CS276: Cryptography September 24, 2015

Encryption Schemes cont.

Instructor: Alessandro Chiesa *Scribe: Eleanor Cawthon*

Message Indistinguishability

First, we review our intuitive and formal definitions of encryption schemes. Informally, we consider Alice and Bob, who share a secret key \$, and a passive adversary Eve who can read their messages, but not stop, inject, or modify them. An encryption scheme provides the following properties:

1. **Functionality**: if Alice sends m, then Bob can recover m.
2. **Security**: Eve learns nothing about m.

Formally,

Definition 1 *an **encryption scheme** is a tuple (E, D) of probabilistic polynomial time algorithms that satisfies the following:*

1. ***Completeness** : for all $k \in \mathbb{N}$ and for all secret keys $sk \in \{0,1\}^{\ell(k)}$, for every message $m \in \{0,1\}^{n(k)}$, we have $D(1^k, sk, E(1^k, sk, m)) = m$.*
2. ***Security** : For every pair of families of messages with the same length $\left(\text{that is, } \forall \left\{m_k^{(0)}\right\}_k, \left\{m_k^{(1)}\right\}_k \text{ s.t. } m_k^{(i)} \in \{0,1\}^{n(k)}\right)$, we have*

$$\left\{E\left(1^k, U_{\ell(k)}, m_k^{(0)}\right)\right\} = \left\{E\left(1^k, U_{\ell(k)}, m_k^{(1)}\right)\right\}$$

Note: the equality can be either perfect, statistical, or computational.

*This is called **message indistinguishability** .*

One-time Pads

Theorem 2 *There exist E and D that satisfy completeness and perfect message indistinguishability.*

Proof: consider the one-time pad (**OTP**):

$$E(1^k, sk, m) := sk \oplus m$$

$$D(1^k, sk, c) := sk \oplus c$$

This is complete — xor'ing the same secret key encrypts and decrypts. It is also secure — for a uniformly selected sk, xor preserves the uniformity – $U_{\ell(k)} \oplus m \equiv U_{\ell(k)} \oplus m' \forall m, m'$. So this works. □

Limitations:

1. Keys are large (the key has to be as long as the message, $|sk| \geq |m|$).
2. Key is one time — if you reuse some sk to generate two cyphertexts $sk \oplus m_1 = c_1$ and $sk \oplus m_2 = c_2$, the adversary can combine them to get back information about the original message ($c_1 \oplus c_2 = m_1 \oplus m_2$). So, if the attacker already knew m_1, it would learn m_2 this way.

Perfect Message Indistinguishability

Theorem 3 *For every (E, D) that satisfies completeness and perfect message indistinguishability, it holds that $\ell(k) \geq n(k)$.*

Proof: suppose $\ell(k) < n(k)$. Pick $m_k^{(0)} \in \{0,1\}^{n(k)}, sk \in \{0,1\}^{\ell(k)}$. Get $c := E(sk, m^0)$. Now pick $m^{(1)} \in \{0,1\}^{n(k)}, sk' \in \{0,1\}^{\ell(k)}$ such that $D(sk', c) \neq m^{(1)}$. (We know $m^{(1)}$ exists because $|\bigcup_{\tilde{sk}} D(\tilde{sk}, c)| \leq 2^{\ell(k)} < 2^{n(k)}$.

By completeness, $E(sk', m^{(1)}) \neq c$.

$$Pr_{\tilde{sk}}\left[E\left(\tilde{sk}, m^{(0)}\right) = c\right] > 0$$

$$Pr_{\tilde{sk}}\left[E\left(\tilde{sk}, m^{(1)}\right) = c\right] = 0$$

□

From now on, we care about computational, not perfect or statistical.

Semantic security

Definition 4 *Eve wants to know $f(m)$ and already knows some information $I(m)$. e.g. Eve might know that Alice draws m from well-known $\mathscr{M}$ (such as the set of all English sentences).*

An encryption scheme provides ***semantic security*** *if, for every message distribution $\mathscr{M}_k \in \Delta(\{0,1\}^{n(k)})$, for every goal function $f_k : \{0,1\}^{n(k)} \to \{0,1\}^*$, for every partial information $I_k : \{0,1\}^{n(k)} \to \{0,1\}^*$, and for every ppt A,*

$$|Pr\left[A\left(1^k, I_k\left(\mathscr{M}_k\right), E\left(1^k, U_{\ell(k)}, \mathscr{M}_k\right)\right) = f_k\left(\mathscr{M}_k\right)\right] - Pr\left[S\left(1^k, I_k\left(\mathscr{M}_k\right)\right) = f_k\left(\mathscr{M}_k\right)\right]|$$

is negligible in k.

