Privacy Enhancing Technologies FS2025 Background

Florian Tramèr

1 Complexity

1.1 Asymptotic Notation

1.1.1 Big O Notation

Big O notation is used to describe the upper bound of the growth rate of a function.

Definition 1. We say that $f(n) = O(g(n))$ for functions $f, g : \mathbb{N} \to \mathbb{R}^+$ if $\exists c > 0, n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$:

$$f(n) \leq c \cdot g(n).$$

1.1.2 Little o Notation

Little o notation provides a strict upper bound, stronger than Big O.

Definition 2. We say that $f(n) = o(g(n))$ for functions $f, g : \mathbb{N} \to \mathbb{R}^+$ if $\forall c > 0, \exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$:

$$f(n) < c \cdot g(n).$$

1.1.3 Omega Notation

Omega notation describes the lower bound of the growth rate of a function.

Definition 3. We say that $f(n) = \Omega(g(n))$ for functions $f, g : \mathbb{N} \to \mathbb{R}^+$ if $\exists c > 0, n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$:

$$f(n) \geq c \cdot g(n).$$

1.2 P vs NP

1.2.1 Definitions

Let Σ be a finite alphabet and $\mathcal{L} \subseteq \Sigma^*$ be a language.

Definition 4 (P)**.** P is the class of languages decidable in polynomial time by a deterministic Turing machine. Formally:

$$\mathtt{P} = \bigcup_{k \in \mathbb{N}} \mathrm{TIME}(n^k)$$

where $\mathrm{TIME}(t(n))$ is the class of languages decidable by a deterministic Turing machine in $O(t(n))$ time.

Definition 5 (NP)**.** NP is the class of languages verifiable in polynomial time by a deterministic Turing machine. Formally, $\mathcal{L} \in \mathtt{NP}$ if there exists a deterministic, polynomial-time algorithm M and a polynomial p such that:

$$x \in \mathcal{L} \iff \exists w \in \{0,1\}^{\mathsf{poly}(|x|)} \text{ s.t. } M(x,w) = 1.$$

The element w is called a witness or certificate.

1.2.2 Relationship

We have $\mathtt{P} \subseteq \mathtt{NP}$, as any language decidable in polynomial time is also verifiable in polynomial time. The central question in complexity theory is whether $\mathtt{P} = \mathtt{NP}$ or $\mathtt{P} \neq \mathtt{NP}$.

2 Basic Cryptographic Primitives

We say a function $f(n)$ is negligible if $f(n) = o(n^{-c})$ for all constants $c \in \mathbb{N}$.

2.1 Pseudorandom Number Generator (PRG)

A PRG $G : \{0,1\}^\lambda \to \{0,1\}^\ell$ where $\ell > \lambda$ is a deterministic poly-time algorithm. It takes a short random seed $s \in \{0,1\}^\lambda$ and expands it into a long "random looking" string $G(s) \in \{0,1\}^\ell$ where $\ell > \lambda$. G is secure if for all poly-time distinguishers $\mathcal{D}$:

$$|\Pr[s \leftarrow\!\$\ \{0,1\}^\lambda : \mathcal{D}(G(s)) = 1] - \Pr[r \leftarrow\!\$\ \{0,1\}^\ell : \mathcal{D}(r) = 1]| \leq \mathrm{negl}(\lambda)$$

2.2 Pseudorandom Function (PRF)

A PRF is a deterministic algorithm that takes as input a key $k \in \mathcal{K}$ and an input $x \in \mathcal{X}$ and outputs a value $y \in \mathcal{Y}$. Let $\mathrm{Funcs}[\mathcal{X}, \mathcal{Y}]$ be the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$. A PRF $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ is secure if it "looks like" a random function from $\mathcal{X}$ to $\mathcal{Y}$. Formally, for all poly-time distinguishers $\mathcal{D}$:

$$|\Pr[k \leftarrow\!\$\ \mathcal{K} : \mathcal{D}^{F(k,\cdot)}() = 1] - \Pr[f \leftarrow\!\$\ \mathrm{Funcs}[\mathcal{X}, \mathcal{Y}] : \mathcal{D}^{f}() = 1]| \leq \mathrm{negl}(\lambda)$$

Here, the notation $\mathcal{D}^f$ means that the distinguisher is given oracle access (i.e., query access) to f.

2.3 Cryptographic Hash Function

A Cryptographic Hash Function is a (family of) function(s) $H_\lambda : \mathcal{M}_\lambda \to \mathcal{T}_\lambda$ from some large message space $\mathcal{M}_\lambda$ into a small digest space $\mathcal{T}_\lambda$, with the following properties:

- Collision resistance: for all poly-time adversaries, the probability that it can find a collision $m_1 \neq m_2 \in \mathcal{M}_\lambda$ such that $H_\lambda(m_1) = H_\lambda(m_2)$ is negligible in λ.
- Preimage resistance (one-wayness): Given $h = H_\lambda(m)$ for a randomly chosen $m \in \mathcal{M}_\lambda$, the probability that a poly-time adversary can find $m' \in \mathcal{M}_\lambda$ such that $H_\lambda(m') = h$ is negligible in λ.

