The LPN Problem with Auxiliary Input

Yu Yu

Université catholique de Louvain – Crypto Group
B-1348 Louvain-la-Neuve – Belgium
yu.yu@uclouvain.be

Abstract. This paper investigates the Learning from Parity with Noise (LPN) problem under the scenario that the unknowns (secret keys) are only unpredictable instead of being uniformly random to the adversaries. In practice, this corresponds to the case where an adversary already possesses some additional knowledge about the secret key. In the information-theoretic setting, we show that the problem is robust against arbitrary leakages as long as the unknowns remain some sufficient amount of min-entropy. In the computational setting, we prove Dodis et al.'s [STOC'09] conjecture that the auxiliary-input LPN assumption is implied by the standard LPN assumption, *i.e.*, encryption schemes based on the standard LPN assumption is secure against any exponentially hard-to-invert auxiliary input.

Our setting is more general than the traditional model of auxiliary input which deals with secret keys of sufficient min-entropy, in particular, we allow leakages that information-theoretically determine their secret keys as long as the keys remain a linear amount of unpredictability pseudoentropy. Further, unlike most other schemes, our result is reducible to a well-known hardness problem and does not quantify over all hard-to-invert auxiliary functions.

1 Introduction

1.1 The LPN problem

The learning from parity with noise (LPN) problem refers to solve the following system of equations:

$$\begin{aligned}
&a_{11}\cdot x_1 + a_{12}\cdot x_2 + \cdots + a_{1n}\cdot x_n + e_1 = b_1 \pmod 2\\
&a_{21}\cdot x_1 + a_{22}\cdot x_2 + \cdots + a_{2n}\cdot x_n + e_2 = b_2 \pmod 2\\
&\vdots\\
&a_{t1}\cdot x_1 + a_{t2}\cdot x_2 + \cdots + a_{tn}\cdot x_n + e_t = b_t \pmod 2
\end{aligned}$$

where $x = [x_1, x_2, \cdots, x_n]$ are unknowns uniform over $\{0,1\}^n$, co-efficient Boolean matrix $[a_{ij}]_{t\times n}$ is also uniform over $\{0,1\}^{t\times n}$, both $[a_{ij}]_{t\times n}$ and vector $[b_1, \cdots, b_t]$

are public, and error vector $e = [e_1, e_2, \cdots, e_t]$ is secret but with a known distribution, namely each e_i is independently distributed as below [1]:

$$\begin{cases} e_i = 0, \text{ with probability } \gamma \\ e_i \in U_1, \text{ with probability } 1-\gamma \end{cases} \tag{1}$$

for $0 < \gamma < 1$.

Hardness of the LPN Problem. While the noise-free case (*i.e.* $\gamma = 1$) can be efficiently solved by Gaussian elimination, the problem appears to be significantly harder (in asymptotic order) for any $\gamma \in (0, 1)$. In particular, the LPN problem is known to be NP-Hard [1], and is hard even within an approximation ratio of two [2]. Blum, Kalai and Wasserman [3] gave the first sub-exponential algorithm that solves the problem in time $2^{O(n/\log n)}$ with $2^{O(n/\log n)}$ equations. Regev [4] generalized the problem to the higher moduli case, namely the LWE (learning with error) problem, and showed that any solution to LWE implies a quantum solution to worst-case lattice problems such as SVP and SIVP, for which the best known polynomial-time algorithms only yield subexponential approximation factors [5–7].

1.2 Cryptographic Applications

The LPN Assumption. The hardness of the LPN problem can be translated into the LPN assumption[2], on which efficient encryption schemes can be constructed. The (decisional) LPN assumption [8]: for every constant $0 < \gamma < 1$ and for every polynomial $t = poly(n)$, the two ensembles are computationally indistinguishable:

$$\{A, (A \cdot X + E)\}_{n\in\mathbb{N}} \overset{c}{\approx} \{A, U_t\}_{n\in\mathbb{N}} \tag{2}$$

where A, X and U_t are uniform over $\{0,1\}^{t\times n}$, $\{0,1\}^n$ and $\{0,1\}^t$ respectively, each bit of E is independent and identically distributed as in (1), "$+$" denotes bitwise XOR, and $A{\cdot}X$ refers to matrix-vector multiplication over GF(2).

Cryptography with LPN. To make the above a useful (decryptable) CPA-secure symmetric encryption scheme, plaintexts have to be encoded against random noise, *e.g.*, the encryption algorithm $\mathcal{E}_{x,\gamma}$ on encryption key x, noise rate γ, and message $\boldsymbol{m} = b_1 \cdots b_{t/l}$, outputs

$$\mathcal{E}_{x,\gamma}(\boldsymbol{m}) = (A, Ax + e + \mathsf{enc}(b_1)\mathsf{enc}(b_2)\cdots\mathsf{enc}(b_{t/l})), \text{ where } \mathsf{enc}(b_i) \overset{def}{=} \underbrace{b_i \cdots b_i}_{l}$$

[1] It is equivalent to say that each e_i follows a Bernoulli distribution of parameter $(1-\gamma)/2$, *i.e.*, $\Pr[e_i = 1] = (1-\gamma)/2$ and $\Pr[e_i = 0] = (1+\gamma)/2$.

[2] We typically assume that the LPN problem cannot be efficiently solved (with any non-negligible probability), called the computational LPN assumption, which is equivalent to the decisional LPN assumption [8] (see Theorem 2).

and the decryption algorithm $\mathcal{D}$ takes as input a key x and a ciphertext $(A,\boldsymbol{c})$, computes $[\mathsf{enc}'(b_1),\cdots,\mathsf{enc}'(b_{t/l})]\leftarrow A\cdot x+\boldsymbol{c}$, and decodes each $\mathsf{enc}'(b_i)$ to 0 if it has Hamming weight less than $l/2$, and to 1 otherwise, where the decoding success rate is more than $1-exp^{-\frac{l\cdot\gamma^2}{2}}$ (due to the Chernoff bound). We refer to [9, 10] for other encryption schemes with additional properties, and [11–13] for other cryptographic applications.

1.3 Related work

One can observe that if any constant portion (say α) of the unknowns are leaked, the LPN assumption still holds, with the security parameter decreased from n to $(1-\alpha)n$ (see also [14] for analogous results for the LWE assumption). Dodis, Kalai and Lovett [8] further conjectured the auxiliary-input LPN assumption.

