Basing PRFs on Constant-Query Weak PRFs: Minimizing Assumptions for Efficient Symmetric Cryptography*

Ueli Maurer and Stefano Tessaro

Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland
{maurer,tessaros}@inf.ethz.ch

Abstract. Although it is well known that all basic private-key cryptographic primitives can be built from one-way functions, finding weak assumptions from which practical implementations of such primitives exist remains a challenging task. Towards this goal, this paper introduces the notion of a *constant-query weak PRF*, a function with a secret key which is computationally indistinguishable from a truly random function when evaluated at a *constant* number s of known random inputs, where s can be as small as two.

We provide iterated constructions of (arbitrary-input-length) PRFs from constant-query weak PRFs that even improve the efficiency of previous constructions based on the stronger assumption of a weak PRF (where polynomially many evaluations are allowed).

One of our constructions directly provides a new mode of operation using a constant-query weak PRF for IND-CPA symmetric encryption which is essentially as efficient as conventional PRF-based counter-mode encryption. Furthermore, our constructions yield efficient modes of operation for keying hash functions (such as MD5 and SHA-1) to obtain iterated PRFs (and hence MACs) which rely solely on the assumption that the underlying compression function is a constant-query weak PRF, which is the weakest assumption ever considered in this context.

1 Introduction

1.1 Minimizing Assumptions: Constant-Query Weak PRFs

Most cryptographic security proofs are *reductions*: Under the *assumption* that a primitive P exists, the existence of a second primitive P' is shown by means of a concrete construction that uses an implementation of P (usually in a black-box manner) to implement P'. For example, P' could be a *pseudorandom function* (PRF), i.e. a function with a secret key which is computationally indistinguishable from a truly random function under arbitrary (adaptive) access. These functions are central primitives as they provide a direct solution to the problems of provably secure symmetric encryption and message authentication.

* This research was partially supported by the Swiss National Science Foundation (SNF), project no. 200020-113700/1.

J. Pieprzyk (Ed.): ASIACRYPT 2008, LNCS 5350, pp. 161–178, 2008.

Ideally, one would like the underlying primitive P to be as *weak* as possible, as in practice it is more likely that an efficient and secure candidate is successfully designed. Also, it is a safe practice to assume that already existing cryptographic functions (such as block ciphers or compression functions of hash functions) only fulfill weaker properties than what they have been originally designed for. Sometimes, however, reductions to weak assumptions turn out to be inefficient and involve large security losses (cf. [14] for a typical example), and hence designers of cryptographic systems are frequently confronted with a *trade-off* between the strength of the underlying assumption and the complexity of the resulting construction.

With the aim of proposing new weak assumptions for the purpose of building symmetric-key primitives, this paper introduces the notion of *constant-query weak pseudorandom functions*: Informally, for some constant s, a function $F : \{0,1\}^\kappa \times \{0,1\}^m \to \{0,1\}^n$ with $\kappa < s \cdot n$ is an *s-query weak PRF* (s-WPRF) if $F(K,\cdot)$ (under a secret key K) is indistinguishable from a random function when evaluated at s independent *known* random inputs.[1] This notion weakens significantly the regular concept of a *weak pseudorandom function* (WPRF) [19], where indistinguishability for polynomially many random inputs is required. We point out that a WPRF is by itself already much weaker than a PRF, as it possibly exhibits several non-random properties (such as having weak inputs or being commutative, i.e. $F(k, F(k', x)) = F(k', F(k, x))$). On top of this, an s-WPRF allows for even more structure: For instance, any $s+1$ distinct inputs $x_1, \ldots, x_{s+1}$ and the corresponding outputs $F(k, x_1), \ldots, F(k, x_{s+1})$ under a secret key k may satisfy an easily verifiable relation with no impact on the pseudorandomness of the function.

In this work, we address the problem of using s-WPRFs to construct PRFs. Since s-WPRFs imply the existence of one-way functions, a straightforward construction can be obtained using the results of [13, 14]. However, the inefficiency and the security loss of the resulting reduction make this approach unsuitable for any practical use, even if the underlying s-WPRF is both highly efficient and secure. For this reason, this paper deals with the question of finding *efficient* constructions of PRFs from s-WPRFs: Surprisingly, we are able to provide constructions which are more efficient than existing reductions of PRFs to WPRFs, while only requiring the underlying function to be an s-WPRF, for s as low as two. Furthermore, our constructions are iterated and can process inputs of arbitrary input length. This structure makes them well suited to be derived from properly keyed hash functions with very weak compression functions.

The next two sections are devoted to discussing previous work in the contexts of building PRFs from WPRFs and of iterated PRFs and MACs, respectively, and to relating it to our results.

[1] The assumption that s-WPRFs exist implies the existence of one-way functions, since the mapping $(k, r) \mapsto F(k, r)$ is easily verified to be one-way as long as $\kappa < s \cdot n$. For $\kappa \geq s \cdot n$, such functions can be constructed unconditionally, e.g. using s-wise independent functions. (However, optimal unconditional constructions with $\kappa = s \cdot n$ are not known for all parameters m).

1.2 Construction of PRFs from Weak PRFs

The first construction of a PRF from a WPRF is due to Naor and Reingold [19], and a further construction was later proposed by Maurer and Sjödin [17]. Both assume[2] a *length-preserving* underlying function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ (which can be obtained e.g. from a block cipher) and realize a keyed function mapping ℓ-bit strings to n-bit strings (for a fixed input length ℓ).

The Naor-Reingold Construction [19]. The construction NR_ℓ takes an ℓ-bit input (with ℓ being a power of two) and its secret key consists of 2ℓ n-bit strings $k_{1,0}, k_{1,1}, \ldots, k_{\ell,0}, k_{\ell,1}$. The computation on input $x = (x_1, \ldots, x_\ell)$ proceeds as follows: First, we define $y_i^{(\log \ell + 1)} := k_{i,x_i}$ for all $i = 1, \ldots, \ell$. Then, for all $j = \log \ell, \ldots, 1$ we compute $y_i^{(j)} := F(y_{2i-1}^{(j+1)}, y_{2i}^{(j+1)})$ for all $i = 1, \ldots, 2^{j-1}$ and finally output $y_1^{(1)}$. In other words, the elements of the key corresponding to the individual input bits are chosen as the values of the ℓ leaves of a complete binary tree which is evaluated in a bottom-up fashion by computing the value of each inner vertex as $F(y_l, y_r)$, where y_l and y_r are the values of its children, and finally outputting the value of the root. Hence, one evaluation of the construction needs $\frac{\ell}{2} + \frac{\ell}{4} + \cdots + \frac{\ell}{\ell} = \ell - 1$ calls to the underlying function F. A more involved construction (which we call $\overline{\mathsf{NR}}_{s,\ell}$) by the same authors uses a key consisting of s n-bit values and improves the total number of calls to roughly $\ell / \log s$ per evaluation, but only accepts ℓ and $\log s$ to have the form $2^j + 2$ for some $j \geq 0$. (For both constructions, other input lengths can be achieved through appropriate paddings.)

The IC-Construction [17]. The construction IC_ℓ takes a $(\kappa + 2n)$-bit key consisting of three values $k_1 \in \{0,1\}^\kappa$ and $r, r' \in \{0,1\}^n$. (The value r' can even be made public.) It first precomputes the values $k_i := F(k_{i-1}, r')$ for all $i = 2, \ldots, \ell$. Furthermore, on an ℓ-bit input $x = (x_1, \ldots, x_\ell)$, it sets $y_0 := r$, and for all $j = 1, \ldots, \ell$, computes $y_j := F(k_j, y_{j-1})$ if $x_j = 1$, and $y_j := y_{j-1}$ else. Finally, it outputs y_ℓ. The construction IC_ℓ requires $w(x)$ calls to F when evaluated on input x, where $w(x) \leq \ell$ is the hamming weight of x. If memory restrictions do not allow storage of the keys $k_2, \ldots, k_\ell$, their values have to be computed at each evaluation and thus the construction requires $(\ell - 1) + w(x)$ calls to F per evaluation, which can be as high as $2\ell - 1$.

