Indistinguishability of Random Systems

Ueli Maurer*

ETH Zurich
Department of Computer Science
maurer@inf.ethz.ch

Abstract. An $(\mathcal{X}, \mathcal{Y})$-random system takes inputs $X_1, X_2, \ldots \in \mathcal{X}$ and generates, for each new input X_i, an output $Y_i \in \mathcal{Y}$, depending probabilistically on $X_1, \ldots, X_i$ and $Y_1, \ldots, Y_{i-1}$. Many cryptographic systems like block ciphers, MAC-schemes, pseudo-random functions, etc., can be modeled as random systems, where in fact Y_i often depends only on X_i, i.e., the system is stateless. The security proof of such a system (e.g. a block cipher) amounts to showing that it is indistinguishable from a certain perfect system (e.g. a random permutation).
We propose a general framework for proving the indistinguishability of two random systems, based on the concept of the equivalence of two systems, conditioned on certain events. This abstraction demonstrates the common denominator among many security proofs in the literature, allows to unify, simplify, generalize, and in some cases strengthen them, and opens the door to proving new indistinguishability results.
We also propose the previously implicit concept of quasi-randomness and give an efficient construction of a quasi-random function which can be used as a building block in cryptographic systems based on pseudo-random functions.

Key words. Indistinguishability, random systems, pseudo-random functions, pseudo-random permutations, quasi-randomness, CBC-MAC.

1 Introduction

1.1 Indistinguishability

Indistinguishability of two systems, introduced by Blum and Micali [7] for defining pseudo-random bit generators, is a central concept in cryptographic security definitions and proofs. The simplest distinguisher problem is that for two random variables: The success probability (or advantage) of the optimal distinguisher is just the distance of the two probability distributions. As a slight generalization, one can define indistinguishability for infinite sequences of random variables, e.g. of a pseudo-random bit generator from a true random bit generator [7].

It is substantially more difficult to investigate the indistinguishability of two *interactive* random systems **F** and **G** because the distinguisher can *adaptively* choose its inputs (also called queries) to the system, depending on the outputs

* Supported in part by the Swiss National Science Foundation, grant 2000-055466.98

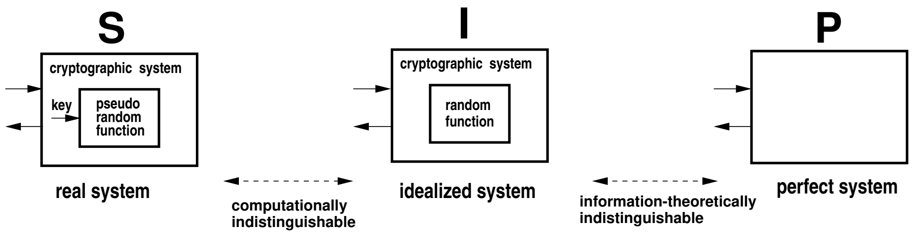

Fig. 1. Real system **S**, idealized system **I**, and perfect system **P**.

seen for previous inputs. Every distinguisher **D** defines a pair of generally very complex random experiments, one when **D** queries **F** and the other one when **D** queries **G**. A security proof requires to prove an upper bound, holding for every **D**, on the difference of the probability of some event in the corresponding two experiments. In general, this is a hard probability-theoretic problem.

1.2 Security Proofs Based on Pseudo-Random Functions

The security of many cryptographic systems (e.g., block ciphers, message authentication codes, challenge-response protocols) is based on the assumption that a certain component (e.g. DES, IDEA, or Rijndael) used in the construction is a pseudo-random function (PRF) [8]. Such systems are proven secure, relative to this assumption, by showing that any algorithm for breaking the system can be transformed into a distinguisher for the PRF. For example, in a classic paper, Luby and Rackoff [10] showed how to construct a secure block cipher from any pseudo-random function, and Bellare et al. [2] proved the security of the CBC-MAC. The following general steps can be used to prove the security of a cryptographic system based on a pseudo-random function (cf. Fig. 1):

1. The attacker's capabilities, i.e., the types and number of allowed queries to **S** are defined. Moreover, security of **S** is defined by specifying what it means for the attacker to break **S**, and a purely theoretical *perfect system* **P** is defined which is trivially secure (see examples below).
2. One considers an *idealized system* **I** obtained from **S** by replacing the PRF by a truly random function and proves that **I** and **P** are information-theoretically indistinguishable: no adaptive computationally unbounded distinguisher algorithm **D** has a non-negligible advantage unless it queries the system for an infeasibly large (e.g. super-polynomial) number of queries.[1]
3. Hence, because **S** is computationally indistinguishable from **I** if the underlying function is pseudo-random, **S** is also computationally indistinguishable from **P**. Because **P** is unbreakable, there exists no breaking algorithm for **S** since it could directly be used as a distinguisher for **S** and **P**.

[1] This is the only technical step in such a proof. It is purely information-theoretic, not involving complexity theory, and is the subject of this paper.

Example 1. For a block cipher the attacker is assumed to obtain the ciphertexts (plaintexts) for adaptively chosen plaintexts (ciphertexts). A perfect block cipher is a truly random permutation on the input space.

Example 2. For a MAC, the attacker may obtain the MAC for arbitrary adaptively chosen messages. A perfect MAC is a random oracle, i.e., a random function from $\{0,1\}^*$, the finite-length bit strings, to the l-bit strings (e.g. $l = 64$).

1.3 Previous Work

Many authors were intrigued by the complexity of certain security proofs in the literature, most notably [10], and have given shorter proofs for these and more general results. It is beyond the scope of this paper to discuss all of these results, but a few are mentioned below. Patarin [14, 15] developed a technique called "coefficient H method" and used it to analyze Feistel ciphers, even with more than four rounds [16]. To the best of our knowledge, the concept of conditioning events in security proofs was first made explicit in [11] and [12] where, using appropriate conditioning events, the proof for the Luby-Rackoff construction and generalizations thereof was shown to boil down to simple collision arguments (but the proof was stated only for non-adaptive distinguishers). Naor and Reingold [18] generalized the Luby-Rackoff constructions. In a sequence of papers (e.g., see [21, 22]), Vaudenay developed decorrelation theory and applied it to the design of block ciphers and the analysis of constructions like the CBC-MAC. Petrank and Rackoff [17] gave a generalized treatment of the CBC-MAC.

1.4 Contributions of the Paper and Sketch of the Framework

This paper defines the natural concept of a random system and proposes a general framework for proving the indistinguishability of two random systems **F** and **G** by identifying internal events such that, conditioned on these events, **F** and **G** are equivalent, i.e., have the identical input-output behavior.

The advantage in distinguishing **F** and **G** with k queries and unbounded computing power is shown to be at most the probability of success in provoking one of these events *not* to occur (Theorem 1). Under a certain condition, adaptive strategies can be shown to be not more powerful than non-adaptive strategies, thus allowing to eliminate the distinguisher from the analysis (Theorem 2 and Corollary 1).

The framework is illustrated for a few application areas and by giving simple and intuitive analyses and generalizations of some classical results. Due to the high level of abstraction, one can apply the basic techniques in settings where previous proof techniques appeared to be too complex or where changing a small detail in the construction requires a complete rehash of the proof.

Moreover, in some cases one can prove stronger bounds. For instance, under certain conditions one can prove that if a construction involves several components, each indistinguishable from a certain perfect system, then the overall system is distinguishable from its perfect counterpart with probability only the *product* (rather than the sum or the maximum) of the maximal distinguishing probabilities of the component systems (Theorem 3).

1.5 A Motivating Example

The security proof [2] for the CBC-MAC (cf. Fig. 6), and several generalizations thereof, will follow as a simple consequence of our framework (see Section 6). Roughly speaking, the proof consists of the following simple steps. First, conditioned on the event that all inputs to the internal random function $\mathbf{R}$ (modeling the PRF used in an actual implementation), corresponding to a final block of a message, are distinct, the CBC-MAC behaves like a random oracle, i.e., a perfect MAC. Second, one can hence restrict attention to algorithms trying to prevent this event from occurring by any adaptive choice of the inputs. Third, since the outputs are independent of the inputs, given this event, one can restrict the analysis to non-adaptive strategies, which turn out to be easy to analyze.

1.6 Quasi-Randomness

The general idea behind such cryptographic constructions is to "package" a given amount of randomness such that it appears to any observer as a random system $\mathbf{S}$ which behaves essentially like a (in some sense) perfect random system $\mathbf{P}$ containing a much larger amount of randomness. If $\mathbf{S}$ is computationally indistinguishable from $\mathbf{P}$, it is generally called pseudo-random (with respect to $\mathbf{P}$). Informally, we call $\mathbf{S}$ *quasi-random* (with respect to $\mathbf{P}$) if it is indistinguishable from $\mathbf{P}$, provided only that the *amount of interaction* (e.g. the number of queries) is bounded, but with otherwise unbounded computational resources.

An important question, addressed in this paper, is how an efficient quasi-random system $\mathbf{S}$ of a certain type can be constructed, using as few random bits as possible, and indistinguishable from the corresponding perfect system $\mathbf{P}$ for as many queries as possible.

1.7 Outline of the Paper

In Section 3 we introduce the concepts of a random automaton and of a random system as well as the equivalence of such systems. We also define monotone conditions and event sequences, the conditional equivalence of random systems, cascades of random systems, and the invocation of a random system by another random system. In Section 4 we define the indistinguishability of random systems, prove a few general results on indistinguishability, and discuss the framework for indistinguishability proofs based on conditional equivalence as well as consequences thereof. In Section 5 we apply the framework to the construction of quasi-random functions, and in Sections 6 and 7 to the analysis and security proofs of MAC's and of pseudo-random permutations, respectively.

The treatment is more general than necessary just for proving the results in Sections 5–7. Due to space limitations, many proofs are omitted (but see [13]).

2 Notation and Preliminaries

Random variables and concrete values they can take on are usually denoted by capital and small letters, respectively. For a set $\mathcal{S}$, an $\mathcal{S}$-*sequence* is an infinite (or possibly finite) sequence $s = s_1, s_2, \ldots$ of elements of $\mathcal{S}$. Prefixes of sequences (of values or random variables) are denoted by a superscript, e.g. s^k denotes the finite sequence $[s_1, s_2, \ldots, s_k]$. For a list L of random variables over the same alphabet, $\text{dist}(L)$ denotes the event that all values in L are distinct. Let $p_{\text{coll}}(n, k)$ denote the probability that k independent random variables with uniform distribution over a set of size n contain a collision, i.e., that they are not all distinct. Of course, $p_{\text{coll}}(n, k) = 1 - \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) < \frac{k^2}{2n}$.

In the context of this paper one considers different random experiments, and when analyzing probabilities it is crucial to be precise about which random experiment is considered. The random experiment is usually defined by one or several defining, usually independent, random variables. We will use these defining random variables as superscripts when denoting probabilities. For example, if $\mathbf{F}$ denotes the system under investigation and $\mathbf{D}$ the distinguisher, then $P^{\mathbf{DF}}$ denotes probabilities in the combined random experiment where $\mathbf{D}$ queries $\mathbf{F}$. In contrast $P^{\mathbf{F}}$ denotes probabilities in the simpler random experiment involving only the selection of $\mathbf{F}$, without even considering a distinguisher. If no superscript is used, the random experiment is clear from the context.

We use the following notation for probability distributions. If $\mathcal{A}$ and $\mathcal{B}$ are events and U and V are random variables with ranges $\mathcal{U}$ and $\mathcal{V}$, respectively, then $P_{U\mathcal{A}|V\mathcal{B}}$ denotes the corresponding conditional probability distribution, a function $\mathcal{U} \times \mathcal{V} \to \mathbf{R}^+$. Thus $P_{U\mathcal{A}|V\mathcal{B}}(u, v)$ for $u \in \mathcal{U}$ and $v \in \mathcal{V}$ is well-defined (except if $P_{V\mathcal{B}}(v) = 0$ in which case it is undefined). Note that $P_{\mathcal{A}}$ is equivalent to $P(\mathcal{A})$. For an event E, $\overline{E}$ denotes the complement of E. Equality of probability distributions means equality as functions, i.e., for all arguments. This extends to the equality of conditional probability distributions, even if one of them contains additional random variables in the conditioning set, meaning that equality holds for all possible values. For example, $P_{Y^i|X^k} = P_{Y^i|X^i}$ for $k > i$ means that for all x^k and y^i, $P_{Y^i|X^k}(y^i, x^k) = P_{Y^i|X^i}(y^i, x^i)$.