Conjecture 1 (The auxiliary-input LPN assumption [8]). For any constant $0<\alpha<1$, and any easy-to-compute but $(1-\alpha)$-exponentially hard-to-invert function f (*i.e.* $\Pr[\mathsf{A}(f(X))=X]\leq 2^{(1-\alpha)n}$ for any PPT adversary A), let $L\stackrel{def}{=}f(X)$, the standard LPN assumption holds even conditioned on L, *i.e.*,

$$\{A,(A\cdot X+E),L\}_{n\in\mathbb{N}} \stackrel{c}{\approx} \{A,U_t,L\}_{n\in\mathbb{N}} \tag{3}$$

They showed that if the above holds, it implies CCA secure encryption schemes (assuming the existence of trapdoor permutations), reusable average-case obfuscators (for point functions) and reusable extractors, which remain secure with exponentially hard-to-invert auxiliary input. However, they were not able to obtain these results from the standard LPN assumption [3], and they only proved the case that the noise is polynomially strong (*i.e.* $\gamma=\frac{1}{poly(n)}$). We note that this would lead to low efficiency as even with noise rate $\gamma=1/n$ each message has to be repeated $l=\Omega(n^2)$ times (see Section 1.2) in codeword to ensure a constant success rate of decoding/decryption.

The problem of securing cryptographic schemes against leakages of arbitrary information was studied in various settings and contexts (e.g. [15–18]). In the context of side channel attacks [19], Ishai et al. considered the case of making circuits provably secure against probing [20] and even tampering with [21] a bounded number of wires. Micali and Reyzin [22] initiated the study of building leak-resilient stream ciphers, followed by the constructions that are secure in the ideal cipher / random oracle model [23, 24] and in the standard model [25, 26]. We mention also some recent constructions of leak-resilient public-key encryption schemes [14, 27] and signatures schemes [28, 29]. To summarize, most of the work mentioned above require the secret key to have some sufficient min-entropy left

[3] Dodis et al. showed that the auxiliary-input LPN assumption is implied by a stronger and less well-studied assumption (than the standard LPN assumption), which they called the Learning Subspace with Noise (LSN) assumption. Due to Raz's attack (see [8] for details), the LSN assumption is likely to hold only for $\gamma=1/poly(n)$ (as opposed to constant noise rate) .

while in this paper we extend the work of Dodis et al. [8], which deals with the situation where the secret key does not necessarily have any min-entropy (*e.g.* secret keys and their leakages can be one-to-one), but is exponentially unpredictable in a computational sense.

1.4 Summary of contributions

In this paper, we show that the LPN problem is robust against arbitrary exponentially hard-to-invert leakages in both settings. In particular, in the computational setting, we show the equivalence between the two versions (decisional and computational) of the standard LPN assumption, and we prove the conjecture in [8] that the auxiliary-input LPN assumption is implied by the standard LPN assumption. Therefore, in the context of cryptography with auxiliary input, we obtain positive results (as opposed to [18]) that: ***(a)*** do not require min-entropy (or HILL pseudo-entropy) sources; ***(b)*** are reducible to a well-known hardness problem (rather than relying on a new assumption [8] for which the constant noise case has been refuted); ***(c)*** do not quantify over all hard-to-invert auxiliary inputs (unlike [15]).

On (non-)uniformity. In general, we consider non-uniform adversaries, but all the reductions are uniform for efficiently computable leakages, which already cover most leakage types in practice. Otherwise said, non-uniformity is needed only for drawing random values from joint distribution $[X,L]$ that is not efficiently sampleable.

2 The LPN Problem in the Information-Theoretic Setting

In this section, we investigate the LPN problem information theoretically and show its robustness against arbitrary exponentially hard-to-invert leakages (in terms of average min-entropy). We note that results obtained in the information-theoretic setting hold only with respect to restrictive values of (n,t,γ), in the sense that they do not give rise to practical encryptions schemes (*i.e.* statistical security is trivial). Therefore, we defer most of the proofs to the appendix.

2.1 Notations, definitions and lemmata

Throughout this paper, we use calligraphic letters $\mathcal{X}$, $\mathcal{Y}$ to denote sets, uppercase letter X, Y to denote random variables, and lower-case letters $\boldsymbol{x}$, $\boldsymbol{y}$ to denote the binary values $x_0 \cdots x_{n-1}$ and $y_0 \cdots y_{n-1}$ that X and Y assume. Set $\{0, 1, \cdots, n-1\}$ is denoted by $[n]$. $|\boldsymbol{a}|$ denotes the length of $\boldsymbol{a}$, and $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ denotes the mod 2 inner product of binary vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Unless otherwise specified, all arithmetic operations are over GF(2).

We denote by $|S|$ the length of S, by $H_W(S)$ the Hamming weight of S, and by $\langle X, Y \rangle$ the inner product of X and Y modulo 2. U_n denotes a random variable uniform over $\{0,1\}^n$ and independent of other random variables under consideration. $\{0,1\}^n/\{0^n\}$ refers to the subset of $\{0,1\}^n$ that excludes zero vector 0^n.

For two distributions X and Y over the same set $\mathcal{S}$, their statistical distance, denoted by $\delta(X;Y)$, is defined as the maximum distinguishing advantage with respect to all adversaries A:

$$\delta(X;Y) \stackrel{def}{=} \max_{\mathsf{A}} |\Pr[\mathsf{A}(X)=1] - \Pr[\mathsf{A}(Y)=1]| = \frac{\sum_{\boldsymbol{s}\in\mathcal{S}} |\Pr[X=\boldsymbol{s}] - \Pr[Y=\boldsymbol{s}]|}{2}$$

whereas their computational distance, denoted by $\delta_s(X;Y)$, limits the complexity of the above A to size s. If distribution X is over $\{0,1\}^n$ then let $d(X) \stackrel{def}{=} \delta(X;U_n)$ and $d_s(X) \stackrel{def}{=} \delta_s(X;U_n)$.

Definition 1 (Min-entropy [30]). *The min-entropy of a random variable X is defined as:*

$$\mathrm{H}_\infty(X) = -\log \max_{x\in\mathcal{X}} \Pr[X=x] \ .$$

Definition 2 (Average min-entropy [31]). *The average min-entropy of X conditioned on L is defined as:*

$$\tilde{\mathrm{H}}_\infty(X|L) = -\log\left(\mathbf{E}_{l\leftarrow L}\Big[\max_{x\in\mathcal{X}} \Pr[X=x|L=l]\Big]\right) \ .$$

Definition 3 (HILL pseudoentropy [32]). *A random variable X has HILL pseudoentropy k, denoted by $\mathrm{H}^{\mathsf{HILL}}_{\epsilon,s}(X)\geq k$, if there exists a random variable Y such that $\mathrm{H}_\infty(Y)\geq k$ and $\delta_s(X;Y)\leq\epsilon$.*

Definition 4 (Conditional HILL pseudoentropy [33]). *For joint distribution (X,L), X has HILL pseudoentropy k conditioned on L, denoted by $\mathrm{H}^{\mathsf{HILL}}_{\epsilon,s}(X|L) \geq k$, if there exists a collection of distributions Y_l (giving rise to a joint distribution [Y,L]) such that $\tilde{\mathrm{H}}_\infty(Y|L)\geq k$ and $\delta_s([X,L];[Y,L])\leq\epsilon$.*

