Digital Signatures with Minimal Overhead from Indifferentiable Random Invertible Functions

Eike Kiltz[1,⋆], Krzysztof Pietrzak[2,⋆⋆], and Mario Szegedy[3]

[1] Horst-Görtz Institute for IT Security, Ruhr-Universität Bochum, Germany
eike.kiltz@rub.de
[2] Institute of Science and Technology, Austria
pietrzak@ist.ac.at
[3] Rutgers University, USA
szegedy@dragon.rutgers.edu

Abstract. In a digital signature scheme with message recovery, rather than transmitting the message m and its signature σ, a single enhanced signature τ is transmitted. The verifier is able to recover m from τ and at the same time verify its authenticity. The two most important parameters of such a scheme are its security and overhead $|\tau| - |m|$. A simple argument shows that for any scheme with "n bits security" $|\tau| - |m| \geq n$, i.e., the overhead is lower bounded by the security parameter n. Currently, the best known constructions in the random oracle model are far from this lower bound requiring an overhead of $n + \log q_h$, where q_h is the number of queries to the random oracle. In this paper we give a construction which basically matches the n bit lower bound. We propose a simple digital signature scheme with $n + o(\log q_h)$ bits overhead, where q_h denotes the number of random oracle queries.

Our construction works in two steps. First, we propose a signature scheme with message recovery having optimal overhead in a new ideal model, the random invertible function model. Second, we show that a four-round Feistel network with random oracles as round functions is tightly "public-indifferentiable" from a random invertible function. At the core of our indifferentiability proof is an almost tight upper bound for the expected number of edges of the densest "small" subgraph of a random Cayley graph, which may be of independent interest.

Keywords: digital signatures, indifferentiability, Feistel, Additive combinatorics, Cayley graph.

1 Introduction

When transmitting a message m over an unauthenticated public channel, one usually appends a string σ to the message that can be used to verify (relative

[⋆] Funded by a Sofja Kovalevskaja Award of the Alexander von Humboldt Foundation and the German Federal Ministry for Education and Research.

[⋆⋆] Supported by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013) / ERC Starting Grant (259668-PSPC).

R. Canetti and J.A. Garay (Eds.): CRYPTO 2013, Part I, LNCS 8042, pp. 571–588, 2013.

to a public key) the authenticity of the message. This string σ is called a *digital signature of* m. More generally, one transforms the message m into an *enhanced signature* τ such that (i) the original message m can be recovered from τ; (ii) the authenticity of m can be verified from τ. This is called a digital signature scheme with message recovery (MR) and is used to save on bandwidth, i.e., to minimize the *signature overhead* informally defined as $\mathsf{O} = |\tau| - |m|$ (signature length minus message length). Standard bodies for signature schemes (e.g. ISO/IEC 9796 and IEEE P1363a) contain several schemes with MR. In this paper we ask the natural question: *what is the minimal overhead required to achieve a desired security level?*

1.1 Bounds on the Overhead

A Trivial Lower Bound for Every Scheme. Following [3], we say that a signature scheme has "n-bit security" if all adversaries A attacking the scheme have success ratio $\mathsf{SR}(\mathsf{A})$ at most 2^{-n}, where $\mathsf{SR}(\mathsf{A}) := \mathsf{success}(\mathsf{A})/\mathsf{time}(\mathsf{A})$. A natural lower bound for the overhead of a signature scheme (with or without message recovery) for n-bit security is $\mathsf{O} \geq n$ bits. This is since for a signature scheme with O bits of overhead any random bit string τ constitutes a valid enhanced signature with probability $2^{-\mathsf{O}}$. Hence an adversary A guessing a single random authenticated message τ has success ratio $\mathsf{SR}(\mathsf{A}) = 2^{-\mathsf{O}}$ which implies $\mathsf{O} \geq n$.

Overhead of Schemes without MR. In standard digital signature schemes (without message recovery) such as RSA full domain hash [5], the probabilistic signature scheme PSS [5], or (pairing-based) BLS signatures [6] the overhead equals the size of a signature. Since classical signatures contain (at least) one group element (e.g., $\mathbb{Z}_N^*$ or an elliptic curve group) whose representation requires at least $2n$ bits (for n bits security, due to generic square-root attacks) we cannot hope to obtain an overhead smaller than $2n$ bits. The above lower bounds do not apply for schemes without such a group structure, in particular schemes based on lattices or codes, but for other reasons these schemes tend to have a very large overhead and/or prohibitively large public parameters.

Overhead of Schemes with MR in the RO model. Computing the overhead for a given signature scheme turns out to be a bit subtle and depends on the security reduction. We exemplify such a calculation for the RSA-based probabilistic signature scheme with message recovery $\mathsf{PSS\text{-}MR}[n_0, n_1]$ [5], which can be seen as a two-round Feistel construction. $\mathsf{PSS\text{-}MR}[n_0, n_1]$ has an overhead of $n_0 + n_1$ bits, where parameter n_0 controls the randomness and n_1 the amount of added redundancy used during signing. The minimal size of n_0 and n_1 providing a given security level can be computed from the security reduction. The security reduction from [5] in the random oracle model [4] transforms an adversary against $\mathsf{PSS\text{-}MR}[n_0, n_1]$ making q_s (online) signing and q_h (offline) hash queries with success probability $\varepsilon_{\mathsf{PSS\text{-}MR}}$ into an adversary against RSA with success probability $\varepsilon_{\mathsf{RSA}}$ such that $\varepsilon_{\mathsf{PSS\text{-}MR}} = \varepsilon_{\mathsf{RSA}} + \varepsilon_{sim}$, where $\varepsilon_{sim} = (q_s + q_h)^2(2^{-n_0} + 2^{-n_1})$.

An easy computation shows that this implies $\mathsf{O}_{\mathsf{PSS\text{-}MR}} = n_0+n_1 \geq 2n+2\log_2(q_h)$ bits of overhead for n bits security.[1] An improved security reduction by Coron gives $\mathsf{O}_{\mathsf{PSS\text{-}MR}} \geq 2n+\log_2(q_h)+\log_2(q_s)$. Recently, an alternative security reduction for PSS-MR was proposed in [15] demonstrating a tight security reduction for $\mathsf{PSS\text{-}MR}[n_0 = 0, n_1]$ with zero-padding from the (stronger) phi-hiding assumption [7]. However, the required overhead is still $\mathsf{O}_{\mathsf{PSS\text{-}MR}} = n + \log_2(q_h)$ bits, stemming from an additive term $\varepsilon_{sim} = q_h^2/2^{n_1}$ in the security reduction.

The Random Invertible Permutation Model. Besides the popular random-oracle model, signature schemes have also been analyzed in other idealized models. In particular, [16,8] propose a digital signature scheme with message recovery, together with optimal security reduction in the ideal *random invertible permutation model.* Unfortunately, unlike for random oracles, there is no standard cryptographic object which could be used to directly instantiate random invertible permutations over a large domain.[2] In order to get a construction in the random oracle model, one can replace the random invertible permutation $\mathcal{P}$ with some construction $\mathsf{C}^{\mathcal{H}}$ (based on a random oracle $\mathcal{H}$) that is *indifferentiable* [19,10] from $\mathcal{P}$. In the context of signature schemes, already a weaker notion called "public-indifferentiability" [23,11,18] is sufficient. In [18] it is proven that a six-round Feistel network with random round functions is public-indifferentiable from a random invertible permutation. (For full indifferentiability more rounds are needed [14].) Unfortunately, the reduction from [18] is not tight in the oracle query complexity (i.e., the number of queries made by the simulator is quadratic in the number of the queries made by the distinguisher), and as a consequence the required overhead is $\log(q_h)$ bits larger than in the ideal permutation model.

Table 1 summarizes the signature overhead and gives concrete parameters for a typical security parameter of $n = 80$ bits and using 1024/2048-bit RSA. (Parameters for $n \in \{128, 192, 256\}$ can be computed accordingly.) We remark that the table is only valid for sufficiently large messages, i.e., if $|M| \geq 1024 - \mathsf{O}$. For smaller messages standard signatures such as BLS naturally outperform any RSA-based signature scheme with MR.

1.2 Our Contribution

Our main contribution is to revisit and affirmatively answer the question whether there exist signature schemes with minimal overhead in the random oracle model. In a first step we show that such a scheme exists in a new ideal model which we call *random invertible function* model, provided that the ideal functions' image

[1] For n-bit security of $\mathsf{PSS\text{-}MR}[n_0, n_1]$ we require $\mathsf{SR}(\mathsf{A}) \leq 2^{-n+1}$ which is implied by $\varepsilon_{\mathsf{RSA}}/\mathsf{time}(\mathsf{A}) \leq 2^{-n}$ and $\varepsilon_{sim}/\mathsf{time}(\mathsf{A}) \leq 2^{-n}$. With $\mathsf{time}(\mathsf{A}) \geq q_s + q_h$ we obtain $n_0 \geq n + \log_2(q_h)$ and $n_1 \geq n + \log_2(q_h)$ and consequently the overhead is $\mathsf{O} = n_0 + n_1 \geq 2n + 2\log_2(q_h)$.

