Fundamentals of Cryptography — Handout 1.

Yu Yu

Notations, probabilities, functions, etc.

Note: I don't write handouts for every lecture, but this one is too fundamental to be skipped. Inform me or Yong Gu of typos, errors, etc.

RANDOM VARIABLES, VALUES, SETS. We use capital letters (e.g. X, Y, A) for random variables, standard letters (e.g. x, y, a) for values, and calligraphic letters (e.g. $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{S}$) for sets (and events).

SET AND ITS OPERATIONS. A set can be defined by enumeration, e.g.,

$$\mathcal{S} \stackrel{\text{def}}{=} \{0, 1, 2, 3, 4, \ldots, 99\}$$

$|\mathcal{S}|$ refers to the cardinality of $\mathcal{S}$ (e.g., $|\mathcal{S}|$=100 for $\mathcal{S}$ defined above). $\mathcal{S}_1 \times \mathcal{S}_2$ refers to the concatenation of sets $\mathcal{S}_1$ and $\mathcal{S}_2$, i.e.,

$$\mathcal{S}_1 \times \mathcal{S}_2 \stackrel{\text{def}}{=} \{(s_1, s_2) : \ s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2 \} \ .$$

If $\mathcal{S}_1$ and $\mathcal{S}_2$ are identical, say they are the same as $\mathcal{S}$, then we use $\mathcal{S}^n$ as the shorthand for the concatenation of $\mathcal{S}$ n times, i.e.

$$\mathcal{S}^n \stackrel{\text{def}}{=} \overbrace{\mathcal{S} \times \mathcal{S} \times \cdots \times \mathcal{S}}^{n}$$

and the most frequently used one is $\{0,1\}^n$, namely, the set of all possible values of an n-bit string. By $\{0,1\}^*$ we refer to the set of all (arbitrarily long) binary strings, i.e.,

$$\{0,1\}^* = \{0,1\} \cup \{0,1\}^2 \cup \cdots \cup \{0,1\}^i \cup \cdots$$

PROBABILITY DISTRIBUTIONS. For any non-empty set $\mathcal{X}$ (called the **sample space**), a (discrete) probability distribution X, defined over set $\mathcal{X}$, refers to the rule that assigns a numeric value (i.e., the probability that $X = x$) to each outcome $x \in \mathcal{X}$. For example,

$\mathcal{X} \stackrel{\text{def}}{=} \{0, 1, 2\}$ define distribution X by rule $\Pr[X = 0] = 0.5$, $\Pr[X = 1] = 0.2$, $\Pr[X = 2] = 0.3$

Note that all the probabilities must sum to unity. We mention that '$\bigcap$', '$\bigcup$', '\' are set operators for intersection, union and minus respectively.

UNIFORM DISTRIBUTIONS AND RANDOM VARIABLES. For distribution X defined over set $\mathcal{X}$, we say that X is uniform (or flat) if for every possible outcome $x \in \mathcal{X}$ it holds that

$$\Pr[X = x] = \frac{1}{|\mathcal{X}|} \ .$$

We often use $U_{\mathcal{X}}$ to denote the uniform distribution over $\mathcal{X}$, and U_n (instead of $U_{\{0,1\}^n}$) to denote uniform distribution over $\{0,1\}^n$.
A **random variable** is a function that maps elements of the sample space to another set (usually, but not necessarily, the set of real numbers, denoted by $\mathbb{R}$). For example, consider a distribution X over $\{0,1\}^n$, we define random variable

$$\mathrm{HW}(x) \stackrel{\text{def}}{=} \text{“the number of 1's in } x \text{”}$$

so that the outcome of $\mathrm{HW}(x)$ is induced by probability distribution $x \leftarrow X$.
Note: We often use “random variables” and “probability distributions” interchangeably as (1) a probability distribution can be considered as a special random variable (where the mapping is the identity function); (2) a random variable (induced by a distribution) can be regarded as another distribution, e.g., $Y \stackrel{\text{def}}{=} \mathrm{HW}(X)$.

Events and independence. In probability theory, an event is a (finite) set of outcomes (a subset of the sample space) to which a probability is assigned. Typically, when the sample space is finite, any subset of the sample space is an event. For example, consider uniform distribution U_n for some even number n, define subset $\mathcal{X} \subset \{0,1\}^n$ to be the set of n-bit strings whose Hamming weight (i.e., the number of 1's) is $n/2$, i.e.,

$$\mathcal{X} \stackrel{\text{def}}{=} \{\ x \in \mathcal{X} : HW(x) = n/2\ \}$$

Then, the event, denoted by $\mathcal{E}$, that an outcome of U_n falls into subset $\mathcal{X}$, has probability

$$\Pr[\mathcal{E}] \ = \ \Pr[U_n \in \mathcal{X}] = \sum_{x \in \mathcal{X}} \Pr[U_{\mathcal{X}} = x] = \frac{|\mathcal{X}|}{2^n} = \frac{\binom{n}{n/2}}{2^n} \quad .$$

Independent events. We say that events $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent iff

$$\Pr[\mathcal{E}_1 \cap \mathcal{E}_2] = \Pr[\mathcal{E}_1] \cdot \Pr[\mathcal{E}_2]$$

or equivalently, $\Pr[\mathcal{E}_2] = \Pr[\mathcal{E}_2|\mathcal{E}_1]$.

