ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 45, Number 4, 2015

A MEASURE OF DEPENDENCE FOR CRYPTOGRAPHIC PRIMITIVES RELATIVE TO IDEAL FUNCTIONS

DANIEL SMITH-TONE AND CRISTINA TONE

ABSTRACT. In this work, we present a modification of a well-established measure of dependence appropriate for the analysis of stopping times for adversarial processes on cryptographic primitives. We apply this measure to construct generic criteria for the ideal behavior of fixed functions in both the random oracle and ideal permutation setting. More significantly, we provide a nontrivial extension of the notion of hash function indifferentiability, transporting the theory from the status of providing security arguments for protocols utilizing ideal primitives into the more realistic setting of protocol assurance with fixed functions. The methodology this measure introduces to indifferentiability analysis connects the security of a hash function with an indifferentiable mode to the security of the underlying compression function in a quantitative way; thus, we prove that dependence results on cryptographic primitives provide a direct means of determining the practical resistance or vulnerability of protocols employing such primitives.

1. Introduction. Many real world phenomena can be studied by associating them with a sequence of identically distributed discrete random variables, corresponding to measurements, which take values in some finite set. If the random variables are independent, the behavior of such processes is well known, both in its long term, or asymptotic behavior and in its behavior over an intermediate finite number of measurements. The situation for dependent processes, however, is less clear. Although the limit theory for dependent random sequences has also been extensively developed, the behavior of dependent processes in the "semi-long term" has received less focus.

Hash analysis provides an excellent example of a process for which this "semi-long term" behavior is of the greatest significance. Specifically, an adaptive adversary trying to discover a weakness in a hash

2010 AMS *Mathematics subject classification.* Primary 94A60.
Received by the editors on July 16, 2013.
DOI:10.1216/RMJ-2015-45-4-1283

function will choose a strategy which depends on the information obtained through interaction with the function. The experiment will continue until some predetermined information is gleaned from this process. Since practical hash functions have a finite output, the stochastic process defined by the interactions of the adversary with the hash function has a stopping time.

The stopping time of such a dependent process is of critical importance in cryptography. An attack on a hash function can take many forms, as one may construe as an attack any process which gathers enough information to distinguish the fixed function from an ideal function with the same codomain. For any meaningful attack, the measure with which we can judge the resistance of such a function is determined by the stopping time of the adaptive adversarial process.

The purpose of this work is to provide a mathematical framework upon which we may study the stopping times of optimal adaptive adversarial strategies. More specifically, we introduce measures of dependence relevant to the study of stopping times for dependent stochastic processes and apply these measures, determining the properties required of a hash primitive to achieve various security criteria.

The manuscript is organized as follows. Section 2 introduces two classical measures of dependence employed in limit theory to determine the asymptotic behavior of mixing random sequences. In Section 3, we present the appropriate terminology required for the cryptographic application. In the following sections, namely Sections 4 and 5, we derive analogues of the measure of dependence appropriate to the study of stopping times and apply the resulting theory to fixed random functions and fixed permutations, respectively. In Section 6, we further employ the dependence condition extending the powerful indifferentiability framework to provide assurance in a more practical setting. Finally, in Section 7, we apply this extension of indifferentiability to some actual hash functions and offer some directions for future work.

2. Mixing coefficients. Strong mixing conditions have been used greatly, both in the context of random sequences and in the broader context of random fields, in order to study phenomena for which observations which are close to each other in time or location may show considerable influence on one another, while observations which are far apart in time or location are almost independent. Many results in the

study of strong mixing conditions for random sequences have been in the literature in connection with various fields of study: Rosenblatt [**25**] and Zhurbenko [**26**], with respect to the estimation of spectral density; Phillips [**23**], for the use of ARMA models in the study of economics; Dabrowski, McDonald and Rösler [**10**], in connection with the flow of electrical signals in the heart or nerve membrane; Halversam and Wise [**15**], related to the detection of a signal in the presence of noise; Philipp [**22**], in connection with some random processes arising from number theory; Kesten and Papanicolaou [**17**], with respect to the motion of a particle in a velocity field; and Davydov [**11**] and Meyn and Tweedie [**19**], for Markov chains.

In 1956, Rosenblatt [**24**] proposed one particularly useful type of dependence which he referred to as the "strong mixing condition" or the α-mixing condition. Since then, a lot of progress has been made in this area (see [**6, 10, 11, 15, 17, 19, 22, 23, 25**] in relation to the following mixing coefficients. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability measure space and $\mathcal{A}$, $\mathcal{B} \subset \mathcal{F}$ are two σ-fields. Define the following measures of dependence:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{\substack{A \in \mathcal{A} \\ B \in \mathcal{B}}} |P(A \cap B) - P(A)P(B)|,$$

and

$$\phi(\mathcal{A}, \mathcal{B}) := \sup_{\substack{A \in \mathcal{A} \\ B \in \mathcal{B} \\ P(A) > 0}} |P(B|A) - P(B)|.$$

It is well known and elementary (see [**5,** Proposition 3.11 (a)]) that

$$0 \leq 2\alpha(\mathcal{A},\ \mathcal{B}) \leq \phi(\mathcal{A},\ \mathcal{B}) \leq 1.$$

It is obvious to notice that if the σ-fields $\mathcal{A}$, $\mathcal{B}$ are independent, then $\alpha(\mathcal{A},\ \mathcal{B}) = 0$ and $\rho(\mathcal{A},\ \mathcal{B}) = 0$, and vice versa.

Suppose $\xi := (\xi_k, k \in \mathbb{Z})$ is a (not necessarily stationary) sequence of random variables. For $-\infty \leq J \leq L \leq \infty$, define the σ-field

$$\mathcal{F}_J^L := \sigma(\xi_k, J \leq k \leq L),$$

the σ-field generated by the random variables $(\xi_k,\ J \leq k \leq L)$. It is understood that the index k is restricted to the integers. This notation will also be used for (not necessarily stationary) "one-sided" random

sequences $\xi := (\xi_1,\ \xi_2,\ \xi_3, \ldots)$, with the obvious modification that σ-field $\mathcal{F}_J^L$ is defined only for $1 \leq J \leq L \leq \infty$. In this paper, for the "one-sided" random sequences $\xi := (\xi_1,\ \xi_2,\ \xi_3, \ldots, \xi_n)$ the σ-field $\mathcal{F}_J^L$ is defined only for $1 \leq J \leq L \leq n$. For each integer n, define the following dependence coefficients:

$$\alpha(n) := \alpha(\xi, n) := \sup_{J \in \mathbb{Z}} \alpha(\mathcal{F}_{-\infty}^J, \mathcal{F}_{J+n}^{\infty})$$

and

$$\phi(n) := \phi(\xi, n) := \sup_{J \in \mathbb{Z}} \phi(\mathcal{F}_{-\infty}^J, \mathcal{F}_{J+n}^{\infty}).$$

The random sequence $\xi := (\xi_k, k \in \mathbb{Z})$ (whether stationary or not) is said to be "strongly mixing" (or α-mixing), respectively "ϕ-mixing" if $\alpha(n) \to 0$ and $\phi(n) \to 0$, respectively, as $n \to \infty$. The strong mixing condition $\alpha(n) \to 0$ was introduced by Rosenblatt **[24]**, and the ϕ-mixing condition $\phi(n) \to 0$ was introduced by Ibragimov **[16]** and also studied by Cogburn **[8]**. While in this paper we are not interested in the asymptotic behavior of the random process, due to the fact that n is an arbitrary fixed integer, it is necessary to emphasize the challenge of showing the behavior for our random process when the number of steps is large and fixed, without looking at its limiting behavior.

3. Cryptographic primitives and modes. In cryptography, hash functions are designed in such a way as to simulate a random function while committing to its output values. For a good hash function, it should be computationally infeasible to find a pre-image of a known hash value, alter a known input/output pair to find a second pre-image, or generate two inputs which hash to the same value. A simple reason for these design criteria is that they are requisite for a function to behave like a random oracle, a theoretical function often used to prove the security of complex protocols.

