Introduction to Cryptography February 15, 2018

Lecture 10: Secret-Key Encryption

Instructor: Vipul Goyal *Scribe: Roger Iyengar*

1 Symmertric key encryption

1. $G(n) = k$
2. $E(m, k) = c$
3. $D(c, k) = m'$

1.1 Requirements

- All three algorithms must be probabalistic polynomial-time (PPT) for every input.
- **Correctness**: $\forall m \ Pr[k \leftarrow G(n), C \leftarrow E(m, k) : D(c, k) = m] = 1$
- Security?

Definition 1 *(Perfect Security)*
(G, E, D) *is perfectly secure if*
$\forall (m_1, m_2), \forall c \ Pr[k \leftarrow G : E(m_1, k) = c] = Pr[k \leftarrow G : E(m_2, k) = c]$

Both messages produce identical ciphertext distributions

Definition 2 *(Indistinguishability based secuirty)*
$\forall (m_1, m_2) \ \{E(m_1, k) : k \leftarrow G(n)\} \approx_c \{E(m_2, k) : k \leftarrow G(n)\}$

Both messages produce computationally indistinguishable ensembles of ciphertext distributions

Definition 3 *(Alternate formulation of Indistinguishability based security)*
$\forall (m_1, m_2) \ \forall PPT \ \mathcal{A} \ Pr[k \leftarrow G(n), b \xleftarrow{\$} \{0, 1\} : \mathcal{A}(E(m_b, k) = b] \leq \frac{1}{2} + negl(n)$

Suppose we had a PPT adversary, $\mathcal{A}$, that was given the encoding of either m_1 or m_2. However, $\mathcal{A}$ is not told which of the two messages were used to generate the encoding. $\mathcal{A}$ must output 0 if the encoding came from m_1 or 1 if the encoding came from m_2. The encryption scheme is secure if $\mathcal{A}$'s chance of outputting the correct value is similar to that of an adversary who outputs a value uniformly at random.

1.2 Equivalence

From the definition of Indistinguishability based security:
$\underbrace{|Pr[k \leftarrow G(n) : D(E(m_0, k)) = 0] - Pr[k \leftarrow G(n) : D(E(m_1, k)) = 0]|}_{p} \leq negl(n)$

For both m_0 and m_1

$$\begin{aligned} Pr[D = 0] &= p \pm negl(n) \\ Pr[D = 1] &= (1 - p) \pm negl(n) \\ Pr[b = D(.)] &= \frac{1}{2} \pm negl(n) \end{aligned}$$

Making use of the fact that the output of D is independent of b

2 One-time Encryption

We will now provide an encryption scheme that can be used to transmit a single message. Note that the two parties using this scheme must have both agreed on a key prior to transmitting a message using this scheme.

Given PRG: $\{0,1\}^n \to \{0,1\}^{l(n)}$

$$\begin{aligned} G(n) &: = \text{sample random } k \xleftarrow{\$} \{0,1\}^n \\ E(m,k) &: = m \oplus PRG(k),\ m \in \{0,1\}^{l(n)} \\ D(c,k) &: = c \oplus PRG(k) \end{aligned}$$

Proof. of security
We want to show that:

$$\{m_0 \oplus PRG(k) : k \xleftarrow{\$} \{0,1\}^n\} \approx_c \{m_1 \oplus PRG(k) : k \xleftarrow{\$} \{0,1\}^n\}$$

We first show that:

$$\{m_0 \oplus PRG(k) : k \xleftarrow{\$} \{0,1\}^n\} \approx_c \{\underbrace{m_0 \oplus R}_{R'} : R \leftarrow \{0,1\}^{l(n)}\} \tag{1}$$

$$\{m_1 \oplus PRG(k) : k \xleftarrow{\$} \{0,1\}^n\} \approx_c \{m_1 \oplus R : R \xleftarrow{\$} \{0,1\}^{l(n)}\} \tag{2}$$

We can show that the ensemble on the left hand side (LHS) of line 1 is indistinguishable from the ensemble on the right hand side (RHS) of line 1 using a reduction. Assume for contradiction that there exists a PPT algorithm $\mathcal{A}$ that can distinguish between the LHS ensemble and the RHS ensemble. This allows us to construct an adversary that can break the indistinguishability of the PRG.[1]

The RHS ensemble on line 1 is indistinguishable from the RHS ensemble on line 2 because one-time pads are secure. Formally:

$$\{m_0 \oplus R : R \leftarrow \{0,1\}^{l(n)}\} \approx_c \{m_1 \oplus R : R \xleftarrow{\$} \{0,1\}^{l(n)}\}$$

The LHS ensemble on line 1 is thus indistinguishable from the LHS ensemble on line 2 because indistinguishability is transitive. ■

[1] This proof is illustrated in detail on page 31 of Abhishek Jain's notes.