[2] For fixed small domain, one might use a block-cipher with a fixed key. Though, the heuristic to replace a random permutation with a block-cipher like AES with fixed known keys is not as well analyzed as replacing a random oracle with a strong cryptographic hash function.

Table 1. Overhead of RSA-based signature schemes with message recovery in the random oracle model for n bits security assuming the adversary makes at most q_h hash and q_s signing queries. The table shows the overhead required for $n = 80$ (and only the trivial upper bound $q_h \leq 2^{80}$) and when we additionally assume that the number of random-oracle/signature queries are upper bounded by $q_h \leq 2^{60}$ and $q_s \leq 2^{40}$, respectively. As the $o(\log q_h)$ term in our bound depends on the domain, we give the bounds for 1024 and 2048 bits RSA.

Type	Required overhead O for n bits security			Security
	asymptotic	$n = 80$	$q_h \leq 2^{60}, q_s \leq 2^{40}$	reduction
2-round Feistel	$2n + 2(\log q_h)$	320	280	Bellare-Rogaway [5]
2-round Feistel	$2n + \log(q_h) + \log(q_s)$	320	240	Coron [9]
2-round Feistel	$n + \log(q_h)$	160	140	Kakvi-Kiltz [15]
6-round Feistel	$n + \log(q_h)$	160	140	[16,8]+[18]
4-round Feistel	$n + o(\log q_h)$	97	93	this work (1024-bit RSA)
4-round Feistel	$n + o(\log q_h)$	92	90	this work (2048-bit RSA)

is sufficiently sparse. Next, we show that a Feistel network with four rounds and random oracles as round functions is public-indifferentiable from a random invertible function *with an almost tight reduction.* Combining the two steps, we obtain a new signature scheme with message recovery with almost minimal overhead in the random oracle model.

Signature scheme with MR from random invertible functions. Given a trapdoor permutation $\mathsf{TDP} = (\mathsf{f}, \mathsf{f}^{-1})$ over $\{0,1\}^k$ and an injective function $\mathcal{F} : \{0,1\}^m \rightarrow \{0,1\}^k$ ($k > m$) that can be queried in both directions, we can define a signature scheme with message recovery $\mathsf{SIG\text{-}MR}^{\mathcal{F}}$ as follows. The enhanced signature τ on a message m is defined as $\tau = \mathsf{f}^{-1}(\mathcal{F}(m))$. Signature recovery first evaluates the trapdoor permutation on τ and checks if the result has a valid pre-image or not, i.e., $\{m, \perp\} = \mathcal{F}^{-1}(\mathsf{f}(\tau))$. If the result is not $\perp$, it returns message m. The overhead of $\mathsf{SIG\text{-}MR}^{\mathcal{F}}$ is $\mathsf{O} = k - m$ bits. It is a straightforward generalization of [16,15], to prove that the resulting signature scheme $\mathsf{SIG\text{-}MR}^{\mathcal{F}}$ is tightly secure (losing an additive factor $q_{\mathcal{F}}/2^{k-m}$, where $q_{\mathcal{F}}$ is the number of queries to $\mathcal{F}$) if $\mathcal{F}$ is chosen at random. (The above scheme can only be proved secure assuming TDP is lossy [22]. Using a trick of [16] we can also prove a slightly modified scheme tightly secure assuming TDP is one-way.)

Instantiating Invertible Random Functions with Random Oracles. To instantiate the above scheme in the random oracle model, we must replace the random invertible function $\mathcal{F} : \{0,1\}^m \rightarrow \{0,1\}^k$ with a construction $\mathsf{C}^{\mathcal{H}}$ that is public-indifferentiable from $\mathcal{F}$.

It is easy to construct a random invertible function $\mathcal{F} : \{0,1\}^m \rightarrow \{0,1\}^k$ from a random invertible permutation $\mathcal{P} : \{0,1\}^k \rightarrow \{0,1\}^k$ (by setting $\mathcal{F}(x) = \mathcal{P}(x \| 0^{k-m})$) with a tight reduction. But as discussed above, we do not know how to instantiate $\mathcal{P}$ in the random oracle model without losing at least a quadratic factor in the oracle query complexity [18]. Furthermore, it is well known that a five (or less) round Feistel network cannot be pub-indifferentiable from a random invertible permutation [18].

A formal definition of pub-indifferentiability is given in Definition 1. The important parameters are the error ε_{sim} and the number of queries q_{S} made by the simulator S, which are both functions in the number of queries q_{D} made by the distinguisher D. In order to get a reduction with optimal overhead, i.e., where the security (in bits) is not much smaller than the overhead $\mathsf{O} = k - m$, we need $q_{\mathsf{S}} \approx q_{\mathsf{D}}$ and $\varepsilon_{sim} \approx q_{\mathsf{D}}/2^{k-m}$.

Two Feistel rounds. As a simple warmup example we show that a two-round Feistel network (with random oracles as round functions) is pub-indifferentiable from $\mathcal{F}$ with

$$\varepsilon_{sim} = q_{\mathsf{D}}^2/2^{k-m} \qquad \text{and} \qquad q_{\mathsf{S}} = q_{\mathsf{D}}.$$

The resulting signature scheme (as explained above) requires an overhead of $\mathsf{O} = n + \log_2(q_h)$ to achieve n bits security. This essentially reproves the overhead of PSS-MR obtained in [15].

Four Feistel rounds. As the main technical result of this paper we give a construction $\mathsf{C}_{4F}^{\mathcal{H}}$ based on a four round Feistel network and prove it pub-indifferentiable from $\mathcal{F}$ with

$$\varepsilon_{sim} \leq q_{\mathsf{D}}^{1+o(1)}/2^{k-m} \qquad \text{and} \qquad q_{\mathsf{S}} = \tilde{O}(q_{\mathsf{D}}). \tag{1}$$

Hence the resulting signature scheme has an overhead of $\mathsf{O} = n + o(\log q_h)$ bits, cf. Table 1. The $o(1)$ term can be computed explicitly and for example leads to 97 bits overhead for $n = 80$ bits security if the domain of the TDP is at least 1024 bits (we get smaller overhead if the domain is larger or we put non-trivial bound on q_h). The $o(1)$ term goes to 0 as the ratio of the security we want to achieve, divided by the domain size of the TDP, decreases.

In the proof of (1), the variable $Q(\mu, q) = \max_{\mathcal{X},\mathcal{Z}} |\{(x, z) \,|x \in \mathcal{X}, z \in \mathcal{Z}, z - x \in \mathcal{B}\}|$ (where $\mathcal{B}, \mathcal{X}, \mathcal{Z}$ are q element subsets of $\mathbb{Z}_\mu$ and $\mathcal{B}$ is sampled uniformly at random) will play a central role. This variable has a natural interpretation in graph theoretical terms, it's the number of edges of the densest "small" subgraph of a random (bipartite) Cayley graph (here the Cayley graph has μ vertices on each side, is of degree q and the subgraph has q vertices on each side.) We prove by a compression argument (Corollary 1) an upper bound

$$\text{for each } 0 < a < 1/4 : Q(\mu, \mu^a) \leq \mu^{a+2a^2} \quad \text{(with probability extremely close to 1).} \tag{2}$$

We believe that this bound may be of independent interest. It complements a result of Alon et al. [2, Th. 4] which states that $Q(\mu, \mu^a) \approx \mu^{3a-1}$ for $2/3 < a \leq 1$, i.e. their bound applies to large subgraphs of size $\geq \mu^{2/3}$.

We show (Theorem 5) that the four round Feistel network $\mathsf{C}_{4F}^{\mathcal{H}}$ is pub-indifferentiable form a random invertible function with a simulator making $q_{\mathsf{S}} = \tilde{O}(q_{\mathsf{D}})$ queries and failing with probability $\varepsilon_{sim} = O(\mathsf{E}[Q(\mu, q_{\mathsf{D}})]/2^{k-m})$. Setting $q_{\mathsf{D}} = \mu^a$ in (2) this gives the claimed bound (1) on the pub-indifferentiability of $\mathsf{C}_{4F}^{\mathcal{H}}$.

