CSE 5852, Modern Cryptography: Foundations Fall 2016

Lecture 9, 2016

Prof. Benjamin Fuller *Scribe: Kirk Gardner*

1 Overview

Last class we introduced semantic security. This was our attempt to bring perfect secrecy to the computational world.

Today, We will begin the proof that semantic security is equivalent to indistinguishable encryption.

Definition 1 (Semantic Security)**.** *[GM84] Let $\mathcal{M}$ be a message space and let K be a distribution.* Enc *is semantically secure if for all PPT $\mathcal{A}$ there exists PPT $\mathcal{A}'$ such that for any distribution M on $\mathcal{M}$ and any $f, h : \mathcal{M} \to \{0,1\}^*$*

$$Pr[\mathcal{A}(c, h(M)) = f(M)] - Pr[\mathcal{A}'(h(M)) = f(M)] \leq \varepsilon.$$

Definition 2 (Indistinguishable Encryption)**.** Enc *has indistinguishable encryption if for all PPT $\mathcal{A}$ and every $m_1, m_2 \in \mathcal{M}$*

$$|Pr_k[\mathcal{A}(\mathsf{Enc}_k(m_1)) = 1] - Pr_k[\mathcal{A}(\mathsf{Enc}_k(m_2)) = 1]| < \varepsilon.$$

Theorem 3. *A private-key encryption scheme has indistinguishable encryptions in the presence of an eavesdropper if and only if it is semantically secure.*

We're going to be doing a few things for the first time in this proof so we are going to be slow and methodical. These are:

1. Working with computationally limited attackers and Turing machines.
2. Building attackers out of one another.
3. Using a proof by contradiction.

Recall that we need to do two proofs here. The first is to show that semantic security implies indistinguishable encryptions and the second is to show that indistinguishable encryptions implies semantic security.

2 Semantic Security $\implies$ Indistinguishable Encryption

Lemma 4. *If a private-key encryption scheme is semantically secure then it has indistinguishable encryptions.*

Proof. Suppose (for the sake of contradiction) that there exists PPT $\mathcal{A}$ and messages $m_2, m_2 \in \mathcal{M}$ such that

$$|\Pr_k[\mathcal{A}(\mathsf{Enc}_k(m_1)) = 1] - \Pr_k[\mathcal{A}(\mathsf{Enc}_k(m_2)) = 1] \geq \frac{1}{p(n)}$$

for some polynomial $p(n)$. Our strategy will be to construct a message distribution M, functions f, h, and an adversary $\mathcal{A}^*$ in which there cannot exist $\mathcal{A}'$ such that

$$\Pr[\mathcal{A}(c, h(M)) = f(M)] - \Pr[\mathcal{A}'(h(M)) = f(M)] \leq \varepsilon.$$

Let M be a uniform message distribution over $\{m_1, m_2\}$ so $\Pr[M = m_1] = \Pr[M = m_2] = \frac{1}{2}$, and define a function $f : \mathcal{M} \to \{0,1\}^*$ which has a different value on these two messages, for example:

$$f(m_1) = 1, f(m_2) = 0.$$

Recall h represents the additional information known to the attacker. Define $h : \mathcal{M} \to \{0,1\}^*$ as any function independent of the message, i.e. the attacker knows no additional information about the message.

$$h(m) = 0 \ldots 0.$$

The following lemma follows from our definition of h and M: $\mathcal{A}'$ knows no additional information about the message, and will therefore always be wrong half the time.

Lemma 5. *For all* $\mathcal{A}'$

$$Pr[\mathcal{A}'(h(M)) = f(M)] = \frac{1}{2}.$$

Note: The above lemma is an exact equality. $\mathcal{A}'$ cannot do better or worse than guessing with probability 1/2 regardless of their strategy.

We now define an attacker $\mathcal{A}^*$ as follows

- **Input:** $c, h(m)$
- **Output:** $\mathcal{A}(c)$.

We now need to show that $\Pr[\mathcal{A}^*(c, h(M)) = f(M)]$ differs greatly from $\frac{1}{2}$ in order to draw a contradiction.

$$\begin{aligned}
\Pr[\mathcal{A}^*(c, h(M)) = f(M)] &= \frac{1}{2}\Pr[\mathcal{A}^*(c, h(M)) = f(M) \mid M = m_1] + \frac{1}{2}\Pr[\mathcal{A}^*(c, h(M)) = f(M) \mid M = m_2] \\
&= \frac{1}{2}\Pr[\mathcal{A}^*(c, h(M)) = 1 \mid M = m_1] + \frac{1}{2}\Pr[\mathcal{A}^*(c, h(M)) = 0 \mid M = m_2] \\
&= \frac{1}{2}\Pr[\mathcal{A}(c) = 1 \mid M = m_1] + \frac{1}{2}\Pr[\mathcal{A}(c) = 0 \mid M = m_2] \\
&= \frac{1}{2}\Pr[\mathcal{A}(\mathsf{Enc}(m_1)) = 1 \mid M = m_1] + \frac{1}{2}(1 - \Pr[\mathcal{A}(\mathsf{Enc}(m_2)) = 1 \mid M = m_2]) \\
&= \frac{1}{2} + \frac{1}{2}\left(\Pr[\mathcal{A}(c) = 1 \mid M = m_1] - \Pr[\mathcal{A}(c) = 0 \mid M = m_1]\right) \\
&> \frac{1}{2} + \frac{1}{2}\frac{1}{p(n)}
\end{aligned}$$

thus, $\Pr[\mathcal{A}^*(c, h(M)) = f(M)] > \frac{1}{2} + \frac{1}{2}\frac{1}{p(n)}$ for some polynomial $p(n)$.

