COM S 687 Introduction to Cryptography	October 12, 2006

Lecture 14: Many-message security

Instructor: Rafael Pass *Scribe: Tudor Marian*

1 Many-message security

Last class we talked about encryption and definitions that were equivalent. In 1981 Goldwasser-Micali were the first to actually give a formal definition of encryption that was analog to Shannon's secrecy. But the problem we noted last class was that in the end, the scheme was not secure if an eavesdropper gets a hold of encryptions of many messages, and therefore the previous definitions were not satisfactory. We shall continue by adapting the definitions to quantify over many messages.

Definition 1 (Secure Encryption, many message indistinguishability) *Let (Gen, Enc, Dec) be an encryption scheme over message space $\mathcal{M}$ and key space $\mathcal{K}$. Then (Gen, Enc, Dec) is said to be many-message secure if $\forall$ non uniform p.p.t. $\mathcal{A}$, $\forall$ polynomial $q(n)$, $\exists$ a negligible function $\epsilon(n)$ such that $\forall m_0, m_1, \ldots, m_{q(n)}, m'_0, m'_1, \ldots, m'_{q(n)}$ for which $|m_0| = \cdots = |m_{q(n)}| = |m'_0| = \cdots = |m'_{q(n)}|$ it holds that*

$$|Pr[k \leftarrow Gen(1^n) : \mathcal{A}(Enc_k(m_0), Enc_k(m_1) \ldots Enc_k(m_{q(n)})) = 1] -$$
$$Pr[k \leftarrow Gen(1^n) : \mathcal{A}(Enc_k(m'_0), Enc_k(m'_1), \ldots Enc_k(m'_{q(n)})) = 1]| \leq \epsilon(n)$$

Last time we have seen that $m \oplus G(s)$ is not a multi-message secure encryption scheme since an attacker can tell apart two messages $m \neq m'$ given that he sees their corresponding distinct cyphertexts $Enc(m) \neq Enc(m')$. In fact the following result is stronger (and not very encouraging):

Theorem 1 *There exists no deterministic stateless many message secure encryption scheme.*

The insight is to use a PRF to generate a new key every time. Consider the first attempt:

- $Gen(1^n) = s \leftarrow \{0,1\}^n$
- $Enc_s(m) = m \oplus f_s(m)$, where $\{f_s\} : \{0,1\}^{|s|} \rightarrow \{0,1\}^{|s|}$ is a PRF.

But this is still a deterministic scheme. Evaluate f on a random element for every message, hence we have the second attempt:

- $Gen(1^n) = s \leftarrow \{0,1\}^n$
- $Enc_s(m) = r||m \oplus f_s(r); r \leftarrow \{0,1\}^n$, where $\{f_s\} : \{0,1\}^{|s|} \rightarrow \{0,1\}^{|s|}$ is a PRF.
- $Dec_s(r,c) = c \oplus f_s(r)$

Theorem 2 *(Gen, Enc, Dec) is a many-message secure encryption scheme.*

Proof. Consider some distinguisher D and two sequences of messages: $m_0, m_1 \ldots, m_{q(n)}$ and $m'_0, m'_1 \ldots, m'_{q(n)}$ respectively. Choose the following sequence of distributions:

- H_1 – real execution with $m_0, m_1 \ldots, m_{q(n)}$
  $$\{s \leftarrow \{0,1\}^n, r_0, \ldots, r_{q(n)} \leftarrow \{0,1\}^n : r_0||m_0 \oplus f_s(r_0), \ldots m_{q(n)} \oplus f_s(r_{q(n)})\}$$
  This is precisely what the adversary sees when receiving the encryptions of $m_0, \ldots m_{q(n)}$.
- H_2 – using a truly random function instead of f on $m_0, m_1, \ldots, m_{q(n)}$
  $$\{R \leftarrow \{\{0,1\}^n \rightarrow \{0,1\}^n\}; r_0, \ldots, r_{q(n)} \leftarrow \{0,1\}^n : r_0||m_0 \oplus R(r_0), \ldots m_{q(n)} \oplus R(r_{q(n)})\}$$
- H_3 – using OTP (one time pad) on $m_0, m_1, \ldots, m_{q(n)}$
  $$\{p_0 \ldots p_{q(n)} \leftarrow \{0,1\}^n; r_0, \ldots, r_{q(n)} \leftarrow \{0,1\}^n : r_0||m_0 \oplus p_0, \ldots m_{q(n)} \oplus p_{q(n)}\}$$
- H_4 – using OTP (one time pad) on $m'_0, m'_1, \ldots, m'_{q(n)}$
  $$\{p_0 \ldots p_{q(n)} \leftarrow \{0,1\}^n; r_0, \ldots, r_{q(n)} \leftarrow \{0,1\}^n : r_0||m'_0 \oplus p_0, \ldots m'_{q(n)} \oplus p_{q(n)}\}$$
- H_5 – using a truly random function instead of f on $m'_0, m'_1, \ldots, m'_{q(n)}$
  $$\{R \leftarrow \{\{0,1\}^n \rightarrow \{0,1\}^n\}; r_0, \ldots, r_{q(n)} \leftarrow \{0,1\}^n : r_0||m'_0 \oplus R(r_0), \ldots m'_{q(n)} \oplus R(r_{q(n)})\}$$
- H_6 – real execution with $m'_0, m'_1 \ldots, m'_{q(n)}$
  $$\{s \leftarrow \{0,1\}^n, r_0, \ldots, r_{q(n)} \leftarrow \{0,1\}^n : r_0||m'_0 \oplus f_s(r_0), \ldots m'_{q(n)} \oplus f_s(r_{q(n)})\}$$

D can distinguish H_1 and $H2$ with at most negligible probability, otherwise it contradicts the pseudo-randomness properties of $\{f_s\}$. The same argument applies for H_6 and H_5.

H_2 and H_3 are "almost" identical except for the case when $\exists i, j$ such that $r_i = r_j$, but this happens with negligible probability, therefore D can distinguish H_2 and H_3 only with negligible probability. The same argument applies for H_4 and H_5.

H_3 and H_4 are identical because of the properties of the one time pad.

By the polyjump lemma, D can distinguish between H_1 and H_6 with at most negligible probability. ■

In what follows we shall explore stronger attack scenarios. This far we had Eve listening to the communication channel between Alice and Bob – suppose Eve has *greater powers*.

2 Stronger Attack Models

Attack models:

- Ciphertext only attack – this is what we considered so far.
- Known plaintext attack – Adversary may get to see pairs of form $(m_0, Enc_k(m_0))\ldots$
- Chosen plain text (CPA) – Adversary gets access to an encryption oracle, before and after selecting messages.
- Chosen ciphertext attack

  CCA1: Adversary has oracle access to decryption before selecting the messages. (due to Naor and. Yung)

  CCA2: CCA1 + adversary also has access to decryption oracle after selecting the messages. It is not allowed to decrypt the challenge however. (introduced by Racko and Simon)

Definition 2 (Secure encryption CPA / CCA1 / CCA2) *Let* $\Pi =$*(Gen, Enc, Dec) be an encryption scheme. Let the random variable* $IND_b^{O_1,O_2}(\Pi, \mathcal{A}, n)$ *where* $\mathcal{A}$ *is a non uniform p.p.t.,* $n \in N,\ b \in \{0,1\}$ *denote the output of the following experiment:*

- $k \leftarrow Gen(1^n)$
- $m_0, m_1, state \leftarrow \mathcal{A}^{O_1(k)}(1^k)$
- $c \leftarrow Enc_k(m_b)$
- *output:* $\mathcal{A}^{O_2(k)}(c, state)$

Π *is said to be respectively CPA / CCA1 / CCA2 secure if* $\forall$ *non uniform p.p.t.* $\mathcal{A}$ *it holds that*

$$\{IND_0^{O_1,O_2}(\Pi, \mathcal{A}, n)\}_n \approx \{IND_1^{O_1,O_2}(\Pi, \mathcal{A}, n)\}_n$$

where $O_1(k), O_2(k)$ *are* $[Enc_k, Enc_k]/[(Enc_k, Dec_k), Enc_k]/[(Enc_k, Dec_k), (Enc_k, Dec_k)]$. *Additionally, in the case of CCA2 attacks one can only quantify over all* $\mathcal{A}$ *that never ask to decrypt* c.