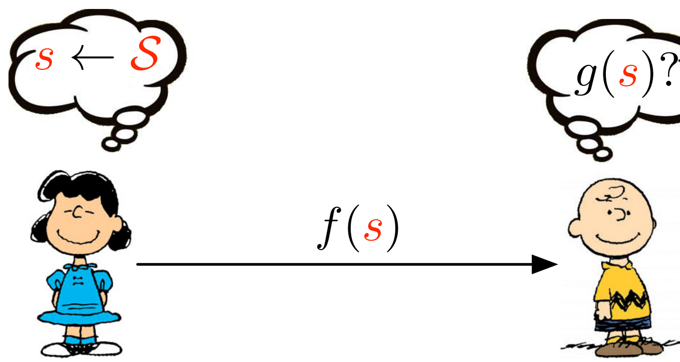
MTAT.07.003 Cryptology II

Semantic Security and Cryptosystems

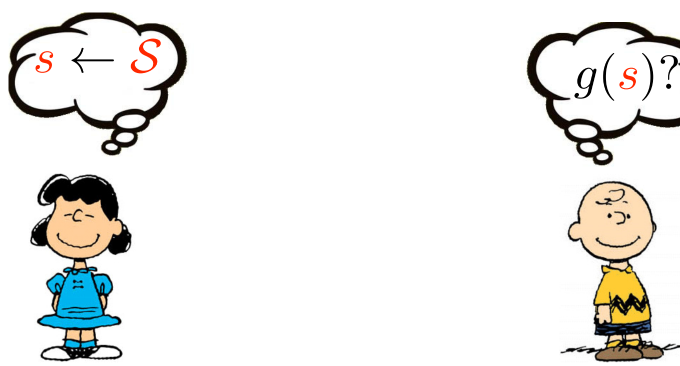
Sven Laur
University of Tartu

Semantic security

Charlie tries to guess $g(s)$ from the description of $\mathcal{S}$ and $f(s)$.



Charlie tries to guess $g(s)$ solely from the description of $\mathcal{S}$.



Indistinguishability implies semantic security

IND-SEM theorem. If for all $s_i, s_j \in \mathrm{supp}(\mathcal{S})$ distributions $f(s_i)$ and $f(s_j)$ are $(2t, \varepsilon)$-indistinguishable, then for all t-time adversaries $\mathcal{A}$:

$$\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) \leq \varepsilon \enspace .$$

Further comments

- Note that function g might be randomised.
- Note that function $g : \mathcal{S} \to \{0,1\}^*$ may extremely difficult to compute.
- It might be even infeasible to get samples from the distribution $\mathcal{S}$.
- The theorem does not hold if $\mathcal{S}$ is specified by the adversary.
- As the proof is non-constructive, there are no *explicit* reductions.

Proof Sketch

A slightly modified formal definition

By definition $\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) = \Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right]$ where

$$\mathcal{G}_0^{\mathcal{A}} \qquad\qquad\qquad\qquad \mathcal{G}_1^{\mathcal{A}}$$

$$\left[\begin{array}{l} s \leftarrow \mathcal{S} \\ g_* \leftarrow \mathcal{A}(f(s)) \\ \textbf{return } [g_* \stackrel{?}{=} g(s)] \end{array}\right. \qquad \left[\begin{array}{l} s \leftarrow \mathcal{S} \\ g_* \leftarrow \operatorname{argmax}_{g_*} \Pr\left[g(s) = g_*\right] \\ \textbf{return } [g_* \stackrel{?}{=} g(s)] \end{array}\right.$$

As a minimising value g_* is *uniquely determined* by $g(\cdot)$, we can express

$$\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) = \Pr\left[s \leftarrow \mathcal{S}_0 : \mathcal{A}(f(s)) = g(s)\right] - \Pr\left[g(s) = g_*\right]$$

Incorrect coin fixing argument

Let $g : \mathcal{S} \times \Omega \to \mathcal{Y}$ is a randomised function. Then by definition

$$\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) = \sum_{\omega \in \Omega} \Pr\left[\omega\right] \cdot \mathsf{Adv}^{\mathsf{sem}}_{f,g_\omega}(\mathcal{A})$$

where $g_\omega(s) \doteq g(s;\omega)$ is a deterministic function.

Hence, the advantage is maximised by a deterministic function, since

$$\sum_{\omega \in \Omega} \Pr\left[\omega\right] \cdot \mathsf{Adv}^{\mathsf{sem}}_{f,g_\omega}(\mathcal{A}) \leq \max_{\omega \in \Omega} \left\{\mathsf{Adv}^{\mathsf{sem}}_{f,g_\omega}(\mathcal{A})\right\} \quad .$$

Formal extension of secret distribution

For a randomised function $g : \mathcal{S} \times \Omega \to \mathcal{Y}$ we can extend secret space