3 Multiple message encryption

A one-time encryption scheme is not secure if it is used to encrypt multiple messages. We therefore define an encryption scheme which can be used to send multiple messages. This scheme requires a variable called *ctr* to store state information.

$$
\begin{aligned}
G(n) &= \text{pick } k \xleftarrow{\$} \{0,1\}^n \\
E(m,k,ctr) &= m \oplus \underline{\text{bits of PRG starting from ctr + 1}} \\
&\quad \text{set } ctr \leftarrow ctr + |m| \\
D(c,k,ctr) &= m \oplus \text{bits of PRG starting from ctr'} \\
&\quad \text{set } ctr' \leftarrow ctr' + |m|
\end{aligned}
$$

Definition 4 *(Multi-message secure encryption)*

$$\forall \{m_0^i\}_{i=1}^q \ \{m_1^i\}_{i=1}^q \text{ where } q = poly(n)$$

$$\{\{E(m_0^i,k)\}_{i=1}^q : k \leftarrow G(n)\} \approx_c \{\{E(m_1^i,k)\}_{i=1}^q : k \leftarrow G(n)\}$$

Imagine you have two sets of q messages, $(m_0^1, m_0^2, \ldots, m_0^q)$ and $(m_1^1, m_1^2, \ldots, m_1^q)$. The encoding of $(m_0^1, m_0^2, \ldots, m_0^q)$ is indistinguishable from the encoding of $(m_1^1, m_1^2, \ldots, m_1^q)$.

Definition 5 *(Alternate formulation of multi-message encryption)*

$$Pr[k \leftarrow G(n), b \xleftarrow{\$} \{0,1\} : A(\{E(m_1^i,k)\}_i) = b] \leq \frac{1}{2} + negl(n)$$

Imagine there exists a PPT adversary $\mathcal{A}$ that is given either $(m_0^1, m_0^2, \ldots, m_0^q)$ or $(m_1^1, m_1^2, \ldots, m_1^q)$. A outputs 0 if it was given $(m_0^1, m_0^2, \ldots, m_0^q)$. $\mathcal{A}$ outputs 1 if it was given $(m_1^1, m_1^2, \ldots, m_1^q)$. A's chance of succeeding is similar to an algorithm that just picks randomly.

4 Stateless Deterministic Encryption

Requiring the two communicating parties to maintain state information is undesirable. We will therefore look at schemes that do not require parties to maintain a state. Unfortunately, we are limited by the following theorem:

Theorem 1 *Any stateless deterministic encryption scheme can't be multi-message secure*

Proof. Consider (m_0, m_0) and (m_1, m_2) where $m_1 \neq m_2$

The adversary is given $\{C_0, C_0\}$ or $\{C_1, C_2\}$. They can determine which they were given by simply checking if the two ciphertexts are equal. ■

5 Stateless Nondeterministic Encryption

We have shown that stateless deterministic encryption schemes cannot be multi-message secure. However, stateless nondeterministic encryption schemes can be multi-message secure. We will show an example of such a scheme.

Construction: Given pseudo-ranfom function (PRF) $\{0,1\}^n X \{0,1\}^n \to \{0,1\}^n$

$$G(n) = k \xleftarrow{\$} \{0,1\}^n$$
$$E(m,k) = \text{Generate } r \xrightarrow{\$} \{0,1\}^n, \text{ output } c = (r, m \oplus PRF(k,r))$$
$$D(c,k) = \text{Parse } C = (c_1, c_2), \text{ output } c_2 \oplus PRF(k, c_1)$$

Proof. of correctness

$$(m \oplus PRF(k,r)) \oplus PRF(k,r) = m$$

■

Proof. that a PPT algorithm can compute this scheme

The PRF can be computed in time polynomial in n. Two exclusive or of two bit strings of length n can also be computed in time polynomial in n ■

Proof. that our stateless nondeterministic encryption scheme is secure

We will use the abbreviation RF to refer to a truly random function.