Definition 5 (Conditional unpredictability pseudoentropy [33]). *For joint distribution (X,L), X has unpredictability pseudoentropy k conditioned on L, denoted by $\mathrm{H}^{\mathsf{unp}}_s(X|L)\geq k$, if for all circuits A of size s it holds that $\Pr[\mathsf{A}(L)=X] \leq 2^{-k}$.*

Lemma 1 (Parity Lemma [34, 35]). *For any random variable X over $\{0,1\}^t$, it holds that*

$$\delta(X;U_t) \leq \sqrt{\sum_{v\in\{0,1\}^t\setminus\{0^t\}} \delta^2(\langle X,v\rangle;\ U_1)}$$

2.2 The LPN problem with auxiliary inputs

We show in Lemma 2 below that for any X with min-entropy k and $t \leq \ln 2\cdot\gamma^{-2}(k+2\log_2\epsilon+2)$, the distribution $[A,(A{\cdot}X+E)]$ is ϵ-close to uniform. Then, Theorem 1 weakens the assumption from min-entropy source X (*i.e.* worst-case randomness) to average min-entropy source (X,L) without incurring further overheads (see Remark 1), where for concreteness one can consider L as an arbitrary input conditioned on which X still has sufficient min-entropy in average.

Lemma 2 (Indistinguishability with weak random unknowns). *Let* $0 < \gamma < 1$ *and* $t \geq n$ *be functions of parameter* n*, let matrix distribution* A *be uniform over* $\{0,1\}^{t\times n}$*, let* X *be over* $\{0,1\}^n$ *with min-entropy* k*, let* E *be over* $\{0,1\}^t$ *with each bit independently distributed as in (1), then it holds that*

$$d([A,(A\cdot X+E)]) \leq \epsilon = 2^{-1-\frac{k}{2}}\sqrt{(1+\gamma^2)^t-1}$$

where ϵ *is upper bounded by* $2^{-1-\frac{k}{2}}\cdot \exp^{\frac{\gamma^2 t}{2}}$ *(see Fact 1 below).*

Proof. Denote the value that A assumes by a, which consists of row vectors $r_1, \cdots, r_t$. To use the parity lemma [34, 35] (see also Section 2.1), consider any non-zero $\boldsymbol{v}=[v_1, \cdots, v_t]\in \{0,1\}^t/\{0^t\}$ with Hamming weight $H_W(\boldsymbol{v})$, we have

$$\langle (a\cdot X+E), \boldsymbol{v}\rangle = \sum_{i=1}^{t} v_i\cdot(\langle r_i, X\rangle + e_i) \bmod 2 = \langle (\sum_{v_i=1}^{1\leq i\leq t} r_i), X\rangle + \sum_{v_i=1}^{1\leq i\leq t} e_i$$

where $\sum\limits_{v_i=1}^{1\leq i\leq t} e_i$ equals to 0 with probability $\gamma^{H_W(v)}$ and U_1 otherwise. Thus,

$$\begin{aligned}
&\sum_{a\in\{0,1\}^{t\times n}} \delta^2(\langle (a\cdot X+E), \boldsymbol{v}\rangle;\ U_1) \\
&= \gamma^{2H_W(v)}\cdot \sum_{a\in\{0,1\}^{t\times n}} \delta^2(\langle \sum_{v_i=1}^{1\leq i\leq t} r_i, X\rangle;\ U_1) \qquad (4)\\
&= \gamma^{2H_W(v)}\cdot 2^{(t-1)n}\cdot \sum_{y\in\{0,1\}^n} \delta^2(\langle y, X\rangle;\ U_1) \qquad (\text{ set } y \stackrel{def}{=} \sum_{v_i=1}^{1\leq i\leq t} r_i\ )
\end{aligned}$$

Then, using the technique in [36] it yields

$$\begin{aligned}
&\delta([A,(A\cdot X+E)];[A,U_t]) \\
&\leq \sum_{a\in\{0,1\}^{t\times n}} \Pr[A=a]\sqrt{\sum_{\boldsymbol{v}\neq 0^t} \delta^2(\langle (a\cdot X+E),\boldsymbol{v}\rangle;\ U_1)} \\
&\qquad\qquad \text{(by the Parity Lemma [34, 35])} \\
&\leq \sqrt{\sum_{\boldsymbol{v}\neq 0^t}\sum_{a\in\{0,1\}^{t\times n}} \Pr[A=a]\cdot \delta^2(\langle (a\cdot X+E),\boldsymbol{v}\rangle;\ U_1)} \\
&\qquad\qquad \text{(by Jensen's inequality, concavity case)} \\
&= \sqrt{2^{(t-1)n}\cdot 2^{-tn}\sum_{\boldsymbol{v}\neq 0^t}\gamma^{2H_W(v)}\sum_{y\in\{0,1\}^n}\delta^2(\langle y, X\rangle;\ U_1)} \qquad \text{(by (4) above)} \\
&= 2^{-1-\frac{k}{2}}\sqrt{\sum_{H_W(v)=1}^{t}\binom{t}{H_W(v)}\gamma^{2H_W(v)}} = 2^{-1-\frac{k}{2}}\sqrt{(1+\gamma^2)^t-1} \quad \text{(by Lemma 4)}
\end{aligned}$$

where Lemma 4 is given and proven in Appendix A. □

Theorem 1 (The LPN problem with auxiliary input). *For n, γ, t A and E as in Lemma 2, and for joint distribution (X,L) with $\tilde{\mathrm{H}}_\infty(X|L) \geq k$, it holds that*

$$\delta([A,(A\cdot X+E),L];[A,U_t,L]) \leq \epsilon = 2^{-1-\frac{k}{2}}\sqrt{(1+\gamma^2)^t-1}$$

Remark 1. Lemma 2 easily extends to Theorem 1 by the fact [31, Lemma 2.3] that any strong worst-case extractor is also an average-case extractor, but that incurs unnecessary costs, *i.e.*, we only obtain in Theorem 1 that the above is bounded by $\frac{1}{\xi}\cdot 2^{-1-\frac{k}{2}}\sqrt{(1+\gamma^2)^t-1}+\xi$ for any $\xi>0$. The proof for a better result is given in the appendix.

Fact 1 (Approximation of ϵ) *For $0<\gamma<1$ and t that are functions of n, it holds that*

$$(1+\gamma^2)^t-1 \begin{cases} < \exp^{\gamma^2 t}-1, \text{ for any } 0<\gamma<1 \\ \approx \gamma^2 t, \text{if } \gamma^2 t = o(1) \text{ with respect to } n \end{cases}$$

3 The Auxiliary-Input LPN Assumption

In the information-theoretic setting, we see that the LPN problem is robust against exponentially hard-to-invert auxiliary input (in the min-entropy sense), but in practice, one still needs the LPN assumption (see Section 1.2) so that with respect to polynomial-size adversaries the indistinguishability (see Equ (2)) holds for much relaxed parameter settings (*i.e.* constant γ and $t = poly(n)$). In this section, we show the equivalence among the decisional LPN assumption, the computational LPN assumption (see footnote 2), and the seemingly stronger auxiliary-input LPN assumption.