A finite random oracle, $ro : A \to B$, is a function chosen uniformly at random from among all functions from the finite set A to the finite set B. To collect information about any function F, one may submit queries, inputs for which F provides outputs, and catalog the results. If a query has never been made to the function previously, it is called a *fresh query*. Random oracles have the property that every fresh query produces an output which is uniformly distributed and independent of all previously cataloged information.

We may define a variable input length random oracle, $RO : \{0,1\}^* \to B$, with a similar property by selecting a sequence of uniformly distributed independent and identically distributed (i.i.d.) B-valued random variables, X_λ, and defining $RO(\lambda) = X_\lambda$. To make this object well defined as a function, it is understood that RO commits to its outputs.

Random oracles are not appropriate for all theoretical applications. In particular, in some contexts, it is important for a permutation to behave unpredictably. In such contexts, we employ ideal permutations. An ideal permutation, $\pi : A \to A$, is a permutation chosen uniformly at random from among all permutations of the finite set A.

One model for building a practical hash function is to chain together fixed finite functions which are designed to mimic the behavior of a finite random oracle or an ideal permutation. A hash mode is a method for creating a variable input length function from such functions of fixed input length. If a particular hash mode is proven secure when its components are assumed to be ideal primitives (either a finite random oracle or an ideal permutation, depending on the design), then the task of creating a secure hash function is reduced to the challenge of creating fixed functions which behave like the ideal functions.

These ideal functions provide a road-map for proving the security of a practical hash function. First, the hash mode is analyzed with the fixed primitives replaced by ideal primitives. If the resulting variable input length function is shown to be indistinguishable from a variable input length random oracle, the mode is considered secure. It then suffices to show that the actual fixed primitives employed in the hash design are indistinguishable from ideal primitives.

4. Fixed random functions. Consider a finite fixed function, F. An adversary, trying to reveal a weakness in F may attempt a collision attack (finding two inputs which map to the same output), a preimage attack (finding a preimage for some given fixed but arbitrary value of the codomain of F), or a second-preimage attack (finding a second preimage for a known input/output pair). For such a fixed function, there may exist some symmetry which a diligent adversary can discover through interaction with the function and possibly exploit to affect one of the above attacks.

Consider, in contrast, the interaction of an adversary with a finite random oracle, *ro*. Clearly, there is no benefit to submitting a single query repeatedly to the oracle; thus, we may assume that the adversary submits fresh queries. Consequently, we can define a sequence of uniformly distributed i.i.d. random variables, $\xi_1,\ \xi_2, \ldots$, corresponding to each successive output of *ro*.

Let the random variable X denote the number of rounds of querying before the adversary is able to form a collision with outputs of *ro*. Then its probability density function is $f(x) = x(n-1)!/(n-x)!n^{-x}$, where $x = 1,\ 2, \ldots, n$. Consequently, the expected value of X is given by

$$E(X) = \sum_{i=1}^{n} i^2 \frac{(n-1)!}{(n-i)!} n^{-i}. \tag{4.1}$$

Classical calculation shows the relative size of this quantity, specifically, by Markov's inequality, $E(X) \geq \alpha P(X \geq \alpha)$, and we may find a lower bound for $E(X)$ by finding α satisfying $P(X \geq \alpha) \geq 1/2$. Clearly, this is equivalent to finding $P(X < \alpha) \leq 1/2$ since, for all $i \leq n$, we have $P(X > i-1) > 0$. We obtain, for all $\alpha \leq n$:

$$P(X < \alpha) \leq \sum_{i=1}^{\alpha-1} P(X = i \mid X > i-1) = \sum_{i=1}^{\alpha-1} \frac{i}{n} = \frac{1}{n}\binom{\alpha}{2}. \tag{4.2}$$

Setting this quantity equal to $1/2$, we may derive that

$$\alpha = (\sqrt{4n+1}+1)/2.$$

This fact implies that

$$E(X) \geq (\sqrt{4n+1}+1)/2 P(X \geq \alpha) \geq (\sqrt{4n+1}+1)/4,$$

which shows that the sum in (4.1) is $\Omega(\sqrt{n})$.

To get the upper bound, notice that the expected number of rounds to get the first collision must be less than or equal to the number of rounds for the expected number of collisions to be at least one, since there can be at most one collision in any given round. Therefore, we can show that this number is of order $\sqrt{n}$. Let $I_{i,j}$ be the indicator that the ith and jth nodes visited coincide. Let Y be the number of

collisions, that is,

$$Y = \sum_{i=1}^{\sigma-1} \sum_{j=i+1}^{\sigma} I_{i,j},$$

where σ represents an arbitrary number of rounds. Therefore, considering the fact that $P(I_{i,j} = 1) = 1/n$, we have:

$$E(Y) = \sum_{i=1}^{\sigma-1} \sum_{j=i+1}^{\sigma} \frac{1}{n} = \binom{\sigma}{2} \frac{1}{n}. \tag{4.3}$$

Thus, for $E(Y) = 1$, we need $\sigma = \lceil \sqrt{2n + 1/4} + 1/2 \rceil$. Thus, our sum is $O(\sqrt{n})$, and therefore, $\Theta(\sqrt{n})$.

The critical question in the context of random functions is, "what happens when we are not guaranteed independence among function output in each of the rounds?" For each $i \in \{1,\ 2, \ldots, n\}$, let us define ξ_i to be a simple random variable taking the values $\{1,\ 2, \ldots, n\}$, where each integer represents the output of a fixed function, F, in round i.

Clearly, whatever the adversary's criteria for a successful attack may be, he or she may only consider the attack successful by referring to information from the current round and past rounds. The adversary may only make judgements in round i based on information from the previous rounds. Therefore, to study the stopping time of the adversarial process, we are specifically interested in a measure of dependence involving past events and events in round i, a measure which consequently bounds the performance of an optimal adversary.

Definition 4.1. Consider the process $\{\xi_k\}$ where $k \in \{1,\ 2, \ldots, n\}$ together with a collection of pairs of sigma fields, $\mathcal{A}_i$ and $\mathcal{B}_i$, where $\mathcal{A}_1$ is the trivial sigma field, $\mathcal{A}_i = \mathcal{F}_1^{i-1}$ for $2 \leq i \leq n$, and $\mathcal{B}_i = \mathcal{F}_i^i$ for $1 \leq i \leq n$. We define the dependence sequence of this process in the following way: for $i \in \{1,\ 2, \ldots, n\}$,

$$\phi_i = \phi(\mathcal{A}_i, \mathcal{B}_i) = \sup_{\substack{A \in \mathcal{A}_i \\ B \in \mathcal{B}_i}} |P(B|A) - P(B)|, \quad P(A) > 0. \tag{4.4}$$

If the ξ_i are the outputs of a random function, F, with a finite codomain of size n, then we say that F is a $\phi[p]$-mixing random function provided that $\phi_i \leq n^{-p}$ for all $i < n$.

By an easy calculation, one may notice that, for each $i \in \{1,2,\ldots,n\}$, if $A \in \sigma(\xi_1,\ldots,\xi_{i-1})$ and $B \in \sigma(\xi_i)$, then

$$|P(A \cap B) - P(A)P(B)| \leq P(A) \cdot \phi_i. \tag{4.5}$$

Now we compute bounds on the probability of obtaining the first collision on round i. Clearly, larger probabilities of collision in early rounds result in a lower expected number of rounds to form a collision. With this in mind, we compute each probability assuming that each previous round attains the theoretical maximum probability of collision. For the first round, we use the formal notation ϕ_1, which simply represents zero, since there are no events occurring before the first round.

In this manner, we compute that the probability of getting a collision in the first round is $1/n = (1/n) + \phi_1$. Therefore, the probability of noncollision in the first round is $(n-1)/n - \phi_1$. Were the outcome in the second round uniformly distributed and independent of the outcome of the first round, the probability of a collision in the second round *and* a noncollision in the first round would be $((n-1)/n - \phi_1)(2/n)$. The definition of ϕ_2 provides us with a possible error in this estimation; specifically, we may be as far off the correct value as $((n-1)/n-\phi_1)\phi_2$. Therefore, the probability that the first collision occurs in round two is bounded by $((n-1)/n - \phi_1)((2/n) + \phi_2)$.