That is, anything eve learns from looking at the message cyphertext, Eve already knew from $\mathscr{M}_k$.

Equivalence of Message Indistinguishability and Semantic Security

Theorem 5 *An encryption scheme provides computational message indistinguishability if and only if it provides semantic security.*

Proof: ($\leftarrow$) Suppose (E, D) is not c.M.I. Then there is some distinguisher D and $\left\{m_k^{(0)}\right\}_k, \left\{m_k^{(1)}\right\}_k$ s.t. $m_k^{(i)} \in \{0,1\}^{n(k)}$ such that

$$\delta(k) = |Pr\left[A(E(U_{r(k)}, m_k^{(0)})) = 1\right] - Pr\left[A(E(U_{r(k)}, m_k^{(1)})) = 1\right]|$$

is not negl(k). Equivalently, we have a distinguisher that allows us to guess a bit with better than $\frac{1}{2}$ probability:

$$Pr\left[A(E(U_{r(k)}, m_k^{(U_1)})) = U_1\right] = \frac{1}{2} + \frac{\delta(k)}{2}$$

We can use this to construct a counterexample for semantic security. Let $\mathscr{M}_k$ be our two one-bit messages $\left\{m_k^{(0)}, m_k^{(1)}\right\}$, and define our goal function $f_k(m_k^b)$ as the function that guesses that bit b. We leave the partial information function empty, $I_k(\cdot) = \varepsilon$. Now consider $A'\left(I_k\left(\mathscr{M}_k\right), E\left(U_{r(k)}, \mathscr{M}_k\right)\right) = A(E(U_{r(k)}, \mathscr{M}_k))$, which we already know has $Pr\left[A(E(U_{r(k)}, \mathscr{M}_k)) = f_k(\mathscr{M}_k)\right] = \frac{1}{2} + \frac{\delta(k)}{2}$. Then for any S with with $Pr\left[S(1^k, I_k(\mathscr{M}_k)) = f_k(\mathscr{M}_k)\right] = \frac{1}{2}$, we have

$$|Pr\left[A'\left(1^k, I_k\left(\mathscr{M}_k\right), E\left(1^k, U_{\ell(k)}, \mathscr{M}_k\right)\right) = f_k\left(\mathscr{M}_k\right)\right] - Pr\left[S\left(1^k, I_k\left(\mathscr{M}_k\right)\right) = f_k\left(\mathscr{M}_k\right)\right]| = \frac{1}{2} + \frac{\delta(k)}{2}$$

which is non-negligible in k. Thus, semantic security implies computational message security.

($\rightarrow$) if a scheme has computational message security, we prove directly that it also provides semantic security. We can fix M_k, I_k, f_k, A to define a simulator

$$S(1^k, I_k(\mathscr{M}_k)) := A(1^k, I_k(m), E(U_{r(k)}, 0))$$

Now, for any m, if the difference between $Pr\left[A(1^k, I_k(m), E(1^k, U_{r(k)}, m)) = f_k(m)\right]$ and $Pr\left[A(1^k, I_k(m), E(1^k, U_{r(k)}, 0)) = f_k(\mathscr{M}_k)\right]$ is non-negligible, then we can use A to construct a distinguisher that distinguishes between $E(1^k, U_{r(k)}, 0)$ and $E(1^k, U_{r(k)}, m)$, violating message indistinguishability. Thus, (E, D) is semantically secure. □

Now back to constructions:

We can construct perfect message indistinguishability from one time pads, but the keys are huge so this is not practical. But, we can construct computaitonal message indistinguishability from one time pads and PRGs to provide (a lot of) randomness

Theorem 6 *If there are pseudorandom generators, we can construct a one-time symmetric encryption system with small keys.*

Proof:

$$E(1^k, sk, m) = g_k(sk) \oplus m$$
$$D(1^k, sk, c) = g_k(sk) \oplus c$$
$$g_k(U_k) \oplus m^{(0)} =^c U_{n(k)} \oplus m^{(0)} = U_{n(k)} = U_{n(k)} \oplus m^{(1)} =^c g_k(U_k) \oplus m^{(1)}$$

we can use the hybrid argument to complete the proof. □

Multi-Message Computational Message Indistinguishability

The above implies our definitions of security don't protect against multiple messages. So, we construct a revised definition for **p-Message indistinguishability** :

Definition 7 *For all* $\left\{\vec{m_k}^{(0)}\right\}_k, \left\{\vec{m_k}^{(1)}\right\}_k$ *s.t.* $m_{k,i}^{(b)} \in \{0,1\}^{n(k)}$, *we have*

$$\left\{E\left(1^k, U_{r(k)}, m_{k,1}^{(0)}\right), \ldots, E\left(1^k, U_{r(k)}, m_{k,p(k)}^{(0)}\right)\right\} =^c \left\{E\left(1^k, U_{r(k)}, m_{k,1}^{(1)}\right), \ldots, E\left(1^k, U_{r(k)}, m_{k,p(k)}^{(1)}\right)\right\}$$

No deterministic, stateless encryption scheme can achieve perfect message indistinguishability.

