Chapter 13

Soundness of Static Equivalence

We will show that under some assumptions on the cryptographic primitives static equivalence is sound with respect to a computational model, i.e. whenever two frames φ_1 and φ_2 are statically equivalent, the distributions corresponding to the implementations of these frames are computationally indistinguishable.

13.1 Security properties of symmetric encryption schemes

We recast first the security definition of IND-CPA (defined in the section 9.1 for public-key encryption), in the case of symmetric key encryption.

We wrile $\mathcal{A}^{\mathcal{O}}$ a Probabilistic Polynomial Time Turing machine, equipped with an oracle $\mathcal{O}$. Let us recall that such machines include in particular a random tape, which is read-only and whose content is drawn uniformly at random when the machine starts. The polynomial time computation should only depend on the input of the machine, not on the actual values on the random tape. We sometimes write $\mathcal{A}(x \mid R)$ for the result of the (deterministic) computation of $\mathcal{A}$ on x with a random tape R.

The machine has also a special tape for oracle calls (and replies). It may write on this tape and, from a special state corresponding to the oracle call, there is a transition of the machine from a configuration γ to a configuration in which only the control state and the content of the oracle tape have changed; if the oracle tape contains m before the call to the oracle, it contains $\mathcal{O}(m)$ after the transition. In case the oracle itself is randomized, it is assumed to be equiped with a random (infinite) string R, which is drawn at its first call. Each time the oracle needs a random input, it takes the appropriate prefix of R and removes this prefix from R. When needed, we write $\mathcal{O}(m \mid R)$ to explicitly state what is the random input of the oracle.

The following definition captures the minimal expectations for a symmetric encryption scheme: key generation/ encryption/decryption can be performed in polynomial time and the decryption with a correct key of the encryption of a plaintext gives back the plaintext.

Definition 13.1 *A* symmetric encryption scheme *consists of three deterministic polynomial time functions* $\mathcal{G}, \mathcal{E}, \mathcal{D}$.

- $\mathcal{G}$ *is the* key generation algorithm. *We assume here that the length of* $\mathcal{G}(x)$ *only depends on the length of* x.
- $\mathcal{E}$ *is an encryption algorithm, that, given* x, k, r *(a plaintext, a key and a random seed) returns* $\mathcal{E}(x, k, r)$. *We assume that the length of* $\mathcal{E}(x, k, r)$ *only depends on the length of* x *for a fixed length key* k. $\mathcal{E}(x, \mathcal{G}(y), r)$ *is also assume to depend only on a prefix of length* $|y|$ *of* r: $\mathcal{E}(x, \mathcal{G}(y), r_1) = \mathcal{E}(x, \mathcal{G}(y), r_2)$ *if* r_1 *and* r_2 *have the same prefix of length* $|y|$.

- $\mathcal{D}$ *is the decryption algorithm. It is assumed to satisfy the equation*

$$\mathcal{D}(\mathcal{E}(x, \mathcal{G}(y), r), \mathcal{G}(y)) = x$$

for all x, y, r.

Note that in case of key mismatches, or a key that is not in the range of $\mathcal{G}$, the result of $\mathcal{D}$ is not specified.

Sometimes it is more convenient to hide the key generation algorithm and to use a key distribution. Given $\eta \in \mathbb{N}$, we write $\mathcal{K}(\eta)$ the distribution defined by the image of the uniform distribution on $\{0,1\}^\eta$ by $\mathcal{G}$: for every a, $\mathbb{P}[k \leftarrow \mathcal{K}(\eta) : k = a] = \mathbb{P}[r \leftarrow U(\{0,1\}^\eta) : \mathcal{G}(r) = a]$.