Theorem 1 (Bayes' Theorem)**.** *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be events over the same sample space and that $\Pr[\mathcal{E}_2] \neq 0$. Then,*

$$\Pr[\mathcal{E}_1|\mathcal{E}_2] = \frac{\Pr[\mathcal{E}_1] \cdot \Pr[\mathcal{E}_2|\mathcal{E}_1]}{\Pr[\mathcal{E}_2]} \quad .$$

Independent random variables. We say that X and Y are independent random variables if for every possible values x and y the events $X = x$ and $Y = y$ are independent, i.e.,

$$\Pr[X = x, Y = y] = Pr[X = x] \cdot \Pr[Y = y] \ .$$

polynomial, super-polynomial and negligible functions.
A function $\mathsf{poly}(\cdot)$ is a **polynomial** (in a single indeterminate) iff it can be represented as

$$\mathsf{poly}(n) = a_c \cdot n^c \ + \ a_{c-1} \cdot n^{c-1} \ + \ a_{c-2} \cdot n^{c-2} \ + \ \cdots \ + \ a_1 \cdot n + a_0$$

where constants $a_c, \cdots, a_0$ are coefficients that uniquely define the polynomial and c is called degree of the polynomial.

A function $\mathsf{superpoly}(\cdot)$ is **super-polynomial** if for every constant $c > 0$ it holds that

$$\mathsf{superpoly}(n) \ > \ n^c$$

for **all sufficiently large** n's.

A function $\mathsf{negl}(\cdot)$ is **negligible** if for every constant $c > 0$ it holds that

$$\mathsf{negl}(n) \ < \ n^{-c}$$

for **all sufficiently large** n's. Superpolynomial and neligible functions are reciprocals of each other.

A function $\mu(\cdot)$ is **non-negligible** (i.e., not negligible) there exists constant $c > 0$ it holds that

$$\mu(n) \ \geq \ n^{-c}$$

for **infinitely many** n's.

A function $\mu(\cdot)$ is **noticeable** if there exists constant $c > 0$ it holds that

$$\mu(n) \ \geq \ n^{-c}$$

for **all sufficiently large** n's.
Note: Non-negligible is not the same as noticeable (you may try to come up an example function which is non-negligible but not noticeable).

FUNCTIONS. Informally, $f(\cdot)$, $g(\cdot, \cdot)$ denote functions that takes one and two inputs respectively. A more formal treatment will specify the domain, range and functionality. For example,

$$g : \{0,1\}^n \to \{0,1\}^n$$

$$g(x) \ \mapsto \ a \cdot x + b$$

where "+" and "$\cdot$" denote addition and multiplication over $\mathrm{GF}(2^n)$ respetively, and $a \in \{0,1\}^n$ and $b \in \{0,1\}^n$ are n-bit constants (interpreted as elements over $\mathrm{GF}(2^n)$) that define function g.

ASYMPTOTIC NOTATOINS. For functions $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$, we say

- $f = O(g)$ if there exists some constant c such that $f(n) \leq c \cdot g(n)$ for all sufficiently large n's.
- $f = \Omega(g)$ if $g = O(f)$.
- $f = \Theta(g)$ iff $f = O(g)$ and $g = O(f)$.
- $f = o(g)$ if for every $\epsilon > 0$ $f(n) \leq \epsilon \cdot g(n)$ for all sufficiently large n's
- $f = \omega(g)$ if $g = o(f)$.

FUNCTION ENSEMBLES. In cryptography, we are often talking about an ensemble of functions, carry out their computation using Turing machine (or equivalent) algorithms, and analyze the efficiency (or the inefficiency of breaking a cryptographic algorithm) using asymptotic notations such as $O(n)$, $\Theta(n^2)$ (or success probability negligble in n, where n is the length of the secret key). For (non-cryptographic) example, the function that "bitwise XOR two equal-length binary strings and outputs their XOR sum" can be represented as an ensemble of function (indexed by n)

$$\{\ f_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n\ \}_{n\in\mathbb{N}}$$

where each f_n handles n-bit strings a and b and outputs $a \oplus b$. It is easy to see that the function is very efficient, namely, computable by a Turing machine in polynomial time (more specifically, time $O(n)$).

