Advanced Topics in Cryptography

Lecture 6: Semantic security, chosen-ciphertext security.

Benny Pinkas
Based on slides of Moni Naor

No class on May 28.

Related papers

- Semantic security
  - Lecture notes of Moni Naor, http://www.cs.ioc.ee/yik/schools/win2004/naor-slides-2.5.ppt
  - Lecture notes of Jonathan Katz, http://www.cs.umd.edu/~jkatz/gradcrypto2/NOTES/lecture2.pdf

To specify security of encryption

- The power of the adversary
  - computational
    - Probabilistic polynomial time machine (PPTM)
  - access to the system
    - Can it change the messages?
- What constitutes a failure of the system
  - What it means to break the system.
    - Reading a message
    - Forging a message?

What is a public-key encryption scheme

- Allows Alice to publish a public key K_P while keeping hidden a secret key K_S
  Key generation: a method $G:\{0,1\}^* \mapsto \{0,1\}^* \times \{0,1\}^*$ that outputs K_P (Public) and K_S (secret)
- ``Anyone" who is given K_P and m can encrypt m
  Encryption: a method
  $$E:\{0,1\}^* \times \{0,1\}^* \times \{0,1\}^* \mapsto \{0,1\}^*$$
  – that takes a public key K_P, a message (plaintext) m and random coins and outputs an encrypted message ciphertext
- Given a ciphertext and the secret key it possible to decrypt it
  Decryption: a method
  $$D:\{0,1\}^* \times \{0,1\}^* \times \{0,1\}^* \mapsto \{0,1\}^*$$
  that takes a secret key K_S, a public key K_P and a ciphertext c and outputs a plaintext m. In general
  $$D(K_S, K_P, E(K_P, m, r)) = m$$

Computational Security of Encryption Indistinguishability of Encryptions

Indistinguishability of encrypted strings:

- Adversary **A** chooses $X_0, X_1 \in \{0,1\}^n$
- receives **encryption** of X_b for $b \in_R \{0,1\}$
- has to decide whether $b = 0$ or $b = 1$.

For every pptm **A,** choosing a pair $X_0, X_1 \in \{0,1\}^n$

$\left| \Pr[\mathbf{A} = \text{'1'} \mid b = 1] - \Pr[\mathbf{A} = \text{'1'} \mid b = 0] \right|$ is *negligible.*

- Probability is over the choice of keys, randomization in the encryption and **A**'s coins.

- In other words:
  the encryptions of X_0, X_1 are indistinguishable
- Note that this holds for any X_0, X_1 that **A** might choose

Computational Security of Encryption Semantic Security

- Whatever Adversary **A** can compute on encrypted string $X \in \{0,1\}^n$, so can **A'** that does **not** see the encryption of X yet simulates **A** 's knowledge with respect to X
- **A** selects:
  - Distribution D_n on $\{0,1\}^n$
  - Relation R(X,Y) - computable in probabilistic polynomial time
- For every pptm **A** choosing a (poly time samplable) distribution D_n on $\{0,1\}^n$ there is an pptm **A'** so that for all pptm relation R, for $X \in_R D_n$
  $\left| \Pr[R(X, \mathbf{A}(E(X))] - \Pr[R(X, \mathbf{A'}(\cdot))] \right|$ is *negligible*(*)
- In other words: The outputs of **A and A'** are indistinguishable even for a test that is aware of X

Note: the presentation of semantic security is non-standard (but equivalent to it)

(*) *$\varepsilon(n)$ is negligible if for $\forall$ polynomial p(n), $\exists$N, s.t. $\forall$n>N $\varepsilon(n) < p(n)$*

Equivalence of Semantic Security and Indistinguishability of Encryptions

- Would like to argue their equivalence
- Must define the attack
  - Otherwise cannot fully talk about an attack
- Chosen plaintext attacks
  - Adversary can obtain the encryption of any message it wishes
  - In an adaptive manner
    - Certainly feasible in a public-key setting
- More severe attacks
  - Chosen ciphertext

Security of public key cryptosystems: exact timing

- Adversary **A** gets to public key K_P
- Then **A** can mount an adaptive attack
  - No need for further interaction since can do all the encryption on its own
- Then **A** chooses
  - In semantic security the distribution D_n and the relation R
  - In indistinguishability of encryptions the pair X_0 , $X_1 \in \{0,1\}^n$
- Then **A** is given the test
  - In semantic security $E(K_P, X, r)$ for $X \in_R D_n$ and $r \in_R \{0,1\}^m$
  - In indistinguishability of encryptions the $E(K_P, X_b, r)$ for $b \in_R \{0,1\}$ and $r \in_R \{0,1\}^m$

When is each definition useful

- Semantic security seems to convey that the message is protected
  - Not the strongest possible definition
- Easier to prove indistinguishability of encryptions

The Equivalence Theorem

- For adaptive chosen plaintext attack in a public key setting:

  a cryptosystem is semantically secure if and only if it has the indistinguishability of encryptions property

Equivalence Proof

If a scheme has the indistinguishability of encryptions property, then it is semantically secure:

- Suppose not, and **A** chooses, some distribution D_n and some relation **R**
- Choose X_0 , $X_1 \in_R D_n$ and run **A** twice on
  - $C_0 = E(K_P, X_0, r_0)$ call the output $Y_0=A(E(K_P, X_0, r_0))$
  - $C_1 = E(K_P, X_1, r_1)$ call the output $Y_1=A(E(K_P, X_1, r_1))$
- For X_0 , $X_1 \in_R D_n$ let
  - $\alpha_0 = \text{Prob}[\mathbf{R}(X_0, Y_0)]$
  - $\alpha_1 = \text{Prob}[\mathbf{R}(X_0, Y_1)]$

Here we Use the power to generate encryptions

- If $|\alpha_0 - \alpha_1|$ is **non** negligible, then can distinguish between an encryption of X_0 and X_1
  - This contradicts the indistinguishability property, and therefore the assumption
- If $|\alpha_0 - \alpha_1|$ is negligible, then can run **A'** with *no* access to encryption
  - We want to compete with R(X,A(E(X)).
  - sample $X' \in_R D_n$ and $C' = E(K_P, X', r)$
  - Run **A** on C' and output Y'.
  - $|\Pr(R(X,A(E(X))) - \Pr(R(X,Y'))| = |\alpha_0 - \alpha_1|$ and is negligible.

Equivalence Proof…

If a scheme is semantically secure, then it has the indistinguishability of encryptions property:

- Suppose not, and **A** chooses
  - A pair X_0 , $X_1 \in \{0,1\}^n$
  - For which it can distinguish with advantage ε
- Choose
  - distribution $D_n = \{X_0 , X_1 \}$
  - Relation **R** which is "equality with X"
- For any **A'** that does not get C = E(K_P, X ,r) and outputs Y'

  Prob[**R(**X, Y')]= ½
- By simulating **A** and outputting Y= X_b for guess $b \in \{0,1\}$

  Prob[**R(**X, Y)] ≥ ½ + ε

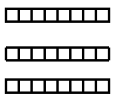
Concatenations

- If (G,E,D) is a semantically secure cryptosystem, then an Adversary **A** which
  - Chooses X_0, $X_1 \in \{0,1\}^n$
  - Receives k independent encryptions of X_b for $b \in_R \{0,1\}$
  - has to decide whether b = 0 or b = 1.
- Cannot have a non-negligible advantage. Namely,

  | Pr(A(E(X_0),…,E(X_0))=1) - Pr(A(E(X_1),…,E(X_1))=1) | is negligible.
- Proof: hybrid argument
  - Let H_j be a hybrid where A receives j encryptions of X_0 followed by k-j encryptions of random X_1
  - Suppose **|** Pr(A(H_k)=1) **-** Pr(A(H_0)=1) **|** is not negligible.
  - Then ∃j s.t. **|** Pr(A(H_{j+1})=1) **-** Pr(A(H_j)=1) **|** is not negligible.
  - Can use it to distinguish between E(X_0) and E(X_1)

From single bit to many bits

- If there is an encryption scheme that can hide E(K_P, 0 ,r) from E(K_P, 1 ,r), then we can construct a full blown (for any length messages) semantically secure cryptosystem by concatenation.
- The construction:
  - Each bit in the message $m \in \{0,1\}^k$ is encrypted separately
- Proof: a hybrid argument
  - Using definition of indistinguishability of encryption
  - Suppose adversary chooses X_0 , $X_1 \in \{0,1\}^k$
  - Let:
    - D_0 be the distribution on encryptions of X_0
    - D_k be the distribution on encryptions of X_1
    - D_i be the distribution where the first i bits are from X_0 and the last k-i bits are from X_1



A construction that fails

- Trapdoor one-way permutation f_p: $\{0,1\}^n \rightarrow \{0,1\}^n$
  - K_P (Public) and K_S (secret) are the keys of the trapdoor permutation.
  - Computing f_p is easy given K_p.
  - Computing f_p^{-1} is easy given K_S. Hard otherwise.
- Why not encrypt m by sending f_p(m)?
  - f_p(m) might reveal partial information about m.
  - For example, if f_p(m) is trapdoor one-way, so is g_p: $\{0,1\}^{2n} \rightarrow \{0,1\}^{2n}$, defined as $g_p(x,y)=(x,f_p(y))$.
  - g_p(m) is not semantically secure, since it reveals half the bits of m.
- In fact, any deterministic encryption scheme cannot provide semantic security

Construction: from trapdoor one-way permutation

- Key generation: K_P (Public) and K_S (secret) are the keys of a trapdoor permutation
- Encryption: to encrypt a message $m \in \{0,1\}^k$
  - select $x \in_R \{0,1\}^n$ and $r \in_R \{0,1\}^n$
  - Compute $g(x) = [x \cdot r, f_P(x) \cdot r, f_P^{(2)}(x) \cdot r, \ldots f_P^{(k-1)}(x) \cdot r]$
  - Send m xored with g(x), and in addition $y=f_P^{(k)}(x)$ and r

$$(g(x) \oplus m, f_P^{(k)}(x), r)$$

- Decryption: given (c, y, r)
  - extract $x = f_P^{(-k)}(y)$ using K_S
  - compute g(x) using r
  - extract m by xoring c with g(x)

Security of construction

Claim: given $y=f_P^{(k)}(x)$, the value of g(x) is indistinguishable from random

Proof:

- it is sufficient to show that given $y=f_P(x)$, r, for a randomly chosen r, the value of $x \cdot r$ is indistinguishable from random (this is the Goldreich-Levin hardcore predicate)
- If the adversary could have reconstructed $x \cdot r$ exactly, it could have revealed x (given sufficient samples)
- We can only assume that for many x's, the adversary can use y to guess $x \cdot r$ with probability $½+\varepsilon$
- The GL proof shows a reconstruction algorithm, that given such an adversary constructs a short *list* of candidates for x. It then checks which of these values satisfies $f_p(x)=y$.

Example

- Blum-Goldwasser cryptosystem
  - Based on the Blum, Blum, Shub pseudo-random generator
  - The permutation defined by $N= P \cdot Q$, where P,Q = 3 mod 4
  - The trapdoor is P,Q
  - For $x \in Z_N^*$, x is a quadratic residue

$$f_N(x)=x^2 \bmod N$$

One-way encryption is sufficient for semantic security against chosen plaintext attack

Call an encryption scheme **one-way** if given $c=E(K_P, m, s)$ for random m and s it is hard to find m

- This is the weakest form of security one can expect from a ``self-respecting" cryptosystem

- Can use it to construct a single-bit indistinguishable scheme:
- To encrypt a bit $b \in \{0,1\}$:
  - choose random x, s and r
  - Send (c,r,b') where
    - $c=E(K_P, x, s)$
    - $b'= x \cdot r \oplus b$

Security: from the Goldreich-Levin reconstruction algorithm