3 Random Systems and Monotone Event Sequences

3.1 Sources, Random Automata, and Random Systems

Definition 1. *An* $\mathcal{X}$-source $\mathbf{S}$ *is an infinite sequence* $\mathbf{S} = S_1, S_2, \ldots$ *of random variables* $S_i \in \mathcal{X}$*, characterized by the sequence* $P^{\mathbf{S}}_{S_i|S^{i-1}}$ *of conditional probability distributions. This also defines the distributions* $P^{\mathbf{S}}_{S^i} := \prod_{j=1}^{i} P^{\mathbf{S}}_{S_i|S^{i-1}}$.

In the following we consider systems which take inputs (or queries) $X_1, X_2, \ldots \in \mathcal{X}$ and generate, for each new input X_i, an output $Y_i \in \mathcal{Y}$. Such a system can be deterministic or probabilistic, and it can be stateless or contain internal memory. A stateless deterministic system is simply a function $\mathcal{X} \to \mathcal{Y}$.

Fig. 2. Left: An $(\mathcal{X},\mathcal{Y})$-random system $\mathbf{F}$ takes inputs $X_1, X_2, X_3, \ldots \in \mathcal{X}$ and outputs $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}$, where Y_i is generated after receiving input X_i. It is characterized by the sequence of conditional probability distributions $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}}$ for $i \geq 1$. Right: Random system $\mathbf{F}$ with a monotone event sequence $\mathcal{A} = A_0, A_1, A_2, \ldots$, denoted $\mathbf{F}^{\mathcal{A}}$.

Definition 2. A *random function* $\mathcal{X} \to \mathcal{Y}$ is a random variable which takes as values functions $\mathcal{X} \to \mathcal{Y}$. A deterministic system with state space Σ is called an $(\mathcal{X},\mathcal{Y})$-*automaton* and is described by an infinite sequence $f_1, f_2, \ldots$ of functions, with $f_i : \mathcal{X} \times \Sigma \to \mathcal{Y} \times \Sigma$, where $(Y_i, S_i) = f_i(X_i, S_{i-1})$, S_i is the state at time i, and an initial state S_0 is fixed. An $(\mathcal{X},\mathcal{Y})$-*random automaton* $\mathbf{F}$ is like an automaton but $f_i : \mathcal{X} \times \Sigma \times \mathcal{R} \to \mathcal{Y} \times \Sigma$ (where $\mathcal{R}$ is the space of the internal randomness), together with a probability distribution over $\mathcal{R} \times \Sigma$ specifying the internal randomness and the initial state.[2]

A large variety of constructions and definitions in the cryptographic literature can be interpreted as random functions, including pseudo-random functions, pseudo-random permutations, and MAC schemes. We consider the more general concept of a (stateful) random system because this is just as simple and because distinguishers can also be modeled as random systems.

The observable input-output behavior of a random automaton $\mathbf{F}$ is referred to as a random system. In the following we use the terms random automaton and random system interchangeably when no confusion is possible.

Definition 3. An $(\mathcal{X},\mathcal{Y})$-*random system* $\mathbf{F}$ is an infinite[3] sequence of conditional probability distributions $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}}$ for $i \geq 1$.[4] Two random automata $\mathbf{F}$ and $\mathbf{G}$ are *equivalent*, denoted $\mathbf{F} \equiv \mathbf{G}$, if they correspond to the same random system, i.e., if $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}} = P^{\mathbf{G}}_{Y_i|X^iY^{i-1}}$ for $i \geq 1$.[5]

The above definition is very general and captures systems that answer several types of queries (in which case the input set $\mathcal{X}$ is the union of the query sets) and for which the behavior depends on the index i. Note that a source can be interpreted as a special type of random system for which the input is ignored, i.e., the outputs are independent of the inputs. We will often assume that the input and output alphabets of a random system are clear from the context.

[2] $\mathbf{F}$ can also be considered as a random variable taking on as values $(\mathcal{X},\mathcal{Y})$-automata.
[3] Random systems with finite-length input sequences could also be defined.
[4] $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}}$ is a function $\mathcal{Y}\times\mathcal{X}^i\times\mathcal{Y}^{i-1}\to\mathbf{R}^+$ such that, for all $x^i\in\mathcal{X}^i$ and $y^{i-1}\in\mathcal{Y}^{i-1}$, $\sum_{y_i\in\mathcal{Y}} P^{\mathbf{F}}_{Y_i|X^iY^{i-1}}(y_i, x^i, y^{i-1}) = 1$.
[5] The distribution $P^{\mathbf{F}}_{Y^i|X^i} = \prod_{j=1}^{i} P^{\mathbf{F}}_{Y_j|X^jY^{j-1}}$ is also defined. $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}}(y_i, x^i, y^{i-1})$ can be undefined for values x^i and y^{i-1} with $P^{\mathbf{F}}_{Y^{i-1}|X^i}(y^{i-1}, x^i) = 0$.

Let us discuss a few special examples of random systems. Throughout, the symbols **B**, **R**, **P**, and **O** are used exclusively for the systems defined below.

Definition 4. An $(\mathcal{X},\mathcal{Y})$-*beacon* [19] $\mathbf{B}$ is a random system (actually a source) for which $Y_1, Y_2, \ldots$ are independent and uniformly distributed over $\mathcal{Y}$, independent of the inputs $X_1, X_2, \ldots$. 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 $\mathcal{X}$). A $\mathcal{Y}$*-random oracle* $\mathbf{O}$ is a random function with input alphabet $\mathcal{X} = \{0,1\}^*$ with $P^{\mathbf{O}}_{Y_i|X_i}(y,x) = 1/|\mathcal{Y}|$ for all $i \geq 1$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

3.2 Monotone Conditions and Event Sequences

For a given $(\mathcal{X},\mathcal{Y})$-random function or automaton $\mathbf{F}$, the evaluation of Y_i usually requires the evaluation of some internal random variables.[6] Consider the internal sequence of random variables $U_1, U_2, \ldots$. In the sequel it is very useful to consider an internal condition defined, for each i, after input X_i is entered. As a simple example, the condition could be $\mathrm{dist}(U^i)$, i.e., that $U_1, \ldots, U_i$ are all distinct.

Such an internal condition can be modeled as a binary random variable, say Z_i, indicating whether the condition is satisfied ($Z_i = 1$) or not ($Z_i = 0$) after input X_i has been given. If Z_i is taken as part of the ith output of $\mathbf{F}$, i.e., the ith output is the pair (Y_i, Z_i) instead of just Y_i, then this corresponds to a $(\mathcal{X}, \mathcal{Y}\times\{0,1\})$-random system.[7] One can also define several such conditions for $\mathbf{F}$, each corresponding to a binary random variable.

We will only consider *monotone* conditions, meaning that once it fails to be satisfied it remains so for all future inputs. For example, the condition $\mathrm{dist}(U^i)$ is obviously monotone. If U_i is a vector in some vector space, another monotone condition is that $U_1, \ldots, U_i$ are linearly independent.

For a random automaton $\mathbf{F}$ and a given monotone internal condition we will often be interested in $\mathbf{F}$'s behavior only as long as the condition is satisfied. For example, a URF behaves like a beacon as long as the inputs are distinct. We therefore consider the monotone sequence $\mathcal{A} = A_0, A_1, A_2, \ldots$ of events, where A_i is the event that the condition is satisfied (and $\overline{A_i}$ is the complementary event) and where A_0 is for convenience defined to be the certain event (cf. Fig. 2).

We will also consider two or more monotone conditions simultaneously. For two monotone event sequences (MES) $\mathcal{A}$ and $\mathcal{B}$ defined for $\mathbf{F}$, $\mathcal{A}\wedge\mathcal{B}$ denotes the MES defined by $(\mathcal{A}\wedge\mathcal{B})_i = A_i \wedge B_i$ for $i \geq 1$, and $\mathcal{A}\vee\mathcal{B}$ is defined analogously.

Definition 5. For MESs $\mathcal{A}$ and $\mathcal{C}$ defined for random automata $\mathbf{F}$ and $\mathbf{G}$, respectively, $\mathbf{F}$ *with* $\mathcal{A}$ *is equivalent to* $\mathbf{G}$ *with* $\mathcal{C}$, denoted $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{C}}$, if $P^{\mathbf{F}}_{Y_iA_i|X^iY^{i-1}A_{i-1}} = P^{\mathbf{G}}_{Y_iC_i|X^iY^{i-1}C_{i-1}}$ for $i \geq 1$.[8]

[6] For example, in the CBC-MAC U_i could be the input to the internal random function corresponding to the last block of the ith message.

[7] One can also think of an internal device (or genie) in $\mathbf{F}$ which beeps when the condition fails to be satisfied ($Z_i = 0$).

[8] Note that $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{C}}$ does not imply $\mathbf{F} \equiv \mathbf{G}$.

We refer to later sections for examples.

Definition 6. For a random system $\mathbf{F}$ with MES $\mathcal{A} = A_0, A_1, A_2, \ldots$, $\mathbf{F}$ *conditioned on* $\mathcal{A}$ *is equivalent to* $\mathbf{G}$, denoted $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}$, if $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}A_i} = P^{\mathbf{G}}_{Y_i|X^iY^{i-1}}$ for $i \geq 1$, for all arguments for which $P^{\mathbf{F}}_{Y_i|X^iY^{i-1}A_i}$ is defined. More generally, if $\mathcal{A}$ and $\mathcal{B}$ are defined for $\mathbf{F}$, then we write $\mathbf{F}^{\mathcal{B}}|\mathcal{A} \equiv \mathbf{G}^{\mathcal{C}}$ if $P^{\mathbf{G}}_{Y_iC_i|X^iY^{i-1}C_{i-1}} = P^{\mathbf{F}}_{Y_iB_i|X^iY^{i-1}B_{i-1}A_i}$ for $i \geq 1$.

Definition 7. One can *adjoin* an MES $\mathcal{C}$ to a random system $\mathbf{G}$ by defining C_i as depending probabilistically on X^i and Y^i, i.e., by a sequence of distributions $P^{\mathbf{G}}_{C_i|X^iY^iC_{i-1}}$. If an MES $\mathcal{C}$ is already defined for $\mathbf{G}$, then one can adjoin a further MES $\mathcal{D}$ according to a sequence $P^{\mathbf{G}}_{D_i|X^iY^iC_iD_{i-1}}$ of distributions.[9]

Lemma 1. (i) *If* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{C}}$*, then* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}|\mathcal{C}$[10] *(but not vice versa).*
(ii) *If* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}$*, then* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{C}}$ *for some MES* $\mathcal{C}$ *adjoined to* $\mathbf{G}$.
(iii) *More generally, if* $\mathbf{F}^{\mathcal{B}}|\mathcal{A} \equiv \mathbf{G}^{\mathcal{C}}$*, then* $\mathbf{F}^{\mathcal{A}\wedge\mathcal{B}} \equiv \mathbf{G}^{\mathcal{C}\wedge\mathcal{D}}$ *for some MES* $\mathcal{D}$.
(iv) *If* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}|\mathcal{C}$ *and* $P^{\mathbf{F}}_{A_i|X^iY^{i-1}A_{i-1}} \leq P^{\mathbf{G}}_{C_i|X^iY^{i-1}C_{i-1}}$ *for* $i \geq 1$ *(and for all* x^i *and* y^{i-1}*), then one can adjoin an MES* $\mathcal{D}$ *to* $\mathbf{G}$ *such that* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{C}\wedge\mathcal{D}}$.