A central remark is that in order for all the aforementioned constructions to be secure PRFs for adversaries issuing q queries, the underlying WPRF must also be secure when evaluated at q random inputs. (The concrete security bounds for these constructions are discussed in the full version.) Moreover, in this paper we will focus on iterated constructions of PRFs and MACs where candidates for WPRFs may arise from (keyed) compression functions of hash functions, which have the form $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$ (where e.g. $\kappa = 160$ and $n = 512$ for SHA-1). The above constructions can all be extended in a straightforward

[2] In fact, the construction of [19] relies on an intermediate primitive, called a *synthesizer*, but a WPRF $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ is in fact a synthesizer.

way[3] to handle such functions as well, but for the same input length ℓ the number of calls would increase considerably if $n > \kappa$ (roughly, by a factor of $\lceil \frac{n}{\kappa} \rceil$ with respect to the case $n = \kappa$, which is e.g. 4 for SHA-1). This holds even if we just want κ-bit outputs. Hence, this calls for a construction for which the condition $n > \kappa$ does not have a negative impact on the efficiency of the construction.

1.3 Assumptions in Iterated MACs and PRFs

Bellare et al. [2] proposed two efficient message authentication codes called HMAC and NMAC, obtained by appropriately keying an iterated[4] hash functions $H : \{0,1\}^\kappa \times \{0,1\}^* \to \{0,1\}^\kappa$ (where the first input is the initialization value) as $\mathsf{HMAC}(k_1 \| k_2, x) := H(IV, k_2 \| H(IV, k_1 \| x))$ (for a fixed known IV and $|k_1|, |k_2|$ both equal to the block length of H) and as $\mathsf{NMAC}(k_1 \| k_2, x) := H(k_2, H(k_1, x))$, respectively.[5] (Note that HMAC only requires black-box usage of H.) Even though alternative designs of MACs exist (such as CBC-MAC [5] and UMAC [8] to name a few), these constructions have enjoyed widespread usage due to the large availability of hash function implementations (both in hardware and in software). From the theoretical standpoint, security of HMAC/NMAC has been first proved [2] under the assumption that the compression function of H is a PRF (when keyed through the chaining value), and that H is *weakly collision resistant*, i.e. it is hard to find two distinct messages x, x' with $H(K, x) = H(K, x')$ for a secret key K (given oracle access to $H(K, \cdot)$). Bellare [1] subsequently proved HMAC/NMAC to be an arbitrary-input-length PRF under the sole assumption of the compression function being a PRF. We point out that the *cascade construction* by Bellare et al. [3] can also be seen as a way to key a hash function with a single key to obtain a PRF under the same assumption, at the expense of using a prefix-free encoding of the inputs. More recently, Fischlin [12] presented security proofs for HMAC/NMAC (when used as a MAC rather than as a PRF) relying on non-malleability properties of the underlying compression function. A further recent line of research [15, 22] has been concerned with increasing the efficiency of the HMAC/NMAC constructions by imposing slightly stronger requirements on the underlying compression function (i.e. pseudorandomness under mild types of related-key attacks).

The bottom line is that in order to deploy one of these constructions in practice, it is relevant to assess the level of confidence one is willing to put in the given compression function, but in view of continuous cryptanalytic achievements this is far from being a simple task. This issue motivates us to take steps in the

[3] One can simply base the above constructions on the function $F' : (k_1 \| \ldots \| k_c, r) \mapsto F(k_1, r) \| \ldots \| F(k_c, r)$ (possibly chopping some bits) where $c = \lceil n/\kappa \rceil$ (the function F' can be shown to be a WPRF). Note that more involved range-extension techniques (such as those from [11, 17, 20]) do not work here, as they require a length-preserving function beforehand.

[4] i.e. based on the Merkle-Damgård construction [10, 18], cf. also Section 2.

[5] Practical implementations usually consider single-keyed versions which, for simplicity, are not discussed here.

opposite direction: We raise the question of constructing iterated MACs (and PRFs) with *very low* requirements on the given compression function, while guaranteeing limited impact on the performance when compared with constructions with stronger underlying security assumptions. In particular, we consider constructions which only require the underlying compression function to be an s-WPRF (for s as small as two).

1.4 Contributions and Outline of This Paper

This paper initiates the study of constant-query WPRFs, and in particular investigates the problem of constructing efficient PRFs from these primitives.

- In Section 3, we present our first construction (called the RC-construction) of an arbitrary-input-length PRF from any s-WPRF $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$ (for some constant $s \geq 2$). As a special case of our construction, one obtains a fixed-input-length PRF which, for input length ℓ, requires $\approx \frac{\ell}{\log s}$ calls to F per evaluation, hence improving on earlier constructions despite the weaker underlying assumption of an s-WPRF.
- Careful instantiation of the RC-construction yields efficient counter-mode symmetric encryption relying on the sole assumption of an s-WPRF (for some $s \geq 2$), while requiring (on average) only $1 + \frac{1}{s-1}$ calls to F per κ-bit block of encrypted data and minimal storage overhead. Furthermore, the RC-construction directly yields constructions of efficient PRGs from s-WPRFs.
- Section 4 presents a further construction, called the *nested* RC-*construction*, which improves the throughput of the RC-construction for long messages making a novel use of pairwise independence, while still solely relying on the underlying function being an s-WPRF.
- Finally, Section 5 addresses the problem of deriving our constructions by keying iterated hash functions (such as SHA-1 or MD5) whose compression function is an s-WPRF: If minimal (and natural) regularity properties are additionally guaranteed by the compression function, the keying can be done in an entirely black-box way. Furthermore, this is the weakest assumption on the compression function for which modes of operations leading to secure PRFs and MACs have ever been considered.

The basic tools needed in the rest of the paper are reviewed in Section 2.

2 Preliminaries

2.1 Notational Conventions

Throughout this paper, for a set $\mathcal{U}$, we denote as $\mathcal{U}^n$, $\mathcal{U}^*$, and $\mathcal{U}^+$ the sets of *sequences* $s = (u_1, u_2, \ldots, u_{|s|})$ of elements of $\mathcal{U}$ of length $|s| = n$, of arbitrary length with the empty sequence ϵ, and of arbitrary length $|s|$ without the empty sequence ϵ, respectively. (For the case $\mathcal{U} = \{0,1\}$ we usually talk of *strings*.) The notation $s \| s'$ stands for the concatenation of sequences s and s', and u^r is the

sequence $(u, u, \ldots, u)$ consisting of r repetitions of the symbol $u \in \mathcal{U}$. Given a two-argument function $F : \mathcal{U} \times \mathcal{V} \to \mathcal{Y}$ we denote by $F(u, \cdot)$ the function $\mathcal{V} \to \mathcal{Y}$ obtained by fixing the first input to u. Finally, $A^{\mathcal{O}}(r)$ denotes the (oracle) algorithm A which runs on input r with access to the oracle $\mathcal{O}$. Algorithms are in general randomized, and throughout this paper we fix some RAM model of computation for these algorithms. In particular, an algorithm A is said to have *running time* t if the sum of its description length and the worst-case number of steps it takes (counting oracle queries as single steps), taken over all randomness values, all inputs, and all compatible oracles, is at most t.

