Randomness in Cryptography January 17, 2013

Lecture 2: Optimality of One-time MACs and Shannon Impossibility

Lecturer: Yevgeniy Dodis *Scribe: Benjamin Fuller and Sasha Golovnev*

Today we conclude with our study with one-time message authentication codes. In Lecture 1, we defined one-time secure MACs and constructed these MACs using $\delta - \mathsf{AXU}$ functions. Furthermore, we showed that the security of this construction lost security exponentially with as the min-entropy of the key decreased. We show that the constructions achieved in Lecture 1 were essentially tight. We will do this by showing that any one-time secure MAC must have a certain amount of entropy in its key (regardless of the key length). We then transition to our first privacy application: one-time encryption. We then move to privacy applications and encryption. We then discuss information theoretic notions of distance and show a generalized version of Shannon impossibility [6, 1].

1 Last Time

We recall the definition of a one-time MAC:

Definition 1 Let function $\mathsf{Tag} : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^\lambda$ be a function and let G_r parameterized by $r \in \{0,1\}^m$. There are two players: a challenger C who receives r as input, and an adversary E who receives no input. G_r has the following three steps.

1. E chooses $x \in \{0,1\}^n$ and sends x to C.
2. C computes and sends $t := \mathsf{Tag}_r(x)$ to E.
3. E outputs $(x', t') \in \{0,1\}^n \times \{0,1\}^\lambda$.

We say that E *wins* G_r if $x' \neq x$ and $\mathsf{Tag}_r(x') = t'$, and write $\mathsf{Adv}_E(r) := \Pr[E \text{ wins } G_r]$ to denote E's *advantage*. Let R be a distribution on $\{0,1\}^m$ and $\delta > 0$. Tag is a (R,δ)-*secure one-time MAC* if for every E,

$$\mathbb{E}_{r \leftarrow R}[\mathsf{Adv}_E(r)] \leq \delta.$$

When $R \equiv U_m$, we simply say δ-*secure.* ♢

Recall we constructed one-time MACs from δ-AXU functions. Furthermore, we showed that reducing the min-entropy of the key reduces security only exponentially:

Theorem 1 (Theorem 3 from Lecture 1) *If* Tag *is a δ-secure MAC with key length m, then for every $R, \mathbf{H}_\infty(R) = k \leq m$ it is also a $(R, 2^{m-k}\delta)$-secure MAC.*

2 Optimality of one-time MACs

We begin by recalling Theorem 4 from last lecture using our δ-AXU construction:

Theorem 2 (Theorem 4 from Lecture 1) *For any k such that $m/2 + \log n < k \leq m$, there is an efficient function $\mathsf{Tag} : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^\lambda$ that is a $(k, n \cdot 2^{m/2-k})$-secure MAC with tag length $\lambda = m/2$.*

In other words, for every n and δ, every $m \geq 2\log(n/\delta)$, and every k such that $m/2 + \log(n/\delta) \leq k \leq m$, there exists a (k, δ)-secure MAC with tag length $\lambda = m/2$.

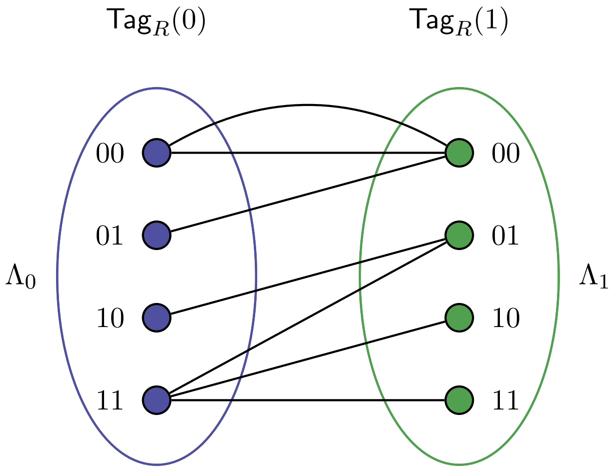
Today we will show that the result of Theorem 2 is optimal (for $n = 1$).

Theorem 3 *Let $\mathsf{Tag} : \{0,1\}^m \times \{0,1\} \to \{0,1\}^\lambda$ be a function. For any $k \leq m$ there exists R with $\mathbf{H}_\infty(R) \geq k$ and an attacker E such that:*

a) If $k \leq m/2$ then $\mathsf{Adv}_E(R) = 1$.

b) If $k > m/2$ then $\mathsf{Adv}_E(R) \geq 2^{m/2-k}$

Proof: Our proof will visualize $\mathsf{Tag}_R(\cdot)$ as creating a bipartite graph. The left nodes (denoted as Λ_0) will be the values of $\mathsf{Tag}_R(0)$, the right nodes (denoted Λ_1) will be the values of $\mathsf{Tag}_R(1)$. We will draw an edge between a left node, t_ℓ, and a right node, t_r if there exists an R such that $t_\ell = \mathsf{Tag}_R(0)$ and $t_r = \mathsf{Tag}_R(1)$. If there exists r, r' such that $\mathsf{Tag}_r(0) = \mathsf{Tag}_{r'}(0)$ and $\mathsf{Tag}_r(1) = \mathsf{Tag}_{r'}(1)$, then we have a bipartite multigraph. For convenience, we will remove vertices that have degree 0. We present an example graph below:



We note that $|\Lambda_0|, |\Lambda_1| \leq 2^\lambda$. For convenience, we denote the number of edges by $M = 2^m$. We now consider two possible cases:
1) There are few values in Λ_0 and thus E can predict the value of Λ_0
2) There are many values in Λ_0 but knowing Λ_0 gives significant information about Λ_1.

