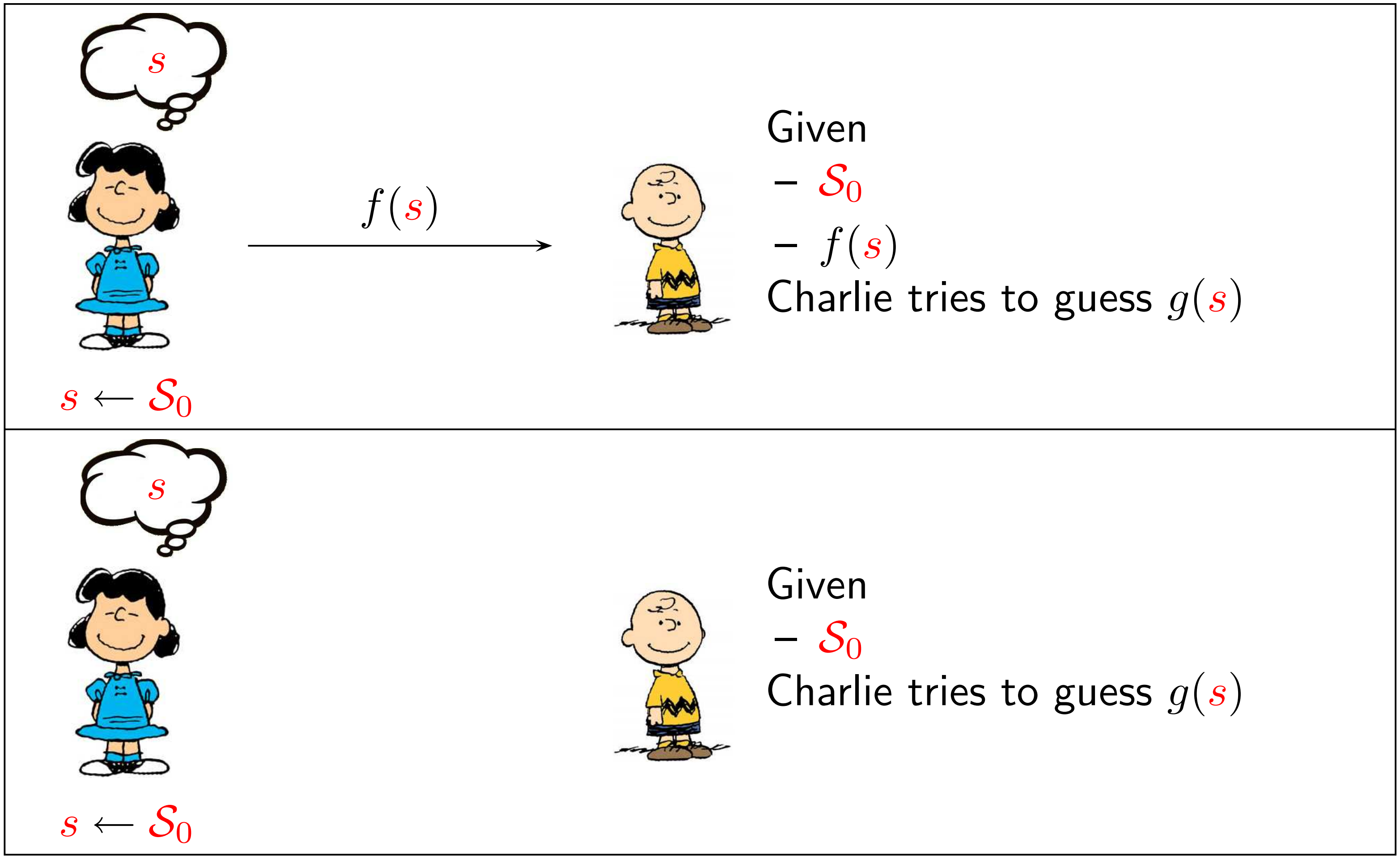
# IND$\Rightarrow$SEM Proof Explained

Sven Laur
swen@math.ut.ee

University of Tartu

# Theoretical Background

# Semantic security



# Formal definition

Consider the following games:

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

$$\left[\begin{array}{l} s \leftarrow \mathcal{S}_0 \\ g' \leftarrow \mathcal{A}(f(s)) \\ \text{return } [g' \stackrel{?}{=} g(s)] \end{array}\right.$$