3.1 The equivalence of the two standard LPN assumptions

It is more natural to assume the computational LPN assumption as it directly relates to the hardness of the LPN problem, but in practice it is convenient to use the decisional case for building encryption schemes. We show their equivalence in the theorem below.

Theorem 2 (Equivalence of the two LPN assumptions). *The decisional and computational LPN assumptions are equivalent.*

Proof. ***On the one hand:*** Suppose that the decisional LPN assumption does not hold, *i.e.*, there exists a polynomial p and a family of distinguishers $\{D_n\}_{n\in\mathbb{N}}$ so that for infinitely many n's it holds that

$$\Pr[\mathsf{D}_n([A,A\cdot X+E])=1]-\Pr[\mathsf{D}_n([A,U_t])=1] \geq \frac{1}{p(n)}$$

For convenience, write A in t rows $[R_1,\cdots,R_t]$ and E in t bits $[E_1,\cdots,E_t]$, and hence write $[A,A\cdot X+E]$ as a t-tuple $[(R_1,\langle\ R_1,\ X\ \rangle+E_1),\cdots,(R_t,\langle\ R_t,\ X\ \rangle$

$+E_t)]$ and by a hybrid argument there exists a $0 \leq i \leq t-1$ so that conditioned on the first i-tuple D_n distinguishes $(R_{i+1}, \langle\ R_{i+1},\ X\ \rangle + E_{i+1})$ from $(R_{i+1},\ U_1)$ with probability at least $\frac{1}{t \cdot p(n)}$ (the rest are padded by uniform randomness), namely, given the first i-tuple D_n can predict $\langle\ R_{i+1},\ X\ \rangle$ with probability $\frac{2}{\gamma \cdot t \cdot p(n)} = \frac{1}{poly(n)}$ for randomly chosen R_{i+1}. Then, employ the Goldreich-Levin list-decoding technique one can efficiently recover X with a non-negligible probability, which contradicts the computational LPN assumption.

On the other hand: Suppose that the computational LPN assumption does not hold, *i.e.*, there exists $\{\mathsf{A}_n\}_{n \in \mathbb{N}}$ and a polynomial p so that

$$\Pr[\mathsf{A}_n([A, A \cdot X + E]) = X] \geq \frac{1}{p(n)}$$

holds for infinitely many n's. Then define a distinguisher D_n that on input $[A, Y]$ (either $Y = A \cdot X + E$ or $Y = U_t$), invokes A_n on $[A, Y]$ to get output X', and outputs 1 if and only if the Hamming weight $H_W(A \cdot X' + Y)$ is no more than $(\frac{1-(\gamma/2)}{2})t$. In case that Y is $A \cdot X + E$, let q_1 be the success probability of A_n conditioned on $H_W(E) \leq (\frac{1-(\gamma/2)}{2})t$, and q_2 be that on $H_W(E) > (\frac{1-(\gamma/2)}{2})t$. Then,

$$\Pr[H_W(E) \leq (\frac{1-(\gamma/2)}{2})t] \cdot q_1 + \Pr[H_W(E) > (\frac{1-(\gamma/2)}{2})t] \cdot q_2 \geq \frac{1}{p(n)}$$

By Hoeffding's inequality,

$$\Pr[H_W(E) > (\frac{1-(\gamma/2)}{2})t] = \Pr\left[H_W(E) - \mathbb{E}[H_W(E)] > \frac{\gamma t}{4}\right] \leq \exp(-\frac{t\gamma^2}{8})$$

and hence

$$\Pr[\mathsf{D}_n([A, A \cdot X + E]) = 1] \geq \Pr[H_W(E) \leq (\frac{1-(r/2)}{2})t] \cdot q_1 \geq \frac{1}{p(n)} - \exp(-\frac{t\gamma^2}{8}) \tag{5}$$

In the other case that Y is U_t, for any fixed value a (that A takes), the random variable $a \cdot X'_a$ assumes at most 2^n values. Take into account all the errors of up to $(\frac{1-(r/2)}{2})t$ bits, on any a and random variable U_t, distinguisher D_n outputs 1 for at most

$$\begin{aligned} & 2^n \cdot (\binom{t}{0} + \binom{t}{1} + \cdots + \binom{t}{\frac{1-(\gamma/2)}{2}t}) \\ & < 2^n \cdot 2^t \cdot (1 - \frac{\gamma^2}{18})^{\frac{t}{2}} \qquad \text{(see Fact 2 in Appendix)} \end{aligned}$$

values that U_t assumes, where $b \stackrel{def}{=} (1 - \frac{\gamma^2}{18}) < 1$ is a constant. Hence,

$$\Pr[\mathsf{D}_n(A, U_t) = 1] < \frac{2^n \cdot 2^t \cdot b^{\frac{t}{2}}}{2^t} < b^{\frac{t}{2} + n\log_b 2} \tag{6}$$

is negligible in n for constant γ and $t = poly(n)$. Therefore, (5)$-$(6) $> \frac{1}{2p(n)}$ for infinitely many n's, which contradicts the decisional LPN assumption. □

3.2 Types of leakage

Prior to proving the auxiliary-input LPN assumption, we generalize and extend the leakage type considered in [8] (see Conjecture 1). For X defined over $\{0,1\}^n$ and constant $0 < \alpha < 1$, we define arbitrary $2^{(1-\alpha)n}$-hard-to-invert leakage L in the following ways:

1. $\tilde{\mathrm{H}}_\infty(X|L) \geq (1-\alpha)n$ (see Definition 2).
2. $\mathrm{H}^{\mathsf{HILL}}_{\epsilon,s}(X|L) \geq (1-\alpha)n$ (see Definition 4).
3. $\mathrm{H}^{\mathsf{unp}}_{s}(X|L) \geq (1-\alpha)n$ (see Definition 5).

Any type of the above is more general than (and implied by) the one(s) preceding to it, but not vice versa [33]. Type 1 is the most common leakage assumed by most existing work (*e.g.* [25, 14, 26]) and type 2 is its computational analogue. Type 3 is similar to, but more general than, the setting of Conjecture 1 as it also incorporates leakages that are not efficiently computable. Therefore, by dealing with leakages of type 3 (which covers all other leakage types), we obtain indistinguishability by using encryption keys of zero min-entropy and HILL pseudoentropy but only with linear unpredictability pseudoentropy (see [33] for their relationships).

