Encryption Schemes

Notes by Yael Kalai

MIT - 6.5610
Lecture 3 (February 12, 2024)

Warning: This document is a rough draft, so it may contain bugs. Please feel free to email me with corrections.

Outline

- Definition of symmetric encryption
- Construction of symmetric encryption from any PRF family

Defining an Encryption Scheme

Last time Henry defined and constructed block ciphers (DES and AES). Recall that a block cipher is a pseudo-random permutation (PRP). In what follows, we first recall the definition of a pseudorandom function (PRF) and PRP.

Definition 1 (Pseudorandom function)**.** A pseudorandom function family consists of a family of functions $\{F_\lambda\}_{\lambda \in \mathbb{N}}$, where for every $\lambda \in \mathbb{N}$, $F_\lambda : \mathcal{K}_\lambda \times \mathcal{X}_\lambda \to \mathcal{Y}_\lambda$, and for every PPT algorithm $\mathcal{A}$ there exists a negligible function $\mu(\cdot)$ such that for every $\lambda \in \mathbb{N}$,

$$|\Pr[\mathcal{A}^{F_\lambda(k,\cdot)}(1^\lambda) = 1] - \Pr[\mathcal{A}^{R_\lambda(\cdot)}(1^\lambda) = 1]| \leq \mu(\lambda)$$

where $k \overset{R}{\leftarrow} \mathcal{K}_\lambda$ and $R_\lambda : \mathcal{X}_\lambda \to \mathcal{Y}_\lambda$ is a truly random function; $\mathcal{A}$ has oracle access to $F_\lambda(k,\cdot)$ or $R_\lambda(\cdot)$, and can make arbitrary oracle calls to its function. These oracle calls $x_1, \ldots, x_t \in \mathcal{X}_\lambda$ can be adaptively chosen based on the values returned by the oracle thus far.

λ is the security parameter. The larger the security parameter the more secure the scheme is, but also the less efficient it is.

In complexity theory, we model an efficient algorithm as polynomial time (or probabilistic polynomial time). We think of negligible as "practically never." $\mathcal{A}$ takes as input 1^λ since it is a PPT algorithm, and we allow it to run in time poly(λ). This is a notational hack used by theoreticians.

For concreteness, we can think of $\mathcal{X} = \{0,1\}^n$ and $\mathcal{Y} = \{0,1\}^m$.

Definition 2. A function $\mu : \mathbb{N} \to \mathbb{N}$ is said to be negligible if for every $c \in \mathbb{N}$ there exists $n_c \in \mathbb{N}$ such that for every $n > n_c$ it holds that $\mu(n) < n^{-c}$.

A pseudorandom permutation (PRP) family is the same as a PRF family except that $F_\lambda(k,\cdot) : \mathcal{X}_\lambda \to \mathcal{X}_\lambda$ is a permutation and there is an efficient algorithm for computing its inverse $F_\lambda^{-1}(k,\cdot)$. Why do we care about PRF and PRP families? What are they good for?

Use a PRP *for encryption?*

It is tempting to use a PRP for encryption. Let us first define the syntax of an encryption scheme. We start with symmetric encryption (also known as secret-key encryption).

Definition 3. [Take 1:] A symmetric encryption scheme is associated with a key space $\{\mathcal{K}_\lambda\}_{\lambda\in\mathbb{N}}$, a message space $\{\mathcal{M}_\lambda\}_{\lambda\in\mathbb{N}}$ and a ciphertext space $\{\mathcal{C}_\lambda\}_{\lambda\in\mathbb{N}}$, and with two algorithm $(\mathsf{Enc},\mathsf{Dec})$, where

$$\mathsf{Enc}_\lambda : \mathcal{K}_\lambda \times \mathcal{M}_\lambda \rightarrow \mathcal{C}_\lambda$$

and

$$\mathsf{Dec}_\lambda : \mathcal{K}_\lambda \times \mathcal{C}_\lambda \rightarrow \mathcal{M}_\lambda$$

such that for every $\lambda \in \mathbb{N}$, every $m \in \mathcal{M}_\lambda$, and every $k \in \mathcal{K}_\lambda$,

$$\mathsf{Dec}_\lambda(k, \mathsf{Enc}_\lambda(k, m)) = m.$$

What about security? How do we define security? Before giving a security definition, lets first consider the following natural way for using a PRP for encryption:

$$\mathsf{Enc}_\lambda(k, m) = F_\lambda(k, m) \quad \text{and} \quad \mathsf{Dec}_\lambda(k, c) = F_\lambda^{-1}(k, c)$$

where the message space and ciphertext space is $\mathcal{X}_\lambda$. This scheme is simple and nice but it does not have the security guarantees we would like, even if F is an ideal PRP. The reason is that the scheme is *deterministic* so an adversary can tell if the same message is encrypted twice. This may leak sensitive information.

Defining Security

When defining security one needs to define what the adversarial goal is and what is its power. In the case of an encryption scheme, the adversarial goal is to break the encryption of *any* message. A weaker goal would be to break the encryption of random messages. This may be too weak since in practice we do not encrypt random messages. Therefore, instead we allow the encryption algorithm to be *randomized*. We therefore need to restate the completeness condition in Definition 3, as follows: For every $\lambda \in \mathbb{N}$ and for every $m \in \mathcal{M}_\lambda$ and every $k \in \mathcal{K}_\lambda$,

$$\Pr[\mathsf{Dec}_\lambda(k, \mathsf{Enc}_\lambda(k, m)) = m] = 1$$

where the probability is over the random coin tosses of Enc. How about the following security definition.