Proof. Claim (i) is obvious. Claim (ii) follows from (iii), which follows by defining the MES $\mathcal{D}$ via $P^{\mathbf{G}}_{D_i|X^iY^iC_iD_{i-1}} = P^{\mathbf{F}}_{A_i|X^iY^{i-1}A_{i-1}B_{i-1}}$. The proof uses $P^{\mathbf{G}}_{Y_iC_i|X^iY^{i-1}C_{i-1}D_{i-1}} = P^{\mathbf{G}}_{Y_iC_i|X^iY^{i-1}C_{i-1}}$ (since $P^{\mathbf{G}}_{D_{i-1}|X^iY^iC_i} = P^{\mathbf{G}}_{D_{i-1}|X^iY^{i-1}C_{i-1}}$) and $P^{\mathbf{G}}_{Y_iC_i|X^iY^{i-1}C_{i-1}} = P^{\mathbf{F}}_{Y_iB_i|X^iY^{i-1}B_{i-1}A_i}$ (from $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{C}}$). The proof of (iv) is omitted. □

The following lemma states the trivial fact that given that all inputs are distinct, a random function behaves like a beacon. The proof is obvious.

Lemma 2. *Let* $\mathcal{C}$ *(*$\mathcal{D}$*) be an MES defined on the inputs (outputs) of a system.*
(i) $\mathbf{F}|\mathcal{C} \equiv \mathbf{F}$ *for every random system* $\mathbf{F}$.
(ii) *If* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$*, then* $\mathbf{F}^{\mathcal{A}\wedge\mathcal{C}} \equiv \mathbf{G}^{\mathcal{B}\wedge\mathcal{C}}$ *and* $\mathbf{F}^{\mathcal{A}\wedge\mathcal{D}} \equiv \mathbf{G}^{\mathcal{B}\wedge\mathcal{D}}$.
(iii) *If* C_i *implies that the first* i *inputs are distinct, then* $\mathbf{R}^{\mathcal{C}} \equiv \mathbf{B}^{\mathcal{C}}$ *and* $\mathbf{R}|\mathcal{C} \equiv \mathbf{B}$.

3.3 Cascades and Invocations of Random Systems

Definition 8. The *cascade* of an $(\mathcal{X}, \mathcal{Y})$-random system $\mathbf{F}$ and a $(\mathcal{Y}, \mathcal{Z})$-random system $\mathbf{G}$, denoted $\mathbf{FG}$, is the $(\mathcal{X}, \mathcal{Z})$-random system defined as applying $\mathbf{F}$ to the input sequence and $\mathbf{G}$ to the output of $\mathbf{F}$ (cf. Fig. 3). For MESs $\mathcal{A}$ and $\mathcal{B}$ defined for $\mathbf{F}$ and $\mathbf{G}$, respectively, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{A}\wedge\mathcal{B}$ are defined naturally for $\mathbf{FG}$.

---

[9] Informally, one connects an *independent* component, characterized by $P^{\mathbf{G}}_{D_i|X^iY^iC_iD_{i-1}}$, to the input and output of $\mathbf{G}$ and to the indicator random variable of $\mathcal{C}$ which generates the indicator random variable for $\mathcal{D}$.

[10] $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}|\mathcal{C}$ should be read as: there exists $\mathbf{H}$ such that $\mathbf{F}|\mathcal{A} \equiv \mathbf{H}$ and $\mathbf{G}|\mathcal{C} \equiv \mathbf{H}$.

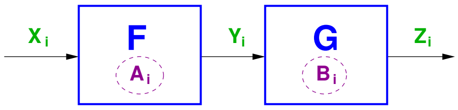

Fig. 3. The cascade of an $(\mathcal{X}, \mathcal{Y})$-random system $\mathbf{F}$ and a $(\mathcal{Y}, \mathcal{Z})$-random system $\mathbf{G}$, denoted $\mathbf{FG}$. For $\mathbf{F}^{\mathcal{A}}$ and $\mathbf{G}^{\mathcal{B}}$, $\mathbf{FG}^{\mathcal{A}\wedge\mathcal{B}}$ is defined naturally.

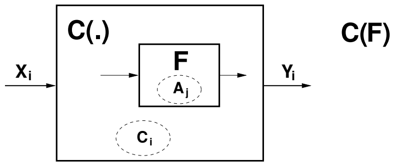

Fig. 4. A random system $\mathbf{C}(.)$ invoking an internal random system $\mathbf{F}$, then the combined random system is $\mathbf{C}(\mathbf{F})$.

Lemma 3. (i) *For any source* $\mathbf{S}$ *and any (compatible)* $\mathbf{E}$ *we have* $\mathbf{ES} \equiv \mathbf{S}$.
(ii) *If* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}$*, then* $\mathbf{EF}|\mathcal{A} \equiv \mathbf{EG}$ *for any compatible* $\mathbf{E}$.

We denote by $\mathbf{C}(.)$ a random system that *invokes* an internal random system (with specified input and output alphabets). If the internal system is $\mathbf{F}$, then the combined random system is $\mathbf{C}(\mathbf{F})$ (cf. Fig. 4). For the evaluation of the output Y_i for a given input X_i to $\mathbf{C}(\mathbf{F})$, $\mathbf{F}$ is called zero, one, or several times, where the inputs to $\mathbf{F}$ and even the number of such inputs may depend on the state of $\mathbf{C}(.)$, hence on $X_1, \ldots, X_i$.[11]

An MES, say $\mathcal{C} = C_0, C_1, C_2, \ldots$, can be defined also for such a system $\mathbf{C}(.)$. If $\mathcal{A}$ is an MES defined for the invoked $\mathbf{F}$, one can associate a natural corresponding MES $\tilde{\mathcal{A}} = \tilde{A}_0, \tilde{A}_1, \tilde{A}_2, \ldots$ with $\mathbf{C}(\mathbf{F})$, where $\tilde{A}_i$ is the event that the A-event occurs for $\mathbf{F}$ up to the evaluation of the ith input to $\mathbf{C}(\mathbf{F})$. If $\mathbf{F}$ is called t times for each input to $\mathbf{C}(\mathbf{F})$, then $\tilde{A}_i = A_{ti}$. Let $m_{\mathbf{C}(.)}(k)$ be the maximal number of evaluations of any internal system $\mathbf{F}$ for any sequence of k inputs to $\mathbf{C}(\mathbf{F})$, if it is defined.

The following lemma states the simple fact that by replacing a random system by an equivalent random system, the overall behavior of a system does not change. Let $\mathbf{C}(.)$ be any random system and let $\mathbf{F}$ and $\mathbf{G}$ be input/output compatible with $\mathbf{C}(.)$. Let $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ be defined for $\mathbf{C}(.)$, $\mathbf{F}$ and $\mathbf{G}$, respectively.

Lemma 4. (i) *If* $\mathbf{F} \equiv \mathbf{G}$*, then* $\mathbf{C}(\mathbf{F}) \equiv \mathbf{C}(\mathbf{G})$ *and* $\mathbf{C}(\mathbf{F})^{\mathcal{C}} \equiv \mathbf{C}(\mathbf{G})^{\mathcal{C}}$.
(ii) *If* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$*, then* $\mathbf{C}(\mathbf{F})^{\tilde{\mathcal{A}}} \equiv \mathbf{C}(\mathbf{G})^{\tilde{\mathcal{B}}}$ *and* $\mathbf{C}(\mathbf{F})^{\tilde{\mathcal{A}}\wedge\mathcal{C}} \equiv \mathbf{C}(\mathbf{G})^{\tilde{\mathcal{B}}\wedge\mathcal{C}}$.[12]

[11] Formally, $\mathbf{C}(.)$ is not a random system without specifying an argument $\mathbf{F}$.
[12] Note, however, that $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}$ does not imply $\mathbf{C}(\mathbf{F})|\tilde{\mathcal{A}} \equiv \mathbf{C}(\mathbf{G})$.

Fig. 5. Distinguishing two $(\mathcal{X}, \mathcal{Y})$-random systems $\mathbf{F}$ and $\mathbf{G}$ by means of a distinguisher $\mathbf{D}$. The figure shows the two random experiments under consideration.

Proof. The lemma follows directly from the fact that the probability distribution of all random variables and events occurring in $\mathbf{C}(.)$, when including $\mathcal{A} = A_0, A_1, A_2, \ldots$ (or $\mathcal{B} = B_0, B_1, B_2, \ldots$), is the product of conditional distributions defined by the random system and by $\mathbf{C}(.)$. The conditional distributions defined by $\mathbf{C}(.)$ are trivially identical and those defined by $\mathbf{F}$ (or $\mathbf{G}$) are identical in both cases because of $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$. □

4 Indistinguishability Proofs for Random Systems

4.1 Distinguishers for Random Systems

We consider the problem of distinguishing two $(\mathcal{X}, \mathcal{Y})$-random systems $\mathbf{F}$ and $\mathbf{G}$ by means of a computationally unbounded, possibly probabilistic adaptive distinguisher algorithm (or simply distinguisher) $\mathbf{D}$ asking at most k queries, for some k (cf. Fig. 5). The distinguisher generates X_1 as an input to $\mathbf{F}$ (or $\mathbf{G}$), receives the output Y_1, then generates X_2, receives Y_2, etc. Finally, after receiving Y_k, it outputs a binary decision bit. More formally:

Definition 9. A *distinguisher for* $(\mathcal{X}, \mathcal{Y})$*-random systems* is a $(\mathcal{Y}, \mathcal{X})$-random system $\mathbf{D}$ together with an initial value $X_1 \in \mathcal{X}$ which outputs a binary decision value after some specified number k of queries to the system. Without loss of generality we can assume that $\mathbf{D}$ outputs a binary value after every query and that this sequence is monotone (0 never followed by 1), i.e., we can define the MES $\mathcal{E} = E_0, E_1, E_2, \ldots$ where E_i is the event that $\mathbf{D}$ outputs 1 after the i-th query. Application of $\mathbf{D}$ to a random system $\mathbf{F}$ (cf. Fig. 5) means that X_1 is the first input to $\mathbf{F}$, the i-th input and output of $\mathbf{D}$ are Y_i and $\tilde{X}_i$, respectively, and $X_i := \tilde{X}_{i-1}$ for $i \geq 2$ is the i-th input to $\mathbf{F}$.

Definition 10. The maximal advantage, of any distinguisher issuing k queries, for distinguishing $\mathbf{F}$ and $\mathbf{G}$, is

$$\Delta_k(\mathbf{F}, \mathbf{G}) \ := \ \max_{\mathbf{D}} \left| P^{\mathbf{DF}}(E_k) - P^{\mathbf{DG}}(E_k) \right| .$$

We summarize a few simple facts used in many security proofs. The inequalities hold for any compatible random automata or random systems.

Lemma 5. (i) $\;\Delta_k(\mathbf{F}, \mathbf{H}) \leq \Delta_k(\mathbf{F}, \mathbf{G}) + \Delta_k(\mathbf{G}, \mathbf{H})$.
(ii) $\;\Delta_k(\mathbf{C}(\mathbf{F}), \mathbf{C}(\mathbf{G})) \leq \Delta_{k'}(\mathbf{F}, \mathbf{G})$*, where* $k' = m_{\mathbf{C}(.)}(k)$.

(iii) $\Delta_k(\mathbf{FF}', \mathbf{GG}') \leq \Delta_k(\mathbf{F}, \mathbf{G}) + \Delta_k(\mathbf{F}', \mathbf{G}')$.
(iv) (Informal.) *If* $\Delta_k(\mathbf{F}, \mathbf{G})$ *is negligible in* k *and* $\mathbf{G}$ *is computationally indistinguishable from* $\mathbf{H}$*, then* $\mathbf{F}$ *is also computationally indistinguishable from* $\mathbf{H}$.

