6.889 New Developments in Cryptography March 8, 2011

Extractors and the Leftover Hash Lemma

Instructors: Shafi Goldwasser, Yael Kalai, Leo Reyzin, Boaz Barak, and Salil Vadhan

Lecturer: Leo Reyzin *Scribe: Thomas Steinke, David Wilson*

1 Statistical Distance

We need random keys for cryptography. Ideally, we want to be able to sample uniformly random and independent bits. However, in practice we have imperfect sources of randomness. We would still like to be able to say that keys that are ε-close to uniform are "good enough". This requires the following definition. The material in the first three sections is based on [V11].

Definition 1 *Let X and Y be two random variables with range U. Then the* statistical distance *between X and Y is defined as*

$$\Delta(X,Y) \equiv \frac{1}{2}\sum_{u\in U}\left|\mathbb{P}\left[X=u\right]-\mathbb{P}\left[Y=u\right]\right|.$$

*For $\varepsilon \geq 0$, we also define the notion of two distributions being ε-*close*:*

$$X \approx_\varepsilon Y \iff \Delta(X,Y) \leq \varepsilon.$$

The statistical distance satisfies the following nice properties.

Lemma 2 *Let X and Y be random variables with range U.*

(i) $\Delta(X,Y) = \max_{T\subset U}(\mathbb{P}\left[X\in T\right]-\mathbb{P}\left[Y\in T\right])$.

(ii) For every $T \subset U$, $\mathbb{P}\left[Y\in T\right] \leq \mathbb{P}\left[X\in T\right]+\Delta(X,Y)$.

(iii) $\Delta(X,Y)$ is a metric on distributions.

(iv) For every randomized function F, $\Delta(F(X),F(Y)) \leq \Delta(X,Y)$. Moreover, we have equality if each realisation f of F is one-to-one.

Property (ii) is particularly useful: Suppose that T is the set of random inputs that cause an algorithm to fail and the failure probability on a truly random input X is $\mathbb{P}\left[X\in T\right]$. If Y is an imperfect random input, then the failure probability when using Y instead of X is $\mathbb{P}\left[Y\in T\right] \leq \mathbb{P}\left[X\in T\right]+\Delta(X,Y)$.

Property (iv) is also very powerful and we will make use of it later. It allows us to reason about the combination of multiple random variables. For example, it allows us to prove that, for any random variables X, X', Y, Y',

$$\max\left\{\Delta(X,X'),\Delta(Y,Y')\right\} \leq \Delta((X,Y),(X',Y')) \leq \Delta(X,X')+\Delta(Y,Y').$$

To prove this, for the left-hand side we take the deterministic projections $f_1(x,y) = x$ and $f_2(x,y) = y$:

$$\Delta(X,X') = \Delta(f_1(X,Y), f_1(X',Y')) \leq \Delta((X,Y),(X',Y'));$$
$$\Delta(Y,Y') = \Delta(f_2(X,Y), f_2(X',Y')) \leq \Delta((X,Y),(X',Y')).$$

For the right-hand side we use the triangle inequality (from part (iii)) and the random functions $F_1(y) = (X,y)$ and $F_2(x) = (x,Y')$:

$$\begin{aligned}\Delta((X,Y),(X',Y')) &\leq \Delta((X,Y),(X,Y')) + \Delta((X,Y'),(X',Y')) \\ &= \Delta(F_1(Y), F_1(Y')) + \Delta(F_2(X), F_2(X')) \\ &\leq \Delta(Y,Y') + \Delta(X,X').\end{aligned}$$

Proof of Lemma 2: Let X and Y be random variables with range U.

