Composition of Random Systems: When Two Weak Make One Strong

Ueli Maurer and Krzysztof Pietrzak

ETH Zürich
Department of Computer Science
{maurer,pietrzak}@inf.ethz.ch

Abstract. A new technique for proving the *adaptive* indistinguishability of two systems, each composed of some component systems, is presented, using only the fact that corresponding component systems are *non-adaptively* indistinguishable. The main tool is the definition of a special monotone condition for a random system **F**, relative to another random system **G**, whose probability of occurring for a given distinguisher **D** is closely related to the distinguishing advantage ε of **D** for **F** and **G**, namely it is lower and upper bounded by ε and $\varepsilon(1 + \ln \frac{1}{\varepsilon})$, respectively.
A concrete instantiation of this result shows that the cascade of two random permutations (with the second one inverted) is indistinguishable from a uniform random permutation by adaptive distinguishers which may query the system from both sides, assuming the components' security only against non-adaptive one-sided distinguishers.
As applications we provide some results in various fields as almost k-wise independent probability spaces, decorrelation theory and computational indistinguishability (i.e., pseudo-randomness).

1 Introduction

1.1 Random Systems and the Distinguishing Problem

The statistical distance δ of two random variables A and B has a natural interpretation: The success probability of an optimal distinguisher in telling apart the two random variables A and B is $(1 + \delta)/2$.

It is much more intricate to deal with the indistinguishability of *random systems*[1] which take inputs $X_1, X_2, \ldots$ and generate, for each new input X_i, an output Y_i which depends probabilistically on the inputs and outputs seen so far. As always, we consider a distinguisher **D** which may interactively query a random system and, after some number k of queries, outputs a decision bit. For two random systems **F** and **G** and a distinguisher **D** one considers the two random experiments where **D** queries **F** and where **D** queries **G**, respectively,

[1] The term "random" is used here in the same sense as it is used in the term "random variable". It does not imply some kind of uniformity.

for some $k \geq 1$ queries. The advantage of $\mathbf{D}$ in distinguishing $\mathbf{F}$ and $\mathbf{G}$ is defined as difference of the probabilities of $\mathbf{D}$ outputting 1, in both random experiments.

Usually one is interested in the indistinguishability of a random system from some *perfect* random system with respect to *any* distinguisher from some general class of distinguishers (e.g. the class of all adaptive or the class of all non-adaptive distinguishers). In this work we will consider the problem of whether one can compose two or more random systems to obtain a new system whose security is superior to the security of any of its components. This is best illustrated by an example.

1.2 Composition of Random Systems: An Example

Let $\mathbf{E}$ (and likewise $\mathbf{F}$) be a random permutation[2] where the advantage of any *non-adaptive* distinguisher[3] for $\mathbf{E}$ and a uniform random permutation (URP) $\mathbf{P}$ is at most ε_k (where k is the number of queries). We can build a new random permutation $\mathbf{E} \circ \mathbf{F}$ by using $\mathbf{E}$ and $\mathbf{F}$ in a cascade (see Figure 1). Intuitively, this construction should be even "closer" to $\mathbf{P}$ than $\mathbf{E}$ or $\mathbf{F}$ individually. Indeed, Vaudenay [7] proved that the *non-adaptive* indistinguishability of $\mathbf{E} \circ \mathbf{F}$ is $2\varepsilon_k^2$, i.e., the distinguishing advantages are multiplied. The same statement holds if we replace (both occurrences) of non-adaptive with adaptive in the above [8].

If $\mathbf{E}$ and $\mathbf{F}$ are secure against *non-adaptive* distinguishers, can we say something about the *adaptive* security of $\mathbf{E} \circ \mathbf{F}$? The intuition here is that adaptivity cannot help too much as the output of $\mathbf{E}$ in the cascade is obscured by $\mathbf{F}$ and the input to $\mathbf{F}$ is randomized by the leading $\mathbf{E}$. This intuition is indeed correct. We will prove that if the *non-adaptive* security of $\mathbf{E}$ and $\mathbf{F}$ is ε_k, then $\mathbf{E} \circ \mathbf{F}$ has *adaptive* security $2\varepsilon_k(1 + \ln \frac{1}{\varepsilon_k})$. A lower bound of $\Omega(\varepsilon_k)$ for this advantage can easily be shown, in contrast to the above stated $O(\varepsilon_k^2)$ when only non-adaptive security is required. This leaves us (as an open problem) a gap on the order of $\ln \frac{1}{\varepsilon_k}$ between the upper and lower bound.

1.3 From Indistinguishability to Monotone Conditions and Back

The framework of [3] is based on the concept of monotone conditions defined for a random system. Intuitively, after each query to the system, such a condition can either be satisfied or can fail to be satisfied. Monotonicity means that once the condition has failed, it is never satisfied in the future. For example, such a condition could be that at a certain point internally in the system, for example at the input to a component, no collision has occurred. This no-collision condition is obviously monotone.

[2] By a random permutation (over some set $\mathcal{X}$) we mean a system which was chosen according to some distribution from all possible permutations over this set. If this distribution is uniform this system is called a *uniform* random permutation (URP).

[3] A non-adaptive distinguisher must choose the queries without seeing the outputs of the invoked system.

Consider two random systems $\mathbf{F}$ and $\mathbf{G}$ with compatible input and output alphabets. In this paper we will consider a monotone condition $\mathcal{A}$ for $\mathbf{F}$, denoted $\mathbf{F}^{\mathcal{A}}$, such that for any fixed input-output behaviour, the probability that $\mathbf{F}$ shows this behaviour *and* the condition occurs is upper bounded by the probability that $\mathbf{G}$ shows this behaviour. This will be denoted as $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$. Lemma 6 shows that if $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$, then for any distinguisher, its advantage in distinguishing $\mathbf{F}$ from $\mathbf{G}$ is upper bounded by the probability that it can make the condition $\mathcal{A}$ fail in $\mathbf{F}$.

One can intuitively think of such a monotone condition as a lamp placed on the system which goes on as soon as the condition fails. More radically, one could think of failure of the condition as a trigger for the system to explode. If the failure of a condition in a system $\mathbf{F}$ is interpreted as such a visible effect, then distinguishing $\mathbf{F}$ from another system $\mathbf{G}$ (without such a trigger) is trivial, provided the trigger event occurs, i.e., the condition fails.

In very many indistinguishability proofs in the literature, such monotone conditions lie at the core of the argument, although this is sometimes obscured in complicated arguments. In [3] it is shown how complex systems with several internal subsystems, each with a monotone condition, can be analysed. However, if one only knows that the two systems are *ε-indistinguishable* from a URF, without knowing a corresponding condition, then the technique of [3] fails. A main goal of this paper is therefore to define a special monotone condition $\mathcal{A}$ (called the maximum condition) for a random system $\mathbf{F}$, relative to a system $\mathbf{G}$, such that $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ and such that its probability ρ of not occurring (for any distinguisher $\mathbf{D}$) is closely related to the distinguishing advantage ε of $\mathbf{F}$ and $\mathbf{G}$ (for $\mathbf{D}$). More precisely, we provide two lemmas (Lemma 6 mentioned before and Lemma 9) which show that $\varepsilon \leq \rho \leq \varepsilon(1 + \ln \frac{1}{\varepsilon})$. This allows to prove the indistinguishability of two systems consisting of subsystems, knowing only that the subsystems are indistinguishable from a certain ideal system, but using the powerful framework based on monotone conditions.