- Case 1: $\Lambda_0 \leq \sqrt{M}$. Since there at most $\sqrt{M}$ nodes in Λ_0 the average degree of a node $t_\ell \in \Lambda_0$ is at least $\sqrt{M}$ (since there are M edges in total). We consider the two conditions of Theorem 3 in turn.

  a) $k \leq m/2$ or equivalently $2^k \leq \sqrt{M}$. Since the average degree is at least $\sqrt{M}$, there is a node t_ℓ with $\mathsf{deg}(t_\ell) \geq \sqrt{M}$. We define the set R as $R = \{r \in \{0,1\}^m | \mathsf{Tag}_r(0) = t_\ell\}$, so that $|R| = \mathsf{deg}(t_\ell) \geq \sqrt{M}$. Without possibility of confusion we also denote by R the uniform distribution over this set. Note that $\mathbf{H}_\infty(R) \geq m/2 \geq k$, but E can simply output t_ℓ as $\mathsf{Tag}_R(0)$ and achieve $\mathsf{Adv}_E(R) = 1$.

b) $k > m/2$. We can no longer assume there is a single node that is of high enough degree. We take the largest degree nodes $t_1, ..., t_\ell$ such that $\sum_{i=1}^{\ell} \mathsf{deg}(t_i) \geq 2^k$. Since $\mathbb{E}_{t \leftarrow \Lambda_0} \mathsf{deg}(t) \geq \sqrt{M}$ the number of needed nodes is $\ell \leq \frac{2^k}{\sqrt{M}}$. As before define $R = \{r | \mathsf{Tag}_r(0) = t_i \text{ for some } 1 \leq i \leq \ell\}$ and the uniform distribution over this set. Then E randomly selects $1 \leq i \leq \ell$ and outputs t_i as $\mathsf{Tag}_R(0)$. Then $\mathsf{Adv}_E(R) \geq 1/\ell \geq \sqrt{M}/2^k = 2^{m/2-k}$.

- Case 2: $\Lambda_0 \geq \sqrt{M}$. Now given $\mathsf{Tag}_R(0)$, E will be able to forge $\mathsf{Tag}_R(1)$. As before we consider the two statements of the theorem:

1. $k \leq m/2$. For all $t \in \Lambda_0$ select an arbitrary incident edge e and add e to R and let R be uniform over these $|\Lambda_0| \geq \sqrt{M}$ edges. Note that $\mathbf{H}_\infty(R) \geq m/2 \geq k$. Now E given $\mathsf{Tag}_R(0)$ traverses the unique edge that was included in R and outputs the corresponding element in Λ_1. Note that $\mathsf{Adv}_R(E) = 1$.
2. $k > m/2$ or equivalently $2^k \geq \sqrt{M}$. Take a subset S of size exactly 2^k edges, that contains at least $\sqrt{M}$ vertices (denote these vertices as V) in Λ_0 and let R be the uniform distribution over this set S. Then given $t = \mathsf{Tag}_R(0)$ output a random $\mathsf{Tag}_R(1)$ adjacent to t according to S. Then $\Pr[\text{Success } E|t] \geq 1/\mathsf{deg}(t)$. Averaging over t one has:

$$\mathsf{Adv}_R(E) = \sum_{t=1}^{|V|} \frac{\mathsf{deg}(t)}{2^k} \frac{1}{\mathsf{deg}(t)} = \frac{|V|}{2^k} \geq \frac{\sqrt{M}}{2^k} = 2^{m/2-k}.$$

□

We have as an immediate corollary:

Corollary 1 *If* $\mathsf{Tag} : \{0,1\}^m \times \{0,1\} \to \{0,1\}^\lambda$ *is a* δ*-secure MAC then* $m \geq 2\log 1/\delta$

Proof: Suppose that $k = m$ in Theorem 3. Then $\delta \geq 2^{m/2-m} = 2^{-m/2}$. This implies that $m \geq 2\log 1/\delta$. □

One can also show a stronger statement that (R, δ)-security implies $\mathbf{H}_{sh}(R) \geq 2\log 1/\delta$. This leads to the following tempting conjecture:

Conjecture 4 *Let* $\mathsf{Tag} : \{0,1\}^n \times \{0,1\} \to \{0,1\}^\lambda$ *be a function. If* Tag *is* (R, δ) *secure then* $\mathbf{H}_\infty(R) \geq 2\log 1/\delta$.