For each $i \in \{1,\ 2,\ldots,n\}$, we may compute the probability of C_i, the event that the first collision occurs in round i, given that each collision event is assigned the maximal probability sequentially,

$$P(C_i) = \left(\frac{i}{n} + \phi_i\right) \prod_{1 \leq k < i} \left(\frac{n-k}{n} - \phi_k\right). \tag{4.6}$$

Consider this expression as a polynomial in $\mathbb{R}[\phi_1]\ldots[\phi_i]$. For each $i \in \{1,\ 2,\ldots,n\}$, we use $\Psi_i^{(k)}$ to denote the sum of the total degree k terms in this bound. Then we obtain

$$\Psi_i^{(0)} = \frac{i}{n^i} \frac{(n-1)!}{(n-i)!}. \tag{4.7}$$

We denote the terms contained in $\Psi_i^{(1)}$ by

$$\Psi_{i,j}^{(1)} = \begin{cases} \frac{\phi_i}{n^i}\frac{n!}{(n-i)!} & \text{if } j = i \\ -\frac{i\phi_j}{n^i}\frac{n!}{(n-i)!(n-j)} & \text{if } j < i, \end{cases} \tag{4.8}$$

where i represents the number of rounds and j represents the index of the variable ϕ used. Thus, equation (4.6) becomes: for each $i \in \{1,\ 2, \ldots, n\}$,

$$P(C_i) = \Psi_i^{(0)} + \sum_{1 \le j \le i} \Psi_{i,j}^{(1)} + \text{higher degree terms.} \tag{4.9}$$

Now we can explicitly write down a lower bound formula on the expected number of rounds, X, required to form a collision relative to the ϕ_i's. Specifically, we have that:

$$E(X) = \sum_{i=1}^{n} i\Psi_i^{(0)} + \sum_{i=1}^{n} i \sum_{j=1}^{i} \Psi_{i,j}^{(1)} + \text{sum of higher degree terms.} \tag{4.10}$$

From the birthday paradox, we obtain $\sum_{i=1}^{n} i\Psi_i^{(0)} = \Theta(\sqrt{n})$. To evaluate the second summation in equation (4.10), we divide into two cases:

Case 1: $j = i$. Under this condition, the second summation in (4.10) contributes positively to the expected value. Specifically, this sum is bounded below by:

$$n \min_{l \le n}(\phi_l) \sum_{i=1}^{n} \frac{i}{n^i} \frac{(n-1)!}{(n-i)!}.$$

Notice that each term in the sum above is exactly $\Psi_i^{(0)}$, the probability of forming a collision among the n outputs at step i. Therefore, this entire term simplifies to $n \min_{l \le n}(\phi_l)$, which is nonnegative. On a side note, in conjunction with our subsequent calculations, this lower bound for the summand provides a natural upper bound on the dependence coefficients, ϕ_i. Note that $E(X) = O(\sqrt{n})$, which implies that any fixed function has some dependence coefficient smaller than $Cn^{-1/2}$. Case 1 is complete.

Case 2: $j < i$. Under this condition, the second sum in equation (4.10) becomes:

(4.11)
$$-\sum_{i=2}^{n}\sum_{j=1}^{i-1}\frac{\phi_j i^2}{n^i}\frac{n!}{(n-i)!(n-j)} = -\sum_{j=1}^{n-1}\frac{n\phi_j}{n-j}\sum_{i=j+1}^{n}\frac{i^2}{n^i}\frac{(n-1)!}{(n-i)!}.$$

Note that the inner sum on the index i above is a partial sum with terms $i\Psi_i^{(0)}$; therefore, it is bounded by $C\sqrt{n}$. Thus, the right-hand side of (4.11) is lower bounded by

$$-Cn^{3/2}\sum_{j=1}^{n-1}\frac{\phi_j}{n-j} = -Cn^{3/2}\sum_{j=1}^{n-1}\frac{\phi_{n-j}}{j} \geq -Cn^{3/2}\max_i \phi_i \sum_{j=1}^{n-1}\frac{1}{j}.$$

Consequently, since the partial sum of the harmonic series above is bounded by $\ln(n)+1$, we obtain that, for $n > 1$,

$$-Cn^{3/2}\max_i \phi_i \sum_{j=1}^{n-1}\frac{1}{j} \geq -Cn^{3/2}(\ln(n)+1)\max_i \phi_i.$$

Therefore, the condition $\phi_i < n^{-1-\epsilon}$ is sufficient to conclude that the second sum in (4.10) is $o(n^{1/2})$ and, hence, the lower bound of the expected value is $\Omega(\sqrt{n})$, provided that the higher degree terms are negligible for such values of ϕ_i.

The following theorem will show the fact that the higher degree terms are indeed negligible if $\phi_i < n^{-1-\epsilon}$ for all $i \in \{1,\ 2, \ldots, n\}$.

Theorem 4.2. *A function, F, has ideal collision resistance provided it is $\phi[1+\epsilon]$-mixing for some $\epsilon > 0$.*

Proof. From above, we have that $E(X) = \Omega(\sqrt{n})+$ is a sum of degree two and higher terms. We need to show that the sum of the terms with total degree greater than or equal to two is $o(\sqrt{n})$. For this, since there are fewer than n possible total degrees, it suffices to show that the sum restricted to each total degree is $o(n^{-1/2})$. Consider the degree $d \geq 2$ sum, $\sum_{i=1}^{n} i\Psi_i^{(d)}$. Allowing $j_1 < j_2 < \cdots < j_d \leq i$, we denote the terms

in $\Psi_i^{(d)}$ by

(4.12)
$$\Psi^{(d)}_{i,j_1,\ldots,j_d} = \begin{cases} (-1)^{d+1} \frac{\phi_i \phi_{j_1} \cdots \phi_{j_{d-1}}}{n^{i-1}} \frac{n!}{(n-i)!(n-j_1)\cdots(n-j_{d-1})} & \text{if } j_d = i \\ (-1)^{d} \frac{i \phi_{j_1} \cdots \phi_{j_d}}{n^{i-1}} \frac{n!}{(n-i)!(n-j_1)\cdots(n-j_d)} & \text{if } j_d < i. \end{cases}$$

We may bound the absolute values of sums of these terms with $j_d = i$ and with $j_d < i$ individually. As before, we consider the following two cases:

Case 1: $j_d = i$. Under this condition, we obtain that

$$\sum_{i=d}^{n} \sum_{j_{d-1}=d-1}^{i-1} \cdots \sum_{j_1=1}^{j_2-1} |i\Psi_{i,j_1,\ldots,j_{d-1},i}| = \sum_{j_{d-1}=d-1}^{n-1} \cdots \sum_{j_1=1}^{j_2-1} \sum_{i=j_{d-1}+1}^{n} |i\Psi_{i,j_1,\ldots,j_{d-1},i}|. \tag{4.13}$$

By (4.12), replacing the ϕ_k's with their maximum values and multiplying and dividing by n on the right-hand side of (4.13), we obtain:

$$n \max_{l \le n}(\phi_l^d) \sum_{j_{d-1}=d-1}^{n-1} \frac{1}{n-j_{d-1}} \cdots \sum_{j_1=1}^{j_2-1} \frac{1}{n-j_1} \sum_{i=j_{d-1}+1}^{n} \frac{i}{n^i} \frac{n!}{(n-i)!}.$$

Since the innermost sum is the probability that it takes more than j_{d-1} rounds to have a collision in the outputs of a random oracle, our quantity above is bounded by

$$n \max_{l \le n}(\phi_l^d) \sum_{j_{d-1}=d-1}^{n-1} \frac{1}{n-j_{d-1}} \cdots \sum_{j_1=1}^{j_2-1} \frac{1}{n-j_1} \le n \max_{l \le n}(\phi_l^d)(\ln(n)+1)^{d-1}.$$

The condition $\phi_i < n^{-1-\epsilon}$ assures that the sum in (4.13) is $o(n^{1-d})$, which for $d \ge 2$ means it is $o(n^{-1})$ and, in particular, $o(n^{-1/2})$. As a consequence, Case 1 is complete.