$\underline{H_0}$: The adversary is given $\{r_i, m_0^i \oplus PRF(r_i)\}_{i=1}^q$

$\underline{H_1}$: The adversary is given $\{r_i, m_0^i \oplus RF(r_i)\}_{i=1}^q$

$\underline{H_2}$: Adversary is given $\{r_i, m_0^i \oplus R_i' : R_i' \xleftarrow{\$} \{0,1\}^n\}_i$

$\underline{H_3}$: Adversary is given $\{r_i, m_1^i \oplus R_i' : R_i' \xleftarrow{\$} \{0,1\}^n\}_i$

$\underline{H_4}$: The adversary is given $\{r_i, m_1^i \oplus RF(r_i)\}_{i=1}^q$

$\underline{H_5}$: The adversary is given $\{r_i, m_1^i \oplus PRF(r_i)\}_{i=1}^q$

Lemma 2 *If the probability of $\mathcal{A}$ outputing 0 in H_0 is $\frac{1}{2} + negl(k)$, this probability remains $\frac{1}{2} + negl(k)$ in H_1*

Suppose not: Say $P[\mathcal{A}$ outputing 0] in H_0 is p_0 and $P[\mathcal{A}$ outputing 0] in H_1 is p_1 such that $p_1 - p_0 \to$ noticable(n). This means that there is a PPT algorithm B with the following properties:

- B is given either $\{r_i, m_0^i \oplus PRF(r_i)\}_{i=1}^q$ or $\{r_i, m_0^i \oplus RF(r_i)\}_{i=1}^q$ but not told which
- B outputs 0 with probability p_0 if it was given $\{r_i, m_0^i \oplus PRF(r_i)\}_{i=1}^q$
- B outputs 0 with probability p_1 if it was given $\{r_i, m_0^i \oplus RF(r_i)\}_{i=1}^q$

We will now use B to construct a PPT algorithm C that wins a game we will refer to as the "PRF Game." C is given oracle access to some function $F(.)$. To win the game, C will output 0 if F is a truly random function or 1 if F is a PRF. Note that B takes a set of (seed, encoded message) pairs as input. C instead can query x_i to get $F(x_i)$.

C works as follows:

for $i = 1$ ***to*** q **do**
  Sample $r_i \xleftarrow{\$} \{0,1\}^n$;
  get $F(r_i)$ from oracle;
end
Prepare Cyphertext (CT) vector $\{r_i, m_0^i \oplus F(r_i)\}_{i=1}^q$;
Run B on CT vector and get output;
Return the output received from B;

$$\begin{aligned} Pr[C \text{ wins } | F = PRF] &= Pr[C \text{ wins } | F = PRF \text{ and } B = 0] * Pr[B = 0 | F = PRF] \\ &\quad + Pr[C \text{ wins } | F = PRF \text{ and } B = 1] * Pr[B = 1 | F = PRF] \\ &= 0 * p_0 + \frac{1}{2} * (1 - p_0) \end{aligned}$$

$$Pr[C \text{ wins } | F \text{ is a random function}] = 1 * p_1 + \frac{1}{2} * (1 - p_1)$$

$$Pr[C \text{ wins }] = \frac{1}{2} + \frac{(p_1 - p_0)}{2} = \frac{1}{2} + \text{noticable}(n)$$

However, this violates the definition of a PRF. Therefore B cannot be a PPT algorithm. Thus the ensemble in H_0 and the ensemble in H_1 are indistinguishable.

We will now compare H_1 and H_2. In H_1, we will never sample the same random string more than once. However, this might happen in H_2. Formally, we have to consider the following case:

$\exists(i, j)$ such that $r_i = r_j$ and $R_i = R_j$ in H_2 but R_i' and R_j' are still random in H_1.

$$Pr[\mathcal{A} = 0] = Pr[\mathcal{A} = 0 | \text{collision}] * \underbrace{Pr[\text{collision}]}_{\text{negl(n)}} + Pr[\mathcal{A} = 0 | \text{no collision}] * Pr[\text{no collision}]$$

The probability of sampling the same string twice in H_2 is negligible. H_1 and H_2 are identical in all other cases. Therefore:

$$Pr[\text{Adversary produces the correct output for } H_1] - Pr[\text{Adversary produces the correct output for } H_2] \leq negl(n)$$

The H_2 and H_3 ensembles are indistinguishable because one-time pads are secure. The H_3 and H_4 and H_4 and H_5 ensembles are indistinguishable due to transitivity. ■