We leave it is an interesting open problem whether our techniques can be used to prove better bounds for constructions of *permutations* from random oracles. As mentioned above, currently all such constructions suffer from a quadratic increase in the oracle query complexity. Another interesting question is, whether random invertible functions can be used to build chosen-ciphertext secure encryption with optimal overhead. Interestingly, the construction from [1] also uses a four round Feistel network, but the proven security suffers from a quadratic loss in running time.

2 Preliminaries

For $n \in \mathbb{N}$, we write 1^n for the string of n ones, and $[n]$ for $\{1, \ldots, n\}$. $|x|$ denotes the length of a bitstring x, while $|S|$ denotes the size of a set S. $s \leftarrow S$ denotes sampling an element s uniformly at random from the set S. For an algorithm A, we write $z \leftarrow \mathsf{A}(x, y, \ldots)$ to indicate that A is a (probabilistic) algorithm that outputs z on input $(x, y, \ldots)$. In the following we will introduce some basic cryptographic objects that (for simplicity) are defined over bit-strings (rather than arbitrary domains).

2.1 Ideal Primitives and Indifferentiability

Throughout, we use the letter $\mathcal{H}$ to denote a random oracle [4], $\mathcal{P}$ for a random invertible permutation and $\mathcal{F}$ for a random invertible function.

A random oracle $\mathcal{H} : \mathcal{D} \to \mathcal{R}$ with input domain $\mathcal{D} \subset \{0,1\}^*$ and range $\mathcal{R} \subset \{0,1\}^*$ is a function chosen uniformly at random from all functions $\mathcal{D} \to \mathcal{R}$. A random invertible function $\mathcal{F} : \mathcal{D} \to \mathcal{R}$ is a function chosen uniformly at random from all injective functions (i.e., all functions where $x \neq x' \Rightarrow \mathcal{F}(x) \neq \mathcal{F}(x')$). A random invertible permutation $\mathcal{P}$ is a random injective function where $\mathcal{D} \equiv \mathcal{R}$.

Unlike for $\mathcal{H}$, which can only be queried in forward direction, whenever we consider algorithms with oracle access to $\mathcal{F}$ (or $\mathcal{P}$), it is always understood that $\mathcal{F}$ can be queried also in inverse direction. Technically, we can think of $\mathcal{F}$ as being given by two oracles $\mathcal{F}$ and $\mathcal{F}^{-1}$, where $\mathcal{F}^{-1}(\mathcal{F}(x)) = x$ and $\mathcal{F}^{-1}(y) = \perp$ if y is not in the range of $\mathcal{F}$.

Below we define a pub-indifferentiable [11,23] construction of $\mathcal{F}$ from $\mathcal{H}$. The public indifferentiability notion differs from the standard indifferentiability notion [19,10] by the fact that in the public notion the simulator S gets to see all queries made by D.

Definition 1 (pub-indifferentiability). *A $(q_{\mathsf{D}}, q_{\mathsf{S}}, \varepsilon_{sim}, t_{sim})$-public indifferentiable construction of a random invertible function $\mathcal{F}$ from a random oracle $\mathcal{H}$ is a stateless oracle circuit C and a (stateful, probabilistic) simulator S such that for any distinguisher D making at most q_{D} oracle queries, S makes at most q_{S} oracle queries, runs in time at most t_{sim} and the following holds:*

$$|\Pr[\mathsf{D}^{\mathcal{F},\mathsf{S}^{\mathcal{F}}}(1^n) = 1] - \Pr[\mathsf{D}^{\mathsf{C}^{\mathcal{H}},\mathcal{H}}(1^n) = 1]| \leq \varepsilon_{sim},$$

here the second oracle $\mathsf{S}^{\mathcal{F}}$ gets to see also the queries made by $\mathsf{D}^{\mathcal{F},\mathsf{S}^{\mathcal{F}}}$ to the first oracle $\mathcal{F}$.

2.2 Digital Signatures with Message Recovery

A digital signature scheme with message recovery $\mathsf{SIG\text{-}MR} = (\mathsf{G}_{\mathsf{SIG\text{-}MR}}, \mathsf{Sign}, \mathsf{Recover})$ consists of three algorithms and two function families $m(n), k(n)$ describing message space $\{0,1\}^{m(n)}$ and signature space $\{0,1\}^{k(n)}$. Key generation $\mathsf{G}_{\mathsf{SIG\text{-}MR}}$ generates a keypair $(pk, sk) \leftarrow \mathsf{G}(1^n)$ for a secret signing key sk and a public verification key pk. The signing algorithm Sign, on input a message $M \in \{0,1\}^{m(n)}$ and the secret signing key, returns an enhanced signature $\tau \leftarrow \mathsf{Sign}_{sk}(M) \in \{0,1\}^{k(n)}$ of the message. The recovery algorithm $\mathsf{Recover}$ takes a verification key pk and an enhanced signature τ as input and returns $M \leftarrow \mathsf{Recover}_{pk}(\tau)$, where $M \in \{0,1\}^{m(n)} \cup \{\perp\}$. We require that $\Pr[\mathsf{Recover}_{pk}(\mathsf{Sign}_{sk}(M)) = M] = 1$.

The security of the signature scheme can be analyzed in a model where an idealized primitive exists, for example a random oracle or a random invertible function. In that case the adversary and the scheme get access to the idealized primitive $\mathcal{O}$ by making oracle calls.

Security. Let us recall the *existential unforgeability against chosen message attacks* (EUF-CMA) security game [12] relative to the ideal primitive $\mathcal{O}$, played between a challenger and a forger A.

1. The challenger runs $\mathsf{G}_{\mathsf{SIG\text{-}MR}}(1^n)$ to generate a keypair (pk, sk). Forger A receives pk as input.
2. Forger A may ask the challenger to sign a number of messages and evaluate the ideal object $\mathcal{O}$. To query the i-th signature, A submits a message $M_i \in \{0,1\}^{m(n)}$ to the challenger. The challenger returns an enhanced signature τ_i under sk for this message. For the j-th query to $\mathcal{O}$, A submits a query x_j to the challenger who returns the values $\mathcal{O}(x_j)$.
3. Forger A outputs an enhanced signature τ^*.

Let $M^* \leftarrow \mathsf{Recover}(pk, \tau^*)$ be the recovered message of A's forgery. The game outputs 1 (meaning forger A wins the game) if $M^* \neq \perp$ (i.e., τ^* is a valid enhanced signature) and $M^* \neq M_i$ for all i. The success probability of A is the probability that the game outputs 1.

Definition 2 (Security and Overhead of $\mathsf{SIG\text{-}MR}$). *Let $\mathcal{O}$ be an ideal primitive and let $\mathsf{SIG\text{-}MR}^{\mathcal{O}}$ be a signature scheme with message recovery, where $\{0,1\}^{m(n)}$ is the message and $\{0,1\}^{k(n)}$ is the signature space. Let $t_{sig}, q_s, q_o, \varepsilon_{sig}$ be functions of a security parameter n.*

Security: $\mathsf{SIG\text{-}MR}^{\mathcal{O}}$ *is $(t_{sig}, q_s, q_o, \varepsilon_{sig})$-secure relative to $\mathcal{O}$, if all adversaries A running in time at most t_{sig} making at most q_s signing queries and q_o queries to $\mathcal{O}$ (this includes direct queries to $\mathcal{O}$, but also the queries to $\mathcal{O}$ done during evaluation of the signature queries), have success probability at most ε_{sig}. If $\mathcal{O}$ is a random oracle (random invertible function), then we say that $\mathsf{SIG\text{-}MR}^{\mathcal{O}}$ is secure in the random oracle (random invertible function) model.*

n-bit security: *We say $\mathsf{SIG\text{-}MR}^{\mathcal{O}}$ has n bits of security against q_s, q_o queries if it is $(t_{sig}, q_s, q_o, \varepsilon_{sig})$-secure for all $t_{sig}, \varepsilon_{sig}$ satisfying $\varepsilon_{sig}/t_{sig} \leq 2^{-n}$. We simply say it has n bits security if it has n bits security for any q_s, q_o (we can always assume the trivial upper bound $q_o \leq t_{sig} \leq 2^n$.[3])*

Overhead: *The overhead is defined as $k(n) - m(n)$. $\mathsf{O}_{\mathsf{SIG\text{-}MR}^{\mathcal{O}}}(n, q_s, q_o)$ denotes the overhead required in the construction $\mathsf{SIG\text{-}MR}^{\mathcal{O}}$ to reach n bits security against q_s and q_o queries. $\mathsf{O}_{\mathsf{SIG\text{-}MR}^{\mathcal{O}}}(n)$ is short for $\mathsf{O}_{\mathsf{SIG\text{-}MR}^{\mathcal{O}}}(n, 2^n, 2^n)$ (i.e., when putting no upper bounds on q_o, q_s).*

In the following we will propose a scheme with finite message space. In the full version [17] we show how to do domain extension in order to get a scheme that can sign arbitrary longer messages with the same security and overhead.