Static *vs.* adaptive leakages. In the auxiliary-input LPN assumption, L has to be independent of A and E, otherwise indistinguishability is not possible as one could ask for a leakage that corresponds to the first bit of $A \cdot X + E$. Thus, under the assumption the encryption scheme in Section 1.2 is CPA-secure *w.r.t.* static leakage. However, this does not rule out the possibility of building CCA-secure encryption scheme against (properly defined) adaptive leakages. We refer to [8] for the corresponding definition of CCA-secure symmetric encryption schemes in the adaptive auxiliary input setting, and the technique for building such schemes under the auxiliary-input LPN assumption and the existence of trapdoor permutations.

3.3 Combating leakages by sampling from subspace

A weakened version of the auxiliary-input LPN assumption. In this subsection, we show that the standard LPN assumption implies a weakened version of the auxiliary-input LPN assumption. That is, instead of sampling A from uniform, we sample the rows of A' from a random subspace of dimension βn, then Lemma 3 below states that $A'X+E$ is pseudorandom conditioned on A' and any leakage L with $\mathrm{H}^{\mathsf{unp}}_{s}(X|L) \geq (1-\alpha)n$, where α and β are arbitrary constants satisfying $\alpha+\beta<1$. Such a construction looks "artificial" as it has at most βn bits of security even in the leakage-free case, but it is an important step towards reducing the auxiliary-input LPN to the standard LPN assumption.

Sampling from random subspace. For a $t \times n$ matrix A', we can efficiently sample its rows from a randomly chosen subspace of dimension βn using uniform randomness, *e.g.*, by performing matrix multiplication $U_{t\times\beta n} \cdot U_{\beta n \times n}$ for uniform distributions $U_{t\times\beta n}$ and $U_{\beta n\times n}$. One can consider $U_{\beta n\times n}$ as the randomness to

sample the basis of a βn-dimensional subspace, and $U_{t\times\beta n}$ as the randomness to sample t vectors from that subspace. By linear algebra, the rank of the product of two matrices is less than or equal to the minimum of the rank of each factor, so A' has rank up to βn.

Lemma 3 (Preliminary results). *Under the LPN assumption, let t be polynomial in n, and let α (the portion of leakage), β (βn quantifies the remaining security) and γ be any constants satisfying $\alpha+\beta< 1$ and $0 < \gamma < 1$, let the rows of A' be sampled from a random βn-dimensional subspace, let joint distribution $[X, L]$ satisfy either (a) $\tilde{\mathrm{H}}_\infty(X|L) \geq (1-\alpha)n$, or (b) $\mathrm{H}^{\mathsf{HILL}}_{\epsilon,s}(X|L)\geq(1-\alpha)n$, or (c) $\mathrm{H}^{\mathsf{unp}}_{s}(X|L) \geq(1-\alpha)n$, then the following ensembles are computationally indistinguishable:*

$$\{A', (A'\cdot X + E), L\}_{n\in\mathbb{N}} \overset{c}{\approx} \{A', U_t, L\}_{n\in\mathbb{N}} \tag{7}$$

Proof. By Claim 1 below, we have

$$\{U_{\beta n\times n}, (U_{\beta n\times n}\cdot X), L\}_{n\in\mathbb{N}} \overset{c}{\approx} \{U_{\beta n\times n}, U_{\beta n}, L\}_{n\in\mathbb{N}}$$

holds for all the 3 types of leakages (with different overheads). By left-multiplying the above with $U_{t\times\beta n}$ (note that $A' \overset{def}{=} U_{t\times\beta n}\cdot U_{\beta n\times n}$), it yields

$$\{A', (A'\cdot X + E), L\}_{n\in\mathbb{N}} \overset{c}{\approx} \{A', (U_{t\times\beta n}\cdot U_{\beta n} + E), L\}_{n\in\mathbb{N}}$$

Then, by the LPN assumption (note that L is independent of $U_{\beta n}$), it holds that

$$\{A', (U_{t\times\beta n}\cdot U_{\beta n} + E), L\}_{n\in\mathbb{N}} \overset{c}{\approx} \{A', U_t, L\}_{n\in\mathbb{N}}$$

which completes the proof. □

Claim 1 (Extractor for all conditional pseudoentropy sources)

$$\delta_{s'}(\ [U_{\beta n\times n}, (U_{\beta n\times n}\cdot X), L];\ [U_{\beta n\times n}, U_{\beta n}, L]\ ) \leq \epsilon' \tag{8}$$

where constant $c \overset{def}{=} (1-\alpha-\beta)$ and

$$(s', \epsilon') = \begin{cases} (\infty, 2^{-1-\frac{c\cdot n}{2}}) & \text{for } \tilde{\mathrm{H}}_\infty(X|L) \geq (1-\alpha)n; \\ (s - O(n^2), \epsilon + 2^{-1-\frac{c\cdot n}{2}}) & \text{for } \mathrm{H}^{\mathsf{HILL}}_{\epsilon,s}(X|L)\geq(1-\alpha)n; \\ (\Omega(\frac{\epsilon^2}{n^2}\cdot s), O(\beta n\cdot\sqrt[3]{n^2\epsilon})) & \text{for } \mathrm{H}^{\mathsf{unp}}_{s}(X|L)\geq(1-\alpha)n \text{ and } \epsilon \geq 2^{-c\cdot n}. \end{cases}$$

Proof. The min-entropy case is implied by Theorem 2 by setting $k = (1-\alpha)n$, $\gamma = 1$ (noise-free), and $t = \beta n$. The HILL pseudoentropy case then follows by Definition 4 and a triangle inequality. The unpredictability pseudoentropy case follows from the Goldreich-Levin Theorem [37, 35] and a hybrid argument (for any $0\leq i \leq \beta n$ it holds that $\mathrm{H}^{\mathsf{unp}}_{s-O(n)}(X|L, U_{i\times n}\cdot X)\geq(1-\alpha)n - i$). We note that the Goldreich-Levin Theorem (see the general version in [38, Lemma 9] quantitatively) does not require L to be efficiently computable from X (although it improves the efficiency of list-decoding by a polynomial factor). □

3.4 The auxiliary-input LPN assumption

We now define a distribution $F_{\beta n\times n}$ that is uniform over all full-rank $\beta n\times n$ Boolean matrices. It is not hard to see that Lemma 3 and Claim 1 still hold if $U_{\beta n\times n}$ is replaced by $F_{\beta n\times n}$ (*i.e.* $A' = U_{t\times\beta n}\cdot F_{\beta n\times n}$). This is because in Claim 1, any $u_{\beta n\times n}\in\mathcal{U}_{\beta n\times n}$ but $\notin\mathcal{F}_{\beta n\times n}$ is a "bad" value conditioned on which adversaries can efficiently distinguish (8) with advantage 1/2 (by exploiting the linear dependency in $u_{\beta n\times n}$).