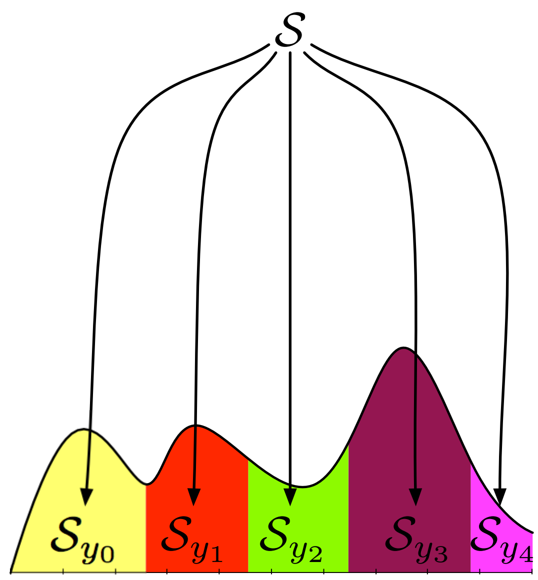
$$f_* : \mathcal{S} \times \Omega \to \mathcal{X} \qquad g_* : \mathcal{S} \times \Omega \to \mathcal{Y}$$

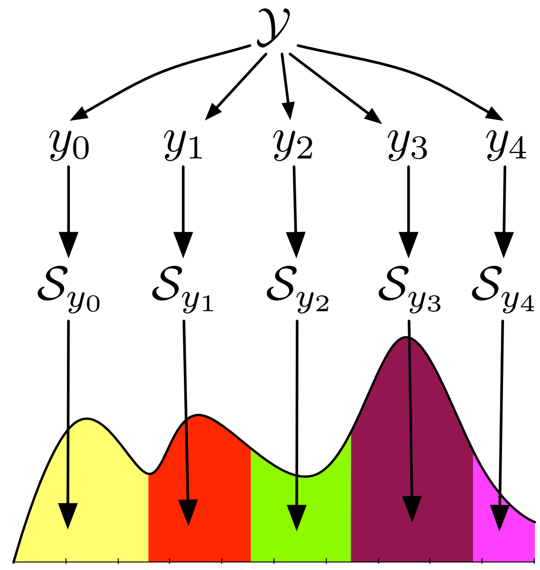
so that observable values $f_*(s, \omega) = f(s)$ do not change and $g_*(s, \omega)$ is deterministic. Now it is easy to see

$$\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) = \mathsf{Adv}^{\mathsf{sem}}_{f_*,g_*}(\mathcal{A}) \enspace .$$

Since the distribution of observable values is same for $f(s)$ and $f_*(s, \omega)$ then the assumptions of the theorem carry over to the extended didtribution.

Sampling idiom





Let $\mathcal{S}_{y_i}$ be the conditional distribution over the set $\{s \in \mathcal{S} : g(s) = y_i\}$ and $\mathcal{Y}$ distribution of final outcomes $g(s)$. Then we get the distribution $\mathcal{S}$ if we first draw y from $\mathcal{Y}$ and then choose s according to $\mathcal{S}_y$.

Resulting guessing game

By using the sampling idiom, we can transform the game into a new form

$$\begin{array}{l}\mathcal{G}_0^{\mathcal{A}} \\ \left[\begin{array}{l} y \leftarrow \mathcal{Y} \\ s \leftarrow \mathcal{S}_y \\ \textbf{return } [g(s) \stackrel{?}{=} \mathcal{A}(f(s))] \end{array}\right.\end{array}$$

where the adversary $\mathcal{A}$ must choose between hypotheses $\mathcal{H}_{y_0} = [y \stackrel{?}{=} y_0]$ for all possible outcomes $y \in \mathcal{Y}$. The success bound for guessing games yields

$$\Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] \leq \max_{y_0, y_1 \in \mathcal{Y}} \mathsf{cd}^{2t}_{f(s)}(\mathcal{H}_{y_0}, \mathcal{H}_{y_1}) + \max_{y_* \in \mathrm{supp}(\mathcal{Y})} \Pr\left[y \leftarrow \mathcal{Y} : y = y_*\right] \enspace .$$

Indistinguishability of conditional distributions

Fix $y_0, y_1 \in \mathcal{Y}$ and let $\mathcal{S}_{y_0}$ and $\mathcal{S}_{y_1}$ be the corresponding distributions. Then for any $2t$-time $\mathcal{B}$ the acceptance probabilities are

$$p_i = \sum_{s_0,s_1} \Pr\left[s \leftarrow \mathcal{S}_{y_0} : s = s_0\right] \Pr\left[s \leftarrow \mathcal{S}_{y_1} : s = s_1\right] \Pr\left[\mathcal{B}(f(s_i)) = 1\right] \quad .$$

Now the difference of acceptance probabilities can be bounded

$$\begin{aligned} |p_0 - p_1| &\leq \sum_{s_0,s_1} \Pr\left[s_0\right] \Pr\left[s_1\right] \left|\Pr\left[\mathcal{B}(f(s_0)) = 1\right] - \Pr\left[\mathcal{B}(f(s_1)) = 1\right]\right| \\ &\leq \max_{s_0,s_1} \left|\Pr\left[\mathcal{B}(f(s_0)) = 1\right] - \Pr\left[\mathcal{B}(f(s_1)) = 1\right]\right| \leq \varepsilon \end{aligned}$$

since all states in $\mathcal{S}$ are $(2t, \varepsilon)$-indistinguishable.