Now we give the definition of a cryptographic function as example (we will use it in subsequent lectures)

Definition 1 (one-way functions)**.** $f \stackrel{\text{def}}{=} \{f_n : \{0,1\}^n \to \{0,1\}^{l(n)}\}_{n\in\mathbb{N}}$ is a one-way function (ensemble) if

- (Easy-to-Compute). f can be computed by some algorithm in time $\mathsf{poly}(n)$.
- (Hard-to-Invert). For any probabilistic polynomial (in n) time (PPT) A, there exists a negligible function $\mathsf{negl}(\cdot)$ such that

$$\Pr_{x\leftarrow U_n,\ x'\leftarrow \mathsf{A}(1^n, f(x))}[f(x') = f(x)] \ \le\ \mathsf{negl}(n).$$

  where $x \leftarrow U_n$ denotes sample a random x from U_n, and the above probability is taken over the choice of x over U_n and the internal coins of A (which is probabilistic).

Note: hereafter we will write $f : \{0,1\}^n \to \{0,1\}^{l(n)}$ instead of $f \stackrel{\text{def}}{=} \{f_n : \{0,1\}^n \to \{0,1\}^{l(n)}\}_{n\in\mathbb{N}}$ for simplicity.

Loosely speaking, the above primitive is easy to compute, but for any polynomial-time adversary who sees only its outputs (on random inputs), it will be hard for her to invert the function (i.e., to find any x' satisfying $f(x') = f(x)$), where the hardness refers to the fact that the success probability is negligible in n. Thus, we just need to set an appropriate value for n to get efficiency and security at the same time (a trade-off between them). It is one of the most fundamental primitives of modern cryptography.

USEFUL INEQUALITIES.

Theorem 2 (Union bound)**.** *if $\mathcal{S}$ is a sample space and $\mathcal{E}_1$, $\mathcal{E}_2 \subseteq \mathcal{S}$ are two events over $\mathcal{S}$. Then we have*

$$\Pr[\mathcal{E}_1 \cup \mathcal{E}_2] \ \le\ \Pr[\mathcal{E}_1] \ +\ \Pr[\mathcal{E}_2]$$

where the equality holds iff $\Pr[\mathcal{E}_1 \cap \mathcal{E}_2] = 0$.

Theorem 3 (Markov Inequality)**.** *Let X be any random variable that takes non-negative real numbers. Then, for any $\delta > 0$*

$$\Pr[\ X\ \ge\ \delta \cdot \mathbb{E}[X]\ ] \le \frac{1}{\delta}\quad .$$

where $\mathbb{E}[X]$ denotes the expectation of X.

Proof. Denote $\mu = \mathbb{E}[X]$ and let $\mathcal{X}$ be the sample space of X. Define set

$$\mathcal{X}_1 \stackrel{\text{def}}{=} \{x \in \mathcal{X} : \Pr[X = x] \geq \delta\mu\}$$

Then,

$$\mu \;=\; \mathbb{E}[X] \;= \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot x \;\geq\; \sum_{x \in \mathcal{X}_1} \Pr[X = x] \cdot x \;\geq\; \delta\mu \cdot \Pr[X \in \mathcal{X}_1]$$

and thus $\Pr[X \in \mathcal{X}_1] \leq 1/\delta$, which is essentially the statement desired. □

Theorem 4 (Chebyshev's inequality.)**.** *Let X be any random variable (taking real number values) with expectation μ and standard deviation σ (i.e., $Var[X] = \sigma^2 = \mathbb{E}[(X - \mu)^2]$). Then, for any $\delta > 0$ we have*

$$\Pr[\; |X - \mu| \;\geq \delta\sigma] \;\leq\; 1/\delta^2$$

Proof. Define non-negative random variable $Y \stackrel{\text{def}}{=} (X - \mu)^2$ with expectation $\mathbb{E}[Y] = \sigma^2$. Applying Markov inequality to Y yields

$$\Pr[Y \geq \delta^2\sigma^2] \leq 1/\delta^2$$

which completes the proof. □

We will often use the following versions of the Chernoff bound and (the more general) Hoeffding bound.

Theorem 5 (Chernoff bound)**.** *Let $X_1, \ldots, X_n$ be independent variables with $0 \leq X_i \leq 1$ for all $1 \leq i \leq n$, denote $\mu = \mathbb{E}[\; (\sum_{i=1}^n X_i)/n \;]$. Then, for any $\epsilon > 0$*

$$\Pr\left[\; \left| \frac{\sum_{i=1}^n X_i}{n} - \mu \right| \;>\; \epsilon \;\right] \;<\; 2^{-\epsilon^2 \cdot n} \quad .$$

Theorem 6 (Hoeffding bound)**.** *Let $X_1, \ldots, X_n$ be independent variables with $b_i \leq X_i \leq a_i$ for all $1 \leq i \leq n$, denote $\mu = \mathbb{E}[\; (\sum_{i=1}^n X_i)/n \;]$. Then, for any $\epsilon > 0$*

$$\Pr\left[\; \left| \frac{\sum_{i=1}^n X_i}{n} - \mu \right| \;>\; \epsilon \;\right] \;<\; 2\exp^{-\frac{2\epsilon^2 \cdot n^2}{\sum_{i=1}^n (b_i - a_i)^2}} \quad .$$

Other inequalities that are used in cryptography include the Jensen's inequality, Cauchy-Schwarz, etc. Ask Google or John for details and more.