In fact, we can even provide the following tempting "proof":

Tempting (but false) Proof of Conjecture: To argue that $\mathbf{H}_\infty(R) \geq 2\log 1/\delta$ we would like to make the following (*invalid*) argument:

$$2^{-k} = \mathsf{Pred}(R) \leq \mathsf{Pred}(\mathsf{Tag}_R(0), \mathsf{Tag}_R(1)) \leq \mathsf{Pred}(\mathsf{Tag}_R(0)) \cdot \mathsf{Pred}(\mathsf{Tag}_R(1) | \mathsf{Tag}_R(0)) \leq \delta \cdot \delta = \delta^2$$

Unfortunately, this line of reasoning crucially relies on the "chain" rule for min-entropy $\mathbf{H}_\infty(A, B) \geq \mathbf{H}_\infty(B) + \mathbf{H}_\infty(A|B)$, which is not true. In fact, we can actually explicitly disprove the conjecture, by giving the following counter-example of (R, δ)-secure MAC where $\mathbf{H}_\infty(R) \approx \log 1/\delta$ (which is clearly the lowest it can get, since obviously $\mathbf{H}_\infty(R) \geq \log 1/\delta$, as otherwise one can guess R with probability δ and forge both $\mathsf{Tag}_R(0)$ and $\mathsf{Tag}_R(1)$). Take a $2^{-m/2}$-secure MAC where $m \gg 2\log(1/\delta)$. Let be R be such that $\Pr[R = 0^m] = \delta$ and otherwise R is uniform over $\{0,1\}^m \setminus 0^m$. Then $\mathbf{H}_\infty(R) = \log 1/\delta$, but the resulting MAC is still $\delta' \approx \max(\delta, 2^{-m/2}) = \delta$ secure, since Eve can either win (for sure) if $R = 0^m$ (but this happens with probability only δ), or otherwise can win with probability at most (essentially) $2^{-m/2}$ by the original $2^{-m/2}$ security of our MAC under the uniform distribution.

Fortunately, we can use a related notion of entropy called *Collision or Rényi entropy.* We will denote this notion as $\mathbf{H}_2(R)$ and can show that $\mathbf{H}_2(R) \geq 2\log 1/\delta$ in our setting. Furthermore, unlike Shannon entropy, a bound between $\mathbf{H}_2(R)$ and $\mathbf{H}_\infty(R)$ is known, namely:

$$\mathbf{H}_{sh}(R) \geq \mathbf{H}_2(R) \geq \mathbf{H}_\infty(R) \geq \frac{\mathbf{H}_2(R)}{2}.$$

Interestingly (and unfortunately), Collision entropy will still *not* satisfy the natural variant of the "chain rule" (except in an important special case; see below). Fortunately, it will satisfy a weak form of the chain rule, which will suffice for our goal to recover the tempting argument above.

Theorem 5 *Let* $\mathsf{Tag} : \{0,1\}^n \times \{0,1\} \to \{0,1\}^\lambda$ *be a function. If* Tag *is* (R, δ)*-secure then* $\mathbf{H}_2(R) \geq 2\log(1/\delta)$.

We now define everything formally, study properties of Rényi entropy, and then return to the above Theorem.

Definition 2 We define the *collision probability* of a random variable R, denoted as $\mathsf{Col}(R)$ as $\mathsf{Col}(R) = \Pr_{r,r' \leftarrow R}[r = r'] = \sum_{r \in \{0,1\}^m} \Pr[R = r]^2$. Then, the *Collision entropy* of R, is $\mathbf{H}_2(R) = \log 1/\mathsf{Col}(R)$. ♢

Properties of Collision entropy:

1. $\mathbf{H}_2(R) \geq \mathbf{H}_\infty(R) \geq \mathbf{H}_2(R)/2$:

   Proof: $\mathbf{H}_2(R) \geq \mathbf{H}_\infty(R)$. Suffices to show that $\mathsf{Col}(R) \leq \mathsf{Pred}(R)$.

$$\begin{aligned}\mathsf{Col}(R) &= \sum_{r \in \{0,1\}^m} \Pr[R = r]^2 \leq \sum_{r \in \{0,1\}^m} \max_{r' \in \{0,1\}^m} \Pr[R = r'] \Pr[R = r] \\ &= \max_{r' \in \{0,1\}^m} \Pr[R = r'] \sum_{r \in \{0,1\}^m} \Pr[R = r] = \max_{r' \in \{0,1\}^m} \Pr[R = r'] = \mathsf{Pred}(R)\end{aligned}$$

$\mathbf{H}_2(R) \leq 2\mathbf{H}_\infty(R)$. Suffices to show that $\mathsf{Col}(R) \geq \mathsf{Pred}(R)^2$. One has

$$\mathsf{Col}(R) = \sum_{r \in \{0,1\}^m} \Pr[R = r]^2 \geq \max_{r' \in \{0,1\}^m} \Pr[R = r']^2 = \mathsf{Pred}(R)^2.$$

□

2. $0 \leq \mathbf{H}_2(R) \leq m$ with equality on the left only for a point distribution and equality on the right only for the uniform distribution.

