Lecture 14: Hardness Assumptions

Instructor: Omkant Pandey

Spring 2017 (CSE 594)

Today

- Some background
- Some hardness assumptions
  - Discrete logarithm
  - RSA
  - LWE
- Scribe notes volunteers?

Modular arithmetic

- $\mathbb{N}$ and $\mathbb{R}$ set of natural and real numbers respectively.
- $\mathbb{Z}$ = set of integers, $\mathbb{Z}^+, \mathbb{Z}^-$ for +ve and -ve integers.
- For $n \in \mathbb{N}$, $\mathbb{Z}_N$ denotes set of integers **modulo** N; i.e.:

$$\mathbb{Z}_N := \{0, 1, 2, \ldots, N-1\}$$

- We can perform "arithmetic in $\mathbb{Z}_N$":

* if we divide integer a by N, the **remainder** (say r) is in $\mathbb{Z}_N$; we write $r = a \mod N$.

* Addition becomes $(a + b) \mod N = (a \mod N) + (b \mod N) \mod N$

* Multiply becomes $(a \times b) \mod N = (a \mod N) \times (b \mod N) \mod N$

- We say that "a is *congruent* to b modulo N" if a, b have the same remainder and write:

$$a \equiv b \mod N$$

- $a \equiv 0 \mod N$ if and only $N|a$ ("N divides a").

Greatest Common Divisor (GCD)

- If a, b are two integers, $\gcd(a, b)$ denotes their greatest common divisor.
- a, b are **relatively prime** if they are non-zero and have no common factors, i.e., $\gcd(a, b) = 1$
- gcd is easy to compute for any two integers a, b.
- Extended Euclidean: $\forall a, b \in \mathbb{Z}$ there exist integers $x, y \in \mathbb{Z}$ (which are also easy to compute) s.t. $ax + by = \gcd(a, b)$.
- If a, b are relatively prime then $ax + by = 1$. $\implies ax \equiv 1 \mod b$.
- $\mathbb{Z}_N^*$ = set of integers mod N that are relatively prime to N:

$$\mathbb{Z}_N^* = \{1 \leqslant x \leqslant N - 1 : \gcd(x, N) = 1\}.$$

$\implies \forall a \in \mathbb{Z}_N^* \; \exists x : ax = 1 \mod N.$

- Such an x is called the **inverse** of a.

Integers modulo a prime

- Of special interest is the case when N is a prime number, say p.
- This defines:

$$\begin{aligned}
\mathbb{Z}_p &= \{0, 1, 2, \ldots, p-1\} \\
\mathbb{Z}_p^* &= \{1 \leqslant x \leqslant p-1 : \gcd(x, p) = 1\} \\
&= \{1, 2, \ldots, p-1\} \\
|\mathbb{Z}_p^*| &= p-1.
\end{aligned}$$

Fermat's Little Theorem

If p is a prime, then for any $a \in \mathbb{Z}_p^*$:

$$a^{p-1} \mod p = 1.$$

Euler's generalization

- Recall: $\mathbb{Z}_N^*$ = integers mod N that are relatively prime to N

$$\mathbb{Z}_N^* = \{1 \leqslant x \leqslant N-1 : \gcd(x, N) = 1\}.$$

- Euler's theorem: for any $N \in \mathbb{N}$ and $a \in \mathbb{Z}_N^*$:

$$a^{\phi(N)} \mod N = 1.$$

where $\phi(N)$ is Euler's totient function: $\phi(N) = |\mathbb{Z}_N^*|$.

- Fundamental Theorem of Arithmetic: every integer N can be written as

$$N = \prod_{i=1}^{k} p_i^{e_i}$$

for primes $p_1 < p_2 < \ldots < p_k$ (called factors) and positive integers $e_i > 0$. This factorization is unique (with empty product taken to be 1).

$$\phi(N) = N \cdot \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)$$

- If $N = pq$ for distinct primes p, q, then $\phi(N) = (p-1) \cdot (q-1)$.

Groups

- Groups: a set G with a "group operation" $\vdots G \times G \to G$ satisfying:
  - Closure: $\forall a, b \in G$, $a \odot b \in G$,
  - Identity: $\exists e \in G$ (idenitity) s.t. $\forall a \in G$: $a \odot e = e \odot a = a$.
  - Associativity: $\forall a, b, c \in G$: $(a \odot b) \odot c = a \odot (b \odot c)$.
  - Inverse: $\forall a \in G$ $\exists b \in G$ s.t. $a \odot b = b \odot a = e$ (identity).
- (Abelian group): a group with *commutative* property — $\forall a, b \in G$: $a \odot b = b \odot a$.
- Examples: $(\mathbb{Z}_N, +)$, $(\mathbb{Z}_N^*, \times)$ are "additive" and "multiplicative" groups for all N.
- (Corollary of Lagrange's Theorem): $\mathbf{x}^{|\mathbf{G}|} = \mathbf{e}$.
- (Generator): $g \in G$ is a *generator* of G if the set $\{g, g^2, \ldots\} = G$. The set of of all generators of G will be denoted by Gen_G.

Discrete Logarithm Problem

- Roughly speaking: given (p, g, y) such that p is a large prime, $g, y \in \mathbb{Z}_p^*$ find x such that $y = g^x \mod p$.
- Not hard for many cases, e.g., if $g = 1$, or p is a "special prime", e.g., if $p - 1$ has small factors.
- However, if g is a *generator* the problem is believed to be hard.
- Normally we want to work with a group such that $|G|$ = number of elements in G is **prime**. ($|G|$ is also called the *order* of the group)
- $\mathbb{Z}_p^*$ has $p - 1$ elements which is not prime.
- However, suppose that $p = 2q + 1$ and q is also a prime. Such primes are called "safe primes"
- Now consider a subset $G_q = \{x^2 : x \in \mathbb{Z}_p^*\}$. It is easy to prove that G_q is a group of prime order q.

