness $1 - \delta/20n$ and soundness $\delta/20n$. More specifically, they show that the following decision problem is in AM:

Input: A pair of circuits $D, D' : \{0,1\}^n \to \{0,1\}$, a number $1 \leq t \leq 2^n$, a target number $0 \leq \tilde{N} \leq 2^n$, and a fraction $0 < \delta \leq 1$ (represented in unary).
Yes instances: $(D, D', t, \tilde{N}, \delta)$ such that $N(t) \geq \tilde{N}$.
No instances: $(D, D', t, \tilde{N}, \delta)$ such that $N((1-\delta)t) < (1-\delta)\tilde{N}$.

Following the ideas of [BBM11] we have the following protocol for statistical closeness:

An AM **protocol for** SC**:** On input ℓ, r, D, D': Set $\delta = (r - \ell)/4$ and

P: Send claims $\tilde{N}_i$ for the values $N_i = N((1-\delta)^{-i}), 0 \leq i \leq en/\delta$.

P, V: Run the AM lower bound protocol for N_i on inputs $(D, D', (1-\delta)^{-i}, \tilde{N}_i, \delta)$ for every $1 \leq i \leq en/\delta$. If all of them pass accept, otherwise reject.

V: Accept if $\sum_{i=0}^{en/\delta}(\tilde{N}_i - \tilde{N}_{i+1})(1-\delta)^{-i} \geq (1-\delta)(1-\ell) \cdot 2^n$.

The completeness and soundness of the protocol rely on the following approximation from [BBM11]:

$$\sum_{i=0}^{en/\delta}(N_i - N_{i+1})(1-\delta)^{-i} \leq (1 - \mathrm{sd}(D, D'))2^n \leq \sum_{i=0}^{en/\delta}(N_i - N_{i+1})(1-\delta)^{-(i+1)}.$$

Claim 20 (Completeness)**.** *If $\mathrm{sd}(D, D') \leq \ell$ then the protocol accepts with probability $2/3$.*

Proof. Assume the honest prover claims that $\tilde{N}_i = N_i$ for every i. By the completeness of the lower bound protocol and a union bound, with probability at least $2/3$ none of the lower bound protocol rejects. In this case, using the above approximation we have

$$\sum_{i=0}^{en/\delta}(\tilde{N}_i - \tilde{N}_{i+1})(1-\delta)^{-i} \geq (1-\delta)(1 - \mathrm{sd}(D, D')) \cdot 2^n. \qquad \square$$

Claim 21 (Soundness)**.** *If the protocol accepts with probability at least* $1/3$ *then* $\mathrm{sd}(D, D') \leq r$*.*

Proof. Assume the verifier accepts with probability at least $1/3$. By the soundness of the lower bound protocol for $N(t)$ and a union bound, there exists at least one setting of the randomness of the verifier for which $N_{i-1} \geq (1-\delta)\tilde{N}_i$ for all i (where $N_{-1} = N_0$) and the verifier accepts. Now (using the fact that the value $N_{en/\delta+1}$ is zero):

$$\begin{aligned}
\sum_{i=-1}^{en/\delta}(N_i - N_{i+1})(1-\delta)^{-i} &= N_{-1}(1-\delta) + \sum_{i=0}^{en/\delta} N_i((1-\delta)^{-i} - (1-\delta)^{-i+1}) \\
&\geq \tilde{N}_0(1-\delta)^2 + (1-\delta)\sum_{i=0}^{en/\delta} \tilde{N}_{i+1}((1-\delta)^{-i} - (1-\delta)^{-i+1}) \\
&= (1-\delta) \cdot \sum_{i=0}^{en/\delta}(\tilde{N}_i - \tilde{N}_{i+1})(1-\delta)^{-i+1} \\
&\geq (1-\delta)^3(1-\ell) \cdot 2^n
\end{aligned}$$

so we get that $1 - \mathrm{sd}(D, D') \geq 1 - \ell$ and therefore $\mathrm{sd}(D, D') \leq 1 - (1-\delta)^3(1-\ell) \leq \ell + 4\delta$. Setting $\delta = (r - \ell)/4$ proves the claim. □