(i) Let

$$T = \{u \in U : \mathbb{P}[X=u] - \mathbb{P}[Y=u] > 0\}.$$

Clearly T maximizes $\mathbb{P}[X \in T] - \mathbb{P}[Y \in T]$: We have $\mathbb{P}[X \in T] - \mathbb{P}[Y \in T] = \sum_{u \in T} \mathbb{P}[X=u] - \mathbb{P}[Y=u]$ and the sum is obviously maximized by selecting the positive terms. Now

$$\begin{aligned}2\Delta(X,Y) &= \sum_{u \in U} |\mathbb{P}[X=u] - \mathbb{P}[Y=u]| \\ &= \sum_{u \in T} (\mathbb{P}[X=u] - \mathbb{P}[Y=u]) + \sum_{u \in U \backslash T} -(\mathbb{P}[X=u] - \mathbb{P}[Y=u]) \\ &= \mathbb{P}[X \in T] - \mathbb{P}[Y \in T] - \mathbb{P}[X \in U \backslash T] + \mathbb{P}[Y \in U \backslash T] \\ &= \mathbb{P}[X \in T] - \mathbb{P}[Y \in T] - (1 - \mathbb{P}[X \in T]) + (1 - \mathbb{P}[Y \in T]) \\ &= 2(\mathbb{P}[X \in T] - \mathbb{P}[Y \in T]) \\ &= 2 \max_{T' \subset U} (\mathbb{P}[X \in T'] - \mathbb{P}[Y \in T']).\end{aligned}$$

(iv for deterministic functions) Fix $f : U \to V$. Then

$$\begin{aligned}\Delta(f(X), f(Y)) &= \max_{S \subset V} \mathbb{P}[f(X) \in S] - \mathbb{P}[f(Y) \in S] \\ &= \max_{S \subset V} \mathbb{P}\left[X \in f^{-1}(S)\right] - \mathbb{P}\left[Y \in f^{-1}(S)\right] \\ &\leq \max_{T \subset U} \mathbb{P}[X \in T] - \mathbb{P}[Y \in T] \\ &= \Delta(X,Y).\end{aligned}$$

Note that, if f is one-to one, then we have equality, as $f^{-1}(f(T)) = T$.

(ii) Fix $T \subset U$. Then

$$\mathbb{P}[Y \in T] - \mathbb{P}[X \in T] \leq \max_{T' \subset U} \mathbb{P}[Y \in T'] - \mathbb{P}[X \in T'] = \Delta(X,Y).$$

(iii) Note that $\Delta(X,Y)$ is simply half the L_1-distance between the distribution vectors of X and Y;[1] it is therefore clear that we have a metric, as the L_1-norm induces a metric.[2]

[1] The distribution vector π of a random variable X is given by $\pi_i = \mathbb{P}[X = u_i]$, where $U = \{u_i : i\}$.

[2] Note that L_1 refers to the norm where $||x||_1 = \sum_i |x_i|$. Also L_2 refers to the norm with $||x||_2 = \sqrt{\sum_i x_i^2}$.

(iv) Let F be a random variable over functions $f : U \to V$ for some fixed V and let $\mathcal{F}$ be the range of F. Then

$$\begin{aligned}
2\Delta(F(X), F(Y)) &= \sum_{v \in V} |\mathbb{P}[F(X) = v] - \mathbb{P}[F(Y) = v]| \\
&= \sum_{f \in \mathcal{F}} \mathbb{P}[F = f] \sum_{v \in V} |\mathbb{P}[F(X) = v | F = f] - \mathbb{P}[F(Y) = v | F = f]| \\
&= \sum_{f \in \mathcal{F}} \mathbb{P}[F = f]\, 2\Delta(f(X), f(Y)) \\
&\leq \sum_{f \in \mathcal{F}} \mathbb{P}[F = f]\, 2\Delta(X, Y) \qquad \text{(because } f \text{ is deterministic)} \\
&= 2\Delta(X, Y).
\end{aligned}$$

Note that, if each $f \in \mathcal{F}$ is one-to-one, then the above gives equality. □

2 Randomness Extractors

Our goal is to extract good random bits from bad ones. Our input X is a random variable that we know very little about. We assume that we do not know its exact distribution; we only know that it comes from some class $\mathcal{C}$ of distributions. We would like $\mathcal{C}$ to be as general as possible, while still being able to extract good random bits. This leads to the following requirement.