Now, as in the section 9.1, we are going to use encryption oracles. Given a symmetric encryption scheme $\mathcal{G}, \mathcal{E}, \mathcal{D}$ and a key k in $\mathcal{G}(\{0,1\}^\eta)$, we define the two randomized oracles $\mathcal{O}_k^1(\_ \mid R)$ and $\mathcal{O}_k^2(\_ \mid R)$ as follows:

- on an input $x\#y$ where $x, y \in \{0,1\}^*$ and $|x| = |y|$, $\mathcal{O}_k^1(x\#y \mid R)$ returns $\mathcal{E}(x, k, r)$ and $\mathcal{O}_k^2(x\#y \mid R)$ returns $\mathcal{E}(y, k, r)$, for a bitstring r that is taken from R.
- If the input does not have the above format, $\mathcal{O}_k^1$ (resp. $\mathcal{O}_k^2$) returns 0.

Note that two successive calls to $\mathcal{O}_k^i$ with the same input may return different values, as the value r is drawn at each call.

Finally, the security cannot be ensured for fixed length keys (there is always then an attacker); it is rather an asymptotic property. We therefore use the following definition of negligible functions:

Definition 13.2 *A function* $f : \mathbb{N} \to \mathbb{Q}$ *is* negligible *if, for any positive polynomial* P *in one variable, there is an* $N \in \mathbb{N}$ *such that, for every* $\eta > N$, $f(\eta) < \frac{1}{P(\eta)}$.

We are now ready to define IND-CPA:

Definition 13.3 *Let* $\mathcal{S} = (\mathcal{G}, \mathcal{E}, \mathcal{D})$ *be a symmetric encryption scheme.*
Given any oracle PPT machine $\mathcal{A}$ *and any security parameter* $\eta \in \mathbb{N}$ *we define*

$$\begin{aligned}\mathrm{Adv}(\mathcal{A}, \eta) = \\ &|\mathbb{P}[k \leftarrow \mathcal{K}(\eta), R_1, R_2 \leftarrow U : \mathcal{A}^{\mathcal{O}_k^1(\_|R_2)}(0^\eta \mid R_1) = 1] \\ &\quad -\mathbb{P}[k \leftarrow \mathcal{K}(\eta), R_1, R_2 \leftarrow U : \mathcal{A}^{\mathcal{O}_k^2(\_|R_2)}(0^\eta \mid R_1) = 1]|\end{aligned}$$

$\mathcal{S}$ *is* IND-CPA *if, for any PPT machine* $\mathcal{A}$, $\mathrm{Adv}(\mathcal{A}, \eta)$ *is a negligible function of* η.

This property states intuitively that an attacker cannot distinguish between the encryption of two plaintexts of his choice.

Exercice 48
Show that the following is a symmetric encryption scheme and is not IND-CPA:

- $\mathcal{G}$ is the identity
- $\mathcal{E}(x, k, r) = x$
- $\mathcal{D}(y, k) = y$

Exercice 49
Show that any encryption scheme, in which $\mathcal{E}(x, k, r)$ is independent of r is not IND-CPA.

Exercice 50
Show that there is no encryption scheme that satisfies

$$\forall P, \exists N, \forall \mathcal{A}, \forall \eta > N. \quad \mathrm{Adv}(\mathcal{A}, \eta) < \frac{1}{P(\eta)}$$

Where $\mathcal{A}$ ranges over PPTs and P over positive polynomials in one variable.

Exercice 51
Show that there is no symmetric encryption scheme that satisfies

$$\forall \mathcal{A}, \exists N, \forall \eta > N.\ \mathrm{Adv}(\mathcal{A}, \eta) < \frac{1}{2^\eta}$$

where $\mathcal{A}$ ranges over PPTs.

Exercice 52
Given a symmetric encryption scheme, we let

$$\mathrm{Adv}'(\mathcal{A}, \eta) = |2 \times \mathbb{P}[b \leftarrow U(\{1,2\}), k \leftarrow \mathcal{K}(\eta), R_1, R_2 \leftarrow U :\ \mathcal{A}^{\mathcal{O}_k^b(\_|R_1)}(0^\eta \mid R_2) = 1] - 1|$$

Show that IND-CPA is equivalent to:

For every PPT $\mathcal{A}$, $\mathrm{Adv}'(\mathcal{A}, \eta)$ is a negligible function of η.