2.4 Symmetric Encryption with Semantic Security

For a symmetric encryption scheme $(\texttt{Gen}, \texttt{Enc}, \texttt{Dec})$, semantic security means that an adversary cannot distinguish between encryptions of two different messages. This property is also called *indistinguishability under chosen plaintext attacks* (IND-CPA).

For every probabilistic polynomial time algorithm $\mathcal{A}$ and two arbitrary messages m_0, m_1, we have:

$$|\Pr[\mathcal{A}(\texttt{Enc}_k(m_0)) = 1 \ : \ k \leftarrow \texttt{Gen}()] - \Pr[\mathcal{A}(\texttt{Enc}_k(m_1)) = 1 \ : \ k \leftarrow \texttt{Gen}()]| \leq \text{negl}(\lambda)$$

3 Number Theory

3.1 Groups

A group $(\mathbb{G}, \cdot)$ is a set $\mathbb{G}$ with a binary operation "$\cdot$" satisfying:

- Closure: $\forall a, b \in \mathbb{G}, a \cdot b \in \mathbb{G}$
- Associativity: $\forall a, b, c \in \mathbb{G}, (a \cdot b) \cdot c = a \cdot (b \cdot c)$
- Identity: $\exists e \in \mathbb{G}$ such that $\forall a \in \mathbb{G}, \ e \cdot a = a \cdot e = a$
- Inverse: $\forall a \in \mathbb{G}, \ \exists a^{-1} \in \mathbb{G}$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$

3.2 Generator and Order

For a finite group $\mathbb{G}$, an element $g \in \mathbb{G}$ is a generator if $\{g^k : k \in \mathbb{Z}\} = \mathbb{G}$. The order of $\mathbb{G}$ is $|\mathbb{G}|$, the number of elements in $\mathbb{G}$.

3.3 Hardness Assumptions in Groups

The following problems are believed to be hard in some groups that are widely used in cryptography. Let g be a randomly chosen generator of a group $\mathbb{G}$ of prime order q.

Discrete Logarithm: Given a uniformly random element $h \in \mathbb{G}$, it is hard to find $x \in \mathbb{Z}_q$ such that $g^x = h$.

Computational Diffie-Hellman (CDH): Given $g^\alpha, g^\beta \in \mathbb{G}$ for uniformly random $\alpha, \beta \in \mathbb{Z}_q$, it is hard to compute $g^{\alpha\beta}$.

Decisional Diffie-Hellman (DDH): It is hard to distinguish between the distributions $(g^\alpha, g^\beta, g^{\alpha\beta})$ and $(g^\alpha, g^\beta, g^\gamma)$ for uniformly random $\alpha, \beta, \gamma \in \mathbb{Z}_q$.

3.4 Finite Fields

A finite field is a finite set $\mathbb{F}$ with two operations "$+$" and "$\cdot$", such that:

- Addition and multiplication are both associative and commutative.
- There exists an additive identity 0 and multiplicative identity 1.
- Every element has an additive inverse.
- Every non-zero element has a multiplicative inverse.
- Multiplication distributes over addition.

The integers $\mathbb{Z}_p$ modulo a prime p form a finite field.

4 Probability and Statistics

4.1 Expectation and Variance

For a discrete random variable X:

- Expectation: $\mathbb{E}[X] = \sum_x x \cdot \Pr[X = x]$
- Variance: $\mathbf{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

4.2 Probability Inequalities

- Union Bound: $\Pr[\bigcup_i A_i] \leq \sum_i \Pr[A_i]$
- Markov's Inequality: For non-negative X and $a > 0$, $\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$
- Chebyshev's Inequality: For a random variable X with mean μ and variance σ^2, and $k > 0$, $\Pr[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$
- Chernoff / Hoeffding Bound: Let $X_1, \ldots, X_n$ be independent random variables taking values in $[a, b]$. Let $X = \sum_i X_i$ and $\mu = \mathbb{E}[X]$. Then:

$$\Pr[|X - \mu| \geq t] \leq 2\exp\left(-\frac{2t^2}{n(b-a)^2}\right).$$

Equivalently,

$$\Pr\left[\left|\frac{X}{n} - \frac{\mu}{n}\right| \geq t\right] \leq 2\exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$

If we remove the absolute value, we can remove the factor 2 in front of the exponential.

4.3 Standard Probability Distributions

- Bernoulli: $\texttt{Ber}(p)$: $\Pr[X = 1] = p, \Pr[X = 0] = 1 - p$, $\mathbb{E}[X] = p$, $\mathbf{Var}[X] = p(1-p)$
- Binomial: $\texttt{Bin}(n, p)$: $\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, $\mathbb{E}[X] = np$, $\mathbf{Var}[X] = np(1-p)$
- Gaussian $\mathcal{N}(\mu, \sigma^2)$: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mathbb{E}[X] = \mu$, $\mathbf{Var}[X] = \sigma^2$