Randomness in Cryptography January 24, 2013

Lecture 3: Privacy and Weak Sources

Lecturer: Yevgeniy Dodis *Scribe: Hamidreza Jahanjou*

Secure encryption requires entropy (Theorem 6, Lecture 2). In particular, even for a uniform distribution $X \equiv U_m$, there are strong bounds on either m or $\mathbf{H}_\infty(R)$. Today, we set out to answer the following question: is full entropy really necessary? As we shall see, the answer is yes.

1 Last Time

Let's quickly review some results from previous lectures.

DEFINITION 1 Let $(\mathsf{Enc}, \mathsf{Dec})$ denote an encryption scheme where $\mathsf{Enc} : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^\lambda$ and $\mathsf{Dec} : \{0,1\}^m \times \{0,1\}^\lambda \to \{0,1\}^n$ are functions. A correct encryption scheme satisfies $\forall r, x\ \mathsf{Dec}_r(\mathsf{Enc}_r(x)) = x$. Let R and X denote distributions on r and x respectively. $(\mathsf{Enc}, \mathsf{Dec})$ is said to be (k, ε)-secure if it is (R, ε)-secure for every k-source R; i.e. $\mathbf{H}_\infty(R) \geq k$. Clearly, $\varepsilon = 0$ means perfect security on k-sources. ♢

Lemma 1 *One-time pad (OTP) is $(m, 0)$-secure.*

But what can be said when $k < m$? Recall that for $n = 1$, (k, ε)-security means that $\mathsf{SD}(\mathsf{Enc}_R(0), \mathsf{Enc}_R(1)) \leq \varepsilon$. It turns out that nothing good is possible even when $k = m - 1$.

2 Warm-Up

Let's begin by showing the impossibility of deterministic bit extractors.

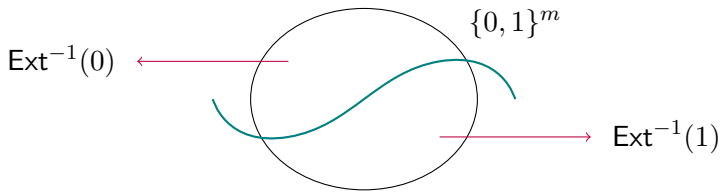
DEFINITION 2 A bit extractor $\mathsf{Ext} : \{0,1\}^m \to \{0,1\}$ is called (k, ε)-secure if for every k-source R, $\mathsf{Bias}(\mathsf{Ext}(R)) \leq \varepsilon$ where, for a distribution B, $\mathsf{Bias}(B) := |2\Pr[B = 0] - 1|$. ♢

We note that if a (k, ε)-secure bit extractor exists then a (k, ε)-secure bit encryption exists; namely, $\mathsf{Enc}_r(b) := b + \mathsf{Ext}(r)$.

In theorem 1, the constant 0.99 can be replaced by any number strictly less than 1.

Theorem 1 *No $(m - 1, 0.99)$-secure bit extractor exists.*

Proof: Let S_0 and S_1 denote the set of preimages of 0 and 1 under Ext respectively.



Without loss of generality, we assume $|S_0| \geq |S_1|$. Observe that U_{S_0} has min-entropy at least $m - 1$. Yet, $\mathsf{Ext}(U_{S_0}) = 0$; therefore, the bias is 1. □

3 Impossibility of Secure Encryption Schemes

Theorem 2 *Given two functions $f, g : \{0,1\}^m \to \mathcal{C}$ where $\mathcal{C}$ is any universe. Let*

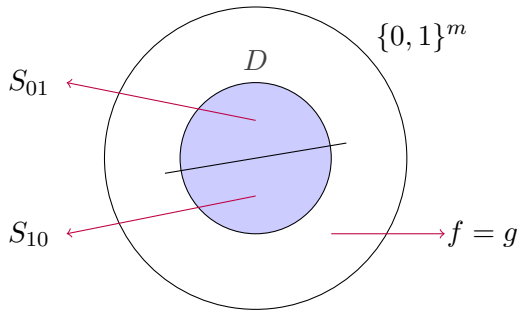
$$\Pr_{r \leftarrow U_m}[f(r) \neq g(r)] \geq 2^{-t}$$

for $0 \leq t \leq m$. Then there exist sources R_1 and R_2 such that

(a) $\mathbf{H}_\infty(R_1) \geq m - t - 1$ *and* $\mathsf{SD}(f(R_1), g(R_1)) \geq \frac{1}{2}$,

(b) $\mathbf{H}_\infty(R_2) \geq m - t - 2$ *and* $\mathsf{SD}(f(R_2), g(R_2)) \geq 1$.