Definition 3 *Let X be a random variable. The* min-entropy *of X is defined as*

$$H_\infty(X) \equiv \min_{u \in U} \{-\log(\mathbb{P}[X = u])\} = -\log\left(\max_{u \in U} \mathbb{P}[X = u]\right).$$

*We say that X is a k-*source *if $H_\infty(X) \geq k$.*

We demand that X has a high min-entropy. To see that this requirement is necessary, consider the following distribution for X with low min-entropy.

$$\mathbb{P}[X = 0] = \frac{1}{2}, \quad \forall 1 \leq l \leq L\ \mathbb{P}[X = l] = \frac{1}{2L}.$$

Then X has min-entropy 1 because of the single high-probability event. Indeed, what can the extractor do with this input? Half the time it receives $X = 0$, which bears effectively no information. The extractor will fail with probability $1/2$. Note that the Shannon entropy of X is $1 + \log(L)/2$.

We have identified an appropriate class of distributions—those with high min-entropy. Now we must define what an extractor is.

Here is our first attempt: A *deterministic ε-extractor* is a function $f : U \to \{0,1\}^n$ such that, for all k-sources X,

$$f(X) \approx_\varepsilon U_n,$$

where U_n is a uniformly distributed random variable over n bits. However, this definition is vacuous; we cannot obtain even one bit: Suppose that f is a deterministic 0.499-extractor. Without loss of generality, assume $|f^{-1}(0)| \geq |f^{-1}(1)|$. Let X be uniformly distributed on $f^{-1}(0)$. Then $H_\infty(X) =$

$\log(|f^{-1}(0)|) \geq \log(|U|/2)$, so X has high min-entropy. But $f(X) = 0$ is deterministic, whence $\Delta(f(X), U_1) = 1/2$—a contradiction.

We conclude that an extractor must itself be a random function. This is because we do not know the distribution of X, only its class. And X may be chosen adversarially from this class.

Definition 4 *Let the seed U_d be uniformly distributed on $\{0,1\}^d$. We say that a function $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, ε)* (strong) extractor *if, for all random variables X on $\{0,1\}^n$ independent of U_d with $H_\infty(X) \geq k$,*

$$(Ext(X, U_d), U_d) \approx_\varepsilon (U_m, U_d),$$

where U_m is uniformly distributed on $\{0,1\}^m$ independent of X and U_d.

Having randomized randomness extractors seems like cheating. However, we demand that the output of the extractor is independent of the seed. This means that the seed can be 'recycled' and, in a cryptographic setting, the seed can be made public without compromizing the output—our key.

3 The Leftover Hash Lemma

The leftover hash lemma (LHL) shows how to construct extractors from universal hash families, which are defined as follows.

Definition 5 *A keyed hash function or, equivalently, a family $\mathcal{H}$ of hash functions of size 2^d from $\{0,1\}^n$ to $\{0,1\}^m$ is called* universal *if, for every $x, y \in \{0,1\}^n$ with $x \neq y$,*

$$\mathbb{P}_{h \in \mathcal{H}}[h(x) = h(y)] \leq 2^{-m}.$$

Now we give an example of a universal hash family for $d = n$. Consider strings in $\{0,1\}^n$ as elements of the field $GF(2^n)$. Let $h_s(x) = sx$ with the output truncated to m bits. (Any truncation is fine.) Then this is a universal hash family. Note that this construction requires a long seed.

Theorem 6 (Leftover Hash Lemma (LHL) [ILL89]) *Let X be a random variable with universe U and $H_\infty(X) \geq k$. Fix $\varepsilon > 0$. Let $\mathcal{H}$ be a universal hash family of size 2^d with output length $m = k - 2\log(1/\varepsilon)$. Define*

$$Ext(x, h) = h(x).$$

Then Ext is a strong $(k, \varepsilon/2)$ extractor with seed length d and output length m.

The leftover hash lemma simply states that a universal hash family gives an extractor. The seed is used to choose a hash function and the output is simply the hash of the input.

Proof: The proof is comprised of three parts. First we bound the collision probability of the extractor. Then we use this to bound the L_2-distance between the extractor's output and true randomness. Finally, we convert the L_2-distance into statistical distance.

Let $M = 2^m$, $K = 2^k$, and $D = 2^d$. Let H be a random hash function chosen using the seed.