Continuing the example of Section 1.2, let us discuss intuitively how this maximum condition allows to upper bound the *adaptive* security ε_k of $\mathbf{E} \circ \mathbf{F}$ assuming that the *non-adaptive* security of $\mathbf{E}$ (and likewise of $\mathbf{F}$) is at least γ_k (the k refers to the number of queries the distinguisher is allowed to make). Let $\mathcal{A}$ be the maximum condition for $\mathbf{E}$ relative to a URP $\mathbf{P}$, and let $\mathcal{B}$ be the maximum condition for $\mathbf{F}$ relative to $\mathbf{P}$. One can show (using Lemma 6) that $\varepsilon_k \leq \alpha_k$, where α_k is an upper bound on the maximal success probability of any adaptive distinguisher in making either $\mathcal{A}$ or $\mathcal{B}$ fail when querying $\mathbf{E}^{\mathcal{A}} \circ \mathbf{F}^{\mathcal{B}}$. Then using $\mathbf{E}^{\mathcal{A}} \preceq \mathbf{P}$ and $\mathbf{F}^{\mathcal{B}} \preceq \mathbf{P}$ one can show that this probability is at most the success probability of any *adaptive* distinguisher in making $\mathcal{A}$ fail in $\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}$ plus the probability of making $\mathcal{B}$ fail in $\mathbf{P} \circ \mathbf{F}^{\mathcal{B}}$. But in $\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}$ (and likewise in $\mathbf{P} \circ \mathbf{F}^{\mathcal{B}}$) adaptive strategies cannot be better than non-adaptive ones in making $\mathcal{A}$ fail as the output of $\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}$ is completely independent of the output of the internal system $\mathbf{E}$ on which $\mathcal{A}$ is defined. So $\varepsilon_k \leq 2\beta_k$ where β_k is an upper bound on the probability of any *non-adaptive* distinguisher in making $\mathcal{A}$ fail in $\mathbf{E}$ (and likewise $\mathcal{B}$ in $\mathbf{F}$). As $\mathcal{A}$ and $\mathcal{B}$ are maximum conditions we now obtain (from Lemma 9) $\beta_k \leq \gamma_k(1 + \ln \frac{1}{\gamma_k})$ and thus $\varepsilon_k \leq 2\gamma_k(1 + \ln \frac{1}{\gamma_k})$.

1.4 Outline of the Paper

In Section 2 the definitions of random systems, monotone conditions, the $\preceq$ relation and of distinguishers are given. In Section 3 first the *maximum condition* for two random systems is defined. Then we lower and upper bound (Lemmas 6 and 9) the success probability of a distinguisher in making the maximum condition fail (as described in Section 1.3).

As an application of our framework, in Section 4 we provide two theorems bounding the adaptive security of two systems (parallel execution and XOR of random functions and cascades of permutations) in terms of the non-adaptive security of the component systems. We also give an application for each of the theorems, the first is about k-wise independent sample spaces, the second about the cascade of random involutions. Section 5 discusses some more implications of the results. Section 6 states some open problems.

1.5 Notation

We denote sets by capital calligraphic letters (e.g. $\mathcal{X}$) and the corresponding capital letter X denotes a random variable taking values in $\mathcal{X}$. Concrete values for X are usually denoted by the corresponding small letter x. For a set $\mathcal{X}$ we denote by $\mathcal{X}^k$ the set of ordered k-tuples of elements from $\mathcal{X}$. $X^k = (X_1, X_2, \ldots, X_k)$ denotes a random variable taking values in $\mathcal{X}^k$ and a concrete value is usually denoted by $x^k = (x_1, x_2, \ldots, x_k)$.

Because we will consider different random experiments where the same random variables appear, we extend the standard notation for probabilities and expectations (e.g. $\mathsf{P}_V(v)$, $\mathsf{P}_{V|W}(v,w)$, $\mathsf{E}[V]$) by explicitly writing the considered random experiment $\mathcal{E}$ as a superscript, e.g. $\mathsf{P}^{\mathcal{E}}_V(v)$, $\mathsf{P}^{\mathcal{E}}_{V|W}(v,w)$ and $\mathsf{E}^{\mathcal{E}}[V]$. Equality of distributions means equality for all arguments, e.g.

$$\mathsf{P}^{\mathcal{E}_1}_V = \mathsf{P}^{\mathcal{E}_2}_V \iff \forall v \in \mathcal{V} : \mathsf{P}^{\mathcal{E}_1}_V(v) = \mathsf{P}^{\mathcal{E}_2}_V(v).$$

We sometimes use the notation $\mathsf{P}^{\mathcal{E}}_{\xi}$ instead of $\mathsf{P}^{\mathcal{E}}(\xi)$ to denote the probability of the event ξ.

2 Random Systems, Conditions and Distinguishers

2.1 Random Systems

Many cryptographic systems correspond to a probabilistic, possibly stateful (but often stateless) system which takes inputs $X_1, X_2, \ldots$ and generates, for each new input X_i, an output Y_i which depends probabilistically on X_i and the internal state.

In communication theory, a memoryless (i.e., stateless) communication channel with input X and output Y is modelled by a conditional probability distribution $P_{Y|X}$. In other words, $P_{Y|X}$ precisely captures the input-output behaviour of the channel, and it is unnecessary to consider the internals of the channel. In the same spirit, a possibly stateful and probabilistic system $\mathbf{F}$ that takes inputs $X_1, X_2, \ldots$ and generates an output Y_i for each new input X_i is modelled as a so-called random system [3], defined as a sequence of conditional probability distributions $P_{Y_i|X_1\cdots X_i,Y_1\cdots Y_{i-1}}$.

Definition 1 An $(\mathcal{X},\mathcal{Y})$-*random system* $\mathbf{F}$ is a (generally infinite) sequence of conditional probability distributions $\mathsf{P}^{\mathbf{F}}_{Y_i|X^iY^{i-1}}$, for $i \geq 1$. Two systems $\mathbf{F}$ and $\mathbf{G}$ are *equivalent*, denoted $\mathbf{F} \equiv \mathbf{G}$, if they correspond to the same random system, i.e., if $\mathsf{P}^{\mathbf{F}}_{Y_i|X^iY^{i-1}} = \mathsf{P}^{\mathbf{G}}_{Y_i|X^iY^{i-1}}$ for all $i \geq 1$.

The sequence $\mathsf{P}^{\mathbf{F}}_{Y_i|X^iY^{i-1}}$ for $i \geq 1$ also defines the sequence $\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}$ by

$$\mathsf{P}^{\mathbf{F}}_{Y^i|X^i} = \prod_{j=1}^{i} \mathsf{P}^{\mathbf{F}}_{Y_j|X^jY^{j-1}},$$

and vice versa by

$$\mathsf{P}^{\mathbf{F}}_{Y_i|X^iY^{i-1}} = \frac{\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}}{\mathsf{P}^{\mathbf{F}}_{Y^{i-1}|X^{i-1}}}.$$

As special classes of random systems we will consider random functions and random permutations, which are stateless random systems.

Definition 2 A *random function* $\mathcal{X} \to \mathcal{Y}$ (*random permutation* on $\mathcal{X}$) is a random variable which takes as values functions $\mathcal{X} \to \mathcal{Y}$ (permutations on $\mathcal{X}$). Throughout the paper the symbols $\mathcal{R}$ and $\mathcal{P}$ are used for the set of all random functions and the set of all random permutations respectively.

A *uniform random function (URF)* $\mathbf{R} : \mathcal{X} \to \mathcal{Y}$ (A *uniform random permutation (URP)* $\mathbf{P}$ on $\mathcal{X}$) is a random function with uniform distribution over all functions from $\mathcal{X}$ to $\mathcal{Y}$ (permutations on $\mathcal{X}$). Throughout the paper, the symbols $\mathbf{R}$ and $\mathbf{P}$ are used exclusively for the systems defined above.

2.2 Monotone Conditions

The concept of *monotone conditions* for random systems was introduced in [3]. A monotone condition $\mathcal{A}$ for a random-system $\mathbf{F}$ is a sequence $a_0, a_1, a_2, \ldots$ of events, where a_0 is the certain event and where a_i ($\overline{a}_i$) denotes the event that the condition is satisfied (failed) after the i'th query to $\mathbf{F}$ has been processed. As described above, monotone means that once the condition has failed, it can never hold again (i.e., $a_i \Rightarrow a_{i-1}$). A natural example of a monotone condition is a no-collision condition. As we are not interested in the behaviour of a random system after the condition has failed, and in fact this behaviour need in general not be defined, the definition below specifies the probability distribution of Y_i, given X^i and Y^{i-1}, only *together* with the event a_i, and conditioned on a_{i-1}. More formally, a random system with a monotone condition is defined like a random system, but the (conditional) probability distributions generally do not sum to 1. We use the term "partial" to denote such distributions which are not actually probability distributions.

Definition 3 An $(\mathcal{X},\mathcal{Y})$*-random system* $\mathbf{F}$ with a monotone condition $\mathcal{A}$, denoted $\mathbf{F}^{\mathcal{A}}$, is an infinite sequence of partial conditional probability distributions $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y_i|X^i Y^{i-1} a_{i-1}}$ for $i \geq 1$.

For any x^i and y^{i-1} we have

$$\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|X^i Y^{i-1} a_{i-1}}(x^i, y^{i-1}) = \sum_{y_i \in \mathcal{Y}} \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y_i|X^i Y^{i-1} a_{i-1}}(y_i, x^i, y^{i-1}) \leq 1.$$

The sequence $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y_i|X^i Y^{i-1} a_{i-1}}$ for $i \geq 1$ also defines the sequence $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y^i|X^i}$ by

$$\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y^i|X^i} = \prod_{j=1}^{i} \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_j Y_j|X^j Y^{j-1} a_{j-1}},$$

and vice versa.

Definition 4 We introduce a partial order $\preceq$ on input-output compatible random systems with monotone conditions, as follows:

$$\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}^{\mathcal{B}} \iff \forall i \geq 1: \ \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y^i|X^i} \leq \mathsf{P}^{\mathbf{G}^{\mathcal{B}}}_{b_i Y^i|X^i}.$$

In other words, $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}^{\mathcal{B}}$ if for all $i \geq 1$ and all $x^i \in \mathcal{X}^i$, $y^i \in \mathcal{Y}^i$, the probability that $\mathbf{F}^{\mathcal{A}}$ outputs y^i on input x^i and the condition $\mathcal{A}$ holds is at most the probability that $\mathbf{G}^{\mathcal{B}}$ will output y^i on input x^i and the condition $\mathcal{B}$ holds. We also define $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ (here one may think of $\mathbf{G}$ having a condition which never fails):

$$\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G} \iff \forall i \geq 1: \ \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y^i|X^i} \leq \mathsf{P}^{\mathbf{G}}_{Y^i|X^i}.$$

2.3 Distinguishers and Their Advantage

Definition 5 A *distinguisher* for an $(\mathcal{X},\mathcal{Y})$-random systems is a $(\mathcal{Y},\mathcal{X})$-random system $\mathbf{D}$ which can interactively query $(\mathcal{X},\mathcal{Y})$-random systems and finally outputs a bit.[4] For an $(\mathcal{X},\mathcal{Y})$-random system $\mathbf{F}$ we denote by $\mathbf{D}\diamond\mathbf{F}$ the random experiment where $\mathbf{D}$ interactively queries $\mathbf{F}$.

This definition refers to adaptive distinguishers. A non-adaptive distinguisher must fix all inputs $X_1, \ldots, X_k$ before seeing the outputs $Y_1, \ldots, Y_k$.

For the case of random permutations, we will consider mono-directional and bidirectional distinguishers (the latter only in the adaptive version). A bidirectional distinguisher can query the system from both sides.

Definition 6 The *advantage* of $\mathbf{D}$ in distinguishing $\mathbf{F}$ from $\mathbf{G}$, after k queries, denoted $\Delta^{\mathbf{D}}_k(\mathbf{F},\mathbf{G})$, is the absolute value of the difference of the probability of $\mathbf{D}$ outputting 1 in the two random experiments $\mathbf{D}\diamond\mathbf{F}$ and $\mathbf{D}\diamond\mathbf{G}$.

[4] An initial random variable $X_1 \in \mathcal{X}$ must also be defined.

Assuming without loss of generality that, after the query phase, $\mathbf{D}$ makes the optimal decision based on X^k and Y^k, we have[5]

$$\Delta_k^{\mathbf{D}}(\mathbf{F},\mathbf{G}) = \frac{1}{2}\sum_{\mathcal{X}^k\times\mathcal{Y}^k}\left|\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^kY^k} - \mathsf{P}^{\mathbf{D}\diamond\mathbf{G}}_{X^kY^k}\right|.$$

We denote the advantages of the best adaptive and the best non-adaptive distinguisher as follows:

$$\Delta_k(\mathbf{F},\mathbf{G}) \stackrel{\text{def}}{=} \max_{\mathbf{D}} \Delta_k^{\mathbf{D}}(\mathbf{F},\mathbf{G})$$

and

$$\begin{aligned}\delta_k(\mathbf{F},\mathbf{G}) &\stackrel{\text{def}}{=} \max_{\text{non-adaptive }\mathbf{D}} \Delta_k^{\mathbf{D}}(\mathbf{F},\mathbf{G})\\ &= \max_{x^k\in\mathcal{X}^k}\frac{1}{2}\sum_{y^k\in\mathcal{Y}^k}\left|\mathsf{P}^{\mathbf{F}}_{Y^k|X^k}(y^k,x^k) - \mathsf{P}^{\mathbf{G}}_{Y^k|X^k}(y^k,x^k)\right|.\end{aligned}$$

Definition 7 For a random system $\mathbf{F}^{\mathcal{A}}$ with a monotone condition, we let

$$\nu_k^{\mathbf{D}}(\mathbf{F},\overline{a}_k) \stackrel{\text{def}}{=} 1 - \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{a_k}$$

be the probability that $\mathbf{D}$ makes the condition fail with at most k queries. Furthermore, let

$$\nu_k(\mathbf{F},\overline{a}_k) \stackrel{\text{def}}{=} \max_{\mathbf{D}}\ \nu_k^{\mathbf{D}}(\mathbf{F},\overline{a}_k)$$

be the maximal probability in provoking $\overline{a}_k$ using any adaptive $\mathbf{D}$, and analogously for non-adaptive $\mathbf{D}$:

$$\mu_k(\mathbf{F},\overline{a}_k) \stackrel{\text{def}}{=} \max_{\text{non-adaptive }\mathbf{D}} \nu_k^{\mathbf{D}}(\mathbf{F},\overline{a}_k).$$

2.4 Random Systems as Components in Random Experiments

In this section we propose two lemmas which we will need several times in the sequel. Consider the random experiment $\mathcal{E}(\mathbf{F})$ where a random system $\mathbf{F}$, defined by a sequence of distributions $\mathsf{P}^{\mathbf{F}}_{Y_i|X^iY^{i-1}}$, is interacting with an environment $\mathcal{E}(\cdot)$, given by a sequence of distributions $\mathsf{P}^{\mathcal{E}(\cdot)}_{X_i|X^{i-1}Y^{i-1}}$.[6] Here $\mathcal{E}(\cdot)$ sends a query X_1 to $\mathbf{F}$ which answers with Y_1, then $\mathcal{E}(\cdot)$ sends a query X_2 and so on. So after k queries this random experiment defines a random variable X^kY^k.

[5] This definition has a natural interpretation in the random experiment where we first toss a uniform random coin $C \in \{0,1\}$. Then we let $\mathbf{D}$ (which has no a priori information on C) make k queries to a system $\mathbf{H}$ where $\mathbf{H}\equiv\mathbf{F}$ if $C=0$ and $\mathbf{H}\equiv\mathbf{G}$ if $C=1$. Here the expected probability that an optimal guess on C based on the k inputs and outputs of $\mathbf{H}$ will be correct is $1/2 + \Delta_k^{\mathbf{D}}(\mathbf{F},\mathbf{G})/2$.

[6] This definition of environment $\mathcal{E}(\cdot)$ is exactly the definition of an adaptive distinguisher. We will consider environments where a distinguisher is part of the environment, so as to avoid ambiguities we introduce the term environment here.

Lemma 1 For $\mathcal{E}(.)$ as just defined

$$\mathsf{P}^{\mathcal{E}(\mathbf{F})}_{X^kY^k} = \mathsf{P}^{\mathcal{E}(\cdot)}_{X^k|Y^{k-1}} \mathsf{P}^{\mathbf{F}}_{Y^k|X^k}.$$

Proof: This follows directly from the definition of this random experiment:

$$\mathsf{P}^{\mathcal{E}(\mathbf{F})}_{X^kY^k} = \prod_{j=1}^{k} \mathsf{P}^{\mathcal{E}(\cdot)}_{X_j|X^{j-1}Y^{j-1}} \mathsf{P}^{\mathbf{F}}_{Y_j|X^jY^{j-1}} = \mathsf{P}^{\mathcal{E}(\cdot)}_{X^k|Y^{k-1}} \mathsf{P}^{\mathbf{F}}_{Y^k|X^k}.$$

□

For example for the random experiment $\mathbf{D}\diamond\mathbf{F}$ (see Definition 5) we have

$$\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^iY^i} = \mathsf{P}^{\mathbf{D}}_{X^i|Y^{i-1}} \mathsf{P}^{\mathbf{F}}_{Y^i|X^i}. \tag{1}$$

For $\mathcal{E}(\cdot)$ as just defined we can also consider the random experiment $\mathcal{E}(\mathbf{F}^{\mathcal{A}})$.[7] It is straight-forward to prove the following lemma.