If we only have security for random messages then one can encrypt a message m by choosing a random message $r \overset{R}{\leftarrow} \mathcal{M}$ and outputting $(\mathsf{Enc}(k, r), r \oplus m$. This can be thought of as a way of enhancing the security of an encryption scheme.

Sometimes this completeness condition is weakened and the probability is allowed to be $1 - \text{negl}(\lambda)$.

Definition 4 (Take 1)**.** An encryption scheme is said to be secure if for every $\lambda \in \mathbb{N}$ and for every messages $m_0, m_1 \in \mathcal{M}_\lambda$ it holds that

$$\mathsf{Enc}(k, m_0) \approx \mathsf{Enc}(k, m_1)$$

where $k \overset{R}{\leftarrow} \mathcal{K}_\lambda$ and where $\mathsf{Enc}_\lambda(k, m_b)$ is a random variable distributed over the random coins of Enc_λ.

Definition 5. Let $\mathcal{A} = \{\mathcal{A}_\lambda\}_{\lambda\in\mathbb{N}}$ and $\mathcal{B} = \{\mathcal{B}_\lambda\}_{\lambda\in\mathbb{N}}$ be two families of distributions. We say that $\mathcal{A}$ and $\mathcal{B}$ are computationally indistinguishable, denoted by $\mathcal{A} \approx \mathcal{B}$, if for every PPT distinguisher $\mathcal{D}$ there exists a negligible function μ such that for every $\lambda \in \mathbb{N}$,

$$|\Pr[\mathcal{D}(a) = 1] - \Pr[\mathcal{D}(b) = 1]| \leq \mu(\lambda)$$

where $a \leftarrow \mathcal{A}_\lambda$ and $b \leftarrow \mathcal{B}_\lambda$.

We denote by $a \leftarrow \mathcal{A}_\lambda$ if a is sampled from the distribution $\mathcal{A}_\lambda$. We denote by $k \overset{R}{\leftarrow} \mathcal{K}_\lambda$ is k is randomly chosen from the set $\mathcal{K}_\lambda$.

Is Definition 4 strong enough? Yes, if all the adversary has access to is this one ciphertext. However, what if the adversary has more information about the secret key k? For example, maybe the adversary sees many ciphertexts? Note that these cirphertexts are functions of the secret key and therefore may leak information about the secret key. We therefore strengthen the security property given in Definition 4, as follows.

Definition 6. An encryption scheme $(\mathsf{Enc}, \mathsf{Dec})$ is said to be secure against *chosen plaintext attacks* (CPA secure) if for every PPT adversary $\mathcal{A}$ there exists a negligible function μ such that for every $\lambda \in \mathbb{N}$, $\mathcal{A}$ wins in the following game with probability at most $\frac{1}{2} + \mu(\lambda)$:

- The challenger chooses a key $k \leftarrow \mathcal{K}_\lambda$.
- The adversary $\mathcal{A}$ given 1^λ chooses a message $m_i \in \mathcal{M}_\lambda$ and receives $c_i \leftarrow \mathsf{Enc}_\lambda(k, m_i)$.

  This step can be repeated polynomially many times.
- The adversary $\mathcal{A}$ chooses $m_0, m_1 \in \mathcal{M}_\lambda$.
- The challenger chooses a random bit $b \leftarrow \{0, 1\}$, generates $c \leftarrow \mathsf{Enc}(k, m_b)$, and sends the ciphertext c to the adversary.
- The adversary given c outputs a bit b'.

We say that $\mathcal{A}$ wins if $b' = b$.

This seems like a super strong security guarantee! However, the golden standard definition is even stronger! It also allows the adversary to see decryptions of ciphertexts of its choice. This is referred to as security against chosen ciphertext attacks (CCA-security).

This definition ensures security even if the adversary can obtain a ciphertext corresponding to any message of its choice!

Using a PRF *to construct a* CPA-*secure encryption scheme*

Let $F = \{F_\lambda\}_{\lambda\in\mathbb{N}}$ be any PRF family where $F_\lambda : \mathcal{K}_\lambda \times \mathcal{X}_\lambda \to \mathcal{Y}_\lambda$. Suppose $\mathcal{Y} = \{0, 1\}^{m(\lambda)}$. We use F to construct a symmetric encryption scheme where the key-space is $\mathcal{K}_\lambda$, the message space is $\{0,1\}^{m(\lambda)}$, and the ciphertext space is $\mathcal{X}_\lambda \times \{0,1\}^{m(\lambda)}$. Specifically,

$$\mathsf{Enc}_\lambda(k,m) = (r, m \oplus F(k,r))$$

where $r \overset{\text{R}}{\leftarrow} \mathcal{X}_\lambda$.

$$\mathsf{Dec}(k,(r,c)) = F(k,r) \oplus c.$$

Note that we did not need to use a PRP rather it suffices to use a PRF. Nevertheless, in practice we use AES for the PRF.

The CPA security of this scheme follows immediately from the definition of a PRF.

References