Cryptography: Semantic security and Message indistinguishability

Shashi Narayan and Arpit Kumar

January 11, 2009

1 Semantic Security

The notion of semantic security was first put forward by Goldwasser and Micali in 1982. It is a widely-used definition for security in an asymmetric key encryption algorithm. For a cryptosystem to be semantically secure, it must be infeasible for a computationally-bounded adversary to derive significant information about a message (plaintext) when given only its ciphertext and the corresponding public encryption key. Semantic security considers only the case of a "passive" attacker.

Definition:

For every distriution X over $\{0,1\}^n$, for every partial information $h : \{0,1\}^n \to \{0,1\}^n$, for every interesting information $f : \{0,1\}^n \to \{0,1\}^*$, for every attacking algorithm A running in time $t' \leq t(n)$, [$t(n)$ is a polynomial in n], there exits a simulating algorithm S such that:

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

Here $\varepsilon(n)$ is a negligile quantity which depends upon n. For example $\varepsilon(n)$ may be $\frac{1}{p(n)}$.

Semantic security tries to attempt ideal security. In simple language, it says that in spite of oserving the ciphertext, attacker obtains no extra interesting oservation than the case when he has not seen the ciphertext.

2 Message Indistinguishability

For every two messages $m_0, m_1 \in \{0,1\}^n$ and for every attacking algorithm A that runs in time $\leq t(n)$

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

3 Equivalence of SS and MI

Goldwasser and Micali demonstrated that Semantic Security (SS) is equivalent to the property of message Indistinguishability (MI). This equivalence allowed for security proofs of practical cryptosystems, and consequently the indistinguishability definition is used more commonly than the original definition of semantic security.

3.1 SS $\Rightarrow$ MI

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, such that

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

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

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

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

$$Pr_{(p_k,S_k)\in G(n)}[A(E(m_1, p_k), p_k) = a] - Pr_{(p_k,S_k)\in G(n)}[A(E(m_0, p_k), p_k) = a] \leq 2 \in (n)$$

Let's call above equation as $(*)$. Then we can say that,

$$(t, \in) - MI \Rightarrow (*) \equiv\sim (t, \in) - MI$$

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

So, $Pr_{\substack{i \in \{0,1\} \\ (p_k, S_k) \leftarrow G(n)}} [A'(E(m_i, p_k), p_k) = i]$

$$\begin{aligned}
&= \frac{1}{2} Pr_{(p_k,S_k)\leftarrow G(n)}[A'(E(m_0, p_k), p_k) = 0] + \frac{1}{2} Pr_{(p_k,S_k)\leftarrow G(n)}[A'(E(m_1, p_k), p_k) = 1] \\
&= \frac{1}{2}(1 - Pr_{(p_k,S_k)\leftarrow G(n)}[A(E(m_0, p_k), p_k) = a]) + \frac{1}{2} Pr_{(p_k,S_k)\leftarrow G(n)}[A(E(m_1, p_k), p_k) = a] \\
&= \frac{1}{2} + \frac{1}{2} Pr_{(p_k,S_k)\leftarrow G(n)}[A(E(m_1, p_k), p_k) = a] - Pr_{(p_k,S_k)\leftarrow G(n)}[A(E(m_0, p_k), p_k) = a] \\
&> \frac{1}{2} + \in (n) \Rightarrow (t, \in) - \text{MI is violated.}
\end{aligned}$$

3.2 MI ⇒ SS

Let's assume that (We will prove it later)

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

then,

$$\sim (t', 2\varepsilon) - SS \Rightarrow \sim (t, \varepsilon) - MI$$

Now define $S(z)$, where z is some information on m and $(p_k, S_k) \in G(n)$ has been selected randomly, returns $A(E(0, p_k), p_k, Z)$.

$\sim (t', 2\varepsilon) - SS$

$$\Rightarrow \quad Pr_{\substack{m \leftarrow X \\ (p_k, S_k) \leftarrow G(n)}} [A(E(m, p_k), p_k, h(m)) = f(m)] > Pr_{(m \leftarrow X)}[S(h(m)) = f(m)] + 2\varepsilon(n)$$

$$\text{or,} \quad Pr_{\substack{m \leftarrow X \\ (p_k, S_k) \leftarrow G(n)}} [A(E(m, p_k), p_k, h(m)) = f(m)]$$

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

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

$$-Pr_{(p_k, S_k) \leftarrow G(n)}[A(E(0, p_k), p_k, h(X)) = f(X)]) > 2\varepsilon(n)$$

$$\Rightarrow \quad \exists m' \in X, \text{st.}(Pr_{(p_k, S_k) \leftarrow G(n)}[A(E(m', p_k), p_k, h(m')) = f(m')]$$

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

$$\Rightarrow \quad \text{as there exist a pair of message for which } (*) \text{ does not hold.}$$

$$\Rightarrow \quad (t, \in) - MI \text{does not hold.}$$