Lemma 2 For $\mathcal{E}(\cdot)$ as above let τ be any event defined on $\mathcal{E}(\cdot)$. Let a_τ be the event that the condition $\mathcal{A}$ holds at the timepoint where τ occurs. If $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ then

$$\mathsf{P}^{\mathcal{E}(\mathbf{F}^{\mathcal{A}})}_{\tau\wedge a_\tau} \leq \mathsf{P}^{\mathcal{E}(\mathbf{G})}_{\tau}$$

3 The Maximum Condition

Definition 8 For two $(\mathcal{X},\mathcal{Y})$-random systems $\mathbf{F}$ and $\mathbf{G}$, $\mathbf{F}$ with the *maximum condition* (relative to $\mathbf{G}$) is the random system with monotone condition $\mathbf{F}^{\mathcal{A}}$ defined by

$$\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|X^iY^i} = \min_{1\leq j\leq i}{}^{*} \left\{ \frac{\mathsf{P}^{\mathbf{G}}_{Y^j|X^j}}{\mathsf{P}^{\mathbf{F}}_{Y^j|X^j}} \right\}$$

and

$$\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_iY^i|X^i} = \mathsf{P}^{\mathbf{F}}_{Y^i|X^i} \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|X^iY^i}$$

for $i \geq 1$, where $\min^*$ means that the constant 1 is included among the terms to be minimised over, i.e., a $\min^*$ expression is always upper bounded by 1. We denote the maximum condition for $\mathbf{F}$ and $\mathbf{G}$ by $\mathbf{F}{\downarrow}\mathbf{G}$ and often give it a short name (e.g. $\mathcal{A} := \mathbf{F}{\downarrow}\mathbf{G}$).

The term "maximum condition" is motivated by the following lemma.

Lemma 3 For $\mathcal{A} := \mathbf{F}{\downarrow}\mathbf{G}$,

$$\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}.$$

Moreover, for all $\mathbf{F}^{\mathcal{B}}$,

$$\mathbf{F}^{\mathcal{B}} \preceq \mathbf{G} \implies \mathbf{F}^{\mathcal{B}} \preceq \mathbf{F}^{\mathcal{A}}.$$

[7] Note that formally, this is not a random experiment since it is only partially defined, but the notion of a probability of an event in this random experiment is naturally defined, provided the condition that $\mathcal{A}$ holds at the timepoint when the event occurs is taken as part of the event.

Proof: We first observe that the condition is monotone, because of the minimisation which implies $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|Y^iX^i} \leq \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_{i-1}|Y^{i-1}X^{i-1}}$. To prove $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$, observe that $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|Y^iX^i} \leq \mathsf{P}^{\mathbf{G}}_{Y^i|X^i}/\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}$ implies $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_iY^i|X^i} = \mathsf{P}^{\mathbf{F}}_{Y^i|X^i}\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|X^iY^i} \leq \mathsf{P}^{\mathbf{G}}_{Y^i|X^i}$.

To see that $\mathbf{F}^{\mathcal{B}} \preceq \mathbf{G}$ implies $\mathbf{F}^{\mathcal{B}} \preceq \mathbf{F}^{\mathcal{A}}$, note that for the maximum condition $\mathcal{A}$ the distribution $\mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|Y^iX^i}$ has everywhere the largest possible value still satisfying both requirements. So for any $\mathbf{F}^{\mathcal{B}} \preceq \mathbf{G}$ we have $\mathsf{P}^{\mathbf{F}^{\mathcal{B}}}_{b_i|Y^iX^i} \leq \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i|Y^iX^i}$ and thus $\mathbf{F}^{\mathcal{B}} \preceq \mathbf{F}^{\mathcal{A}}$. □

For the remainder of this section, let $\mathbf{F}$ and $\mathbf{G}$ be any $(\mathcal{X},\mathcal{Y})$-random systems. For each $i \geq 0$ we define the function $\lambda_i^{\mathbf{F},\mathbf{G}} : \mathcal{X}^i \times \mathcal{Y}^i \to [0,1]$ as

$$\lambda_i^{\mathbf{F},\mathbf{G}}(x^i,y^i) \stackrel{\text{def}}{=} \max\left\{\frac{\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}(y^i,x^i) - \mathsf{P}^{\mathbf{G}}_{Y^i|X^i}(y^i,x^i)}{\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}(y^i,x^i)}, 0\right\}.$$

In a random experiment where the random variables X^i and Y^i are defined we can consider the random variables Z_i and $\widetilde{Z}_i$ defined as

$$Z_i \stackrel{\text{def}}{=} \lambda_i^{\mathbf{F},\mathbf{G}}(X^i,Y^i) \qquad \text{and} \qquad \widetilde{Z}_i \stackrel{\text{def}}{=} \max_{0\leq j\leq i} Z_j. \tag{2}$$

The next two lemmas state that the expectation of these random variables in the random experiment $\mathbf{D}\diamond\mathbf{F}$ are the distinguishing advantage of $\mathbf{D}$ for $\mathbf{F}$ and $\mathbf{G}$ and the probability that $\mathbf{D}$ provokes the maximum condition for $\mathbf{F}$ (relative to $\mathbf{G}$) to fail, respectively.

Lemma 4

$$\Delta_k^{\mathbf{D}}(\mathbf{F},\mathbf{G}) = \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_k].$$

Proof:

$$\begin{aligned}
\Delta_k^{\mathbf{D}}(\mathbf{F},\mathbf{G}) &= \frac{1}{2}\sum_{\mathcal{X}^k\times\mathcal{Y}^k} |\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^kY^k} - \mathsf{P}^{\mathbf{D}\diamond\mathbf{G}}_{X^kY^k}| \\
&= \sum_{\mathcal{X}^k\times\mathcal{Y}^k} \max\left\{\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^kY^k} - \mathsf{P}^{\mathbf{D}\diamond\mathbf{G}}_{X^kY^k}, 0\right\} \\
&= \sum_{\mathcal{X}^k\times\mathcal{Y}^k} \mathsf{P}^{\mathbf{D}}_{X^k|Y^{k-1}} \max\left\{\mathsf{P}^{\mathbf{F}}_{Y^k|X^k} - \mathsf{P}^{\mathbf{G}}_{Y^k|X^k}, 0\right\} \\
&= \sum_{\mathcal{X}^k\times\mathcal{Y}^k} \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^kY^k} \max\left\{\frac{\mathsf{P}^{\mathbf{F}}_{Y^k|X^k} - \mathsf{P}^{\mathbf{G}}_{Y^k|X^k}}{\mathsf{P}^{\mathbf{F}}_{Y^k|X^k}}, 0\right\} \\
&= \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_k].
\end{aligned}$$

□

Lemma 5 For $\mathcal{A} := \mathbf{F}{\downarrow}\mathbf{G}$,

$$\nu_k^{\mathbf{D}}(\mathbf{F}^{\mathcal{A}}, \overline{a}_k) = \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[\widetilde{Z}_k].$$

Proof:

$$
\begin{aligned}
\nu_k^{\mathbf{D}}(\mathbf{F}^{\mathcal{A}}, \overline{a}_k) &= 1 - \sum_{\mathcal{X}^k \times \mathcal{Y}^k} \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}^{\mathcal{A}}}_{a_k X^k Y^k} \\
&= \sum_{\mathcal{X}^k \times \mathcal{Y}^k} \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^k Y^k} (1 - \mathsf{P}^{\mathcal{A}}_{a_k | X^k Y^k}) \\
&= \sum_{\mathcal{X}^k \times \mathcal{Y}^k} \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^k Y^k} \left(1 - \min_{1 \le j \le k}{}^* \left\{ \frac{\mathsf{P}^{\mathbf{G}}_{Y^j | X^j}}{\mathsf{P}^{\mathbf{F}}_{Y^j | X^j}} \right\} \right) \\
&= \sum_{\mathcal{X}^k \times \mathcal{Y}^k} \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^k Y^k} \max_{1 \le j \le k}{}^* \left\{ \frac{\mathsf{P}^{\mathbf{F}}_{Y^j | X^j} - \mathsf{P}^{\mathbf{G}}_{Y^j | X^j}}{\mathsf{P}^{\mathbf{F}}_{Y^j | X^j}} \right\} \\
&= \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[\widetilde{Z}_k].
\end{aligned}
$$

Here $\max^*$ means that the constant 0 is included among the terms to be minimised over, i.e., a $\max^*$ expression is always non-negative. □

Lemma 6 If $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$, then

$$\Delta_k^{\mathbf{D}}(\mathbf{F}, \mathbf{G}) \le \nu_k^{\mathbf{D}}(\mathbf{F}^{\mathcal{A}}, \overline{a}_k).$$

Proof: Let $\mathcal{B} := \mathbf{F}{\downarrow}\mathbf{G}$. Using the Lemmas 4 and 5 we get

$$
\begin{aligned}
\Delta_k^{\mathbf{D}}(\mathbf{F}, \mathbf{G}) &= \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_k] \\
&\le \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[\widetilde{Z}_k] \\
&= \nu_k^{\mathbf{D}}(\mathbf{F}^{\mathcal{B}}, \overline{b}_k) \\
&\le \nu_k^{\mathbf{D}}(\mathbf{F}^{\mathcal{A}}, \overline{a}_k).
\end{aligned}
$$

The last step is easily verified using $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{F}^{\mathcal{B}}$, which follows from Lemma 3. □

Definition 9 A sequence of random variables $V_0, V_1, \ldots,$ is a *sub-martingale* if for all $i \ge 0$

$$\mathsf{E}[V_{i+1} | V_0, \ldots, V_i] \ge V_i.$$

The proofs of the Lemmas 7 and 8 below can be found in Appendix A.