Exercice 53
Given a symmetric encryption scheme, we define

$$\mathrm{Adv}''(\mathcal{A}, \eta) = \mathbf{Average}[k \leftarrow \mathcal{K}(\eta) :\ \begin{array}{l} |\mathbb{P}[R_1, R_2 \leftarrow U : \mathcal{A}^{\mathcal{O}_k^1(\_|R_1)}(0^\eta \mid R_2) = 1] \\ \quad - \quad \mathbb{P}[R_1, R_2 \leftarrow U : \mathcal{A}^{\mathcal{O}_k^2(\_|R_1)}(0^\eta \mid R_2) = 1]| \end{array}]$$

Is IND-CPA equivalent to the following property:

For every PPT $\mathcal{A}$, $\mathrm{Adv}''(\mathcal{A}, \eta)$ is a negligible function of η.

13.2 The symbolic model

We consider here a fixed set of function symbols $\mathcal{F}$: symmetric encryption $\{\_\}_\_$, pairing $\langle \_, \_ \rangle$, symmetric decryption $\mathsf{dec}(\_, \_)$, projections $\pi_1(\_), \pi_2(\_)$, as well as a collection of constants $\mathcal{W}$. In addition, $\mathcal{N}$ is a set of names. This set of names can be partitioned into different sets, for instance keys, random seeds and nonces. For simplicity, we are going to consider in what follows only one name sort.

We also consider the equational theory E:

$$\begin{array}{ll} \mathsf{dec}(\{x\}_k^r, k) = x & \text{For every } k, r \in \mathcal{N} \\ \pi_1(\langle x, y \rangle) = x & \pi_2(\langle x, y \rangle) = y \end{array}$$

Orienting the equations from left to right, we get a (recursive) convergent rewriting system: every term u in $T(\mathcal{F}, \mathcal{X})$ has a unique normal form $u\!\downarrow$.

In what follows, we consider only (for simplicity) the set of *valid terms* $\mathcal{M}_0$, that is the least set of terms such that:

- $\mathcal{N} \cup \mathcal{W} \subseteq \mathcal{M}_0$
- if $u, v \in \mathcal{M}_0$, then $\mathsf{pair}uv \in \mathcal{M}_0$
- if $u \in \mathcal{M}_0$, $r, k \in \mathcal{N}$, then $\{u\}_k^r \in \mathcal{M}_0$

For convenience, we re-define the static equivalence (and we will see later that it may match the definition of chapter 6: instead of only checking equalities, we give the attacker the ability to check some other predicates.

Valid messages: the unary predicate symbol M is assumed to check the well-formedness of messages.

Its interpretation in our message structure is set M^I of ground terms u such that $u\downarrow \in \mathcal{M}_0$.

Equality: the binary predicate EQ checks the equality of messages: its interpretation EQ^I is the set of pairs of messages (u, v) such that $u \in M^I$, $v \in M^I$ and $u\downarrow = v\downarrow$.

Note that, for instance $(\mathsf{dec}(k,k), \mathsf{dec}(k,k)) \notin EQ^I$: pairs of ill-formed terms are not considered as equal.

Equal keys: as we will see, we will need a predicate EK checking that two ciphertexts use the same encryption keys (the indistinguishability of two such ciphertexts is not guaranteed by IND-CPA).

EK^I is true on pairs of ciphertexts that are using the same encryption key: $(u, v) \in SK^I$ iff there is a $k \in \mathcal{N}$, there are $r_1, r_2 \in \mathcal{N}$, there are terms $u_1, v_1 \in \mathcal{M}_0$ such that $u\downarrow = \{u_1\}_k^{r_1}$ and $v\downarrow = \{v_1\}_k^{r_2}$.

Equal lengths: in the computational model, if two plaintexts have different lengths, then the corresponding ciphertexts have different lengths. Hence we need to reflect this ability to distinguish messages in the symbolic model. Formally, we use the binary predicate symbol EL, whose interpretation will be formally defined later in this section. Informally, EL^I is the pair of terms (u, v) such that there are terms $u_1, v_1 \in \mathcal{M}_0$, names k_1, k_2, r_1, r_2 such that $u\downarrow = \{u_1\}_{k_1}^{r_1}$, $v\downarrow = \{v_1\}_{k_2}^{r_2}$ and, for every $\eta \in \mathbb{N}$, $l(u_1, \eta) = l(v_1, \eta)$.