   Since $\mathbf{H}_2(R)$ is a sum of probabilities all of which are in $[0,1]$ it is nonnegative. The equations: $\sum_{r=1}^{M} p_r^2 = 1, \sum_{r=1}^{M} p_r = 1, p_r \in [0,1]$ are satisfiable if and only some $p_r = 1$ that is if $\Pr[R = r'] = 1$ for some r'.

   The fact that $\mathbf{H}_2(R) \leq m$ follows from the Cauchy-Schwartz inequality as $\sum_{r=1}^{M} p_r^2 \geq (\sum_{r=1}^{M} p_r)^2/M$ (here $p_r = \Pr[R = r]$ and $M = 2^m$). Finally, $\mathsf{Col}(U_m) = \sum_{r \in \{0,1\}^m} \Pr[U_m = r]^2 = \sum_{r \in \{0,1\}^m} 2^{-2m} = \frac{2^m}{2^{2m}} = 2^{-m}$.

3. Let $f : \{0,1\}^m \to \{0,1\}^{m'}$ be a function. Then $\mathbf{H}_2(f(R)) \leq \mathbf{H}_2(R)$. It suffices to show that $\mathsf{Col}(f(R)) \geq \mathsf{Col}(R)$. One has:

$$\mathsf{Col}(f(R)) = \Pr[f(R) = f(R')] \geq \Pr[R = R'] = \mathsf{Col}(R)$$

Conditional Rényi Entropy. Define $\mathsf{Col}(A|B) = \mathbb{E}_{b \leftarrow B}[\mathsf{Col}(A|B = b)]$ and $\mathbf{H}_2(A|B) = \log(1/\mathsf{Col}(A|B))$.

We start with a "positive" property of this definition: $\mathbf{H}_2(A) \geq \mathbf{H}_2(A|B) \geq \mathbf{H}_\infty(A|B)$. (The proof is left as an exercise.)

Moving to the "negative", it is very tempting to make the following *incorrect* claim, which would be the (false) chain rule for Rényi entropy:

$$\mathbf{H}_2(A,B) = \mathbf{H}_2(B) + \mathbf{H}_2(A|B) \quad \Longleftrightarrow \quad \mathsf{Col}(A,B) = \mathsf{Col}(B) \cdot \mathsf{Col}(A|B)$$

To show this fallacy, we could proceed as follows:

$$\begin{aligned}
\mathsf{Col}(A,B) &= \Pr[A = A' \ \wedge \ B = B'] \\
&= \Pr[B = B'] \cdot \Pr[A = A' \mid B = B'] \\
(\text{incorrectly}) &= \mathsf{Col}(B) \cdot \mathop{\mathbb{E}}_{b \leftarrow B}[\Pr[A = A'|B = B' = b]] \\
&= \mathsf{Col}(B) \cdot \mathop{\mathbb{E}}_{b \leftarrow B}[\mathsf{Col}(A|B = b)] \\
&= \mathsf{Col}(B) \cdot \mathsf{Col}(A|B)
\end{aligned}$$

The mistake come from the fact that the following two experiments are *not the same in general.*

Experiment 1: sample independent $b \leftarrow B$ and $b' \leftarrow B$, until $b = b'$. Sample random A conditioned on $B = b$. Sample random A' conditioned on $B = b$.

Experiment 2: sample $b \leftarrow B$ once. Sample random A conditioned on $B = b$. Sample random A' conditioned on $B = b$.

Indeed, while the last two steps of both experiments are the same, the first step gives possibly different marginal distributions on B. While the second distribution is the original marginal of B, the first re-assignes the probability of $B = b$ to $\Pr[B = b]^2 / \sum_{b'} \Pr[B = b']^2$, which could be different from the original $\Pr[B = b]$.

We leave it as an exercise to find examples where Experiments 1 and 2 are very different (and, in particular, it is possible for either $\mathbf{H}_2(A, B) > \mathbf{H}_2(B) + \mathbf{H}_2(A|B)$ or $\mathbf{H}_2(A, B) < \mathbf{H}_2(B) + \mathbf{H}_2(A|B)$). However, we would like to point out one very useful special case where the Experiments are *indeed the same.* This happens when *the marginal distribution of B is uniform.* Indeed, in this case

$$\frac{\Pr[B = b]^2}{\sum_{b'} \Pr[B = b']^2} = \frac{1}{|\mathcal{B}|} = \Pr[B = b]$$

We get the following very useful lemma.

Lemma 2 *Assume the joint distribution (A, B) is such that the marginal on B is uniform. Then*

$$\mathbf{H}_2(A, B) = \mathbf{H}_2(B) + \mathbf{H}_2(A|B) \quad (or) \quad \mathsf{Col}(A, B) = \mathsf{Col}(B) \cdot \mathsf{Col}(A|B)$$

For general (A, B), we now state the following form of a "weak chain rule" for Rényi entropy:

Lemma 3 $\mathbf{H}_2(A, B) \geq \mathbf{H}_2(A|B) + \mathbf{H}_\infty(B)$.

Proof:

$$\mathsf{Col}(A|B) \cdot \mathsf{Pred}[B] = (\sum_b \Pr(B = b) \sum_a \Pr[A = a|B = b]^2) \cdot \max_b \Pr[B = b] \geq$$
$$\sum_{a,b} \Pr(B = b)^2 \Pr[A = a|B = b]^2 = \sum_{a,b} \Pr[A = a, B = b]^2 = \mathsf{Col}(A, B).$$

□

We can now use this weak chain rule to establish Theorem 5.

Proof: Using the monotonicity and the weak chain rule (Lemma (3)) of Rènyi entropy, coupled with $\mathbf{H}_2(A|B) \geq \mathbf{H}_\infty(A|B)$, we get

$$\begin{aligned}
\mathbf{H}_2(R) &\geq \mathbf{H}_2(\mathsf{Tag}_R(0), \mathsf{Tag}_R(1)) \\
&\geq \mathbf{H}_2(\mathsf{Tag}_R(1)|\mathsf{Tag}_R(0)) + \mathbf{H}_\infty(\mathsf{Tag}_R(0)) \\
&\geq \mathbf{H}_\infty(\mathsf{Tag}_R(1)|\mathsf{Tag}_R(0)) + \mathbf{H}_\infty(\mathsf{Tag}_R(0))
\end{aligned}$$

Indeed, the δ-security of the MAC and the definition of predictability of the MAC immediately imply that $\mathsf{Pred}(\mathsf{Tag}_R(1)|\mathsf{Tag}_R(0)), \mathsf{Pred}(\mathsf{Tag}_R(0)) \leq \delta$.

□

Remark 1 *We now consider MACs that are secure for more than one use. One can show that if* $\mathsf{Tag} : \{0,1\}^n \times \{0,1\} \to \{0,1\}^\lambda$ *is a* (R, δ)*-secure* c*-time MAC then* $\delta \geq 2^{\frac{cm}{c+1}-k}$ *and* $k \geq \frac{cm}{c+1}$*. These bounds can be achieved by using a* $(c+1)$*-wise independent hash function. Another related result (using weak chain rule) is that* $m \geq \mathbf{H}_2(R) \geq (c+1)\log\frac{1}{\delta}$.

We now present several questions related to information theoretic MACs.

Quesject 1 *Can the construction of one-time secure MACs be improved by using interaction instead of non-interactive* (x, t) *format? In particular, (a) is it still true that* $\delta \geq 2^{\frac{m}{2}-k}$*?; and (b) for uniform* R*, is it still true that* $m \geq \mathbf{H}_2(R) \geq 2\log\frac{1}{\delta}$*?*

Interestingly, Naor, Segev, and Smith [5] show that $\mathbf{H}_{sh}(R) \geq 2\log 1/\delta$*, but in part (b) we are asking for such a bound of* $\mathbf{H}_2(R)$*, which is stronger than* $\mathbf{H}_{sh}(R)$*. Also, Gemmell and Naor [4] show an interactive MAC construction where* $m = 2\log 1/\delta + O(1)$ *removing the* $\log n$ *term from prior constructions. However, there construction requires* perfect local randomness *to be sampled by Alice and Bob. Interestingly, once local randomness is allowed, one can show that (i)* $\delta \geq 2^{\frac{m}{2}-k}$ *is still true for non-interactive protocols; (ii) one can achieve non-trivial message authentication in two rounds for any* $k = \Omega(\log(1/\delta))$*, which is optimal, and totally beats the* $k > m/2$ *bound (meaning that question (a) is false for interactive protocols with local randomness). We will study point (ii), proved by Dodis and Wichs [3], later in the course, but mention that the following questions remain open for interactive MACs: (b') for uniform* R*, is it still true that* $m \geq \mathbf{H}_2(R) \geq 2\log\frac{1}{\delta}$ *even if local randomness is allowed?; (c) can we match Gemmell and Naor [4] bound* $m = 2\log 1/\delta + O(1)$ *(i.e., save* $\log n$*) without using local randomness?*

Quesject 2 *The counterexample presented in Theorem 3 the adversary* E *used a distribution that was not efficiently sampleable and* E *itself was not efficient. The question is whether we can present an* E *that is efficient given oracle access to* Tag*. A good midpoint is making only the sampler or* E *efficient (but not the other).*

3 Privacy Applications

We will now move from predicting applications (MACs) to a distinguishing application (encryption). In this area, we will have much stronger impossibility results when we try and reduce the entropy of the key. Before beginning we will cover notions of distance, as our security definitions will use these notions.