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

$$\left[\begin{array}{l} s \leftarrow \mathcal{S}_0 \\ g' \leftarrow \operatorname{argmax}_{g'} \Pr\left[g(s) = g'\right] \\ \text{return } [g' \stackrel{?}{=} g(s)] \end{array}\right.$$

Then we can define a true guessing advantage

$$\begin{aligned} \mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) &= \Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right] \\ &= \Pr\left[s \leftarrow \mathcal{S}_0 : \mathcal{A}(f(s)) = g(s)\right] - \max_{g'} \Pr\left[g(s) = g'\right] \enspace . \end{aligned}$$

# IND $\Longrightarrow$ SEM

**Theorem.** If for all $s_i, s_j \in \mathrm{supp}(\mathcal{S}_0)$ distributions $f(s_i)$ and $f(s_j)$ are $(t, \varepsilon)$-indistinguishable, then for all $t$-time adversaries $\mathcal{A}$:

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

Note that

▷ function $g$ might be randomised,

▷ function $g : \mathcal{S}_0 \to \{0,1\}^*$ may extremely difficult to compute,

▷ it might be even infeasible to get samples from the distribution $\mathcal{S}_0$.

# Proof in Small Steps

# Mixture of distributions

Consider a following sampling algorithm

$$\begin{array}{l}\text{GetSample}()\\ \left[\begin{array}{l} i \leftarrow \mathcal{D} \\ s \leftarrow \mathcal{S}_i \\ \text{return } s \end{array}\right.\end{array}$$

where $\mathcal{D}$ is a distribition over the set $\{0, 1, \dots, t\}$ and $\mathcal{S}_0, \dots, \mathcal{S}_t$ are just some distributions. Then

$$\Pr\left[\text{GetSample}() = s_0\right] = \sum_{i_0=0}^{t} \Pr\left[i \leftarrow \mathcal{D} : i = i_0\right] \cdot \Pr\left[s \leftarrow \mathcal{S}_{i_0} : s = s_0\right]$$

# Classical sampling idiom (1/2)

We can reverse the process. Assume that $s$ is sampled from the distribution $\mathcal{S}$ and let $g : \mathcal{S} \to \{0, 1, \ldots, t\}$ be a deterministic function. Then

$$\Pr\left[s \leftarrow \mathcal{S} : s = s_0\right] = \sum_{i_0=1}^{t} \Pr\left[s \leftarrow \mathcal{S} : g(s) = i_0\right] \cdot \Pr\left[s_0 | g(s) = i_0\right]$$

where by definition

$$\Pr\left[s_0 | g(s) = i_0\right] = \frac{\Pr\left[s \leftarrow \mathcal{S} : s = s_0 \land g(s) = i_0\right]}{\Pr\left[s \leftarrow \mathcal{S} : g(s) = i_0\right]}$$

# Classical sampling idiom (2/2)

Let now $\mathcal{D}$ be the distribution over $\{0, 1, \ldots, t\}$ such that

$$\Pr\left[i \leftarrow \mathcal{D} : i = i_0\right] = \Pr\left[s \leftarrow \mathcal{S} : g(s) = i\right]$$

and let $\mathcal{S}_{i_0}$ be defined so that

$$\Pr\left[s \leftarrow \mathcal{S}_i : s = s_0\right] = \Pr\left[s_0 | g(s) = i_0\right] \quad .$$

Then the the output od the sampling procedure GetSample() coincides with the distribution $\mathcal{S}$.

# Slightly modified security game

Let $\mathcal{D}$ and $\mathcal{S}_0, \ldots, \mathcal{S}_t$ be the distributions defined in the previous slide. Then we can rewrite the game $\mathcal{G}_0$ without changing its meaning:

$$\mathcal{G}_0^{\mathcal{A}}\left[\begin{array}{l} i \leftarrow \mathcal{D} \\ s \leftarrow \mathcal{S}_i \\ g' \leftarrow \mathcal{A}(f(s)) \\ \text{return } [g' \stackrel{?}{=} i] \end{array}\right.$$

In other words $\mathcal{A}$ must distinguish between following hypotheses

$$\mathcal{H}_0 = [i \stackrel{?}{=} 0], \mathcal{H}_1 = [i \stackrel{?}{=} 1], \ldots, \mathcal{H}_t = [i \stackrel{?}{=} t] \enspace .$$

It is a guessing game between many hypotheses.