Case 2: $j_d < i$. In this second case, we have:

$$\sum_{i=d+1}^{n} \sum_{j_d=d}^{i-1} \cdots \sum_{j_1=1}^{j_2-1} |i\Psi_{i,j_1,\ldots,j_d}| = \sum_{j_d=d}^{n-1} \cdots \sum_{j_1=1}^{j_2-1} \sum_{i=j_d+1}^{n} |i\Psi_{i,j_1,\ldots,j_d}|. \tag{4.14}$$

We can rewrite the right-hand expression above as

$$\sum_{j_d=d}^{n-1} \cdots \sum_{j_1=1}^{j_2-1} \frac{n\phi_{j_1} \cdots \phi_{j_d}}{(n-j_1)\cdots(n-j_d)} \sum_{i=j_d+1}^{n} \frac{i^2}{n^i} \frac{n!}{(n-i)!}.$$

Again, since the inner-most sum is a partial sum of the expected number of rounds required to form a collision in the outputs of a random oracle, we can bound it by $Cn^{1/2}$. Therefore, we have

$$\begin{aligned} Cn^{3/2} \sum_{j_d=d}^{n-1} \frac{\phi_{j_d}}{n-j_d} \cdots \sum_{j_1=1}^{j_2-1} \frac{\phi_{j_1}}{n-j_1} &\leq Cn^{3/2} \max_{l \leq n}(\phi_l^d) \left(\sum_{i=1}^{n-1} \frac{1}{i} \right)^d \\ &\leq Cn^{3/2} \max_{l \leq n}(\phi_l^d)(\ln(n)+1)^d. \end{aligned}$$

Using the given bound, $\phi_i < n^{-1-\epsilon}$, we obtain a bound of $o(n^{3/2-d})$ for the sum in (4.14), which, for $d \geq 2$, it implies the sum is $o(n^{-1/2})$. Therefore, Theorem 4.2 is complete. □

Corollary 4.3. *A fixed function, F, has ideal collision resistance if the dependence sequence satisfies $\phi_i < n^{-1-\epsilon}$ for $1 \leq i \leq C\sqrt{n}$ for some constant C.*

Proof. For any fixed function, F, let us denote

$$E(X|X \leq \sigma) := \sum_{i=0}^{\sigma} iP(X = i|X \leq \sigma).$$

Now define $E_{X \leq \sigma}(X) = E(X \mid X \leq \sigma)P(X \leq \sigma)$, which is exactly the σth partial sum of $E(X)$. Therefore, $\{E_{X \leq \sigma}\}$ forms a monotonically increasing sequence with index σ. Thus, all "little-oh" bounds from Theorem 4.2 and before hold for the negative parts of $\sum i\Psi_i^{(d)}$, $d \geq 1$. Hence, it suffices to show that, for a random oracle, $E_{X \leq C\sqrt{n}}(X) = \Omega(\sqrt{n})$.

Fix σ to be the smallest integer such that $P(X \leq \sigma) \geq 1/2$. From the discussion at the beginning of this section, $\sigma = C\sqrt{n}$. We need only show that $E(X \mid X \leq \sigma) = \Omega(\sqrt{n})$.

Again, by Markov's inequality, $E(X \mid X \leq \sigma) \geq \alpha P(X \geq \alpha \mid X \leq \sigma)$. We find an α such that $P(X \geq \alpha \mid X \leq \sigma) \geq 1/2$; this is equivalent to finding α such that $P(X < \alpha \mid X \leq \sigma) \leq 1/2$, which for

$\alpha < \sigma$ becomes $P(X < \alpha) \leq [P(X \leq \sigma)]/2$. Since $P(X \leq \sigma) \geq 1/2$, it suffices to find an α such that $P(X < \alpha) = 1/4$.

As in this section's prologue,

$$P(X < \alpha) \leq \frac{1}{n}\binom{\alpha}{2},$$

and we find that α must be roughly $\sqrt{n/2}$ to satisfy our inequality. Therefore, the partial sums of $E(X)$ for a random oracle are $\Omega(\sqrt{n})$, provided that there are at least $\Omega(\sqrt{n})$ summands. Thus, if $\phi_i < n^{-1-\epsilon}$ for $1 \leq i \leq C\sqrt{n}$, $E(X) > E_{X \leq C\sqrt{n}}(X) = \Omega(\sqrt{n})$, and the corollary is complete. □

5. Fixed permutations. Clearly, collision resistance is not a concern for a permutation; however, we can analyze a random walk model similar to the random oracle process of the previous section to simulate the process of deriving a cycle in the permutation. Obviously, having an easily derived short cycle is a weakness for a permutation, since such a cycle may produce a free-start collision in a naive mode using the permutation. With this in mind, we compute the expected number of queries required to form a cycle in an ideal permutation.

Let $F : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be an ideal permutation. Since the properties of an ideal permutation imply that each element in the domain has the same probability of belonging to a cycle of any specified length, we may analyze a random walk starting with the initial value IV, computing $F(IV), F(F(IV)), \ldots, F^i(IV) = IV$.

To determine the probability that $F^i(IV) = IV$, we first count the number of permutations for which $F^j(IV) \neq IV$ for all $0 < j < i$. There are $(n-1)!/(n-i)!$ choices for a prefix of an i-cycle beginning with IV, the choice $F^i(IV) = IV$ is forced, and there are $(n-i)!$ ways of permuting the remaining $n-i$ elements. Therefore, there are $(n-1)!$ permutations in which $F^i(IV) = IV$. Thus, for all i, $P(F^i(IV) = IV) = 1/n$.

To find the expected number of queries, X, required to form a cycle, we calculate:

$$E(X) = \sum_{i=1}^{n} iP(X = i) = \sum_{i=1}^{n} \frac{i}{n} = \frac{n+1}{2}.$$

Definition 5.1. A permutation, F, is said to have the *ideal cycle property* if the number of queries, Q, required to discover a cycle, $F^{(i)}$, is $\Theta(n)$, the number of queries to perform the same task for an ideal permutation.

For an ideal permutation, the hypotheses of Theorem 4.2 are not satisfied. In particular, for each $i \in \{1,\ 2, \ldots, n\}$, consider calculating the dependence sequence, $\{\phi_i\}$, corresponding to the ideal permutation. Again, $\phi_1 = 0$ since there are no events occurring before the first round. For the second dependence coefficient,

$$\phi_2 = \sup_{\substack{A \in \sigma(\xi_1) \\ B \in \sigma(\xi_2)}} |P(B|A) - P(B)|, \quad P(A) > 0.$$

Consider the events $A = [\xi_1 = IV]$ and $B = [\xi_2 \neq IV]$. Clearly, $P(A) = 1/n$. By the above calculation, $P(B) = (n-1)/n$. Notice, however, that $A \cap B = \emptyset$, since if $\xi_1 = IV$, then $\xi_2 = IV$. Therefore, $\phi_2 \geq (n-1)/n$. Although Theorem 4.2 is not applicable, we can, however, still apply the techniques from Section 4 and generalize the results to give us a meaningful result.

In contrast to a random oracle, the outputs of fresh queries to an ideal permutation do not form an i.i.d. sequence of random variables. In particular, the outputs of the permutation are definitively not independent. If we consider the dependence sequence of the previous section in a new light, however, we can generalize this technique to construct a useful measure of variation between the output distributions of a fixed permutation and an ideal permutation.

Note that if the ith input to a random oracle is fresh, then, for all $A \in \sigma(\xi_1, \ldots, \xi_{i-1})$ and $B \in \sigma(\xi_i)$, we have that $P(B \mid A) = P(B)$. Therefore, we can consider the dependence sequence of the previous section as a measure of variation between the distribution of outputs of the fixed function, F, and the random oracle RO. To make this formal, we need to define a sequence of relative dependence coefficients depending on two sequences of random variables.

