Computational Independence*

Björn Fay
mail@bfay.de

December 20, 2014

Abstract

We will introduce different notions of independence, especially computational independence (or more precise independence by polynomial-size circuits (PSC)), which is the analog to computational indistinguishability. We will give some first implications and will show that an encryption scheme having PSC independent plaintexts and ciphertexts is equivalent to having indistinguishable encryptions.

Keywords: Independence, indistinguishability, computational, encryption

1. Introduction

One of the basic principles in modern cryptography is the notion of computational indistinguishability, but for independence only the stochastic independence is used. We introduce the computational analogon, namely computational independence, which is quite unknown, but not totally new. The only other approach known to the author is given in [Yao82]. Yao uses a construction with "effective conditional entropy" and "effective mutual information" to define effectively independent random variables. With this notion of independence he says that an encryption scheme is computationally secure if the plaintext and ciphertext are computationally independent. This is the computational equivalent of Shannon's perfect secrecy [Sha49].

In this paper we will introduce a general framework to work with different kinds of independence, where the range is from perfect independence to computational independence, matching the well known flavors of indistinguishability. The definitions provided are a bit

*This work is an extended and updated extract of the basics in [Fay08].

simpler than the one by Yao and more generic in the sense that they are quite similar to or based on the definitions of indistinguishability and hence can be used similar as the stochastic independence and random variables with the same distribution. This framework can also help to analyze protocols and algorithms, which was the original reason to define it (in [Fay08]).

We will also show that an encryption scheme having PSC independent (by polynomial-size circuits) plaintexts and ciphertexts is equivalent to having indistinguishable encryptions (non-uniform), see section 4.

The rest of the paper is structured as follows. In section 2 we introduce some notions and basic definitions. We show how to work with these new definitions in section 3 by providing some implications. A first application is given in section 4, where we show the relationship to secure encryptions. Finally in section 5 we give some open questions, which might be motivation for some further research. In appendix A we also give some alternative definitions.

2. Notation and Definitions

In this paper we use sequences of random variables, e. g. $(X_n)_{n\in\mathbb{N}}$ is such a sequence, where X_n is a random variable for all $n \in \mathbb{N}$. Since we only use integer values as index, we often shorten this notation to (X_n). We also restrict the random variables to have a countable range, because in the computational cases this is what we have anyhow and we need it for some arguments. If two random variables X, Y have the same distribution we write $X \sim Y$ and if we have two sequences $(X_n), (Y_n)$ for which is $X_n \sim Y_n$ for all $n \in \mathbb{N}$, we write $(X_n) \sim (Y_n)$. If two random variables X, Y are stochastically independent we write $X \perp\!\!\!\perp Y$ and if we have two sequences $(X_n), (Y_n)$ for which is $X_n \perp\!\!\!\perp Y_n$ for all $n \in \mathbb{N}$, we write $(X_n) \perp\!\!\!\perp (Y_n)$.

Further more we use the standard notion of negligibility: A function $f : \mathbb{N} \to \mathbb{R}$ is negligible if for all positive polynomials p there exists an $N \in \mathbb{N}$ so that for all $n > N$ it is $|f(n)| < 1/p(n)$. If f is explicitly given as $f(n)$, we say that $f(n)$ is negligible in n, e.g. $k/(nm)$ is negligible in n or m, but not in k. So we explicitly give the variable to avoid possible ambiguity. We say that a function $f : \mathbb{N} \to \mathbb{R}$ is polynomially bounded if there is a positive polynomial p so that $|f(n)| < p(n)$ for all $n \in \mathbb{N}$.

Before we begin with some standard definitions of indistinguishability, we also introduce some abbreviations based on the notions used in [Gol03] and [Gol04] to specify the computational model which we are using. If we are in the non-uniform complexity setting we use polynomial-size circuits or probabilistic polynomial-size circuits, which we abbreviate with PSC and PPSC. In the uniform complexity setting, which is normally modeled using Turing machines, we use the abbreviations PT and PPT for polynomial time and probabilistic

polynomial time.

Definition 2.1. *Two sequences of random variables* $(X_n)_{n\in\mathbb{N}}$ *and* $(Y_n)_{n\in\mathbb{N}}$ *are* statistically indistinguishable *(or* statistically close*) if and only if*

$$\sum_{\alpha} |P(X_n = \alpha) - P(Y_n = \alpha)|$$

is negligible in n*. The notation for this is* $(X_n) \sim_s (Y_n)$*.*

Definition 2.2. *Two sequences of random variables* $(X_n)_{n\in\mathbb{N}}$ *and* $(Y_n)_{n\in\mathbb{N}}$ *are* indistinguishable by polynomial-size circuits *(PSC indistinguishable) if and only if for all sequences* $(C_n)_{n\in\mathbb{N}}$ *of probabilistic polynomial-size circuits (PPSC) the difference*

$$|P(C_n(X_n) = 1) - P(C_n(Y_n) = 1)|$$

is negligible in n*. The notation for this is* $(X_n) \sim_p (Y_n)$*.*

For this definition it is irrelevant if we use probabilistic or deterministic polynomial-size circuits see theorem A.1

Definition 2.3. *Two sequences of random variables* $(X_n)_{n\in\mathbb{N}}$ *and* $(Y_n)_{n\in\mathbb{N}}$ *are* computationally indistinguishable *if and only if for all PPT (probabilistic polynomial time) algorithms* D *the difference*

$$|P(D(1^n, X_n) = 1) - P(D(1^n, Y_n) = 1)|$$

is negligible in n*. The notation for this is* $(X_n) \sim_c (Y_n)$*.*

Note that all three of these relations are equivalence relations and that for two sequences of random variables $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ we have

$$(X_n) \sim (Y_n) \quad \Rightarrow \quad (X_n) \sim_s (Y_n) \quad \Rightarrow \quad (X_n) \sim_p (Y_n) \quad \Rightarrow \quad (X_n) \sim_c (Y_n).$$

The inverse implications are false in general.

We now introduce the new notions of independence. In figure 1 you can see the general setup of the definitions.