Let us recall that a *frame* is an expression $\nu\overline{n}.\{x_1 \mapsto s_1, \ldots, x_m \mapsto s_m\}$ where $s_1, \ldots, s_m$ are ground terms, $x_1, \ldots, x_m$ are distinct variables and $\overline{n}$ is a sequence of distinct names.

The *free names* $fn(\phi)$ of a frame $\phi = \nu\overline{n}.\{x_1 \mapsto s_1, \ldots, x_m \mapsto s_m\}$. are the names appearing in $s_1, \ldots, s_m$, that are not in $\overline{n}$. If $\phi = \nu\overline{n}.\{x_1 \mapsto s_1, \ldots, x_m \mapsto s_m\}$, we write σ_ϕ the substitution $\{x_1 \mapsto s_1, \ldots, x_m \mapsto s_m\}$.

A frame is defined up to the renaming of the names in $\overline{n}$: $\phi = \nu n_1, \ldots, n_k.\sigma_\phi$ is considered to be the same frame as $\phi' = \nu n'_1, \ldots, n'_k.\sigma_{\phi'}$ if $fn(\phi) = fn(\phi')$ and $\sigma_{\phi'}$ is obtained by replacing each n_i with n'_i in σ_ϕ.

Definition 13.4 *Given a set of predicate symbols $\mathcal{P}$, two frames $\phi_1 = \nu\overline{n_1}.\{x_1 \mapsto s_1, \ldots, x_k \mapsto s_k\}$ and $\phi_2 = \nu\overline{n_2}.\{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\}$, such that $fn(\phi_1) \cap \overline{n_2} = fn(\phi_2) \cap \overline{n_1} = \emptyset$, are* statically equivalent*, which we write $\phi_1 \sim \phi_2$, if $k = m$ and*

$$\forall P \in \mathcal{P}, \quad \forall u_1, \ldots, u_i \in T(\mathcal{F} \cup (\mathcal{N} \setminus (\overline{n_1} \cup \overline{n_2})), \{x_1, \ldots, x_k\}),$$
$$(u_1\sigma_{\phi_1}, \ldots, u_i\sigma_{\phi_1}) \in P^I \quad \Leftrightarrow \quad (u_1\sigma_{\phi_2}, \ldots, u_i\sigma_{\phi_2}) \in P^I$$

Exercice 54
Let $\mathcal{F}$ be the set of function symbols that has been defined in the beginning of this section and $\mathcal{P}$ be $\{M, EQ, EK\}$. Show that the above definition coincides with the definition of chapter 6, for a well chosen (recursive) equational theory $\mathcal{E}$.

13.3 Indistinguishability of ensembles

Two sequences of distributions (called *ensembles*) parametrized by $\eta \in \mathbb{N}$ are indistinguishable, if any PPT adversary, when faced to the two experiments, cannot guess with a significant advantage with which of the two experiments it is faced:

Definition 13.5 *Let $D = \{D_\eta\}_{\eta\in\mathbb{N}}$ and $D' = \{D'_\eta\}_{\eta\in\mathbb{N}}$ be two ensembles. D and D' are* computationally indistinguishable*, which is written $D \approx D'$ if, for any PPT $\mathcal{A}$, the advantage:*

$$\epsilon(\mathcal{A},\eta) = |\mathbb{P}[x \leftarrow D_\eta, r \leftarrow U : \mathcal{A}(x, 0^\eta \mid r) = 1] - \mathbb{P}[x \leftarrow D'_\eta, r \leftarrow U : \mathcal{A}(x, 0^\eta \mid r) = 1]|$$

is a negligible function of η.

Exercice 55
Show that the Dirac ensemble defined by $\mathbb{P}[x \leftarrow \delta_\eta : x = 0^\eta] = 1$ and the uniform ensemble $\mathbb{P}[x \leftarrow U(\{0,1\}^\eta) : x = a] = \frac{1}{2^\eta}$ for every $a \in \{0,1\}^\eta$ are distinguishable.