Proof: Let's start with the special case where $\mathcal{C} = \{0,1\}$. Define $D = \{z : f(z) \neq g(z)\}$. By assumption, $|D| \geq 2^{m-t}$. In the picture below, let S_{01} denote the subset of $\{0,1\}^m$ on which the value of f is 0 and the value of g is 1. Similarly, let S_{10} denote the subset of $\{0,1\}^m$ on which the value of f is 1 and the value of g is 0.



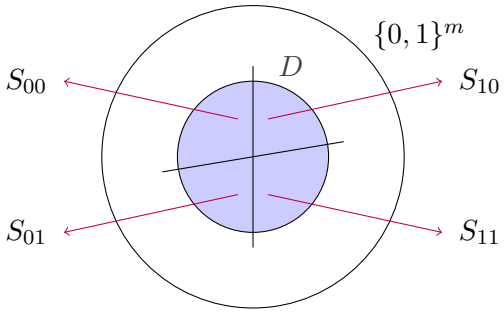
Without loss of generality, we can assume $|S_{01}| \geq |S_{10}|$. Therefore,

$$|S_{01}| \geq \frac{|D|}{2} \geq 2^{m-t-1}.$$

Set $R_1 := U_{S_{01}}$. We get $\mathbf{H}_\infty(R_1) \geq m - t - 1$. Moreover, $f(R_1) \equiv 0$ and $g(R_1) \equiv 1$ which implies that $\mathsf{SD}(f(R_1), g(R_1)) = 1$. Obviously, This result is strong enough to satisfy both parts (a) and (b).

Now, let's focus on the general case. Let $H = \{h : \mathcal{C} \to \{0,1\}\}$ be a family of universal hash functions; i.e. $\Pr_{H \leftarrow h}[h(z) \neq h(z')] = \frac{1}{2}$ for all $z \neq z'$. Also, define

$$S_{\alpha\beta}(h) = \{r \in D \,|\, h(f(r)) = \alpha \text{ and } h(g(r)) = \beta\}.$$



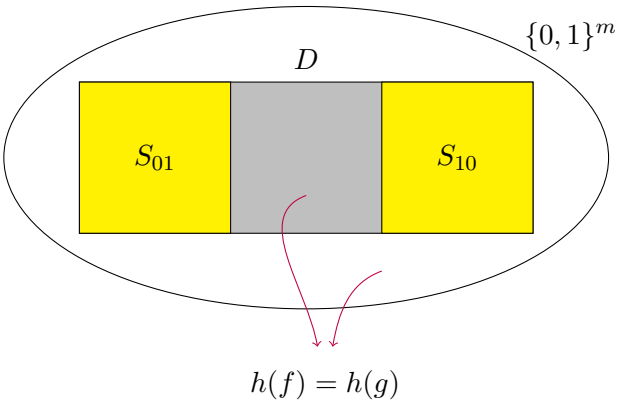
We'd like to compute

$$\begin{aligned}\mathbb{E}_H[|S_{01}| + |S_{10}|] &= \mathbb{E}_H\Big[\sum_{r\in D}\big(\chi_{S_{01}\cup S_{10}}(r)\big)\Big] \\ &= \sum_{r\in D}\Pr_H[r \in S_{01}\cup S_{10}] \\ &= \sum_{r\in D}\Pr_H\big[H(f(r)) \neq H(g(r))\big].\end{aligned}$$

Note that χ_A denotes the characteristic function of a set A. Moreover, we define

$$B(h) = \begin{cases} 0 & \text{if } |S_{01}(h)| \geq |S_{10}(h)| \\ 1 & \text{otherwise} \end{cases}$$

Now, there exists $h^* : \mathcal{C} \to \{0,1\}$ such that $|S_{01}(h^*) \cup S_{10}(h^*)| \geq \frac{|D|}{2} \geq 2^{m-t-1}$. Set $b^* = B(h^*)$. Without loss of generality, we can assume $b^* = 0$.