Lemma 7 Let $V_0, V_1, \ldots$ be a *sub-martingale* where $0 \le V_i \le 1$ for all i, and let $\widetilde{V}_n \stackrel{\text{def}}{=} \max_{0 \le j \le n} V_j$. Then

$$\mathsf{E}[\widetilde{V}_n] \le \mathsf{E}[V_n] \cdot (1 - \ln(\mathsf{E}[V_n]))\,.$$

Lemma 8 The sequence $Z_0, Z_1, \ldots$ as defined in (2) is a sub-martingale sequence in the random experiment $\mathbf{D}\diamond\mathbf{F}$, i.e.,

$$\forall i \ge 0: \ \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_{i+1} | Z_0, \ldots, Z_i] \ge Z_i.$$

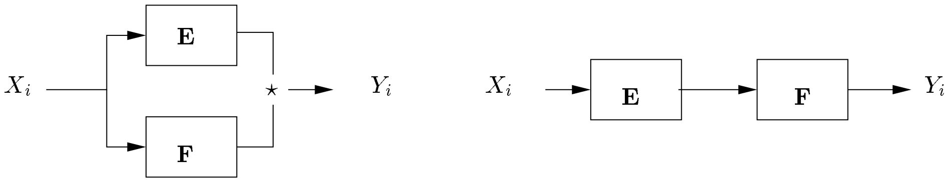

Fig. 1. The random systems $\mathbf{E} \star \mathbf{F}$ (left) and $\mathbf{E} \circ \mathbf{F}$ (right).

Lemma 9 For $\mathcal{A} := \mathbf{F}{\downarrow}\mathbf{G}$,

$$\nu_k^{\mathbf{D}}(\mathbf{F}^{\mathcal{A}}, \overline{a}_k) \leq \Delta_k^{\mathbf{D}}(\mathbf{F}, \mathbf{G}) \left(1 - \ln\left(\Delta_k^{\mathbf{D}}(\mathbf{F}, \mathbf{G})\right)\right).$$

Proof: Using Lemmas 8 and 7 we get

$$\mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[\widetilde{Z}_k] \leq \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_k] \cdot \left(1 - \ln\left(\mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_k]\right)\right).$$

Now one can apply the Lemmas 4 and 5. □

4 Stronger Security by Composition

Definition 10 A composition operator $\bowtie$ for a class of random systems $\mathcal{Q}$ is a binary operator $\mathcal{Q} \times \mathcal{Q} \to \mathcal{Q}$ which, given two random systems $\mathbf{E}, \mathbf{F} \in \mathcal{Q}$, defines how to combine $\mathbf{E}$ and $\mathbf{F}$ into a random system $\mathbf{E} \bowtie \mathbf{F} \in \mathcal{Q}$ where, on any invocation of $\mathbf{E} \bowtie \mathbf{F}$, the internal random systems $\mathbf{E}$ and $\mathbf{F}$ are invoked once. In this paper we will consider the two composition operators $\star$ and $\circ$ described below.

- Let $\mathbf{E}, \mathbf{F} \in \mathcal{R}$ be random functions $\mathcal{X} \to \mathcal{Y}$ (see Definition 2) and let $\star$ denote some group operation on $\mathcal{Y}$. We denote by $\mathbf{E} \star \mathbf{F} \in \mathcal{R}$ the random function defined by applying the input to $\mathbf{E}$ and $\mathbf{F}$ and then applying $\star$ to the outputs (see Figure 1, left).
- Let $\mathbf{E}, \mathbf{F} \in \mathcal{P}$ be random permutations over $\mathcal{X}$ (see Definition 2). We denote by $\mathbf{E} \circ \mathbf{F} \in \mathcal{P}$ the random permutation defined by applying the input to $\mathbf{E}$ and $\mathbf{F}$ to the output of $\mathbf{E}$ (see Figure 1, right).

Lemma 10 Consider a class $\mathcal{Q}$ of random systems and a composition operator $\bowtie$ on $\mathcal{Q}$. If there is a random system $\mathbf{I} \in \mathcal{Q}$ such that for all $\mathbf{F} \in \mathcal{Q}$ the following two conditions are satisfied

1. $\mathbf{I} \bowtie \mathbf{I} \equiv \mathbf{I}$.
2. $\nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{I}, \overline{a}_k) = \mu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{I}, \overline{a}_k)$ and $\nu_k(\mathbf{I} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{b}_k) = \mu_k(\mathbf{I} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{b}_k)$.[8]

[8] This means that whenever one of the two system $\bowtie$ takes as input is the "perfect" system $\mathbf{I}$, then the best adaptive distinguisher has no advantage over the best non-adaptive distinguisher in provoking some event defined on the other system.

Then for any $\mathbf{E}, \mathbf{F} \in \mathcal{Q}$ and any $k \geq 1$ we have

$$\delta_k(\mathbf{E}, \mathbf{I}) \leq \varepsilon \quad \wedge \quad \delta_k(\mathbf{F}, \mathbf{I}) \leq \varepsilon \quad \Longrightarrow \quad \Delta_k(\mathbf{E} \bowtie \mathbf{F}, \mathbf{I}) \leq 2\varepsilon(1 + \ln \tfrac{1}{\varepsilon}).$$

Proof: Let $\mathcal{A}$ ($\mathcal{B}$) be the maximum condition for $\mathbf{E}$ ($\mathbf{F}$), relative to $\mathbf{I}$, i.e.,

$$\mathcal{A} := \mathbf{E}{\downarrow}\mathbf{I} \qquad \text{and} \qquad \mathcal{B} := \mathbf{F}{\downarrow}\mathbf{I}.$$

Now we have (here $b_{\overline{a}}$, and likewise $a_{\overline{b}}$, denote the event that at any timepoint where the condition $\mathcal{B}$ holds, also the condition $\mathcal{A}$ holds)

$$\begin{aligned}
\Delta_k(\mathbf{E} \bowtie \mathbf{F}, \mathbf{I}) &= \Delta_k(\mathbf{E} \bowtie \mathbf{F}, \mathbf{I} \bowtie \mathbf{I}) \\
&\leq \nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{a}_k \vee \overline{b}_k) \\
&\leq \nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{a}_k \wedge b_{\overline{a}}) + \nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{b}_k \wedge a_{\overline{b}}) \\
&\leq \nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{I}, \overline{a}_k) + \nu_k(\mathbf{I} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{b}_k) \\
&= \mu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{I}, \overline{a}_k) + \mu_k(\mathbf{I} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{b}_k) \\
&\leq \mu_k(\mathbf{E}^{\mathcal{A}}, \overline{a}_k) + \mu_k(\mathbf{F}^{\mathcal{B}}, \overline{b}_k) \\
&\leq \delta_k(\mathbf{E}, \mathbf{I})\,(1 - \ln(\delta_k(\mathbf{E}, \mathbf{I}))) + \delta_k(\mathbf{F}, \mathbf{I})\,(1 - \ln(\delta_k(\mathbf{F}, \mathbf{I}))) \\
&\leq 2\varepsilon\left(1 + \ln \tfrac{1}{\varepsilon}\right).
\end{aligned}$$

The first step above follows from the first condition in the statement of the lemma. As for the second step, let $(\mathbf{E} \bowtie \mathbf{F})^{\mathcal{M}}$ be given by the partial distributions

$$\forall i \; : \; \mathsf{P}^{(\mathbf{E} \bowtie \mathbf{F})^{\mathcal{M}}}_{m_i Y^i | X^i} := \mathsf{P}^{\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{F}^{\mathcal{B}}}_{a_i \wedge b_i Y^i | X^i}.$$