2.2 Cryptographic Functions

Pseudorandom Functions (PRFs). For some set $\mathcal{X}$ (generally $\mathcal{X} = \{0,1\}^{\ell}$ or $\mathcal{X} = \{0,1\}^*$) we consider *keyed* functions of the form $F : \{0,1\}^{\kappa} \times \{0,1\}^{\rho} \times \mathcal{X} \to \{0,1\}^n$, where the first and the second parameters are called the *public* and the *private* part of the *key*,[6] respectively. The third parameter is the *input* of F. We define the *PRF advantage* of D in distinguishing F from random as the quantity

$$\mathbf{Adv}_F^{\mathsf{PRF}}(D) := \left| \mathsf{P}\left[D^{F(K,R,\cdot)}(R) = 1\right] - \mathsf{P}\left[D^{\mathbf{R}_{\mathcal{X},n}}(R) = 1\right] \right|,$$

where K and R are independent and uniformly chosen from $\{0,1\}^{\kappa}$ and $\{0,1\}^{\rho}$, respectively, whereas $\mathbf{R}_{\mathcal{X},n}$ is a *random function* mapping elements of $\mathcal{X}$ to n-bit strings, i.e. an oracle which associates with each $x \in \mathcal{X}$ a uniformly-distributed independent n-bit string. (Whenever $\mathcal{X}$ is finite, this is equivalent to a randomly chosen function $\mathcal{X} \to \{0,1\}^n$.) For notational convenience we introduce the shorthand $\mathbf{Adv}_F^{\mathsf{PRF}}(t, q)$ to indicate the best advantage taken over all distinguishers with running time t and making at most q queries. Informally, F is a PRF if $\mathbf{Adv}_F^{\mathsf{PRF}}(t, q)$ is "negligible" for all t and q polynomial in some (understood) security parameter.[7] We often consider the case $\mathcal{X} = \{0,1\}^*$: Such a PRF is called an *arbitrary-input-length PRF* (AIL-PRF), and for this case we define $\mathbf{Adv}_F^{\mathsf{PRF}}(t, q, \ell)$ as the maximal advantage taken over all distinguishers with running time t making at most q queries each of length at most ℓ.

Message Authentication Codes (MACs). A keyed function $F : \{0,1\}^{\kappa} \times \{0,1\}^{\rho} \times \{0,1\}^* \to \{0,1\}^n$ is a MAC if it is "unpredictable" under a secret key. Formally, for an adversary A, we define its *MAC advantage* as

$$\mathbf{Adv}_F^{\mathsf{MAC}}(A) := \mathsf{P}[A^{F(K,R,\cdot)}(R) = (x, y) \wedge F(K, R, x) = y \wedge x \text{ new}],$$

where K and R are random independent κ- and ρ-bit strings, respectively, and "x new" means that x was not queried by A to the given oracle. We define $\mathbf{Adv}_F^{\mathsf{MAC}}(t, q, \ell)$ to be the best advantage of an adversary with running time t

[6] We take this unconventional point of view as the constructions of this paper will allow part of the key to be publicly revealed with no harm to their security, and there are settings where this is a useful feature.

[7] If one considers both parts of the key as a single secret key, this implies that F is a PRF according to the usual definition considered in the literature.

issuing at most $q-1$ queries to $F(K,R,\cdot)$, each of length at most ℓ (and the message x output has also length at most ℓ). It is a well-known fact that a secure AIL-PRF is also a good MAC, namely $\mathbf{Adv}^{\mathsf{MAC}}_F(t,q,\ell) \leq \mathbf{Adv}^{\mathsf{PRF}}_F(t',q,\ell) + \frac{1}{2^n}$, where $t \approx t'$.

Weak Pseudorandom Functions (WPRFs). This notion weakens a PRF to only withstand attacks where the function is queried on independent random *known* inputs. (Sometimes, this is called a *known-plaintext attack* (KPA) in the literature.) Formally, for some function g, we let $\mathcal{S}^g$ be the oracle that returns an ordered pair $(r, g(r))$ for a fresh random r each time it is invoked. Then, for a keyed function $F : \{0,1\}^\kappa \times \{0,1\}^m \to \{0,1\}^n$ we define the *WPRF advantage* of the distinguisher D in distinguishing F from random as

$$\mathbf{Adv}^{\mathsf{WPRF}}_F(D) := \left| \mathsf{P}[D^{\mathcal{S}^{F(K,\cdot)}} = 1] - \mathsf{P}[D^{\mathcal{S}^{\mathbf{R}_{m,n}}} = 1] \right|,$$

where $\mathbf{R}_{m,n}$ is a random function mapping m-bit strings to n-bit strings and K is a random κ-bit secret key.[8] Additionally $\mathbf{Adv}^{\mathsf{WPRF}}_F(t,q)$ stands for the best advantage taken over all distinguishers with running time t making at most q queries. For a constant s, we call a function $F : \{0,1\}^\kappa \times \{0,1\}^m \to \{0,1\}^n$ with $\kappa < s \cdot n$ an *s-weak pseudorandom function* (s-WPRF) if $\mathbf{Adv}^{\mathsf{WPRF}}_F(t,s)$ is negligible for all polynomial running times t, and we simply call it a *weak pseudorandom function* (WPRF) if $\mathbf{Adv}^{\mathsf{WPRF}}_F(t,q)$ is negligible for all polynomially bounded t and q.

Cascade and Iterated Hash Functions. For $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$, it is convenient to define its *cascade* $F^* : \{0,1\}^\kappa \times (\{0,1\}^n)^+ \to \{0,1\}^\kappa$ as the function which, on input $k \in \{0,1\}^\kappa$ and $(x_1,\ldots,x_\lambda) \in (\{0,1\}^n)^+$ (with $x_1,\ldots,x_\lambda \in \{0,1\}^n$) first computes $y_0 := k$ and $y_i = F(y_{i-1}, m_i)$ for all $i = 1,\ldots,\lambda$, and subsequently outputs y_λ. In this work we also consider *iterated hash functions* [10, 18] $H : \{0,1\}^* \to \{0,1\}^\kappa$ with underlying *compression function* $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$ (n is generally called the *block length*) and *initialization value* $IV \in \{0,1\}^\kappa$ which are defined such that every input $x \in \{0,1\}^*$ is first padded as $(x_1,\ldots,x_\lambda) \in (\{0,1\}^n)^+$ and subsequently the value $F^*(IV,(x_1,\ldots,x_\lambda))$ is output. In general, the last block x_λ contains some padding bits as well as the length of the message (the so-called *MD-strengthening*) to preserve collision resistance of the compression function. Examples of such functions are those from the MD and the SHA families.

Universal Hashing. Let $H : \{0,1\}^\kappa \times \{0,1\}^* \to \{0,1\}^n$, and let $\delta : \mathbb{N} \to \mathbb{R}^+$. We say that H is *δ-almost universal* (δ-AU) if

$$\mathsf{p}^{\mathsf{COLL}}_H(x,x') := \mathsf{P}[H(K,x) = H(K,x')] \leq \delta(\max\{|x|,|x'|\})$$

for all distinct $x, x' \in \{0,1\}^*$, where K is a randomly chosen κ-bit key. We stress that we extend the standard notion [9, 21] to deal with arbitrary input lengths

[8] In contrast to the definitions of PRFs and MACs, here we only consider a fully-secret key.

by letting δ be a function of the message length. The following lemma extends to the arbitrary-input-length case the well-known fact that δ-AU hash functions can be used to extend the domain of PRFs. (We omit its proof which follows the lines of the fixed-input-length case.)

Lemma 1. *Let $H : \{0,1\}^\kappa \times \{0,1\}^* \to \{0,1\}^m$ be δ-AU, and let $F : \{0,1\}^\kappa \times \{0,1\}^\rho \times \{0,1\}^m \to \{0,1\}^n$ be a keyed function. Define $HF : \{0,1\}^{\kappa+\kappa'} \times \{0,1\}^\rho \times \{0,1\}^* \to \{0,1\}^n$ such that $HF(k\|k', r, x) := F(k', r, H(k, x))$. Then we have*

$$\mathbf{Adv}^{\mathsf{PRF}}_{HF}(t, q, \ell) \leq \mathbf{Adv}^{\mathsf{PRF}}_{F}(t', q) + \tfrac{1}{2} \cdot q^2 \cdot \delta(\ell),$$

where $t' = t + q \cdot t_H(\ell)$, with $t_H(\ell)$ being the time needed to evaluate H on inputs of length at most ℓ.

3 The Randomized Cascade Construction

3.1 Description and Security of the Construction

In this section, we present the first iterated construction of this paper. It is reminiscent of the cascade construction of Bellare et al. [3], but only requires the underlying function $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$ to be an s-WPRF with $s \geq 2$ being a parameter of the construction. As in [3], the construction relies on the concept of a prefix-free encoding, which we briefly introduce.

Prefix-free Encodings. For a set $\mathcal{X}$, the efficiently computable function $\mathsf{ENC} : \mathcal{X} \to \{1, \ldots, s\}^+$ (i.e. outputting a non-empty sequence of elements of $\{1, \ldots, s\}$) is a *prefix-free encoding scheme* if for all distinct $x, x' \in \mathcal{X}$ the sequence $\mathsf{ENC}(x)$ is not a prefix of the sequence $\mathsf{ENC}(x')$. (In particular, ENC must be injective.) If $\mathcal{X} = \{0,1\}^*$, a prefix-free encoding scheme is e.g. obtained by encoding canonically the input as a sequence in $\{1, \ldots, s-1\}^*$, and then appending the symbol s to the sequence. Other variants exist, but it is generally desirable that ENC operates *on-line*, i.e. the encoding is progressively output while the input bits are provided, without the need to know the entire input before starting the encoding process. If $\mathcal{X} = \{0,1\}^\ell$ for some fixed ℓ, then prefix-freeness is achieved "for free" by encoding all inputs as sequences in $\{1, \ldots, s\}^*$ of equal length $\lceil \frac{\ell}{\log_2 s} \rceil$.

Construction. The *randomized cascade construction* with parameter s and input set $\mathcal{X}$ (where usually either $\mathcal{X} = \{0,1\}^*$ or $\mathcal{X} = \{0,1\}^\ell$ for a fixed ℓ) for the function F and prefix-free encoding scheme ENC, denoted $\mathsf{RC}^F_{s,\mathcal{X},\mathsf{ENC}}$, is a mapping $\{0,1\}^\kappa \times \{0,1\}^{sn} \times \mathcal{X} \to \{0,1\}^\kappa$: It takes a key consisting of a κ-bit private part k and an sn-bit long public part, which is interpreted as the concatenation of s n-bit strings $r_1, \ldots, r_s$. On input $x \in \mathcal{X}$, the κ-bit output is computed through the following two steps:

1. Compute $\mathsf{ENC}(x) = (m_1, \ldots, m_\lambda) \in \{1, \ldots, s\}^+$;
2. Output $F^*(k, (r_{m_1}, \ldots, r_{m_\lambda}))$.

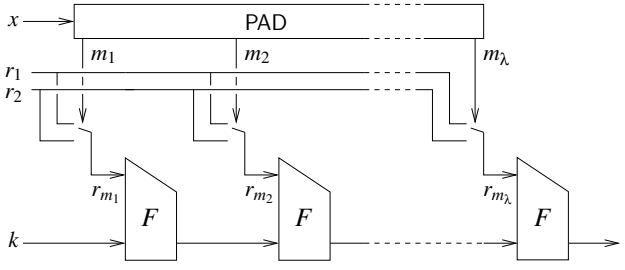

Fig. 1. The construction $\mathsf{RC}^F_{2,\mathsf{ENC}}$

As an example, the construction is depicted in Figure 1 for the special case $s = 2$. For notational convenience, we use the shorthands $\mathsf{RC}^F_{s,\mathsf{ENC}}$ for $\mathcal{X} = \{0,1\}^*$ (and omit the prefix-free encoding when it is generally understood from the context), as well as $\mathsf{RC}^F_{s,\ell}$ for $\mathcal{X} = \{0,1\}^\ell$ (where the canonical encoding described above is used). We also generically refer to the construction as the RC-construction.

Efficiency Comparisons. A fair comparison between the RC-construction and previous results can be undertaken for the fixed-input-length construction $\mathsf{RC}_{s,\ell}$ only. In the length-preserving case ($\kappa = n$), the construction $\mathsf{RC}_{\ell,s}$ is comparable to (for the case $s = 2$) the NR- and the IC-constructions in terms of calls to F, and outperforms them for $s > 2$. Furthermore, we obtain the same space-time trade-off of the $\overline{\mathsf{NR}}_{s,\ell}$-construction, but we allow for all possible values of s. Our construction also limits the effects of possibly very long input paddings in the NR- and $\overline{\mathsf{NR}}$-constructions. The efficiency improvement of our construction is however more evident in the case where $n > \kappa$, as even if $s = 2$, the number of calls to F of (the extended versions of) all other constructions is larger at least by a factor $\lceil \frac{n}{\kappa} \rceil$ (the factor is e.g. 4 when instantiating F with the compression function of SHA-1). Finally, because of the iterated structure, efficient sequential evaluation of $\mathsf{RC}_{s,\ell}$ requires (beside sufficient storage for the key material) κ bits only to store the "chaining value".

Security. In order to give precise security bounds for the RC-construction, it is convenient to think of the prefix-free encoding ENC in terms of a (possibly infinite) directed tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ consisting of all sequences $(m_1, \ldots, m_j)$ which are a prefix of $\mathsf{ENC}(x)$ for some input x (in particular, including the encodings themselves and the empty sequence ϵ). Furthermore, for each $(m_1, \ldots, m_j) \in \mathcal{V}$ there exists a directed edge to $(m_1, \ldots, m_j, m_{j+1})$ for all $m_{j+1} \in \{1, \ldots, s\}$ such that $(m_1, \ldots, m_{j+1}) \in \mathcal{V}$. Hence, it is easy to see that ϵ is the root of the directed tree and its leaves are exactly the encodings of the inputs. We provide two examples of such trees in Figure 2.

Every sequence of queries to the RC-construction defines a subtree of $\mathcal{T}$ consisting of the paths from the root to the encodings of the queries: For notational convenience, we define the shorthand $L(x_1, \ldots, x_q)$, for q inputs $x_1, \ldots, x_q$, to be the amount of inner vertices (i.e. vertices which are not leaves) of the sub-tree induced by the evaluations of $x_1, \ldots, x_q$. It is easy to verify that for $\mathsf{RC}_{s,\ell}$ we

have $L(x_1, \ldots, x_q) \leq 1 + q(\lceil \frac{\ell}{\log s} \rceil - 1)$. Also, for the case where the inputs are strings with arbitrary length, we define (always with respect to the understood encoding) $L(q, \ell) := \max_{x_1,\ldots,x_q : |x_i| \leq \ell} L(x_1, \ldots, x_q)$.

Consequently, one can see an interaction with the RC-construction as a process where the tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ defined by ENC is traversed and κ-bit values are assigned to all visited vertices: While the root ϵ is assigned a random κ-bit value, the value of each visited vertex $(m_1, \ldots, m_j)$ is set to $F(z, r_{m_j})$, with z being the value of the parent vertex $(m_1, \ldots, m_{j-1})$. A query with input x is answered with the value at the corresponding leaf $\mathsf{ENC}(x)$. By the definition of an s-WPRF, it is easy to see that evaluating F under some given (pseudo-)random secret key at s independent random inputs produces s pseudorandom outputs,[9] and hence intuitively the above process sets the values of all visited vertices to pseudorandom values (and in particular this holds for the leaves). However, to formalize this intuition, we have to show that it is indeed possible to recycle the same values $r_1, \ldots, r_s$ for each invocation of F.

The following theorem formally captures the main security statement for the RC-construction (for a general input set $\mathcal{X}$).

Theorem 1. *Let $s \geq 2$, let $\mathcal{X}$ be a set, and let $\mathsf{ENC} : \mathcal{X} \to \{1, \ldots, s\}^+$ be a prefix-free encoding scheme. Furthermore, let $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$. For all L and all distinguishers D with running time t and with $L(x_1, x_2, \ldots) \leq L$ for all possible query sequences $x_1, x_2, \ldots \in \mathcal{X}$, there exists a distinguisher $D' = D'(D)$ such that*

$$\mathbf{Adv}^{\mathsf{PRF}}_{\mathsf{RC}^F_{s,\mathcal{X},\mathsf{ENC}}}(D) \leq L \cdot \left[\mathbf{Adv}^{\mathsf{WPRF}}_F(D') + s^2 \cdot 2^{-(n+1)} \right],$$

where D' makes exactly s queries and has running time $t' = t + \mathcal{O}(L \cdot t_F)$, with t_F being the time needed to evaluate F.

In Appendix A we provide a precise description of the distinguisher D', and refer the reader to the full version of this paper for the complete proof.

We remark that the term $s^2 2^{-(n+1)}$ is negligible, as s is assumed to be constant. Combined with the above observations on L, the theorem directly yields the following security bounds for the specialized variants of the RC-construction:

$$\mathbf{Adv}^{\mathsf{PRF}}_{\mathsf{RC}^F_s}(t, q, \ell) \leq L(q, \ell) \cdot \left[\mathbf{Adv}^{\mathsf{WPRF}}_F(t', s) + s^2 \cdot 2^{-(n+1)} \right],$$

$$\mathbf{Adv}^{\mathsf{PRF}}_{\mathsf{RC}^F_{s,\ell}}(t, q) \leq \left[1 + q \left(\left\lceil \frac{\ell}{\log s} \right\rceil - 1 \right) \right] \cdot \left[\mathbf{Adv}^{\mathsf{WPRF}}_F(t'', s) + s^2 \cdot 2^{-(n+1)} \right],$$

with $t' = t + \mathcal{O}(L(q, \ell) \cdot t_F)$ and $t'' = t + \mathcal{O}((1 + q\,(\lceil \ell / \log s \rceil - 1)) \cdot t_F)$.

The most important observation is that all variants of the RC-construction require F to be only an s-WPRF. A minor positive aspect of the randomized cascade construction (if compared with other constructions) is the absence of any q-dependent birthday-like term in the above inequalities. Furthermore, if

[9] Except in the case where two of the random inputs $r_1, \ldots, r_s$ collide, which happens with small probability only.

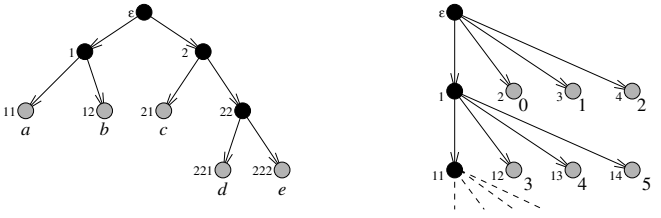

Fig. 2. Example trees associated with prefix-free encodings. Left: Encoding mapping inputs a, b, c, d, and e to sequences $(1,1)$, $(1,2)$, $(2,1)$, $(2,2,1)$, and $(2,2,2)$, respectively. Right: Encoding CTRENC used for efficient counter-mode evaluation.

we assume that F is indeed secure against q queries, the security of the $\mathsf{RC}_{s,\ell}$-construction is comparable to the one of the IC_ℓ-construction if we assume (in fact, very optimistically) that the best WPRF-distinguishing advantage grows linearly in the number of queries, i.e. $\mathbf{Adv}_F^{\mathsf{WPRF}}(t,q) = \Theta(q \cdot \mathbf{Adv}_F^{\mathsf{WPRF}}(t,s))$.

Larger Output Sizes. It is easy to increase the output size of the RC-construction (if needed) with the addition of a minor number of invocations of F per evaluation, which is independent of the input length: To obtain a construction $\overline{\mathsf{RC}}^F : \{0,1\}^\kappa \times \{0,1\}^{ns} \times \mathcal{X} \to \{0,1\}^{\phi\kappa}$ with output size $\phi \cdot \kappa$, we fix ϕ distinct strings $a_1, \ldots, a_\phi \in \mathcal{X}$ such that $L(a_1, \ldots, a_\phi)$ is minimal. Then, given key with private part k and public part $r_1, \ldots, r_s$, on input $x \in \mathcal{X}$, to compute $\overline{\mathsf{RC}}^F(k, r_1 \| \ldots \| r_s, x)$ we first compute $k' := \mathsf{RC}^F(k, r_1 \| \ldots \| r_s, x)$ and finally output $\mathsf{RC}^F(k', r_1 \| \ldots \| r_s, a_1) \| \ldots \| \mathsf{RC}^F(k', r_1 \| \ldots \| r_s, a_\phi)$. Security of this construction can be inferred by the fact that evaluating it at input x accounts to evaluating at inputs $(x, a_1), \ldots, (x, a_\phi)$ a variant of the RC-construction with input set $\mathcal{X} \times \{a_1, \ldots, a_\phi\}$ and prefix-free encoding $\mathsf{ENC}'(x, a) := \mathsf{ENC}(x) \| \mathsf{ENC}(a)$.

3.2 Efficient Encryption and PRGs from the RC-Construction

This section addresses two important applications of the RC-construction. For lack of space, we omit the proofs of the technical claims (which are mostly corollaries of Theorem 1 or are based on standard techniques).

Symmetric Encryption from the RC-Construction. Given a PRF $F : \{0,1\}^\kappa \times \{0,1\}^m \to \{0,1\}^n$ (in practice usually realized by a block cipher) one obtains an efficient stateful IND-CPA[10] encryption scheme for arbitrary-length messages by using F in so-called *counter-mode*, i.e. given a secret key k, we keep a counter ctr (initially 0), and the plaintext x (padded such that $|x|$ is a multiple of n) is encrypted as $[\mathsf{ctr}, x \oplus (F(k, \mathsf{ctr}) \| F(k, \mathsf{ctr}+1) \| \ldots \| F(k, \mathsf{ctr} + |x|/n - 1))]$

[10] Informally, a (stateful or randomized) encryption scheme (E, D) is *IND-CPA secure* [4, 16] if for a secret key K no polynomial-time adversary can distinguish the encryptions $E(K, x_0)$ and $E(K, x_1)$ for any two equally long messages x_0, x_1 of its choice even if it can obtain adaptively chosen encryptions $E(K, x)$ for arbitrary x's.

(and ctr is increased by $|x|/n$), where integers are canonically mapped to m-bit strings. Note in particular that we need one call to F for each n-bit block of encrypted data. Variants of *randomized stateless* counter-mode encryption (where one chooses a fresh random counter at every encryption instead of keeping a state) based on any WPRF $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ were presented in [11, 17]. As with a full PRF, these schemes only require one call per n-bit block of encrypted data, but the underlying WPRF must be secure against as many queries as the amount of encrypted message blocks.

One can substantially weaken the assumption to an s-WPRF by using the RC-construction in stateful counter mode (with any encoding scheme). However, a dramatic increase of efficiency is achieved using a prefix-free encoding scheme $\mathsf{CTRENC} : \mathbb{N} \to \{1,\ldots,s\}^+$ tailored at this mode of operation, defined as

$$\mathsf{CTRENC}(i) := 1^{i \operatorname{div} s-1} \| (2 + (i \bmod s - 1)).$$

The tree arising from this encoding scheme is illustrated in Figure 2: In particular, it is clear that the sequence of values $\mathsf{RC}^F_{s,\mathsf{CTRENC}}(0), \mathsf{RC}^F_{s,\mathsf{CTRENC}}(1), \ldots$ can be computed very efficiently in an iterated way using only $\kappa + sn$ bits of memory and needing approximately $1 + \frac{1}{s-1}$ calls to F per κ-bit block of encrypted data. Furthermore, the values $r_1, \ldots, r_s$ can be chosen publicly by one communicating party (provided an authenticated channel is available), hence reducing the cost of key establishment to the generation of the κ-bit private part of the key. Security against (adaptive) chosen-ciphertext attacks based on any s-WPRF can be then obtained by standard techniques appending a MAC of the ciphertext [7] (e.g. using any of the PRF constructions presented in this paper).

Pseudorandom Generators from s-WPRFs. Recall that a *pseudorandom generator* (PRG) is a length-expanding function $G : \{0,1\}^\kappa \to \{0,1\}^m$ such that $G(K)$ is computationally indistinguishable from a random m-bit string under a random K. Surprisingly, constructing a good PRG from a WPRF (or an s-WPRF) turns out not to be a straightforward task: In contrast to PRFs, a WPRF F does not generally allow to find few "good" inputs $x_1, \ldots, x_t$ such that the mapping $k \mapsto F(k, x_1) \| \ldots \| F(k, x_t)$ is a PRG. However, one can use this approach employing the RC-construction as the underlying PRF: For any t *fixed* inputs $x_1, \ldots, x_t$ ($t > 2$) the mapping $\mathsf{G}^F : \{0,1\}^{sn+\kappa} \to \{0,1\}^{sn+t\kappa}$ such that $\mathsf{G}^F(r_1, \ldots, r_s, k)$ equals

$$r_1 \| \cdots \| r_s \| \mathsf{RC}^F_s(k, r_1 \| \ldots \| r_s, x_1) \| \cdots \| \mathsf{RC}^F_s(k, r_1 \| \ldots \| r_s, x_t)$$

is a PRG if F is an s-WPRF. (The order of the strings in the concatenation is irrelevant.) Note that an important advantage is that the strings $r_1, \ldots, r_s$ can be output as well. For example, given a 2-WPRF $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$, the mapping $\overline{\mathsf{G}}^F : \{0,1\}^{3n} \to \{0,1\}^{6n}$ such that $\overline{\mathsf{G}}^F(k, r_0, r_1)$ is set to

$$r_0 \| F(F(k, r_0), r_0) \| F(F(k, r_0), r_1) \| F(F(k, r_1), r_0) \| F(F(k, r_1), r_1) \| r_1 \qquad (1)$$

is a length-doubling PRG which requires 6 calls to F. In particular, 3 calls are necessary in order to input only one both halves of the output. This improves a construction given in [17], which needed 3 and 4 calls, respectively.

An alternative approach to building a PRF from an s-WPRF F would consist of first constructing a length-doubling PRG G from F, and subsequently using the well-known GGM-construction [13] to build a PRF with a κ-bit key and ℓ-bit inputs by outputting, on input $x = (x_1, \ldots, x_{\ell-1}, x_\ell) \in \{0,1\}^\ell$ and key k, the κ-bit value $G_{x_\ell}(G_{x_{\ell-1}}(\cdots G_{x_1}(k)\cdots))$, where $G_i(k)$ for $i = 0, 1$ gives the first and the second half of the output of G, respectively. However, it is not hard to see that all constructions following this approach turn out to be less efficient than using the RC-construction directly (e.g. using the PRG of Equation 1 one needs 3 calls of F per input bit).

4 The Nested Randomized Cascade Construction

Even though the RC-construction can be practically efficient in special instantiation scenarios discussed earlier, its throughput is a major bottleneck in the case where the construction is used as a PRF (or a MAC) which is invoked at arbitrary inputs with variable lengths. Furthermore, the prefix-free encoding can be a limiting factor in the arbitrary-input-length case. This section presents a construction with better efficiency for long messages (i.e. longer than κ bits) and with no prefix-freeness requirements. Its core ingredient is a novel use of pairwise independence.

Pairwise-Independent Mappings. Recall that a mapping[11] $M : \{0,1\}^\kappa \times \{0,1\}^m \to \{0,1\}^n$ is *pairwise independent* if the values $M(K, x)$ and $M(K, x')$ are independent and uniformly distributed for all distinct $x, x' \in \{0,1\}^m$ under a random κ-bit key K. Most pairwise-independent mappings satisfy the following property, which will be central in our construction.

Definition 1. *A pairwise-independent mapping $M : \{0,1\}^\kappa \times \{0,1\}^m \to \{0,1\}^n$ is* key programmable *if there exists a (possibly randomized) algorithm* SAMPLE *which on input (x, x', y, y') (where possibly $x = x'$, $y = y'$) returns a uniformly chosen element from the set $\{k \mid M(k, x) = y, M(k, x') = y'\}$.*

If M is key programmable, the following two random experiments are equivalent to sampling a random κ-bit key K: (i) For some m-bit string x, sample Y as a uniform random n-bit string and $K := \mathsf{SAMPLE}(x, x, Y, Y)$; and (ii) For n-bit strings $x \neq x'$, sample Y, Y' as independent random n-bit strings and $K := \mathsf{SAMPLE}(x, x', Y, Y')$. Both the last two sampling strategies are used to ensure that $M(K, x) = Y$ (and possibly $M(K, x') = Y'$) for values $Y, Y' \in \{0,1\}^n$ which, although uniform and independent, are provided externally.

We provide two examples of key-programmable pairwise-independent mappings.

Example 1. Let M be such that given $k_1, k_2 \in \{0,1\}^n$ and the input $x \in \{0,1\}^n$, the output $M(k_1 \| k_2, x)$ equals $k_1 \oplus (k_2 \odot x)$, where $\oplus$ and $\odot$ are addition and multiplication of n-bit strings interpreted as elements of the extension field $GF(2^n)$.

[11] We use the word mapping, rather than hash function, to stress the fact that $m = n$ may also hold.

The unique $k_1 \| k_2$ such that $M(k_1 \| k_2, x) = y$ and $M(k_1 \| k_2, x') = y'$ (with $x \neq x'$) can efficiently be found solving the corresponding system of two equalities. Is only a single constraint $M(k_1 \| k_2, x) = y$ given, one chooses a random n-bit string k_2 and sets $k_1 := (k_2 \odot x) \oplus y$.

Example 2. An alternative is the mapping M' whose $(nm + n)$-bit key consists of an $(m \times n)$-binary matrix $\mathbf{A}$ and of a n-dimensional binary column vector $\mathbf{b}$, and on input x the output is $\mathbf{A}x + \mathbf{b}$, where x is interpreted as an m-dimensional column vector, and addition and multiplications are modulo 2. The function M' needs a larger key than M described above, but avoids finite-field multiplications.

Construction. The main idea of the nested RC-construction (called NRC, for short) is to combine an iterated phase where blocks are processed at a higher rate (but which satisfies a property weaker than pseudorandomness) with a second phase where the $\mathsf{RC}_{s,\kappa}$-construction (for fixed input length κ and a parameter s) is invoked on the output of the first phase (with independent key material).

More precisely, let $M : \{0,1\}^{\kappa'} \times \{0,1\}^m \to \{0,1\}^n$ be a key-programmable pairwise-independent mapping and let $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$ be the given compression function. The construction $\mathsf{PI}^F_M : \{0,1\}^{\kappa+\kappa'} \times \{0,1\}^* \to \{0,1\}^\kappa$ takes a key $k \| k'$, where $k \in \{0,1\}^\kappa$ and $k' \in \{0,1\}^{\kappa'}$. On input $x \in \{0,1\}^*$, it pads[12] x as $(x_1, \ldots, x_\lambda)$, where $x_1, \ldots, x_\lambda \in \{0,1\}^m$, and outputs $F^*(k, (M(k', x_1), \ldots, M(k', x_\lambda)))$.

Moreover, given the additional parameter s, we define the nested construction $\mathsf{NRC}^F_{M,s} : \{0,1\}^{2\kappa+\kappa'} \times \{0,1\}^{sn} \times \{0,1\}^* \to \{0,1\}^\kappa$ such that

$$\mathsf{NRC}^F_{M,s}(k_1 \| k_2 \| k', r_1 \| \ldots \| r_s, x) := \mathsf{RC}^F_{s,\kappa}(k_1, r_1 \| \ldots \| r_s, \mathsf{PI}^F_M(k_2 \| k', x)).$$

It is easy to verify that in order to process a message x, the construction needs totally $\left\lceil \frac{|x|+1}{m} \right\rceil + \lceil \frac{\kappa}{\log s} \rceil$ calls to the underlying function F.

It is tempting to increase the throughput of the construction by choosing a mapping M with m much larger than n. However, all known constructions of pairwise-independent hash functions (in particular key-programmable ones) require keys twice as long as the *input* (rather than the output), and hence such an approach would entail a much longer key. In fact, we believe the length-preserving mapping M presented above to be a viable practically efficient solution: This special case of the construction is depicted in Figure 3.

Security. The following theorem precisely quantifies the security of the NRC-construction. We give only a compact statement, as well as an overview of the proof. The complete proof and the concrete reduction arising from it are given in the full version.

[12] According to the canonical padding which pads a string x to have length being a multiple of m by appending a 1 and sufficiently many 0's: The resulting padded string consists hence of $\left\lceil \frac{|x|+1}{m} \right\rceil$ m-bit blocks.

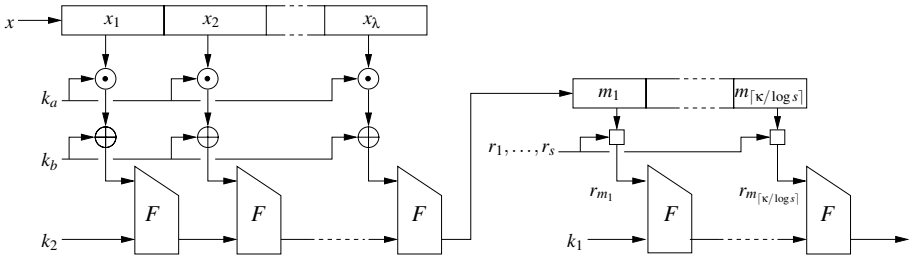

Fig. 3. The construction $\mathsf{NRC}^F_{M,s}$ for the special case $M(k_a \| k_b, x) = (k_a \odot x) \oplus k_b$

Theorem 2. *Let $M : \{0,1\}^{\kappa'} \times \{0,1\}^m \to \{0,1\}^n$ be a key-programmable pairwise-independent mapping, and $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$. For all $s \geq 2$ and for all t, q, and ℓ we have*

$$\mathbf{Adv}^{\mathsf{PRF}}_{\mathsf{NRC}^F_{M,s}}(t,q,\ell) \leq \left(1 + q\left(\left\lceil \tfrac{\kappa}{\log s} \right\rceil - 1\right)\right) \cdot \left(\mathbf{Adv}^{\mathsf{WPRF}}_F(t',s) + s^2 \cdot 2^{-(n+1)}\right) + \left\lceil \tfrac{\ell+1}{m} \right\rceil \cdot q^2 \cdot \left(\mathbf{Adv}^{\mathsf{WPRF}}_F(t'',2) + 2^{-n}\right) + q^2 \cdot 2^{-(\kappa+1)},$$

where $t' = t + \mathcal{O}(q(\frac{\ell}{m} + \frac{\kappa}{\log s}) \cdot t_F)$ and $t'' = \mathcal{O}\left(\frac{2\ell}{m} \cdot t_F\right)$, with t_F being the time needed for an evaluation of F.

The core of the proof consists of showing that whenever F is a WPRF for two-query adversaries, the PI-construction is δ-AU for a suitable function δ to be computed below. In the following, given two inputs x, x' with corresponding padded strings $(x_1, \ldots, x_\lambda)$ and $(x'_1, \ldots, x'_{\lambda'})$ (where without loss of generality $\lambda < \lambda'$), let λ^* be maximal with the property that $x_1 = x'_1, \ldots, x_{\lambda^*} = x'_{\lambda^*}$ (in particular, $\lambda^* := 0$ if $x_1 \neq x'_1$), and define the quantity $\Lambda(x,x')$ as $\lambda + \lambda' - \lambda^* - 1$ if $(x_1, \ldots, x_\lambda)$ is not a prefix of $(x'_1, \ldots, x'_{\lambda'})$, and as $\lambda + 1$ otherwise. In particular, note that $\Lambda(x,x') \leq \lambda + \lambda' \leq 2\max\{\lambda, \lambda'\} \leq 2\lceil \frac{\ell+1}{m} \rceil$ if $|x|, |x'| \leq \ell$.

The following lemma provides a precise upper bound on the collision probability of the PI-construction in terms of the WPRF distinguishing advantage of a distinguisher $D_{x,x'}$ (which in particular only depends on x and x') for F. We refer the reader to the full version of this paper for its proof.

Lemma 2. *For all distinct inputs $x, x' \in \{0,1\}^*$, there exists a two-query distinguisher $D_{x,x'}$ such that*

$$\mathsf{p}^{\mathsf{COLL}}_{\mathsf{PI}^F_M}(x,x') \leq \Lambda(x,x') \cdot \left(\mathbf{Adv}^{\mathsf{WPRF}}_F(D_{x,x'}) + 2^{-n}\right) + 2^{-\kappa},$$

where $D_{x,x'}$ has running time $\mathcal{O}(\Lambda(x,x') \cdot t_F)$.

In particular, given some ℓ, let $t'' = \mathcal{O}\left(\frac{2\ell}{m} \cdot t_F\right)$ be the maximal running time of the distinguisher $D_{x,x'}$ taken over all x, x' with $|x|, |x'| \leq \ell$. We define $\delta(\ell) := 2\left\lceil \frac{\ell+1}{m} \right\rceil \cdot (\mathbf{Adv}^{\mathsf{WPRF}}_F(t'',2) + 2^{-n}) + 2^{-\kappa}$. The function PI^F_M is δ-universal by Lemma 2, and this implies Theorem 2 using Lemma 1 and Theorem 1.

5 Black-Box Keying of Iterated Hash Functions

The iterated structure of the RC- and the NRC-constructions makes compression functions ideal candidates for instantiating the underlying s-WPRF. In general, however, we may be constrained to only have black-box access to an implementation of an iterated hash function $H : \{0,1\}^* \to \{0,1\}^\kappa$ (cf. Section 2) with direct access neither to the initialization value IV nor to the underlying compression function $F : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^\kappa$. To overcome this obstacle, we encode (as in HMAC) an n-bit key as the first block of the input to the hash function H. More precisely, given the prefix-free encoding scheme $\mathsf{ENC} : \{0,1\}^* \to \{1,\ldots,s\}^+$, we consider the construction $\mathsf{HRC}^F_{s,\mathsf{ENC}}$ which takes a key with private part $k \in \{0,1\}^n$ and public part $r_1,\ldots,r_s \in \{0,1\}^n$, and on input x with $\mathsf{ENC}(x) = (m_1,\ldots,m_\lambda)$ outputs the value

$$\mathsf{HRC}^H_{s,\mathsf{ENC}}(k, r_1\|\ldots\|r_s, x) := H(k\|r_{m_1}\|\ldots\|r_{m_\lambda}),$$

and analogously we define $\mathsf{HRC}_{s,\ell}$ for inputs of fixed-length ℓ (using the canonical encoding to the base s). Furthermore, with $M : \{0,1\}^{\kappa'} \times \{0,1\}^m \to \{0,1\}^n$ being a key-programmable pairwise-independent mapping, we consider the construction $\mathsf{HNRC}^H_{M,s}$ which takes a key with private part $k_1, k_2 \in \{0,1\}^n$, $k' \in \{0,1\}^{\kappa'}$ and public parts $r_1,\ldots,r_s$. On input input x (padded as $(x_1,\ldots,x_\lambda)$) it outputs

$$\mathsf{HNRC}^H_{M,s}(k_1\|k_2\|k', r_1\|\ldots\|r_s, x) := \\ \mathsf{HRC}^H_{s,\kappa}(k_1, r_1\|\ldots\|r_s, H(k_2\|M(k', x_1)\|\ldots\|M(k', x_\lambda))).$$

In order to lift the security statements of the RC- and the NRC-constructions to both the HRC- and HNRC-constructions, the assumption that F is an s-WPRF is not sufficient: First, it is necessary that the κ-bit output $F(IV, K)$ is computationally indistinguishable from a uniformly-distributed random string of length κ (under a secret random K); This guarantees that the chaining value obtained after the first evaluation of F is pseudorandom and can be used as the "key" for the RC- or the PI-construction. A further problem is due to the fact that we generally cannot enforce the last n-bit block processed by F to be random because of the padding introduced by H, and this issue should not destroy the pseudorandomness of the outputs. To our rescue, however, comes the fact that each such block is processed keying F with a *fresh* pseudorandom value: It is hence enough to additionally guarantee that for an arbitrary *fixed* n-bit string x and a random secret κ-bit string K, the string $F(K, x)$ is computationally indistinguishable from a random κ-bit string.

We stress that both these extra properties are very weak requirements: In fact, a good compression function should satisfy them even unconditionally. It is sufficient, for example, that $F(IV, \cdot)$ and $F(\cdot, x)$ (for all $x \in \{0,1\}^n$) are all (nearly-)regular functions. (We refer the reader to [6] for a discussion on regularity-properties of hash functions.). With these two additional assumptions on the compression function F of H, the security bounds of the RC and the NRC-construction can be lifted to their black-box counterparts. For lack of space, we omit the proofs, which are very similar to the ones of the original constructions.

6 Conclusions and Open Problems

We have shown that efficient arbitrary-input-length PRFs (and consequently MACs and encryption schemes) can be constructed under very weak assumptions, i.e. weak PRFs where security holds only for a limited number of queries. Our results provide new insights into the property of weak pseudorandomness.

A natural open question is whether there exist constructions of PRFs from WPRFs which take explicit advantage of more secure WPRFs (i.e. tolerating many queries) to achieve more efficient constructions than what we propose and what was considered in the literature (e.g. processing linearly-many bits per invocation even for short inputs). We conjecture, however, that this is not possible. A further direction arising from our work consists of finding further examples of cryptographic primitives where restricting adversaries in terms of queries leads to interesting phenomena such as those observed in this paper for weak pseudorandomness.

References

1. Bellare, M.: New proofs for NMAC and HMAC: Security without collision-resistance. In: Dwork, C. (ed.) CRYPTO 2006. LNCS, vol. 4117, pp. 602–619. Springer, Heidelberg (2006)
2. Bellare, M., Canetti, R., Krawczyk, H.: Keying hash functions for message authentication. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 1–15. Springer, Heidelberg (1996)
3. Bellare, M., Canetti, R., Krawczyk, H.: Pseudorandom functions revisited: The cascade construction and its concrete security. In: FOCS 1996, pp. 514–523 (1996)
4. Bellare, M., Desai, A., Jokipii, E., Rogaway, P.: A concrete security treatment of symmetric encryption. In: FOCS 1997, pp. 394–403 (1997)
5. Bellare, M., Kilian, J., Rogaway, P.: The security of the cipher block chaining message authentication code. Journal of Computer and System Sciences 61(3), 362–399 (2000)
6. Bellare, M., Kohno, T.: Hash function balance and its impact on birthday attacks. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 401–418. Springer, Heidelberg (2004)
7. Bellare, M., Namprempre, C.: Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In: Okamoto, T. (ed.) ASIACRYPT 2000. LNCS, vol. 1976, pp. 531–545. Springer, Heidelberg (2000)
8. Black, J., Halevi, S., Krawczyk, H., Krovetz, T., Rogaway, P.: UMAC: Fast and secure message authentication. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 216–233. Springer, Heidelberg (1999)
9. Carter, J.L., Wegman, M.N.: Universal classes of hash functions. Journal of Computer and System Sciences 18(2), 143–154 (1979)
10. Damgård, I.B.: A design principle for hash functions. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 416–427. Springer, Heidelberg (1990)
11. Damgård, I.B., Nielsen, J.B.: Expanding pseudorandom functions; or: From known-plaintext security to chosen-plaintext security. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 449–464. Springer, Heidelberg (2002)

12. Fischlin, M.: Security of NMAC and HMAC based on non-malleability. In: Malkin, T.G. (ed.) CT-RSA 2008. LNCS, vol. 4964, pp. 138–154. Springer, Heidelberg (2008)
13. Goldreich, O., Goldwasser, S., Micali, S.: How to construct random functions. In: FOCS 1984, pp. 464–479 (1984)
14. Håstad, J., Impagliazzo, R., Levin, L.A., Luby, M.: A pseudorandom generator from any one-way function. SIAM Journal on Computing 28(4), 1364–1396 (1999)
15. Hirose, S., Park, J.H., Yun, A.: A simple variant of the Merkle-Damgård scheme with a permutation. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 113–129. Springer, Heidelberg (2007)
16. Katz, J., Yung, M.: Complete characterization of security notions for probabilistic private-key encryption. In: STOC 2000, pp. 245–254 (2000)
17. Maurer, U., Sjödin, J.: A fast and key-efficient reduction of chosen-ciphertext to known-plaintext security. In: Naor, M. (ed.) EUROCRYPT 2007. LNCS, vol. 4515, pp. 498–516. Springer, Heidelberg (2007)
18. Merkle, R.C.: A certified digital signature. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 218–238. Springer, Heidelberg (1990)
19. Naor, M., Reingold, O.: Synthesizers and their application to the parallel construction of pseudo-random functions. Journal of Computer and System Sciences 58(2), 336–375 (1999)
20. Pietrzak, K., Sjödin, J.: Range extension for weak PRFs; the good, the bad, and the ugly. In: Naor, M. (ed.) EUROCRYPT 2007. LNCS, vol. 4515, pp. 517–533. Springer, Heidelberg (2007)
21. Stinson, D.R.: Universal hashing and authentication codes. In: Feigenbaum, J. (ed.) CRYPTO 1991. LNCS, vol. 576, pp. 74–85. Springer, Heidelberg (1992)
22. Yasuda, K.: Boosting Merkle-Damgård hashing for message authentication. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 216–231. Springer, Heidelberg (2007)

A Description of D' in the Proof of Theorem 1

We define $L+1$ hybrid experiments $H_0, H_1, \ldots, H_L$ where D is given random inputs $r_1, \ldots, r_s$ and interacts which a (randomized) oracle $\mathcal{X} \to \{0,1\}^\kappa$ that keeps track of all vertices of the subtree of $\mathcal{T}$ induced by the queries of D. In particular, it assigns to all *internal* vertices v of this subtree increasing integer values $l(v)$ according to the order in which they are visited for the first time, with $l(\epsilon) := 0$. Furthermore, it associates κ-bit values $z(v)$ with all visited vertices: Initially only $z(\epsilon)$ is defined and set to a random value. In H_i an oracle query $x \in \mathcal{X}$ (with $\mathsf{ENC}(x) = (m_1, \ldots, m_\lambda)$) by D is answered by looking for the highest λ^* such that $z(m_1, \ldots, m_{\lambda^*})$ is defined and for all $j = \lambda^* + 1, \ldots, \lambda$ assigning to $z(m_1, \ldots, m_j)$ a fresh random value if $l(m_1, \ldots, m_{j-1}) < i$ and $F(z(m_1, \ldots, m_{j-1}), r_{m_j})$ otherwise. Finally, $z(m_1, \ldots, m_\lambda)$ is returned to D as the oracle's output. Clearly, H_0 behaves as the experiment where D interacts with the RC-construction, whereas H_L answers all queries of D randomly.

For all $i = 0, \ldots, L-1$ one then constructs a distinguisher D_i for $\mathcal{S}^{F(k,\cdot)}$ and $\mathcal{S}^{\mathbf{R}_{n,\kappa}}$ which first issues s queries to the given oracle, obtaining s pairs $(r_1, y_1), \ldots, (r_s, y_s)$ and subsequently simulates the interaction of D with H_i, except that $z(m_1, \ldots, m_j)$ is set to y_{m_j} whenever $l(m_1, \ldots, m_{j-1}) = i$. Finally, the distinguisher $D'(D)$ chooses a random $i \in \{0, \ldots, L-1\}$ and runs D_i.

We refer the reader to the full version for the concrete analysis of the distinguishing advantage of D'.