Observe that

$$|S_{01}| \geq \max(2^{m-t-2}, |S_{01}|) \text{ and } |S_{01}| + |S_{10}| \geq 2^{m-t-1}.$$

Define $R_2 := U_{S_{01}}$. Clearly, $\mathbf{H}_\infty(R_2) \geq m-t-2$; also, $h^*(f(R_2)) \equiv 0$ and $h^*(g(R_2)) \equiv 1$. Hence $\mathsf{SD}(f(R_2), g(R_2)) = 1$. Similarly, define $R_1 := U_{S_{10}}$. We have $\mathbf{H}_\infty(R_2) \geq m-t-1$. Finally,

$$\begin{aligned}\mathsf{SD}(f(R_1), g(R_1)) &\geq \Pr[h^*(f(R_1)) = 0] - \Pr[h^*(g(R_1)) = 0] \\ &\geq \Pr[R_1 \in S_{01}] \geq 1/2.\end{aligned}$$

□

Note that one can define Eve as $\mathsf{Eve}_{\mathsf{h}^*}(\mathsf{C}) = 1 \Leftrightarrow \mathsf{h}^*(\mathsf{C}) = 0$. Eve is efficient.

Exercise 1 *Is it possible to achieve* $\mathsf{SD}(f(R_1), g(R_1)) = 1$ *for all universes* $\mathcal{C}$*?*

In our case $f(r) = \mathsf{Enc}_r(0)$, $g(r) = \mathsf{Enc}_r(1)$ and $t = 0$ since for all secret keys $r \in \{0,1\}^m$ it holds that $\mathsf{Enc}_r(0) \neq \mathsf{Enc}_r(1)$. More precisely, we have the following result.

Theorem 3 *Even if $n = 1$, there is no $(m-1, 1/2)$- and $(m-2, 0.99)$-secure encryption scheme.*

Proof: Define $f(r) := \mathsf{Enc}_r(0)$ and $g(r) := \mathsf{Enc}_r(1)$. Clearly, $\forall r : f(r) \neq g(r)$. Since $\mathsf{Dec}_r(f(r)) \neq \mathsf{Dec}_r(g(r))$, we conclude $T = 1$ and $t = 0$. The lemma implies that there exist sources R_1 and R_2 such that $\mathsf{SD}(\mathsf{Enc}_{R_1}(0), \mathsf{Enc}_{R_2}(1)) \geq 1/2$ and $\mathsf{SD}(\mathsf{Enc}_{R_2}(0), \mathsf{Enc}_{R_2}(1)) = 1$. Moreover, $\mathbf{H}_\infty(R_i) \geq m - i$ for $i \in \{1, 2\}$. □

Although this result is stated in case of encryption, it seems to hold for most traditional privacy primitives. Some examples are in order.

Example 1 (Commitment) *Commitment can be viewed as a function $\mathsf{Com}(b; r) = C$ satisfying the following two properties*

(a) **Hiding***: $\mathsf{SD}(\mathsf{Com}(0; R), \mathsf{Com}(1; r)) \leq \varepsilon$,*

(b) **Binding***: intuitively, it is "hard" to find an r such that $\mathsf{Com}(0; r) \neq \mathsf{Com}(1; r)$.*

Instead of the binding property, we use the weak binding property which states that

$$\Pr_{r \in U_m}\left[\mathsf{Com}(0; r) \neq \mathsf{Com}(1; R)\right] \geq 1/2.$$

Now, setting $f(r) = \mathsf{Com}(0; r)$ and $g(r) = \mathsf{Com}(1; r)$ with get the impossibility of $(m - 1, 0, 99)$-commit.

Example 2 (Secret Sharing) *$(2, T)$ secret sharing ($T \geq 2$) even of one bit is impossible. We have secret shares $S_1 = \mathsf{Share}_1(b; r)$, $S_2 = \mathsf{Share}_2(b; r)$, ..., $S_T = \mathsf{Share}_T(b; r)$ with the following properties*

(a) $\mathsf{Rec}(S_1, ..., S_T) = b$ where Rec is the recovery function.