Definition 2.4. *Two sequences of random variables* $(X_n)_{n\in\mathbb{N}}$ *and* $(Y_n)_{n\in\mathbb{N}}$ *are* statistically almost independent *if and only if there exists a sequence* $(\tilde{X}_n, \tilde{Y}_n)_{n\in\mathbb{N}}$ *of pairs of random variables such that* $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n)$ *and* $(X_n, Y_n) \sim_s (\tilde{X}_n, \tilde{Y}_n)$*. The notation for this is* $(X_n) \perp\!\!\!\perp_s (Y_n)$*.*

$$\left.\begin{array}{c}(X_n)\\ \perp\!\!\!\perp_*\\ (Y_n)\end{array}\right\} \sim_* \left\{\begin{array}{c}(\tilde{X}_n)\\ \perp\!\!\!\perp\\ (\tilde{Y}_n)\end{array}\right.$$

Figure 1: Definition of independence $\perp\!\!\!\perp_*$

Definition 2.5. *Two sequences of random variables $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are* independent for polynomial-size circuits *(PSC independent) if and only if there exists a sequence $(\tilde{X}_n, \tilde{Y}_n)_{n\in\mathbb{N}}$ of pairs of random variables such that $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n)$ and $(X_n, Y_n) \sim_p (\tilde{X}_n, \tilde{Y}_n)$. The notation for this is $(X_n) \perp\!\!\!\perp_p (Y_n)$.*

Definition 2.6. *Two sequences of random variables $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are* computationally independent *if and only if there exists a sequence $(\tilde{X}_n, \tilde{Y}_n)_{n\in\mathbb{N}}$ of pairs of random variables such that $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n)$ and $(X_n, Y_n) \sim_c (\tilde{X}_n, \tilde{Y}_n)$. The notation for this is $(X_n) \perp\!\!\!\perp_c (Y_n)$.*

We will see, that these notions behave like one expects them to do. That is if two sequences are independent and are indistinguishable from two further independent sequences (one by one), then these pairs of sequences are indistinguishable (as pairs). This holds for all four kinds of independence and indistinguishability (cf. figure 2).

Since the definitions of independence rely on the definitions of indistinguishability, the above mentioned implications hold also for the kinds of independence, that is

$$(X_n) \perp\!\!\!\perp (Y_n) \quad \Rightarrow \quad (X_n) \perp\!\!\!\perp_s (Y_n) \quad \Rightarrow \quad (X_n) \perp\!\!\!\perp_p (Y_n) \quad \Rightarrow \quad (X_n) \perp\!\!\!\perp_c (Y_n).$$

All definitions of independence above can be generalized to sets of sequences of random variables in a canonical way, for pairwise independence and mutually independence.

Instead of definition 2.4 we also could have used the formulation of the following theorem, which is more similar to stochastic independence.

Theorem 2.7. *Two sequences of random variables $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are* statistically almost independent *if and only if*

$$\sum_{x_n, y_n} |P(X_n = x_n \wedge Y_n = y_n) - P(X_n = x_n) \cdot P(Y_n = y_n)|$$

is negligible in n.

Proof. "⇒": To make the proof easier to read we will use the abbreviation $\tilde{X}$ for $\tilde{X}_n = x_n$ (and similar for other variables) inside the parentheses of a probability.

If we have $(X_n) \perp\!\!\!\perp_s (Y_n)$, there exist $(\tilde{X}_n)$ and $(\tilde{Y}_n)$ with $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n)$ and $(\tilde{X}_n, \tilde{Y}_n) \sim_s (X_n, Y_n)$. This implies that

$$\begin{aligned}
&\sum_{x_n,y_n} \Big| P(\tilde{X}_n = x_n \wedge \tilde{Y}_n = y_n) - P(X_n = x_n \wedge Y_n = y_n) \Big| \\
&= \sum_{x_n,y_n} \Big| P(\tilde{X}) \cdot P(\tilde{Y}) - P(X \wedge Y) \Big| \\
&= \sum_{x_n,y_n} \Big| P(\tilde{X}) \cdot (P(\tilde{Y}) - P(Y) + P(Y)) - P(X \wedge Y) \Big| \\
&= \sum_{x_n,y_n} \Big| P(\tilde{X}) \cdot (P(\tilde{Y}) - P(Y)) + P(\tilde{X}) \cdot P(Y) - P(X \wedge Y) \Big| \\
&= \sum_{x_n,y_n} \Big| P(\tilde{X}) \cdot (P(\tilde{Y}) - P(Y)) + (P(\tilde{X}) - P(X) + P(X)) \cdot P(Y) - P(X \wedge Y) \Big| \\
&= \sum_{x_n,y_n} \Big| P(\tilde{X}) \cdot (P(\tilde{Y}) - P(Y)) + (P(\tilde{X}) - P(X)) \cdot P(Y) + P(X) \cdot P(Y) - P(X \wedge Y) \Big|
\end{aligned}$$

is negligible in n. The sum

$$\sum_{x_n,y_n} \Big| P(\tilde{X}) \cdot (P(\tilde{Y}) - P(Y)) \Big| = \sum_{x_n} P(\tilde{X}) \cdot \sum_{y_n} \Big| (P(\tilde{Y}) - P(Y)) \Big| = \sum_{y_n} \Big| (P(\tilde{Y}) - P(Y)) \Big|$$

is negligible in n (see $(*)$ below), as well as $\sum_{x_n,y_n} \Big| (P(\tilde{X}) - P(X)) \cdot P(Y) \Big|$. This shows that the remaining sum $\sum_{x_n,y_n} |P(X) \cdot P(Y) - P(X \wedge Y)|$ is negligible in n.

$(*)$ In general it is

$$\begin{aligned}
&\sum_{x_n,y_n} \Big| P(\tilde{X}_n \wedge \tilde{Y}_n) - P(X_n \wedge Y_n) \Big| \\
&\geq \sum_{y_n} \Big| \sum_{x_n} \Big(P(\tilde{X}_n \wedge \tilde{Y}_n) - P(X_n \wedge Y_n) \Big) \Big| \\
&= \sum_{y_n} \Big| P(\tilde{Y}_n) - P(Y_n) \Big|
\end{aligned}$$

"$\Leftarrow$": If we have that $\sum_{x_n,y_n} |P(X_n = x_n \wedge Y_n = y_n) - P(X_n = x_n) \cdot P(Y_n = y_n)|$ is negligible in n, there exist $(\tilde{X}_n)$ and $(\tilde{Y}_n)$ so that $(\tilde{X}_n) \sim (X_n)$, $(\tilde{Y}_n) \sim (Y_n)$, and $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n)$. With that and the same argumentation as above, just in the other direction, we get that

$$\sum_{x_n,y_n} \Big| P(\tilde{X}_n = x_n \wedge \tilde{Y}_n = y_n) - P(X_n = x_n \wedge Y_n = y_n) \Big|$$

is negligible in n, which shows $(\tilde{X}_n, \tilde{Y}_n) \sim_s (X_n, Y_n)$ and hence $(X_n) \perp\!\!\!\perp_s (Y_n)$ because $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n)$. $\square$

$$\left.\begin{array}{c}(X_n)\\ \perp\!\!\!\perp_*\\ (Y_n)\end{array}\right\} \begin{array}{c}\sim_*\\ \sim_*\\ \sim_*\end{array} \left\{\begin{array}{c}(X'_n)\\ \perp\!\!\!\perp_*\\ (Y'_n)\end{array}\right.$$

Figure 2: Implicated indistinguishability $\sim_*$

3. Implications

In this section we will see some implications, which can be used to ease the usage of the different flavors of independence and indistinguishability. In figure 2 you can see the general setup of the implications, which are shown in the following subsections.

3.1. Implications for stochastic independence

For the plain stochastic case the above mentioned behavior is already known and easy to see.

Theorem 3.1. *Let X, Y, X', Y' be random variables. If $X \perp\!\!\!\perp Y$, $X' \perp\!\!\!\perp Y'$ and $X \sim X'$, $Y \sim Y'$, then $(X, Y) \sim (X', Y')$.*

Proof. For all x and y,

$$\begin{aligned} P((X,Y) = (x,y)) &= P(X = x \wedge Y = y) \\ &= P(X = x) \cdot P(Y = y) \\ &= P(X' = x) \cdot P(Y' = y) \\ &= P(X' = x \wedge Y' = y) \\ &= P((X',Y') = (x,y)). \end{aligned}$$

□

3.2. Implications for statistical almost independence

We now want to show similar implications for the other three cases. First we take the statistical case.

Theorem 3.2. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ be sequences of random variables. If $(X_n) \perp\!\!\!\perp_s (Y_n)$, $(X'_n) \perp\!\!\!\perp_s (Y'_n)$ and $(X_n) \sim_s (X'_n)$, $(Y_n) \sim_s (Y'_n)$, then $(X_n, Y_n) \sim_s (X'_n, Y'_n)$.*

Proof. Because of $(X_n) \perp\!\!\!\perp_s (Y_n)$ and $(X'_n) \perp\!\!\!\perp_s (Y'_n)$ there exist $(\tilde{X}_n), (\tilde{Y}_n)$ and $(\tilde{X}'_n), (\tilde{Y}'_n)$ such that $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{Y}_n), (\tilde{X}'_n) \perp\!\!\!\perp (\tilde{Y}'_n)$ and

$$(X_n, Y_n) \sim_s (\tilde{X}_n, \tilde{Y}_n), \qquad (X'_n, Y'_n) \sim_s (\tilde{X}'_n, \tilde{Y}'_n),$$

which implies

$$(\tilde{X}_n) \sim_s (X_n) \sim_s (X'_n) \sim_s (\tilde{X}'_n), \qquad (\tilde{Y}_n) \sim_s (Y_n) \sim_s (Y'_n) \sim_s (\tilde{Y}'_n).$$

For the rest of the proof we introduce some abbreviations to make the formulas better to read. We write $P(X)$ for $P(X_n = \alpha)$, $P(Y)$ for $P(Y_n = \beta)$ and $P(X, Y)$ for $P((X_n, Y_n) = (\alpha, \beta))$. These abbreviations are for all variants of X and Y. Hence we have

$$\begin{aligned}
&\sum_{\alpha,\beta} |P(X,Y) - P(X',Y')| \\
&= \sum_{\alpha,\beta} |P(X,Y) - P(\tilde{X},\tilde{Y}) + P(\tilde{X},\tilde{Y}) - P(\tilde{X}',\tilde{Y}') + P(\tilde{X}',\tilde{Y}') - P(X',Y')| \\
&\leq \underbrace{\sum_{\alpha,\beta} |P(X,Y) - P(\tilde{X},\tilde{Y})|}_{\text{negligible in } n} + \sum_{\alpha,\beta} |P(\tilde{X},\tilde{Y}) - P(\tilde{X}',\tilde{Y}')| + \underbrace{\sum_{\alpha,\beta} |P(\tilde{X}',\tilde{Y}') - P(X',Y')|}_{\text{negligible in } n}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{\alpha,\beta} |P(\tilde{X},\tilde{Y}) - P(\tilde{X}',\tilde{Y}')| \\
&= \sum_{\alpha,\beta} |P(\tilde{X})P(\tilde{Y}) - P(\tilde{X}')P(\tilde{Y}')| \\
&= \sum_{\alpha,\beta} |P(\tilde{X})P(\tilde{Y}) - P(\tilde{X})P(\tilde{Y}') + P(\tilde{X})P(\tilde{Y}') - P(\tilde{X}')P(\tilde{Y}')| \\
&\leq \underbrace{\sum_{\alpha} P(\tilde{X})}_{=1} \cdot \underbrace{\sum_{\beta} |P(\tilde{Y}) - P(\tilde{Y}')|}_{\text{negligible in } n} + \underbrace{\sum_{\beta} P(\tilde{Y}')}_{=1} \cdot \underbrace{\sum_{\alpha} |P(\tilde{X}) - P(\tilde{X}')|}_{\text{negligible in } n}
\end{aligned}$$

which shows $(X_n, Y_n) \sim_s (X'_n, Y'_n)$. □

3.3. Implications for PSC independence

For this case we must first make some observations. Since we now want to study the setup with polynomial-size circuits, we need to restrict the values of the random variables to values that could be generated by such circuits. This is not really a restriction, because in practice all the random variables are either generated by an encryption scheme (or another

real world algorithm in a computer, i.e. a PPT algorithm) or by an adversary, who will be restricted to use only polynomial-size circuits. This general restriction of the random variables is given by the following definition.

Definition 3.3. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is* constructible by polynomial-size circuits (PSCC)*, if and only if there exists a sequence $(C_n)_{n\in\mathbb{N}}$ of PPSC such that for all $n \in \mathbb{N}$, $C_n \sim X_n$.*

Now we can start to examine the implications for this case.

Lemma 3.4. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}$ be PSCC sequences of random variables, such that $X_n \perp\!\!\!\perp Y_n$ and $X'_n \perp\!\!\!\perp Y_n$ for all $n \in \mathbb{N}$. If $(X_n) \sim_p (X'_n)$, then $(X_n, Y_n) \sim_p (X'_n, Y_n)$.*

Proof. Assume that the theorem is false, then there would exist a sequence $(D_n)_{n\in\mathbb{N}}$ of PPSC, such that

$$|P(D_n(X_n, Y_n) = 1) - P(D_n(X'_n, Y_n) = 1)|$$

would not be negligible in n. Let S_n be a sequence of PPSC such that $S_n \sim Y_n$ and let $(D'_n)_{n\in\mathbb{N}}$ be the sequence of PPSC that is constructed by $D'_n(x) = D_n(x, S_n)$. Let R_n be the range of Y_n and S_n. Then

$$\begin{aligned}
&|P(D'_n(X_n) = 1) - P(D'_n(X'_n) = 1)| \\
&= |P(D_n(X_n, S_n) = 1) - P(D_n(X'_n, S_n) = 1)| \\
&= \left| \sum_{y\in R_n} P(D_n(X_n, y) = 1) \cdot P(S_n = y) - \sum_{y\in R_n} P(D_n(X'_n, y) = 1) \cdot P(S_n = y) \right| \\
&= \left| \sum_{y\in R_n} P(D_n(X_n, y) = 1) \cdot P(Y_n = y) - \sum_{y\in R_n} P(D_n(X'_n, y) = 1) \cdot P(Y_n = y) \right| \\
&= |P(D_n(X_n, Y_n) = 1) - P(D_n(X'_n, Y_n) = 1)|
\end{aligned}$$

is negligible in n because of $(X_n) \sim_p (X'_n)$ which yields a contradiction. □

Lemma 3.5. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ be PSCC sequences of random variables, such that $(X_n) \perp\!\!\!\perp (Y_n)$ and $(X'_n) \perp\!\!\!\perp (Y'_n)$. If $(X_n) \sim_p (X'_n)$ and $(Y_n) \sim_p (Y'_n)$, then $(X_n, Y_n) \sim_p (X'_n, Y'_n)$.*