Here $(\mathbf{E} \bowtie \mathbf{F})^{\mathcal{M}} \preceq \mathbf{I} \bowtie \mathbf{I}$ (which follows from $\mathbf{E}^{\mathcal{A}} \preceq \mathbf{I}$ and $\mathbf{F}^{\mathcal{B}} \preceq \mathbf{I}$) and we can apply Lemma 6 as $(\mathbf{E} \bowtie \mathbf{F})^{\mathcal{M}} \leq \nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{m}_k) = \nu_k(\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{F}^{\mathcal{B}}, \overline{a}_k \vee \overline{b}_k)$. The third step uses the union bound. Note that $\overline{a}_k \wedge b_{\overline{a}_k}$ is the event that the $\mathcal{A}$-condition fails before the $\mathcal{B}$-condition fails. The fourth step follows from Lemma 2. The fifth step follows by the second condition in the statement of the lemma. The sixth step follows as a non-adaptive distinguisher which queries $\mathbf{E}^{\mathcal{A}}$ (and likewise $\mathbf{F}^{\mathcal{B}}$) can simply "simulate" the system $\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{I}$ ($\mathbf{I} \bowtie \mathbf{F}^{\mathcal{B}}$).[9] The seventh step follows from Lemma 9, and the final step from the assumption of the lemma. □

Theorem 1 For random functions $\mathbf{E}, \mathbf{F} \in \mathcal{R}$ and $\star$ as in Definition 10,

$$\delta_k(\mathbf{E}, \mathbf{R}) \leq \varepsilon \quad \wedge \quad \delta_k(\mathbf{F}, \mathbf{R}) \leq \varepsilon \quad \Longrightarrow \quad \Delta_k(\mathbf{E} \star \mathbf{F}, \mathbf{R}) \leq 2\varepsilon\left(1 + \ln \tfrac{1}{\varepsilon}\right).$$

Proof: The Theorem follows from Lemma 10 by setting $\mathbf{I} \leftarrow \mathbf{R}$, $\mathcal{Q} \leftarrow \mathcal{R}$ and $\bowtie \leftarrow \star$. We only have to verify that the two points required by Lemma 10 are satisfied. As for the first point, $\mathbf{R} \star \mathbf{R} \equiv \mathbf{R}$ clearly holds. For the second point,

[9] Here we need that a query to $\mathbf{E}^{\mathcal{A}} \bowtie \mathbf{I}$ results in exactly one invocation of each subsystem. This guarantees we have no feedback which could not be simulated by a non-adaptive distinguisher.

note that the output of $\mathbf{E}^{\mathcal{A}} \star \mathbf{R}$ is independent of the output of the internal system $\mathbf{E}^{\mathcal{A}}$ on which our event is defined. So seeing the output cannot help in making the condition fail and we have $\nu_k(\mathbf{E}^{\mathcal{A}} \star \mathbf{R}, \overline{a}_k) = \mu_k(\mathbf{E}^{\mathcal{A}} \star \mathbf{R}, \overline{a}_k)$. By symmetry, also $\nu_k(\mathbf{R} \star \mathbf{F}^{\mathcal{B}}, \overline{b}_k) = \mu_k(\mathbf{R} \star \mathbf{F}^{\mathcal{B}}, \overline{b}_k)$ holds. □

As an application for this theorem one can consider an adaptive version of almost k-wise independent distributions (see [5], and [1] for simpler constructions). These are distributions over $\{0,1\}^n$ such that the bits at any k fixed positions are close (say some $\varepsilon > 0$ far away) to uniform.

It is natural to consider an adaptive version of ε-almost k-wise independence where the positions can be chosen adaptively by a distinguisher.

Definition 11 A distribution over $\{0,1\}^n$ is *adaptively* ε-almost k-wise independent if even an adaptive distinguisher, selecting the k positions adaptively, cannot distinguish the k bits from uniformly random with advantage more than ε.

Corollary 1. The distribution over $\{0,1\}^n$ defined by XOR-ing two ε-almost k-wise independent distributions is adaptively $2\varepsilon(1 + \ln \frac{1}{\varepsilon})$-almost k-wise independent.

The following theorem is inspired by Lemma 3 from [4]. We use the notation of [3] to denote bidirectional random permutations. If $\mathbf{F}$ is a random permutation, then $\langle \mathbf{F} \rangle$ is like $\mathbf{F}$, but it can be queried from both sides. The distinguisher can thus also issue a direction bit, in addition to the query, to indicate from which side it is supposed to be applied as input.

Theorem 2 For two random permutations $\mathbf{E}, \mathbf{F} \in \mathcal{P}$ and $\circ$ as in Definition 10,

$$\delta_k(\mathbf{E}, \mathbf{P}) \leq \varepsilon \quad \wedge \quad \delta_k(\mathbf{F}, \mathbf{P}) \leq \varepsilon \quad \Longrightarrow \quad \Delta_k(\mathbf{E} \circ \mathbf{F}, \mathbf{P}) \leq 2\varepsilon \left(1 + \ln \tfrac{1}{\varepsilon}\right).$$

If we take the inverse $\mathbf{F}^{-1}$ of $\mathbf{F}$ as the second element in the cascade, we additionally obtain security against bidirectional distinguishers:

$$\delta_k(\mathbf{E}, \mathbf{P}) \leq \varepsilon \quad \wedge \quad \delta_k(\mathbf{F}, \mathbf{P}) \leq \varepsilon \quad \Longrightarrow \quad \Delta_k(\langle \mathbf{E} \circ \mathbf{F}^{-1} \rangle, \langle \mathbf{P} \rangle) \leq 2\varepsilon \left(1 + \ln \tfrac{1}{\varepsilon}\right).$$

Proof: The first statement of the theorem follows from Lemma 10 by setting $\mathcal{Q} \leftarrow \mathcal{P}$, $\mathbf{I} \leftarrow \mathbf{P}$ and $\bowtie \leftarrow \circ$. For the second statement we must set $\mathcal{Q} \leftarrow \mathcal{P}$, $\mathbf{I} \leftarrow \langle \mathbf{P} \rangle$ and $\bowtie$ to be the mapping $\mathbf{E}, \mathbf{F} \rightarrow \langle \mathbf{E} \circ \mathbf{F}^{-1} \rangle$.

We will only prove the (slightly more involved) second statement of the theorem. Note that this statement is somewhat stronger than a direct application of Lemma 10 would imply: the precondition is $\delta_k(\mathbf{E}, \mathbf{P}) \leq \varepsilon \wedge \delta_k(\mathbf{F}, \mathbf{P}) \leq \varepsilon$, and not $\delta_k(\langle \mathbf{E} \rangle, \langle \mathbf{P} \rangle) \leq \varepsilon \wedge \delta_k(\langle \mathbf{F} \rangle, \langle \mathbf{P} \rangle) \leq \varepsilon$ as one would expect (we will come back to that point later).

We must verify that the two points required by Lemma 10 are satisfied. As for the first point, $\langle \mathbf{P} \circ \mathbf{P}^{-1} \rangle \equiv \langle \mathbf{P} \rangle$ clearly holds. For the second point, note that in $\langle \mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1} \rangle$ a query from the $\mathbf{P}^{-1}$ side results in a random value on the input and output of $\mathbf{E}^{\mathcal{A}}$. Thus a query from this side can be replaced by a random query from the $\mathbf{E}^{\mathcal{A}}$ side without changing the probability of an event defined on

$\mathbf{E}^{\mathcal{A}}$, and we have $\nu_k(\langle \mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1} \rangle, \overline{a}_k) = \nu_k(\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1}, \overline{a}_k)$. Now the output of $\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1}$ is completely independent of the output of the internal system $\mathbf{E}^{\mathcal{A}}$. So adaptive strategies cannot help in provoking an event defined on $\mathbf{E}^{\mathcal{A}}$, i.e., $\nu_k(\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1}, \overline{a}_k) = \mu_k(\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1}, \overline{a}_k)$. We have shown that $\nu_k(\langle \mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1} \rangle, \overline{a}_k) = \mu_k(\mathbf{E}^{\mathcal{A}} \circ \mathbf{P}^{-1}, \overline{a}_k)$, and by symmetry we get $\nu_k(\langle \mathbf{P} \circ (\mathbf{F}^{\mathcal{B}})^{-1} \rangle, \overline{a}_k) = \mu_k(\mathbf{F}^{\mathcal{B}} \circ \mathbf{P}^{-1}, \overline{a}_k)$. This is more than what is actually required by the second condition of Lemma 9. An inspection of the proof of the lemma shows that with this we also get a stronger statement (as mentioned before). □

As an application of this theorem, consider the cascade of two uniform random involutions over $\mathcal{X}$. An involution is a permutation which is its own inverse, and a uniform random involution (URI) on $\mathcal{X}$ is a permutation selected at random from the set of all involutions on $\mathcal{X}$. A URI $\mathbf{I}$ is non-adaptively indistinguishable from a URP $\mathbf{P}$ (the advantage is very small even for a large number of queries, actually $O(\sqrt{|\mathcal{X}|})$ queries are required to achieve a constant advantage), but an adaptive distinguisher can easily distinguish $\mathbf{I}$ from $\mathbf{P}$ simply by using any query X_1, setting $X_2 := Y_1$, and checking whether $Y_2 = X_1$. For a URI, this condition is always satisfied, whereas for a URP, it is satisfied only with exponentially small probability. We get the following corollary from Theorem 2

Corollary 1 Any adaptive bidirectional distinguisher must make in the order of $\sqrt{|\mathcal{X}|}$ queries to achieve a constant distinguishing advantage for a cascade of two uniform random involutions over $\mathcal{X}$ and a uniform random permutation over $\mathcal{X}$.