Collision Probability: Define the collision probability $\mathrm{CP}(Z)$ of a random variable Z to be the probability that two independent samples of Z are equal. Note that a distribution Z can be seen as

a vector π, where $\pi_i = \mathbb{P}[Z = i]$. With this in mind, the collision probability $\mathrm{CP}(Z)$ is equal to the square of the L_2-norm of π—that is[3],

$$\mathrm{CP}(Z) = \mathbb{P}_{Z\sim Z'}[Z = Z'] = \sum_i \mathbb{P}[Z = i]^2 = \sum_i \pi_i^2 = ||\pi||_2^2.$$

Part 1: The probability that two independent output-seed pairs will collide is

$$\begin{aligned}
\mathrm{CP}((H(X), H)) &= \mathbb{P}_{h_1,h_2\in\mathcal{H},x_1,x_2\in U}[(h_1(x_1), h_1) = (h_2(x_2), h_2)] \\
&= \mathbb{P}[h_1 = h_2]\,\mathbb{P}[h_1(x_1) = h_2(x_2)|h_1 = h_2] \\
&= \frac{1}{D}\left(\begin{array}{l}\mathbb{P}[h_1(x_1) = h_2(x_2)|h_1 = h_2, x_1 \neq x_2]\,\mathbb{P}[x_1 \neq x_2] \\ +\mathbb{P}[h_1(x_1) = h_2(x_2)|h_1 = h_2, x_1 = x_2]\,\mathbb{P}[x_1 = x_2]\end{array}\right) \\
&\leq \frac{1}{D}\left(\frac{1}{M} + \frac{1}{K}\right) \\
&= \frac{1+\varepsilon^2}{DM},
\end{aligned}$$

as $M = K\varepsilon^2$.

Part 2: Let v be the vector corresponding to the difference between the distributions of $(H(X), H)$ and (U_m, U_d). We have

$$\begin{aligned}
||v||_2^2 &= ||(H(X), H) - (U_m, U_d)||_2^2 \\
&= \sum_{h,h(x)} \left(\mathbb{P}[H(X) = h(x) \wedge H = h] - \mathbb{P}[U_m = h(x) \wedge U_d = h]\right)^2 \\
&= \sum_{h,h(x)} \left(\mathbb{P}[H(X) = h(x) \wedge H = h]^2 - \frac{2}{MD}\mathbb{P}[H(X) = h(x) \wedge H = h] + \frac{1}{M^2D^2}\right) \\
&= \mathrm{CP}((H(X), H)) - \frac{2}{MD} + \frac{1}{MD} \\
&\leq \frac{\varepsilon^2}{MD}.
\end{aligned}$$

Part 3: Let $u_i = \mathrm{sign}(v_i)$. Then

$$\begin{aligned}
||v||_1 &= \sum_i |v_i| \\
&= \langle v, u\rangle \\
&\leq ||v||_2\,||u||_2 \qquad \text{(by Cauchy-Schwarz)} \\
&\leq \frac{\varepsilon}{\sqrt{MD}}\sqrt{MD} \\
&= \varepsilon,
\end{aligned}$$

since $\mathrm{dimension}(u) = \mathrm{dimension}(U_m)\mathrm{dimension}(U_d) = MD$. This completes the proof, as

$$\Delta((H(X), H), (U_m, U_d)) = \frac{||v||_1}{2} \leq \frac{\varepsilon}{2}.$$

[3]The notation $Z \sim Z'$ means that Z and Z' are independent random variables with the same distribution. So $\mathbb{P}_{Z\sim Z'}[Z = Z']$ is simply the probability that two independent samples from the distribution of Z collide.

□

The leftover hash lemma and the construction we gave for a universal hash family give (k, ε) extractors with $n = d$ and $m = k - 2\log(1/\varepsilon)$. Ideally we want d to be smaller. Nonconstructively ([V11] Theorem 6.17) we can achieve $m = k - 2\log(1/\varepsilon) - O(1)$ and $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$, which is much better. Constructively ([GUV] Theorem 1.5 or [V11] Theorem 6.36), we can achieve $d = O_\alpha(\log(n/\varepsilon))$ and $m \geq (1-\alpha)k$ for any constant $\alpha > 0$.