Example:
$$\begin{array}{ll} (E(1^k, U_{r(k)}, 0), E(1^k, U_{r(k)}, 0)) & \vec{m_k}^{(0)} = (0,0) \\ (E(1^k, U_{r(k)}, 0), E(1^k, U_{r(k)}, 1)) & \vec{m_k}^{(1)} = (0,1) \end{array}$$

Note that E is both deterministic and stateless. Can we fix it with statefullness?

Stateful Approaches

$$E(1^k, sk, m, i) = g_k^{(i)}(sk) \oplus m$$
$$D(1^k, sk, c, i) = g_k^{(i)}(sk) \oplus c$$
$$g_k^{(1)}(U_k), g_k^{(2)}(U_k), \cdots =^c U_{n(k)}^{(1)}, U_{n(k)}^{(2)}, \ldots$$

Even though using the same generator/seed, each individual element is indistinguishable from the others!

Chosen plaintext attacks

Now we want a stronger version of this that is secure against **chosen plaintext attacks** .

Say $\mathscr{F} = \{F_k\}$ with $F_k = \{f_s | s \leftarrow \{0,1\}^k\}$ is a pseudorandom function.

$$E(1^k, sk, m) = f_{sk}(r) \oplus m$$
$$D(1^k, sk, c) = f_{sk}(c_0) \oplus c_1$$

So, we can have syncrhonization on a much smaller amount of randomness but just as secure.

Definition 8 *Security against chosen plaintext attacks:*

(E, D) is **secure against CPA** if for all pairs of messages $\left\{\vec{m_k}^{(0)}\right\}_k, \left\{\vec{m_k}^{(1)}\right\}_k$ such that $m_k^b \in \{0,1\}^{n(k)}$, we have that for all ppt A:

$$|Pr\left[A^{E(1^k, U_{r(k)}, \cdot)}(1^k, E(1^k, U_{r(k)}, m_k^{(0)})) = 1\right] - Pr\left[A^{D(1^k, U_{r(k)}, \cdot)}(1^k, E(1^k, U_{r(k)}, m_k^{(1)})) = 1\right]|$$

is negligible (k).

So, Eve has access to an encryption oracle and can encrypt whatever she wants and find out what the ciphertext looks like.

Now, we'll prove that CPA security implies $p-$message indistinguishability

Theorem 9 *An encryption scheme that is secure against chosen plaintext attacks is also p-message indistinguishable for all p.*

Proof: suppose there is some polynomial p, messages $\left\{\vec{m_k}^{(0)}\right\}_k, \left\{\vec{m_k}^{(1)}\right\}_k$, and ppt A, such that A distinguishes the messages.

Consider the hybrids: $H_k^{(i)} := E(Ur(k), m_{k,1}^{(0)}), \ldots, E(U_{r(k)}, m_{k,i}^{(0)}\, E(Ur(k), m_{k,i+1}^{(1)}), \ldots, E(U_{r(k)}, m_{k,p(k)}^{(1)}$

There is an i such that A distinguishes $H_k^{(1)}$ and H_k^{i+1} with probability $\frac{\delta(k)}{p(k)}$.

Let $\tilde{A}^{E(1^k, sk, \cdot)}(1^k, c) :=$

1. $j = 1, \ldots, i : c_j \leftarrow E(1^k, sk, m_{k,j}^{(0)})$
2. $j = i+1, \ldots, p(k) : c_j \leftarrow E(1^k, sk, m_{k,j}^{(1)})$
3. $\vec{c} = (c_1 \ldots c_i c c_{i+1} \ldots (p_k))$
4. outputs $A(\vec{c})$

Note that this time we can't sample ciphertexts — so CPA is stronger that p-M.I.

We couldn't go from p to 1 because of samplability. We fix this with the oracle.

$$X \equiv^c Y \implies (X^{(1)}, X^{(2)}) \equiv^c (Y^{(1)}, Y^{(2)})$$

□