Semantic security of a single encryption

Let $f : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ is a $(2t, \varepsilon)$-pseudorandom function family. Then it is difficult to approximate a function $g(m)$ given only a value $f(m; k)$. In particular, for all t-time adversaries $\mathcal{A}$ and message distributions $\mathcal{M}_0$:

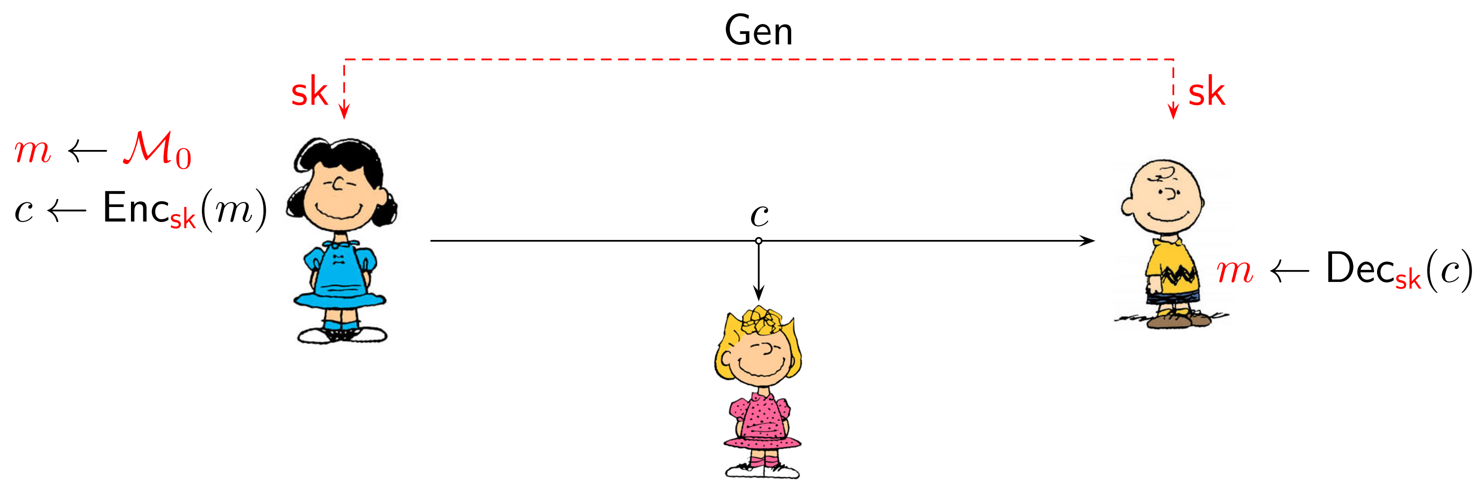
$$\Pr\left[\mathcal{A}(f(m,k)) = g(m)\right] \leq \max_{m_* \in \mathrm{supp}(\mathcal{M}_0)} \Pr\left[g(m_*)\right] + \varepsilon \enspace .$$

Remarks

- We have to consider f as randomised function $f(m) = f(m; k)$.
- The theorem does not hold if $\mathcal{M}_0$ is specified by the adversary.
- The result cannot be generalised for longer multi-block messages.