4 Cryptographic Context

Suppose Alice and Bob share a secret w and want a secret key, but w is not uniform. For example w could be a password or the result of a Diffie-Hellman key exchange. They can send a *public* seed and use an extractor to obtain an almost uniform secret key. We only require that no single password is too likely. So we can deal with bad distributions.

What if instead we have a good distribution for X, but the adversary knows some correlated information E about it? The conditional distribution $\mathbb{P}[X|E=e]$ may not be good. As above, we can extract a good distribution from X. Thus, we see a connection between imperfect randomness and leakage resiliance. This will be made formal in the next section.

5 Average min-entropy

Definition 7 [4] *Let X, E be a joint distribution. Then we define the* average min-entropy *of X conditioned on E—denoted $H_\infty(X|E)$—as*

$$H_\infty(X|E) \equiv -\log \mathbb{E}\left[\max_x \mathbb{P}\left[(X|E=e)\right]\right]$$

Here $\mathbb{E}[Y]$ denotes the expected value of Y and is not to be confused with the distribution E. The expectation is taken over the values e of E.

5.1 Justification

Note that one might naively assume that the "average min-entropy" would be $\mathbb{E}[H_\infty(X|E=e)]$ (the average of the min-entropy of X under each value of the distribution E). However, this is not a very useful measure. Consider the case where X has k bits of min-entropy a priori, but is entirely predictable under one-half of the possible values $e \in E$, and uniformly distributed under the rest. Taking the min-entropy of each individual case and averaging would imply that X still has $k/2$ average bits of min-entropy given a corresponding sample from E. However, one-half the time, we can guess the value of X exactly!

Instead, we use the intuitive notion of "guessability." A distribution X has min-entropy of k bits if an adversary cannot guess a sample from X with probability greater than 2^{-k}. Thus, the min-entropy of a distribution is the negative logarithm of the guessability of that distribution. If the adversary gains some additional information (represented by getting a sample e from a correlated distribution), their best strategy is to guess the most likely possibility for x given their sample

[4] This definition and several of the theorems proved below are taken from [DORS08].

e. Thus, the guessability given a sample e is $\max_x \mathbb{P}[X|E=e]$, and the average guessability is the expected value over e of this value. The average min-entropy is then calculated by taking the negative logarithm of this expectation.

5.2 Leftover Hash Lemma for average min-entropy

Lemma 8 *The average min-entropy of X conditioned on E is at least the min-entropy of the joint distribution minus the number of bits required to express E. Mathematically,*
$H_\infty(X|E) \geq H_\infty(X,E) - \log|support(E)|$
$H_\infty(X|E_1,E_2) \geq H_\infty(X,E_1|E_2) - \log|support(E_1)|$

The Leftover Hash Lemma, seen earlier for standard min-entropy, has a direct correlation with average min-entropy.

Theorem 9 *If $H_\infty(X|E) \geq k$ and $m = k - 2\log(1/\varepsilon)$, then $(H(X), H, E) \approx_{\varepsilon/2} (U_m, U_d, E)$.*

Here m is, as before, the output length of a universal hash function H, which is a random variable. The hash function H should not be confused with H_∞, the entropy.

This formalizes the application of extractors discussed in Section 4: X should be interpreted as a secret and E as some leaked information about X. The above theorem shows that, as long as there is enough secret information left in X (in particular, that not too much information has leaked out through E), we can extract an almost-uniform secret using a universal hash function. Note that a weaker form of this result is true for any extractor. However, the bounds are tighter if it comes from a universal hash family.

5.3 Improved LHL for Cryptographic Applications

Suppose we have a signature scheme that is δ-secure against adversary A assuming the keys are chosen using uniform randomness. (By δ-secure, we mean that the probability of an adversary forging a signature is at most δ.) If we choose the keys using extracted randomness that is ε-close to uniform, the scheme is $(\delta+\varepsilon)$-secure. This follows directly from the fact that if $X \approx_\varepsilon Y$ and f is any (possibly randomized) function, then $f(X) \approx_\varepsilon f(Y)$.