3.1 Notions of Distance

We want to ask what is the best way to distinguish two random variables A and B (possibly given C).

Definition 3 The *statistical distance* between random variables A and B, denoted $\mathsf{SD}(A, B)$

is

$$\begin{aligned}\mathsf{SD}(A,B) &= \max_{Eve} |\Pr[Eve(A)=1] - \Pr[Eve(B)=1]| \\ &= \max_{T\subseteq\{0,1\}^m} \Pr[A\in T] - \Pr[B \in T] \\ &= \frac{1}{2}\sum_{r\in\{0,1\}^m} |\Pr[A=r]-\Pr[B=r]|\end{aligned}$$

$\diamondsuit$

We could also use a game based definition where a challenge picks a bit $c \leftarrow \{0,1\}$, and if $c=0$ then samples $a \leftarrow A$ and gives a to E, else if $c=1$, sample $b \leftarrow B$ and give b to E. Then, $\mathsf{SD}(A,B) = \max_E \Pr[E \text{ guesses } c] - \frac{1}{2}$. This is equivalent to the above definition.

Definition 4 We will define *conditional statistical distance* between A and B conditioned on C, denoted $\mathsf{SD}(A,B|C)$ as $\mathsf{SD}(A,B|C) = \mathsf{SD}((A,C),(B,C)) = \mathbb{E}_{c\leftarrow C}[\mathsf{SD}(A|C=c, B|C=c)]$. $\diamondsuit$

Small statistical distance says that the additive difference between outcomes in A and B cannot be two large. We can also define a distance measure with respect to multiplication. Two distributions are close if $\Pr[A=y]/\Pr[B=y] \approx 1$ for all y. This definition is inspired by differential privacy [2]. We make the notion formal below.

Definition 5 We say that the *relative distance* between A and B is ε, denoted $\mathsf{RD}(A,B) = \varepsilon$ if ε is the smallest number such that $\forall y \in \{0,1\}^m$:

$$\Pr[B=y] \in [e^{-\varepsilon}\cdot\Pr[A=y]; e^{\varepsilon}\cdot\Pr[A=y]].$$

More generally, *conditional relative distance* between A and B conditioned on C, denoted $\mathsf{RD}(A,B|C)$, is defined as $\mathsf{RD}(A,B|C) = \mathsf{RD}((A,C),(B,C))$. $\diamondsuit$

For small values of $\varepsilon < 1/10$ we can approximate $e^{\pm\varepsilon} \approx 1\pm\varepsilon$. For notational convenience we will $\Pr[B=y] \in [e^{-\varepsilon}\Pr[A=y], e^{\varepsilon}\Pr[A=y]]$ as $\Pr[B=y]\in[e^{\pm\varepsilon}\Pr[A=y]]$.

As an easy exercise, we can also restate this definition in terms of an adversary Eve as follows:

Lemma 4 *$\mathsf{RD}(A,B) \le \varepsilon$ if and only if $\forall E$, $\Pr[E(A)=1] \in [e^{\pm\varepsilon}\Pr[E(B)=1]]$.*

Notice, for statistical distance restricting E to a computationally efficient machine produces a new notion of computational distance that is *weaker* than statistical distance. Indeed, most of modern cryptography is based thus this new computational distance. However, for relative distance, there is some singleton event y that maximally splits the two distributions. Thus, for non-uniform E the computational and unbounded versions of relative distance are equivalent. On the other hand, relative distance is a stricter notion than statistical distance. As an example, two distributions must have equal supports for relative distance to be bounded. This is formalized below (simple proof omitted).

Lemma 5 *If $\mathsf{RD}(A,B)\le\varepsilon$ then $\mathsf{SD}(A,B) \le e^{\varepsilon}-1 \approx \varepsilon$.*

We also remark that both notions are well defined distances (obeying the triangle inequality):

$$\begin{aligned}\mathsf{SD}(A,C) &\le \mathsf{SD}(A,B)+\mathsf{SD}(B,C)\\ \mathsf{RD}(A,C) &\le \mathsf{RD}(A,B)+\mathsf{RD}(B,C)\end{aligned}$$

To discuss privacy we will want to say that a ciphertext does not reveal much information about the underlying plaintext. Thus, we define the statistical/relative independence of two random variables.

Definition 6 The *statistical/relative independence* of A and B is $\mathsf{SI}(A;B) \triangleq \mathsf{SD}((A,B), A\times B)$ (resp. $\mathsf{RI}(A;B) \triangleq \mathsf{RD}((A,B), A\times B)$). $\diamondsuit$

Notice that $\mathsf{RI}(A;B)=0 \iff \mathsf{SI}(A;B)=0 \iff (A,B)\equiv A\times B$.

Also, $\mathsf{RI}(X,C)=\mathsf{RD}((X,C),X\times C)\le\varepsilon$ if and only if $\forall x,c$ we have $\Pr[X=x\wedge C=c]\in[e^{\pm\varepsilon}\Pr[X=x]\Pr[C=c]$. This easily implies

Exercise 1 *If* $\mathsf{RI}(C,X)\le\varepsilon$ *then* $\mathbf{H}_\infty(C,X)\ge\mathbf{H}_\infty(C)+\mathbf{H}_\infty(X)-\varepsilon\ln 2$.