Proof. We take two PSCC sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$ of stochastically independent (pairwise and from the rest) random variables such that $X_n \sim \tilde{X}_n, Y_n \sim \tilde{Y}_n$ for all $n \in \mathbb{N}$. Then by lemma 3.4 we have

$$(X_n, Y_n) \sim_p (\tilde{X}_n, Y_n) \sim_p (\tilde{X}_n, \tilde{Y}_n) \sim_p (X'_n, \tilde{Y}_n) \sim_p (X'_n, Y'_n).$$

□

Lemma 3.6. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ be PSCC sequences of random variables, such that $(X_n) \perp\!\!\!\perp_p (Y_n)$. Then there exist PSCC sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$ of random variables such that $\tilde{X}_n \perp\!\!\!\perp \tilde{Y}_n$ and $(\tilde{X}_n, \tilde{Y}_n) \sim_p (X_n, Y_n)$.*

Proof. Per definition of $(X_n) \perp\!\!\!\perp_p (Y_n)$ there exist sequences $(X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ of random variables such that $X'_n \perp\!\!\!\perp Y'_n$ and $(X'_n, Y'_n) \sim_p (X_n, Y_n)$. Because $(X_n), (Y_n)$ are PSCC there also exist PSCC sequences $(S_n), (T_n)$ such that $S_n \sim X_n \sim_p X'_n$ and $T_n \sim Y_n \sim_p Y'_n$. Their outputs (of S_n and T_n) are stochastically independent and with lemma 3.5 we have $(S_n, T_n) \sim_p (X'_n, Y'_n) \sim_p (X_n, Y_n)$. So $(S_n), (T_n)$ are the claimed $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$. □

Theorem 3.7. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ be PSCC sequences of random variables, such that $(X_n) \perp\!\!\!\perp_p (Y_n)$ and $(X'_n) \perp\!\!\!\perp_p (Y'_n)$. If $(X_n) \sim_p (X'_n)$ and $(Y_n) \sim_p (Y'_n)$, then $(X_n, Y_n) \sim_p (X'_n, Y'_n)$.*

Proof. Per lemma 3.6 there exist PSCC sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}, (\tilde{X}'_n)_{n\in\mathbb{N}}, (\tilde{Y}'_n)_{n\in\mathbb{N}}$ of random variables such that $\tilde{X}_n \perp\!\!\!\perp \tilde{Y}_n$, $\tilde{X}'_n \perp\!\!\!\perp \tilde{Y}'_n$ and $(\tilde{X}_n, \tilde{Y}_n) \sim_p (X_n, Y_n)$, $(\tilde{X}'_n, \tilde{Y}'_n) \sim_p (X'_n, Y'_n)$.

Hence $\tilde{X}_n \sim_p X_n \sim_p X'_n \sim_p \tilde{X}'_n$ and $\tilde{Y}_n \sim_p Y_n \sim_p Y'_n \sim_p \tilde{Y}'_n$. With lemma 3.5 we have $(\tilde{X}_n, \tilde{Y}_n) \sim_p (\tilde{X}'_n, \tilde{Y}'_n)$ and then

$$(X_n, Y_n) \sim_p (\tilde{X}_n, \tilde{Y}_n) \sim_p (\tilde{X}'_n, \tilde{Y}'_n) \sim_p (X'_n, Y'_n).$$

□

3.4. Implications for computational independence

The computational case is similar to the PSC setup, but now we have only PPT algorithms instead of PSCs. This reflects the real world use case where everything (every random variable) is generated by a computer. This general restriction of the random variables is given by the following definition.

Definition 3.8. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is* polynomial-time-constructible (PTC)*, if and only if there exists a PPT algorithm S such that for all $n \in \mathbb{N}$, $S(1^n) \sim X_n$.*

Now we can start to examine the computational case.

Lemma 3.9. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}$ be PTC sequences of random variables, such that $X_n \perp\!\!\!\perp Y_n$ and $X'_n \perp\!\!\!\perp Y_n$ for all $n \in \mathbb{N}$. If $(X_n) \sim_c (X'_n)$, then $(X_n, Y_n) \sim_c (X'_n, Y_n)$.*

Proof. Assume that the theorem is false, then there would exist a PPT algorithm D, such that

$$|P(D(1^n, X_n, Y_n) = 1) - P(D(1^n, X'_n, Y_n) = 1)|$$

would not be negligible in n. Let S be a PPT algorithm such that $S(1^n) \sim Y_n$ and let D' be the algorithm that is constructed by $D'(1^n, x) = D(1^n, x, S(1^n))$. This is also an PPT algorithm. Let R_n be the range of Y_n and $S(1^n)$. Then

$$\begin{aligned}
&|P(D'(1^n, X_n) = 1) - P(D'(1^n, X'_n) = 1)| \\
&= |P(D(1^n, X_n, S(1^n)) = 1) - P(D(1^n, X'_n, S(1^n)) = 1)| \\
&= \left| \sum_{y \in R_n} P(D(1^n, X_n, y) = 1) \cdot P(S(1^n) = y) - \sum_{y \in R_n} P(D(1^n, X'_n, y) = 1) \cdot P(S(1^n) = y) \right| \\
&= \left| \sum_{y \in R_n} P(D(1^n, X_n, y) = 1) \cdot P(Y_n = y) - \sum_{y \in R_n} P(D(1^n, X'_n, y) = 1) \cdot P(Y_n = y) \right| \\
&= |P(D(1^n, X_n, Y_n) = 1) - P(D(1^n, X'_n, Y_n) = 1)|
\end{aligned}$$

is negligible in n, because of $(X_n) \sim_c (X'_n)$, which yields a contradiction. □

Lemma 3.10. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ be PTC sequences of random variables, such that $(X_n) \perp\!\!\!\perp (Y_n)$ and $(X'_n) \perp\!\!\!\perp (Y'_n)$. If $(X_n) \sim_c (X'_n)$ and $(Y_n) \sim_c (Y'_n)$, then $(X_n, Y_n) \sim_c (X'_n, Y'_n)$.*

Proof. We take two PTC sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$ of stochastically independent (pairwise and from the rest) random variables such that $X_n \sim \tilde{X}_n, Y_n \sim \tilde{Y}_n$ for all $n \in \mathbb{N}$. Then by lemma 3.9 we have

$$(X_n, Y_n) \sim_c (\tilde{X}_n, Y_n) \sim_c (\tilde{X}_n, \tilde{Y}_n) \sim_c (X'_n, \tilde{Y}_n) \sim_c (X'_n, Y'_n).$$

□

Lemma 3.11. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ be PTC sequences of random variables, such that $(X_n) \perp\!\!\!\perp_c (Y_n)$. Then there exist PTC sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$ of random variables such that $\tilde{X}_n \perp\!\!\!\perp \tilde{Y}_n$ and $(\tilde{X}_n, \tilde{Y}_n) \sim_c (X_n, Y_n)$.*