Exercice 56
Fix $k \in \mathbb{N}$. Show that the two following ensembles are indistinguishable: the uniform distribution on $\{0,1\}^\eta$ and the distribution U^k_η defined by:

$$\mathbb{P}[x \leftarrow U^k_\eta : x = a] = \begin{cases} 0 & \text{if } a = b0^{n-k} \text{ for some } b \\ \frac{1}{2^n - 2^k} & \text{Otherwise} \end{cases}$$

13.4 The computational interpretation of terms

We let $\mathcal{G}, \mathcal{E}, \mathcal{D}$ be a symmetric encryption scheme and assume that p is a polynomial time pairing function on bitstrings: p is an injection from $\{0,1\}^* \times \{0,1\}^*$ intp $\{0,1\}^*$, whose two inverses p_1^{-1} and p_2^{-1} are also polynomially computable and such that, forall $x, y \in \{0,1\}^*$, $p_1^{-1}(p(x,y)) = x$ and $p_2^{-1}(p(x,y)) = y$. We also assume that, if $|x_1| = |x_2|$ and $|y_1| = |y_2|$, then $|p(x_1,x_2)| = |p(y_1,y_2)|$.

We define here, for each security parameter $\eta \in \mathbb{N}$ the interpretation of terms as bitstrings. First, each $w \in \mathcal{W}$ is interpreted as $[\![w]\!] \in \{0,1\}^*$. Typically, the constants w denote some specific bitstrings and we could have $[\![0101]\!] = 0101$.

Next, given η, we let τ be a mapping from $\mathcal{N}$ to $\{0,1\}^\eta$. Then $[\![\cdot]\!]^\tau_\eta$ is the homomorphism from $T(\mathcal{F} \cup \mathcal{N})$ to $\{0,1\}^*$ that extends τ:

- If $w \in \mathcal{W}$, $[\![w]\!]_\eta = ([\![w]\!])^\eta$
- If $n \in \mathcal{N}$, $[\![n]\!]^\tau_\eta = \tau(n)$
- If $k, r \in \mathcal{N}$ and $u \in \mathcal{M}_0$, then $[\![\{u\}^r_k]\!]^\tau_\eta = \mathcal{E}([\![u]\!]^\tau_\eta, [\![k]\!]^\tau_\eta, [\![r]\!]^\tau_\eta)$
- If $u, v \in \mathcal{M}_0$, $[\![\langle u, v\rangle]\!]^\tau_\eta = p([\![u]\!]^\tau_\eta, [\![v]\!]^\tau_\eta)$.
- $[\![\mathsf{dec}(u,v)]\!]^\tau_\eta = \mathcal{D}([\![u]\!]^\tau_\eta, [\![v]\!]^\tau_\eta)$
- $[\![\pi_1(u)]\!]^\tau_\eta = p_1^{-1}([\![u]\!]^\tau_\eta)$
- $[\![\pi_2(u)]\!]^\tau_\eta = p_2^{-1}([\![u]\!]^\tau_\eta)$

If the names occurring in a ground term u are partitioned into $\mathcal{N}_1$ and $\mathcal{N}_2$ and τ_1 is a mapping from $\mathcal{N}_1$ to $\{0,1\}^\eta$ and τ_2 is a mapping from $\mathcal{N}_2$ to $\{0,1\}^\eta$, then $[\![u]\!]^{\tau_1}_\eta$ defines a distribution: $\mathbb{P}[x \leftarrow [\![u]\!]^{\tau_1}_\eta : x = a] = \mathbb{P}[\tau_2 : [\![u]\!]^{\tau_1 \cup \tau_2}_\eta = a]$. As a particular case, $[\![u]\!]_\eta$ is an ensemble, in which all names in u are sampled in $\{0,1\}^\eta$ (according to a distribution that is not precised here, and which may be assumed to be uniform, for simplicity).