Remark 2 *This is comparable to the average case notion of mutual information* $\mathbf{I}(A;B)$ *for Shannon's entropy:*

$$\mathbf{I}(A;B)=\sum_{a\in A}\sum_{b\in B}\Pr[A=a\wedge B=b]\log\frac{\Pr[(A,B)=(a,b)}{\Pr[A=a]\Pr[B=b]}.$$

Indeed, this quantity can be stated in terms of Shannon entropy: $\mathbf{I}(A;B)=\mathbf{H}_{sh}(A)+\mathbf{H}_{sh}(B)-\mathbf{H}_{sh}(A,B)$*, which is similar to the* RI *definition above.*

Finally, the following exercise shows that low relative information implies closeness for both the min-entropy and collision entropy, as follows.

Exercise 2 *If* $\mathsf{RI}(A,B|C)\le\varepsilon$ *then*

- $\mathsf{Pred}(A|C)\in[e^{\pm\varepsilon}\mathsf{Pred}(B|C)]$. *Equivalently,* $\mathbf{H}_\infty(A|C)\in\mathbf{H}_\infty(B|C)\pm\varepsilon\ln 2$.
- $\mathsf{Col}(A|C)\in[e^{\pm 2\varepsilon}\mathsf{Col}(B|C)]$. *Equivalently,* $\mathbf{H}_2(A|C)\in\mathbf{H}_2(B|C)\pm 2\varepsilon\ln 2$.

3.2 Generalized Shannon Bounds for Encryption

We will now define an encryption scheme. We will have a correctness and privacy requirement.

Definition 7 Let $\mathsf{Enc}:\{0,1\}^m\times\{0,1\}^n\to\{0,1\}^\lambda$ and $\mathsf{Dec}:\{0,1\}^m\times\{0,1\}^\lambda\to\{0,1\}^n$ be functions. For convenience we write $\mathsf{Enc}(r,x)$ as $\mathsf{Enc}_r(x)$ and $\mathsf{Dec}(r,c)$ as $\mathsf{Dec}_r(c)$. We say pair of functions are a *correct* encryption scheme if $\forall r,x,\mathsf{Dec}_r(\mathsf{Enc}_r(x))=x$. Let R,X be distributions and define $C=\mathsf{Enc}_R(X)$. We say that $(\mathsf{Enc},\mathsf{Dec})$ is (R,ε)*-relatively (resp. statistically) secure* on X if $\mathsf{RI}(X;C)\le\varepsilon$ (resp. $\mathsf{SI}(X;C)\le\varepsilon$). $\diamondsuit$

By Lemma 5 we know that any (R, ε)-relatively secure scheme X is (essentially) (R, ε)-statistically secure on X. Also, the more entropy X has, the harder it is to satisfy our definition. We will omit "on X" if a scheme is secure *for all* distributions X, but for our lower bounds security for (only) the uniform distribution will suffice. Also, it is easy to see that to achieve security for all distributions it suffices to achieve security for all min-entropy 1 distributions (i.e., uniform distributions over any two messages x_0 and x_1). Finally, we will simply say "secure" to mean *statistically* secure (this is the more standard notion).

We now have the following theorem about the difficulty of encryption with imperfect randomness:

Theorem 6 *If* $(\mathsf{Enc}, \mathsf{Dec})$ *is*

a) (R, ε)-relatively secure on X then $\mathbf{H}_\infty(R) \geq \mathbf{H}_\infty(X) - \varepsilon \ln 2$. In particular, if $X \equiv U_n$ then $\mathbf{H}_\infty(R) \geq n - \varepsilon \ln 2$. Furthermore, the adversary that breaks (R, ε) relative security when $\mathbf{H}_\infty(R) < n - \varepsilon \ln 2$ is efficient.

b) (R, ε)-secure, then $2^m \geq 2^{\mathbf{H}_\infty(X)}(1 - \varepsilon)$ or equivalently, $m \geq \mathbf{H}_\infty(X) - \log\left(\frac{1}{1-\varepsilon}\right) \geq \mathbf{H}_\infty(X) - 2\varepsilon$. In particular, if $X \equiv U_n$ then $m \geq n - 2\varepsilon$.

This shows the efficiency of the one-time pad $\mathsf{Enc}_r(x) = x \oplus r$ is essentially tight as it is $(U_m, 0)$-secure on $\{0,1\}^m$.