If we use keys extracted using the LHL, we can express the security as

$$\delta + \frac{\varepsilon}{2} = \delta + \frac{1}{2}\sqrt{2^{-\lambda}} \tag{1}$$

where $\lambda = k - m$ is the entropy loss. (This is the result of substituting for ε using the earlier equation $M = K\varepsilon^2$.)

However, this was only using the result of the LHL analysis as a black box rather than looking at this specific problem; we can get a better security bound by reanalyzing in light of this specific application. In part 3 of the proof of the LHL, we used a vector u containing only 1 and -1 values and took its dot product with v in order to bound the L1 norm of v. For the application to signatures, however, we ultimately want to bound the adversary's chance of forging a signature. So we can restrict our analysis to outputs that lead to forgery; we can 'zero-out' the good outputs. Thus,

instead of using u, we define u' to be the vector of success probabilities of the adversary over all possible keys (ignoring signs)—that is,

$$u'_i = \text{sign}(v_i)\mathbb{P}\left[\text{The adversary can forge a signature with } i \text{ as the output of the extractor.}\right].$$

Assume, for now, that the adversary's success or failure is deterministic with a fixed key; then u' is a vector of 0s, 1s and -1s. Furthermore, since the adversary's overall success probability for uniform keys is δ, we have that $||u'||_2 \leq \sqrt{\delta DM}$. Following the Cauchy-Schwarz analysis from before, we conclude that $||v||_1 \leq \varepsilon\sqrt{\delta}$, and therefore get a security of

$$\delta + \frac{1}{2}\sqrt{\delta 2^{-\lambda}}.$$

So we have an extra $\sqrt{\delta}$ factor for security. This allows us to have $\lambda \leq 0$ without disaster, as would be the case with (1). So we can run the extractor without entropy loss; or, indeed, with entropy gain.

If the adversary is not deterministic, then instead we have $\mathbb{E}\left[||u'||_1\right] \leq \delta DM$. By convexity and Hölder's inequality,

$$\mathbb{E}\left[||u'||_2\right]^2 \leq \mathbb{E}\left[||u'||_2^2\right] \leq \mathbb{E}\left[||u'||_1 ||u'||_\infty\right] \leq \mathbb{E}\left[||u'||_1\right] \leq \delta DM.$$

Thus we get the same result for randomized adversaries.

6 Extractor Game

We then proceeded to define an "alternating extractor" game between two parties, based on the work on Dziembowski and Pietrzak[DP07]. (An additional and perhaps simpler explanation of this idea can be found in [DW09]).

We assume that two parties, Q and W, each have an extremely long string (also referred to as Q and W). Q also starts with a short string S_1, and starts the game by sending S_1 to W. Whenever either party receives a message, they use that message as a seed to an extractor, apply the extractor to their string, and send the result to the other party. Thus, the interaction looks like this:

$$\begin{array}{lcl} Q & \xrightarrow{\qquad S_1 \qquad} & W \\ & \xleftarrow{R_1 = Ext(W; S_1)} & \\ & \xrightarrow{S_2 = Ext(Q; R_1)} & \\ & \xleftarrow{R_2 = Ext(W; S_2)} & \\ & \cdots & \end{array}$$

Theorem 10 *(Informal statement) If the total amount of information received by W up through round i is less than $|Q|$ (minus some small factor), then S_i is close to uniform from the point of view of W.*

A more formal statement of the theorem, along with a proof, is expected in the next lecture.

References

[DORS08] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, *Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data*. In *SIAM Journal on Computing* 38(1):97-139, 2008.

[DP07] S. Dziembowski and K. Pietrzak, *Intrusion-Resilient Secret Sharing*. In *FOCS*, pages 227-237, 2007.

[DW09] Y. Dodis and D. Wichs, *Non-Malleable Extractors and Symmetric Key Cryptography from Weak Secrets*. In *STOC*, pages

[GUV] V. Guruswami, C. Umans, and S. Vadhan, *Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes*. In J. ACM 56, 4, Article 20, 2009

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby *Pseudo-random generation from one-way functions*. In Proceedings of the twenty-first annual ACM symposium on Theory of computing (STOC '89) pages 12-24, 1989

[V11] S. Vadhan, *Pseudorandomness*, Draft: `http://people.seas.harvard.edu/~salil/pseudorandomness/`