Proof. Per definition of $(X_n) \perp\!\!\!\perp_c (Y_n)$ there exist sequences $(X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ of random variables such that $X'_n \perp\!\!\!\perp Y'_n$ and $(X'_n, Y'_n) \sim_c (X_n, Y_n)$. Because $(X_n), (Y_n)$ are PTC there exist PPT algorithms S, T such that $S(1^n) \sim X_n \sim_c X'_n$ and $T(1^n) \sim Y_n \sim_c Y'_n$. Their outputs (of $S(1^n)$ and $T(1^n)$) are stochastically independent and with lemma 3.10 we have $(S(1^n), T(1^n)) \sim_c (\tilde{X}_n, \tilde{Y}_n) \sim_c (X_n, Y_n)$. So $(S(1^n)), (T(1^n))$ are the claimed $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$. □

Theorem 3.12. *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (X'_n)_{n\in\mathbb{N}}, (Y'_n)_{n\in\mathbb{N}}$ be PTC sequences of random variables, such that $(X_n) \perp\!\!\!\perp_c (Y_n)$ and $(X'_n) \perp\!\!\!\perp_c (Y'_n)$. If $(X_n) \sim_c (X'_n)$ and $(Y_n) \sim_c (Y'_n)$, then $(X_n, Y_n) \sim_c (X'_n, Y'_n)$.*

Proof. Per lemma 3.11 there exist PTC sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}, (\tilde{X}'_n)_{n\in\mathbb{N}}, (\tilde{Y}'_n)_{n\in\mathbb{N}}$ of random variables such that $\tilde{X}_n \perp\!\!\!\perp \tilde{Y}_n$, $\tilde{X}'_n \perp\!\!\!\perp \tilde{Y}'_n$ and $(\tilde{X}_n, \tilde{Y}_n) \sim_c (X_n, Y_n)$, $(\tilde{X}'_n, \tilde{Y}'_n) \sim_c (X'_n, Y'_n)$.

Hence $\tilde{X}_n \sim_c X_n \sim_c X'_n \sim_c \tilde{X}'_n$ and $\tilde{Y}_n \sim_c Y_n \sim_c Y'_n \sim_c \tilde{Y}'_n$. With lemma 3.10 we have $(\tilde{X}_n, \tilde{Y}_n) \sim_c (\tilde{X}'_n, \tilde{Y}'_n)$ and then

$$(X_n, Y_n) \sim_c (\tilde{X}_n, \tilde{Y}_n) \sim_c (\tilde{X}'_n, \tilde{Y}'_n) \sim_c (X'_n, Y'_n).$$

□

4. A First Application: Secure Encryptions

Perfect secrecy for an encryption scheme was defined by Shannon in [Sha48] and it says that for perfect secrecy the ciphertext has to be stochastically independent of the plaintext. We want to generalize this to different types of independence.

Note that we only examine private-key encryption schemes here. We use some variations of the definitions provided in [Gol04] with some explanation why they are equivalent.

Definition 4.1. *An encryption scheme is a triple (G, E, D) of PPT algorithms satisfying the following two conditions:*

1. *On input 1^n, algorithm G (called the key-generator) outputs a bit string.*
2. *For every k in the range of $G(1^n)$, and for every $\alpha \in \{0,1\}^*$, algorithm E (encryption) and D (decryption) satisfy*

$$P(D_k(E_k(\alpha)) = \alpha) = 1.$$

Here we have only reduced the definition 5.1.1 in [Gol04] to the private-key case.

Before we start to study the relationship between the different flavors of independence and secure encryption we should note that the length of the plain- and/or ciphertexts is a quite sensitive variable for several reasons:

- Longer plaintexts correspond also to longer ciphertexts, at least in general. So to some extend information about the plaintext length can be deduced from the ciphertext length.
- Perfect secrecy can only exist if the plaintext is not longer than the key. Similar holds for "almost perfect secrecy" in the case where we replace stochastic independence by statistical almost independence.

- For the two computational definitions of secure encryptions (uniform and non-uniform complexity) no such boundary exists, the length just has to be polynomially bounded. Therefore the relationship has a slightly different form there regarding the length.

Now let us start with the first case.

4.1. Stochastic Independence

Just for completeness we show the equivalence of stochastically independent plain- and ciphertexts (secure encryptions in this case) and equality of ciphertext distributions.

Theorem 4.2. *Let (G, E, D) be an encryption scheme. Then for every positive, polynomially bounded function ℓ the following two statements are equivalent:*

1. *For every sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with $X_n \in \{0,1\}^{\ell(n)}$ it is*

$$(X_n) \perp\!\!\!\perp (E_{G(1^n)}(X_n)).$$

2. *For all sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ with $x_n, y_n \in \{0,1\}^{\ell(n)}$ it is*

$$(E_{G(1^n)}(x_n)) \sim (E_{G(1^n)}(y_n)).$$

Proof. So let us start with $1 \Rightarrow 2$. Then we have for every $n \in \mathbb{N}$, $x_n \in \{0,1\}^{\ell(n)}$ and $e \in \{0,1\}^*$ that

$$P(E_{G(1^n)}(x_n) = e) = P(E_{G(1^n)}(X_n) = e \mid X_n = x_n) = P(E_{G(1^n)}(X_n) = e).$$

And the same holds for $P(E_{G(1^n)}(y_n) = e)$ (for all $y_n \in \{0,1\}^{\ell(n)}$), so that we have $(E_{G(1^n)}(x_n)) \sim (E_{G(1^n)}(y_n))$.

Let us now look at $2 \Rightarrow 1$. Let (x_n) be a sequence with $x_n \in \{0,1\}^{\ell(n)}$, then we have for every $n \in \mathbb{N}$ and $e \in \{0,1\}^*$ that

$$\begin{aligned}
P(E_{G(1^n)}(X_n) = e) &= \sum_{x\in\{0,1\}^{\ell(n)}} P(E_{G(1^n)}(x) = e \mid X_n = x) \cdot P(X_n = x) \\
&= \sum_{x\in\{0,1\}^{\ell(n)}} P(E_{G(1^n)}(x) = e) \cdot P(X_n = x) \\
&= P(E_{G(1^n)}(x_n) = e) \cdot \sum_{x\in\{0,1\}^{\ell(n)}} P(X_n = x) \\
&= P(E_{G(1^n)}(x_n) = e) \\
&= P(E_{G(1^n)}(X_n) = e \mid X_n = x_n).
\end{aligned}$$

So we have $(X_n) \perp\!\!\!\perp (E_{G(1^n)}(X_n))$. □

4.2. Statistical Almost Independence

Theorem 4.3. *Let (G, E, D) be an encryption scheme. Then for every positive, polynomially bounded function ℓ the following two statements are equivalent:*

1. *For every sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with $X_n \in \{0,1\}^{\ell(n)}$ it is*

$$(X_n) \perp\!\!\!\perp_s (E_{G(1^n)}(X_n)).$$

2. *For all sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ with $x_n, y_n \in \{0,1\}^{\ell(n)}$ it is*

$$(E_{G(1^n)}(x_n)) \sim_s (E_{G(1^n)}(y_n)).$$

Proof. Before we start, we introduce a notation to simplify the proof. Two sequences of functions $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are almost equal if $\sum_x |f_n(x) - g_n(x)|$ is negligible in n and we write $f_n(x) \approx_n^x g_n(x)$ for explicit definitions of functions. We can use telescoping series and triangle inequality to show that this is an equivalence relation. If the sum of differences is 0, then we write $f_n(x) =_n^x g_n(x)$.

