Introduction to Security Reduction

Lecture 7: Analysis

(Towards A Correct Reduction)

Adversary

My IQ is up to 186.

My interest is breaking schemes.

You want me to help you solve problem?

Fool me first!

Computational Complexity Theory

Outline

1 Overview

2 Step 1: Indistinguishable Simulation

- Random and Independent
- Simulation with a General Function
- Simulation with a Linear System
- Simulation with a Polynomial

3 Step 2: Indistinguishable Attack

- Attack Revisited
- Requirements
- Absolutely Hard Problems

4 Correctness of Analysis

Outline

Overview

Correct Security Reduction: Even if the attack on the simulated scheme is launched by an adversary who is

- **Malicious**
- **Computationally unbounded**,

the advantage of solving the underlying hard problem in polynomial time must be still non-negligible.

Overview

The analysis of correctness is to analyze that

- **The simulation is indistinguishable.**
- **The attack is a useful attack with non-negligible probability.**

Note: The simulation requires to be indistinguishable before the adversary launches a successful attack. However, it is not necessary to program the whole simulation indistinguishable from real attack.

Outline

Indistinguishable Simulation

A simulation is indistinguishable if

- **All responses to queries are correct.**
- **All simulated random numbers are random and independent.**

Note: We can analyze the correctness of responses in the simulated scheme after computing each response. Therefore, proving the "indistinguishable simulation" in the analysis is to analyze random and independent.

Outline

Random and Independent

Random numbers (including random group elements) are very common in constructing cryptographic schemes. They are used in

- Key Generation.
- Signature Generation. (could be)
- Ciphertext Generation

Suppose each number in the set $\{A_1, A_2, \cdots, A_n\} \in \mathbb{Z}_p$ is a random number. This means that each number is chosen randomly and independently from $\mathbb{Z}_p$, and all numbers are uniformly distributed in $\mathbb{Z}_p$.

In a simulated scheme, if random numbers are simulated with a function, we must prove that these simulated random numbers are also random and independent from the point of view of the adversary.

Random and Independent

Let (A, B, C) be three random integers chosen from the space $\mathbb{Z}_p$. The concept of *random and independent* can be explained as follows.

- **Random.** C is equal to any integer in $\mathbb{Z}_p$ with probability $\frac{1}{p}$.
- **Independent.** C cannot be computed from A and B.

Suppose an adversary is only given A and B. The adversary then has no advantage in guessing the integer C and can only guess the integer C correctly with probability $\frac{1}{p}$.

Note: If A, B are two integers randomly chosen from the space $\mathbb{Z}_p$ and $C = A + B \bmod p$, we still have that C is equivalent to a random number chosen from $\mathbb{Z}_p$. However, A, B, C are not independent, because C can be computed from A and B.

Outline

Important Lemma

Lemma

Suppose a real scheme and a simulated scheme generate integers (A, B, C) with different methods described as follows.

- *In the real scheme, (A, B, C) are randomly chosen from $\mathbb{Z}_p$.*
- *In the simulated scheme, (A, B, C) are computed by a function with random (w, x, y, z) from $\mathbb{Z}_p$ denoted by $(A, B, C) = F(w, x, y, z)$.*

Suppose the adversary knows the function F from the reduction algorithm but not (w, x, y, z). The simulated scheme is indistinguishable from the real scheme if for any given (A, B, C) from $\mathbb{Z}_p$, the number of solutions (w, x, y, z) satisfying $(A, B, C) = F(w, x, y, z)$ is the same.

That is, any (A, B, C) from $\mathbb{Z}_p$ will be generated with the same probability in the simulated scheme. This lemma will be applied in next lemmas.

Example (1)

$$(A, B, C) = F(x,y) = (x, y, x+y) \tag{1}$$

Example (1)

$$(\ A,\ B,\ C\ ) \quad = \quad F(x,y) = (\ x\ ,\ y\ ,\ x+y\ ) \tag{1}$$

Distinguishable. In this function, we have

$$\begin{aligned} x &= A, \\ y &= B, \\ x+y &= C. \end{aligned}$$

If the given (A,B,C) satisfies $A+B=C$, the function has one solution

$$< x,y >=< A,B > .$$

Otherwise, there is no solution. Therefore, the simulated A,B,C are not random and independent. To be precise, C can be computed from $A+B$.

Example (2)

$$(A,\ B,\ C) \quad = \quad F(x,y,z) = (\ x\ ,\ \ y\ ,\ \ z+3\ ) \tag{2}$$

Example (2)

$$(\ A,\ B,\ C\ ) \quad = \quad F(x,y,z) = (\ x\ ,\ y\ ,\ z+3\ ) \tag{2}$$

Indistinguishable. In this function, we have

$$\begin{aligned} x &= A, \\ y &= B, \\ z+3 &= C. \end{aligned}$$

For any given (A,B,C), the function has one solution

$$< x,y,z > = < A, B, C-3 > .$$

Therefore, A,B,C are random and independent.

Example (3)

$$(\ A,\ B,\ C\ ) \quad = \quad F(x,y,z) = (\ x\ ,\ y\ ,\ z + 4 \cdot xy\ ) \tag{3}$$

Example (3)

$$(A,\ B,\ C\ ) \quad = \quad F(x,y,z) = (\ x\ ,\ y\ ,\ z + 4 \cdot xy\ ) \tag{3}$$

Indistinguishable. In this function, we have

$$\begin{aligned} x &= A, \\ y &= B, \\ z + 4xy &= C. \end{aligned}$$

For any given (A,B,C), the function has one solution

$$< x,y,z > = < A,B,C-4AB > .$$

Therefore, A,B,C are random and independent.

Example (4)

$$(\,A,\ B,\ C\,) \quad = \quad F(w,x,y,z) = (\ x+w\ ,\ \ y\ ,\ \ z+w\cdot x\ ) \tag{4}$$

Example (4)

$$(\,A,\ B,\ C\,) \quad = \quad F(w,x,y,z) = (\ x+w\ ,\ y\ ,\ z+w\cdot x\ ) \tag{4}$$

Indistinguishable. In this function, we have

$$\begin{aligned} x+w &= A, \\ y &= B, \\ z+w\cdot x &= C. \end{aligned}$$

For any given (A,B,C), the function has p different solutions

$$< w,x,y,z >=< w, A-w, B, C-w(A-w) >,$$

where w can be any integer from $\mathbb{Z}_p$. Therefore, A, B, C are random and independent.

Outline

Simulation with a Linear System

A general system of n linear equations (or linear system) over $\mathbb{Z}_p$ with n unknown secrets $(x_1, x_2, \cdots, x_n)$ can be written as

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = y_2 \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = y_n \end{array},$$

where the a_{ij} are the coefficients of the system, and $y_1, y_2, \cdots, y_n$ are constant terms from $\mathbb{Z}_p$. We define $\mathbb{A}$ as the coefficient matrix,

$$\mathbb{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}.$$

Lemma

Lemma

Suppose a real scheme and a simulated scheme generate integers $(A_1, A_2, \cdots, A_n)$ with different methods described as follows.

- *In the real scheme, $(A_1, A_2, \cdots, A_n)$ are random integers from $\mathbb{Z}_p$.*
- *In the simulated scheme, let $(A_1, A_2, \cdots, A_n)$ be computed by*

$$(A_1, A_2, \cdots, A_n)^\top = \mathbb{A} \cdot X^\top = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \bmod p,$$

where $x_1, x_2, \cdots, x_n$ are random integers chosen from $\mathbb{Z}_p$.

Suppose the adversary knows $\mathbb{A}$ but not X. If the determinant of $\mathbb{A}$ is nonzero, the simulated scheme is indistinguishable from the real scheme.

Example

$$(A_1, A_2, A_3) = (x_1 + 3x_2 + 3x_3 , x_1 + x_2 + x_3 , 3x_1 + 5x_2 + 5x_3) \quad (5)$$

Example

$$(A_1, A_2, A_3) = (x_1 + 3x_2 + 3x_3 , x_1 + x_2 + x_3 , 3x_1 + 5x_2 + 5x_3) \quad (5)$$

Distinguishable. In this function, we have

$$\begin{aligned} x_1 + 3x_2 + 3x_3 &= A_1, \\ x_1 + x_2 + x_3 &= A_2, \\ 3x_1 + 5x_2 + 5x_3 &= A_3. \end{aligned}$$

It is easy to verify that the determinant of the coefficient matrix satisfies

$$\begin{vmatrix} 1 & 3 & 3 \\ 1 & 1 & 1 \\ 3 & 5 & 5 \end{vmatrix} = 0.$$

Therefore, (A_1, A_2, A_3) are not random and independent. To be precise, given A_1 and A_2, we can compute A_3 by $A_3 = A_1 + 2A_2$.

Outline

Polynomial

Let $f(x) \in \mathbb{Z}_p[x]$ be a $(q-1)$-degree polynomial function defined as

$$f(x) = a_{q-1}x^{q-1} + a_{q-2}x^{q-2} + \cdots + a_1 x + a_0,$$

where there are q coefficients, and all coefficients a_i are randomly chosen from $\mathbb{Z}_p$. There are q number of coefficients.

Note: We assume that the simulator randomly chooses a_i. Therefore, the polynomial $f(x)$ is unknown to the adversary.

Lemma

Lemma

Suppose a real scheme and a simulated scheme generate integers $(A_1, A_2, \cdots, A_n)$ with different methods described as follows.

- *In the real scheme, let $(A_1, A_2, \cdots, A_n)$ be random integers from $\mathbb{Z}_p$.*
- *In the simulated scheme, let $(A_1, A_2, \cdots, A_n)$ be computed by*

$$(A_1, A_2, \cdots, A_n) = (f(m_1), f(m_2), \cdots, f(m_n)),$$

where $m_1, m_2, \cdots, m_n$ are n distinct integers in $\mathbb{Z}_p$ and f is a $(q-1)$-degree polynomial.

Suppose the adversary knows $m_1, m_2, \cdots, m_n$ but not $f(x)$. The simulated scheme is indistinguishable from the real scheme if $q \geq n$.

Lemma Explanation

We can rewrite the simulation as

$$
\begin{aligned}
& (A_1, A_2, \cdots, A_n)^\top \\
= \; & (f(m_1), f(m_2), \cdots, f(m_n))^\top \\
= \; & \begin{pmatrix} m_1^{q-1} & m_1^{q-2} & m_1^{q-3} & \cdots & m_1^0 \\ m_2^{q-1} & m_2^{q-2} & m_2^{q-3} & \cdots & m_2^0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ m_n^{q-1} & m_n^{q-2} & m_n^{q-3} & \cdots & m_n^0 \end{pmatrix} \cdot \begin{pmatrix} a_{q-1} \\ a_{q-2} \\ \vdots \\ a_0 \end{pmatrix} \text{ mod } p.
\end{aligned}
$$

The coefficient matrix is the Vandermonde matrix, whose determinant is nonzero. The number of solutions for each $(A_1, A_2, \cdots, A_n)$ is the same. Therefore the simulated scheme is indistinguishable from the real scheme.

Outline

1. Overview

2. Step 1: Indistinguishable Simulation
   - Random and Independent
   - Simulation with a General Function
   - Simulation with a Linear System
   - Simulation with a Polynomial

3. **Step 2: Indistinguishable Attack**
   - Attack Revisited
   - Requirements
   - Absolutely Hard Problems

4. Correctness of Analysis

A Signature Scheme As Example

KeyGen: The key pair is $pk = (g, g^\alpha, g^\beta), \ \ sk = (\alpha, \beta)$.

Sign: The signature on $m \in \mathbb{Z}_p$ is

$$\sigma_m = \left(r, g^{\frac{\beta - r}{\alpha - m}}\right),$$

where r is randomly chosen and unique for each message.

Verify: The signature σ_m is valid if $e(\sigma_m, g^\alpha g^{-m}) = e(g^\beta g^{-r}, g)$.

Proof. Given $(g, g^a, g^{a^2}, \cdots, g^{a^q})$, the simulator chooses a q-degree polynomial $f(x) \in \mathbb{Z}_p[x]$ and sets $pk = (g, g^a, g^{f(a)})$.

Outline

1. Overview
2. Step 1: Indistinguishable Simulation
   - Random and Independent
   - Simulation with a General Function
   - Simulation with a Linear System
   - Simulation with a Polynomial
3. **Step 2: Indistinguishable Attack**
   - **Attack Revisited**
   - Requirements
   - Absolutely Hard Problems
4. Correctness of Analysis

Useful Attack and Useless Attack

The adversary can launch an adaptive attack to break the scheme.

For example: The forged signature on m^* can be any one as follows:

$$\left(r_1, g^{\frac{\beta - r_1}{\alpha - m^*}}\right), \left(r_2, g^{\frac{\beta - r_2}{\alpha - m^*}}\right), \left(r_3, g^{\frac{\beta - r_3}{\alpha - m^*}}\right), \cdots, \left(r_n, g^{\frac{\beta - r_n}{\alpha - m^*}}\right)$$

- A forged signature with a distinct r can be seen as a different attack.
- The adversary will adaptively pick $r_i \in \{r_1, \cdots, r_n\}$ in forgery.

Useful Attack and Useless Attack

The adversary can launch an adaptive attack to break the scheme.

For example: The forged signature on m^* can be any one as follows:

$$\left(r_1, g^{\frac{\beta-r_1}{\alpha-m^*}}\right), \left(r_2, g^{\frac{\beta-r_2}{\alpha-m^*}}\right), \left(r_3, g^{\frac{\beta-r_3}{\alpha-m^*}}\right), \cdots, \left(r_n, g^{\frac{\beta-r_n}{\alpha-m^*}}\right)$$

- A forged signature with a distinct r can be seen as a different attack.
- The adversary will adaptively pick $r_i \in \{r_1, \cdots, r_n\}$ in forgery.

Let $(r^*, g^{\frac{\beta-r^*}{\alpha-m^*}})$ be the forged signature,where

$$g^{\frac{\beta-r^*}{\alpha-m^*}} = g^{\frac{f(a)-r^*}{a-m^*}}.$$

- If $r^* = f(m^*)$, the forged signature is a useless attack.
- If $r^* \neq f(m^*)$, the forged signature is a useful attack.

Useful Attack and Useless Attack

Let $(r^*, g^{\frac{\beta - r^*}{\alpha - m^*}})$ be the forged signature,where

$$g^{\frac{\beta-r^*}{\alpha-m^*}} = g^{\frac{f(a)-r^*}{a-m^*}}.$$

- If $r^* = f(m^*)$, the forged signature is a useless attack because the forged signature is also computable by the simulator.
- If $r^* \neq f(m^*)$, the forged signature is a useful attack because

$$g^{\frac{\beta-r^*}{\alpha-m^*}} = g^{\frac{f(a)-r^*}{a-m^*}} = g^{\frac{f(a)-f(m^*)}{a-m^*}} \cdot g^{\frac{f(m^*)-r^*}{a-m^*}}.$$

Since $(a - m^*)|(f(a) - f(m^*))$ and $f(m^*) - r^* \neq 0$, we have

$$(-m^*, g^{\frac{1}{a-m^*}}),$$

which can be computed and it is the problem solution of q-SDH problem.

Outline

Requirements

Let $Attack_i$ be one specific way for breaking the proposed scheme. That is, launching a specific way of queries and challenge.

- Let $\{Attack_1, Attack_2, \cdots, Attack_n\}$ be the set of all potential attacks.
- Some attacks are useful attacks and some are useless attacks.
- The adversary will launch an adaptive $Attack^*$ from the set.

Requirement: The adversary has no advantage in identifying which attack is a useless attack. Otherwise, the adaptive attack will be useless.

Requirements

Let $(r^*, g^{\frac{\beta - r^*}{\alpha - m^*}})$ be the forged signature,where

$$g^{\frac{\beta - r^*}{\alpha - m^*}} = g^{\frac{f(a) - r^*}{a - m^*}}.$$

- If $r^* = f(m^*)$, the forged signature is a useless attack.
- If $r^* \neq f(m^*)$, the forged signature is a useful attack.

Requirement: We prove that the adversary has **no advantage** in computing $f(m^*)$ from what it knows.

The corresponding approach is called absolutely hard problem.

Outline

1. Overview
2. Step 1: Indistinguishable Simulation
   - Random and Independent
   - Simulation with a General Function
   - Simulation with a Linear System
   - Simulation with a Polynomial
3. **Step 2: Indistinguishable Attack**
   - Attack Revisited
   - Requirements
   - **Absolutely Hard Problems**
4. Correctness of Analysis

Absolutely Hard Problems

A computing problem is absolutely hard if computationally unbounded adversary has no advantage in solving it. It is also known as information-theoretic security.

For example

Given: (g, g^{x+a^2}), where x, a are random integers from $\mathbb{Z}_p$.
Compute: a

The problem is absolutely hard because given a problem instance, any integer in $\mathbb{Z}_p$ can be the potential solution with the same probability.

Absolutely Hard Problems

KeyGen: The key pair is $pk = (g, g^{\alpha}, g^{\beta}), \quad sk = (\alpha, \beta)$.
Sign: The signature on $m \in \mathbb{Z}_p$ is

$$\sigma_m = \left(r, g^{\frac{\beta - r}{\alpha - m}}\right)$$

Verify: The signature σ_m is valid if $e(\sigma_m, g^{\alpha} g^{-m}) = e(g^{\beta} g^{-r}, g)$.

Proof. Given $(g, g^{a}, g^{a^2}, \cdots, g^{a^q})$, the simulator chooses a q-degree polynomial $f(x) \in \mathbb{Z}_p[x]$ and sets $pk = (g, g^{a}, g^{f(a)})$.

Given: $a, f(a)$ (from the view of computationally unbounded adversary)
Compute: $f(m^*)$ (to launch a useless attack)

Absolutely Hard Problems

Question: How to prove that a computing problem is absolutely hard?

Absolutely Hard Problems

Question: How to prove that a computing problem is absolutely hard?

Answer: We prove that the problem instance and the problem solution are random and independent.

In the previous example, we prove that

$$a, f(a), f(m^*)$$

are random and independent.

Note: The problem instance is generated by the reduction algorithm that the adversary knows.

Example (1)

Suppose (a, Z, c, x) satisfies $Z = ac + x \bmod p$, where $a, x \in \mathbb{Z}_p$ and $c \in \{0, 1\}$ are randomly chosen.

Given (a, Z), the adversary has no advantage in distinguishing whether Z is computed from either $a \cdot 0 + x$ or $a \cdot 1 + x$ except with probability $1/2$.

The reason is that a, Z, c are random and independent.

Example (2)

Suppose $(a, Z_1, Z_2, \cdots, Z_{n-1}, Z_n, x_1, x_2, \cdots, x_n)$ satisfies $Z_i = a + x_i \bmod p$, where a, x_i for all $i \in [1, n]$ are randomly chosen from $\mathbb{Z}_p$.

Given $(a, Z_1, Z_2, \cdots, Z_{n-1})$, the adversary has no advantage in computing $Z_n = a + x_n$ except with probability $1/p$.

The reason is that $a, Z_1, Z_2, \cdots, Z_n$ are random and independent.

Example (3)

Suppose $(f(x), Z_1, Z_2, \cdots, Z_n, x_1, x_2, \cdots, x_n)$ satisfies $Z_i = f(x_i)$, where $f(x) \in \mathbb{Z}_p[x]$ is an n-degree polynomial randomly chosen from $\mathbb{Z}_p$.

Given $(Z_1, Z_2, \cdots, Z_n,\ x_1, x_2,\ \cdots, x_n)$, the adversary has no advantage in computing a pair $(x^*, f(x^*))$ for a new x^* different from x_i except with probability $1/p$.

The reason is that $Z_1, Z_2, \cdots, Z_n, f(x^*)$ are random and independent.

Example (4)

Suppose $(\mathbb{A}, Z_1, Z_2, \cdots, Z_{n-1}, Z_n, x_1, x_2, \cdots, x_n)$ satisfies $|\mathbb{A}| \neq 0 \bmod p$ and Z_i is computed by $Z_i = \sum_{j=1}^{n} a_{i,j} x_j \bmod p$, where

- $\mathbb{A}$ is an $n \times n$ matrix whose elements are from $\mathbb{Z}_p$, and
- x_j for all $j \in [1, n]$ are randomly chosen from $\mathbb{Z}_p$.

Given $(\mathbb{A}, Z_1, Z_2, \cdots, Z_{n-1})$, the adversary has no advantage in computing $Z_n = \sum_{j=1}^{n} a_{n,j} x_j$ except with probability $1/p$.

The reason is that $Z_1, Z_2, \cdots, Z_n$ are random and independent.

Example (5)

Suppose (g, h, Z, x, y) satisfies $Z = g^x h^y$, where $x, y \in \mathbb{Z}_p$ are randomly chosen.

Given $(g, h, Z) \in \mathbb{G}$, the adversary has no advantage in computing (x, y) except with probability $1/p$. Once the adversary finds x, it can immediately compute y with Z.

The reason is that g, h, Z, x are random and independent.

Example (6)

Suppose (g, h, Z, x, c) satisfies $Z = g^x h^c$, where $x \in \mathbb{Z}_p$ and $c \in \{0, 1\}$ are randomly chosen.

Given $(g, h, Z) \in \mathbb{G}$, the adversary has no advantage in distinguishing whether Z is computed from either $g^x h^0$ or $g^x h^1$, except with probability $1/2$.

The reason is that g, h, Z, c are random and independent.

Outline

Analysis Structure

Proof. Suppose there exists an adversary who can break the proposed scheme......

Simulation + Solution

This completes the simulation and the solution. The correctness is analyzed as follows.

Indistinguishable simulation. Analyze that the simulation is indistinguishable when simulation is successful.

Probability of successful simulation and useful attack. Analyze the success probability of solving hard problem.

Advantage and time cost. Analyze advantage and time cost of solving hard problem.

This completes the proof. □

Analysis Structure

Indistinguishable simulation. XXXXXXXXXXXXX

Probability of successful simulation and useful attack. XXXXXXX

Advantage and time cost. XXXXXXXXXXXXXX

- We don't have to follow the above structure to give the analysis.
- However, indistinguishable simulation and non-negligible advantage of solving problem must be analyzed.

I lost !

You deserve my help as long as you can fool me !


END


Introduction
to Security
Reduction