(b) No individual share (R, ε)-leak on b; for all $j \in [T]$ it holds that

$$\mathsf{SD}\big(\mathsf{Share}_j(0; R), \mathsf{Share}_j(1; R)\big) \leq \varepsilon.$$

Theorem 4 *No $(m - \log T - 1, 0.99)$ or $(m - \log T - 2, 1/2)$ secret sharing in possible.*

Proof:

$$\begin{aligned}
&\forall r : (S_1(0, r), ..., S_T(0, r)) \neq (S_1(1, r), ..., S_T(1, r)) \\
&\quad \Rightarrow \exists j = j(r) : S_j(0, r) \neq S_j(1, r) \\
&\quad \Rightarrow \exists j^* : |\{r | j(r) = j^*\}| \geq \frac{2^m}{T}
\end{aligned}$$

Define $f(r) := \mathsf{Share}_j(0, r)$ and $g(r) := \mathsf{Share}_{j^*}(1, r)$. Hence $\Pr[f(z) \neq g(z)] \geq 1/T$. The result follows. □

Question 1 *Can we make the distributions R_1 and R_2 efficiently samplable given oracle access to f and g while keeping Eve efficient?*

Question 2 *Did we have to lose $\log T$ for $(2, T)$ secret sharing?*

4 Getting out of Impossibility

Faced with these impossibility results, the question arises: where to go from here? To get out of impossibility, we need to change the model; i.e. put some restrictions. There seems to be four major directions for us to follow.

1. Make the source more structured by defining block sources. This approach culminates in Santha-Vazirani sources.
2. Is perfect/extractable randomness (to do both authentication and encryption) inherently necessary?
3. Relax correctness (a.k.a. differential privacy).
4. Allow public randomness.

Today we will focus on block devices. In particular we will analyze enhanced Santha-Vazirani sources (e-SVN). The main question we would like to answer is: can we do privacy with this type of sources?

4.1 Block Sources

So far, we were assuming one m-bit source R of min-entropy k. It is sometimes, however, reasonable to assume a sequence of potentially correlated sources R_1, R_2, ... where, for each i, $|R_i| = m$ and each source (block) has some "fresh" entropy given some other block. We're interested in block sources not only because they can potentially help us to overcome the impossibility results but also because they are reasonable sources deserving further study.

DEFINITION 3 A (potentially unbounded) sequence of random variables R_1, R_2, ... is called a (k, m)-block source if $|R_i| = m$ and, for all i and for all fixed values $r_1, ..., r_{i-1} \in \{0,1\}^m$ the following holds

$$\mathbf{H}_\infty(R_i | R_1 = r_1, R_2 = r_2, ..., R_{i-1} = r_{i-1}) \geq k.$$

This means that having fixed r_1, ..., r_{i-1}, Eve can sample any R_i of conditional entropy k. Moreover, m is the block length and the entropy rate, r, is k/m. ♢

Note 1 *We can also consider the more general case in which block length is not fixed (i.e. $\exists m_1, ...$) and/or the conditional min-entropy is not fixed (i.e. $\exists k_1, ...$).*

Note 2 *We have used the worst-case conditional min-entropy in our definition. So, why not use $\mathbf{H}_\infty(R_i | R_1, ..., R_{i-k})$? One reason is that the Chain Rule does not hold for $\mathbf{H}_\infty$; with worse-case min-entropy, it is possible to argue that $R^i := (R_1, R_2, ..., R_i)$ is an (ik)-source. Furthermore, we don't want to condition on "the future." In a way, we don't want the future to pay for the past. As we shall see, enhanced block sources, however, allow such possibility.*

DEFINITION 4 A (potentially unbounded) sequence of random variables R_1, R_2, ... is called a (k, m)-enhanced block source if $|R_i| = m$ and, for all i and for every $I \subseteq [r]$, such that $i \notin I$, and for every fixed value vector $r_I = (r_{i_1}, ..., r_{i_t})$, where i_1, ..., i_t are indices into I,

$$\mathbf{H}_\infty(R_i | R_I = r_I) \geq k.$$

$\diamondsuit$

Example 3 *Suppose $A, B \equiv U_{m/2}$. Then, $R_1 = (A, 0)$ and $R_2 = (A, B)$ are $(m/2, m)$-block sources (but not enhanced).*

The simplest case $m = 1$, in the above example, corresponds to Santha-Vazirani sources. It's the friendliest (possibly enhanced) block source.