Proof. (i) follows by a simple application of the triangle inequality $|c-a| \leq |b-a| + |c-b|$ for any real a, b, and c, applied to $a = P^{\mathbf{DF}}(E_k)$, $b = P^{\mathbf{DG}}(E_k)$, and $c = P^{\mathbf{DH}}(E_k)$ for any distinguisher $\mathbf{D}$. To prove (ii), suppose for the sake of contradiction that there exists a distinguisher for $\mathbf{C}(\mathbf{F})$ and $\mathbf{C}(\mathbf{G})$, asking at most k queries, with advantage greater than $\Delta_{k'}(\mathbf{F}, \mathbf{G})$. By simulating $\mathbf{C}(.)$ one can construct a distinguisher for $\mathbf{F}$ and $\mathbf{G}$ with the same advantage, asking at most k' queries. This is a contradiction. Now we prove (iii). From (ii) we have $\Delta_k(\mathbf{FF}', \mathbf{GF}') \leq \Delta_k(\mathbf{F}, \mathbf{G})$ and $\Delta_k(\mathbf{GF}', \mathbf{GG}') \leq \Delta_k(\mathbf{F}', \mathbf{G}')$. Now we apply (i) to the random systems $\mathbf{FF}'$, $\mathbf{GF}'$, and $\mathbf{GG}'$. The proof of (iv) is omitted. □

It is easy to see that the described view of a distinguisher $\mathbf{D}$ is equivalent to an alternative view where $\mathbf{D}$ is given access to a blackbox containing $\mathbf{F}$ or $\mathbf{G}$ with probability $\frac{1}{2}$ each, where $\mathbf{D}$ must guess which of the two is the case. The best success probability with k queries is $\frac{1}{2} + \frac{1}{2}\Delta_k(\mathbf{F}, \mathbf{G})$.

4.2 Indistinguishability Proofs Based on Conditional Equivalence

In this section we prove that if $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}$ for some MES $\mathcal{A}$ (or if $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$), then a distinguisher $\mathbf{D}$ for distinguishing $\mathbf{F}$ from $\mathbf{G}$ with k queries (according to the view described above) *must* provoke the event $\overline{A_k}$ in $\mathbf{F}$ in order to have a non-zero advantage. Informally this could be proved by assuming a genie sitting inside $\mathbf{F}$ and beeping when it sees that $\overline{A_i}$ occurs for some i. The genie's help can only help since it could always be ignored, and given the genie's help, the optimal strategy would be to guess "$\mathbf{F}$" if the genie beeps and to flip a fair coin between $\mathbf{F}$ and $\mathbf{G}$ otherwise. Therefore we consider distinguishers $\mathbf{D}$ that try to provoke the event $\overline{A_k}$.

Definition 11. For a random system $\mathbf{F}$ with MES $\mathcal{A}$, let

$$\nu(\mathbf{F}, \overline{A_k}) \ := \ \max_{\mathbf{D}} P^{\mathbf{DF}}(\overline{A_k})$$

be the maximal probability, for any adaptive strategy $\mathbf{D}$, of provoking $\overline{A_k}$ in $\mathbf{F}$. Moreover, let

$$\mu(\mathbf{F}, \overline{A_k}) \ := \ \max_{x^k} P^{\mathbf{F}}_{\overline{A_k}|X^k}(x^k)$$

be the maximal probability of $\overline{A_k}$ for non-adaptive algorithms querying $\mathbf{F}$.

Lemma 6. **(i)** $\mu(\mathbf{F}, \overline{A_k}) \leq \nu(\mathbf{F}, \overline{A_k})$.
(ii) *If* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$*, then* $\nu(\mathbf{F}, \overline{A_k}) = \nu(\mathbf{G}, \overline{B_k})$.
(iii) $\nu(\mathbf{F}, \overline{A_k} \vee \overline{B_k}) \leq \nu(\mathbf{F}, \overline{A_k}) + \nu(\mathbf{F}, \overline{B_k})$ *if* $\mathcal{A}$ *and* $\mathcal{B}$ *are defined for* $\mathbf{F}$.
(iv) *For any system* $\mathbf{C}(.)$ *with MES* $\mathcal{C}$*, invoking* $\mathbf{F}$*,* $\nu(\mathbf{C}(\mathbf{F}), \overline{C_k}) \leq \nu(\mathbf{C}(.), \overline{C_k})$[13]

[13] $\nu(\mathbf{C}(.), \overline{C_k})$ is defined as the maximal probability of provoking event $\overline{C_k}$ in $\mathbf{C}(.)$ for algorithms with full control of the input to $\mathbf{C}(.)$ and the internal interface.

and $\nu(\mathbf{C}(\mathbf{F}), \widetilde{\overline{A_k}}) \leq \nu(\mathbf{F}, \overline{A_{k'}})$, *where* $k' = m_{\mathbf{C}(.)}(k)$.
(v) *If* $\mathcal{A}$ *is defined on the inputs of* $\mathbf{F}$, *then* $\mu(\mathbf{EF}, \overline{A_k}) = \mu(\mathbf{E}, \overline{A_k})$ *for any* $\mathbf{E}$.

Proof. (i) holds because the set of adaptive strategies includes the non-adaptive ones. Claim (ii) follows from $\nu(\mathbf{F}, \overline{A_k}) = 1 - \nu(\mathbf{F}, A_k)$ and $\nu(\mathbf{G}, \overline{B_k}) = 1 - \nu(\mathbf{G}, B_k)$, using $\nu(\mathbf{F}, A_k) = \nu(\mathbf{G}, B_k)$ which follows from Lemma 4. Claim (iii) is a simple application of the union bound together with the fact that if different systems $\mathbf{D}$ can be used to provoke $\overline{A_k}$ and $\overline{B_k}$, this can only improve the success probability. Claim (iv) follows from the fact that $\mathbf{C}(.)$ can be used as a possible algorithm for provoking $\overline{A_k}$ in $\mathbf{F}$, and similarly $\mathbf{F}$ can be used as the random system in an algorithm for provoking $\overline{B_k}$ in $\mathbf{C}(.)$. Claim (v) is trivial. □

Lemma 7. *If* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$, *then for any (compatible) distinguisher* $\mathbf{D}$ *and any event* E_k *defined in* $\mathbf{D}$ *after* k *queries,*

$$\left| P^{\mathbf{DF}}(E_k) - P^{\mathbf{DG}}(E_k) \right| \leq P^{\mathbf{DF}}(\overline{A_k}) = P^{\mathbf{DG}}(\overline{B_k}).$$

Proof. Lemma 4 gives $P^{\mathbf{DF}}(E_k \wedge A_k) = P^{\mathbf{DG}}(E_k \wedge B_k) \leq P^{\mathbf{DG}}(E_k)$. Thus

$$P^{\mathbf{DF}}(E_k) = P^{\mathbf{DF}}(E_k \wedge A_k) + P^{\mathbf{DF}}(E_k \wedge \overline{A_k}) \leq P^{\mathbf{DG}}(E_k) + P^{\mathbf{DF}}(\overline{A_k}).$$

$P^{\mathbf{DG}}(E_k) \leq P^{\mathbf{DF}}(E_k) + P^{\mathbf{DG}}(\overline{B_k})$ follows by symmetry, and $P^{\mathbf{DF}}(\overline{A_k}) = P^{\mathbf{DG}}(\overline{B_k})$ follows from Lemma 4. □

Theorem 1. (i) *If* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ *or* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}$, *then* $\Delta_k(\mathbf{F}, \mathbf{G}) \leq \nu(\mathbf{F}, \overline{A_k})$.
(ii) *If* $\mathbf{F}^{\mathcal{B}}|\mathcal{A} \equiv \mathbf{G}^{\mathcal{C}}$, *then* $\Delta_k(\mathbf{F}, \mathbf{G}) \leq \nu(\mathbf{F}, \overline{A_k} \vee \overline{B_k}) \leq \nu(\mathbf{F}, \overline{A_k}) + \nu(\mathbf{G}, \overline{C_k})$.
(iii) *If* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}|\mathcal{C}$ *and* $P^{\mathbf{F}}_{A_i|X^iY^{i-1}A_{i-1}} \leq P^{\mathbf{G}}_{C_i|X^iY^{i-1}C_{i-1}}$ *for* $i \geq 1$, *then* $\Delta_k(\mathbf{F}, \mathbf{G}) \leq \nu(\mathbf{F}, \overline{A_k})$.

Proof. The first claim of (i) is a special case of Lemma 7, where $\mathbf{D}$ is the distinguisher with MES $\mathcal{E}$. The second claim of (i) is a special case of (ii), which is proved as follows. According to Lemma 1 (iii) we have $\mathbf{F}^{\mathcal{A} \wedge \mathcal{B}} \equiv \mathbf{G}^{\mathcal{C} \wedge \mathcal{D}}$ for some MES $\mathcal{D}$ defined for $\mathbf{G}$. Thus we can apply (i). The last inequality of (ii) follows because for any $\mathbf{D}$, $P^{\mathbf{DF}}(\overline{A_k} \vee \overline{B_k}) \leq P^{\mathbf{DF}}(\overline{A_k}) + P^{\mathbf{DF}}(\overline{B_k}|A_k)$, and since if $P^{\mathbf{DF}}(\overline{A_k})$ and $P^{\mathbf{DF}}(\overline{B_k}|A_k)$ can be maximized separately by choices of $\mathbf{D}$, this is an upper bound on $\max_{\mathbf{D}} P^{\mathbf{DF}}(\overline{A_k} \vee \overline{B_k})$. Moreover, $\max_{\mathbf{D}} P^{\mathbf{DF}}(\overline{B_k}|A_k) = \max_{\mathbf{D}} P^{\mathbf{DG}}(\overline{C_k}) = \nu(\mathbf{G}, \overline{C_k})$. To prove (iii), adjoin the MES $\mathcal{D}$ to $\mathbf{G}$ as in Lemma 1 (iv) and apply (i) of this theorem. □

4.3 Adaptive Versus Non-Adaptive Strategies

It is generally substantially easier to analyze non-adaptive as opposed to adaptive strategies, e.g. for distinguishing two random systems. The following theorem states simple and easily checkable conditions for a random system $\mathbf{F}$ with MES $\mathcal{A}$ which implies that no adaptive strategy for provoking $\overline{A_k}$ is better than the best non-adaptive strategy. The optimal strategy hence selects (one of) the fixed input sequence(s) x^k that minimizes $P^{\mathbf{F}}_{A_k|X^k}(x^k)$ (or equivalently, maximizes $P^{\mathbf{F}}_{\overline{A_k}|X^k}(x^k)$). Hence the system $\mathbf{D}$ (over choices of which the definition of $\nu(\mathbf{F}, \overline{A_k})$ maximizes) can be eliminated from the analysis.