Using a composition theorem [19], we can express the security of a signature scheme proven secure in the invertible function model when we replace the invertible random function $\mathcal{F}$ with an pub-indifferentiable constructions $\mathsf{C}^{\mathcal{H}}$ as follows.

Theorem 1. *If $\mathsf{SIG\text{-}MR}^{\mathcal{F}}$ is $(t_{sig}, q_s, q_h, \varepsilon_{sig})$-secure in the random invertible function model, and C is a $(q_{\mathsf{D}} = q_h, q_{\mathsf{S}}, \varepsilon_{sim}, t_{sim})$-pub-indifferentiable construction of $\mathcal{F}$ from $\mathcal{H}$ (cf. Def.1), then $\mathsf{SIG\text{-}MR}^{\mathsf{C}^{\mathcal{H}}}$ is $(t_{sig} - t_{sim}, q_s, q_{\mathsf{S}}, \varepsilon_{sig} + \varepsilon_{sim})$-secure in the random oracle model.*

2.3 Trapdoor Permutations

A trapdoor permutation $\mathsf{TDP} = (\mathsf{G}_{\mathsf{TDP}}, \mathsf{f}, \mathsf{f}^{-1})$ over domain $\mathcal{D}(n) = \{0,1\}^{k(n)}$ consists of three ppt algorithms. The key generation algorithm $\mathsf{G}_{\mathsf{TDP}}$ generates a keypair $(ek, td) \leftarrow \mathsf{G}_{\mathsf{TDP}}(1^n)$ of evaluation key and trapdoor. For every (ek, td) in the domain of $\mathsf{G}_{\mathsf{TDP}}(1^n)$, $\mathsf{f}(ek, \cdot)$ and $\mathsf{f}^{-1}(td, \cdot)$ compute permutations $\mathsf{f}_{ek}(\cdot), \mathsf{f}_{td}^{-1}(\cdot)$ on $\{0,1\}^{k(n)}$ s.t. for all $x \in \{0,1\}^{k(n)}$: $\mathsf{f}_{td}^{-1}(\mathsf{f}_{ek}(x)) = x$. We say TDP is homomorphic if $(\mathcal{D}(n), \circ)$ is a group and for all $x_1, x_2 \in \mathcal{D}(n)$, $\mathsf{f}_{ek}(x_1) \circ \mathsf{f}_{ek}(x_2) = \mathsf{f}_{ek}(x_1 \circ x_2)$.

We now recall the security properties of one-wayness and regular lossiness [15,22].

Definition 3 (Security of TDP). *Let $t = t(n)$ and $\varepsilon_{one-way} = \varepsilon_{one-way}(n)$ be functions of a security parameter n. TDP is $(\varepsilon_{one-way}, t)$-one-way if for all adversaries A running in time at most t, $\Pr[\mathsf{A}(ek, \mathsf{f}_{ek}(x)) = x] \leq \varepsilon_{one-way}$, where $(ek, td) \leftarrow \mathsf{G}_{\mathsf{TDP}}(1^n)$, $x \leftarrow \{0,1\}^{k(n)}$.*

Definition 4 (Lossy TDP). *Let $t_{lossy} = t_{lossy}(n)$, $\ell = \ell(n)$ and $\varepsilon_{lossy} = \varepsilon_{lossy}(n)$ be functions of a security parameter n. A trapdoor permutation TDP over domain $\{0,1\}^{k(n)}$ is regular $(\varepsilon_{lossy}, t_{lossy}, \ell)$-lossy if there exists a ppt algorithm G_{lossy} (the lossy key generator) that on input 1^n outputs ek' such that*

[3] As $\varepsilon \leq 1$, $\varepsilon_{sig}/t_{sig} \leq 2^{-n}$ for every $t_{sig} \geq 2^n$, so we only have to look at the case $t_{sig} \leq 2^n$.

1. *(indistinguishability of real and lossy keys) for all adversaries* A *running in time at most* t_{lossy}, $\Pr[\mathsf{A}(ek) = 1] - \Pr[\mathsf{A}(ek') = 1] \leq \varepsilon_{lossy}$, *where* $(ek, td) \leftarrow \mathsf{G}_{\mathsf{TDP}}(1^n)$ *and* $ek' \leftarrow \mathsf{G}_{lossy}(1^n)$*;*
2. *(lossiness)* $\mathsf{f}_{ek'}(\cdot)$ *is* ℓ*-to-*1*, i.e.* $\forall x \in \{0,1\}^{k(n)} : |\{z \ : \ f_{ek'}(z) = f_{ek'}(x)\}| = \ell$

For any $\ell \geq 1$, a lossy trapdoor permutation is collision-resistant when instantiated in lossy mode [22]. The most important example of a trapdoor permutation is RSA with domain $\mathbb{Z}_N^*$, defined as $\mathsf{f}_{N,e}(x) = x^e \bmod N$. It is homomorphic with respect to modular multiplication. It is one-way under the RSA assumption; for any $e < N^{1/4}$ it is furthermore regular e-lossy under the phi-hiding assumption [15], where e is the public RSA exponent. Another example of a (homomorphic and regular lossy) trapdoor function is Paillier [21].

3 Signatures with MR from Random Invertible Functions

Let $k = k(n)$ and $m = m(n)$ be functions with $k(n) \geq m(n)$. Let TDP be a trapdoor permutation over domain $\{0,1\}^k$ and $\mathcal{F} : \{0,1\}^m \to \{0,1\}^k$ be a random invertible function. We build a signature scheme with message recovery $\mathsf{SIG\text{-}MR}^{\mathcal{F}} = (\mathsf{G}_{\mathsf{SIG\text{-}MR}}, \mathsf{Sign}, \mathsf{Recover})$ with message space $\mathcal{M}(n) = \{0,1\}^m$ and signature space $\mathcal{S}(n) = \{0,1\}^k$. $\mathsf{G}_{\mathsf{SIG\text{-}MR}}(1^n)$ runs $(ek, td) \leftarrow \mathsf{G}_{\mathsf{TDP}}(1^n)$. It returns $pk = ek$ and $sk = td$.

```
Algorithm $\mathsf{Sign}_{sk}(M \in \{0,1\}^m)$
    $y := \mathcal{F}(M) \in \{0,1\}^k$
    Return $\tau = \mathsf{f}_{td}^{-1}(y) \in \{0,1\}^k$
```

```
Algorithm $\mathsf{Recover}_{pk}(\tau \in \{0,1\}^k)$
    $y = \mathsf{f}_{ek}(\tau)$
    If $\mathcal{F}^{-1}(y) = \perp$ then return $\perp$
    Else return $M = \mathcal{F}^{-1}(y)$
```

Note that $\mathsf{SIG\text{-}MR}$ has $n_1 = k - m$ bits of redundancy and correctness follows since TDP is a permutation.

The following theorem proves security provided TDP is regular lossy. Its proof is similar to the one of FDH in [15] and postponed to the full version [17].

Theorem 2. *Suppose* TDP *is regular* $(\ell, t_{lossy}, \varepsilon_{lossy})$*-lossy (i.e., lossy by* $\log_2(\ell)$ *bits) and* $\mathcal{F}$ *is a random invertible function from* $\{0,1\}^m$ *to* $\{0,1\}^k$*. Then* $\mathsf{SIG\text{-}MR}^{\mathcal{F}}$ *is* $(t_{sig}, q_s, q_f, \varepsilon_{sig})$ *secure with*

$$t_{sig} \approx t_{lossy}, \quad \varepsilon_{sig} = (2\ell - 1)/\ell \cdot \varepsilon_{lossy} + \frac{q_f}{2^{k-m}}.$$

In case TDP only satisfies the weaker security property of $(t, \varepsilon_{one-way})$-onewayness, we only can obtain a non-tight security reduction [9] with respect to $\varepsilon_{one-way}$. As we will show now, a tight security reduction from one-wayness can be obtained by padding M with one random bit b, using a reduction technique by Katz and Wang [16]. We now define an alternative signature scheme $\mathsf{SIG\text{-}MR}_{\mathrm{ow}}^{\mathcal{F}}$ with message space $\mathcal{M}(n) = \{0,1\}^{m-1}$ which can be proved tightly secure from one-wayness of TDP.

```
Algorithm $\mathsf{Sign}_{sk}(M \in \{0,1\}^{m-1})$
    $b(M) \leftarrow \{0,1\}$
    $y := \mathcal{F}(b\|M) \in \{0,1\}^k$
    Return $\tau = \mathsf{f}_{td}^{-1}(y) \in \{0,1\}^k$
```

```
Algorithm $\mathsf{Recover}_{pk}(\tau \in \{0,1\}^k)$
    $y = \mathsf{f}_{ek}(z)$
    If $\mathcal{F}^{-1}(y) = \perp$ then return $\perp$
    Else compute $b\|M = \mathcal{F}^{-1}(y)$
    Return $M$
```

It is furthermore enforced that Sign always uses the same random bit $b = b(M)$ for message M. (E.g., by defining $b = \mathsf{PRF}_K(M)$.) Note that $\mathsf{SIG\text{-}MR}^{\mathcal{F}}_{\mathrm{ow}}$ has $k - m + 1$ bits redundancy.

The proof of the following theorem is postponed to the full version [17].

Theorem 3. *Suppose* TDP *is homomorphic and* $(t, \varepsilon_{one-way})$*-one-way and* $\mathcal{F}$ *is a random injective function from* $\{0,1\}^m$ *to* $\{0,1\}^k$*. Then the scheme* $\mathsf{SIG\text{-}MR}^{\mathcal{F}}_{ow}$ *is* $(t, q_s, q_f, 2\varepsilon_{one-way} + \frac{q_f}{2^{k-m}})$ *secure.*

4 Pub-Indifferentiable Constructions Based on Feistel Networks

4.1 The Two Round Feistel Network

Consider the two-round construction $\mathsf{C}^{\mathcal{H}}_{2f} : \mathbb{Z}_\mu \to \mathbb{Z}_\mu \times \mathbb{Z}_\rho$ Figure 1 (left) which is derived from an unbalanced two-round Feistel network ϕ_{2f} instantiated with random oracles $\mathcal{H}_1 : \mathbb{Z}_\mu \to \mathbb{Z}_\rho, \mathcal{H}_2 : \mathbb{Z}_\rho \to \mathbb{Z}_\mu$

$$\phi_{2f}(x,v) = (x + \mathcal{H}_2(\mathcal{H}_1(x) + v), \mathcal{H}_1(x) + v)$$
$$\phi_{2f}^{-1}(w,y) = (w - \mathcal{H}_2(y), y - \mathcal{H}_1(w - \mathcal{H}_2(y))$$

$$\text{as} \qquad \mathsf{C}^{\mathcal{H}}_{2f}(x) = \phi_{2f}(x,0) \qquad \mathsf{C}^{\mathcal{H}\,-1}_{2f}(w,y) = \begin{cases} x \text{ if } \phi_{2f}^{-1}(w,y) = (x,0) \\ \perp \text{ otherwise} \end{cases}$$

This will serve as an example of a simple indifferentiability proof and to prepare for our four round Feistel network in the next section.

Theorem 4 (pub-indifferentiability of C_{2f}, implicit in [5]). $\mathsf{C}^{\mathcal{H}}_{2f}$ *as illustrated in Figure 1 (left) is* $(q_{\mathsf{D}}, q_{\mathsf{S}}, \varepsilon_{sim}, t_{sim})$*-pub-indifferentiable from* $\mathcal{F}$ *(cf. Def. 1) where*

$$q_{\mathsf{S}} = q_{\mathsf{D}} \qquad t_{sim} = q_{\mathsf{D}} \cdot polylog(\mu) \qquad \varepsilon_{sim} = q_{\mathsf{D}}^2/\rho,$$

More precisely, we can set $t_{sim} = O(q_{\mathsf{D}} \log(q_{\mathsf{D}}) \log(\mu))$ *using that the cost per (find or insert) operation on a sorted list with* $\leq q_{\mathsf{D}}$ *elements of size* $\log(\mu)$ *bits is* $O(\log(q_{\mathsf{D}}) \log(\mu))$.

The proof of Theorem 4 is postponed to the full version [17]. There we also formally show that a combination with Theorems 2/3 and Theorem 1 leads to the overhead of $\mathsf{O}(n, q_h, q_s) = n + \log(q_h)$ bits for the two schemes $\mathsf{SIG\text{-}MR}^{\mathsf{C}^{\mathcal{H}}_{2f}}[\mathsf{RSA}]$ and $\mathsf{SIG\text{-}MR}^{\mathsf{C}^{\mathcal{H}}_{2f}}_{\mathrm{ow}}[\mathsf{RSA}]$ in the random oracle model.

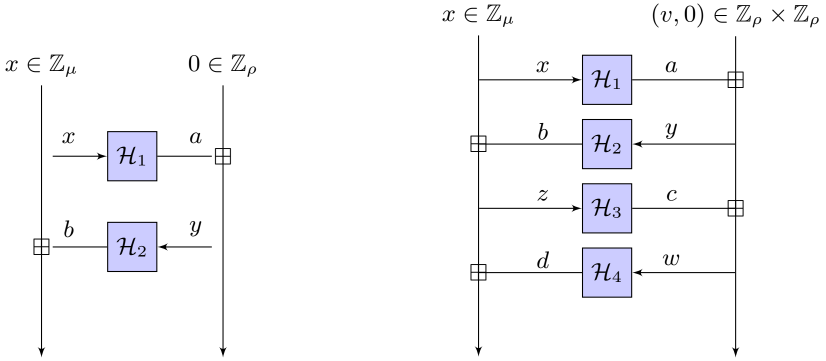

Fig. 1. (left) Two round Feistel network $\phi_{2f} : \mathbb{Z}_\mu \times \mathbb{Z}_\rho \to \mathbb{Z}_\mu \times \mathbb{Z}_\rho$, the construction $\mathsf{C}^{\mathcal{H}}_{2f} : \mathbb{Z}_\mu \to \mathbb{Z}_\mu \times \mathbb{Z}_\rho$ of a random invertible function $\mathcal{F}$ from a random oracle $\mathcal{H}$ is derived from ϕ_{2f} by setting the right part to 0, i.e. $\mathsf{C}^{\mathcal{H}}_{2f}(x) = \phi_{2f}(x, 0)$. $\boxplus$ denotes componentwise addition in the respective domains. **(right)** Four round Feistel network ϕ_{4F}, our main construction is derived from it as $\mathsf{C}^{\mathcal{H}}_{4F}(x, v) = \phi_{4F}(x, v, 0)$.

4.2 The Four Round Feistel Network

We prove the following theorem which bounds the pub-indifferentiability of our main construction $\mathsf{C}^{\mathcal{H}}_{4F}$ as illustrated in Figure 1 (right) in terms of the variable $Q(\mu, q)$ (which we mentioned in the introduction, and will discuss in detail in Section 5).

Theorem 5 (pub-indifferentiability of $\mathsf{C}^{\mathcal{H}}_{4F}$). *$\mathsf{C}^{\mathcal{H}}_{4F}$ as illustrated in Figure 1 (right) is $(q_{\mathsf{D}}, q_{\mathsf{S}}, \varepsilon_{sim}, t_{sim})$-pub-indifferentiable from $\mathcal{F}$ (cf. Def. 1) where*

$$q_{\mathsf{S}} \leq q_{\mathsf{D}} \log(\rho) \qquad t_{sim} = q_{\mathsf{S}} \cdot polylog(\mu)$$

$$\varepsilon_{sim} = \frac{2\mathsf{E}[Q(\mu, q_{\mathsf{D}})]}{\rho} + \frac{2q_{\mathsf{D}}^4}{\mu} + \frac{2q_{\mathsf{D}}^2}{\rho^2} \cdot \left(\frac{\log(\rho)}{\log(\rho/q_{\mathsf{D}})}\right)^2 . \qquad (3)$$

Given Theorem 5 we will now compute the concrete overhead of $\mathsf{SIG\text{-}MR}^{\mathsf{C}^{\mathcal{H}}_{4F}}[\mathsf{RSA}]$ and $\mathsf{SIG\text{-}MR}^{\mathsf{C}^{\mathcal{H}}_{4F}}_{\mathrm{ow}}[\mathsf{RSA}]$. Let $N = pq$ be the RSA modulus with $k = \log N$ and recall that $\log \mu = k - \log \rho$, where $\log \rho$ is the redundancy of the scheme. For all practically relevant values, the first term in ε_{sim} in (3) is the dominating one.[4] With the same argument as in the case of two rounds, by Theorems 2/3 and Theorem 1 the overhead for n-bit security can (up to a small additive constant) be computed as

$$\mathsf{O}(n, q_h, q_s) = n + \log \mathsf{E}[Q(\mu, q_h)] - \log q_h . \qquad (4)$$

[4] Unless one proves an even stronger upper bound on $Q(\mu, q_{\mathsf{D}})$ than we do in this work, in which case the last term might become dominant for large q_{D}.