Definition 5.2. Let F be a fixed function with a finite codomain of size n, and let π be an ideal function. Define ξ_i to be the random sequence of outputs of F under fresh queries. Similarly, define ζ_i to

be the random sequence of outputs of π under fresh queries. Let $\mathcal{F}_J^L := \sigma(\xi_k, J \le k \le L)$ and $\mathcal{G}_J^L := \sigma(\zeta_k, J \le k \le L)$. Define a map $\tau : \mathcal{F}_1^n \to \mathcal{G}_q^n$ by $\tau[(\xi_1, \ldots, \xi_n) \in A] = [(\zeta_1, \ldots, \zeta_n) \in A]$ for any subset, A, of $\{1, \ldots, n\}^n$. For each $i \in \{1, 2, \ldots, n\}$, let $\mathcal{A}_i = \mathcal{F}_1^{i-1}$ where $\mathcal{A}_1$ is the trivial sigma field, let $\mathcal{B}_i = \mathcal{F}_i^i$ and define the relative dependence sequence,

$$\widehat{\phi}_i = \sup_{\substack{A_i \in \mathcal{A}_i \\ B_i \in \mathcal{B}_i}} |P(B_i|A_i) - P(\tau(B_i)|\tau(A_i))|, \tag{5.1}$$

with $P(A_i)P(\tau(A_i)) > 0$. F is said to be $\widehat{\phi}[p]$-mixing if $\widehat{\phi}_i \le n^{-p}$ for all $i < n$.

Note that, if ξ_i and ζ_i are identically distributed and, in addition, ζ_i are also independent, then $\widehat{\phi}_i = \phi_i$. This fact is the analytic expression of the intuition that the measure of dependence ϕ is essentially a measure of variation between the given process and an idealized independent version of the process. As a consequence, $\widehat{\phi}_i = \phi_i$ if τ maps to the sigma field generated by the outputs of a random oracle.

Define an i-singleton event by $A_i = [\xi_1 = a_1, \ldots, \xi_i = a_i] \in \sigma(\xi_1, \ldots, \xi_i)$, and an (i,k)-singleton event by $B_{i,k} = [\xi_{k+1} = b_{k+1}, \ldots, \xi_i = b_i] \in \sigma(\xi_{k+1}, \ldots, \xi_i)$. Clearly, in the case of an ideal permutation, we have that $P(A_i) = (n-i)!/n!$ for all i and $P(B_{i,k}) = (n-i+k)!/n!$ for all $k < i$. Using this relation, we derive the following permutation measuring lemma:

Lemma 5.3. *Fix $1 \le i \le n$. Let F be a fixed permutation. Assume that ξ_i, fresh query outputs for F, and ζ_i, fresh query outputs for an ideal permutation, π, are identically distributed. Let $A_i \in \mathcal{F}_1^{i-1}$ and $B_i \in \mathcal{F}_i^i$. Then $|P(A_i \cap B_i) - n/(n+1-i)P(A_i)P(B_i)| \le P(A_i)\widehat{\phi}_i$.*

Proof. If $P(A_i) = 0$, then the result is obviously valid. If $P(A_i) > 0$, then $P(\tau(A_i)) > 0$; consequently, $|P(B_i|A_i) - P(\tau(B_i)|\tau(A_i))| \le \widehat{\phi}_i$. Note $|P(A_i \cap B_i) - n/(n+1-i)P(A_i)P(B_i)| = |P(A_i \cap B_i) - P(A_i)(nP(B_i))/(n+1-i)|$.

Since $B_i \in \sigma(\xi_i)$, ξ_i and ζ_i are identically distributed, we obtain the relation $P(B_i) = P(\tau(B_i))$. Now $\tau(A_i)$ can be expressed as the disjoint union of a $(i-1)$-singleton events, and $\tau(B_i)$ can be

expressed as a disjoint union of b $(i, i-1)$-singleton events of the form $[\xi_i = k]$, each having probability $1/n$. Therefore, we compute, using the above-mentioned relation for $(i-1)$-singleton events, the quantities $P(\tau(A_i)) = a(n+1-i)!/n!$, and $P(\tau(B_i)) = b/n$.

Moreover, since the event $\tau(A_i) \cap \tau(B_i)$ must be the disjoint union of $a \cdot b$ i-singleton events, it follows that $P(\tau(A_i) \cap \tau(B_i)) = ab(n-i)!/n!$. We, therefore, have that

$$P(\tau(B_i)|\tau(A_i)) = \frac{b}{n+1-i} = \frac{nP(\tau(B_i))}{n+1-i} = \frac{nPr(B_i)}{n+1-i}.$$

We thus obtain

$$|P(A_i \cap B_i) - P(A_i)P(\tau(B_i)|\tau(A_i))| = P(A_i)|P(B_i|A_i) - P(\tau(B_i)|\tau(A_i))|,$$

since $P(A_i) > 0$. This quantity is bounded by $P(A_i)\widehat{\phi}_i$, and, consequently, the lemma holds. □

Using a technique similar to that of Section 4, we compute the probability of obtaining the first cycle on round i. Again, larger probabilities of cycle formation in early rounds result in a lower expected number of rounds to form a cycle; therefore, we compute each probability assuming that every previous round attains the theoretical maximum probability of cycle formation. As in the random oracle case, $\widehat{\phi}_1 = 0$.

The probability of forming a cycle in the first round is $1/n = 1/n + \widehat{\phi}_1$; consequently, the probability of noncollision in the first round is $(n-1)/n - \widehat{\phi}_1$. By Lemma 5.3, the probability of getting a cycle in the second round *and* no cycle in the first round is bounded by

$$P(A)\left(\widehat{\phi}_2 + \frac{n}{n-1}\frac{1}{n}\right),$$

where A represents the event that there is no cycle formed in the first round. This quantity simplifies to $((n-1/n) - \widehat{\phi}_1) \cdot (\widehat{\phi}_2 + (1/n-1))$.

Similarly, for each $i \in \{1,\ 2, \ldots, n\}$, we derive the probability of C_i, the event that the first cycle formation occurs in round i, given that each cycle formation event is assigned the maximal probability sequentially,

$$P(C_i) = \left(\frac{1}{n+1-i} + \widehat{\phi}_i\right) \prod_{1 \le k < i} \left(\frac{n-k}{n+1-k} - \widehat{\phi}_k\right). \tag{5.2}$$

Again, let us consider this expression as a polynomial in $\mathbb{R}[\widehat{\phi}_1]\cdots[\widehat{\phi}_i]$ with $i \in \{1,\ 2, \ldots, n\}$. Let $\Psi_i^{(k)}$ denote the sum of the total degree k terms in this bound. Then we obtain

$$\Psi_i^{(0)} = \frac{1}{n}.$$

We denote the terms in $\Psi_i^{(1)}$ by

$$\Psi_{i,j}^{(1)} = \begin{cases} \frac{(n+1-i)\widehat{\phi}_i}{n} & \text{if } i = j \\ -\frac{(n+1-j)\widehat{\phi}_j}{n(n-j)} & \text{if } j < i. \end{cases}$$

Thus, equation (5.2) becomes:

$$P(C_i) = \Psi_i^{(0)} + \sum_{1\le j\le i} \Psi_{i,j}^{(1)} + \text{higher degree terms.} \tag{5.3}$$

We now derive an explicit lower bound formula for the expected number of rounds, X, required to form a cycle relative to the $\widehat{\phi}_i$'s. We obtain

$$E(X) = \sum_{i=1}^{n} i\Psi_i^{(0)} + \sum_{i=1}^{n} i \sum_{j=1}^{i} \Psi_{i,j}^{(1)} + \text{sum of higher degree terms.} \tag{5.4}$$

From the analysis earlier in this section $\sum_{i=1}^{n} i\Psi_i^{(0)} = (n+1)/2$. The following theorem provides conditions for which the sum of these terms constitutes the dominant term.

Theorem 5.4. *A permutation, F, has the ideal cycle property, provided it is $\widehat{\phi}[1+\epsilon]$-mixing for some $\epsilon > 0$.*

Proof. First, we note that the terms in $\Psi_i^{(d)}$, the sum of the total degree d terms, satisfy:

$$\Psi_{i,j_1,\ldots,j_d}^{(d)} = \begin{cases} (-1)^{d+1} \frac{(n+1-j_1)\cdots(n+1-j_d)\hat{\phi}_{j_1}\cdots\hat{\phi}_{j_d}}{n(n-j_1)\cdots(n-j_{d-1})} & \text{if } j_d = i \\ (-1)^{d} \frac{(n+1-j_1)\cdots(n+1-j_d)\hat{\phi}_{j_1}\cdots\hat{\phi}_{j_d}}{n(n-j_1)\cdots(n-j_d)} & \text{if } j_d < i. \end{cases}$$

By the calculations shown at the beginning of this section, $\sum_{i=1}^{n} i\Psi_i^{(0)} = \Theta(n)$. It therefore suffices to show that the absolute value of the sum of the negative terms, $\sum_{i=d}^{n} i\Psi_i^{(d)}$ with $d > 0$ is $o(n)$.

Now, $\Psi_i^{(d)}$ is negative when either $j_d = i$ and $d > 0$ is even, or $j_d < i$ and d is odd. In particular, since we have taken care of the $d = 0$ case, we consider the absolute value of the sum $\sum_{i=d}^{n} i\Psi_{i,j_1,\ldots,j_d}^{(d)}$ when $j_d < i$ and $d \geq 1$ is odd, and when $j_d = i$ and $d \geq 2$ is even.

Case 1: $j_d = i$ and $d \geq 2$ is even. Under these conditions, we have that

(5.5)
$$\sum_{i=d}^{n} i|\Psi_{i,j_1,\ldots,j_{d-1},i}^{(d)}| \leq \max_{l\leq n} \widehat{\phi}_l^d \sum_{i=1}^{n} \frac{i(n+1-i)}{n} \left(\sum_{j=1}^{n-1} \frac{n+1-j}{n-j} \right)^{d-1}.$$

The first sum on the right-hand side of (5.5) evaluates to

$$\sum_{i=1}^{n} \frac{i(n+1-i)}{n} = \frac{1}{3}\binom{n+1}{2},$$

while the second sum in (5.5) simplifies to

$$\left(\sum_{j=1}^{n-1} \frac{n+1-j}{n-j} \right)^{d-1} = \left(n - 1 + \sum_{k=1}^{n-1} \frac{1}{k} \right)^{d-1} \leq (n + \ln(n))^{d-1}. \tag{5.6}$$

Thus, the right-hand sum in (5.5) is bounded by the quantity

$$\frac{1}{3} \max_{l\leq n} \hat{\phi}_l^d \binom{n+1}{2} (n + \ln(n))^{d-1}.$$

Since, by hypothesis, $\max_{l\leq n} \widehat{\phi}_l < n^{-1-\epsilon}$, we have

$$\frac{1}{3} \max_{l\leq n} \hat{\phi}_l^d \binom{n+1}{2} (n + \ln(n))^{d-1} < \frac{1}{3} n^{-d-d\epsilon} \binom{n+1}{2} (n + \ln(n))^{d-1},$$

which is $O(n^{1-d\epsilon})$, and hence, Case 1 is complete.

Case 2: $j_d < i$ and $d \geq 1$ is odd. Under these conditions, we obtain that

$$\sum_{i=d}^{n} i|\Psi_{i,j_1,\ldots,j_d}^{(d)}| \leq \max_{l\leq n} \widehat{\phi}_l^d \sum_{i=1}^{n} \frac{i}{n} \left(\sum_{j=1}^{n-1} \frac{n+1-j}{n-j} \right)^{d}.$$

Using the bound $\max_{l\leq n} \widehat{\phi}_l < n^{-1-\epsilon}$, this expression simplifies to:

$$n^{-d-d\epsilon} \frac{n+1}{2} (n + \ln(n))^{d} = O(n^{1-d\epsilon}),$$

and Case 2 is complete.

Taking the maximum constant, C, from each of the "big-oh" expressions in the n values of d, we obtain the following bound for the sum of the negative terms in (5.4):

$$C\sum_{d=1}^{n} n^{1-d\epsilon} \leq Cn\frac{1}{n^{\epsilon}-1} = o(n). \tag{5.7}$$

Therefore, the total sum in (5.4) is $\Theta(n)$, and Theorem 5.4 is complete. □

6. Application to indifferentiability. The indifferentiability framework for hash functions provides a security criterion for the replacement of an ideal function with a fixed function. Specifically, we have the following definition based on [**9, 18**].

Definition 6.1. An interactive Turing machine T with oracle access to an ideal primitive $\mathcal{F}$ is said to be $(t_{\mathcal{A}}, t_S, \sigma, \epsilon)$-indifferentiable from an ideal function $\mathcal{G}$ if there exists a simulator S such that, for any distinguisher $\mathcal{A}$, we have:

$$\left|P(\mathcal{A}^{T,\mathcal{F}} \Rightarrow 1) - P(\mathcal{A}^{\mathcal{G},S} \Rightarrow 1)\right| \leq \epsilon.$$

The simulator S is an interactive Turing machine with oracle access to $\mathcal{G}$ running in time at most t_S, the distinguisher $\mathcal{A}$ runs in time at most $t_{\mathcal{A}}$, the number of queries $\mathcal{A}$ is allowed to make is σ, and ϵ is a negligible function of the security parameter of T.

The indifferentiability bound, corresponding to the parameter σ in the definition, provides the order of magnitude of the stopping time for an optimal probabilistic adversarial process distinguishing between the hash mode utilizing an ideal primitive and a truly independent random process. The indifferentiability framework provides the most important theoretical means of verifying the security of a hash mode. In particular, an indifferentiable hash mode, together with an ideal compression function, is resistant to every single-stage generic attack, including any possible generic attack found in the future. As a result, recent years have witnessed a great deal of work devoted to the indifferentiability security of the most significant hash modes. In [**2, 7**], an indifferentia-

bility bound of $n/2$ bits is derived for the hash function BLAKE; the same bound is also derived in [**4, 13, 14, 21**] for Skein, Gröstl, JH and the SHA-3 winner Keccak, respectively.

As an application of the framework developed in the previous sections, we are able to extend any generic indifferentiability proof of hash mode H using an ideal primitive to the case of a $\widehat{\phi}[p]$-mixing primitive. This advancement is due primarily to the observation that the technique advanced in [**20, 21**] can be applied to any indifferentiability proof. Specifically, any indifferentiability proof can be modeled with the use of only three games: Game (RO, S), a game modeling interaction with a variable input length random oracle RO and simulator S, Game (H, ip), a game modeling interaction with the actual hash mode H and ideal primitive ip, and $G1$, a single hybrid game with an input/output distribution identical to that of Game (H, ip) and different from Game (RO, S) on an exhaustively specified set of events defined for each round i, denoted by BAD_i.

The methodology employed in [**20, 21**] recasts the typical quest for a long sequence of games into a purely mathematical problem involving the events BAD_i, denoting the event that some occurrence in the ith round may allow an adversary to determine with which entity she is interacting, and $GOOD_i$, denoting $\bigcap_{j<i} \neg BAD_j$. Here, $\neg BAD_j$ denotes the complement of BAD_j. Specifically, any indifferentiability proof can in this way be reduced to checking that the catalogue of information retrieved by any indifferentiability adversary $\mathcal{A}$ is identical for the two games $G1$ and Game (RO, S) in round i if $GOOD_i$ occurred followed by a simple computation of $\neg GOOD_i$.

For the purpose of analysis of indifferentiability arguments involving $\widehat{\phi}[p]$-mixing primitives one need not be concerned with the specific definition of $GOOD_i$, such as the task of the above standard of indifferentiability proof. The only prerequisite is the relation $GOOD_i = \cap_{k \leq i} GOOD_k$, which necessarily holds for this type of event. We note that the existence of *any* valid indifferentiability proof implies the existence of this collection of events and provides implicit bounds on the probabilities of BAD events, though it would seem to be prudent to include in any indifferentiability analysis an explicit definition of the $GOOD_i$ and BAD_i events, since these are precisely the events relevant to cryptanalysis.

A key concept from this analysis is the fact that, given a fixed view, any adversary $\mathcal{A}$, being a probabilistic Turing machine, has a fixed distribution of state transitions. Consequently, even though, given the same input/output distributions $\mathcal{A}$ may in one instance output 1 while in another instance output 0, the probability of each output is unchanged from one experiment to the next.

Lemma 6.2. *Let* $0 \leq a_1, \ldots, a_k, b_1, \ldots, b_k \leq 1$. *Then* $|\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i| \leq \sum_{i=1}^{k} |a_i - b_i|$.

Proof. The result is clearly true if $k = 1$. Assume the result is true for all values less than k.

$$\begin{aligned} \left| \prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i \right| &= \left| \prod_{i=1}^{k} a_i - b_k \prod_{i=1}^{k-1} a_i + b_k \prod_{i=1}^{k-1} a_i - \prod_{i=1}^{k} b_i \right| \\ &\leq \prod_{i=1}^{k-1} a_i \, |a_k - b_k| + b_k \left| \prod_{i=1}^{k-1} a_i - \prod_{i=1}^{k-1} b_i \right| \\ &\leq |a_k - b_k| + \left| \prod_{i=1}^{k-1} a_i - \prod_{i=1}^{k-1} b_i \right|. \end{aligned} \tag{6.1}$$

Thus, by induction,

$$\left| \prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i \right| \leq \sum_{i=1}^{k} |a_i - b_i| \, .$$

□

Theorem 6.3. *Let* $\mathcal{A}$ *be an indifferentiability adversary interacting with the hybrid game* $G(H, \mathcal{F}_p)$, *which utilizes a* $\widehat{\phi}[p]$*-mixing primitive* $\mathcal{F}_p$, *and* $\text{Game}\,(RO, S)$ *while limited to* σ *queries. If, when interacting with* $\text{Game}\,(RO, S)$ *and the hybrid game* $G(H, ip)$, *which utilizes an ideal primitive* ip, *and* $\mathcal{A}$ *is limited by* σ *queries, we have*

$$\left| P(\mathcal{A}^{G(H,ip)} \Rightarrow 1) - P(\mathcal{A}^{RO,S} \Rightarrow 1) \right| \leq \epsilon(\sigma), \tag{6.2}$$

then the following inequality holds as well:

$$\left| P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1) - P(\mathcal{A}^{RO,S} \Rightarrow 1) \right| \leq 2\sigma n^{-p} + \epsilon(\sigma). \tag{6.3}$$

Proof. The left-hand side of (6.3) can be expanded in the following form:

$$\left|P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1) - P(\mathcal{A}^{G(H,ip)} \Rightarrow 1) + P(\mathcal{A}^{G(H,ip)} \Rightarrow 1) - P(\mathcal{A}^{RO,S} \Rightarrow 1)\right|. \tag{6.4}$$

By (6.2), the equation (6.4) is bounded by

$$\left|P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1) - P(\mathcal{A}^{G(H,ip)} \Rightarrow 1)\right| + \epsilon(\sigma).$$

Now let G_i denote the event $GOOD_i$ for all i. Note that we can split the left summand above into:

$$\begin{aligned}\Big|&P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1|\tau^{-1}(G_\sigma))P(\tau^{-1}(G_\sigma)) \\ &-P(\mathcal{A}^{G(H,ip)} \Rightarrow 1|G_\sigma)P(G_\sigma) \\ &+P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1|\neg\tau^{-1}(G_\sigma)P(\neg\tau^{-1}(G_\sigma)) \\ &-P(\mathcal{A}^{G(H,ip)} \Rightarrow 1|\neg G_\sigma)P(\neg G_\sigma)\Big|.\end{aligned} \tag{6.5}$$

By Lemma 6.2 and the triangle inequality, it suffices to obtain bounds on the following three quantities:

$$\left|P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1|\tau^{-1}(G_\sigma)) - P(\mathcal{A}^{G(H,ip)} \Rightarrow 1|G_\sigma)\right|,$$
$$\left|P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1|\neg\tau^{-1}(G_\sigma)) - P(\mathcal{A}^{G(H,ip)} \Rightarrow 1|\neg G_\sigma)\right|,$$

and

$$\left|P(\tau^{-1}(G_\sigma)) - P(G_\sigma)\right|.$$

By the definition of τ, given either of the event pairs $(\tau^{-1}(G_\sigma), G_\sigma)$ or $(\neg\tau^{-1}(G_\sigma), \neg G_\sigma)$, the view of the adversary in the game $G(H, \mathcal{F}_p)$, $V_{1,\sigma}$ and the view of the adversary in $G(H, ip)$, $V_{2,\sigma}$, are identically distributed. Therefore, since the possible states of $\mathcal{A}$ interacting with $G(H, \mathcal{F}_p)$ and $G(H, ip)$ are identically distributed, the corresponding output distributions are identical. Thus, we have:

$$\left|P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1|\tau^{-1}(G_\sigma)) - P(\mathcal{A}^{G(H,ip)} \Rightarrow 1|G_\sigma)\right| = 0$$

and

$$\left| P(\mathcal{A}^{G(H,\mathcal{F}_p)} \Rightarrow 1 | \neg \tau^{-1}(G_\sigma)) - P(\mathcal{A}^{G(H,ip)} \Rightarrow 1 | \neg G_\sigma) \right| = 0.$$

To analyze the last of the three quantities, recall that G_i has the special property that $G_i = \cap_{k \leq i} G_k$. Therefore,

$$P(G_\sigma) = P(G_1) \prod_{i=2}^{\sigma} P(G_i | G_{i-1}),$$

and $\tau^{-1}(G_\sigma)$ satisfies a similar relation. Thus, by Lemma 6.2, $|P(\tau^{-1}(G_\sigma)) - P(G_\sigma)|$ is bounded by

$$\sum_{i=1}^{\sigma} \left| P(\tau^{-1}(G_i) | \tau^{-1}(G_{i-1})) - P(G_i | G_{i-1}) \right| \leq \sum_{i=1}^{\sigma} \widehat{\phi}_i.$$

Therefore, (6.5) is bounded by $2 \sum_{i=1}^{\sigma} \hat{\phi}_i$, which is bounded by $2\sigma n^{-p}$. □

7. Results, limitations, and open questions. Theorem 6.3 of the previous section provides a quantitative means of determining the indifferentiability security of practical hash functions. In particular, any distinguishing attach on a compression function provides bounds on the dependence sequence.

For example, consider an $(t_{\mathcal{A}}, t_S, \sigma, \epsilon)$-indifferentiable hash mode using an n-bit primitive, $\mathcal{F}$, with an indifferentiability security bound of $\sigma = 2^{n/4}$. If one derives a near-collision attack finding a t-bit near-collision on the n-bit primitive in 2^b queries with probability p, this indicates that

$$\sum_{i=1}^{2^b} \widehat{\phi}_i$$

is at least

$$\left| p - 1 + \frac{(2^n - 2^{b+t} - 1)!}{2^{n(2^b-1)}(2^n - 2^{b+t} - 2^b)!} \right|,$$

by [**1**]. If, furthermore, this quantity exceeds $2^{b-n/4+1}$ and $b < log(\sigma)$, then clearly $\mathcal{F}$ is not $\widehat{\phi}[1/4]$-mixing, and the hash function does not achieve a practical indifferentiability security bound of $2^{n/4}$. This result

on hash security depends only on Theorem 6.3 and an analysis of the primitive.

On the other hand, Theorem 6.3 is valid only if the primitive is treated as a black box function. As a result, differential cryptanalyses on reduced-iteration variants of primitives such as [**12**] cannot be applied to find quantitative bounds on indifferentiability security. For any meaningful full-iteration attack, however, Theorem 6.3 applies to provide bounds on the indifferentiability security.

The obvious question raised by this result asks if we can determine bounds on the dependence coefficients of important hash primitives such as the Keccak compression function, see [**3**]. In particular, we can approach the problem from the reverse direction and ask whether we can determine relations which show the dependence coefficients to be large enough to affect the indifferentiability results on the modes of these hash functions, as seen in [**4, 13, 14, 21**]. Table 1 summarizes the necessary $\widehat{\phi}$-mixing rate to support the current indifferentiability results for some notable hash functions.

TABLE 1. $\widehat{\phi}$-mixing rate bounds required for the indicated indifferentiability bound. $(*)$ indicates that an explicit description of the events $GOOD_i$ and BAD_i is missing from the literature. $(+)$ indicates that the primitive is a 2–1 compression. The primitives *ro*, *ic* and *ip* are shorthand for random oracle, ideal cipher, and ideal permutation, respectively.

Mode of operation	Primitive input (a)	Message block (ℓ)	Rate (ℓ/a)	Indiff. bound	Prim.	$\widehat{\phi}$-Mixing rate
BLAKE^{*+} [**2,7**]	$4n$	$2n$	0.5	$n/2$	*ic*	$\frac{1}{4}$
FWP [**20**]	$2n$	n	0.5	$2n/3$	*ro*	$\frac{1}{3}$
Groestl* [**14**]	$2n$	n	0.5	$n/2$	*ip*	$\frac{1}{4}$
JH [**21**]	$2n$	n	0.5	$n/2$	*ip*	$\frac{1}{4}$
Keccak* [**4**]	$2n$	n	0.5	$n/2$	*ip*	$\frac{1}{4}$

Another interesting direction to explore is whether these results can be extended in the case of hash functions based on block cipher compression functions to analyze a protocol consisting of the hash mode embedded with the block cipher protocol and using the block cipher round function as the primitive. If such a result can be achieved, then many reduced-iteration attacks in the literature can be used directly to compute dependence coefficients and bound indifferentiability results.

This path seems, however, unlikely to produce a viable result based on the irregularity of such composed protocols.

8. Conclusion. The measures of dependence derived in Sections 4 and 5 provide a measure for fixed functions, determining such a function's proximity to an ideal primitive in some sense. This framework supports the direct analysis of cryptographic primitives as exemplified in Sections 4 and 5 as well as providing a means of porting protocol security arguments involving ideal primitives into a more realistic setting.

Perhaps the most interesting application for these dependence coefficients has been the stopping time analysis for an indifferentiability adversary provided in Section 6. An analysis of the dependence sequence for the primitives of the SHA-3 finalists and the winner of the SHA-3 competition, Keccak, may solidify the indifferentiability arguments already present in the literature. Although the direct calculation of the dependence sequence of such fixed functions is itself a challenging task, such functions are often constructed from block ciphers and other iterative protocols which can be further broken down and analyzed.

On the other hand, the determination of lower bounds for the dependence coefficients of these hash primitives is a more immediately approachable problem. The discovery of particularly large dependence coefficients, while not constituting an invalidation of indifferentiability results on the hash mode, may call the practical resistance to distinguishing attacks of protocols employing such primitives into question. In fact, any nontrivial full-iteration relation on such primitives provides a lower bound on the dependence sequence and hence an upper bound on the indifferentiability security of the hash function. Furthermore, this methodology can be utilized for the construction of new heuristic arguments which may influence the construction of primitive random functions.

REFERENCES

1. M. Abramson and W.O.J. Moser, *More birthday surprises*, Amer. Math. Month. **77** (1970), 856-858.

2. E. Andreeva, A. Luykx and B. Mennink, *Provable security of Blake with non-ideal compression function*, IACR Cryptology ePrint Archive **2011** (2011), 620.

3. G. Bertoni, J. Daemen, M. Peeters and G.V. Assche, *Sponge functions*, Ecrypt Hash Workshop 2007, May 2007.

4. ———, *On the indifferentiability of the sponge construction*, in EUROCRYPT, N.P. Smart, ed., Lect. Notes Comp. Sci. **4965** (2008), 181-197.

5. R.C. Bradley, *Introduction to strong mixing conditions*, Volumes 1, 2 and 3, Kendrick Press, Heber City, Utah, 2007.

6. R.C. Bradley and S.A. Utev, *On second-order properties of mixing random sequences and random fields*, B. Grigelionis, J. Kubilius, H. Pragarauskas and V. Statulevicius, eds., VSP Science Publishers, Utrecht, The Netherlands, and TEV Publishers Service Group, Vilnius, Lithuania, 1994.

7. D. Chang, M. Nandi and M. Yung, *Indifferentiability of the hash algorithm blake*, IACR Crypt. ePrint Archive **2011** (2011), 623.

8. R. Cogburn, *Asymptotic properties of stationary sequences*, Univ. Calif. Publ. Stat. **3** (1960), 99-146.

9. J.-S. Coron, Y. Dodis, C. Malinaud and P. Puniya, *Merkle-damgøard revisited*: *How to construct a hash function*, in CRYPTO, V. Shoup, ed., Lect. Notes Comp. Sci. **3621** (2005), 430-448.

10. A.R. Dabrowski, D. McDonald and U. Rösler, *Renewal theory properties of ion channels*, Ann. Stat. **18** (1990), 1091-1115.

11. Y.A. Davydov, *Mixing conditions for Markov chains*, Theor. Prob. Appl. **18** (1973), 312-328.

12. I. Dinur, O. Dunkelman and A. Shamir, *Self-differential cryptanalysis of up to* 5 *rounds of sha*-3, IACR Crypt. ePrint Arch. **2012** (2012), 672.

13. N. Ferguson, S. Lucks, B. Schneier, et al., *The SKEIN hash function*, The 1st SHA-3 Candidate Conference, Leuven, Belgium, 2009.

14. P. Gauravaram, L. Knudsen, K. Matusiewicz, et al., *Groestl - A SHA*-3 *candidate*, The 1st SHA-3 Candidate Conference, Leuven, Belgium, 2009.

15. D.R. Halverson and G.L. Wise, *Approximately optimal memoryless detection of random signals in dependent noise*, IEEE Trans. Inform. Theor. **424** (1984), 420-424.

16. I.A. Ibragimov, *Some limit theorems for stochastic processes stationary in the stricts sense*, Dokl. Akad. Nauk. **125** (1959), 711-714.

17. H. Kesten and G. Papanicolaou, *A limit theorem for turbulent diffusion*, Comm. Math. Phys. **65** (1979), 97-128.

18. U.M. Maurer, R. Renner and C. Holenstein, *Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology*, in TCC, M. Naor, ed., Lect. Notes Comp. Sci. **2951** (2004), 21-39.

19. S. Meyn and R. Tweedie, *Markov chains and stochastic stability*, Springer-Verlag, New York, 1995.

20. D. Moody, S. Paul and D. Smith-Tone, *Indifferentiability security of the fast widepipe hash*: *Breaking the birthday barrier*, IACR Crypt. ePrint Arch. **2011** (2011), 630.

21. D. Moody, S. Paul and D. Smith-Tone, *Improved indifferentiability security bound for the jh mode*, IACR Crypt. ePrint Arch. **2012** (2012), 278.

22. W. Phillip, *Limit theorems for sums of partial quotients of continued fractions*, Monats. Math. **105** (1998), 195-206.

23. P.C. Phillips, *Regression theory for near-integrated time series*, Econometrica **56** (1988), 1021-1043.

24. M. Rosenblatt, *A central limit theorem and a strong mixing condition*, Proc. Natl. Acad. Sci. **42** (1956), 43-47.

25. ———, *Stationary sequences and random fields*, Birkhauser, Boston, 1985.

26. I.G. Zhurbenko, *The spectral analysis of time series*, North-Holland, Amsterdam, 1986.

National Institute of Standards and Technology, Gaithersburg, MD 20899 and Department of Mathematics, University of Louisville, Louisville, Kentucky 40292
Email address: **daniel.smith@nist.gov, daniel-c.smith@louisville.edu**

Department of Mathematics, University of Louisville, Louisville, Kentucky 40292
Email address: **cristina.tone@louisville.edu**