Theorem 3 (The equivalence of the LPN assumption and the auxiliary-input LPN assumption [8]). *Under the LPN assumption, for every constant $0<\gamma<1$, and for every polynomial $t = poly(n)$, let A be uniform distribution over $\{0,1\}^{t\times n}$, let joint distribution $[X,L]$ satisfy either (a) $\tilde{\mathrm{H}}_\infty(X|L)\geq(1-\alpha)n$, or (b) $\mathrm{H}^{\mathsf{HILL}}_{\epsilon,s}(X|L)\geq(1-\alpha)n$, or (c) $\mathrm{H}^{\mathsf{unp}}_{s}(X|L)\geq(1-\alpha)n$ for any constant $0<\alpha<1$, then the following ensembles are computationally indistinguishable:*

$$\{A,(A\cdot X+E),L\}_{n\in\mathbb{N}} \overset{c}{\approx} \{A,U_t,L\}_{n\in\mathbb{N}} \tag{9}$$

Proof. We recall that $A' = U_{t\times\beta n}\cdot F_{\beta n\times n}$ and $A = U_{t\times n} = U_{t\times n}\cdot F_{n\times n}$. By comparing the above with the statement in Lemma 3, it suffices to prove the claim below. □

Claim 2 *For any $\beta n\leq i<n$, if*

$$\delta_s(\ [A_i\overset{def}{=}(U_{t\times i}\cdot F_{i\times n}),A_i\cdot X+E,L];\ \ [A_i,U_t,L]\ )\ \leq\ \epsilon \tag{10}$$

then it holds that

$$\delta_{s-O(n^3)}(\ [A_{i+1}\overset{def}{=}(U_{t\times(i+1)}\cdot F_{(i+1)\times n}),A_{i+1}\cdot X+E,L];\ \ [A_{i+1},U_t,L]\ )\leq\epsilon+2^{-\Omega(n)}$$

Proof (of Claim 2). For convenience, write

$$U_{t\times(i+1)}\overset{def}{=}[U_{t\times i},C_{i+1}],F_{i\times n}\overset{def}{=}\begin{bmatrix}B_1\\ \vdots\\ B_i\end{bmatrix},F_{(i+1)\times n}\overset{def}{=}\begin{bmatrix}F_{i\times n}\\ B_{i+1}\end{bmatrix}$$

where C_{i+1} is the $(i+1)^{th}$ column of $U_{t\times(i+1)}$, and the B's are the basis that generates the rows of A_i (and A_{i+1}). For $B_1,\cdots,B_i$ and A_{i+1}, define an algorithm Alg that takes them as input, and deterministically [4] outputs $B'_{i+1}\overset{def}{=}\mathsf{Alg}(B_1,\cdots,B_i,A_{i+1})$ so that $B_1,\cdots,B_i,B'_{i+1}$ is a basis that generates the row vectors of A_{i+1}. Then, the following two distributions are identical:

$$[A_{i+1},C_{i+1},B_1,\cdots,B_i,B_{i+1}]\sim[A_{i+1},C_{i+1},B_1,\cdots,B_i,B'_{i+1}+\sum_{1\leq j\leq i}U_1^{(j)}\cdot B_j]$$

[4] Due to lack of $U_{t\times i}$, the outputs can be far from unique, we however require Alg to be deterministic to emphasize that B'_{i+1} is determined given $B_1,\cdots,B_i$ and A_{i+1}.

due to that $U_{t\times i}$ is not being conditioned on, where $U_1^{(j)}$ is uniform over {0,1} and $U_1^{(j)}\cdot B_j$ is scalar-vector multiplication. Then, as $A_{i+1}\cdot X = A_i\cdot X+\langle B_{i+1}, X\rangle\cdot C_{i+1}$, we have distribution $[A_{i+1}, A_{i+1}\cdot X + E, B_1, \cdots, B_i, C_{i+1}, X, L]$ is identical to

$$[\ A_{i+1}, A_i\cdot X + E + \langle (B'_{i+1} + \sum_{1\le j\le i} U_1^{(j)}\cdot B_j), X\rangle \cdot C_{i+1}, B_1, \cdots, B_i, C_{i+1}, X, L\ ]$$

where $\langle B'_{i+1}, X\rangle$ is determined given $B_1, \cdots, B_i$, A_{i+1} and X. Thus, $\langle B'_{i+1}, X\rangle + \sum_{1\le j\le i} U_1^{(j)}\cdot\langle B_j, X\rangle$ is identical to U_1 if $[\langle B_1, X\rangle, \cdots, \langle B_i, X\rangle]$ is not all zero, which has probability

$$\Pr_{B_1,\cdots,B_i,X}\left[\ [\langle B_1, X\rangle, \cdots, \langle B_i, X\rangle] \neq 0^i\ \right] = \Pr\left[U_i \neq 0^i\right] = 1 - 2^{-i} = 1 - 2^{-\Omega(n)}$$

Thus, we have

$$\delta([A_{i+1}, A_{i+1}\cdot X + E, L]; [A_{i+1}, A_i\cdot X + E + U_1\cdot C_{i+1}, L]) \le 2^{-\Omega(n)} \qquad (11)$$

and by (10), it holds that

$$\delta_{s-O(n^3)}(\ [A_{i+1}, A_i\cdot X + E + U_1\cdot C_{i+1}, L];\ [A_{i+1}, U_t, L]\ ) \le \epsilon \qquad (12)$$

as $A_{i+1} = A_i + C_{i+1}\cdot B_{i+1}$, and B_{i+1} can be efficiently sampled given A_i. The conclusion follows from (11) and (12) by a triangle inequality. □

4 Concluding Remarks

We answer the open question in [8] whether the auxiliary-input LPN assumption is implied by (and thus equivalent to) the standard LPN assumption. In addition, they are equivalent for any constant noise rate.