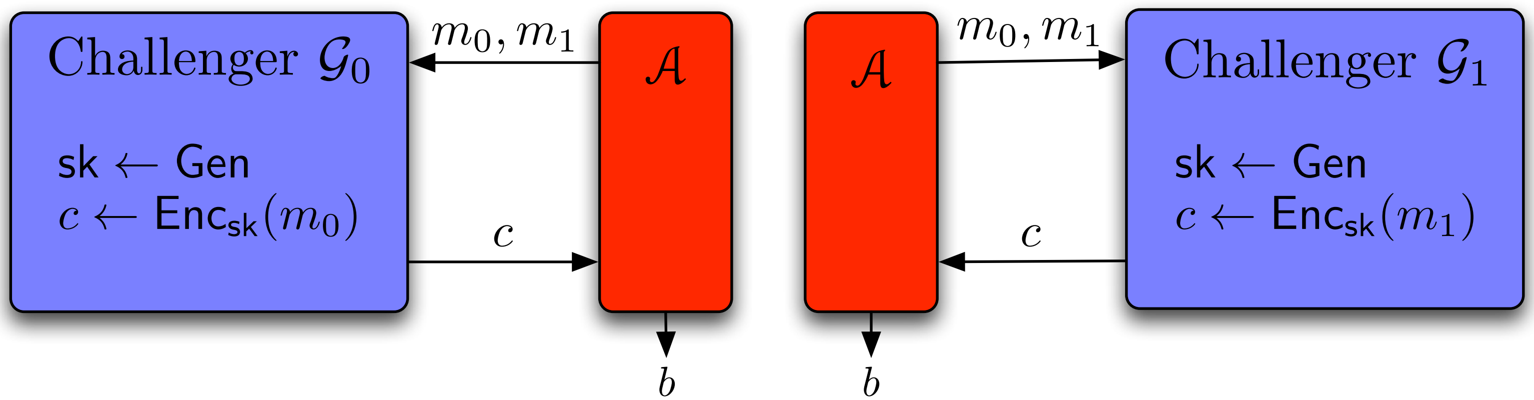
Symmetric Key Encryption

Symmetric key cryptosystem



- ▷ A randomised *key generation algorithm* outputs a *secret key* sk that must be transferred privately to the sender and to the receiver.
- ▷ A randomised *encryption algorithm* $\mathsf{Enc}_{\mathsf{sk}} : \mathcal{M} \rightarrow \mathcal{C}$ takes in a *plaintext* and outputs a corresponding *ciphertext*.
- ▷ A *decryption algorithm* $\mathsf{Dec}_{\mathsf{sk}} : \mathcal{C} \rightarrow \mathcal{M} \cup \{\perp\}$ recovers the plaintext or a special abort symbol $\perp$ to indicate invalid ciphertexts.

Fixed message attack



A cryptosystem $\mathcal{C}$ is *(t, ε)-IND-FPA secure* if for all t-time adversaries $\mathcal{A}$:

$$\mathsf{Adv}^{\mathsf{ind\text{-}fpa}}_{\mathcal{C}}(\mathcal{A}) = \left|\Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right]\right| \leq \varepsilon$$

and thus for any function $g : \mathcal{M} \to \{0,1\}^*$ and for any $\frac{t}{2}$-time adversary $\mathcal{B}$

$$\mathsf{Adv}^{\mathsf{sem}}_{\mathsf{Enc}_{\mathsf{sk}}(\cdot), g}(\mathcal{B}) \leq \varepsilon.$$

Weaknesses of IND-FPA security

Fact I. One-time pad is perfectly IND-FPA secure.

Fact II. If $f : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ is (t, ε)-pseudorandom function, the Electronic Codebook algorithm defined below is $(t, 2\varepsilon)$-IND-FPA secure.

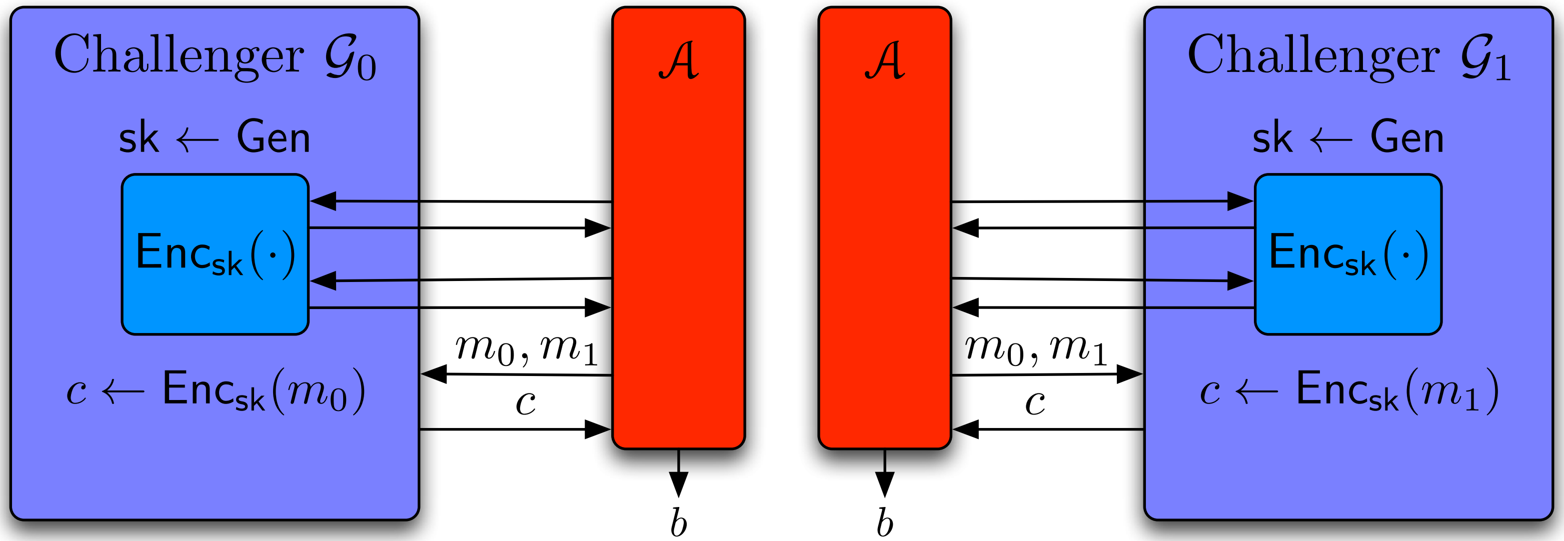
- ▷ **Key generation** Gen: Return $k \xleftarrow[u]{} \mathcal{K}$.
- ▷ **Encryption** $\mathsf{Enc}_{\mathsf{sk}}(\cdot)$: Given $m \in \mathcal{M}$, return $f(m, k)$
- ▷ **Decryption** $\mathsf{Dec}_{\mathsf{sk}}(\cdot)$: Given $c \in \mathcal{C}$, return m such that $f(m, k) = c$.