Discrete Logarithm Problem (continued)

- This means that you can cycle through all q elements of G by applying the group operation to the generator over and over again.
- There are other ways to construct prime order groups, e.g., group formed by points on an appropriate elliptic curve.
- Hard to compute discrete log in prime order groups...

Assumption (Discrete Log Assumption)

If G_q is a group of prime order q then for every non-uniform PPT $\mathcal{A}$ there exists a negligible function μ s.t.:

$$\Pr\left[q \leftarrow \Pi_n; g \leftarrow \mathsf{Gen}_{G_q}; x \leftarrow \mathbb{Z}_q : \mathcal{A}(1^n, g^x) = x\right] \leqslant \mu(n).$$

- Note: not true for all groups, but there are groups where it is believed to be hard.

Diffie-Hellman Problems

- The adversary gets $X = g^x \mod p$, and $Y = g^y \mod p$ and (p, g).
- The Computational Diffie-Hellman (CDH) problem is as follows:

  Given (g, q, g^x, g^y), compute $g^{xy} \in G_q$ where x, y are random and all computations are in G_q.
- When working with a safe $p = 2q + 1$, g can be generator for order q subgroup, and computations can be modulo p.
- CDH Assumption: $\forall$ non-uniform PPT A, $\exists$ negligible μ s.t. $\forall n$: A solves the CDH problem with probability at most $\mu(n)$.

Diffie-Hellman Problems

- In fact, g^{xy} "looks indistinguishable" from a random group element
- Roughly, the **Decisional Diffie-Hellman** problem is:

  Distinguish (g, p, g^x, g^y, g^{xy}) from (g, p, g^x, g^y, g^z) where (x, y, z) are random and all computations are in G_q.
- DDH Assumption: $\forall$ non-uniform PPT "distinguishers" D, $\exists$ negligible μ s.t. $\forall n$: D solves the DDH problem with probability at most $\frac{1}{2} + \mu(n)$.

RSA Function and RSA Assumption

- RSA = Rivest, Shamir, Adleman
- Let p, q be large random primes of roughly the same size.
- Let $N = pq$. N is called a RSA modulus.
- Recall that $\phi(N) = (p-1)(q-1)$
- Recall that: $\phi(N) = |\mathbb{Z}_N^*|$ where:

$$\mathbb{Z}_N^* = \left\{x \in \mathbb{Z}_N : \gcd(x, N) = 1\right\}$$

RSA Function and RSA Assumption

- Let e be an odd number between 1 and $\phi(N)$ such that

$$\gcd(e, \phi(N)) = 1$$

  Therefore, $e \in \mathbb{Z}^*_{\phi(N)}$.

- Let d be such that:

$$e \cdot d = 1 \mod \phi(N).$$

- If $\phi(N)$ is known, you can compute d.
- If $\phi(N)$ is not known, d seems hard to compute!
- Therefore, $\phi(N)$ must be kept secret.

RSA Function and RSA Assumption

- Let N, e, d be as before so that $e \cdot d = 1 \mod \phi(N)$.
- d can be used to compute e**-th root** of numbers modulo N.

  Suppose that $y = x^e \mod N$, then:

$$\begin{aligned} y^d \mod N &= x^{ed} \mod N \\ &= x^{ed \mod \phi(N)} \mod N \\ &= x \mod N. \end{aligned}$$

- Without d, it seems hard to compute e-th roots mod N. (RSA Assumption)
- We can publish (N, e), and it would be hard to compute e-th roots!
- Furthermore, we can use d as a secret trapdoor!

RSA Function and RSA Assumption

Definition (RSA Assumption)

For every non-uniform PPT A there exists a negligible function μ such that for all $n \in \mathbb{N}$:

$$\Pr\left[\begin{array}{l} p, q \leftarrow \Pi_n; N \leftarrow pq; \\ e \leftarrow \mathbb{Z}^*_{\phi(N)}; y \leftarrow \mathbb{Z}^*_N; \\ x \leftarrow A(N, e, y) \end{array} \quad : \quad x^e = y \mod N\right] \leqslant \mu(n)$$

- RSA Function: for N, e as above, the following is called the RSA function
$$f_{N,e}(x) = x^e \mod N$$
- The RSA Function actually yields a **collection of trapdoor one-way permutations**. (Later class)

Learning With Errors (LWE)

- Let $s = (s_1, \ldots, s_n) \in Z_q^n$ some modulus q and a parameter n.
- Suppose you are given many equations for known "a" values:

$$\begin{aligned} a_1 \cdot s_1 + a_2 \cdot s_2 + \ldots + a_n \cdot s_n &= b_1 (\mod q) \\ a'_1 \cdot s_1 + a'_2 \cdot s_2 + \ldots + a'_n \cdot s_n &= b_2 (\mod q) \\ \text{etc.} \end{aligned}$$

- You can solve this by Gaussian elimination.
- However, if the equations **contain errors**, this may not work!

Learning With Errors (LWE)

- In particular, if you add independent error to each equation distributed according to the Normal Distribution with standard deviation $\alpha q > \sqrt{n}$, the problem is believed to be hard.

$$\begin{aligned} a_1 \cdot s_1 + a_2 \cdot s_2 + \ldots + a_n \cdot s_n &\approx b_1(\mod q) \\ a'_1 \cdot s_1 + a'_2 \cdot s_2 + \ldots + a'_n \cdot s_n &\approx b_2(\mod q) \\ \text{etc} & \end{aligned}$$