References

1. Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Transactions on Information Theory **24** (1978) 384–386
2. Håstad, J.: Some optimal inapproximability results. J. ACM **48** (2001) 798–859
3. Blum, A., Kalai, A., Wasserman, H.: Noise-tolerant learning, the parity problem, and the statistical query model. J. ACM **50** (2003) 2003
4. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: STOC 2005. (2005) 84–93
5. Lenstra, A., H.W. Lenstra, J., Lovász, L.: Factoring polynomials with rational coefficients. Mathematische Annalen **261**(4) (1982) 515–534
6. Schnorr, C.P.: A hierarchy of polynomial time lattice basis reduction algorithms. Theor. Comput. Sci. **53**(2-3) (1987) 201–224
7. Ajtai, M., Kumar, R., Sivakumar, D.: A sieve algorithm for the shortest lattice vector problem. In: STOC 2001. (2001) 601–610

8. Dodis, Y., Kalai, Y.T., Lovett, S.: On cryptography with auxiliary input. In: STOC 2009. (2009) 621–630
9. Gilbert, H., Robshaw, M.J.B., Seurin, Y.: How to encrypt with the lpn problem. In: ICALP (2). (2008) 679–690
10. Applebaum, B., Cash, D., Peikert, C., Sahai, A.: Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In: CRYPTO 2009. (2009) 595–618
11. Hopper, N.J., Blum, M.: Secure human identification protocols. In: ASIACRYPT. (2001) 52–66
12. Juels, A., Weis, S.A.: Authenticating pervasive devices with human protocols. In: CRYPTO. (2005) 293–308
13. Gilbert, H., Robshaw, M.J.B., Seurin, Y.: $\mathrm{Hb}^{\#}$: Increasing the security and efficiency of hb^{+}. In: EUROCRYPT. (2008) 361–378
14. Akavia, A., Goldwasser, S., Vaikuntanathan, V.: Simultaneous hardcore bits and cryptography against memory attacks. In: TCC 2009. Volume 5444. (2009) 474–495
15. Canetti, R.: Towards realizing random oracles: Hash functions that hide all partial information. In: CRYPTO 1997, Springer-Verlag (1997) 455–469
16. Canetti, R., Dodis, Y., Halevi, S., Kushilevitz, E., Sahai, A.: Exposure-resilient functions and all-or-nothing transforms. In: EUROCRYPT 2000. (2000) 453–469
17. Dodis, Y., Sahai, A., Smith, A.: On perfect and adaptive security in exposure-resilient cryptography. In: EUROCRYPT 2001. (2001) 301–324
18. Goldwasser, S., Kalai, Y.T.: On the impossibility of obfuscation with auxiliary input. In: FOCS 2005. (2005) 553–562
19. Kocher, P., E, J.J., Jun, B.: Differential power analysis. In: CRYPTO 1999, Springer-Verlag (1999) 388–397
20. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: Securing hardware against probing attacks. In: CRYPTO 2003. (2003) 463–481
21. Ishai, Y., Prabhakaran, M., Sahai, A., Wagner, D.: Private circuits ii: Keeping secrets in tamperable circuits. In: EUROCRYPT. (2006) 308–327
22. Micali, S., Reyzin, L.: Physically observable cryptography. In: TCC 2004, LNCS 2951, Springer (2004) 278–296
23. Petit, C., Standaert, F.X., Pereira, O., Malkin, T., Yung, M.: A block cipher based pseudo random number generator secure against side-channel key recovery. In: ASIACCS. (2008) 56–65
24. Standaert, F.X., Pereira, O., Yu, Y., Quisquater, J.J., Yung, M., Oswald, E.: Leakage resilient cryptography in practice. Cryptology ePrint Archive, Report 2009/341 (2009) `http://eprint.iacr.org/`.
25. Dziembowski, S., Pietrzak, K.: Leakage-resilient cryptography. In: FOCS, IEEE Computer Society (2008) 293–302
26. Pietrzak, K.: A leakage-resilient mode of operation. In: Eurocrypt 2009. (2009) 462–482
27. Naor, M., Segev, G.: Public-key cryptosystems resilient to key leakage. In: CRYPTO. (2009) 18–35
28. Joël Alwen, Y.D., Wichs, D.: Leakage-resilient public-key cryptography in the bounded-retrieval model. In: CRYPTO. (2009) 36–54
29. Katz, J., Vaikuntanathan, V.: Signature schemes with bounded leakage resilience. In: ASIACRYPT (to appear). (2009)
30. Zuckerman, D.: General weak random sources. In: FOCS 1990. (1990) 534–543
31. Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. SIAM J. Comput. **38**(1) (2008) 97–139

32. Hastad, J., Impagliazzo, R., Levin, L.A., Luby, M.: A pseudorandom generator from any one-way function. SIAM Journal on Computing **28** (1999) 12–24
33. Hsiao, C.Y., Lu, C.J., Reyzin, L.: Conditional computational entropy, or toward separating pseudoentropy from compressibility. In: EUROCRYPT. (2007) 169–186
34. Vazirani, U.: Strong communication complexity or generating quasi-random sequences from two communicating semi-random sources. Combinatorica **7**(4) (1987) 375–392
35. Goldreich, O.: Three XOR-Lemmas — An Exposition. Electronic Colloquium on Computational Complexity (ECCC) **2** (1997)
36. Dodis, Y., Elbaz, A., Oliveira, R., Raz, R.: Improved randomness extraction from two independent sources. In: RANDOM-APPROX. (2004) 334–344
37. Goldreich, O., Levin, L.: A hard-core predicate for all one-way functions. In: STOC 1989. (1989) 25–32
38. Goldreich, O., Nisan, N., Wigderson, A.: On yao's xor-lemma. Technical report, Electronic Colloquium on Computational Complexity (1998)
39. Babai, L., Frankl, P., Simon, J.: Complexity classes in communication complexity theory. In: FOCS. (1986) 337–347
40. Worsch, T.: Lower and upper bounds for (sums of) binomial coefficients (1994)

A Lemmata and Proofs Omitted from the Main Body

Lemma 4. *For n-bit random variable X with $H_\infty(X) = k$, it holds that*

$$\sum_{y\in\{0,1\}^n} \delta^2(\langle X, y\rangle; U_1) \le 2^{n-k-2}$$

Proof. Following the proof of the Lindsey Lemma [39], there exists a $2^n \times 2^n$ Hadamard matrix (which can be constructed recursively using Sylvester's technique) $H = [h_{ij}]_{2^n\times 2^n}$ (i.e. a ± 1 matrix with pairwise orthogonal rows and columns indexed by vectors over $\{0,1\}^n$) and the inner product $\langle x, y\rangle \pmod 2$ can be considered as identifying the element h_{xy} on x-th row and y-th column and producing $\frac{h_{xy}+1}{2}$ as output. Let $\boldsymbol{R}_x$ be the row indexed by x and we recall that by orthogonality the Euclidean inner product $\boldsymbol{R}_x^T \boldsymbol{R}_{x'}$ over $\mathcal{R}$ is 2^n if $x = x'$, and 0 otherwise. Define vector $\boldsymbol{R} \in \mathcal{R}^{2^n}$ as $\boldsymbol{R} \stackrel{def}{=} \frac{1}{2}\Sigma_x \Pr[X = x] \cdot \boldsymbol{R}_x$, then we see that the y-th coordinate of $\boldsymbol{R}$, denoted by r_y, is

$$\begin{aligned} r_y &= \frac{1}{2}\Big(\sum\nolimits_x \Pr[X = x, h_{xy} = 1] - \sum\nolimits_x \Pr[X = x, h_{xy} = -1]\Big) \\ &= \delta(\langle X \cdot y\rangle; U_1) \end{aligned}$$