Theorem 2. *If a random system* $\mathbf{F}$ *with MES* $\mathcal{A}$ *satisfies*

$$P^{\mathbf{F}}_{A_i|X^iY^{i-1}A_{i-1}} = P^{\mathbf{F}}_{A_i|X^iA_{i-1}} \tag{1}$$

for $i \geq 1$*, which holds if*

$$P^{\mathbf{F}}_{Y^i|X^iA_i} = P^{\mathbf{G}}_{Y^i|X^i} \tag{2}$$

for $i \geq 1$*, for some system* $\mathbf{G}$ *(actually,* $\mathbf{G} \equiv \mathbf{F}|\mathcal{A}$*), then* $\nu(\mathbf{F}, \overline{A_k}) = \mu(\mathbf{F}, \overline{A_k})$*.*

Corollary 1. **(i)** *If* $\mathcal{A}$ *is defined on the inputs of* $\mathbf{F}$*, then* $\mathbf{F}$ *satisfies (1).*
(ii) *If* $\mathbf{F}$ *with* $\mathcal{A}$ *satisfy (1), then so does* $\mathbf{FG}$ *with* $\mathcal{A}$ *for any (compatible)* $\mathbf{G}$*.*
(iii) *If* $\nu(\mathbf{F}, \overline{A_k}) = \mu(\mathbf{F}, \overline{A_k})$*, then* $\nu(\mathbf{FG}, \overline{A_k}) = \mu(\mathbf{F}, \overline{A_k})$ *for any* $\mathbf{G}$*.*
(iv) *If* $\mathcal{A}$ *is defined on the inputs of* $\mathbf{F}$ *and* $\mathbf{F}|\mathcal{A} \equiv \mathbf{U}$ *for a source* $\mathbf{U}$*, then* $\nu(\mathbf{EF}, \overline{A_k}) = \mu(\mathbf{E}, \overline{A_k})$ *for any* $\mathbf{E}$*.*
(v) *If* A_i *(*B_i*) is defined on the inputs (outputs) of* $\mathbf{F}$ *and* $\mathbf{F}^{\mathcal{B}}|\mathcal{A} \equiv \mathbf{U}^{\mathcal{B}}$ *for a source* $\mathbf{U}$*, then* $\nu(\mathbf{EF}, \overline{A_k} \vee \overline{B_k}) \leq \mu(\mathbf{E}, \overline{A_k}) + \mu(\mathbf{U}, \overline{B_k})$ *for any* $\mathbf{E}$*.*
(vi) *If* $\mathcal{A}$ *is defined on the inputs of* $\mathbf{F}$ *and* $\mathbf{F}|\mathcal{A} \equiv \mathbf{B}$*, then for any random system* $\mathbf{C}(.)$ *such that* $\mathbf{C}(\mathbf{B}) \equiv \mathbf{B}$*,* $\nu(\mathbf{C}(\mathbf{F}), \overline{A_k}) = \mu(\mathbf{C}(\mathbf{F}), \overline{A_k})$*.*

4.4 Exploiting Independent Events

Consider a random system $\mathbf{C}(.,.)$ invoking two independent random systems $\mathbf{F}$ and $\mathbf{G}$ with MESs $\mathcal{A}$ and $\mathcal{B}$, respectively. For each input to $\mathbf{C}(\mathbf{F}, \mathbf{G})$, $\mathbf{F}$ and $\mathbf{G}$ can be called several times. For a given k, let k' and k'' be the maximal number of invocations of $\mathbf{F}$ and $\mathbf{G}$, respectively, for any input sequence to $\mathbf{C}(\mathbf{F}, \mathbf{G})$ of length k.

Theorem 3. *If* $\mathbf{C}(\mathbf{F}, \mathbf{G})|(\tilde{\mathcal{A}} \vee \tilde{\mathcal{B}}) \equiv \mathbf{H}$*, then*

$$\Delta_k(\mathbf{C}(\mathbf{F}, \mathbf{G}), \mathbf{H}) \leq \nu(\mathbf{F}, \overline{\tilde{A}_{k'}}) \cdot \nu(\mathbf{G}, \overline{\tilde{B}_{k''}}).$$

Proof. We have

$$\begin{aligned}
\Delta_k(\mathbf{C}(\mathbf{F}, \mathbf{G}), \mathbf{H}) \leq \nu(\mathbf{C}(\mathbf{F}, \mathbf{G}), \overline{\tilde{A}_{k'}} \wedge \overline{\tilde{B}_{k''}}) &= \max_{\mathbf{D}} P^{\mathbf{DCFG}}(\overline{\tilde{A}_{k'}} \wedge \overline{\tilde{B}_{k''}}) \\
&= \max_{\mathbf{D}} \left(P^{\mathbf{DCFG}}(\overline{\tilde{A}_{k'}}) \cdot P^{\mathbf{DCFG}}(\overline{\tilde{B}_{k''}}|\overline{\tilde{A}_{k'}}) \right) \\
&\leq \underbrace{\max_{\mathbf{D}} P^{\mathbf{DCFG}}(\overline{\tilde{A}_{k'}})}_{=\nu(\mathbf{C}(\mathbf{F},\mathbf{G}),\overline{\tilde{A}_{k'}}) \leq \nu(\mathbf{F},\overline{\tilde{A}_{k'}})} \cdot \underbrace{\max_{\mathbf{D}} P^{\mathbf{DCFG}}(\overline{\tilde{B}_{k''}}|\overline{\tilde{A}_{k'}})}_{\leq \nu(\mathbf{G},\overline{\tilde{B}_{k''}})}.
\end{aligned}$$

The last inequality holds because in the expression on the last line the two maximizations over choices of $\mathbf{D}$ are independent, as opposed to the previous line. We have $\nu(\mathbf{C}(\mathbf{F}, \mathbf{G}), \overline{\tilde{A}_{k'}}) \leq \nu(\mathbf{F}, \overline{\tilde{A}_{k'}})$ by Lemma 6 (iv) and $\max_{\mathbf{D}} P^{\mathbf{DCFG}}(\overline{\tilde{B}_{k''}}|\overline{\tilde{A}_{k'}}) \leq \nu(\mathbf{G}, \overline{\tilde{B}_{k''}})$ because for every particular choices for $\mathbf{D}$, $\mathbf{C}$, and $\mathbf{F}$, the probability of $\overline{\tilde{B}_{k''}}$ is at most $\nu(\mathbf{G}, \overline{\tilde{B}_{k''}})$, whether or not $\overline{\tilde{A}_{k'}}$ occurs for these choices. Thus the bound on $\nu(\mathbf{G}, \overline{\tilde{B}_{k''}})$ also holds on average. □

Corollary 2. *Let* $\mathbf{F}$ *with MES* $\mathcal{A}$ *and* $\mathbf{G}$ *with MES* $\mathcal{B}$ *be random permutations such that* $\mathbf{F}|\mathcal{A} \equiv \mathbf{P}$ *and* $\mathbf{G}|\mathcal{B} \equiv \mathbf{P}$*. Then* $\Delta_k(\mathbf{FG}, \mathbf{P}) \leq \nu(\mathbf{F}, \overline{\tilde{A_{k'}}}) \cdot \nu(\mathbf{G}, \overline{\tilde{B_{k''}}})$.

Proof. We have $\mathbf{FG}|(\mathcal{A} \vee \mathcal{B}) \equiv \mathbf{P}$, hence Theorem 3 can be applied.[14] □

For two $(\mathcal{X}, \mathcal{Y})$-random automata $\mathbf{F}$ and $\mathbf{G}$ and a group operation $\star$ on $\mathcal{Y}$, let $\mathbf{F} \star \mathbf{G}$ denote the random automaton obtained by using $\mathbf{F}$ and $\mathbf{G}$ in parallel (with the same input) and combining the two outputs using $\star$.

Corollary 3. *If* $\mathbf{F}|\mathcal{A} \equiv \mathbf{G}|\mathcal{B} \equiv \mathbf{R}$*, then* $\Delta_k(\mathbf{F} \star \mathbf{G}, \mathbf{R}) \leq \nu(\mathbf{F}, \overline{A_k}) \cdot \nu(\mathbf{G}, \overline{B_k})$.

Proof. We have $(\mathbf{F} \star \mathbf{G})|(\mathcal{A} \vee \mathcal{B}) \equiv \mathbf{R}$, hence Theorem 3 can be applied. □

5 Applications to Quasi-Random Functions

5.1 Quasi-Random Functions

Definition 12. For a function $d : \mathbf{N} \to \mathbf{R}^+$, a random function or random system $\mathbf{F}$ is called a $d(k)$-*quasi-random function* ($d(k)$-QRF for short) if $\Delta_k(\mathbf{F}, \mathbf{R}) \leq d(k)$ for $k \geq 1$. Quasi-random permutations, beacons and oracles are defined analogously, replacing $\mathbf{R}$ by $\mathbf{P}$, $\mathbf{B}$, and $\mathbf{O}$, respectively.

By concatenating, for any w, 2^w outputs of a $d(k)$-QRF $\{0,1\}^l \to \{0,1\}^m$ one obtains a $\tilde{d}(k)$-QRF $\{0,1\}^{l-w} \to \{0,1\}^{2^w m}$ for $\tilde{d}(k) = d(2^w k)$, thus increasing the output size by a factor 2^w at the expense of reducing the input size by w bits.

The problem considered in this section is to *expand* the input size substantially at the sole expense of increasing $d(k)$ moderately, i.e., to expand a given supply of random bits into a much larger supply of apparently random bits.

This general problem is important because the core of a cryptographic system based on a PRF corresponds to the construction of a quasi-random system of the same type from a URF $\mathbf{R}$. In any such construction, $\mathbf{R}$ can be replaced by a QRF, possibly constructed recursively from smaller QRF's, where at the lowest level the randomness is replaced by the PRF. This can for instance be used to avoid the birthday problem when collisions are a security issue (see below).

For any $d(k)$-QRF $\mathbf{G} : \{0,1\}^L \to \{0,1\}^M$ constructed from a URF $\mathbf{R} : \{0,1\}^l \to \{0,1\}^m$ it is obvious that $d(k)$ cannot be negligible for $k > 2^l m/M$, i.e., when the internal randomness is exhausted. One could achieve $d(k) = 0$ for up to $k \approx 2^l m/M$ by defining $\mathbf{G}$ as the evaluation of a polynomial whose coefficients are taken from the function table of $\mathbf{R}$, but this construction would be exponentially inefficient since the entire table of $\mathbf{R}$ must be read for each evaluation of $\mathbf{G}$. Efficiency, i.e., the number of evaluations of $\mathbf{R}$ required for one evaluation of $\mathbf{G}$, is an important parameter of a construction. There is a trade-off between the efficiency and the degree $d(k)$ of indistinguishability.

[14] The corollary also follows from Vaudenay's nice proof [22] (stated in our terminology) that $\Delta_k(\mathbf{FG}, \mathbf{P}) \leq \Delta_k(\mathbf{F}, \mathbf{P}) \cdot \Delta_k(\mathbf{G}, \mathbf{P})$ for two random permutations $\mathbf{F}$ and $\mathbf{G}$.

5.2 An Efficient Construction of a Quasi-Random Function

We now propose the construction of an efficient QRF $\mathbf{C}(\mathbf{F}) : \{0,1\}^L \to \{0,1\}^m$ from a QRF $\mathbf{F} : \{0,1\}^l \to \{0,1\}^m$, for $L \gg l$. The basic idea for the definition of $\mathbf{C}(.)$ is to map an argument to $\mathbf{C}(.)$ to a list of t arguments for $\mathbf{F}$ and to XOR the corresponding values of $\mathbf{F}$. In fact, we can (but need not) use the convention that if a list contains a value more than once, these values are ignored, resulting in fewer than t values being XORed.

One can associate, in a natural manner, with each such set of t values a characteristic vector, with at most t 1-entries, in the vector space $\{0,1\}^{2^l}$. The described XORing operation corresponds to computing the scalar product of the characteristic vector with the function table of $\mathbf{F}$ (interpreted as a vector in $(\{0,1\}^m)2^l$).

Hence Lemma 11 in the Appendix implies that, given the event that these k vectors (for the k arguments to $\mathbf{C}(.)$) are linearly independent, the construction is equivalent to a URF (and also a beacon). Therefore Theorem 1 (i) can be applied.

It only remains to find a mapping $\mathbf{H} : \{0,1\}^L \to S$, where S is the subset of the vector space $\{0,1\}^{2^l}$ consisting of the vectors of weight at most t. The internal randomness of $\mathbf{H}$ can actually be taken from the function table of $\mathbf{F}$ (say for the z highest values, where z is an appropriate small number). For this to be secure, the mapping $\mathbf{H}$ must be restricted slightly to generate vectors with no 1-entry in the last z coordinates.

Lemma 12 in the Appendix shows that $\mathbf{H}$ can be implemented by using a $2t$-wise random function $\mathbf{E} : \{0,1\}^L \times \{1,\ldots,t\} \to \{0,\ldots,2^l - z - 1\}$. For an argument $x \in \{0,1\}^L$ of $\mathbf{H}$, $\mathbf{E}(x,i)$ for $1 \le i \le t$ is evaluated and the corresponding characteristic vector is formed.[15] Note that the z unit vectors with 1-entries in one of the top z positions must also be taken into account in Lemma 12, but they are of course linearly independent of the k vectors discussed above.

Hence we have outlined the proof of the following theorem.