So after the introduction of this notation let us start with $1 \Rightarrow 2$. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be two sequences with $x_n, y_n \in \{0,1\}^{\ell(n)}$ and define a sequence (X_n) of random variables with $X_n = x_n$ and $X_n = y_n$ with probability $\frac{1}{2}$ for all $n \in \mathbb{N}$. Per definition there exist $(\tilde{X}_n), (\tilde{E}_n)$ with $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{E}_n)$ and $(X_n, E_{G(1^n)}(X_n)) \sim_s (\tilde{X}_n, \tilde{E}_n)$. We then have

$$\begin{aligned}
P(E_{G(1^n)}(x_n) = e) &=_n^e P(E_{G(1^n)}(X_n) = e \mid X_n = x_n) \\
&=_n^e \frac{P(E_{G(1^n)}(X_n) = e \wedge X_n = x_n)}{P(X_n = x_n)} \\
&=_n^e 2 \cdot P(E_{G(1^n)}(X_n) = e \wedge X_n = x_n) \\
&\approx_n^e 2 \cdot P(\tilde{E}_n = e \wedge \tilde{X}_n = x_n) \qquad (*) \\
&=_n^e 2 \cdot P(\tilde{E}_n = e) \cdot P(\tilde{X}_n = x_n) \\
&\approx_n^e 2 \cdot P(\tilde{E}_n = e) \cdot P(X_n = x_n) \\
&=_n^e P(\tilde{E}_n = e) \\
&\vdots \\
&\approx_n^e P(E_{G(1^n)}(y_n) = e).
\end{aligned}$$

$(*)$ holds because $(X_n, E_{G(1^n)}(X_n)) \sim_s (\tilde{X}_n, \tilde{E}_n)$ implies this if you take sums over all possible values of Es and Xs. If you do not take all possible values and the sum of differences was negligible before, it is still negligible. Similar holds two lines further down. Hence we have $(E_{G(1^n)}(x_n)) \sim_s (E_{G(1^n)}(y_n))$.

Let us now look at $2 \Rightarrow 1$. So let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with $X_n \in \{0,1\}^{\ell(n)}$ and (z_n) a sequence of values with $P(X_n = z_n) > 0$. Let $(\tilde{X}_n)$ and $(\tilde{E}_n)$ be sequences of random variables with $\tilde{X}_n \sim X_n$, $\tilde{E}_n = E_{G(1^n)}(z_n)$, and $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{E}_n)$. Please note that $E_{G(1^n)}(x)$ and X_n are stochastically independent for all (fixed) x. Then we have

$$\begin{aligned}
P(E_{G(1^n)}(X_n) = e \wedge X_n = x) &=^{x,e}_n P(E_{G(1^n)}(X_n) = e \mid X_n = x) \cdot P(X_n = x) \\
&=^{x,e}_n P(E_{G(1^n)}(x) = e \mid X_n = x) \cdot P(X_n = x) \\
&=^{x,e}_n P(E_{G(1^n)}(x) = e) \cdot P(X_n = x) \\
&\approx^{x,e}_n P(E_{G(1^n)}(z_n) = e) \cdot P(X_n = x) \qquad (*) \\
&=^{x,e}_n P(\tilde{E}_n = e) \cdot P(X_n = x) \\
&=^{x,e}_n P(\tilde{E}_n = e) \cdot P(\tilde{X}_n = x) \\
&=^{x,e}_n P(\tilde{E}_n = e \wedge \tilde{X}_n = x).
\end{aligned}$$

The step $(*)$ might need some further explanations. Note that $P(X_n = x) = 0$ if $x \notin \{0,1\}^{\ell(n)}$ and hence

$$\begin{aligned}
&\sum_{x,e} \left| P(E_{G(1^n)}(x) = e) \cdot P(X_n = x) - P(E_{G(1^n)}(z_n) = e) \cdot P(X_n = x) \right| \\
&= \sum_x \left(P(X_n = x) \cdot \sum_e \left| P(E_{G(1^n)}(x) = e) - P(E_{G(1^n)}(z_n) = e) \right| \right) \\
&\leq \underbrace{\sum_x P(X_n = x)}_{\leq 1} \cdot \underbrace{\sum_e \left| P(E_{G(1^n)}(x_n) = e) - P(E_{G(1^n)}(z_n) = e) \right|}_{\text{negligible}}
\end{aligned}$$

where (x_n) is a sequence of values such that

$$\begin{aligned}
&\sum_e \left| P(E_{G(1^n)}(x_n) = e) - P(E_{G(1^n)}(z_n) = e) \right| \\
&= \max_{x\in\{0,1\}^{\ell(n)}} \sum_e \left| P(E_{G(1^n)}(x) = e) - P(E_{G(1^n)}(z_n) = e) \right|.
\end{aligned}$$

If we summarize this we have shown that $(E_{G(1^n)}(X_n), X_n) \sim_s (\tilde{E}_n, \tilde{X}_n)$ and hence

$$(E_{G(1^n)}(X_n)) \perp\!\!\!\perp_s (X_n).$$

□

4.3. PSC Independence

We want to show that if we use PSC independence for plaintext and ciphertext then this is equivalent to the encryption scheme having indistinguishable encryptions (non-uniform)

and hence is also equivalent to semantic security (non-uniform). See [Gol04] for more details.

Definition 4.4. *An encryption scheme (G, E, D) has indistinguishable encryptions (non-uniform) if for every sequence $(C_n)_{n\in\mathbb{N}}$ of PPSC, for every positive, polynomially bounded function ℓ and positive polynomial p, there exists an $N \in \mathbb{N}$, so that for all $n > N$ and every $x, y \in \{0,1\}^{\ell(n)}$, it is*

$$|P(C_n(E_{G(1^n)}(x)) = 1) - P(C_n(E_{G(1^n)}(y)) = 1)| < \frac{1}{p(n)}.$$

This definition is equivalent to definition 5.2.3 in [Gol04]. There is only one difference: We used a sequence of PPSC instead of PSC, which does not make any difference. This is the same argument as for theorem A.1.

Unfortunately this definition has a slightly different form than the statements in theorem 4.2 and 4.3. So we first show that definition 4.4 can be written in the same form.