In order to bound $\mathsf{E}[Q(\mu, q_h)]$ we assume $n \leq \log \rho \leq 1.25n$ and hence $\log \mu = \log N - 2 \log \rho \geq \log N - 2(1.25n)$. The following table summarizes the overhead $\mathsf{O}(n, q_s, q_h)$ for $n = 80$ bits security using (4) and the bounds on $\Pr[Q(\mu, q_h) \geq q_h 2^s]$ from Theorem 6 in Section 5. We use $\log N \in \{1024, 2048\}$ as bit-length of RSA and $\log q_h \in \{60, 80\}$ as upper bound on the random oracle queries.

$\log N$	$\log q_h$	$t = \log \mu$	s	$\Pr[Q(\mu, q_h) \geq q_h 2^s]$	$\mathsf{O}(n, q_h, q_s)$
1024	80	824	17	2^{-427} $(l = 8)$	≈ 97
1024	60	824	13	2^{-430} $(l = 10)$	≈ 93
2048	80	1848	12	2^{-230} $(l = 16)$	≈ 92
2048	60	1848	10	2^{-92} $(l = 18)$	≈ 90

4.3 Proof Intuition

For space reasons, the proof of Theorem 5 is only given in the full version [17] of this paper. In this section we give a high level intuition of the simulator, and in the next section give a proof of a combinatorial result which is at the heart of our proof.

To prove Theorem 5, we have to define a simulator $\mathsf{S}^{\mathcal{F}}$, which is given access to a random function $\mathcal{F} : \mathbb{Z}_\mu \times \mathbb{Z}_\rho \to \mathbb{Z}_\mu \times \mathbb{Z}_\rho \times \mathbb{Z}_\rho$, such that the pair of oracles $(\mathcal{F}, \mathsf{S}^{\mathcal{F}})$ behaves like $(\mathsf{C}_{4F}^{\mathcal{H}}, \mathcal{H})$.

Our simulator will internally define fake random oracles $\hat{\mathcal{H}}_i, i = 1, \ldots, 4$ by lazy sampling, and always try to make sure that they are consistent with $\mathcal{F}$ in the sense that on inputs x on which $\mathcal{F}$ has been queried, the $\hat{\mathcal{H}}_i$ are defined on all values required to evaluate $\mathsf{C}_{4F}^{\hat{\mathcal{H}}}(x)$ and moreover $\mathsf{C}_{4F}^{\hat{\mathcal{H}}}(x) = \mathcal{F}(x)$.

At some point, the simulator might not be able to define the $\hat{\mathcal{H}}$'s consistently any more due to collisions or more complicated linear relations amongst some of the inputs/outputs of $\mathcal{F}$ and the $\hat{\mathcal{H}}$'s. We can easily bound the probability of most such failure events by roughly q_{D}/ρ or less, except for one case, which we'll outline below.

Consider a q_{D} query adversary D who queries an unbalanced three round Feistel network (as illustrated in Figure 1 on the right side, but ignore the last round, and let the right half of the input be $0 \in \mathbb{Z}_\rho$ not $(v, 0) \in \mathbb{Z}_\rho \times \mathbb{Z}_\rho$). Assume the adversary queried the third oracle $\mathcal{H}_3$ on inputs $\mathcal{Z}$ and the second on inputs $\mathcal{Y}$ (receiving outputs $\mathcal{B}$). Next, D chooses some set $\mathcal{X}$ and queries the network on inputs $(x, 0)$. If for some $x \in \mathcal{X}$ we have $\mathcal{H}_1(x) \in \mathcal{Y}$ and $x + \mathcal{H}_2(\mathcal{H}_1(x)) \in \mathcal{Z}$, then the input to $\mathcal{H}_3$ on this query has been already fixed, and the simulator can't program it so it is consistent with $\mathcal{F}(x)$, we'll refer to this as a bad event below.[5] Any tuple $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ where $x + \mathcal{H}_2(y) = z$ can lead to such a

[5] The reason our actual construction needs one more round, is so we can also program the right half of the output of the network. Moreover the input to the right half contains not just the redundancy $0 \in \mathbb{Z}_\rho$, but another element $\mathbb{Z}_\rho$ which is part of the message. The reason we do this is that now the domain of the right half is large enough so we can upper bound by $q_{\mathsf{D}}^2/\rho^2 \leq q_{\mathsf{D}}/\rho$ terms which come up in the proof that depend on the collision probability of random elements over this domain.

failure with probability $\mathbb{Z}_\rho^{-1}$ (namely, if $\mathcal{H}_1(x) = y$). The number of such tuples (for an optimal choice of $\mathcal{X}, \mathcal{Z}$ for a given $\mathcal{B}$) is

$$Q(\mu, q, \mathcal{B}) = \max_{\mathcal{X},\mathcal{Z} \subset \mathbb{Z}_\mu, |\mathcal{X}|=|\mathcal{Z}|=q} |\{(x, z) \mid x \in \mathcal{X}, z \in \mathcal{Z}, z - x \in \mathcal{B}\}| \tag{5}$$

We can thus bound the probability of this bad event by $Q(\mu, q_\mathsf{D}, \mathcal{B})/\rho$. Simple lower and upper bounds on $Q(\mu, q_\mathsf{D}, \mathcal{B})$ are[6]

$$\forall \mathcal{B} \subset \mathbb{Z}_\mu, |\mathcal{B}| = q \ : \ 2q - 1 \leq Q(\mu, q, \mathcal{B}) \leq q^2.$$

Unfortunately, there exist $\mathcal{B}$ for which the upper bound is almost achieved.[7] Fortunately, the set $\mathcal{B}$ is not adversarially chosen, but the output of a random oracle, which makes it a random subset of $\mathbb{Z}_\mu$. We thus consider the variable $Q(\mu, q_\mathsf{D})$ which denotes $Q(\mu, q_\mathsf{D}, \mathcal{B})$ for a randomly chosen $\mathcal{B}$. In order to get a good upper bound on the probability of the bad event, it thus suffices to give an upper bound on $Q(\mu, q_\mathsf{D})$ that holds with high probability over the choice of $\mathcal{B}$. In Section 5 we give such a bound showing that $Q(\mu, q_\mathsf{D})$ is $q_\mathsf{D}^{1+o(1)}$ with very high probability, here the $o(1)$ term goes to 0 as the ratio q_D/μ decreases.

5 A Bit of Additive Combinatorics

Additive combinatorics deals with questions of the sort that given an Abelian group A, find subsets $\mathcal{Z}, \mathcal{X}$ of given size that minimizes the size of

$$\mathcal{Z} - \mathcal{X} = \{z - x | z \in \mathcal{Z}, \ \ x \in \mathcal{X}\}$$

Often we also want to find out the structure of such optimal (or nearly optimal) $\mathcal{Z}, \mathcal{X}$ pairs. Such pairs are of course special, and we do not have too many of them. Analogous questions are also raised when the '$-$' is replaced with '$+$'.

Here we investigate a variant, where we also have a third set $\mathcal{B} \subseteq A$ with the same size as $\mathcal{Z}$ and $\mathcal{X}$ with the property that $z - x \in \mathcal{B}$ for an unusually large number (say, $|\mathcal{B}|^{3/2}$) of (x, z) pairs with $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. We show that for an adequately small random $\mathcal{B}$ it is very unlikely that we can find *any* $\mathcal{Z}, \mathcal{X}$ such that $\mathcal{Z}, \mathcal{X}$ and $\mathcal{B}$ form a triplet as above. We may interpret our result as a property of the random Cayley graph generated by $\mathcal{B}$.

Remark 1. Although our setting is natural and undoubtedly useful for the application at hand, the problem we raise does not seem to have been studied before. An often-cited work of B. J. Green [13] computes the maximum clique

[6] To see the $2q - 1$ lower bound, add any x to $\mathcal{X}$ and let $\mathcal{Z}$ be $\{z \mid z - x \in \mathcal{B}\}$, this gives us already a set of size $|\mathcal{B}| = q$. We can get $q - 1$ more by adding any $q - 1$ elements x to $\mathcal{X}$ s.t. $x = z - b$ for some $b \in \mathcal{B}$, each increasing the set by at least 1. To see the q^2 upper bound, note that for any of the q distinct $x \in \mathcal{X}$, $z - x \in \mathcal{B}$ can hold for at most q z's as $|\mathcal{B}| = q$. This upper bound holds even if we allow $\mathcal{Z}$ (or $\mathcal{X}$) to be all of $\mathbb{Z}_\mu$.