# Computational distance between hypotheses

Let $\mathcal{A}$ be a $t$-time algorithm that must distinguish hypotheses $\mathcal{H}_i$ and $\mathcal{H}_j$. Then the corresponding security games are following

$$\overline{\mathcal{G}}_i^{\mathcal{A}} \quad \left[\begin{array}{l} s \leftarrow \mathcal{S}_i \\ \text{return } \mathcal{A}(f(s)) \end{array}\right. \qquad \text{and} \qquad \overline{\mathcal{G}}_j^{\mathcal{A}} \quad \left[\begin{array}{l} s \leftarrow \mathcal{S}_j \\ \text{return } \mathcal{A}(f(s)) \end{array}\right.$$

In other words

$$\Pr\left[\overline{\mathcal{G}}_i^{\mathcal{A}} = 0\right] = \sum_{s_0 \in \mathrm{supp}(\mathcal{S}_i)} \Pr\left[s \leftarrow \mathcal{S}_i : s = s_0\right] \cdot \Pr\left[\mathcal{A}(f(s_0)) = 0\right]$$

# Double summation trick

For obvious reasons

$$\sum_{s_0 \in \mathrm{supp}(\mathcal{S}_i)} \Pr\left[s \leftarrow \mathcal{S}_i : s = s_0\right] = 1 = \sum_{s_1 \in \mathrm{supp}(\mathcal{S}_j)} \Pr\left[s \leftarrow \mathcal{S}_j : s = s_1\right]$$

Consequently

$$\begin{aligned}
&\left|\Pr\left[\overline{\mathcal{G}}_i^{\mathcal{A}} = 0\right] - \Pr\left[\overline{\mathcal{G}}_j^{\mathcal{A}} = 0\right]\right| \\
&\le \sum_{\substack{s_0 \in \mathrm{supp}(\mathcal{S}_i) \\ s_1 \in \mathrm{supp}(\mathcal{S}_j)}} \Pr\left[s \leftarrow \mathcal{S}_i : s = s_0\right] \cdot \Pr\left[s \leftarrow \mathcal{S}_j : s = s_1\right] \underbrace{\left|\Pr\left[\mathcal{A}(f(s_0)) = 0\right] - \Pr\left[\mathcal{A}(f(s_1)) = 0\right]\right|}_{\le \varepsilon} \\
&\le \varepsilon
\end{aligned}$$

and thus $\mathsf{cd}_x^t(\mathcal{H}_i, \mathcal{H}_j) \le \varepsilon$.

# Summary

Since modified $\mathcal{G}_0$ is nothing more than guessing game between many hypotheses $\mathcal{H}_0, \ldots, \mathcal{H}_t$ that are $(t, \varepsilon)$-indistinguishable, we have proven the claim for deterministic functions $g$.

# Average-case $\leq$ worst-case(1/2)

For the final proof step, assume $\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) > \varepsilon$ for some randomised function

$$g : \mathcal{S}_0 \times \Omega \rightarrow \{0, \ldots, t\} \quad .$$

Now by definition

$$\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) = \Pr\left[s \leftarrow \mathcal{S}_0, \omega \leftarrow \Omega : \mathcal{A}(f(s)) = g(s,\omega)\right] - \max_{g'} \Pr\left[g(s) = g'\right] \quad .$$

Now

$$\begin{aligned}
&\Pr\left[s \leftarrow \mathcal{S}_0, \omega \leftarrow \Omega : \mathcal{A}(f(s)) = g(s,\omega)\right] \\
&= \sum_{\omega_0 \in \Omega} \Pr\left[\omega \leftarrow \Omega : \omega = \omega_0\right] \cdot \Pr\left[s \leftarrow \mathcal{S}_0 : \mathcal{A}(f(s)) = g(s,\omega_0)\right] \\
&\leq \max_{\omega_0 \in \Omega} \Pr\left[s \leftarrow \mathcal{S}_0 : \mathcal{A}(f(s)) = g(s,\omega_0)\right]
\end{aligned}$$

# Average-case $\leq$ worst-case(2/2)

Let $g_0 : \mathcal{S}_0 \to \mathbb{Z}$ be a deterministic function $g_0(s) = g(s, \omega_0)$ where

$$\omega_0 = \operatorname*{argmax}_{\omega_0 \in \Omega} \Pr\left[s \leftarrow \mathcal{S}_0 : \mathcal{A}(f(s)) = g(s, \omega_0)\right] \quad .$$

Then by construction

$$\mathsf{Adv}^{\mathsf{sem}}_{f,g}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{sem}}_{f,g_0}(\mathcal{A})$$

and thus we can indeed observe only deterministic functions.

QED