Theorem 4.5. *Let (G, E, D) be an encryption scheme. Then the following two statements are equivalent:*

1. *An encryption scheme (G, E, D) has indistinguishable encryptions (non-uniform) as in definition 4.4.*
2. *For every positive, polynomially bounded function ℓ and all sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ with $x_n, y_n \in \{0,1\}^{\ell(n)}$ it is*

$$(E_{G(1^n)}(x_n)) \sim_p (E_{G(1^n)}(y_n)).$$

Proof. For the ease of discussion let us denote

$$\delta := |P(C_n(E_{G(1^n)}(x_n)) = 1) - P(C_n(E_{G(1^n)}(y_n)) = 1)|.$$

Further let us rewrite the two statements in short form (renamed x and y to x_n and y_n):

1. $\forall (C_n), \ell, p : \exists N : \forall n \geq N, x_n \in \{0,1\}^{\ell(n)}, y_n \in \{0,1\}^{\ell(n)} : \delta < 1/p(n)$,
2. $\forall\, \ell, (x_n \in \{0,1\}^{\ell(n)}), (y_n \in \{0,1\}^{\ell(n)}), (C_n), p : \exists N : \forall n \geq N : \delta < 1/p(n)$.

The second statement can be reordered to

$$\forall (C_n), \ell, p, (x_n \in \{0,1\}^{\ell(n)}), (y_n \in \{0,1\}^{\ell(n)}) : \exists N : \forall n \geq N : \delta < 1/p(n).$$

The direction $1 \Rightarrow 2$ is now easy to see, because if 1 holds then the same N exists in 2 and all the x_n and y_n in the two sequences fulfill the conditions in 1 and hence $\delta < 1/p(n)$ holds for them if $n \geq N$.

The other direction $2 \Rightarrow 1$ is is a little bit more tricky. We show this by contradiction. So we first logically invert the short forms:

a) $\exists(C_n), \ell, p : \forall N : \exists n \geq N, x_n \in \{0,1\}^{\ell(n)}, y_n \in \{0,1\}^{\ell(n)} : \delta \geq 1/p(n)$,

b) $\exists(C_n), \ell, p, (x_n \in \{0,1\}^{\ell(n)}), (y_n \in \{0,1\}^{\ell(n)}) : \forall N : \exists n \geq N : \delta \geq 1/p(n)$.

Now we have to show a) $\Rightarrow$ b). So if a) holds then for infinitely many n there are x_n and y_n for which $\delta \geq 1/p(n)$. So we can just take these x_n and y_n and take for the rest of the n randomly chosen $x_n \in \{0,1\}^{\ell(n)}$ and $y_n \in \{0,1\}^{\ell(n)}$. Now we have sequences (x_n) and (y_n) which fulfill b). □

Now let us have look at how this corresponds to PSC independence of plaintext and ciphertext.

Theorem 4.6. *An encryption scheme (G, E, D) has indistinguishable encryptions (non-uniform) if for all positive, polynomially bounded functions ℓ and PSCC sequences $(X_n)_{n\in\mathbb{N}}$ of random variables with $|X_n| = \ell(n)$ it is*

$$(X_n)_{n\in\mathbb{N}} \perp\!\!\!\perp_p (E_{G(1^n)}(X_n))_{n\in\mathbb{N}}.$$

Proof. We prove this by contradiction. Assume that there is a positive, polynomially bounded function ℓ, a positive polynomial p, and a sequence (C_n) of PPSC, so that for infinitely many $n \in \mathbb{N}$ there exist $x, y \in \{0,1\}^{\ell(n)}$ with

$$|P(C_n(E_{G(1^n)}(x)) = 1) - P(C_n(E_{G(1^n)}(y)) = 1)| \geq \frac{1}{p(n)}.$$

Then we have a positive, polynomially bounded function ℓ and we can define a sequence (x_n, y_n) with $x_n, y_n \in \{0,1\}^{\ell(n)}$ by taking x and y from above for the n where such x and y exist. Please note that these x and y have to be different to get a difference in the probabilities. For all other n we take random values in $\{0,1\}^{\ell(n)}$, such that $x_n \neq y_n$. We then define X_n as uniformly distributed random variables in $\{x_n, y_n\}$, which is PSCC.

We want to show now that $(X_n) \not\perp\!\!\!\perp_p (E_{G(1^n)}(X_n))$. Therefore we define $\tilde{X}_n = X_n$ and $\tilde{E}_n = E_{G(1^n)}(S_n)$, where $S_n \sim X_n$, but $S_n \perp\!\!\!\perp X_n$. Note that $\tilde{X}_n$ and $\tilde{E}_n$ are also PSCC with $(\tilde{X}_n) \sim (X_n)$, $(\tilde{E}_n) \sim (E_{G(1^n)}(X_n))$, and $(\tilde{X}_n) \perp\!\!\!\perp (\tilde{E}_n)$. Hence we have to show now that $(X_n, E_{G(1^n)}(X_n)) \not\approx_p (\tilde{X}_n, \tilde{E}_n)$.

Therefore we define

$$C'_n(x, e) := \begin{cases} C_n(e) & \text{if } x = x_n \\ 1 - C_n(e) & \text{else, especially if } x = y_n. \end{cases}$$

With that, the abbreviation $E(x) := E_{G(1^n)}(x)$, and the fact that

$$P(X_n = x_n) = P(X_n = y_n) = \frac{1}{2},$$

we get

$$\begin{aligned}
&2 \cdot |P(C'_n(X_n, E(X_n)) = 1) - P(C'_n(X_n, \tilde{E}_n) = 1)| \\
&= 2 \cdot |P(X_n = x_n) \cdot P(C'_n(x_n, E(x_n)) = 1 \mid X_n = x_n) \\
&\qquad + P(X_n = y_n) \cdot P(C'_n(y_n, E(y_n)) = 1 \mid X_n = y_n) \\
&\qquad - P(X_n = x_n) \cdot P(C'_n(x_n, \tilde{E}_n) = 1 \mid X_n = x_n) \\
&\qquad - P(X_n = y_n) \cdot P(C'_n(y_n, \tilde{E}_n) = 1 \mid X_n = y_n)| \\
&= |P(C'_n(x_n, E(x_n)) = 1) + P(C'_n(y_n, E(y_n)) = 1) \\
&\quad - P(C'_n(x_n, \tilde{E}_n) = 1) - P(C'_n(y_n, \tilde{E}_n) = 1)| \\
&= |P(C_n(E(x_n)) = 1) - 1 + P(C_n(E(y_n)) = 0) \\
&\quad - P(C_n(\tilde{E}_n) = 1) + 1 - P(C_n(\tilde{E}_n) = 0)| \\
&= |P(C_n(E(x_n) = 1) - P(C_n(E(y_n) = 1)|
\end{aligned}$$

which is $\geq \frac{1}{p(n)}$ for infinitely many $n \in \mathbb{N}$. And since $\tilde{X}_n = X_n$, we have that

$$(X_n, E_{G(1^n)}) \not\approx_p (\tilde{X}_n, \tilde{E}_n).$$

□

Theorem 4.7. *If an encryption scheme (G, E, D) has indistinguishable encryptions (non-uniform) then the following holds: For all positive, polynomially bounded functions ℓ and PSCC sequences $(X_n)_{n\in\mathbb{N}}$ of random variables with $|X_n| = \ell(n)$ it is*

$$(X_n)_{n\in\mathbb{N}} \not\perp\!\!\!\perp_p (E_{G(1^n)}(X_n))_{n\in\mathbb{N}}.$$

Proof. We prove this by contradiction. So assume there exists a positive, polynomially bounded function ℓ and a PSCC sequence (X_n) with $|X_n| = \ell(n)$, but $(X_n) \not\perp\!\!\!\perp_p (E_{G(1^n)}(X_n))$.

