Notions of Security - 01.01.2009

Priyesh Jaipuriar[1] Prateek Agarwal[2]

[1]05CS1013 { priyeshjaipuriar@gmail.com }
[2]05CS1021 { prat0318@gmail.com }

I. Introduction

This lecture describes the two fundamental definitions of security and proves their equivalence. The first definition, **Semantic Security** is a computational-complexity analog of Shannon's definition of perfect privacy. It represents the infeasibility to learn anything about the plaintext from the ciphertext.

The second definition, **Message Indistinguishability** interprets security as the infeasibility of distinguishing between encryptions of a given pair of messages.

II. Semantic Security (SS)

Definition 1: Loosely speaking, semantic security means that nothing can be gained by looking at a ciphertext. Following the simulation paradigm, this means that whatever can be efficiently learned from the ciphertext can also be efficiently learned from scratch (or from nothing) i.e. ciphertext reveals no information about the message.

For every distribution X over $\{0, 1\}^n$ and for every partial information h : $\{0, 1\}^n \rightarrow \{0, 1\}^n$

For every interesting information f : $\{0, 1\}^n \rightarrow \{0, 1\}^n$

For every attacking algorithm A running in time t' $\leq$ t(n) [t(n) is a polynomial in n], there exists a simulating algorithm S such that :

$$Pr_{m \leftarrow X, (p_k, s_k) \leftarrow G(n)}[A(E(m, p_k), p_k, h(m)) = f(m)] \leq Pr_{m \leftarrow X}[S(h(m)) = f(m)] + \epsilon(n)$$

where ϵ(n) is a negligible quantity which depends upon value of n. Ex. ϵ(n) may be $\frac{1}{p(n)}$ where p(n) is a polynomial in n of a large degree.

III. Message Indistinguishability (MI)

Definition 2: Given two encryptions of messages m_0 and m_1, the probability of guessing the message is very close to the random probability of guessing the correct message ($\frac{1}{2}$).

The following technical interpretation of security states that it is infeasible to distinguish the encryptions of two plaintexts (of the same length). That is, such ciphertexts are computationally indistinguishable.

For every two messages m_0, $m_1 \in \{0, 1\}^n$

For every algorithm A that runs in time $\leq$ t(n)

$$Pr_{i \in \{0,1\}, (p_k, s_k) \leftarrow G(n)}[A(E(m_i, p_k), p_k) = i] \leq \frac{1}{2} + \epsilon(n)$$

IV. Equivalence of SS and MI

Theorem 1: A private-key encryption scheme is semantically secure if and only if it has indistinguishable encryptions.

A. SS ⇒ MI

Proof: If X = {m_0, m_1}, f : f(m_0) = 0, f(m_1) = 1, h() : empty output string

From SS, for every adversary A there is a simulator S , s.t.

$$Pr_{m\leftarrow X,(p_k,s_k)\leftarrow G(n)}[A(E(m,p_k),p_k)=i] \leq Pr_{m\leftarrow X}[S()=i]+\epsilon(n)$$

Now since the simulator receives no information : Pr[S() = i] = $\frac{1}{2}$, regardless of S.

Thus,

$$Pr_{i\in\{0,1\},(p_k,s_k)\leftarrow G(n)}[A(E(m_i,p_k),p_k)=i] \leq \frac{1}{2}+\epsilon(n)$$

■

B.

For every m_0, $m_1 \in \{0, 1\}^n$, for every algorithm A that runs in time ≤ t(n), for every a ∈ {0, 1}*,

$$Pr_{(p_k,s_k)\in G(n)}[A(E(m_1,p_k),p_k)=a] - Pr_{(p_k,s_k)\in G(n)}[A(E(m_0,p_k),p_k)=a] \leq 2\epsilon(n) \quad ...(^*)$$

$$(t,\epsilon)-MI \Rightarrow \quad {}^* \equiv \neg(^*) \Rightarrow \quad \neg(t,\epsilon)-MI$$

C.

$$Define, \quad A'(c,p) = \begin{cases} 1, & \text{if } A(c,p)=a \\ 0, & \text{otherwise} \end{cases}$$

$$\therefore \quad Pr_{i\in\{0,1\},(p_k,s_k)\leftarrow G(n)}[A'(E(m_i,p_k),p_k)=i]$$

$$= \frac{1}{2}Pr_{(p_k,s_k)\leftarrow G(n)}[A'(E(m_0,p_k),p_k)=0] + \frac{1}{2}Pr_{(p_k,s_k)\leftarrow G(n)}[A'(E(m_1,p_k),p_k)=1]$$

$$= \frac{1}{2}(1-Pr_{(p_k,s_k)\leftarrow G(n)}[A(E(m_0,p_k),p_k)=a]) + \frac{1}{2}Pr_{(p_k,s_k)\leftarrow G(n)}[A(E(m_1,p_k),p_k)=a]$$

$$= \frac{1}{2} + \frac{1}{2}(Pr_{(p_k,s_k)\leftarrow G(n)}[A(E(m_1,p_k),p_k)=a] - Pr_{(p_k,s_k)\leftarrow G(n)}[A(E(m_0,p_k),p_k)=a])$$

$$< \frac{1}{2}+\epsilon(n) \quad \Rightarrow \quad (t,\epsilon)-MI \ is \ violated$$

V. MI $\Rightarrow$ SS

Let us assume that

$$(t, \epsilon) - MI \quad \Rightarrow \quad (t', 2\epsilon) - SS)$$

Thus, $\neg$(t', 2ϵ) - SS $\Rightarrow$ $\neg$(t, ϵ) - MI

Define S(z), where z is some information on m.

Pick(p_k, s_k) $\in$ G(n) at random.

Return A(E(0, p_k), p_k, z)

(Note that the run time of S is running time of A + poly(n).)

$\neg$ (t', 2ϵ) $\Rightarrow$

$$Pr_{m \leftarrow X, (p_k, s_k) \leftarrow G(n)}[A(E(m, p_k), p_k, h(m)) = f(m)] > Pr_{m \leftarrow X}[S(h(m)) = f(m)] + 2\epsilon(n)$$

$$or,\ Pr_{m \leftarrow X, (p_k, s_k) \leftarrow G(n)}[A(E(m, p_k), p_k, h(m)) = f(m)] > Pr_{m \leftarrow X, (p_k, s_k) \leftarrow G(n)}[A(E(0, p_k), p_k, h(m)) = f(m)] + 2\epsilon(n)$$

$$or, \sum_m Pr[X = m](Pr_{p_k, s_k \leftarrow G(n)}[A(E(X, p_k), p_k, h(X)) = f(X)] - Pr_{(p_k, s_k) \leftarrow G(n)}[A(E(0, p_k), p_k, h(X)) = f(X)])$$

$> 2\epsilon$(n) $\Rightarrow$ $\exists$ m' $\in$ X, s.t.

$$Pr_{(p_k, s_k) \leftarrow G(n)}[A(E(m', p_k), p_k, h(m')) = f(m')] - Pr_{p_k, s_k \leftarrow G(n)}[A(E(0, p_k), p_k, h(m')) = f(m')] > 2\epsilon(n)$$

$\Rightarrow$ as there exists a pair of messages for which (*) does not hold

$\Rightarrow$ (t, ϵ) - MI does not hold.