1 Plan for lectures 26

- So far in the course, we have looked at private (symmetric) key cryptography. This includes primitives such as symmetric key encryption and MACs. The main limitation is that the parties need to share a common key. Today, we will start our discussion of public key cryptography with one of the most basic questions: can Alice and Bob share a key without meeting in-person? Alice and Bob execute some (well-defined) protocol, they exchange messages. At the end of the protocol, both have a common key. However, the adversary should not learn anything about the key, even though it can see all the communication between Alice and Bob.

- We will discuss some basic number theory, which will be needed for discussing the above protocol, as well as the primitives that we will discuss in future lectures.

2 The key exchange problem

Diffie and Hellman, in their seminal work titled 'New Directions in Cryptography', introduced the notions of public key encryption and digital signatures. They also gave a solution for the key exchange problem, which is now known as the Diffie-Hellman protocol. The solution relies on the hardness of a number-theoretic problem.

First, we describe the idea at a high level; we will go into the details next lecture.

Alice and Bob wish to choose a common key such that the adversary has no information about the key. Alice chooses a prime p from some distribution. Next, she chooses a number $g \in \{0, 1, \ldots, p-1\}$, a number $a \in \{0, 1, \ldots, p-1\}$, and sends $(p, g, v = g^a \mod p)$ to Bob. Bob chooses a number $b \in \{0, 1, \ldots, p-1\}$ and sends $w = g^b \mod p$. Alice can compute $K_1 = w^a \mod p$, Bob can compute $K_2 = v^b \mod p$.

Correctness of the protocol: At the end of the protocol, both Alice and Bob have the same key, since $w^a \mod p = v^b \mod p = g^{a \cdot b} \mod p$. The correctness proof uses only one elementary property of modular arithmetic: $\alpha \cdot \beta \mod p = ((\alpha \mod p) \cdot (\beta \mod p) \mod p)$.

Security: A key agreement protocol is secure against passive ('read-only') adversaries if the adversary learns nothing about the key, even after seeing the entire communication. In the above protocol, the adversary sees $p, g, g^a \mod p, g^b \mod p$, and wants to learn something about $g^{a \cdot b} \mod p$.

How should Alice and Bob choose p, g, a, b?

- Clearly, the adversary should not learn either a or b given $g^a \mod p$ and $g^b \mod p$. If the adversary can learn either a or b, then it can compute $g^{a \cdot b} \mod p$. Therefore, a and b must be chosen from a sufficiently large set.
- As a result, p must be a large prime.
- There are some bad choices of (p, g) that we must avoid. For instance, suppose we consider a large prime $p = 631$, and take $g = 587$. Then $g^2 \mod p = 43$, but $g^3 \mod p = 1$. Such a (p, g) combination is bad, since $g^{a \cdot b} \mod p$ can only take three values (1, 587 and 43). We must ensure that such p, g are not chosen.

3 Basic Number Theory-Module 1

The focus of this module will be arithmetic modulo prime.

Recall, we defined the set $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ and $\mathbb{Z}_p^* = \{1, 2, \ldots, p-1\}$. The set $\mathbb{Z}_p^*$, together with the operation 'multiplication modulo p' has a number of interesting properties. Let us represent the operation 'multiplication modulo p' using the symbol $\times_p$. Here are a few basic properties:

- (Property 1) Take any two numbers a, b in $\mathbb{Z}_p^*$. Then $a \times_p b$ is also in $\mathbb{Z}_p^*$.

- (Property 2) Take any number $a \in \mathbb{Z}_p^*$. There exists a unique number $b \in \mathbb{Z}_p^*$ such that $a \times_p b = 1$. This follows from the extended Euclidean algorithm. Recall, extended Euclidean algorithm states that for any positive integers v, w, there exist integers x, y such that $v \cdot x + w \cdot y = \gcd(v, w)$. Set $v = a$, $w = p$, and Euclid's algorithm gives us integers x, y such that $a \cdot x + p \cdot y = \gcd(a, p) = 1$ (since $a \in \mathbb{Z}_p^*$). Taking mod p on both sides, we get $b = x \mod p$ such that $a \times_p b = 1$. If there were two distinct integers b_1 and b_2 in $\mathbb{Z}_p^*$ such that $a \times_p b_1 = a \times_p b_2 = 1$, then $a \times_p |b_1 - b_2| = 0$. This violates Property 1 since $|b_1 - b_2| \in \mathbb{Z}_p^*$ but the product of a and $|b_1 - b_2|$ is not in $\mathbb{Z}_p^*$.
- (Property 3) Take any number $a \in \mathbb{Z}_p^*$. Consider the set $S_a = \{a \times_p i \ : \ i \in \mathbb{Z}_p^*\}$. Then $S_a = \mathbb{Z}_p^*$. If S_a is strictly smaller than $\mathbb{Z}_p^*$, then there exist two elements $r, s \in \mathbb{Z}_p^*$ such that $a \times_p r = a \times_p s$. Again, this violates Property 1.

Using multiplication, we can define exponentiation modulo p. Note: even if p is an n-bit prime, for any $a \in \mathbb{Z}_p^*$ and $x \in \mathbb{Z}_p$, we can compute $a^x \mod p$ efficiently (in time polynomial in n) using repeated squaring.

Next, we derive a useful theorem in number theory, called Fermat's Little Theorem.

Theorem 26.01. For any prime p and $a \in \mathbb{Z}_p^*$, $a^{p-1} = 1 \mod p$.