Recalling Lemma 5 implies $\Pr[\mathcal{A}'(h(M)) = f(M)] = \frac{1}{2}$ for all $\mathcal{A}'$ we have

$$\Pr[\mathcal{A}^*(c, h(M)) = f(M)] - \Pr[\mathcal{A}'(h(M)) = f(M)] = \frac{1}{2} + \frac{1}{2}\frac{1}{p(n)} - \frac{1}{2} = \frac{1}{2p(n)}$$

which implies there exists some PPT $\mathcal{A}$ and a message distribution M such that for all $\mathcal{A}'$

$$\Pr[\mathcal{A}^*(c, h(M)) = f(M)] - \Pr[\mathcal{A}'(h(M)) = f(M)] > \varepsilon(n)$$

for some polynomial function $\varepsilon(n) := \frac{1}{2p(n)}$. It follows that if Enc does not have indistinguishable encryption then Enc is not semantically secure, and therefore that semantic security implies indistinguishable encryption by contrapositive. □

3 Indistinguishable Encryption $\implies$ Semantic Security (setup)

Consider some PPT $\mathcal{A}$ and assume Enc has indistinguishable encryption. Our goal is to create a simulator $\mathcal{A}'$ that receives $h(m)$ as input.

We define $\mathcal{A}'$ as follows:

1. **Input:** $h(m)$
2. create $k \leftarrow \mathrm{Gen}(\cdot)$
3. create $c = \mathsf{Enc}_k(1 \ldots 1)$
4. run $A(c, h(m))$

Given this construction of $\mathcal{A}'$ we will prove the following lemma by contradiction.

Lemma 6. *For any polynomial p and all message distributions M*

$$|Pr[\mathcal{A}(\mathsf{Enc}(M), h(M)) = f(M)] - Pr[\mathcal{A}'(h(M)) = f(M)]| \leq \frac{1}{p(n)}$$

Proof. Suppose (for the sake of contradiction) there exists a polynomial p such that

$$|\Pr[\mathcal{A}(\mathsf{Enc}(M), h(M)) = f(M)] - \Pr[\mathcal{A}'(\mathsf{Enc}(1 \ldots 1), h(M)) = f(M)]| > \frac{1}{p(n)}.$$

We note that if this is true for a message distribution M then it must be true for at least one message in the distribution. So there exists $m^* \in M$ such that

$$|\Pr[\mathcal{A}(\mathsf{Enc}(m^*), h(m^*)) = f(m^*)] - \Pr[\mathcal{A}'(\mathsf{Enc}(1 \ldots 1), h(m^*)) = f(m^*)]| > \frac{1}{p(n)}.$$

We need to build an attacker $\mathcal{A}^*$ that doesn't take $h(M^*)$ as input. The strategy will be to hardwire the values of $h(m^*), f(m^*))$.[1]

[1] A small but important note. This hardwiring of $h(m^*)$ and $f(m^*)$ actually makes this a non-uniform reduction. That is that indistinguishable encryptions imply semantic security for non-uniform Turing machines. It is possible to close this gap and show the equivalence for uniform Turing machines.

- **Input:** c
- **Output:** 1 if $A(c, h(m^*)) = f(m^*)$, 0 otherwise.

Next, we will show that A^* breaks indistinguishable encryptions for $m^*, 1 \dots 1$.

4 Semantic Security for Multiple Messages

In the next homework we will convert these definitions to multiple messages. Like in the information-theoretic world security for a single message does not imply security for multiple messages.[2]

We present an extension of semantic security for multiple messages.

Definition 7 (k-Message Semantic Security)**.** *Let $\mathcal{M}$ be a message space. Let K be a distribution.* Enc *is semantically secure for k messages if for all PPT $\mathcal{A}$ there exists PPT $\mathcal{A}'$ such that for any joint message distribution $M = M_1, \dots, M_k$ on $\mathcal{M}$ and any $f, h : \mathcal{M}^k \to \{0,1\}^*$*

$$Pr[\mathcal{A}(C_1, \dots, C_k, h(M_1, \dots, M_k)) = f(M_1, \dots, M_k)] - Pr[\mathcal{A}'(h(M_1, \dots, M_k)) = f(M_1, \dots, M_k)] \leq \varepsilon.$$

for some negligible $\epsilon(n)$.

References

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[2] Looking ahead when we use public-key cryptography single message security does imply multiple message security.