

1. Consider the following two samplers:

- X takes in 5 pure random bits $r_0 r_1 r_2 r_3 r_4$ and outputs the 3 bits

$$x_1 = r_0, x_2 = r_1 r_2, x_3 = r_3 r_4.$$

- Y takes in 5 pure random bits $r_0 r_1 r_2 r_3 r_4$ and outputs the 6 bits $r_0 + r_i r_j$, $1 \leq i < j \leq 4$, namely

$$y_1 = r_0 + r_1 r_2, y_2 = r_0 + r_3 r_4, y_3 = r_0 + r_1 r_3, y_4 = r_0 + r_2 r_4, y_5 = r_0 + r_1 r_4, y_6 = r_0 + r_2 r_3.$$

(a) Calculate the probabilities of the events $x_i = 1$ for all i in sampler X .

Solution: x_1 is a random bit. It is one with probability half. x_2 and x_3 are products of two random bits each. They are one with probability $1/4$.

(b) Show that the test “Does x have Hamming weight at most 1?” distinguishes the output of X from pure random bits. What is its advantage?

Solution: There are four possible outcomes that result in acceptance: 000, 100, 001, 010. Under X , their respective probabilities are $9/32, 9/32, 3/32, 3/32$. The test accepts X with probability $24/32 = 3/4$. If the input is a string of random bits, it is equally likely to have fewer ones than zeros as fewer zeros than ones. The probability the test accepts it is $1/2$. The test’s advantage is $3/4 - 1/2 = 1/4$.

(c) (**Optional**) Is there a distinguisher with higher advantage than the one from part (b)?

Solution: Yes, this is the distinguisher that accepts exactly when $x_2 = x_3 = 0$. On an output of X , it accepts with probability

$$P(r_1 r_2 = 0) \cdot P(r_3 r_4 = 0) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}.$$

It accepts a pure random 3-bit string with probability $1/4$ only. Its advantage is therefore $9/16 - 1/4 = 5/16$. This is the highest possible advantage: the test accepts precisely those outputs whose probability under X is higher than under equally likely outcomes.

(d) Calculate the probabilities of the events $y_i = 1$ for all i in sampler Y .

Solution: For every i , the bit y_i has the form r_0 plus something independent of r_0 . Whatever the value of that something is, adding r_0 to it makes the result purely random. So each of these events has probability half, as if y were truly random.

(e) As Y ’s seed is shorter than its output, its output must be distinguishable from 6 pure random bits. Describe a distinguisher that does the job.

Solution: There are many possible distinguishers. One such D accepts if $y_1 + y_3 = 0$. For samples of Y , this happens whenever $(r_0 + r_1 r_2) + (r_0 + r_1 r_3) = r_1(r_2 + r_3)$ equals zero. As $r_2 + r_3$ is a pure random bit, and r_1 is an independent pure random bit, their product is zero with probability $3/4$. So D accepts an output of Y with probability $3/4$. In contrast, when y_1 and y_3 were purely random, D accepts with probability $1/2$. D ’s advantage is $1/4$.

2. This question is about indistinguishability of single versus multiple samples.

(a) Let S be the sampler that takes in two bits of seed r_1, r_2 and outputs $(r_1, r_2, r_1 + r_2)$. What is the advantage of the best distinguisher that tells the output of S from truly random bits?

Solution: The possible outputs of S are 000, 011, 101, 110. Each has probability $1/4$. The best distinguisher accepts these and rejects all others. Its advantage is $1/2$.

- (b) Let S^2 be the sampler that runs S twice on independent seeds: $S^2(r_1, r_2, r'_1, r'_2) = (S(r_1, r_2), S(r'_1, r'_2))$. Here S^2 produces six bits of output. What is the advantage of the best distinguisher for S^2 ?

Solution: The possible outcomes of S^2 are products of outcomes of S . The first three bits and the last three bits must be outcomes of S . There are 16 equally likely possibilities. Each has probability $1/4^2 = 1/16$. If the six bits were chosen uniformly at random, the probability would have been $1/2^6 = 1/256$. The best distinguisher accepts these 16 outcomes and rejects all others. Its advantage is $1 - 4^2/2^6 = 1 - (1/2)^2 = 3/4$.

- (c) Now S is arbitrary. Assume distinguisher D tells S from pure random bits, and D accepts all outputs of S . How should you use D to distinguish S^2 from random? How does the advantage change?

Solution: Run D on both samples and accept when both runs accept. Samples of S^2 will always be accepted. If D accepts uniform samples with probability p , uniform samples of S^2 will be accepted with probability p^2 . The advantage increases from $q = 1 - p$ to $1 - p^2 = 1 - (1 - q)^2 = 2q(1 - q)$. When q is small, the advantage almost doubles.

- (d) (**Optional**) Now assume distinguisher D' tells S^2 from pure random bits. How can you use D' to distinguish S from random? (**Hint:** Don't aim for maximum advantage.)

Solution: D' distinguishes two independent samples from S , call them S_1S_2 , from two uniform independent samples, call them U_1U_2 . This same D' must therefore either distinguish S_1S_2 from S_1U_2 , or it must distinguish S_1U_2 from U_1U_2 . If the former happens, D' effectively distinguishes S_2 from U_2 , as the S_1 component is fixed and independent of the rest. If the latter happens, D' effectively distinguishes S_1 from U_1 for the same reason. In either case, a sample of S from a uniform sample is distinguished.

Quantitatively, the advantage of D' in distinguishing S_1S_2 from S_1U_2 and its advantage in distinguishing S_1U_2 from U_1U_2 add up to its original advantage:

$$\begin{aligned} & \text{P}(D' \text{ accepts } S_1S_2) - \text{P}(D' \text{ accepts } U_1U_2) \\ &= (\text{P}(D' \text{ accepts } S_1S_2) - \text{P}(D' \text{ accepts } S_1U_2)) + (\text{P}(D' \text{ accepts } S_1U_2) - \text{P}(D' \text{ accepts } U_1U_2)). \end{aligned}$$

If the original advantage was q' , one of the two strategies will have advantage at least $q'/2$.

3. R is a pure random function that maps 4 input bits into one output bit. F is the keyed function

Setup stage: Sample and store a random 5-bit key $k = k_0k_1k_2k_3k_4$.

Evaluation stage: On input $x = x_1x_2x_3x_4$, output $k_0 + k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4$.

- (a) What are the probabilities of the events $R(x) = 1$ and $R(x) + R(y) = 0$ for distinct inputs $x \neq y$?

Solution: For each fixed input x , the value $R(x)$ is a pure random bit, so $\text{P}(R(x) = 1) = 1/2$. For distinct x and y , the bits $R(x)$ and $R(y)$ are independent. Their sum is zero when they are equal, so $\text{P}(R(x) + R(y) = 0) = \text{P}(R(x) = R(y)) = 1/2$.

- (b) What is the probability that $F(x) = 1$ for a given input x ?

Solution: For every fixed x , $F(x)$ is a sum of some nonempty subset of key bits. As the key is uniform this sum is unbiased: $\text{P}(F(x) = 1) = 1/2$.

- (c) What is the probability that $F(x) + F(y) = 0$ for a given distinct pair $x \neq y$?

Solution: For fixed x and y , $F(x) + F(y)$ is again a subset sum of key bits. The key bits that participate in the sum are those for which x_i and y_i differ. Since $x \neq y$ there must be at least one such bit. The sum is again unbiased: $\text{P}(F(x) + F(y) = 1) = 1/2$.

- (d) (**Optional**) If Eve makes only two queries, can she distinguish R from F ?

Solution: No. For any two distinct queries x and y , we saw that $\text{P}(F(x) = 1) = \text{P}(F(y) = 1) = \text{P}(F(x) + F(y) = 1) = 1/2$. The parities of all subsets of these two bits are unbiased. We said that the only random variable with this property is the uniform one: $(F(x), F(y))$ is indistinguishable from two uniform random bits.

If you are not convinced by this argument, you can use the laws of probability to derive the relation

$$P(F(x) + F(y) = 1) = P(F(x) = 1) + P(F(y) = 1) - 2P(F(x) = 1 \text{ and } F(y) = 1).$$

from where $P(F(x) = 1 \text{ and } F(y) = 1) = (1/2 + 1/2 - 1/2)/2 = 1/4$. This is the product of the marginal probabilities. $F(x)$ and $F(y)$ are independent.

- (e) What is the probability that $R(x) + R(y) + R(z) + R(w) = 0$ for distinct inputs x, y, z, w ?

Solution: The four bits $R(x), R(y), R(z), R(w)$ are independent pure random bits. Their parities are unbiased, so the probability of this event is $1/2$.

- (f) Find inputs for which the probability that $F(x) + F(y) + F(z) + F(w) = 0$ is different from that for R .

Solution: Inputs 0000, 0100, 1000, 1100 have this property as

$$F(0000) + F(0100) + F(1000) + F(1100) = k_0 + (k_0 + k_2) + (k_0 + k_1) + (k_0 + k_1 + k_2) = 0.$$

The relation holds with probability 1.

4. Assume F is a pseudorandom function with equal input and output length. The secret key to F is known to Alice and Bob only. Diagnose the problem with each of the following variants of the PRF pad encryption:

- (a) m encrypts to $F(r) + m$ only, where r is chosen at random at encryption time.

Solution: The ciphertext does not include r . Without knowing r , Bob cannot remove $F(r)$ from the ciphertext. Decryption is impossible.

- (b) m encrypts to $(r, F(r) + m)$, where r is chosen once at key generation and reused for every encryption.

Solution: Eve can distinguish the encryption of two identical messages from those of two different messages. In the former case, the encryptions will also be identical as r is reused. In the latter case, the second part will be different.

- (c) m encrypts to $(r, F(r) + m, F(r + m))$.

Solution: Eve can distinguish the encryption of the all-zero message 0 from the encryption of some other message, say 1. If $m = 0$ the second and third component of the ciphertext are identical: $F(r) + 0 = F(r) = F(r + 0)$. If not, they are likely to be distinct. $F(r + 1)$ and $F(r)$ are PRF evaluations at distinct inputs. They are indistinguishable from a pair of independent random values. So will be $F(r + 1)$ and $F(r) + 1$. They are therefore likely to be different. If F has r bits of output, they would collide with probability about 2^{-r} .