Santha-Vazirani Sources

DEFINITION 5 Let B_1, B_2, ... be a potentially unbounded sequence of Boolean random variables.

$$\mathsf{SV}(\gamma) = \{B_1 B_2 ... \mid \forall i \, \forall b_1, ..., b_{i-1} \in \{0,1\}^{i-1} : \\ \Pr[B_i = 0 \mid B_1 = b_1, ..., B_{i-1} = b_{i-1}] \in [\frac{1}{2}(1-\gamma), \frac{1}{2}(1+\gamma)]\}.$$

Equivalently,

$$\mathsf{Bias}(B_i \mid B_{\{1,...,i-1\}} = b_{\{1,...,i-1\}}) \leq \gamma.$$

$\diamondsuit$

Note that $\gamma = 0$ corresponds to the perfect source. Enhanced Santha-Vazirani sources are defined similarly

DEFINITION 6 Let B_1, B_2, ... B_N be a sequence of Boolean random variables,

$$\mathsf{eSV}(\gamma) = \{B_1 ... B_N \mid \forall i \, \forall b_{[N]\setminus\{i\}} : \mathsf{Bias}\big(B_i \mid B_{[N]\setminus\{i\}} = b_{[N]\setminus\{i\}}\big) \leq \gamma\}.$$

The source rate r is $\log(\frac{2}{1+\gamma})$. $\diamondsuit$

eSV is the most structured source without independence. The main question here is the following. Can we do privacy with $\mathsf{eSV}(\gamma, N)$? Unfortunately, as we will see, the answer is negative.

Given that our results will be negative, we consider the friendly case of an enhanced γ-SV source where $\gamma > 0$.

DEFINITION 7 Given $N \geq 1$ and any $S \subseteq \{0,1\}^N$ with $|S| = 2^{N-1}$; let $\mathcal{H}_S(\gamma, N)$ denote the following distribution

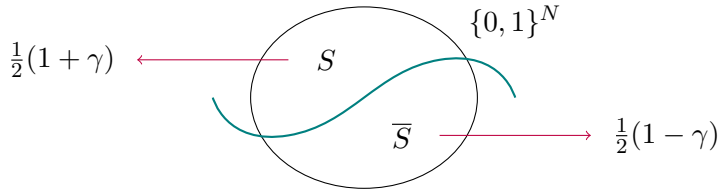
$$R \equiv \mathcal{H}_S(\gamma, N) := \begin{cases} \Pr[R = r] = (1+\gamma)2^{-N} & \text{if } r \in S, \\ \Pr[R = r] = (1-\gamma)2^{-N} & \text{if } r \notin S. \end{cases}$$

$\mathcal{H}_S(\gamma, N)$ is also know as a γ-semi-flat source. Furthermore,

$$\mathcal{H}(\gamma, N) = \{\mathcal{H}_S(\gamma, N) \mid S \subseteq \{0,1\}^N \text{ and } |S| = 2^{N-1}\}.$$

♢

As illustrated, One can view $\mathcal{H}_S(\gamma, N)$ as first using a γ-biased coin to select S or $\overline{S}$ and then selecting a uniform sample from the corresponding set.



Lemma 2 $\mathcal{H}(\gamma, N) \subsetneq \mathsf{eSV}(\gamma, N)$.