Then we have

$$\begin{aligned} &\sum_{y\in\{0,1\}^n} \delta^2(\langle X \cdot y\rangle; U_1) = \sum_{y\in\{0,1\}^n} r_y^2 = \boldsymbol{R}^T \cdot \boldsymbol{R} \\ &= \frac{1}{4}\sum\nolimits_x (\Pr[X = x])^2 \boldsymbol{R}_x^T \cdot \boldsymbol{R}_x + \frac{1}{4}\sum\nolimits_{x\neq x'} \Pr[X = x]\Pr[X = x'] \boldsymbol{R}_x^T \cdot \boldsymbol{R}_{x'} \\ &= \frac{1}{4} 2^n \sum\nolimits_x (\Pr[X = x])^2 \qquad \text{(note that } \Pr[X = x] \le 2^{-k} \text{ for any } x) \\ &\le 2^{n-2} \cdot 2^{-k} \sum\nolimits_x \Pr[X = x] \quad = 2^{n-k-2} \end{aligned}$$

which finishes the proof. □

Proof (of Fact 1).
In case that $\gamma^2 t = o(1)$, it converges to zero for sufficiently large n's and thus

$$\begin{aligned}
&(1+\gamma^2)^t - 1 \\
&= t\gamma^2 + \frac{t(t-1)}{1\times 2}\gamma^4 + \frac{t(t-1)(t-2)}{1\times 2\times 3}\gamma^6 + \cdots + \frac{t(t-1)\cdots 1}{1\times 2\times\cdots\times t}\gamma^{2t} \\
&= t\gamma^2(1 + \frac{(t-1)\gamma^2}{1\times 2} + \frac{(t-1)(t-2)\gamma^4}{1\times 2\times 3} + \cdots + \frac{(t-1)\cdots 1\cdot\gamma^{2(t-1)}}{1\times 2\times\cdots\times t}) \\
&< t\gamma^2(1 + \frac{t\gamma^2}{1\times 2} + \frac{t^2\gamma^4}{1\times 2\times 3} + \cdots + \frac{t^{t-1}\gamma^{2(t-1)}}{1\times 2\times\cdots\times t}) \approx t\gamma^2
\end{aligned}$$

In a general case, to show that $(1+\gamma^2)^t - 1 < exp^{\gamma^2 t} - 1$, it suffices to have the following (using the technique above):

$$\begin{aligned}
&(1+\gamma^2)^{\gamma^{-2}} \\
&< 1 + \gamma^{-2}\gamma^2(1 + \frac{\gamma^{-2}\gamma^2}{1\times 2} + \frac{\gamma^{-4}\gamma^4}{1\times 2\times 3} + \cdots + \frac{\gamma^{-(2\gamma^{-2}-2)}\gamma^{2\gamma^{-2}-2}}{1\times 2\times\cdots\times\gamma^{-2}}) \\
&= 1 + 1 + \frac{1}{1\times 2} + \frac{1}{1\times 2\times 3} + \cdots + \frac{1}{1\times 2\times\cdots\times\gamma^{-1}} \\
&< \exp
\end{aligned}$$

where by Taylor series the mathematical constant exp is the sum of the infinite series: $\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots$, which completes the proof. □

Proof (of Theorem 1). Denote $\max_x \Pr[X = x|L = l]$ by 2^{-k^l}, we have

$$\begin{aligned}
&\delta([A, (A\cdot X + E), L]; [A, U_t, L]) \\
&\leq \sum_l \Pr[L = l] \cdot 2^{-1-\frac{k^l}{2}} \cdot \sqrt{(1+\gamma^2)^t - 1} \quad \text{(by Lemma 1)} \\
&= 2^{-1} \cdot \sqrt{(1+\gamma^2)^t - 1} \cdot \sum_l \sqrt{\Pr[L = l]} \cdot \sqrt{\Pr[L = l]} \cdot 2^{-\frac{k^l}{2}} \\
&\leq 2^{-1} \cdot \sqrt{(1+\gamma^2)^t - 1} \cdot \sqrt{\left(\sum_l \Pr[L = l]\right) \cdot \left(\sum_l \Pr[L = l] \cdot 2^{-k^l}\right)} \\
&\qquad\qquad \text{(by Cauchy's inequality)} \\
&= 2^{-1} \cdot \sqrt{(1+\gamma^2)^t - 1} \cdot \sqrt{\sum_l \Pr[L = l] \cdot 2^{-k^l}} \\
&= 2^{-1} \cdot \sqrt{(1+\gamma^2)^t - 1} \cdot \sqrt{2^{-\tilde{H}_\infty(X|L)}} \leq 2^{-1-\frac{k}{2}} \cdot \sqrt{(1+\gamma^2)^t - 1}
\end{aligned}$$

which completes the proof. □

Fact 2 *For $0 < \gamma < 1$ and $t > 0$ it holds that*

$$\binom{t}{0} + \binom{t}{1} + \cdots + \binom{t}{\frac{1-(\gamma/2)}{2}t} < 2^t \cdot (1 - \frac{\gamma^2}{18})^{\frac{t}{2}}$$

Proof. By the upper bound on the partial sum of binomial coefficients [40],

$$\begin{aligned}
&\binom{t}{0} + \binom{t}{1} + \cdots + \binom{t}{\frac{1-(\gamma/2)}{2}t} \\
&\leq 2^t \cdot ((1 - \frac{\gamma}{2+\gamma}) \cdot \exp^{\frac{\gamma}{2+\gamma}})^{\frac{t(2+\gamma)}{4}} \qquad \text{(see [40, Corollary 4.2])}
\end{aligned}$$

Let $b \stackrel{def}{=} (1 - \frac{\gamma}{2+\gamma}) \cdot \exp^{\frac{\gamma}{2+\gamma}} < 1$. We have

$$\begin{aligned}
b &= (1 - \frac{\gamma}{2+\gamma}) \cdot (1 + \frac{\gamma}{2+\gamma} + \frac{1}{2!}(\frac{\gamma}{2+\gamma})^2 + \frac{1}{3!}(\frac{\gamma}{2+\gamma})^3 + \cdots) \\
&< 1 - \frac{1}{2}(\frac{\gamma}{2+\gamma})^2 + (\frac{1}{3!} - \frac{1}{2!})(\frac{\gamma}{2+\gamma})^3 + (\frac{1}{4!} - \frac{1}{3!})(\frac{\gamma}{2+\gamma})^4 + \cdots \\
&< 1 - \frac{1}{2}(\frac{\gamma}{2+\gamma})^2 < 1 - \frac{1}{2}(\frac{\gamma}{3})^2 < 1 - \frac{\gamma^2}{18}
\end{aligned}$$

Hence $0 < b < 1$ and it follows that $b^{\frac{t(2+\gamma)}{4}} < b^{\frac{t}{2}}$, which completes the proof. □