Observation. If we apply these encryption algorithms for messages m_1, m_2, the resulting ciphertexts c_1, c_2 leak information whether $m_1 = m_2$ or not.

Analysis

- ▷ Separately taken c_1 and c_2 leak no information about m_1 nor m_2.
- ▷ As c_1 is known by the adversary dependence m_1 between m_2 may leak.

Chosen message attack

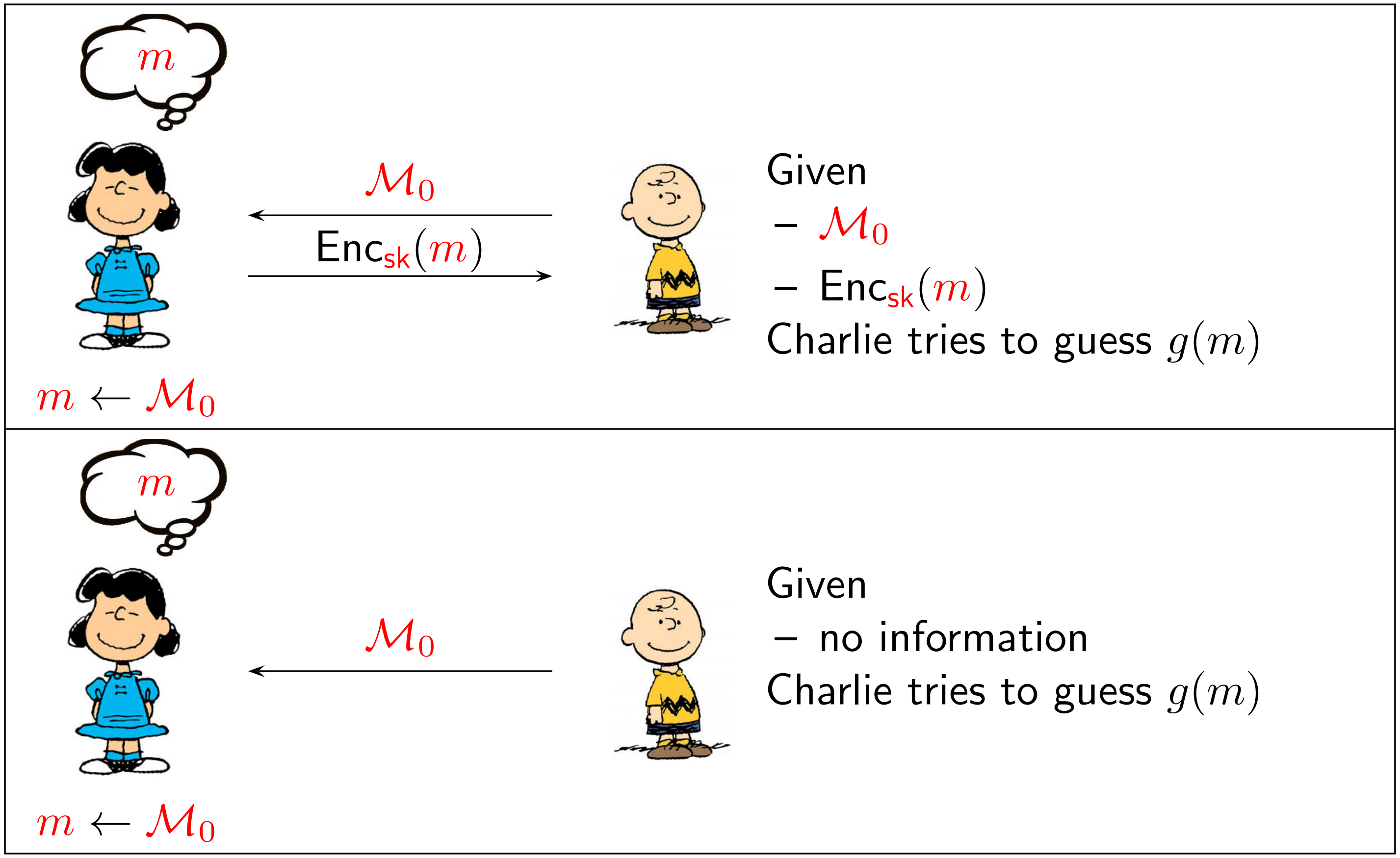


A cryptosystem $\mathcal{C}$ is *(t, ε)-IND-CPA1 secure* if for all t-time adversaries $\mathcal{A}$:

$$\mathsf{Adv}^{\mathsf{ind\text{-}cpa}}_{\mathcal{C}}(\mathcal{A}) = \left|\Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right]\right| \leq \varepsilon \enspace .$$

Semantic Security

Semantic security against adaptive influence



Formal definition

Consider following games:

$$\mathcal{G}_0^{\mathcal{A}}\left[\begin{array}{l}\mathsf{sk} \leftarrow \mathsf{Gen}\\ \mathcal{M}_0 \leftarrow \mathcal{A}^{\mathsf{Enc}_{\mathsf{sk}}(\cdot)}\\ m \leftarrow \mathcal{M}_0\\ c \leftarrow \mathsf{Enc}_{\mathsf{sk}}(m)\\ \textbf{return } [g(m) \stackrel{?}{=} \mathcal{A}(c)]\end{array}\right.$$

$$\mathcal{G}_1^{\mathcal{A}}\left[\begin{array}{l}\mathsf{sk} \leftarrow \mathsf{Gen}\\ \mathcal{M}_0 \leftarrow \mathcal{A}^{\mathsf{Enc}_{\mathsf{sk}}(\cdot)}\\ m \leftarrow \mathcal{M}_0, \boxed{\overline{m} \leftarrow \mathcal{M}_0}\\ \boxed{\overline{c} \leftarrow \mathsf{Enc}_{\mathsf{sk}}(\overline{m})}\\ \textbf{return } [g(m) \stackrel{?}{=} \boxed{\mathcal{A}(\overline{c})}]\end{array}\right.$$

The true guessing advantage is

$$\mathsf{Adv}_g^{\mathsf{sem}}(\mathcal{A}) = \Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right] \quad .$$

IND-CPA $\Rightarrow$ SEM-CPA

Theorem. Assume that g is a t_g-time function and it is always possible to obtain a sample from $\mathcal{M}_0$ in time t_m. Now if the cryptosystem is (t, ε)-IND-CPA1 secure, then for all $(t - t_g - 2t_m)$-time adversaries $\mathcal{A}$:

$$\mathsf{Adv}_g^{\mathsf{sem}}(\mathcal{A}) \leq \varepsilon \ .$$

Note that

- ▷ The function g might be randomised.
- ▷ The function g must be efficiently computable.
- ▷ The distribution $\mathcal{M}_0$ must be efficiently samplable.

The corresponding proof

Let $\mathcal{A}$ be an adversary that can predict the value of g well in SEM-CPA1 game. Now consider a new IND-CPA adversary $\mathcal{B}$:

$$\mathcal{B}^{\mathsf{Enc}_{\mathsf{sk}}(\cdot)}$$

$$\left[\begin{array}{l} \mathcal{M}_0 \leftarrow \mathcal{A}^{\mathsf{Enc}_{\mathsf{sk}}(\cdot)} \\ m_0 \leftarrow \mathcal{M}_0, m_1 \leftarrow \mathcal{M} \\ \textbf{return } (m_0, m_1) \end{array}\right.$$

$$\mathcal{B}(c)$$

$$\left[\begin{array}{l} \mathsf{guess} \leftarrow \mathcal{A}(c) \\ \textbf{return } [\mathsf{guess} \stackrel{?}{=} g(m_0)] \end{array}\right.$$

Running time analysis

The running time of $\mathcal{A}$ is $t_b + t_g + 2t_m$ where t_b is the running time of $\mathcal{B}$.

Further analysis by code rewriting

For clarity, let $\mathcal{Q}_0$ and $\mathcal{Q}_1$ denote the IND-CPA1 security games and $\mathcal{G}_0$ and $\mathcal{G}_1$ IND-SEM security games. Then note

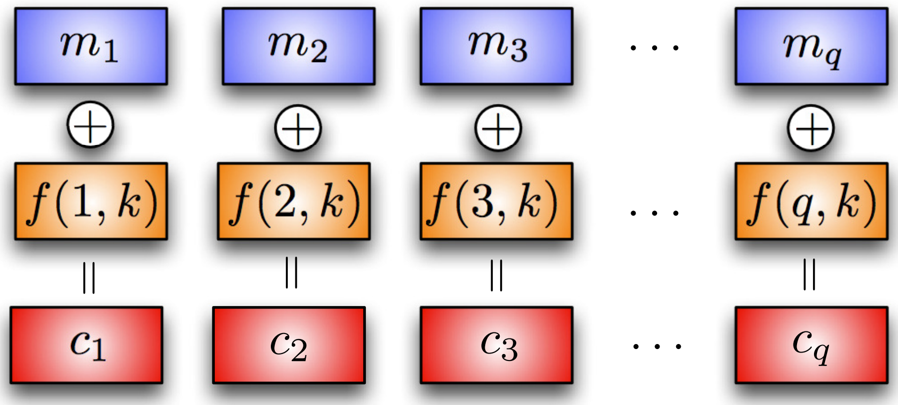
$$\mathcal{Q}_0^{\mathcal{B}} \equiv \mathcal{G}_0^{\mathcal{A}} \qquad \text{and} \qquad \mathcal{Q}_1^{\mathcal{B}} \equiv \mathcal{G}_1^{\mathcal{A}}$$

where

$$\mathcal{Q}_0^{\mathcal{B}} \qquad\qquad \mathcal{Q}_1^{\mathcal{B}}$$

$$\left[\begin{array}{l} \mathsf{sk} \leftarrow \mathsf{Gen} \\ (m_0, m_1) \leftarrow \mathcal{B}^{\mathsf{Enc}_{\mathsf{sk}}(\cdot)} \\ \textbf{return } \mathcal{B}(\mathsf{Enc}_{\mathsf{sk}}(m_0)) \end{array}\right. \qquad \left[\begin{array}{l} \mathsf{sk} \leftarrow \mathsf{Gen} \\ (m_0, m_1) \leftarrow \mathcal{B}^{\mathsf{Enc}_{\mathsf{sk}}(\cdot)} \\ \textbf{return } \mathcal{B}(\mathsf{Enc}_{\mathsf{sk}}(m_1)) \end{array}\right.$$

CTR cipher mode is IND-CPA secure



- **Key generation**: Set ctr $\leftarrow 0$ and choose $k \xleftarrow[u]{} \mathcal{K}$.
- **Encryption**: Given $m \in \mathcal{M}$, increment ctr by 1 and return $m \oplus f(\mathsf{ctr}, k)$
- **Decryption** Given $c \in \mathcal{M}$, increment ctr by 1 and return $c \oplus f(\mathsf{ctr}, k)$.

Theorem. If $f : \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{C}$ is (t, ε)-pseudorandom function, then CTR cipher mode is $(t, 2\varepsilon)$-IND-CPA secure.

Switching Lemma

Motivation

Block ciphers are designed to be pseudorandom permutations. However, it is much more easier to work with pseudorandom functions. Therefore, all classical security proofs have the following structure:

1. Replace pseudorandom permutation family $\mathcal{F}$ with the family $\mathcal{F}_{\mathrm{prm}}$.
2. Use the PRP/PRF switching lemma to substitute $\mathcal{F}_{\mathrm{prm}}$ with $\mathcal{F}_{\mathrm{all}}$.
3. Solve the resulting combinatorial problem to bound the advantage:
   - ▷ All output values $f(x)$ have uniform distribution.
   - ▷ Each output $f(x)$ is independent of other outputs.

More formally, let $\mathcal{G}_0$ the original security game and $\mathcal{G}_1$ and $\mathcal{G}_2$ be the games obtained after replacement steps. Then

$$\mathsf{Adv}^{\mathsf{win}}_{\mathcal{G}_0}(\mathcal{A}) = \Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] \leq \mathsf{cd}^t_\star(\mathcal{G}_0, \mathcal{G}_1) + \mathsf{sd}_\star(\mathcal{G}_1, \mathcal{G}_2) + \Pr\left[\mathcal{G}_2^{\mathcal{A}} = 1\right] \enspace .$$

PRP/PRF switching lemma

Theorem. Let $\mathcal{M}$ be the input and output domain for $\mathcal{F}_{\mathrm{all}}$. Then the permutation family $\mathcal{F}_{\mathrm{prm}}$ is (q, ε)-pseudorandom function family where

$$\varepsilon \leq \frac{q(q-1)}{2\,|\mathcal{M}|} \enspace .$$

Theorem. Let $\mathcal{M}$ be the input and output domain for $\mathcal{F}_{\mathrm{all}}$. Then for any $q \leq \sqrt{|\mathcal{M}|}$ there exists a $\mathrm{O}(q \log q)$ distinguisher $\mathcal{A}$ that achieves

$$\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{F}_{\mathrm{all}},\mathcal{F}_{\mathrm{prm}}}(\mathcal{A}) \geq 0.316 \cdot \frac{q(q-1)}{|\mathcal{M}|} \enspace .$$

Birthday paradox

Obviously $f \notin \mathcal{F}_{\mathrm{prm}}$ if we find a collision $f(x_i) = f(x_j)$ for $i \neq j$.

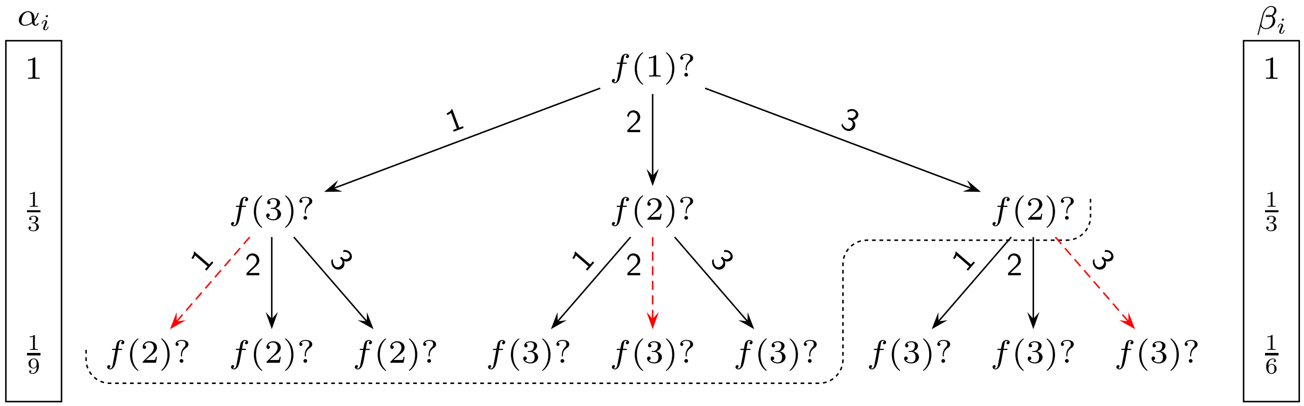
For the proof note that:

- ▷ If $x_1, \ldots, x_q$ are different then the outputs $f(x_1), \ldots, f(x_q)$ have uniform distribution over $\mathcal{M} \times \ldots \times \mathcal{M}$ when $f \xleftarrow[u]{} \mathcal{F}_{\mathrm{all}}$.
- ▷ Hence, the corresponding adversary $\mathcal{A}$ that outputs 0 only in case of collision obtains

$$\begin{aligned}
\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{F}_{\mathrm{all}},\mathcal{F}_{\mathrm{prm}}}(\mathcal{A}) &= \Pr\left[\mathsf{Collision}|\mathcal{F}_{\mathrm{all}}\right] - \Pr\left[\mathsf{Collision}|\mathcal{F}_{\mathrm{prm}}\right] \\
&= \Pr\left[\mathsf{Collision}|\mathcal{F}_{\mathrm{all}}\right] \geq 0.316 \cdot \frac{q(q-1)}{|\mathcal{M}|} \enspace .
\end{aligned}$$

Distinguishing strategy as decision tree

Let $\mathcal{A}$ be a deterministic distinguisher that makes *up to* q oracle calls.



Then $\Pr[\text{Vertex } u|\mathcal{F}_{\mathrm{prm}}]$ and $\Pr[\text{Vertex } u|\mathcal{F}_{\mathrm{all}} \wedge \neg\text{Collision}]$ might differ. However, if $\mathcal{A}$ makes *exactly* q queries then all vertices on decision border are sampled with uniform probability and thus

$$\Pr[\mathcal{A} = 1|\mathcal{F}_{\mathrm{prm}}] = \Pr[\mathcal{A} = 1|\mathcal{F}_{\mathrm{all}} \wedge \neg\text{Collision}] \quad .$$

The corresponding proof

Obviously, the best distinguisher $\mathcal{A}$ is deterministic and makes exactly q oracle calls. Consequently,

$$\begin{aligned}
\Pr\left[\mathcal{A}=1|\mathcal{F}_{\mathrm{all}}\right] &= \Pr\left[\mathsf{Collision}|\mathcal{F}_{\mathrm{all}}\right]\cdot\Pr\left[\mathcal{A}=1|\mathcal{F}_{\mathrm{all}}\wedge\mathsf{Collision}\right]\\
&\qquad+\Pr\left[\neg\mathsf{Collision}|\mathcal{F}_{\mathrm{all}}\right]\cdot\Pr\left[\mathcal{A}=1|\mathcal{F}_{\mathrm{all}}\wedge\neg\mathsf{Collision}\right]\\
&\leq \Pr\left[\mathsf{Collision}|\mathcal{F}_{\mathrm{all}}\right]+\Pr\left[\mathcal{A}=1|\mathcal{F}_{\mathrm{prm}}\right]
\end{aligned}$$

and thus also

$$\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{F}_{\mathrm{all}},\mathcal{F}_{\mathrm{prm}}}(\mathcal{A})\leq\Pr\left[\mathsf{Collision}|\mathcal{F}_{\mathrm{all}}\right]\quad .$$

Now observe

$$\Pr\left[\bigvee_{i\neq j} f(x_i)=f(x_j)\right]\leq\sum_{i\neq j}\Pr\left[f(x_i)=f(x_j)\right]=\frac{q(q-1)}{2}\cdot\frac{1}{|\mathcal{M}|}\quad .$$

Historical references

Nonconstructive IND-SEM theorem was first mentioned in 1982

- **Shaft Goldwasser and Silvio Micali**. Probabilistic Encryption & How To Play Mental Poker Keeping Secret All Partial Information.

Hybrid argument was also first mentioned in 1982

- **Andrew Yao**. Theory and Applications of Trapdoor Functions.

Constructive and modern IND-SEM proof in was given in late 90-ties.

- **Mihir Bellare, Anand Desai, E. Jokipii and Phillip Rogaway**. A Concrete Security Treatment of Symmetric Encryption (1997).
- **Mihir Bellare, Anand Desai, David Pointcheval and Phillip Rogaway**. Relations among Notions of Security for Public-Key Encryption Schemes (1998).