Encryption

Debdeep Mukhopadhyay
IIT Kharagpur

Notion of Security

- "A Good disguise should not reveal the person's height"
  - Shafi Goldwasser and Silvio Micali, 1982

Design of Encryption Algorithms

- Encryption algorithms are used for privacy of data.
  - which means they do not leak any information about the plaintext
- The question is when are we satisfied that the cipher really does not leak?
  - For this we need to know the power of the adversary.

What Shannon said?

- Shannon said in his classical work that using a one-time pad, the cipher achieved “perfect secrecy”
  - no attacker, even with infinite power of computation can obtain any information about the plain-text.
  - But the one-time pad is impractical.

But Cryptographers want provable security

- Lets assume that the attacker is a "probabilistic polynomial time" (PPT) machine
  - that's a more practical assumption!
- So, now the question is can the adversary (attacker) obtain information about the plaintext efficiently?
  - for our purpose efficiently means in polynomial time.

PPT

- Probabilistic Algorithms or randomized algorithms, A, may toss a coin a finite number of times during its computation.
- The output y, and the next step may depend on the results of the preceding coin tosses.
- The coin is in general fair.
- Examples: Primality test algorithms, factoring algorithms etc.

Definition of Semantic Security (SS)

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: $\{0,1\}^n \rightarrow \{0,1\}^*$

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:

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

- Here ε(n) is a negligible quantity.
- Notion tries to attempt ideal security.
- That is the eavesdropper is disconnected from the communication.
- In spite of observing the ciphertext, he obtains no extra interesting observation than the case when he has not seen the ciphertext.

Message Indistinguishability (MI)

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

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

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

- SS and MI are equivalent

Proofs : SS => 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, st.

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

Now, since the simulator receives no information:

$\Pr[S() = i] = 1/2$, regardless of S.

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

SS=>MI

For every $m_0, m_1 \in \{0,1\}^n$, for every algorithm A that

runs in time $\leq t(n)$, for every $a \in \{0,1\}^*$,

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

(*)

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

SS=>MI

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

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

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

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

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

$$>\frac{1}{2}+\in(n)\Rightarrow(t,\in)-MI \text{ is violated.}$$

(t,ε)-MI=>(t',2ε)-SS

- Thus ┐(t',2ε)-SS =>┐ (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) */

(t,ε)-MI=>(t',2ε)-SS

$\neg$(t',2ε)-SS $\Rightarrow$

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

or,

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

or,

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

$$\Rightarrow \exists m' \in X, \text{ st. } \Pr_{(p_k,s_k)\leftarrow G(n)}[A(E(m',p_k),p_k,h(m'))=f(m')] - \Pr_{(p_k,s_k)\leftarrow G(n)}[A(E(0,p_k),p_k,h(m'))=f(m')] > 2\varepsilon(n)$$

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

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