Proof: Proof of a). Recall for security, $\mathsf{RI}(X; C) \leq \varepsilon$. We will construct an E that separates $(C, X), C \times X)$. By Lemma 4 this provides a upper bound on security. Let r^* be the most likely value of R, that is $\Pr[R = r^*] = 2^{-\mathbf{H}_\infty(R)}$. Define $E(x, c)$ to output 1 iff $\mathsf{Dec}_{r^*}(c) = x$. This yields,

$$\Pr[E(X, C) = 1] = \Pr[\mathsf{Dec}_{r^*}(\mathsf{Enc}_R(X)) = X] \geq \Pr[R = r^*] = 2^{-\mathbf{H}_\infty(R)}$$

Now we consider the distribution $C \times X$ where C and X are sampled independently,

$$\Pr_{(x,c) \leftarrow X \times C}[E(x, c) = 1] = \Pr_{(x,c) \leftarrow X \times C}[\mathsf{Dec}_{r^*}(c) = x] \leq 2^{-\mathbf{H}_\infty(X)}$$

because $\mathsf{Dec}_{r^*}(C)$ is independent of X. By relative security $2^{-\mathbf{H}_\infty(R)} \leq e^{\varepsilon} 2^{-\mathbf{H}_\infty(X)}$ or equivalently, $\mathbf{H}_\infty(R) \geq \mathbf{H}_\infty(X) - \varepsilon \ln 2$.

Proof of b) consider the following (inefficient) E: $E(x, c) = 1$ if and only if there exists $r \in \{0,1\}^m$ such that $\mathsf{Dec}_r(c) = x$. Then $\Pr[E(X, C) = 1] = 1$ as the properly sampled key exists. Thus means by statistical security:

$$\begin{aligned} 1 - \varepsilon &\leq \Pr_{(x,c) \leftarrow X \times C}[\exists r^* \text{ s.t. } \mathsf{Dec}_{r^*}(c) = x] \\ &\leq \sum_{r \in \{0,1\}^m} \Pr_{x,c \leftarrow X \times C}[\mathsf{Dec}_r(c) = x] \leq 2^m \cdot 2^{-\mathbf{H}_\infty(X)} \end{aligned}$$

where again $\Pr_{x,c \leftarrow X \times C}[\mathsf{Dec}_r(c) = x] \leq 2^{-\mathbf{H}_\infty(X)}$ since C is independent of X. Rearranging terms, $2^m \geq (1 - \varepsilon) 2^{\mathbf{H}_\infty(X)}$. □

We also present an alternate proof of part a) that only uses the properties of independence and entropy established above:

Proof:

$$\begin{aligned}\mathbf{H}_\infty(R) &= \mathbf{H}_\infty(R, X) - \mathbf{H}_\infty(X)\\ &\geq \mathbf{H}_\infty(C, X) - \mathbf{H}_\infty(X)\\ &\geq \mathbf{H}_\infty(C) - \varepsilon \ln 2\\ &\geq \mathbf{H}_\infty(C|R) - \varepsilon \ln 2\\ &\geq \mathbf{H}_\infty(X|R) - \varepsilon \ln 2\\ &= \mathbf{H}_\infty(X) - \varepsilon \ln 2.\end{aligned}$$

The equality proceeds by independence of R and X. The first inequality because $(C = \mathsf{Enc}_R(X), X)$ is a deterministic function of (R, X), the second inequality because of relative security and Exercise 1, the third inequality since conditioning on R can only reduce the min-entropy, the fourth inequality because $X = \mathsf{Dec}_R(C)$ is deterministic function of R and C, and the last equality is again because X is independent of R. □

We will conclude by considering the implication of this result for modern cryptography. In modern cryptography, we restrict E to be computationally bounded and allow E to have a ε advantage in distinguishing the two distributions. We will consider the strongest starting conditions for Theorem 6, when we have have perfect randomness $R \equiv U_m$. We then consider part a) of Theorem 6 with $\varepsilon = 0$ (where relative and statistical security are the same) and part b) with nonzero ε.

Corollary 6 *Consider* $(\mathsf{Enc}, \mathsf{Dec})$ *as above.*

1. $(U_m, 0)$*-security with efficient* E *implies that* $m \geq n$.

2. (U_m, ε)*-security with inefficient* E *implies that* $m \geq n - 2\varepsilon$.

This means if we want to encryption with keys significantly shorter than messages we need to consider *both* efficient E and allow E nonzero probability of winning. These two restrictions will lead us to modern cryptography.

References

[1] Yevgeniy Dodis. Shannon Impossibility Revisited. In *International Conference on Information Theoretic Security.*

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography Conference 2006.*

[3] Yevgeniy Dodis and Daniel Wichs. Non-malleable Extractors and Symmetric Key Cryptography from Weak Secrets, Symposium on Theory of Computing (STOC), May 2009.

[4] Peter Gemmell and Moni Naor. Codes for interactive authentication. In *CRYPTO 1993.*

[5] Moni Naor, Gil Segev, Adam Smith. Tight Bounds for Unconditional Authentication Protocols in the Manual Channel and Shared Key Models. In *IEEE Transactions on Information Theory 54(6): 2408-2425 (2008).*

[6] Claude Shannon. Communication Theory of Secrecy Systems. In *Bell Systems Technical Journal 1945*.