Theorem 4. *For a $d(k)$-QRF $\mathbf{F}$, $\mathbf{C}(\mathbf{F})$ is a $\tilde{d}(k)$-QRF for $\tilde{d}(k) = k\left(\frac{kt}{2^l}\right)^t + d(tk+z)$.*

The term $k(kt/2^l)^t$ is very small, even for $k \gg 2^{l/2}$ for which collisions among random values in the input space of $\mathbf{F}$ would be very probable. This was called "security beyond the birthday barrier" in [1].[16] Already for moderate values of t, the described construction achieves a negligible $\tilde{d}(k)$ for $k \approx 2^{lt/(t+1)}$, i.e., far beyond the birthday barrier.

The above construction ideas apply in other contexts as well, for instance the use of some values of a PRF as the key of another component in a manner that

---

15 Such a function $\mathbf{E}$ can for instance be obtained by evaluation of a polynomial of degree $2t$ over an appropriate finite field of size at least $t2^L$.

16 This fact was pointed out already in [12], Theorem 2, where the basic idea of XORing several values of a function to go beyond the birthday bound was proposed.

does not compromise security. Note that the security of the XOR-MAC [3] and of other constructions based on linearly independent inputs (e.g. [1]) follow directly from Lemma 11 as well as a (non-adaptive) analysis of the linear independence event. For the XOR-MAC the analysis of this event is trivial.

6 Applications to MAC's

A secure MAC-scheme is a PRF $\mathcal{M} \to \{0,1\}^l$ for $\mathcal{M} = \cup_{i=1}^{L}\{0,1\}^i$ for some maximal message length L and an appropriate security parameter l. If $L = \infty$, then this corresponds to a pseudo-random oracle.

A very natural construction originating in [23] and used in many later papers (e.g. see [5, 20] and the discussion and references therein) is to apply an ϵ-almost universal hash function[17] $\mathbf{U} : \mathcal{M} \to \mathcal{X}$ for some $\mathcal{X}$ to the message and to apply a PRF $\mathbf{F} : \mathcal{X} \to \{0,1\}^l$ to the result. Such a scheme has two keys, those of $\mathbf{U}$ and $\mathbf{F}$, but in fact the $\mathbf{U}$-key can be obtained by evaluating $\mathbf{F}$ for an appropriate number z of fixed arguments, as follows easily from our framework. More precisely, $\mathbf{U}(.)$ is a random system[18] invoking $\mathbf{F}$ some z times to set up the key of $\mathbf{U}$ and then applies it to the input.[19] Of course, the key can be cached so that only one evaluation of $\mathbf{F}$ is needed for each input.

The security proof of such a scheme is trivial in our framework. The following theorem implies that $\mathbf{U}(\mathbf{F})$ is a computationally secure MAC for any PRF $\mathbf{F}$.

Theorem 5. *For a $d(k)$-QRF $\mathbf{F}$, $\mathbf{U}(\mathbf{F})$ is a $\tilde{d}(k)$-QRO for $\tilde{d}(k) = \epsilon(k+z)^2/2 + d(k+z)$.*

Proof. Define A_i as the event that all inputs to $\mathbf{F}$ are distinct, including the z fixed values needed for the key setup for $\mathbf{U}$. Lemma 5 (i) implies $\Delta_k(\mathbf{U}(\mathbf{F}), \mathbf{R}) \leq \Delta_k(\mathbf{U}(\mathbf{F}), \mathbf{U}(\mathbf{R})) + \Delta_k(\mathbf{U}(\mathbf{R}), \mathbf{R})$. Lemma 5 (ii) implies $\Delta_k(\mathbf{U}(\mathbf{F}), \mathbf{U}(\mathbf{R})) \leq d(k+z)$. Moreover, $\mathbf{U}(\mathbf{R})|\mathcal{A} \equiv \mathbf{R}$ and hence, using Theorem 1 (i), $\Delta_k(\mathbf{U}(\mathbf{R}), \mathbf{R}) \leq \nu(\mathbf{U}(\mathbf{R}), \overline{A_k})$. Using Corollary 1 (vi) together with $\mathbf{R}|\mathcal{A} \equiv \mathbf{B}$ and $\mathbf{U}(\mathbf{B}) \equiv \mathbf{B}$ gives $\nu(\mathbf{U}(\mathbf{R}), \overline{A_k}) = \mu(\mathbf{U}(\mathbf{R}), \overline{A_k})$, hence one can restrict attention to non-adaptive strategies. Now, for any fixed input sequence to $\mathbf{U}(\mathbf{R})$, A_k is the union of $\binom{k+z}{2} < (k+z)^2/2$ collision events, each with probability at most ϵ. Application of the union bound concludes the proof. $\square$

As a further demonstration of the general applicability of the framework, we give a simple security proof of a generalized version of the CBC-MAC (e.g., see Fig. 6 and [2]), with which we assume the reader is familiar. We do not wish to make an *a priori* assumption about the maximal message length, hence we need a prefix-free encoding $\sigma : \{0,1\}^* \to \{0,1\}^*$ of the binary strings which does not significantly expand the length. A good choice is to prepend a block encoding

[17] $P(\mathbf{U}(x) = \mathbf{U}(x')) \leq \epsilon$ for any $x \neq x'$. Actually, $\mathbf{U}$ must satisfy $P(\mathbf{U}(x) = y) \leq \epsilon$ for any x and y (which is usually the case).

[18] This is a cascade $\mathbf{UF}$, but this notation is incorrect because $\mathbf{U}$ depends on $\mathbf{F}$.

[19] As an alternative, a fixed value of $\mathbf{F}$ could be used as the key to generate the key of $\mathbf{U}$ pseudo-randomly. The security of such a scheme follows also from our analysis.

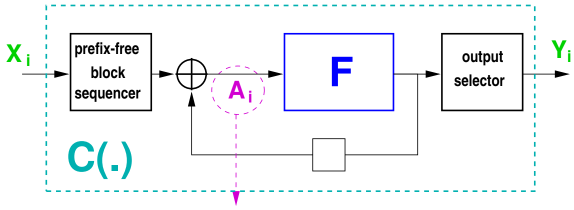

Fig. 6. The CBC-MAC. The $(\{0,1\}^*, \{0,1\}^l)$-random system $\mathbf{C}(\mathbf{F})$ is defined by applying some prefix-free encoding σ to the message, then padding the result with 0's to complete the last block, then applying the CBC feedback construction with a random function (or more generally a random automaton) $\mathbf{F}$, and taking the last output (for a given message) as the MAC-value for that message.

the length of the message, but from a theoretical viewpoint this restricts the message length and hence does not yield a true quasi-random oracle.[20]

Let $\mathbf{C}(\mathbf{F})$ be the $(\{0,1\}^*, \{0,1\}^l)$-random system defined by applying σ to the message, then padding with 0's to fill the last block, and then applying the CBC-MAC with a random function (or more generally a random system) $\mathbf{F}$ (cf. Fig. 6). A result similar in spirit to the following theorem was stated (without proof) independently by Petrank and Rackoff [17].

Theorem 6. *If $\mathbf{F}$ is a $d(k)$-QRF, then $\mathbf{C}(\mathbf{F})$ is a $\tilde{d}(k)$-quasi-random oracle for $\tilde{d}(k) = n^2 2^{-(l+1)} + d(n)$, where n is the total number of blocks of all k messages issued by the distinguisher.*

Proof. Lemma 5 (i) implies $\Delta_k(\mathbf{C}(\mathbf{F}), \mathbf{O}) \leq \Delta_k(\mathbf{C}(\mathbf{F}), \mathbf{C}(\mathbf{R})) + \Delta_k(\mathbf{C}(\mathbf{R}), \mathbf{O})$. Lemma 5 (ii) implies $\Delta_k(\mathbf{C}(\mathbf{F}), \mathbf{C}(\mathbf{R})) \leq d(n)$. Consider the event A_i that all inputs to $\mathbf{F}$ are distinct, up to and including the processing of the i-th message, except those inputs to $\mathbf{F}$ that are trivially equal because the prefix of the actual message processed so far is also a prefix of a previous message. Because due to σ no (encoded) message is a prefix of another message, A_i implies that for a given message x_i the last input to $\mathbf{F}$ (for x_i) is distinct from all previous inputs to $\mathbf{F}$ (for $x_1, \ldots, x_{i-1}$). Hence $\mathbf{C}(\mathbf{R})|\mathcal{A} \equiv \mathbf{O}$ and by Theorem 1 (i) we have $\Delta_k(\mathbf{C}(\mathbf{R}), \mathbf{O}) \leq \nu(\mathbf{C}(\mathbf{R}), \overline{A_k})$. Equation (2) is satisfied (for $\mathbf{G} = \mathbf{B}$) for all i since $P^{\mathbf{C}(\mathbf{R})}_{Y^i|X^i A_i}$ is the uniform distribution over $\{\{0,1\}^l\}^i$ for all input values (resulting in A_i being satisfied). Hence $\nu(\mathbf{C}(\mathbf{R}), \overline{A_k}) = \mu(\mathbf{C}(\mathbf{R}), \overline{A_k})$ and one can restrict attention to non-adaptive strategies, which are easy to analyse.

[20] A true prefix-free encoding $\sigma : \{0,1\}^* \to \{0,1\}^*$ can be obtained as follows. Let $\overline{n}$ be the standard binary representation of the integer n, and let $l(x)$ be the length of the binary string x. It is not difficult to see that the mapping $\sigma : \{0,1\}^* \to \{0,1\}^*$ defined by $r = l(\overline{l(x)}) - 1$ and $\sigma(x) := 0^r 1 || \overline{l(x)} || x$ is prefix-free. For instance, $\sigma(1100010111001) = 0001110111000101111001$. This encoding is efficient: $l(\sigma(x)) \approx l(x) + 2 \log l(x)$. It can be improved to $l(\sigma(x)) \approx l(x) + \log l(x)$ by using the encoding $x \mapsto \sigma(\overline{l(x)}) || x$.

For any given k input messages $x_1, \ldots, x_k$ of arbitrary lengths, but consisting of a total of n blocks, $\overline{A_k}$ corresponds to the event that a collision occurs among $n - w(x^k)$ independent and uniformly random values, where $w(x^k)$ is the total number of blocks in the messages $x_1, \ldots, x_k \in (\{0,1\}^l)^*$ which belong to a prefix (say of x_i) that was also the prefix of a previous message $x_1, \ldots, x_{i-1}$ (see above), i.e., $P^{\mathbf{C(R)}}_{\overline{A_k}|X^k}(x^k) = p_{\text{coll}}(2^l, n - w(x^k)) \le p_{\text{coll}}(2^l, n) \le n^2 2^{-(l+1)}$.[21] $\square$

7 Applications to the Analysis of Random Permutations

7.1 Random Permutations

For a random permutation[22] $\mathbf{Q}$, the inverse is also a random permutation and is denoted by $\mathbf{Q}^{-1}$. Remember that $\mathbf{P}$ denotes a uniform random permutation. Let $(\mathbf{E}, \mathbf{G})$ be any pair of (possibly dependent[23]) random permutations.

Lemma 8. (i) $\mathbf{EPG} \equiv \mathbf{P}$. *Moreover, if* $\mathbf{Q}|\mathcal{A} \equiv \mathbf{P}$*, then* $\mathbf{EQG}|\mathcal{A} \equiv \mathbf{P}$.
(ii) *For a MES* $\mathcal{C}$ *defined on the outputs of* $(\mathcal{X}, \mathcal{Y})$*-random systems such that* C_i *implies that the first* i *outputs are distinct, we have* $\mathbf{R}|\mathcal{C} \equiv \mathbf{P}|\mathcal{C}$ *and* $\mathbf{R}^{\mathcal{C}} \equiv \mathbf{P}^{\mathcal{C} \wedge \mathcal{D}}$ *for some MES* $\mathcal{D}$ *adjoined to* $\mathbf{P}$.

