Privacy Enhancing Technologies
Background Sheet

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.** For functions $f, g : \mathbb{N} \to \mathbb{R}^+$, we say $f(n) = O(g(n))$ 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.** For functions $f, g : \mathbb{N} \to \mathbb{R}^+$, we say $f(n) = o(g(n))$ 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.** For functions $f, g : \mathbb{N} \to \mathbb{R}^+$, we say $f(n) = \Omega(g(n))$ 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 $L \subseteq \Sigma^*$ be a language.

Definition 4 (P). $\texttt{P}$ is the class of languages decidable in polynomial time by a deterministic Turing machine. Formally:

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

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

Definition 5 (NP). $\texttt{NP}$ is the class of languages verifiable in polynomial time by a deterministic Turing machine. Formally, $L \in \texttt{NP}$ if $\exists$ polynomial p and polynomial-time decidable relation $R \subseteq \Sigma^* \times \Sigma^*$ such that:

$$x \in L \iff \exists y \in \Sigma^*, |y| \leq p(|x|) : R(x, y)$$

Here, y is called a witness or certificate.

1.2.2 Relationship

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

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 (PRNG)

A function $G : \{0,1\}^s \to \{0,1\}^n$ where $n > s$ is a secure PRNG if for any probabilistic polynomial-time distinguisher D:

$$|\Pr[D(G(\mathcal{U}_s)) = 1] - \Pr[D(\mathcal{U}_n) = 1]| \leq \text{negl}(s)$$

where $\mathcal{U}_k$ denotes the uniform distribution over $\{0,1\}^k$.

2.2 Pseudorandom Generator (PRG)

A PRG 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 $\ell > \lambda$.

A PRG $G : \{0,1\}^\lambda \to \{0,1\}^\ell$ where $\ell > \lambda$ is a deterministic poly-time algorithm. It is secure if for all poly-time algorithms $\mathcal{A}$:

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

2.3 Cryptographic Hash Function

A Cryptographic Hash Function is a function $H : \{0,1\}^* \to \{0,1\}^n$ with the following properties:

- Collision resistance: It's computationally infeasible to find $x \neq y$ such that $H(x) = H(y)$.
- Preimage resistance: Given y, it's computationally infeasible to find x such that $H(x) = y$.

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 is also called *indistinguishability under chosen plaintext attacks*, or IND-CPA).

We can formalize this as follows. For every probabilistic polynomial time algorithm $\mathcal{A}$ and two arbitrary messages m_0, m_1, we have:

$$\Pr[\mathcal{A}(\texttt{Enc}_k(m_b)) = b \ : \ k \leftarrow \texttt{Gen}()] \leq \frac{1}{2} + \mathsf{negl}(n)(n)$$

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. In each case, let g be a randomly chosen generator of a group $\mathbb{G}$ of order q.

Discrete Logarithm: Given an 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 $g^a, g^b \in \mathbb{G}$ for random $a, b \in \mathbb{Z}_q$, it is hard to compute g^{ab}

Decisional Diffie-Hellman (DDH): It is hard to distinguish between the distributions (g^a, g^b, g^{ab}) and (g^a, g^b, g^c) for random $a, b, c \in \mathbb{Z}_q$.

3.4 Finite Fields

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

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

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 $\mathcal{N}(\mu, \sigma^2)$: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \mathbb{E}[X] = \mu, \mathbf{Var}[X] = \sigma^2$