Proof: First, note that the lemma is equivalent to saying that for all S, $\mathcal{H}_S(\gamma, N)$ is an enhanced γ-SV source. Take any $i \in [N]$, for every $b_i \in \{0,1\}$ and any $b_{-i} \in \{0,1\}^{N-1}$ let $R = \mathcal{H}_S(\gamma, N) = (B_i, B_{-i})$. Consider

$$\frac{\alpha}{\beta} := \frac{\Pr[B_i = 0 \mid B_{-i} = b_{-i}]}{\Pr[B_i = 1 \mid B_{-i} = b_{-i}]} = \frac{\Pr[(B_i, B_{-i}) = (0, b_{-i})]}{\Pr[(B_i, B_{-i}) = (1, b_{-i})]} \in \left[\frac{1-\gamma}{1+\gamma}, \frac{1+\gamma}{1-\gamma}\right]$$

On the other hand, $\alpha + \beta = 1$. Hence, $\alpha, \beta \in [\frac{1}{2}(1-\gamma), \frac{1}{2}(1+\gamma)]$. We've arrived at the definition of an enhanced source and thus the proof is complete. □

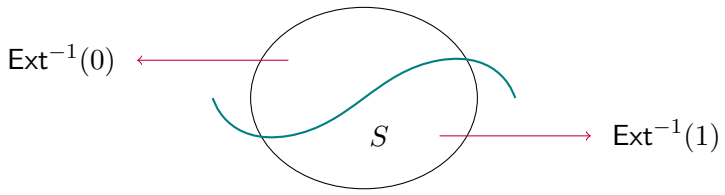
Impossibility of Privacy with Santha-Vazirani Sources

Here we will derive an impossibility result for privacy with $\mathcal{H}_S(\gamma, N)$ for all N which implies one with $\mathsf{SV}(\gamma, N)$ for all N. First, as a warm-up, let's prove a negative result for bit extraction. We present a new proof here. Previous papers used a direct proof for $\mathsf{SV}(\gamma, N)$ using a greedy strategy with induction on N. Our proof, while less intuitive, is shorter and stronger.

Theorem 5 $\forall N \; \forall \; \mathsf{Ext} : \{0,1\}^N \to \{0,1\}$, *there exists a* γ*-semi-flat source* $R \in \mathcal{H}(\gamma, N) \subset \mathsf{eSV}(\gamma, N)$ *such that* $\mathsf{Bias}(\mathsf{Ext}(R)) \geq \gamma$.

Proof: Without of loss of generality, assume $|\mathsf{Ext}^{-1}(1)| \geq |\mathsf{Ext}^{-1}(0)|$. Let S be any 2^{N-1}-subset of $\mathsf{Ext}^{-1}(1)$. Now, let $R := \mathcal{H}_S(\gamma, N)$. By construction we have

$$\Pr[\mathsf{Ext}(R) = 1] \geq \Pr[R \in S] = \frac{1}{2}(1+\gamma).$$

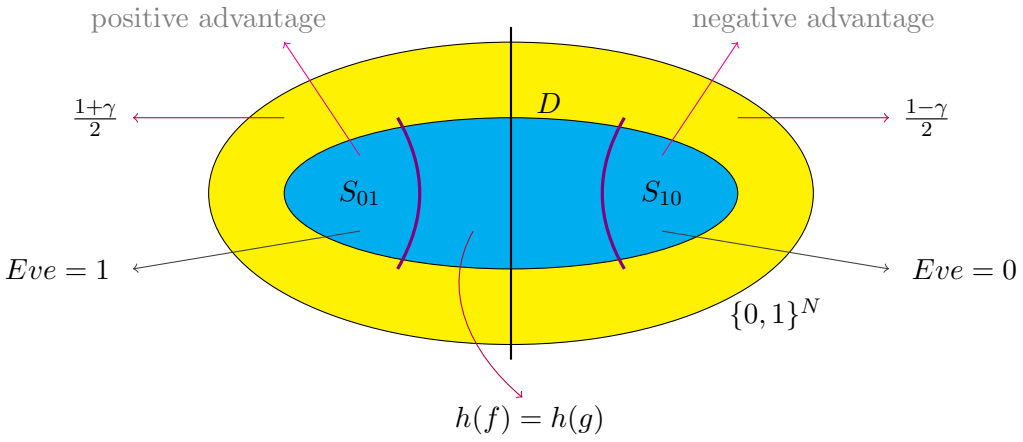


□

Now we present an optimized proof of the following lemma appearing in [2].