Proof. $\mathbf{EPG} \equiv \mathbf{P}$ is a special case of the second statement when A_i is the certain event for all i. We have $E\mathbf{Q}G|\mathcal{A} \equiv E\mathbf{P}G$ for any two fixed permutations E and G because E and G simply correspond to relabelings of the input and output alphabets of $\mathbf{Q}$. Hence this equivalence also holds if the pair (E, G) is a random variable. Now we prove (ii). We have $\mathbf{R}|\mathcal{C} \equiv \mathbf{P}|\mathcal{C}$ since conditioned on the output being distinct, both $\mathbf{R}$ and $\mathbf{P}$ generate completely new random outputs. Moreover, $P^{\mathbf{R}}_{C_i|X^iY^{i-1}C_{i-1}} \le P^{\mathbf{P}}_{C_i|X^iY^{i-1}C_{i-1}}$ is a simple consequence of the fact that for a given X^i with distinct values (i.e., $\text{dist}(X_1, \ldots, X_i)$), only Y^i with distinct values are consistent with $\mathbf{P}$, whereas other values for Y^i are consistent with $\mathbf{R}$, but C_i cannot hold for these Y^i. Now apply Lemma 1 (iv). $\square$

Definition 13. *A pairwise independent permutation (PIP) [18]* $\mathbf{Q}$ *is a random permutation such that for any two inputs* x *and* x', $\mathbf{Q}(x)$ *and* $\mathbf{Q}(x')$ *are a completely random pair of (distinct) values.*[24]

[21] The proof goes through for more general versions of the CBC-MAC. For example, in addition to letting the input to $\mathbf{F}$ be the current message block XORed with the previous output of $\mathbf{F}$, as in the CBC-MAC, one could XOR in any further function of all the previous message blocks and all the previous outputs of $\mathbf{F}$ (except the last). Such a modification could make sense if one considers the risk that $\mathbf{F}$ might not be a PRF and hence wants to build in extra complexity for heuristic security.

[22] Much of this section can be generalized to the more general concept of a permutation random system, i.e., a $(\mathcal{X}, \mathcal{X})$-random system $\mathbf{Q}$ which for all i is a random permutation on $\mathcal{X}^i$.

[23] However, the pair $(\mathbf{E}, \mathbf{G})$ is, as always, assumed to be independent of $\mathbf{Q}$.

[24] A PIP can for instance be implemented by interpreting all quantities as elements of a finite field F and setting $\mathbf{Q}(x) = ax + b$ for random $a, b \in F$ with $a \neq 0$.

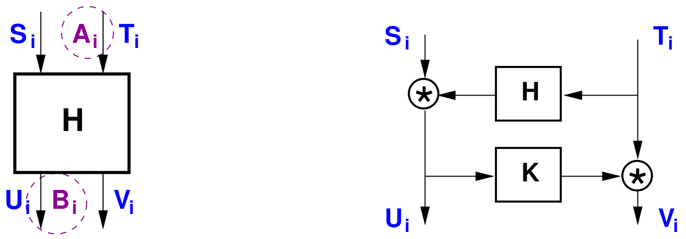

Fig. 7. Left side: Notation for random systems whose inputs and outputs are pairs. $A_i := \mathrm{dist}(T^i)$ and $B_i := \mathrm{dist}(U^i)$. Right side: Special case; two Feistel rounds with random systems $\mathbf{H}$ and $\mathbf{K}$, denoted $\mathbf{M}(\mathbf{H}, \mathbf{K})$.

7.2 Two Feistel Rounds with Random Functions

Let $\mathcal{R}$ be a set and let $\star$ be a group operation on $\mathcal{R}$. Typically $\mathcal{R} = \{0,1\}^l$ for some l and $\star$ is bitwise XOR. We now consider permutations on $\mathcal{R}^2$, i.e., on pairs which can be considered as "left" and "right" halves, or as high and low part when the pair is interpreted as a single element of, say, a field. For any random function $\mathbf{F} : \mathcal{R}^2 \to \mathcal{R}^2$ we can define the following random variables (see Figure 7, left): (S_i, T_i) is the i-th input and (U_i, V_i) are the i-th output. We define two MES, $A_i := \mathrm{dist}(T^i)$ and $B_i := \mathrm{dist}(U^i)$, used throughout Section 7.

For two random functions $\mathcal{R} \to \mathcal{R}$, $\mathbf{H}$ and $\mathbf{K}$, let $\mathbf{M}(\mathbf{H}, \mathbf{K})$ be the $\mathcal{R}^2$-random permutation defined by two Feistel rounds with $\mathbf{H}$ and $\mathbf{K}$ (see Figure 7, right).[25] More precisely, $U_i = S_i \star \mathbf{H}(T_i)$ and $V_i = T_i \star \mathbf{K}(U_i)$. Let $\mathbf{R} : \mathcal{R}^2 \to \mathcal{R}^2$ be a URF, and let $\mathbf{R}'$ and $\mathbf{R}''$ be URF's $\mathcal{R} \to \mathcal{R}$. We have

Lemma 9. $\mathbf{M}(\mathbf{R}', \mathbf{R}'')^{\mathcal{A} \wedge \mathcal{B}} \equiv \mathbf{B}^{\mathcal{A} \wedge \mathcal{B}} \equiv \mathbf{R}^{\mathcal{A} \wedge \mathcal{B}} \equiv \mathbf{P}^{\mathcal{A} \wedge \mathcal{B} \wedge \mathcal{D}}$ *for some MES* $\mathcal{D}$.

Proof. Given A_i, the joint distribution of (U_i, V_i) and B_i is identical for $\mathbf{M}(\mathbf{R}', \mathbf{R}'')$, for $\mathbf{B}$, and for $\mathbf{R}$, independent of the input: U_i and V_i are independent new random values and B_i is determined by U^i. Hence $\mathbf{M}(\mathbf{R}', \mathbf{R}'')^{\mathcal{A} \wedge \mathcal{B}} \equiv \mathbf{B}^{\mathcal{A} \wedge \mathcal{B}} \equiv \mathbf{R}^{\mathcal{A} \wedge \mathcal{B}}$. The last equivalence follows from $\mathbf{R}^{\mathcal{B}} \equiv \mathbf{P}^{\mathcal{B} \wedge \mathcal{D}}$ (Lemma 8 (ii)) and because $\mathcal{A}$ is defined on the inputs and thus Lemma 2 (ii) can be applied. □

7.3 Mono-directional Luby-Rackoff and Naor-Reingold

The following theorem generalizes the one-directional Luby-Rackoff [10] and Naor-Reingold [18] results (cf. Fig. 8 left) and follows easily from our framework.

Theorem 7. *Let* $\mathbf{L} := \mathbf{E}\mathbf{M}(\mathbf{R}', \mathbf{R}'')$ *for some random permutation* $\mathbf{E}$. *Then* $\Delta_k(\mathbf{L}, \mathbf{P}) \le \mu(\mathbf{E}, \overline{A_k}) + p_{\mathrm{coll}}(|\mathcal{R}|, k)$. *If* $\mathbf{E}$ *is a PIP (Naor-Reingold) or if* $\mathbf{E}$ *is a Feistel round with another random function* $\mathbf{R}'''$ *(Luby-Rackoff), then* $\Delta_k(\mathbf{L}, \mathbf{P}) \le 2 \cdot p_{\mathrm{coll}}(|\mathcal{R}|, k) < k^2/|\mathcal{R}|$.

[25] This can easily be generalized from random functions to random automata.

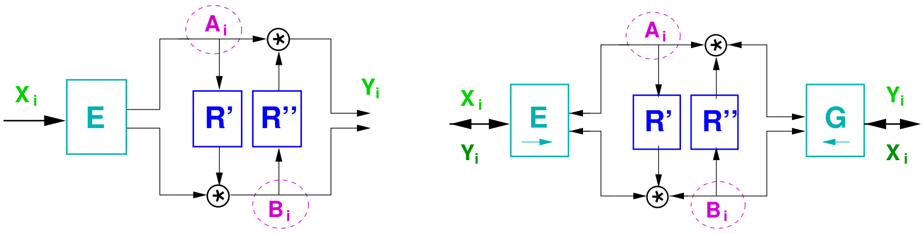

Fig. 8. Illustration for the one-directional (left) and bidirectional (right) Luby-Rackoff and Naor-Reingold results and generalizations thereof.

Proof. Using Lemma 9 and Lemma 4 we obtain

$$\mathbf{L}^{\mathcal{A}\wedge\mathcal{B}} \equiv \mathbf{EB}^{\mathcal{A}\wedge\mathcal{B}} \equiv \mathbf{EP}^{\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{D}} \tag{3}$$

(with the events A_i defined internally). Lemma 8 (i) yields the first step of

$$\Delta_k(\mathbf{L},\mathbf{P}) = \Delta_k(\mathbf{L},\mathbf{EP}) \leq \nu(\mathbf{L}, \overline{A_k} \vee \overline{B_k}) = \nu(\mathbf{EB}, \overline{A_k} \vee \overline{B_k})$$

and the next two steps follow from (3) and Theorem 1 (i), and from (3) and Lemma 6 (ii), respectively. Now obviously (and by Corollary 1 (v)), $\nu(\mathbf{EB}, \overline{A_k} \vee \overline{B_k}) \leq \mu(\mathbf{E}, \overline{A_k}) + \mu(\mathbf{B}, \overline{B_k})$ where $\mu(\mathbf{B}, \overline{B_k}) = p_{\mathrm{coll}}(|\mathcal{R}|, k)$. The second claim follows by a trivial analysis of a collision event among k random values. $\square$

Remark. Theorem 7, besides being more general, is also slightly stronger than that of [18] and [10] (see also [9]) where an additional term $k^2/(|\mathcal{R}|)^2$ appears on the right side. This weaker bound would in our context be obtained by proving $\Delta_k(\mathbf{L},\mathbf{R}) < k^2/|\mathcal{R}|$ and then using $\Delta_k(\mathbf{R},\mathbf{P}) \leq k^2/|\mathcal{R}|^2$. One could also append an additional random permutation $\mathbf{G}$, as follows directly from Corollary 1 (iii).

7.4 Bidirectional Permutations

Definition 14. For an $\mathcal{X}$-random permutation $\mathbf{Q}$, let $\langle\mathbf{Q}\rangle$ be the *bidirectional permutation*[26] $\mathbf{Q}$ with access from both sides (i.e., one can query both $\mathbf{Q}$ and $\mathbf{Q}^{-1}$). More precisely, $\langle\mathbf{Q}\rangle$ is the random function $\mathcal{X} \times \{0,1\} \to \mathcal{X}$ defined as follows:

$$\langle\mathbf{Q}\rangle(U_i, D_i) = \begin{cases} \mathbf{Q}(U_i) & \text{if } D_i = 0 \\ \mathbf{Q}^{-1}(U_i) & \text{if } D_i = 1\ . \end{cases}$$

If $\mathcal{A}$ is defined for $\mathbf{Q}$, $\mathcal{A}$ can also be defined naturally for $\langle\mathbf{Q}\rangle$: Let $V_i := \langle\mathbf{Q}\rangle(U_i, D_i)$, and let X_i and Y_i be the i-th input and output of $\mathbf{Q}$ (i.e., if $D_i = 0$, then $X_i = U_i$ and $Y_i = V_i$, and if $D_i = 1$, then $Y_i = U_i$ and $X_i = V_i$). Recall that $P^{\mathbf{Q}}_{Y_i A_i|X^i Y^{i-1} A_{i-1}} = P^{\mathbf{Q}}_{Y_i|X^i Y^{i-1} A_{i-1}} \cdot P^{\mathbf{Q}}_{A_i|X^i Y^i A_{i-1}}$. Now we let $P^{\langle\mathbf{Q}\rangle}_{A_i|X^i Y^i A_{i-1}} := P^{\mathbf{Q}}_{A_i|X^i Y^i A_{i-1}}$.

[26] This definition is motivated by considering a block cipher which in a mixed chosen-plaintext and chosen-ciphertext attack can be queried from both sides.