Proof. The proof of this theorem follows from Property 3. We know that $S_a = \mathbb{Z}_p^*$. Consider the following two products: $\alpha_1 = \left(\prod_{y \in S_a} y\right) \mod p$ and $\alpha_2 = \left(\prod_{i \in \mathbb{Z}_p^*} i\right) \mod p$. Since $S_a = \mathbb{Z}_p^*$, these two products are equal. As a result,

$$\begin{aligned}
\alpha_1 &= \left(\prod_{y \in S_a} y\right) \mod p \\
&= \left(\prod_{i \in \mathbb{Z}_p^*} a \cdot i \mod p\right) \mod p \\
&= \left(a^{p-1} \cdot \left(\prod_{i \in \mathbb{Z}_p^*} i\right)\right) \mod p \\
&= \left(\left(a^{p-1} \mod p\right) \cdot \left(\prod_{i \in \mathbb{Z}_p^*} i \mod p\right)\right) \mod p \\
&= \left(a^{p-1} \mod p\right) \times_p \alpha_2 = \left(a^{p-1} \mod p\right) \times_p \alpha_1
\end{aligned}$$

The proof follows from Property 2. □

3.1 Generators of $\mathbb{Z}_p^*$

In this subsection, we will consider special subsets of $\mathbb{Z}_p^*$, which are 'generated' by an element of $\mathbb{Z}_p^*$. Fix any $a \in \mathbb{Z}_p^*$, and consider the set $\langle a \rangle_p = \{1, a, a^2, \ldots\}$ (here everything is modulo p). We will establish some properties of these sets, known as 'subgroups' of $\mathbb{Z}_p^*$.

Let us start with an example: $\mathbb{Z}_{13}^*$. Consider the following set:

$$\langle 2 \rangle_{13} = \left\{2^0, 2^1, 2^2, \ldots, \right\}$$

where all the exponentiations are happening modulo 13. Clearly, the size of this set cannot be larger than 12 (using Property 1, this is a subset of $\mathbb{Z}_{13}^*$). However, for any $a \in \mathbb{Z}_p^*$, the set $\langle a \rangle_p$ has some structure, and we will explore that below, starting with a few examples.

- $\langle 2 \rangle_{13} = \mathbb{Z}_p^*$
- $\langle 3 \rangle_{13} = \{1, 3, 9\}$
- $\langle 5 \rangle_{13} = \{1, 5, 12, 8\}$
- $\langle 12 \rangle_{13} = \{1, 12\}$
- $\langle 10 \rangle_{13} = \{1, 10, 9, 12, 3, 4\}$

Note that the size of all these subsets of $\mathbb{Z}_p^*$ divides 12. This is no coincidence.

Theorem 26.02. Fix any prime p. For any $a \in \mathbb{Z}_p^*$, there exists a divisor d of $(p-1)$ such that $|\langle a \rangle_p| = d$.

Proof. Suppose z is the **smallest** positive number such that $a^z = 1 \mod p$. Clearly, $z \leq p-1$ (using Fermat's Little Theorem). Suppose $z \nmid (p-1)$. Then, $p-1 = q \cdot z + r$ for some integer $0 < r < z$.

$$\begin{aligned}
1 &= a^{p-1} \mod p \\
&= a^{q \cdot z + r} \mod p \\
&= ((a^{q \cdot z} \mod p) \cdot (a^r \mod p)) \mod p \\
&= (a^r \mod p)
\end{aligned}$$

In the last step, we use the observation that if $a^z = 1 \mod p$, then $a^{z \cdot q} = 1 \mod p$. Since $a^r \mod p = 1$ and $0 < r < z$, it violates the assumption that z is the smallest number s.t. $a^z = 1 \mod p$.

□

The converse of Theorem 26.02 also holds, but we will not prove it here. You can assume it as a fact.

Theorem 26.03. Fix any prime p. For any divisor d of $p-1$, there exists $a \in \mathbb{Z}_p^*$ such that $|\langle a \rangle_p| = d$.

As a corollary, it follows that for any prime p, there exists an integer $a \in \mathbb{Z}_p^*$ such that $\langle a \rangle_p = \mathbb{Z}_p^*$. Such an integer is called a generator of $\mathbb{Z}_p^*$.

Definition 26.01. Take any prime p. A number $g \in \mathbb{Z}_p^*$ is called the generator of $\mathbb{Z}_p^*$ if $\langle g \rangle_p = \mathbb{Z}_p^*$.

4 Lecture summary, plan for next lecture, additional resources

Summary: We first saw Diffie-Hellman's key agreement protocol at a high level. The security of this protocol relies on the hardness of number-theoretic problems. We started discussing modular arithmetic, in particular arithmetic modulo primes. We saw the following useful theorems/facts (below, p is a prime):

- For any $a \in \mathbb{Z}_p^*$, $a^{p-1} \mod p = 1$.
- For any $a \in \mathbb{Z}_p^*$, let $\langle a \rangle_p = \{1, a, a^2, \ldots\} \subseteq \mathbb{Z}_p^*$. The size of this set is a divisor of $p-1$.
- For any divisor d of $p-1$, there exists an $a \in \mathbb{Z}_p^*$ such that $|\langle a \rangle_p| = d$. In particular, there exists an $a \in \mathbb{Z}_p^*$ such that $|\langle a \rangle_p| = p-1$. Such an element is called a generator of $\mathbb{Z}_p^*$.

Next lecture: Next lecture, we will see that if p is appropriately chosen, then $\mathbb{Z}_p^*$ has many generators, and therefore we can pick a uniformly random element of $\mathbb{Z}_p^*$, and it will be a generator with good probability. Next, we will discuss the first number-theoretic problem that is believed to be computationally hard: the discrete logarithm problem. We will see a few applications of the discrete logarithm problem, and then discuss the key agreement protocol in some more detail.