Lecture 19: Random Functions and Encrypting Long Messages

Random Functions

- Let $\mathcal{F}_{m,n}$ be the set of all function from the domain $\{0,1\}^m$ to the range $\{0,1\}^n$
- Each function $f \in \mathcal{F}_{m,n}$ can be uniquely represented by a list of length $\{0,1\}^m$ where the i-th entry in the list is the entry $f(i)$, for $i \in \{0,1\}^m$
- So, each entry in the list has 2^n options. And, there are a total of 2^m such entries. So, the total number of distinct functions from the set $\{0,1\}^m \to \{0,1\}^n$ is

$$\overbrace{(2^n) \times \cdots \times (2^n)}^{2^m\text{-times}} = 2^{n2^m}$$

- So, we can conclude that each function $f \in \mathcal{F}_{m,n}$ can be described using $n2^m$ bits

Crucial Property of Random Functions

Intuition.

- Suppose we pick a random $f \xleftarrow{\$} \mathcal{F}_{m,n}$
- Then the evaluation of f at any input x_1 is uniformly random over $\{0,1\}^n$.
- Further, the evaluation of f at any other input x_2 given $f(x_1)$ is again uniformly random over $\{0,1\}^n$.
- In particular, the evaluation of f at an input x_t given $f(x_1), \ldots, f(x_{t-1})$ is uniformly random
- Intuitively, the evaluation of a random f is completely unpredictable at any <u>new</u> input

Formally. For any distinct inputs $x_1, \ldots, x_t \in \{0,1\}^m$ and any outputs $y_1, \ldots, y_t \in \{0,1\}^n$, the following holds

$$\mathbb{P}_{f \xleftarrow{\$} \mathcal{F}_{m,n}} \left[f(x_t) = y_t | f(x_1) = y_1, \ldots, f(x_{t-1}) = y_{t-1} \right] = \frac{1}{2^n}$$

Secret-key Encryption using Random Functions

Consider the following private-key encryption scheme

1. Gen(): Return $\mathsf{sk} = f \xleftarrow{\$} \mathcal{F}_{m,n}$
2. $\mathrm{Enc}_f(m)$: Pick a random $r \xleftarrow{\$} \{0,1\}^m$. Return $(m \oplus f(r), r)$, where $m \in \{0,1\}^n$.
3. $\mathrm{Dec}_f(\widetilde{c}, \widetilde{r})$: Return $\widetilde{c} \oplus f(\widetilde{r})$.

Features. Suppose the messages $m_1, \ldots, m_u$ are encrypted as the cipher-texts $(c_1, r_1), \ldots, (c_u, r_u)$.

- As long as the $r_1, \ldots, r_u$ are all distinct, each one-time pad $f(r_1), \ldots, f(r_u)$ are uniform and independent of others. So, this encryption scheme is perfectly secure!
- The probability that any two of the randomness in $r_1, \ldots, r_u$ are <u>not</u> distinct is <u>very small</u> (We shall prove this later as "Birthday Bound")
- This scheme is a "state-less" encryption scheme. Alice and Bob do not need to remember any private state (except the secret-key sk)!

Bottleneck of using Random Functions

- The secret-key sk needs $n2^m$ bits to represent it, which is exponentially large.
- We shall replace "random functions" using "pseudorandom functions" to construct an encryption scheme with short keys that remains secure against computationally bounded adversaries!

Birthday Bound

- Suppose we have a set $S = \{s_1, s_2, \ldots, s_n\}$
- Suppose we sample an element x_1 uniformly at random from the set S.
- Replace this element back in the set S and sample an element x_2 uniformly at random from the set S
- This process of sampling is referred to as "sampling with replacement"
- Suppose we sampled elements $\{x_1, x_2, \ldots, x_k\}$
- We are interested in understanding how likely it is that there are two elements $x_i = x_j$, such that $i \neq j$. Intuitively, we are interested in finding the probability that k elements, when sampled uniformly at random from S (with replacement), encounter a collision

- Why are we studying this probability? Recall that earlier in this lecture, we noted that if all the random r's chosen in the encryption algorithm are distinct, then the encryption scheme remains secure against computationally bounded eavesdroppers. So, the probability that we are computing shall help us determine the length of the randomness so that it is highly unlikely to encounter collisions.
- Okay, let us start by studying the complementary event. We are interested in the event that all the samples $\{x_1, x_2, \dots, x_k\}$ are distinct
- Note that the probability that x_1 is distinct from all previous samples is 1
- Conditioned on the fact that $\{x_1\}$ is distinct, the probability that x_2 is distinct from all previous samples is $\left(1 - \frac{1}{n}\right)$

Birthday Bound III

- Conditioned on the fact that $\{x_1, x_2\}$ are distinct, the probability that x_3 is distinct from all previous samples is $\left(1 - \frac{2}{n}\right)$
- Extrapolating these observations, we can conclude the following. Conditioned on the fact that $\{x_1, x_2, \dots, x_{i-1}\}$ are distinct, the probability that x_i is distinct from all previous samples is $\left(1 - \frac{i-1}{n}\right)$
- So, using the chain rule, we can conclude the following. The following product is the probability that $\{x_1, \dots, x_k\}$ are all distinct.

$$1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)$$

- This expression is the product that we saw in the midterm. We shall use the fact that $\exp(-x) \approx 1 - x$ when $0 \leqslant x \ll 1$. This fact can be made more mathematically precise using Taylor's Remainder Theorem, which is beyond the scope of this course. So, in this course, we shall proceed by using $\exp(-x) \approx 1 - x$
- So, let us begin the manipulation of the expression above

$$
\begin{aligned}
& 1 \cdot \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \\
\approx\, & \exp(-0)\exp(-1/n)\exp(-2/n) \cdots \exp(-(k-1)/n) \\
=\, & \exp\left(-0 - \frac{1}{n} - \frac{2}{n} - \cdots - \frac{k-1}{n}\right) \\
=\, & \exp\left(-\frac{k(k-1)}{2n}\right) \approx \exp(-k^2/2n) = \exp(-k^2/2|S|)
\end{aligned}
$$

- Suppose we set $k = \sqrt{|S|}/100$. Substituting this value of k in the formula above, note that the probability that all the samples are distinct is $\approx \exp(-1/20000)$, which is very close to 1!
- Suppose we set $k = 100\sqrt{|S|}$. Substituting this value of k in the formula above, note that the probability that all the samples are distinct is $\approx \exp(-5000)$, which is very close to 0!
- Intuitively, it says that if $k \leqslant \sqrt{|S|}/100$, all samples are very likely to be distinct. On the other hand, if $k \geqslant 100\sqrt{|S|}$, it is highly unlikely that all samples are distinct (that is, there are two identical samples, or collision occurs)

A Numerical Example of the Birthday Bound

- Suppose we are picking uniform random strings from the set $\{0,1\}^n$
- Our objective is that 2^{1000} random samples have a collision with probability at most 2^{-80}
- What value of n should we use?

- So, we have $S = \{0,1\}^n$. The size of the set S is 2^n.
- The probability that k samples are all distinct is $\exp(-k^2/2|S|) = \exp(-k^2/2^{n+1})$. The problem states that we shall pick $k = 2^{1000}$ samples.
- Our objective is to have collision probability $\leqslant 2^{-80}$. The probability of all samples being distinct is $\geqslant 1 - 2^{-80}$.
- So, we have the following equation, and we need to solve for n

$$\exp(-k^2/2^{n+1}) = \exp(-2^{2000}/2^{n+1}) \geqslant 1 - 2^{-80} \approx \exp(-2^{-80})$$

- Solving this equation is left as an exercise