Lemma 10. *For any random permutatio* $\mathbf{F}$ *and* $\mathbf{G}$*,*
(i) $\Delta_k(\mathbf{F},\mathbf{G}) \le \Delta_k(\langle\mathbf{F}\rangle,\langle\mathbf{G}\rangle)$.[27]
(ii) *If* $\mathbf{F} \equiv \mathbf{G}$*, then* $\mathbf{F}^{-1} \equiv \mathbf{G}^{-1}$ *and* $\langle\mathbf{F}\rangle \equiv \langle\mathbf{G}\rangle$*.*
(iii) *More generally,* $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$ *implies* $\langle\mathbf{F}\rangle^{\mathcal{A}} \equiv \langle\mathbf{G}\rangle^{\mathcal{B}}$*.*

Proof. Claim (i) follows from the fact that being able to query from both sides can only help the distinguisher. Proof of claim (ii): the behavior of a random permutation $\mathbf{Q}$ uniquely determines the behavior of $\mathbf{Q}^{-1}$ and hence also of $\langle\mathbf{Q}\rangle$. Claim (iii) follows because if $\mathbf{F}^{\mathcal{A}} \equiv \mathbf{G}^{\mathcal{B}}$, then $P^{\mathbf{F}}_{A_i|X^iY^iA_{i-1}} = P^{\mathbf{G}}_{B_i|X^iY^iB_{i-1}}$ and thus $P^{\langle\mathbf{F}\rangle}_{A_i|U^iD^iV^iA_{i-1}} = P^{\langle\mathbf{G}\rangle}_{B_i|U^iD^iV^iB_{i-1}}$. □

The following theorem generalizes Theorem 3.2 of [18] in several ways. The proof is omitted.

Theorem 8. *Let* $\mathbf{L}$ *be defined as* $\mathbf{L} := \mathbf{EM}(\mathbf{R}',\mathbf{R}')\mathbf{G}^{-1}$ *(cf. Fig. 8 right).*
(i) *If* $\mathbf{E}$ *and* $\mathbf{G}^{-1}$ *are independent PIP's, then* $\Delta_k(\langle\mathbf{L}\rangle,\langle\mathbf{P}\rangle) < k^2/|\mathcal{R}|$*.*
(ii) *If* $\mathbf{E}$ *is a PIP and* $\mathbf{G} = \mathbf{E}^{-1}$*, then* $\Delta_k(\langle\mathbf{L}\rangle,\langle\mathbf{P}\rangle) < 4k^2/|\mathcal{R}|$*.*
(iii) *If* $\mathbf{R}' = \mathbf{R}''$*, i.e.,* $\mathbf{L} := \mathbf{EM}(\mathbf{R}',\mathbf{R}')\mathbf{E}^{-1}$*, then* $\Delta_k(\langle\mathbf{L}\rangle,\langle\mathbf{P}\rangle) < 8k^2/|\mathcal{R}|$*.*
(iv) *Moreover, if* $\mathcal{R} = GF(q)$ *is a field and* $\mathbf{E}$ *is also derived from* $\mathbf{R}'$ *by a linear polynomial* $ax+b$ *over* $GF(q^2)$ *with* a *and* b *defined by* $a = (\mathbf{R}(\xi_1)||\mathbf{R}(\xi_2))$ *and* $b = (\mathbf{R}(\xi_3)||\mathbf{R}(\xi_4))$ *for some fixed* $\xi_1,\xi_2,\xi_3,\xi_4 \in GF(q)$*, then* $\Delta_k(\langle\mathbf{L}\rangle,\langle\mathbf{P}\rangle) < 8(k+1)^2/|\mathcal{R}| + 1/|\mathcal{R}|^2$*.*

8 Conclusions

We have described a general framework for indistinguishability proofs of the most general form of random systems. The purpose of the framework is to prove results at the most general and abstract level, and this leads to substantial simplifications in actual security proof (making them for example tractable for a textbook) and to new security proofs that before may have appeared unrealistic. It would be a pleasure to see the framework at work in future security proofs.

We suggest as an open problem to find constructions of QRF's from QRF's better than that of Section 5, i.e., with either higher security (degree of indistinguishability) or lower complexity (number of evaluations of $\mathbf{F}$), or both. However, it is possible that this construction is quite close to optimal.

Acknowledgments

I would like to thank Thomas Holenstein, Olaf Keller, Krzysztof Pietrzak, and Renato Renner for many very helpful comments and for a careful proofreading, and Markus Stadler for discussions at an early stage of this work.

[27] $\Delta_k(\langle\mathbf{F}\rangle,\langle\mathbf{G}\rangle)$ can be much larger than $\Delta_k(\mathbf{F},\mathbf{G})$ because inverse queries may help the distinguisher significantly.

References

1. M. Bellare, O. Goldreich, and H. Krawczyk, Stateless evaluation of pseudorandom functions: security beyond the birthday barrier, *Advances in Cryptology - CRYPTO '99*, Lecture Notes in Computer Sc., vol. 1666, pp. 270–287, Springer-Verlag, 1999.
2. M. Bellare, J. Kilian, and P. Rogaway, The security of the cipher block chaining message authentication code, *Advances in Cryptology - CRYPTO '94*, Lecture Notes in Computer Science, vol. 839, pp. 341–358, Springer-Verlag, 1995.
3. M. Bellare, J. Guérin, and P. Rogaway, XOR MACs: New methods for message authentication using finite pseudorandom functions, *Advances in Cryptology - CRYPTO '95*, Lecture Notes in Computer Science, vol. 963, Springer-Verlag, 1994.
4. D. J. Bernstein, How to stretch random functions: The security of protected counter sums, *Journal of Cryptology*, vol. 12, pp. 185–192, Springer-Verlag, 1999.
5. J. Black, S. Halevi, H. Krawczyk, T. Krovetz, and P. Rogaway, UMAC: Fast and secure message authentication, *Advances in Cryptology - CRYPTO '99*, Lecture Notes in Computer Science, vol. 1666 pp. 216–233, Springer-Verlag, 1999.
6. R. E. Blahut, *Principles and practice of information theory*, Addison-Wesley Publishing Company, 1988.
7. M. Blum and S. Micali, How to generate cryptographically strong sequences of pseudo-random bits, *SIAM J. on Computing*, vol. 13, no. 4, pp. 850–864, 1984.
8. O. Goldreich, S. Goldwasser, and S. Micali, How to construct random functions, *Journal of the ACM*, vol. 33, no. 4, pp. 210–217, 1986.
9. M. Luby, *Pseudorandomness and Cryptographic Applications*, Princeton University Press, 1996.
10. 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.
11. U. M. Maurer, Conditionally-perfect secrecy and a provably-secure randomized cipher, *Journal of Cryptology*, vol. 5, pp. 53–66, Springer-Verlag, 1992.
12. ——, A simplified and generalized treatment of Luby-Rackoff pseudo-random permutation generators, *Advances in Cryptology - EUROCRYPT '92*, Lecture Notes in Computer Science, vol. 658, pp. 239–255, Springer-Verlag, 1992.
13. ——, Extended version of this paper, see `www.crypto.ethz.ch/publications/`.
14. J. Patarin, Etude des générateurs de permutations basés sur le Schéma du D.E.S., Ph. D. Thesis, INRIA, Le Chesnay, France, 1991. An extract appeared in: J. Patarin, New results on pseudorandom permutation generators based on the DES scheme, *Advances in Cryptology – CRYPTO'91*, J. Feigenbaum (ed.), Lecture Notes in Computer Science, Vol. 576, Springer-Verlag, pp. 301–312, 1992.
15. ——, How to construct pseudorandom permutations from a single pseudorandom function, *Advances in Cryptology - EUROCRYPT '92*, R. Rueppel (ed.), Lecture Notes in Computer Science, vol. 658, pp. 256–266, Springer-Verlag, 1992.
16. ——, About Feistel schemes with six (or more) rounds, *Fast Software Encryption*, Lecture Notes in Computer Science, vol. 1372, pp. 103–121, Springer-Verlag, 1998.
17. E. Petrank and C. Rackoff, CBC MAC for real-time data sources, *Journal of Cryptology*, vol. 13, no. 3, pp. 315–338, 2000.
18. M. Naor and O. Reingold, On the construction of pseudorandom permutations: Luby-Rackoff revisited, *Journal of Cryptology*, vol. 12, no. 1, pp. 29–66, 1999.
19. M. O. Rabin, Transaction protection by beacons, *J. Comp. Sys. Sci.*, vol. 27, pp. 256–267, 1983.
20. V. Shoup, On fast and provably secure message authentication based on universal hashing, *Advances in Cryptology - CRYPTO '96*, Lecture Notes in Computer Science, vol. 1109, pp. 313–328, Springer-Verlag, 1996.

21. 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.
22. ——, On provable security for conventional ciphers, in *Proc. of ICISC'99*, Lecture Notes in Computer Science, Springer-Verlag, 1999.
23. M. N. Wegman and J. L. Carter, New hash functions and their use in authentication and set equality, *J. of Computer and System Sciences*, vol. 22, pp. 265–279, 1981.

Appendix

Lemma 11. *Let* $\mathbf{U} = [U_1, \ldots, U_n]$ *with* $U_i \in GF(q)$ *be a vector of random variables with uniform distribution* $GF(q)^n$*, and define the random function* $\mathbf{K} : GF(q)^n \to GF(q)$ *as the scalar product of the input vector* $\mathbf{x} = [x_1, \ldots, x_n] \in GF(q)^n$ *and* $\mathbf{U}$*,*

$$\mathbf{K}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{U} \rangle = \sum_{j=1}^{n} x_j U_j.$$

Then $\mathbf{K}^{\mathcal{A}} \equiv \mathbf{R}^{\mathcal{A}} \equiv \mathbf{B}^{\mathcal{A}}$ *with* A_i *as the event that* $\mathbf{x}_1, \ldots, \mathbf{x}_i$ *are linearly independent.*

Proof. For a list $\mathbf{v}^k = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ of vectors in a finite-dimensional vector space, let $span(\mathbf{v}^k)$ denote the subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_k$ and let $dim(\mathbf{v}^k)$ denote its dimension. If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent, then $dim(\mathbf{v}^k) = k$.

Let $T \subseteq GF(q)^n$ be a set of input vectors to $\mathbf{K}$, and let $\mathbf{K}(T)$ denote the corresponding list of values of $\mathbf{K}$. We prove[28] that $H(\mathbf{K}(T)) = dim(T)r$, where $r = \log q$. This clearly implies that for any set of linearly independent vectors the corresponding function values have maximal entropy, as is to be proved. Linear dependence implies functional dependence, hence $H(\mathbf{K}(T)) = H(\mathbf{K}(span(T))) = H(\mathbf{K}(span(B)))$, where B is any basis of $span(T)$ and has cardinality $B = dim(T)$. Thus $H(\mathbf{K}(T)) \leq dim(T)r$. On the other hand, it follows from linear algebra that T can be complemented by a set T' of size $n - dim(T)$ such that $T \cup T'$ spans the entire space $GF(q)^n$. Hence $H(\mathbf{K}|\mathbf{K}(T)) \leq (n - dim(T))r$. Because $H(\mathbf{K}) = H(\mathbf{K}(T)) + H(\mathbf{K}|\mathbf{K}(T)) = nr$ we must have equality in the two previous inequalities. $\square$

Let $S_n := \{1, \ldots, n\}$. The characteristic vector in $\{0,1\}^n$ of a subset S' of S_n has a 1 at position i if and only if $i \subseteq S'$. For multi-sets or lists of elements of S_n, we define the characteristic vector to have a 1-entry only for those elements of S_n that occur *exactly once.*

The proof of following lemma is straight-forward.

Lemma 12. *If* kt *elements of* S_n *are selected* b*-wise independently (for* $b \geq 2t$*) and interpreted as* k *lists of* t *elements,* $V_i = [V_{i1}, \ldots, V_{it}]$ *for* $1 \leq i \leq k$*, then their characteristic vectors* $W_1, \ldots, W_k$ *are linearly independent with probability at least* $1 - k\left(\frac{kt}{n}\right)^t$.

[28] See [6] for definitions of the entropy $H(X)$ and the conditional entropy $H(X|Y)$.