Theorem 6 *Given two functions $f, g : \{0,1\}^N \to \mathcal{C}$. Let $\Pr_{r \leftarrow U_m}[f(r) \neq g(r)] \geq 1/T$ for some $T \geq 1$ (e.g. $T = 2^t$). Then there exists a subset $S \subseteq \{0,1\}^N$ of size 2^{N-1} such that, setting $R := \mathcal{H}_S(\gamma, N)$, $\mathsf{SD}(f(R), g(R)) \geq \gamma/2T$.*

Proof: First recall that (in theorem 2) we defined $D := \{r \mid f(r) \neq g(r)\}$ and showed the existence of an efficient hash function $h^* : \mathcal{C} \to \{0,1\}$ giving rise to the situation illustrated below.



In particular $|S_{01}| \geq \max(2^N/4T, |S_{10}|)$ and $|S_{01}| + |S_{10}| \geq 2^N/2T$. As before, Eve is defined as $\mathsf{Eve}(C) = 1 \Leftrightarrow h^*(C) = b^* = 0$. Now, we have

$$\begin{aligned}
\mathsf{SD}(f(R), g(R)) &\geq \Pr[\mathsf{Eve}(f(R)) = 0] - \Pr[\mathsf{Eve}(g(R)) = 0] \\
&\geq \Pr[R \in S_{01}] - \Pr[R \in S_{10}] \\
&= (1+\gamma)2^{-N} \cdot |S_{01}| - (1-\gamma)2^{-N} \cdot |S_{10}| \\
&= 2^{-N}\big[\underbrace{|S_{01}| - |S_{10}|}_{\geq 0} + \gamma \cdot (\underbrace{|S_{01}| + |S_{10}|}_{\geq |D|/2})\big] \\
&\geq \frac{\gamma}{2^N} \cdot \frac{|D|}{2} \\
&= \frac{\gamma}{2T}
\end{aligned}$$

□

Note 3 *For $\mathcal{C} = \{0,1\}$, the lower bond can be improved to γ/T since in this case h^* can be identity and $1/2$ can be saved.*

Note 4 *This is the same lemma as theorem 2 except that here we have N in place of m and we have to settle for a smaller but still non-trivial advantage $\Omega(\gamma/T)$ instead of 1 or $1/2$.*

Note 5 *For most privacy primitives, such as commitment, secret sharing, etc., the length N of source can't achieve $(\mathsf{eSV}(\gamma, N), \Omega(\gamma/T))$ security. Consequently, for a non-negligible γ (and T) we can't achieve "negligible" security.*

Note 6 *For all N and for all $\mathsf{Ext} : \{0,1\}^N \to \{0,1\}^N$,*

$$\exists R \in eSV(\gamma, N) \text{ s.t. } \Pr[\mathsf{Ext}(R) = 0] \notin \big(\frac{1}{2}(1-\gamma), \frac{1}{2}(1+\gamma)\big).$$

That is, the best 1-bit extractor for γ-SV sources is r_1; the bias can not be reduced below γ.

We end today's lecture with some open questions and projects.

Quesject 1 *Can we make the distribution R efficiently samplable, given oracle access to f and g (and still keep Eve efficient)? To get some ideas see the recent papaer of Austen et al. [1].*

5 MACs with SV/block or eSV/eblock Sources

Project 1 *Investigate MACs with block and SV sources as well as their enhanced versions. In the case of enhanced sources, this seems to be easy. The regular version, however, seems to be much harder.*

Exercise 2 *For the enhanced version, prove that $\gamma = 2^{-\Omega(kn)}$, where $N = O(n)$, using a simple coding scheme.*

Quesject 2 *What is the general capacity of γ-SV sources?*

References

[1] Per Austrin, Kai-Min Chung, Mohammad Mahmoody, Rafael Pass, and Karn Seth. On the (Im)Possibility of Tamper-Resilient Cryptography: Using Fourier Analysis in Computer Viruses. *In preparation, 2012*

[2] Yevgeniy Dodis, Adriana Lpez-Alt, Ilya Mironov and Salil P. Vadhan. Differential Privacy with Imperfect Randomness. In *CRYPTO 2012: 497-516*