[7] Consider the case $\mathcal{B} = \mathcal{X} = \mathcal{Z} = \{0, \ldots, q-1\}$, which shows $|Q(\mu, q, \mathcal{B})| \geq q(q+1)/2$.

size of (dense) random Cayley graphs of cyclic groups and of $\mathbb{Z}_2^n$. Other authors e.g. Christofides and N. Alon have also investigated random Cayley graphs, but with focus on Hamiltonicity, chromatic number, etc. The size of the generator set, unlike in our case, in most studies are either very small (poly($\log |A|$)) or very large ($\Omega(|A|)$). Since spectra of random Cayley graphs have been studied, it is conceivable that there is a shorter analytic proof to our statement. We use simple combinatorics to prove our theorem.

We (non-crucially) set the Abelian group A to be the cyclic group $\mathbb{Z}_\mu$, where μ is a prime. Let $1 \leq q \leq \mu$ arbitrary, but we will think of it as a small constant power of μ, for instance $q = \mu^{0.1}$. For a set $\mathcal{B} \subset \mathbb{Z}_\mu$, $|\mathcal{B}| = q$ define

$$Q(\mu, q, \mathcal{B}) = \max_{\mathcal{X},\mathcal{Z}\subset\mathbb{Z}_\mu, |\mathcal{X}|=|\mathcal{Z}|=q} |\{(x, z) \mid z \in \mathcal{Z},\ x \in \mathcal{X},\ z - x \in \mathcal{B}\}| \qquad (6)$$

Expression (6) becomes a random variable $Q(\mu, q, .)$ as $\mathcal{B}$ ranges over all uniformly random $\mathcal{B} \subseteq \mathbb{Z}_\mu$ of size q. The minimum value of this random variable is at least q, because for any $\mathcal{B}$ one can choose $\mathcal{Z} = \mathcal{B}$ and $0 \in \mathcal{X}$. We show that if q is a small power of μ, the probability of the event that this random variable much exceeds q is small. To obtain practical expressions in the theorem and simpler formulas in the proof, we introduce $\mu = 2^t$ and $q = 2^r$.

Theorem 6. *For $0 < r < t/4$, and for every $s, l > 0$, $2^s \geq l^2$ it holds that*

$$\Pr[Q(2^t, 2^r) \geq 2^{r+s}] \leq 2^{-DB+t}$$

where $D = \lceil 2^{s-\frac{r}{l}}/(2l + 2)\rceil$ and $B = t - l(r + 1)$.

Corollary 1. *Let $q = \mu^a$, where $a \leq 1/4$. If q is large enough (while parameter a is fixed), then*

$$\Pr[Q(\mu, q) \geq q^{1+2a}] \leq 2^{-q^a/2}$$

We defer the proof of the corollary to after that of the theorem.

Proof. (of Theorem 6) Let $\mu = 2^t$ denote the size of the group, which we assume to be $\mathbb{Z}_\mu$, but this is not essential. We prove Theorem 6 by an information compression argument. What we show is that a set $\mathcal{B}$ satisfying $|\mathcal{B}| = 2^r$, $Q(2^t, 2^r, \mathcal{B}) \geq 2^{r+s}$ has a lot of constant size linear relations between its elements, which allows us to describe it with significantly less than $\log \binom{2^t}{2^r}$ bits.

In order to encode a $\mathcal{B} \subseteq \mathbb{Z}_\mu$ for which $|\mathcal{B}| = 2^r$, $Q(2^t, 2^r, \mathcal{B}) \geq 2^{r+s}$ efficiently, we show that any such $\mathcal{B}$ has a decomposition $\mathcal{B} = \mathcal{D} \cup \overline{\mathcal{D}}$, where $|\mathcal{D}| = D$ as in the theorem, $\overline{\mathcal{D}} = \mathcal{B} \setminus \mathcal{D}$, and there exist fixed $x, z \in \mathbb{Z}_\mu$ that the elements b of $\mathcal{D}$ can be ordered suitably and be expressed as

$$b = \epsilon(z - x) - \epsilon_1 b_1 - \ldots - \epsilon_{l-1} b_{l-1}, \qquad (7)$$

where $b_1, \ldots, b_{l-1}$ are either from $\overline{\mathcal{D}}$ or from elements of $\mathcal{D}$ that are expressed earlier. The numbers $\epsilon, \epsilon_1, \ldots, \epsilon_{l-1}$ are all in $\{-1, 1\}$. The saving per every item in $\mathcal{D}$ is the difference measured in bits between its description length via (7) versus their default information cost per item. The latter is:

$$\frac{\log\binom{2^t}{2^r} - \log\binom{2^t}{2^r - D}}{D} \sim t - r$$

Since the sequence $\epsilon_1, b_1, \ldots, \epsilon_{l-1}, b_{l-1}$ together with ϵ can be described with $(l-1)(r+1)+1$ bits (each b_i is element of $\mathcal{B}$ which is already on our list, so has an r-bit description), our saving per item is

$$B = t - r - (l-1)(r+1) - 1 = t - l(r+1)$$

bits. Our total saving is then $DB - t$, since we also need t bits to describe $z - x$ (once for the entire $\mathcal{D}$). The upper bound on the probability of the event $Q(2^t, 2^r, \mathcal{B}) \geq 2^{r+s}$ is then 2^{-DB+t}.

We are left to construct the $(\mathcal{D}, \overline{\mathcal{D}})$ decomposition and to calculate D. Consider a $\mathcal{B}$ that satisfies $Q(2^t, 2^r, \mathcal{B}) \geq 2^{r+s}$. Then there are $\mathcal{X}, \mathcal{Z} \subseteq \mathbb{Z}_\mu$, $|\mathcal{X}| = |\mathcal{Z}| = 2^r$ such that $|\{(x,z) \mid x \in \mathcal{X}, z \in \mathcal{Z}, z - x \in \mathcal{B}\}| \geq 2^{r+s}$. We fix such an $\mathcal{X}, \mathcal{Z}$ pair. Let G be the bipartite graph with bipartition $(\mathcal{X}, \mathcal{Z})$ and edge set

$$e(G) = \{(x,z) \mid \ x \in \mathcal{X}, z \in \mathcal{Z}, z - x \in \mathcal{B}\}.$$

By our assumption $|e(G)| \geq 2^{r+s}$. If we iteratively remove the minimum degree vertex from G until all degrees of the resulting graph are at least $2^s/2$ (i.e. the average degree of G divided by two), it is easy to show that this process ends up with a non-empty graph G' with minimum degree at least $2^s/2$. Fix a vertex $x \in \mathcal{X} \cap V(G')$. Our proof hinges upon the following construction:

Definition 5. *Let P_i for $i = 1, 2, \ldots$ be the set of all those (not necessarily simple) paths π of length i in G' (the length is the number of edges) that satisfy:*

1. *π starts at x*
2. *No two edges edges of π have identical labels, where a label of an edge (v, w) ($v \in \mathcal{X}$, $w \in \mathcal{Z}$) is by definition $w - v$.*

Let π be a path in P_i and let $d = d(\pi)$ denote the degree of its end point z. All edges incident to z have distinct labels, so the number of those edges incident to z whose label do not coincide with any labels we already have in π is at least $d - i$. Thus π has $d - i \geq \frac{2^s}{2} - i$ continuations in P_{i+1}. Therefore, by induction, for $i \geq 1$:

$$|P_i| \geq \prod_{j=0}^{i-1}\left(\frac{2^s}{2} - j\right) > \frac{1}{e}\frac{2^{is}}{2^i}.$$

Consider the set P_l. Notice that if l is odd, then every path in P_l end in $\mathcal{Z}$, otherwise they all end in $\mathcal{X}$. Since the nodes of G' are from $\mathcal{X} \cup \mathcal{Z}$ and $|\mathcal{X}|, |\mathcal{Z}| = 2^r$, there must be a $z \in \mathcal{X}$ (if l is even) or $z \in \mathcal{Z}$ (if l is odd) such that at least $\frac{|P_l|}{2^r} \geq \frac{1}{e}\frac{2^{ls-r}}{2^l}$ paths from P_l end in z.