5 Discussion

We discuss a few implications of the results of this paper.

5.1 Pseudorandomness

As discussed in [3], essentially all proofs of computational indistinguishability of random systems consist basically of an information-theoretic indistinguishability proof. The results of this paper therefore have direct applications to computational settings. For example, in order to design a bidirectionally secure pseudorandom permutation (i.e., a block cipher secure against a combined chosen-message and chosen-ciphertext attack) from any pseudorandom function, it suffices to design an only non-adaptively secure random permutation $\mathbf{F}$ from a random function, then to replace the random function by a pseudorandom function, and to apply the construction twice with one of them inverted. More generally, this paper allows for new constructions of quasi-random systems, as discussed in [3].

5.2 Generalizing Indistinguishability Theory

This paper proposes two generalisations of the framework of [3], where the following technique to bound the indistinguishability $\Delta_k(\mathbf{F}, \mathbf{G})$ of two random systems $\mathbf{F}$ and $\mathbf{G}$ is used:

– Find conditions $\mathcal{A}$ and $\mathcal{B}$ such that $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$, which is defined as

$$\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}} \iff \forall i \geq 1: \quad \mathsf{P}^{\mathbf{F}^{\mathcal{A}}}_{a_i Y_i | X^i Y^{i-1} a_{i-1}} = \mathsf{P}^{\mathbf{G}^{\mathcal{B}}}_{b_i Y_i | X^i Y^{i-1} b_{i-1}}.$$

– Prove an upper bound on $\nu_k(\mathbf{F}^{\mathcal{A}}, \overline{a}_k)$, the success probability of any distinguisher in making the condition fail with k queries. Now (by Lemma 7 from [3]) $\Delta_k(\mathbf{F}, \mathbf{G}) \leq \nu_k(\mathbf{F}^{\mathcal{A}}, \overline{a}_k)$ and we are done.

The first generalisation is that by Lemma 6 we may replace the requirement $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ with the weaker requirement $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ and the second point still holds. As $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ implies $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ but $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ does not imply the existence of $\mathcal{B}$ such that $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$, this requirement is strictly weaker.

The second generalisation is that, due to Lemma 9, one can go from indistinguishability to monotone conditions: If $\Delta_k(\mathbf{F}, \mathbf{G}) \leq \varepsilon$, then there always exists a monotone condition (i.e. the maximum condition for $\mathbf{F}$ and $\mathbf{G}$) $\mathcal{A}$ such that $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ and $\nu_k(\mathbf{F}^{\mathcal{A}}, \overline{a}_k) \leq \varepsilon(1 + \ln \frac{1}{\varepsilon})$. So using the above framework (with $\mathbf{F}^{\mathcal{A}} \preceq \mathbf{G}$ instead of $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ in the first step) does not inherently restrict the set of provable statements.

This is in sharp contrast to the original $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ requirement, as there are, for any $\varepsilon > 0$, random systems $\mathbf{F}$ and $\mathbf{G}$ where (for some k, or rather some range for k) $\Delta_k(\mathbf{F}, \mathbf{G}) \leq \varepsilon$, but for any conditions $\mathcal{A}$ and $\mathcal{B}$ which satisfy $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ we have $\nu_k(\mathbf{F}^{\mathcal{A}}, \overline{a}_k) \geq 1 - \varepsilon$. For such systems this framework (with the original $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ requirement in the first step) is not applicable.

As an example for such systems, let the first be a source of uniform random bits and the second be a source where each bit is not completely uniform but has some small bias α. Here $\Delta_k(\mathbf{F}, \mathbf{G}) \approx \sqrt{k}\alpha$ (see [6]) and $\nu_k(\mathbf{F}^{\mathcal{A}}, \overline{a}_k) \approx 1 - (1-\alpha)^{k/2} \approx \alpha k/2$. Thus choosing α small and k large enough we can achieve any $\varepsilon > 0$ as described.

5.3 Decorrelation Theory

Decorrelation theory was introduced by Vaudenay as a tool to prove security of block ciphers against d-iterated attacks, this class of attacks includes linear and differential cryptanalysis. Loosely speaking, in a d-iterated attack a distinguisher, which tries to distinguish the block cipher from a uniform random permutation, is limited to look at blocks of at most d queries at the same time. Decorrelation theory is based on different matrix norms. We refer to [7] for the definition of these norms and note that

For a random permutation $\mathbf{E}$ over $\mathcal{M}$ let $[E]^d$ denote the $\mathcal{M}^d \times \mathcal{M}^d$ matrix where the $(x^d, y^d) \in \mathcal{M}^d \times \mathcal{M}^d$ entry of $[E]^d$ is $\mathsf{P}^{\mathbf{E}}_{Y^d|X^d}(x^d, y^d)$. Now let D be a distance over the matrix space $\mathbb{R}^{\mathcal{M}^d \times \mathcal{M}^d}$. The d-wise decorrelation bias of the permutation E is the distance (C^* denotes the distribution of the uniform random permutation)

$$\mathrm{DecP}^d_D(E) = D([E]^d, [C^*]^d).$$

In the above definition the distance D can be replaced by a matrix norm. The matrix norms considered are denoted $||\cdot||_\infty$, $||\cdot||_a$ and $||\cdot||_s$. These norms have a natural interpretation as they are exactly twice the advantage of the best (non-adaptive, adaptive or bidirectional) distinguisher making at most d queries in distinguishing $\mathbf{E}$ from a URP, i.e. (note that here the first terms are in our notation)

$$\delta_d(\mathbf{E},\mathbf{P}) = \frac{1}{2}||[E]^d - [C^*]^d||_\infty = \frac{1}{2}\mathrm{DecP}^d_\infty(E) \tag{3}$$

$$\Delta_d(\mathbf{E},\mathbf{P}) = \frac{1}{2}||[E]^d - [C^*]^d||_a = \frac{1}{2}\mathrm{DecP}^d_a(E) \tag{4}$$

$$\Delta_d(\langle\mathbf{E}\rangle,\langle\mathbf{P}\rangle) = \frac{1}{2}||[E]^d - [C^*]^d||_s = \frac{1}{2}\mathrm{DecP}^d_s(E) \tag{5}$$

The main theorem of [7] states that if a block cipher has small $2d$-wise ($||\cdot||_\infty$,$||\cdot||_a$ or $||\cdot||_s$) decorrelation bias it is secure against any d-iterated attack performed by any (non-adaptive, adaptive or bidirectional) distinguisher.

We can plug in (3), (4) and (5) directly into Theorem 1 and get the first nontrivial relations known among this norms.

Corollary 2

$$\mathrm{DecP}^d_\infty(E) \le \varepsilon \;\wedge\; \mathrm{DecP}^d_\infty(F) \le \varepsilon \;\Rightarrow\; \mathrm{DecP}^d_a(E\circ F) \le 2\varepsilon\left(1+\ln\tfrac{2}{\varepsilon}\right)$$

$$\mathrm{DecP}^d_\infty(E) \le \varepsilon \;\wedge\; \mathrm{DecP}^d_\infty(F) \le \varepsilon \;\Rightarrow\; \mathrm{DecP}^d_s(E\circ F^{-1}) \le 2\varepsilon\left(1+\ln\tfrac{2}{\varepsilon}\right).$$

The second statement of the corollary now implies that using a block-cipher with small $2d$-wise decorrelation bias in the ∞ norm against non-adaptive chosen plaintext d-iterated attacks in a cascade (with independent keys, the second time in decrypt mode) results in a block cipher which is secure against adaptive combined chosen plaintext and ciphertext $2d$-iterated attacks.

6 Conclusions and Open Problems

It would be interesting to have a similar framework as the one proposed in this paper for the computational setting. For example, the computational analog of Theorem 2 would state that the cascade of two block-ciphers, each secure against non-adaptively chosen plaintext attacks, is secure against adaptive chosen plaintext/ciphertext adversaries.

As already mentioned in the introduction, there is a gap in the order of $\ln\frac{1}{\varepsilon}$ between the $O(\varepsilon\ln\frac{1}{\varepsilon})$ bound proven in Theorems 1 and 2 and an easy to show $\Omega(\varepsilon)$ lower bound for the respective terms. However, Lemmas 6 and 9 can be shown to be tight up to a (small) multiplicative constant, so we cannot hope to close this gap (i.e. showing an upper bound of $O(\varepsilon)$) by improving on them. But trying to find a concrete example (of random systems) for which a matching $\Omega(\varepsilon\ln\frac{1}{\varepsilon})$ lower bound can be proven seems promising to us.

References

1. N. Alon, O. Goldreich, J. Hastad, R. Peralta, Simple construction of almost k-wise independent random variables, *Random Structures and Algorithms*, vol. 3, no. 3, pp. 289–304, 1992.
2. M. Luby and C. Rackoff, How to construct pseudo-random permutations from pseudo-random functions, *SIAM J. on Computing*, vol. 17, no. 2, pp. 373–386, 1988.
3. U. Maurer, Indistinguishability of random systems, *Advances in Cryptology - EUROCRYPT '02*, Lecture Notes in Computer Science, vol. 2332, pp. 110–132, Springer-Verlag, 2002.
4. U. Maurer and K. Pietrzak, The security of many-round Luby-Rackoff pseudo-random permutations, *Advances in Cryptology - EUROCRYPT '03*, Lecture Notes in Computer Science, vol. 2656, pp. 544–561, Springer-Verlag, 2003.
5. J. Naor and M. Naor, Small-bias probability spaces: Efficient constructions and applications, *SIAM Journal on Computing*, vol. 22, no. 4, pp. 838–356, 1993.
6. R. Renner, The Statistical Distance of Independently Repeated Experiments, Manuscript, available at http://www.crypto.ethz.ch/~renner/publications.html
7. S. Vaudenay, Provable security for block ciphers by decorrelation, *Proceedings of STACS'98*, Lecture Notes in Computer Science, vol. 1373, Springer-Verlag, pp. 249–275, 1998.
8. S. Vaudenay, Adaptive-attack norm for decorrelation and super-pseudorandomness, *Proc. of SAC'99*, Lecture Notes in Computer Science, vol. 1758, pp. 49–61, Springer-Verlag, 2000.

A Martingales

In what follows, let $\widetilde{V}_n \stackrel{\text{def}}{=} \max_{0\le j\le n} V_j$. The following lemma is known as the Kolmogorov-Doob inequality.

Lemma 11 *Let $V_0, V_1, \ldots$ be a sub-martingale sequence where the V_i are non-negative. Then, for every n,*

$$\mathsf{P}[\widetilde{V}_n \ge \lambda] \le \frac{\mathsf{E}[V_n]}{\lambda}.$$

Proof of Lemma 7: We restate the lemma for the reader's convenience: If $V_0, V_1, \ldots$ is a *sub-martingale* sequence where $0 \le V_i \le 1$ for all i, then

$$\mathsf{E}[\widetilde{V}_n] \le \mathsf{E}[V_n] \cdot (1 - \ln(\mathsf{E}[V_n])).$$

Let $\psi(r)$ denote the function

$$\psi(r) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } r < \mathsf{E}[V_n] \\ \mathsf{E}[V_n]/r & \text{if } \mathsf{E}[V_n] \le r \le 1 \\ 0 & \text{if } r > 1 \end{cases}$$

With Lemma 11 and $0 \le \widetilde{V}_n \le 1$ (which follows from $0 \le V_i \le 1$) we see that

$$\forall r : \mathsf{P}[\widetilde{V}_n \ge r] \le \psi(r).$$

So we can upper bound $\mathsf{E}[\widetilde{V}_n]$ as

$$
\begin{aligned}
\mathsf{E}[\widetilde{V}_n] &\leq -\int_{-\infty}^{\infty} \psi'(r)\, r\, dr \\
&= -\int_{\mathsf{E}[V_n]}^{1} \left(\frac{\mathsf{E}[V_n]}{r}\right)' r\, dr + \mathsf{E}[V_n] \\
&= \int_{\mathsf{E}[V_n]}^{1} \frac{\mathsf{E}[V_n]}{r^2}\, r\, dr + \mathsf{E}[V_n] \\
&= -\ln(\mathsf{E}[V_n]) \cdot \mathsf{E}[V_n] + \mathsf{E}[V_n].
\end{aligned}
$$

□

Proof of Lemma 8: We restate the lemma for the reader's convenience: $Z_1, Z_2, \ldots$ as defined in (2) is a sub-martingale sequence in the random experiment $\mathbf{D}\diamond\mathbf{F}$, i.e.,

$$\forall i \geq 0 : \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_{i+1}|Z_0, \ldots, Z_i] \geq Z_i.$$

Because the $Z_0, \ldots, Z_i$ are determined by X^iY^i, we can prove the (stronger) statement

$$\forall i \geq 0 : \mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_{i+1}|X^iY^i] \geq Z_i$$

instead. Below the sums over $\mathcal{X} \times \mathcal{Y}$ always apply to the random variables X_{i+1} and Y_{i+1}. Lemma 1 is used several times.

$$
\begin{aligned}
&\mathsf{E}^{\mathbf{D}\diamond\mathbf{F}}[Z_{i+1}|X^iY^i] \\
&= \sum_{\mathcal{X}\times\mathcal{Y}} \underbrace{\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X_{i+1}Y_{i+1}|X^iY^i}}_{\mathsf{P}^{\mathbf{D}}_{X_{i+1}|X^iY^i}\mathsf{P}^{\mathbf{F}}_{Y_{i+1}|X^{i+1}Y^i}} \overbrace{\max\left\{\frac{\mathsf{P}^{\mathbf{F}}_{Y^{i+1}|X^{i+1}} - \mathsf{P}^{\mathbf{G}}_{Y^{i+1}|X^{i+1}}}{\underbrace{\mathsf{P}^{\mathbf{F}}_{Y^{i+1}|X^{i+1}}}_{\mathsf{P}^{\mathbf{F}}_{Y_{i+1}|X^{i+1}Y^i}\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}}}, 0\right\}}^{Z_{i+1}} \\
&= \frac{1}{\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}} \sum_{\mathcal{X}\times\mathcal{Y}} \underbrace{\mathsf{P}^{\mathbf{D}}_{X_{i+1}|X^iY^i}}_{\mathsf{P}^{\mathbf{D}}_{X^{i+1}|Y^i}/\mathsf{P}^{\mathbf{D}}_{X^i|Y^i}} \max\left\{\mathsf{P}^{\mathbf{F}}_{Y^{i+1}|X^{i+1}} - \mathsf{P}^{\mathbf{G}}_{Y^{i+1}|X^{i+1}}, 0\right\} \\
&= \frac{1}{\mathsf{P}^{\mathbf{D}}_{X^i|Y^i}\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}} \sum_{\mathcal{X}\times\mathcal{Y}} \max\left\{\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^{i+1}Y^{i+1}} - \mathsf{P}^{\mathbf{D}\diamond\mathbf{G}}_{X^{i+1}Y^{i+1}}, 0\right\} \\
&\geq \frac{1}{\mathsf{P}^{\mathbf{D}}_{X^i|Y^i}\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}} \max\left\{\underbrace{\sum_{\mathcal{X}\times\mathcal{Y}} \mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^{i+1}Y^{i+1}}}_{\mathsf{P}^{\mathbf{D}\diamond\mathbf{F}}_{X^iY^i}} - \underbrace{\sum_{\mathcal{X}\times\mathcal{Y}} \mathsf{P}^{\mathbf{D}\diamond\mathbf{G}}_{X^{i+1}Y^{i+1}}}_{\mathsf{P}^{\mathbf{D}\diamond\mathbf{G}}_{X^iY^i}}, 0\right\} \\
&= \frac{\mathsf{P}^{\mathbf{D}}_{X^i|Y^i}}{\mathsf{P}^{\mathbf{D}}_{X^i|Y^i}} \max\left\{\frac{\mathsf{P}^{\mathbf{F}}_{Y^i|X^i} - \mathsf{P}^{\mathbf{G}}_{Y^i|X^i}}{\mathsf{P}^{\mathbf{F}}_{Y^i|X^i}}, 0\right\} \\
&= Z_i.
\end{aligned}
$$

□