There exists a sequence (S_n) of PPSC with $S_n = (S_n^{(1)}, S_n^{(2)}) \sim (X_n, E_{G(1^n)}(X_n))$ and $S_n^{(2)}$ is computed by $S_n^{(2)} = E_{G(1^n)}(S_n^{(1)})$. Let $(\tilde{X}_n) = (X_n)$ and $(\tilde{E}_n) = (S_n^{(2)})$. Then $(\tilde{X}_n) \not\perp\!\!\!\perp (\tilde{E}_n)$ and hence $(\tilde{X}_n, \tilde{E}_n) \not\approx_p (X_n, E_{G(1^n)}(X_n))$, because otherwise the conditions of definition 2.5 would be fulfilled.

That means it exists a sequence (C_n) of PPSC so that

$$|P(C_n(\tilde{X}_n, \tilde{E}_n) = 1) - P(C_n(X_n, E_{G(1^n)}(X_n)) = 1)| \qquad (*)$$

is not negligible in n.

We now show that then the scheme (G, E, D) does not have indistinguishable encryptions. Since $(*)$ is not negligible there must be at least one instance (x_n, y_n) of $(X_n, S_n^{(1)})$ so that

$$|P(C_n(x_n, E_{G(1^n)}(y_n)) = 1) - P(C_n(x_n, E_{G(1^n)}(x_n)) = 1)|$$

is not negligible in n (otherwise it would be negligible for all instances and hence $(*)$ would be negligible, analog as in the proof to theorem A.1; here $(X_n, S_n^{(1)})$ takes the role of R_n). Let (C'_n) be a sequence of PPSC with $C'_n(e) = C_n(x_n, e)$, then we have

$$\begin{aligned}&|P(C'_n(E_{G(1^n)}(x_n)) = 1) - P(C'_n(E_{G(1^n)}(y_n)) = 1)| \\ &= |P(C_n(x_n, E_{G(1^n)}(x_n)) = 1) - P(C_n(x_n, E_{G(1^n)}(y_n)) = 1)|\end{aligned}$$

which is not negligible and hence (G, E, D) does not have indistinguishable encryptions. □

If we summarize the last two theorems, this yields the following theorem.

Theorem 4.8. *An encryption scheme (G, E, D) has indistinguishable encryptions (non-uniform) if and only if the following holds: For all positive, polynomially bounded functions ℓ and PSCC sequences $(X_n)_{n\in\mathbb{N}}$ of random variables with $|X_n| = \ell(n)$:*

$$(X_n)_{n\in\mathbb{N}} \perp\!\!\!\perp_p (E_{G(1^n)}(X_n))_{n\in\mathbb{N}}.$$

Or as alternative formulation:

Theorem 4.9. *Let (G, E, D) be an encryption scheme. Then the following two statements are equivalent:*

1. *For every positive, polynomially bounded function ℓ and every sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with $X_n \in \{0,1\}^{\ell(n)}$ it is*

$$(X_n) \perp\!\!\!\perp_p (E_{G(1^n)}(X_n)).$$

2. *For every positive, polynomially bounded function ℓ and all sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ with $x_n, y_n \in \{0,1\}^{\ell(n)}$ it is*

$$(E_{G(1^n)}(x_n)) \sim_p (E_{G(1^n)}(y_n)).$$

5. Some Open Questions

After having clarified the relationship between PSC independence and indistinguishable encryptions (non-uniform), there still remains the question if there is a similar relationship between computational independence and indistinguishable encryptions (uniform). The standard definition of a secure encryption includes also multiple messages and public key systems. It is also not clear if the relationship can be generalized to these cases.

A. Different Definitions

In this section we will have a look at the definition of PSC indistinguishability. It is obvious that if two sequences of random variables are indistinguishable by PPSC then they are also indistinguishable by PSC, because every PSC is also a PPSC. So we show only the opposite direction.

Theorem A.1. *Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be two sequences of random variables. If for all sequences $(C_n)_{n\in\mathbb{N}}$ of PSC*

$$|P(C_n(X_n) = 1) - P(C_n(Y_n) = 1)|$$

is negligible in n, then it holds that for all sequences $(D_n)_{n\in\mathbb{N}}$ of PPSC

$$|P(D_n(X_n) = 1) - P(D_n(Y_n) = 1)|$$

is negligible in n.

Proof. We proof this by contradiction. So assume that there is a sequence (D_n) of PPSC so that

$$|P(D_n(X_n) = 1) - P(D_n(Y_n) = 1)|$$

is not negligible in n.

Let us denote the internal randomness of the PPSC with R_n, so that $D_n(x) = D'_n(R_n, x)$, where D' is only a PSC and R_n a random variable, which is independent of X_n and Y_n and has polynomial length (in n). Then we have

$$\begin{aligned}
&|P(D_n(X_n) = 1) - P(D_n(Y_n) = 1)| \\
&= |P(D'_n(R_n, X_n) = 1) - P(D'_n(R_n, Y_n) = 1)| \\
&= \left|\sum_{r_n} P(R_n = r_n) \cdot P(D'_n(r_n, X_n) = 1) - \sum_{r_n} P(R_n = r_n) \cdot P(D'_n(r_n, Y_n) = 1)\right| \\
&= \sum_{r_n} P(R_n = r_n) \cdot |P(D'_n(r_n, X_n) = 1) - P(D'_n(r_n, Y_n) = 1)|
\end{aligned}$$

which is not negligible in n. So there must be at least one sequence $(\tilde{r}_n)$ for which

$$|P(D'_n(\tilde{r}_n, X_n) = 1) - P(D'_n(\tilde{r}_n, Y_n) = 1)|$$

is not negligible in n, otherwise the sum would be negligible, because $\sum_{r_n} P(R_n = r_n) = 1$.

If we construct C_n so that $C_n(x) = D'_n(\tilde{r}_n, x)$ for this particular sequence than C_n is PSC and

$$|P(C_n(X_n) = 1) - P(C_n(Y_n) = 1)| \quad = \quad |P(D'_n(\tilde{r}_n, X_n) = 1) - P(D'_n(\tilde{r}_n, Y_n) = 1)|$$

is not negligible in n, which is exactly what we wanted to show. □

References

[Fay08] Björn Fay. *Neue Ansätze für die Sicherheit der Random-Oracle-Methodik*. PhD thesis, Justus-Liebig-Universität, Otto-Behaghel-Str. 8, 35394 Gießen, 2008.

[Gol03] Oded Goldreich. *Foundations of Cryptography*, volume I, Basic Tools. Cambridge University Press, reprinted with corrections edition, 2003.

[Gol04] Oded Goldreich. *Foundations of Cryptography*, volume II, Basic Applications. Cambridge University Press, 2004.

[Sha48] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[Sha49] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *FOCS*, pages 80–91. IEEE Computer Society, 1982.