Let T be the set of the paths in P_l that end in this z. We will use the paths in T to find a lot of small linear relations among the elements of $\mathcal{B}$. For a path π let $\ell(\pi)$ denote the set of labels that occur on its edges, and define $\mathcal{D}_0 = \cup_{\pi \in T} \ell(\pi)$, which is just the collection of all labels that ever occur in those paths of P_l that end in z. Of course, $\mathcal{D}_0 \subseteq \mathcal{B}$, because all labels along the edges of G' are in $\mathcal{B}$. In order to estimate $|\mathcal{D}_0|$ we view a path $\pi \in P_l$ as an ordered sequence of labels. Each $\pi \in P_l$ uniquely corresponds to such a sequence of length l (although not necessarily every element of $\mathcal{D}_0^l$ is an element of P_l). Since from an alphabet of size $|\mathcal{D}_0|$ we can create at most $|\mathcal{D}_0|^l$ different sequences of length l, we have that

$$|\mathcal{D}_0| \geq |T|^{1/l} \geq \left(\frac{1}{e}\frac{2^{ls-r}}{2^l}\right)^{1/l} \geq 2^{s-r/l}/(2+2/l).$$

We are now ready to define the decomposition $\mathcal{B} = \mathcal{D} \cup \overline{\mathcal{D}}$ as promised in the beginning. The role of x and z in expression (7) will be played by the common starting- and end-point of all paths in T. For any path $\pi \in T$ we have that

$$z - x = b_1 - b_2 + b_3 - \ldots + b_l \quad \text{(if } l \text{ is odd, otherwise the last sign is a minus)}$$

It is a trivial matter to transform the above equation into (7), where b is one of the b_is (our choice which one). What remains is to show is that starting from a subset of T we can to generate all remaining elements by (7) such, that the number of generated elements is no less than the bound we require. A combinatorial lemma will help us in this.

Definition 6. *We say that a set $\{h_1, \ldots, h_{l-1}\}$ of nodes in an undirected hyper-graph $\mathcal{H}$ generates node h, if $\{h_1, \ldots, h_{l-1}, h\}$ is a hyper-edge. A generator set for $\mathcal{H}$ is a subset of nodes from which we can iteratively generate the entire vertex set of $\mathcal{H}$.*

Lemma 1. *Let $\mathcal{H}$ be an hyper-graph on m nodes such that every edge is a set of size at most l, and every node is contained in at least one hyper-edge. Then $\mathcal{H}$ has a generator of size at most $\frac{(l-1)m}{l}$.*

Proof. The proof is by induction on l. The claim is trivial for $l = 1$. Take a minimal generator set X for $\mathcal{H}$. If it does not satisfy our condition, then $|X| > \frac{(l-1)m}{l}$. Consider the hyper-graph $\mathcal{H}'$ we get from $\mathcal{H}$ by restricting all of its nodes and edges to X. Since a minimal generator set in $\mathcal{H}$ cannot properly contain any hyper-edge, every hyper-edge in $\mathcal{H}'$ has size at most $l-1$. Thus by induction $\mathcal{H}'$ has a generator set Y of size at least $\frac{(l-2)|X|}{l-1}$. But $Y \cup \overline{X}$ generates $\mathcal{H}$, and it has size at most $\frac{l-2}{l-1}|X| + m - |X| \leq \frac{(l-1)m}{l}$.

We now apply this lemma for the hyper-graph, which has vertex set $\mathcal{D}_0$ and edge set $\{\ell(\pi) \mid \pi \in T\}$. We get a generator set of size $(l-1)|\mathcal{D}_0|/l$. We put the elements of this generator set into $\overline{\mathcal{D}}$, as well as the elements of $\mathcal{B}$ that are not in $\mathcal{D}_0$. We can generate the remaining elements of $\mathcal{D}_0$ out of these via (7), and we let these form the set $\mathcal{D}$. The size of $\mathcal{D}$ is $|\mathcal{D}_0|/l = \frac{2^{s-r/l}}{(2l+2)}$.

Proof. (of Corollary 1) In Theorem 6 we set $a = r/t$, $s = 2r^2/t$, $l = \frac{3t}{4(r+1)}$. This gives $B = t/4$ and

$$D = \frac{\exp_2\left(2\frac{r^2}{t} - \frac{4r(r+1)}{3t}\right)}{2l+2} = q^{2\frac{r}{t} - \frac{4(r+1)}{3t}}/(2l+2) \geq q^{a/2}$$

if q is large enough (above $\exp_2(z)$ is by definition 2^z). Thus $2^{-DB+t} \leq 2^{-q^a/2}$ when q is sufficiently large.

Acknowledgements. We thank Mihir Bellare for suggesting the open problem to us. Furthermore, we are grateful to Yannick Seurin for pointing out a gap in the definition of the simulator in a previous version of this paper.

References

1. Abe, M., Kiltz, E., Okamoto, T.: Chosen ciphertext security with optimal ciphertext overhead. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 355–371. Springer, Heidelberg (2008)
2. Alon, N., Kaufman, T., Krivelevich, M., Ron, D.: Testing triangle-freeness in general graphs. SIAM J. Discrete Math. 22(2), 786–819 (2008)
3. Bellare, M., Ristenpart, T.: Simulation without the artificial abort: Simplified proof and improved concrete security for waters' IBE scheme. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 407–424. Springer, Heidelberg (2009)
4. Bellare, M., Rogaway, P.: Random oracles are practical: A paradigm for designing efficient protocols. In: Ashby, V. (ed.) ACM CCS 1993, pp. 62–73. ACM Press (November 1993)
5. Bellare, M., Rogaway, P.: The exact security of digital signatures: How to sign with RSA and Rabin. In: Maurer, U.M. (ed.) EUROCRYPT 1996. LNCS, vol. 1070, pp. 399–416. Springer, Heidelberg (1996)
6. Boneh, D., Lynn, B., Shacham, H.: Short signatures from the weil pairing. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 514–532. Springer, Heidelberg (2001)
7. Cachin, C., Micali, S., Stadler, M.A.: Computationally private information retrieval with polylogarithmic communication. In: Stern, J. (ed.) EUROCRYPT 1999. LNCS, vol. 1592, pp. 402–414. Springer, Heidelberg (1999)
8. Chevallier-Mames, B., Phan, D.H., Pointcheval, D.: Optimal asymmetric encryption and signature paddings. In: Ioannidis, J., Keromytis, A.D., Yung, M. (eds.) ACNS 2005. LNCS, vol. 3531, pp. 254–268. Springer, Heidelberg (2005)
9. Coron, J.-S.: On the exact security of full domain hash. In: Bellare, M. (ed.) CRYPTO 2000. LNCS, vol. 1880, pp. 229–235. Springer, Heidelberg (2000)
10. Coron, J.-S., Dodis, Y., Malinaud, C., Puniya, P.: Merkle-Damgård revisited: How to construct a hash function. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 430–448. Springer, Heidelberg (2005)
11. Dodis, Y., Ristenpart, T., Shrimpton, T.: Salvaging Merkle-Damgård for practical applications. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 371–388. Springer, Heidelberg (2009)

12. Goldwasser, S., Micali, S., Rivest, R.L.: A digital signature scheme secure against adaptive chosen-message attacks. SIAM Journal on Computing 17(2), 281–308 (1988)
13. Green, B.: Counting sets with small sumset, and the clique number of random cayley graphs. Combinatorica, 307–326 (2005)
14. Holenstein, T., Künzler, R., Tessaro, S.: The equivalence of the random oracle model and the ideal cipher model, revisited. In: Fortnow, L., Vadhan, S.P. (eds.) ACM STOC, pp. 89–98. ACM Press (June 2011)
15. Kakvi, S.A., Kiltz, E.: Optimal security proofs for full domain hash, revisited. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 537–553. Springer, Heidelberg (2012)
16. Katz, J., Wang, N.: Efficiency improvements for signature schemes with tight security reductions. In: Jajodia, S., Atluri, V., Jaeger, T. (eds.) ACM CCS 2003, pp. 155–164. ACM Press (October 2003)
17. Kiltz, E., Pietrzak, K., Szegedy, M.: Digital signatures with minimal overhead. Cryptology ePrint Archive, Report 2012/658 (2012), http://eprint.iacr.org/
18. Mandal, A., Patarin, J., Seurin, Y.: On the public indifferentiability and correlation intractability of the 6-round feistel construction. In: Cramer, R. (ed.) TCC 2012. LNCS, vol. 7194, pp. 285–302. Springer, Heidelberg (2012)
19. Maurer, U.M., Renner, R.S., Holenstein, C.: Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In: Naor, M. (ed.) TCC 2004. LNCS, vol. 2951, pp. 21–39. Springer, Heidelberg (2004)
20. Naor, A., Verstraëte, J.: A note on bipartite graphs without 2k-cycles. Comb. Probab. Comput. 14(5-6), 845–849 (2005)
21. Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes. In: Stern, J. (ed.) EUROCRYPT 1999. LNCS, vol. 1592, pp. 223–238. Springer, Heidelberg (1999)
22. Peikert, C., Waters, B.: Lossy trapdoor functions and their applications. In: Ladner, R.E., Dwork, C. (eds.) ACM STOC, pp. 187–196. ACM Press (May 2008)
23. Yoneyama, K., Miyagawa, S., Ohta, K.: Leaky random oracle. IEICE Transactions 92-A(8), 1795–1807 (2009)