Similarly, if $u_1, \ldots, u_k$ is a sequences of terms and τ is a partial interpretation of the names occurring in $u_1, \ldots, u_k$, $[\![u_1, \ldots, u_k]\!]^\tau_\eta$ is an ensemble: for each η, it defines a distribution on k-uples of bitstrings.

Exercice 57
Assume that the names occurring in $s_1, \ldots, s_n$ are disjoints from the names occurring in $t_1, \ldots, t_m$ then show that $[\![s_1, \ldots, s_n, t_1, \ldots, t_m]\!]_\eta = [\![s_1, \ldots, s_n]\!]_\eta \times [\![t_1, \ldots, t_m]\!]_\eta$.

Show, (using a uniform distribution of name interpretations) that it is not always true when the assumption on the disjointness of the set of names is dropped.

We may now precise the interpretation of EL. We first observe that, according to the assumptions on $\mathcal{E}, p$ and the name samples, $|[\![u]\!]^\tau_\eta|$ is independent of τ (that interprets all names occurring in u) and only depends on η. We let then $l(u, \eta) = |[\![u]\!]^\tau_\eta|$, which completes the definition of EL.

13.5 Preliminary indistinguishability results relying on the property of the encryption scheme

Lemma 13.1 *Fix an interpretation τ of the names occurring in a term u. Assume that the encryption scheme is IND-CPA. Let $u \in \mathcal{M}_0$ be such that k, r do not occur in u and are not in the domain of τ. Then:*

$$[\![\{u\}^r_k]\!]^\tau_\eta \approx [\![\{0^{l(u,\eta)}\}^r_k]\!]^\tau_\eta$$

Proof : Let $\mathcal{A}$ be a PPT machine and

$$\epsilon(\mathcal{A}, \eta) = |\mathbb{P}[x \leftarrow [\![\{u\}^r_k]\!]^\tau_\eta, R \leftarrow U : \mathcal{A}(x, 0^\eta \mid R) = 1] - \mathbb{P}[x \leftarrow [\![\{0^{l(u,\eta)}\}^r_k]\!]^\tau_\eta, R \leftarrow U : \mathcal{A}(x, 0^\eta \mid R) = 1]|$$

Consider now the oracle PPT machine $\mathcal{B}$ (which may depend on τ) such that:

1. Computes $[\![u]\!]^\tau_\eta$ and stores this in y
2. Submits the pair $(y, 0^{l(u,\eta)})$ to the oracle
3. Simulates $\mathcal{A}$ on the reply x of the oracle.

$\mathcal{B}$ runs in polynomial time,as each computation step runs in polynomial time. Furthermore, $|y| = l(u, \eta)$. Hence the machine $\mathcal{B}$ is an attacker on $IND - CPA$, whose advantage is exactly $\mathrm{Adv}(\mathcal{B}, \eta) = \epsilon(\mathcal{A}, \eta)$. Therefore $\epsilon(\mathcal{A}, \eta)$ is negligible.

This is easily generalized to sequences of ciphertexts:

Lemma 13.2 *Let $u_1, \ldots, u_m \in \mathcal{M}_0$. Fix an interpretation τ of the names occurring in $u_1, \ldots, u_m$. Assume that the encryption scheme is IND-CPA. If the names $k, r_1, \ldots, r_m$ are distinct and do not occur in $u_1, \ldots, u_m$, then*

$$[\![\{u_1\}^{r_1}_k, \ldots, \{u_m\}^{r_m}_k]\!]^\tau_\eta \approx [\![\{0^{l(u_1,\eta)}\}^{r_1}_k, \ldots, \{0^{l(u_m,\eta)}\}^{r_m}_k]\!]^\tau_\eta$$

Exercice 58
Complete the proof of the above lemma, using the same ideas as in the proof of the lemma 13.1.

The condition on the occurrences of k, r in u is necessary for lemma 13.1, as shown by the following

Exercice 59
Assume that there exists at least one IND-CPA symmetric encryption scheme. Construct another IND-CPA encryption scheme such that $[\![\{k\}^r_k]\!]_\eta \not\approx [\![\{0^\eta\}^r_k]\!]_\eta$.

We may, however, relax a little bit the assumptions: