Electronic Colloquium on Computational Complexity, Report No. 17 (2002)

Cryptographic Hardness based on the Decoding of Reed-Solomon Codes with Applications

Aggelos Kiayias* Moti Yung†

Abstract

We investigate the decoding problem of Reed-Solomon Codes (aka: the Polynomial Reconstruction Problem – PR) from a cryptographic hardness perspective. First, following the standard methodology for constructing cryptographically strong primitives, we formulate a decisional intractability assumption related to the PR problem. Then, based on this assumption we show: (i) *hardness of partial information extraction*: an adversary who wishes to predict the value of some computable function on a new point of the solution of a given PR-instance, has no more than a negligible advantage over an adversary who wishes to do the same without seeing the PR-instance (for any probability distribution of the new point), and ii) *pseudorandomness*: PR-instances are pseudorandom in the sense that they are indistinguishable from totally random sets of points over the finite field.

The above results lay the theoretical framework for the exploitation of PR as a basic cryptographic tool. In fact, there are several advantages of cryptographic primitives built over this tool. For example, in PR, the size of the corrupted codeword (which corresponds to the size of a ciphertext and the plaintext) and the size of the index of error locations (which corresponds to the size of the key) are independent and can even be super-polynomially related. We know of no other problem that allows such a property. Subsequently, we present concrete constructions of primitives: First, we construct a direct one-way function that behaves as a "large secure envelope." Then, we use the one-way function as a building block in a non-interactive commitment scheme for large values which is the first scheme with sublinear decommitment witness size. Further, we construct a semantically secure stateful-cipher that possesses unique properties: it allows keys to be inverse super-polynomially shorter than the encrypted messages and it satisfies "computational perfect secrecy", "forward secrecy" and "key-equivalence."

1 Introduction

Finding new problems based on which we can design cryptographic primitives is an important research area. Given a presumably hard problem it is usually non-trivial to exploit it directly in cryptography. Many times, in order to serve as the base for secure cryptographic primitives, we need to find related hard decision problems (predicates). This is the fundamental methodology initiated by Goldwasser and Micali in [GM84] where they started the quest for formal notions and proofs of security in cryptography. The decision problem's hardness, typically seems related to (or at times proved in some sense related or, even better, reducible from) the

*The Graduate Center, CUNY, NY USA, akiayias@gc.cuny.edu
†CertCo, NY USA, moti@cs.columbia.edu

ISSN 1433-8092

hardness of the original problem. Hard predicate assumptions allow formal security proofs (in the form of reductions) for advanced cryptographic primitives such as pseudorandomness and semantically secure encryption. The first example of a decisional assumption is the Quadratic-Residuosity, which is related to (but not known to be reducible from) Factoring and was employed in designing the first semantically secure encryption scheme [GM84]. Another such assumption is the Decisional Diffie-Hellman which implies the security of ElGamal encryption and other advanced cryptographic primitives (e.g., [NR97]), and is related to (but not known to be reducible from) the Diffie-Hellman problem.

In this work, our goal is to investigate the possibility of cryptographic primitives whose security is based on the problem of *Polynomial Reconstruction* (PR). Recall that the problem of Polynomial Reconstruction is defined as follows: Given n points over a (large) finite field $\mathbb{F}$, such that at least t of them belong to the graph of a polynomial p of degree less than k, recover such a polynomial (where $n > t > k$).

We note that Polynomial Reconstruction is essentially equivalent to the decoding problem of Reed-Solomon codes and naturally has received much attention from a "positive" (coding theoretic) perspective: Starting from the classical algorithm of Berlekamp and Welch([BW86]) which solves Polynomial Reconstruction provided that $t \geq \frac{n+k}{2}$ (error correcting bound for Reed-Solomon Codes), to the recent work of Guruswami and Sudan [GS98] which solves it when $t \geq \sqrt{kn}$ (many solutions are possible in the worst case). The current state of knowledge suggests that for values of t below $\sqrt{kn}$ the problem is hard.

Regarding our goal, Polynomial Reconstruction as is, does not seem to be ready for direct cryptographic exploitation: even if presumed hard, it is not at all clear how to build advanced cryptographic primitives whose security can be reduced to it. Indeed, when Naor and Pinkas [NP99] first employed the problem cryptographically in a context of protocol design, they actually introduced a related pseudorandomness assumption. The relation of this assumption to PR also motivates further investigation.

In this work, we first identify a decisional problem naturally related to PR. This problem is based on the following basic question: given a PR-instance that contains n points and an index $i \in \{1, \ldots, n\}$, does the i-th point of the instance belong in the graph of the polynomial solution or not? (note that in the range of our parameters, a PR-instance has a unique solution with very high probability). We formalize the hardness of this predicate for all indices i as the "Decisional-PR-Assumption" (DPR).

Based on the DPR-Assumption we show: (i) *hardness of partial information extraction*: an adversary with access to a PR-instance who wishes to predict the value of some computable function on a new point of the polynomial-solution, gains only negligible advantage compared to an adversary who wishes to predict the same value without seeing the instance — this holds true even if the point follows an adversarially chosen probability distribution; also: (ii) *pseudo-randomness*: PR-instances are pseudorandom in the sense that they are indistinguishable from random sets of points, for any poly-time observer. These results suggest that PR is quite robust in the cryptographic sense and is suitable for employment in cryptographic constructions.

There are several possible advantages of the PR problem which can be exploited by cryptographic primitives built on it, for example: (i) The natural dichotomy and independence exhibited between the key-size (index of error locations) and the size of Reed-Solomon encoded message (or concealed information in PR-based systems) allows key-sizes to be selected independently of (and possibly super-polynomially smaller than) the message size; we know

of no other problem that allows such a property in the cryptographic literature. (ii) The PR problem enjoys a unique algebraic structure. (iii) The operation of polynomial interpolation which is basic in PR cryptographic primitives can be implemented quite efficiently (especially in special purpose hardware).

With the above advantages in mind, we apply our results to the design of PR-based cryptographic primitives. First we define a one-way function based on Polynomial Reconstruction. Under DPR, our one-way function has strong partial-information concealment properties that make it suitable as a building block in designing large-value commitment schemes. In particular, our commitment scheme yields a non-interactive scheme with sublinear decommitment witness size, i.e. it allows the decommitment information to be substantially shorter than the committed values and the commitment information. This property allows substantial reduction of the private storage space required for decommitment. Then, we introduce a new semantically-secure stateful cipher based on Polynomial Reconstruction. Our cipher demonstrates the exploitation of structural properties possessed by the PR problem and exhibits unique properties, which are:

- (i) *Forward Secrecy*: this property suggests that if a total security breach occurs at a certain time (e.g. the key is revealed), this affects the security only of future messages while the previously sent messages are semantically secure in the view of the perpetrator.
- (ii) *Computational perfect secrecy.* Consider the following two attacks against a cryptosystem: an existential attack is a chosen-plaintext attack that reveals an encrypted message whereas a universal attack is a chosen-plaintext attack that reveals the key, and thus all messages (from some point on in a forward-secure cipher). A cipher for which the two attacks are inter-reducible is said to satisfy "computational perfect secrecy." This property is motivated by Shannon's early information-theoretic work and by the work of Goldwasser and Blum [BG84] who introduced a variant of it in the computational sense by exhibiting a remarkable cryptosystem where violating semantic security implies factoring of the composite key.
- (iii) *Short key-size*: this property suggests that the plaintext can be *superpolynomial* in the key-size (the security parameter).
- (iv) *Built-in error correction.* The decryption operation incorporates error-correction capabilities in a direct manner.
- (v) *Key-equivalence.* There are no "weak" (prone to specialized attacks) families of keys.

Notation. All computations are performed in a (large) finite field $\mathbb{F}$. Tuples in $\mathbb{F}^n$ are denoted by $\mathbf{x}$ and $(\mathbf{x})_i$ denotes the i-th coordinate of $\mathbf{x}$. Denote by $(n)_k := n(n-1)\ldots(n-k+1)$, and if A is a set denote by $(A)_k$ the set of all k-tuples over A without repetitions. PPT stands for "probabilistic polynomial-time." All algorithms mentioned in the paper are PPT Turing Machines, and denoted by $\mathcal{A}, \mathcal{B}$ etc. For any PPT $\mathcal{A}$ that uses randomness $r \in \mathcal{R}$ and input $x \in D$, if y is in the range of $\mathcal{A}$ we will denote by $\mathbf{Prob}_{r \in_U \mathcal{R}; x \in_U D}[\mathcal{A}(r,x) = y]$ the probability that $\mathcal{A}$ returns y when r and x are uniformly distributed over their respective ranges (note that y may be a function of x). A function $\alpha(n) : \mathbb{N} \to \mathbb{R}$ is negligible if for all c it holds that $\alpha(n) < n^{-c}$ for sufficiently large n. A function $\beta(n) : \mathbb{N} \to \mathbb{R}$ is called non-negligible if it is not negligible for all large enough inputs, namely there is a c s.t. $\beta(n) \geq n^{-c}$ for all n sufficiently

large. When the probability of an event is greater equal to $1 - \epsilon(n)$ where $\epsilon(n)$ is negligible, then we write that the event happens "with overwhelming probability."

2 The Problem

Definition 2.1 Polynomial Reconstruction (PR). *Given n, k, t and $\{\langle z_i, y_i\rangle\}_{i=1}^{n}$ with $z_i \neq z_j$ for $i \neq j$, output all $\langle p(x), I\rangle$ such that $p \in \mathbb{F}[x]$, degree$(p) < k$, $I \subseteq \{1, \ldots, n\}$, $|I| \geq t$ and $\forall i \in I(p(z_i) = y_i)$.*

PR as a coding theoretic problem asks for all messages that agree with at least t positions of the received Reed-Solomon codeword. For a general treatment on the subject the interested reader is referred to [Ber68] or [MS77]. Note that $k < n$ since k/n is the message rate of the code, and that we further require that at least one solution $\langle p(x), I\rangle$ exists.

When $t \geq \frac{n+k}{2}$ then PR$[n, k, t]$ has only one solution and it can be found with the algorithm of Berlekamp and Welch [BW86] ($\frac{n+k}{2}$ is the error-correction bound of the Reed-Solomon codes). When t is beyond the error-correction bound then having more than one solution is possible. Sudan proposed an algorithm that solves the PR beyond the error-correction bound when $t \geq \sqrt{2kn}$ in [Sud97] and later in [GS98], Guruswami and Sudan presented an algorithm that solves the PR for $t > \sqrt{kn}$. In [GSR95] it was proven that when $t > \sqrt{kn}$ the number of solutions is bounded by a polynomial. In [GS98] it is pointed out that the possibility of an algorithm that solves instances for smaller values of t might be limited. We note here that the solvability of PR (and related problems) was also studied in the context of lattices, see [BN00]. Consequently the current state of knowledge implies that PR$[n, k, t]$ is hard for the choice of parameters $t < \sqrt{kn}$.

2.1 Structure of the Instance Space

An instance of PR will be denoted by $X := \{\langle z_i, y_i\rangle\}_{i=1}^{n}$; the set of all instances with parameters n, k, t will be denoted by $\mathcal{S}_{n,k,t}$. In order to refer to PR with parameters n, k, t we will write PR$[n, k, t]$. Note that unless stated otherwise we assume that n is polynomially related to $\log |\mathbb{F}|$.

Let $I \subseteq \{1, \ldots, n\}$ with $|I| = t$. We denote by $\mathcal{S}_{n,k,t}(I)$ the subset of $\mathcal{S}_{n,k,t}$ so that for any $X \in \mathcal{S}_{n,k,t}(I)$ it holds that X has a solution of the form $\langle p, I\rangle$. It is clear that $\mathcal{S}_{n,k,t} = \cup_{|I|=t}\mathcal{S}_{n,k,t}(I)$, but $\{\mathcal{S}_{n,k,t}(I)\}_{|I|=t}$ does not constitute a partition of $\mathcal{S}_{n,k,t}$. Nevertheless concentrating on instance sets of the form $\mathcal{S}_{n,k,t}(I)$ is helpful in understanding the structure of $\mathcal{S}_{n,k,t}$.

Lemma 2.2 *For any $I \subseteq \{1, \ldots, n\}$ with $|I| = t$ it holds that $\#\mathcal{S}_{n,k,t}(I) = (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}$.*

Proof. Straightforward since $n - t + k$ are exactly the degrees of freedom that each element of $\mathcal{S}_{n,k,t}(I)$ has. ■

Clearly if a PR-instance $X \in \mathcal{S}_{n,k,t}$ has two distinct solutions $\langle p_1, I_1\rangle$ and $\langle p_2, I_2\rangle$, it holds that $X \in \mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)$. To determine the likelihood that a given PR-instance has a single solution or more, the following lemma is helpful:

Lemma 2.3 *(i) For all $I_1, I_2 \subseteq \{1,\ldots,n\}$, with $|I_1| = |I_2| = t$, $I_1 \neq I_2$, it holds that $\#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)) \leq (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k-1}$.*
(ii) The total number of PR-instances of $\mathcal{S}_{n,k,t}$ that have more than one solution is less than $\binom{n}{t}^2 (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k-1}$.

Proof. (i) Let $|I_1 \cap I_2| = m$; note that $m \in \{0,\ldots,t-1\}$. The $\langle z_1,\ldots,z_n\rangle$ values contribute $(\mathbb{F})_n$ choices. The "free" (noise) points contribute $|\mathbb{F}|^{n-2t+m}$ choices. It remains to find the number of choices due to the y-elements that correspond to the positions $I_1 \cup I_2$. There first solution contributes $|\mathbb{F}|^k$ choices, whereas the second solution, if $m < k$, it contributes $|\mathbb{F}|^{k-m}$. If $m \geq k$ no second solution is feasible. So we have two cases: $m < k$, where $\#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)) = (|\mathbb{F}|)_n |\mathbb{F}|^{n-2t+2k}$, and $m \geq k$, where $\#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)) = (|\mathbb{F}|)_n |\mathbb{F}|^{n-2t+m+k}$, with $m \in \{k,\ldots,t-1\}$. As a result, independently of the choice of I_1, I_2, $\#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)) \leq (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k-1}$ (recall that $t > k$).
(ii) it follows easily from the fact that the set of all instances of $\mathcal{S}_{n,k,t}$ that have more than one solution is a subset of $\cup_{I_1 \neq I_2} \mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)$. ■

The following lemma compares the number of elements of $\mathcal{S}_{n,k,t}$ and $\mathcal{S}_{n,k,t}(I)$ and in combination with the previous lemma it provides an estimate to the number of elements of $\mathcal{S}_{n,k,t}$.

Lemma 2.4 *Suppose $\log|\mathbb{F}| \geq 3n$. For any $I \subseteq \{1,\ldots,n\}$, $|I| = t$, it holds that $\binom{n}{t} - 2^{-n} \leq \frac{\#\mathcal{S}_{n,k,t}}{\#\mathcal{S}_{n,k,t}(I)} \leq \binom{n}{t}$.*

Proof. By definition it holds that $\mathcal{S}_{n,k,t} = \cup_{|I|=t} \mathcal{S}_{n,k,t}(I)$. It follows from lemma 2.2 that $\#\mathcal{S}_{n,k,t}(I) = \#\mathcal{S}_{n,k,t}(I')$ for all I, I'. Now fix some $I \subseteq \{1,\ldots,n\}$, $|I| = t$. It follows that,

$$\binom{n}{t} \#\mathcal{S}_{n,k,t}(I) - \sum_{I_1 \neq I_2} \#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)) \leq \#\mathcal{S}_{n,k,t} \leq \binom{n}{t} \#\mathcal{S}_{n,k,t}(I)$$

Next using the upper bound on $\sum_{I_1 \neq I_2} \#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2))$ that follows from lemma 2.3, it follows that (using the facts $\log|\mathbb{F}| \geq 3n$, $\binom{n}{t} < 2^n$)

$$\sum_{I_1 \neq I_2} \#(\mathcal{S}_{n,k,t}(I_1) \cap \mathcal{S}_{n,k,t}(I_2)) < \frac{\binom{n}{t}^2 \#\mathcal{S}_{n,k,t}(I)}{|\mathbb{F}|} < \frac{\#\mathcal{S}_{n,k,t}(I)}{2^n}$$

It follows that

$$(\binom{n}{t} - \frac{1}{2^n}) \#\mathcal{S}_{n,k,t}(I) \leq \#\mathcal{S}_{n,k,t} \leq \binom{n}{t} \#\mathcal{S}_{n,k,t}(I)$$

which completes the proof. ■

As a result we can draw the following corollary:

Corollary 2.5 *The number of elements of $\mathcal{S}_{n,k,t}$ can be approximated (within negligible error) by $\binom{n}{t}(|\mathbb{F}|)_n |\mathbb{F}|^{n-k+t}$.*

Clearly sampling the uniform distribution over $\mathcal{S}_{n,k,t}(I)$ is straightforward (based on the fact that the uniform distribution over the finite field $\mathbb{F}$ can be sampled — something that can be shown easily). Next we proceed to show that the uniform distribution of PR instances is actually *samplable* (with negligible statistical error). We start with a standard definition:

Definition 2.6 *A probability distribution $\mathcal{D}$ over some space R of objects of size polynomial in n is called (polynomial time) samplable if there is a* PPT $S_{\mathcal{D}} : \mathcal{R}_{\mathcal{D}} \rightarrow R$ *so that the probability assigned to any $y \in R$ by $\mathcal{D}$ is* $\mathbf{Prob}_{\mathcal{D}}[y] = \mathbf{Prob}_{x \in_U \mathcal{R}_{\mathcal{D}}}[S_{\mathcal{D}}(x) = y]$.

Consider the following procedure S that samples $\mathcal{S}_{n,k,t}$: first select n random distinct elements of $\mathbb{F}$, $z_1, \ldots, z_n$. Then, select a random I such that $|I| = t$ and then select a random polynomial p of degree less than k (e.g. by selecting k random elements of $\mathbb{F}$ as its coefficients). Set $y_i := p(z_i)$ for $i \in I$ and select the remaining y_i for $i \notin I$ at random. The output of S is $\{\langle z_i, y_i \rangle\}_{i=1}^{n}$. The following lemma suggests that the described procedure S essentially samples the uniform distribution over $\mathcal{S}_{n,k,t}$.

Lemma 2.7 *Let* $\log |\mathbb{F}| \geq 3n$. *The probability distribution defined by* S *is statistically indistinguishable from the uniform over $\mathcal{S}_{n,k,t}$. More specifically,* $A := \sum_{X \in \mathcal{S}_{n,k,t}} | \mathbf{Prob}[\mathsf{S}(1^n) = X] - \frac{1}{\#\mathcal{S}_{n,k,t}} | < 2n2^{-n}$.

Proof. Fix some $X \in \mathcal{S}_{n,k,t}$. If only a single solution $\langle p, I \rangle$ with $|I| = t$ exists in X then it follows easily that there is a unique assignment of the random choices of S that yields X. As a result in this case it holds that $\mathbf{Prob}[\mathsf{S}(1^n) = X] = \frac{1}{\binom{n}{t} \cdot (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}}$. Let us partition $\mathcal{S}_{n,k,t}$ to the set $\mathcal{S}_1$ that contains instances X with a single solution as above and let $\mathcal{S}_2 := \mathcal{S}_{n,k,t} - \mathcal{S}_1$. If $\frac{1}{2}A$ is the statistical distance between the two distributions then it follows that:

$$A = A_1 + A_2 = \sum_{X \in \mathcal{S}_1} | \frac{1}{\binom{n}{t} (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}} - \frac{1}{\#\mathcal{S}_{n,k,t}} | + \sum_{X \in \mathcal{S}_2} | \mathbf{Prob}[\mathsf{S}(1^n) = X] - \frac{1}{\#\mathcal{S}_{n,k,t}} |$$

From lemma 2.4 it holds that

$$| \frac{1}{\binom{n}{t} (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}} - \frac{1}{\#\mathcal{S}_{n,k,t}} | = \frac{1}{\binom{n}{t} \#\mathcal{S}_{n,k,t}} | \frac{\#\mathcal{S}_{n,k,t}}{(|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}} - \binom{n}{t} | < \frac{1}{2^n \binom{n}{t} \#\mathcal{S}_{n,k,t}}$$

It follows that:

$$A_1 = \sum_{X \in \mathcal{S}_1} | \frac{1}{\binom{n}{t} (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}} - \frac{1}{\#\mathcal{S}_{n,k,t}} | < \frac{1}{2^n \binom{n}{t}}$$

and as a result A_1 is negligible. Next we proceed to show that A_2 is also negligible. Note that this will follow immediately by the following two facts:
(i) $\sum_{X \in \mathcal{S}_2} \mathbf{Prob}[\mathsf{S}(1^n) = X] < (n-t)2^{-n}$. To see this, let n_X be such that $\mathbf{Prob}[\mathsf{S}(1^n) = X] = \frac{n_X}{\binom{n}{t} \cdot (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}}$. It follows that $\sum_{X \in \mathcal{S}_2} n_X = \binom{n}{t} \cdot (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k} - \#\mathcal{S}_1$. Since $\mathcal{S}_1$ contains all those PR instances that contain exactly one solution it follows easily that $\#\mathcal{S}_1 > \binom{n}{t} (|\mathbb{F}|)_n |\mathbb{F}|^k (|\mathbb{F}| - \binom{n}{k}))^{n-t}$. As a result (using the facts $\log |\mathbb{F}| \geq 3n$, $\binom{n}{k} < 2^n$)

$$\sum_{X \in \mathcal{S}_2} n_X < \binom{n}{t} \cdot (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k} \left(1 - \left(\frac{|\mathbb{F}| - \binom{n}{k}}{|\mathbb{F}|}\right)^{n-t}\right)$$

$$\Longrightarrow \sum_{X \in \mathcal{S}_2} \mathbf{Prob}[\mathsf{S}(1^n) = X] < 1 - \left(1 - \frac{\binom{n}{k}}{|\mathbb{F}|}\right)^{n-t} < 1 - \left(1 - \frac{1}{2^n}\right)^{n-t} = \sum_{i=1}^{n-t} \binom{n-t}{i} \frac{(-1)^{i+1}}{2^{ni}}$$

the sum on the right hand side is easily shown to be less than $(n-t)2^{-n}$.

(ii) $\sum_{X \in \mathcal{S}_2} \frac{1}{\#\mathcal{S}_{n,k,t}} < 2^{-n}$. Indeed the sum equals to $\frac{\#\mathcal{S}_2}{\#\mathcal{S}_{n,k,t}}$ and the stated result follows from lemma 2.3(ii).

Finally we conclude that $A < \frac{1}{2^n \binom{n}{t}} + \frac{n-t}{2^n} + \frac{1}{2^n} < 2n2^{-n}$. ■

Lemma 2.8 *Suppose that* $\log|\mathbb{F}| \geq 2n$. *The ratio of the number of PR-instances of* $\mathcal{S}_{n,k,t}$ *with more than one solution, over* $\#\mathcal{S}_{n,k,t}$ *is less than* 2^{-n}.

Proof. Because of lemma 2.4 it holds that $(\binom{n}{t} - 2^{-n})(|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k} \leq \#\mathcal{S}_{n,k,t} \leq \binom{n}{t}(|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k}$. The number of PR-instances of $\mathcal{S}_{n,k,t}$ with more than one solution is less than $\binom{n}{t}^2 (|\mathbb{F}|)_n |\mathbb{F}|^{n-t+k-1}$ (from lemma 2.3ii). It follows that the ratio is less than $\binom{n}{t}^2 (\binom{n}{t} - 2^{-n})^{-1} |\mathbb{F}|^{-1} < 2^{-n}$. ■

It is an immediate corollary from the above lemma that any PPT which samples the uniform distribution over $\mathcal{S}_{n,k,t}$ will select an instance X that has a unique solution with overwhelming probability $1 - 2^{-n}$. Consequently any instance $X \in \mathcal{S}_{n,k,t}$ uniquely defines a polynomial p (with overwhelming probability) such that degree$(p) < k$. We denote this polynomial by s_X (for solution of X). The set of indices that corresponds to the graph of p which we call "the index-solution set" is denoted by $I(X)$. Obviously, the recovery of s_X implies the recovery of $I(X)$ and vice-versa.

2.2 Security Parameters

In our exposition we will use n as be the security parameter. The parameters k, t are functions in n, so that $k < t < n$ and $t < \sqrt{nk}$. The straightforward brute-force algorithm for solving PR$[n, k, t]$ requires checking all possibilities and as a result has complexity proportional to $\min(\binom{n}{k}, \binom{n}{t})$. The parameters $[n, k(n), t(n)]$ are called *sound* for PR$[n, k, t]$ if $k(n)$ and $t(n)$ are chosen so that $t < \sqrt{kn}$ and $\min(\binom{n}{k}, \binom{n}{t})$ is exponential in n. Note that we will suppress (n) in $k(n), t(n)$. Observe that if $[n, k, t]$ are sound parameters then it also holds that $[n, k+1, t]$ are sound parameters (provided that $k+1 < t$). Intuitively this means that allowing the degree of the solution-polynomial to be greater without changing the other parameters it cannot make the problem easier. We will assume sound parameters throughout.

2.3 Partial Random Self-Reducibility

As it is noted in [NP99], Polynomial Reconstruction enjoys a partial self-reducibility property, namely that given an $X := \{\langle z_i, y_i \rangle\}_{i=1}^n \in \mathcal{S}_{n,k,t}$ it is possible to randomize the polynomial solution of X: choose a random polynomial p' of degree less than k and compute the instance $Y := \{\langle z_i, y_i + p(z_i) \rangle\}_{i=1}^n$. Nevertheless this is not at all sufficient to show that the problem is randomly self-reducible. This is because the procedure does not randomize the points that do not lie in the index-solution-set. Polynomial Reconstruction enjoys yet another partial random self-reducibility property, namely that the choice of the index-solution-set is not important. Informally this can be seen by the fact that one can permute the points of PR-instance by applying a random n-permutation. This fact is of importance from a cryptographic viewpoint since in many settings the index-solution-set plays the role of a cryptographic key. This second partial random self-reducibility is formalized and strengthened in the next section.

2.4 Altering The Distribution of PR-Instance Solutions

Suppose that some points of a polynomial solution of PR$[n,k,t]$ instance follow a given (non-uniform over $\mathbb{F}$) probability distribution. If h points of the polynomial solution follow a certain probability distribution we will fix these points to be the values of the polynomial over $\{0,\ldots,h-1\}$. Without loss of generality we assume that $0,\ldots,h-1$ are not equal to any of the $\langle z_1,\ldots,z_n\rangle$ values in a PR-instance (this is an event of negligible probability). Note that alternative "base" values $w_1,\ldots,w_h$ can be used instead of $0,\ldots,h-1$.

Let $[n,k-h,t]$ be sound parameters for some $0<h<k$. Let $\mathcal{D}_h$ be a samplable probability distribution over $\mathbb{F}^h$. We can extend $\mathcal{D}_h$ to be a samplable probability distribution over $\mathcal{S}_{n,k,t}$ by modifying the sampler S of section 2.1 so that it selects h values of the polynomial solution following $\mathcal{D}_h$ (instead of at random). We use the notation $\mathsf{S}_{\mathcal{D}_h}$ to denote this generalized sampler over $\mathcal{S}_{n,k,t}$. Note that we will use the same notation $\mathcal{D}_h$ for both probability distributions (over $\mathbb{F}^h$ and $\mathcal{S}_{n,k,t}$). Defining $\mathcal{D}_h$ over $\mathcal{S}_{n,k,t}(I)$ can be done in a similar manner as above, and the sampler will be denoted by $\mathsf{S}^I_{\mathcal{D}_h}$. If the base values are set to $\{w_1,\ldots,w_h\}$ the derived probability distribution over $\mathcal{S}_{n,k,t}$ and $\mathcal{S}_{n,k,t}(I)$ will be denoted by $\mathcal{D}_h^{w_1,\ldots,w_h}$.

The next lemma reveals that even under such a "modified solution distribution", the particular choice of the index-solution-set does not affect the output behavior of a certain procedure that operates on PR-instances. The core of the proof below is that given a PR-instance with unknown solution one can randomly permute the points in the instance.

Lemma 2.9 *Let $\mathcal{D}_h$ be a probability distribution over $\mathbb{F}^h$, with $h\in\{0,\ldots,k\}$. Let $\mathcal{A}:\mathcal{S}_{n,k,t}\rightarrow V$ be some PPT. Then it holds that there exists a PPT $\mathcal{A}'$ s.t. for all $v\in V$ and $I\subseteq\{1,\ldots,n\}$ with $|I|=t$,*

$$|\ \mathbf{Prob}_{X\in_{\mathcal{D}_h}\mathcal{S}_{n,k,t}}[\mathcal{A}(X)=v]-\mathbf{Prob}_{X\in_{\mathcal{D}_h}\mathcal{S}_{n,k,t}(I)}[\mathcal{A}'(X)=v]\ |$$

is negligible in n.

Proof. Fix some samplable distribution $\mathcal{D}_h$ over $\mathbb{F}^h$, a $v\in V$, and some $I\subseteq\{1,\ldots,n\}$ with $|I|=t$. Let $S_{\mathcal{D}_h}:\mathcal{R}_{\mathcal{D}_h}\rightarrow\mathbb{F}^h$ be the PPT that samples $\mathcal{D}_h$. Let $\rho\in\Re$ be the randomness used by $\mathsf{S}_{\mathcal{D}_h}$, to sample an element of $\mathcal{S}_{n,k,t}$, i.e. $\rho:=\langle I,z_1,\ldots,z_n,m_1,\ldots,m_{k-h},r,y_1,\ldots,y_{n-t}\rangle$; it holds that $\#\Re=\binom{n}{t}(\mathbb{F})_n|\mathbb{F}|^{k-h}\mathcal{R}_{\mathcal{D}_h}|\mathbb{F}|^{n-t}$. Similarly denote by $\rho'\in\Re'$ to be the randomness used by $\mathsf{S}^I_{\mathcal{D}_h}$, i.e. $\rho':=\langle z_1,\ldots,z_n,m_1,\ldots,m_{k-h},r,y_1,\ldots,y_{n-t}\rangle$; it holds that $\#\Re'=(\mathbb{F})_n|\mathbb{F}|^{k-h}\mathcal{R}_{\mathcal{D}_h}|\mathbb{F}|^{n-t}$. It follows $\#\Re=\#\Re'\binom{n}{t}$. Regarding the probability of $\mathcal{A}$ to return v, we have that:

$$\mathbf{Prob}_{X\in_{\mathcal{D}_h}\mathcal{S}_{n,k,t}}[\mathcal{A}(X)=v]=\mathbf{Prob}_{\rho\in_U\Re}[\mathcal{A}(\mathsf{S}_{\mathcal{D}_h}(\rho))=v]$$

Now consider the PPT $\mathcal{A}'$ that on input X, first it selects a random permutation π, it permutes the pairs of X according to π to obtain X^π and then it simulates $\mathcal{A}$.

$$\mathbf{Prob}_{X\in_{\mathcal{D}_h}\mathcal{S}_{n,k,t}(I);\pi\in_U\mathrm{Perm}(n)}[\mathcal{A}'(\pi,X)=v]=\mathbf{Prob}_{\rho'\in_U\Re';\pi\in_U\mathrm{Perm}(n)}[\mathcal{A}([\mathsf{S}^I_{\mathcal{D}_h}(\rho')]^\pi)=v]$$

Assume that $\mathcal{A}$ does $q(n)$ coin-tosses and define $C:=\{\langle b,\rho\rangle\mid\mathcal{A}(b,\mathsf{S}_{\mathcal{D}_h}(\rho))=v\}$ and $D:=\{\langle b,\rho',\pi\rangle\mid\mathcal{A}(b,[\mathsf{S}^I_{\mathcal{D}_h}(\rho)]^\pi)=v\}$, where $b\in\{0,1\}^{q(n)}$.

It follows that

$$\mathbf{Prob}_{\rho\in_U\Re}[\mathcal{A}(\mathsf{S}_{\mathcal{D}_h}(\rho))=v]=\frac{\#C}{2^{q(n)}\#\Re}$$

and

$$\mathbf{Prob}_{\rho' \in_U \Re'; \pi \in_U \mathrm{Perm}(n)}[\mathcal{A}([\mathsf{S}^I_{\mathcal{D}_h}(\rho')]^\pi) = v] = \frac{\#D}{2^{q(n)} n! \# \Re'}$$

Consider a mapping $J : \Re' \times \mathrm{Perm}(n) \to \Re$ so that if $\rho = J(\rho', \pi)$ with $\rho = \langle I^\rho, z_1^\rho, \ldots, z_n^\rho, m_1^\rho, \ldots, m_{k-h}^\rho, r^\rho, y_1^\rho, \ldots, y_{n-t}^\rho \rangle$ and $\rho' = \langle z_1^{\rho'}, \ldots, z_n^{\rho'}, m_1^{\rho'}, \ldots, m_{k-h}^{\rho'}, r^{\rho'}, y_1^{\rho'}, \ldots, y_{n-t}^{\rho'} \rangle$ it holds that $z_i^\rho = z_i^{\rho'}$, $m_j^\rho = m_j^{\rho'}$, $r^\rho = r^{\rho'}$ and $y_\ell^\rho = y_\ell^{\rho'}$, for $i = 1, \ldots, n$, $j = 1, \ldots, k-h$ and $\ell = 1, \ldots, n-t$ and additionally $I^\rho = \{\pi(i) \mid i \in I\}$. It is easy to see that a certain $\rho \in \Re$ has $t!(n-t)!$ pre-images under J. It follows that $\#D = t!(n-t)!\#C$ and as a result:

$$\mathbf{Prob}_{\rho \in_U \Re}[\mathcal{A}(\mathsf{S}_{\mathcal{D}_h}(\rho)) = v] = \mathbf{Prob}_{\rho' \in_U \Re'; \pi \in_U \mathrm{Perm}(n)}[\mathcal{A}([\mathsf{S}^I_{\mathcal{D}_h}(\rho')]^\pi) = v]$$

the result of the theorem follows. ■

Note that in the statement of the lemma above the choice of the points $\{0, \ldots, h-1\}$ as the ones that will be distributed according to some probability distribution is arbitrary as it is very easy to reformulate the above result so that some other collection of "base" values is selected. Additionally the value v used above can be generalized to being a function of X in a straightforward manner, without any modifications in the proof.

2.5 The Intractability Assumption

A decision problem that relates naturally to the hardness of solving an instance X of PR$[n, k, t]$ is the following: given X and an index $i \in \{1, \ldots, n\}$ decide whether $i \in I(X)$. We postulate that such decision is computationally hard to make whenever PR is hard. Since this has to hold true for all indices we will use a counter-positive argument to formalize the related decisional intractability assumption. In the definition below we describe a pair of predicates that refutes the assumption by "revealing" one of the points that belongs in the graph of the solution-polynomial (note that we formulate probabilities independently of the index-solution-set since given any PR-instance the index-solution-set can be randomized — see lemma 2.9):

Definition 2.10 *A pair of* PPT *predicates $\mathcal{A}_1, \mathcal{A}_2$ is called a gap-predicate-pair for the parameters n, k, t if for all $I \subseteq \{1, \ldots, n\}$ with $|I| = t$ it holds that:*

$$|\ \mathbf{Prob}[\mathcal{A}_1(i, X) = 1] - \mathbf{Prob}[\mathcal{A}_2(i, X) = 1]\ | = \begin{cases} \text{negligible} & \forall i \notin I \\ \text{non–negligible} & \text{for some } i \in I, i \leq n-k \end{cases}$$

where the probabilities are taken over all choices of $X \in \mathcal{S}_{n,k,t}(I)$ and internal coin-tosses of the predicates $\mathcal{A}_1, \mathcal{A}_2$.

A gap-predicate-pair when given a PR instance X and $i \in \{1, \ldots, n\}$ exhibits a measurable difference for at least one $i \in I(X)$, where at the same time it exhibits no measurable difference for indices outside $I(X)$. Using this, we formulate the Decisional-PR-Assumption as follows:

Decisional-PR-Assumption. (DPR$[n, k, t]$)
For any sound parameters $[n, k, t]$ there does not exist a gap-predicate-pair.

The relation of DPR to the Polynomial Reconstruction problem is revealed in the following two facts which are used to underline the justification for our intractability assumption. The first is straightforward:

Fact 2.11 *The existence of a polynomial-time algorithm for* PR$[n,k,t]$ *violates* DPR$[n,k,t]$.

To state the second fact we need a definition: a predicate $\mathcal{A} : \cup_j D_j \rightarrow \{0,1\}$ is called *independently samplable* over $\cup_j D_j$ if there is a PPT $S_{\mathcal{A}}$ that given $u \in \mathbb{N}$ and $X \in D_j$, it draws u independently sampled values of $\mathcal{A}$ over the space D_j. In particular, given $X \in D_j$, it holds that $S_{\mathcal{A}}(u,X) := \langle c_1, \ldots, c_u \rangle$ where each c_i is distributed over $\{0,1\}$ according to $\mathcal{A}(Y)$ where Y is uniformly selected over D_j. We denote by $\sum S_{\mathcal{A}}(u,X)$ the sum $\sum_{i=1}^{u} c_i$.

Lemma 2.12 *If there exists a gap-predicate-pair* $\mathcal{A}_1, \mathcal{A}_2$ *so that the predicates are independently samplable over the space* $\cup_I \mathcal{S}_{n,k,t}(I)$*, it follows that* PR$[n,k,t]$ *is solvable with overwhelming probability.*

Proof. First we show how to obtain an $i \in I$ with overwhelming probability. Let $\mathcal{A}_1, \mathcal{A}_2$ be a gap-predicate-pair and denote the non-negligible probability of revealing an index of the index-solution-set by $\alpha(n)$. Suppose we are given some $X \in \mathcal{S}_{n,k,t}$, let $I := I(X)$.

Since $\alpha(n)$ is non-negligible it follows that $\alpha(n) \geq \frac{1}{n^c}$ for some c and sufficiently large n. Let $N := n^{2c+1}$. Consider the following procedure $\mathcal{B}$: first compute the values $a_i := \sum S_{\mathcal{A}_1}(N,i,X)$ and $a'_i := \sum S_{\mathcal{A}_2}(N,i,X)$ for all $i = 1, \ldots, n$ (note that $S_{\mathcal{A}_1}(N,i,X) := \langle \mathcal{A}_1(i,X_1), \ldots, \mathcal{A}_1(i,X_N) \rangle$ and similarly for $S_{\mathcal{A}_2}$). For all $i \in \{1, \ldots, n\}$ check the difference $|a_i - a'_i|$. If it is discovered that for some i, $|a_i - a'_i| \geq \frac{2n^{c+1}}{3}$, output i as a "good" index (i.e. an index that belongs in I). If no such i is discovered the procedure fails.

We show that for any $X \in \mathcal{S}_{n,k,t}$ the above procedure returns an element of $I(X)$ with overwhelming probability. Let A_i and A'_i be the random variables that correspond to the computed values a_i and a'_i. Let μ_i and μ'_i denote the expected values of A_i and A'_i. By definition it holds that A_i, A'_i follow the Binomial probability distribution over N Bernoulli trials with probability of success $p_i := \mathbf{Prob}_{X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{A}_1(i,X) = 1]$ and $p'_i := \mathbf{Prob}_{X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{A}_2(i,X) = 1]$ respectively. Using the Chernoff bound we have that for $\epsilon > 0$, $\mathbf{Prob}[|A_i - Np_i| > \epsilon N] \leq 2e^{-2\epsilon^2 N}$ and $\mathbf{Prob}[|A'_i - Np'_i| > \epsilon N] \leq 2e^{-2\epsilon^2 N}$. Now observe that:

$$|Np_i - Np'_i| - |A_i - Np_i| - |A'_i - Np'_i| \leq |A_i - A'_i| \leq |Np_i - Np'_i| + |A_i - Np_i| + |A'_i - Np'_i|$$

Consider the following two facts:
(a) Suppose that $i \notin I$; then it holds that $q_i := \mathbf{Prob}[|A_i - A'_i| \geq \frac{n^{c+1}}{2}]$ is negligible in n. Indeed, $q_i \leq \mathbf{Prob}[|A_i - Np_i| + |A'_i - Np'_i| + |Np_i - Np'_i| \geq \frac{n^{c+1}}{2}]$. Now because $|p_i - p'_i|$ is negligible it follows that for sufficiently large n it holds that $|p_i - p'_i| < \frac{1}{6n^c}$. As a result $q_i \leq \mathbf{Prob}[|A_i - Np_i| + |A'_i - Np'_i| \geq \frac{n^{c+1}}{3}]$. It follows that $q_i \leq \mathbf{Prob}[|A_i - Np_i| \geq \frac{n^{c+1}}{6}] + \mathbf{Prob}[|A'_i - Np'_i| \geq \frac{n^{c+1}}{6}]$ and using the Chernoff bound for $\epsilon := \frac{1}{6n^{c-1}}$ we conclude that $q_i \leq 4e^{-(2/36)n^2}$ which is clearly negligible.
(b) Suppose that $i_0 \in I$ is the index for which $\alpha(n) = |p_{i_0} - p'_{i_0}|$ is non-negligible. The probability $q_{i_0} := \mathbf{Prob}[|A_{i_0} - A'_{i_0}| \leq \frac{2n^{c+1}}{3}]$ is negligible in n: first observe that $q_{i_0} \leq \mathbf{Prob}[|A_{i_0} - Np_{i_0}| + |A'_{i_0} - Np'_{i_0}| \geq N|p_{i_0} - p'_{i_0}| - \frac{2n^{c+1}}{3}]$. We know that $|p_{i_0} - p'_{i_0}| \geq \frac{1}{n^c}$ for sufficiently large n. As a result $q_{i_0} \leq \mathbf{Prob}[|A_{i_0} - Np_{i_0}| + |A'_{i_0} - Np'_{i_0}| \geq \frac{n^{c+1}}{3}]$. This probability was shown in case (a) above to be negligible.

Using the above two facts we deduce the following about the procedure $\mathcal{B}$:
1. The procedure fails with negligible probability. This is because of fact (b).

2. The procedure will report an index that is not in the index-solution set with negligible probability. This is because of fact (a).

It follows that, given *any* $X \in \mathcal{S}_{n,k,t}$, $\mathcal{B}$ reports an index of the index solution set $I(X)$ with overwhelming probability. Moreover for such index i_1 it will hold that $i_1 \leq n-k$ with overwhelming probability (because of the corresponding property of the gap-predicate-pair).

Now we modify the instance X as follows: we substitute the i_1-th point with the n-th point to obtain the altered instance X_2. Subsequently we repeat the procedure $\mathcal{B}$ that will recovers an index of the index-solution-set (different from i_1). By repeating the above k times we obtain k points of the solution polynomial of X and the solution follows by interpolation. This will be done with overwhelming probability. ■

Fact 2.13 *Violating the* DPR *by an independently samplable gap-predicate-pair with parameters* $[n,k,t]$ *implies that* PR$[n,k,t]$ *is solvable with overwhelming probability.*

3 Hardness of Recovering Partial Information of any Specific Polynomial Value

In this section we show that PR$[n,k,t]$ "leaks no partial information" about any specific polynomial value under the DPR-Assumption. In particular, we show that for some fixed value $w \in \mathbb{F}$, given an instance $X := \{\langle z_i, y_i\rangle\}_{i=1}^{n} \in \mathcal{S}_{n,k,t}$ with $w \notin \{z_1,\ldots,z_n\}$, we get no polynomial advantage in predicting the value of *any* function g over the polynomial value $s_X(w)$ for $s_X(w)$ drawn from any polynomially samplable probability distribution $\mathcal{D}$, unless the DPR fails for parameters $[n,k-1,t]$. In the remaining of the section we will fix $w \in \mathbb{F}$ and we will assume that $\mathcal{S}_{n,k,t}$ does not contain instances with w among the z-values (which is a negligible probability event). The generality of the proof stems from the fact that we can map a PR$[n,k-1,t]$-instance X into a PR$[n,k,t]$-instance X' of which we can select the value $s_{X'}(w)$. Then, we can use any algorithm that makes a non-negligible prediction regarding some property of $s_{X'}(w)$ to extract a parameterized predicate that is sensitive to a parameter choice inside the index-solution-set. This predicate yields a gap-predicate-pair that violates DPR$[n,k-1,t]$.

For the rest of the section fix some value $w \in \mathbb{F}$. Next, we formalize the concept of "leaking no partial information." Informally, we can describe the definition as follows: for any PPT that predicts the value of $g(s_X(w))$ given a PR instance, there is another algorithm with essentially the same functionality that operates *without* the PR instance.

Definition 3.1 PR$[n,k,t]$ *leaks no partial information means that for all poly-time computable* $g : \mathbb{F} \to R$ *and all polynomial-time samplable probability distributions* $\mathcal{D}_1$ *over* $\mathbb{F}$ *it holds: for all* PPT $\mathcal{A}$ *there exists a* PPT $\mathcal{A}'$ *such that the following is negligible in* n*:*

$$|\ \mathbf{Prob}_{r\in_U\mathcal{R};X\in_{\mathcal{D}_1^w}\mathcal{S}_{n,k,t}}[\mathcal{A}(r,X)=g(s_X(w))] - \mathbf{Prob}_{r'\in_U\mathcal{R}';u\in_{\mathcal{D}_1}\mathbb{F}}[\mathcal{A}'(r')=g(u)]\ |$$

A consequence of lemma 2.9 is that the definition above can be made more specific so that: for all PPT $\mathcal{A}$ there exists a PPT $\mathcal{A}'$ so that for all $I \subseteq \{1,\ldots,n\}$ with $|I| = t$ it holds that the following is negligible in n:

$$|\ \mathbf{Prob}_{r\in_U\mathcal{R};X\in_{\mathcal{D}_1^w}\mathcal{S}_{n,k,t}(I)}[\mathcal{A}(r,X)=g(s_X(w))] - \mathbf{Prob}_{r'\in_U\mathcal{R}';u\in_{\mathcal{D}_1}\mathbb{F}}[\mathcal{A}'(r')=g(u)]\ |$$

So, the probability of success of any PPT $\mathcal{A}$ is taken over $\mathcal{S}_{n,k,t}(I)$ following the distribution $\mathcal{D}_1^w$, independently of the index-solution-set I. The core of the proof that PR leaks no partial information is the following lemma:

Lemma 3.2 *Suppose that there is a poly-time computable $g : \mathbb{F} \rightarrow R$ and a probability distribution $\mathcal{D}_1$ for which* PR$[n,k,t]$ *leaks partial information. Then there exists a* PPT $\mathcal{B}$ *such that for all $I \subseteq \{1,\ldots,n\}$ with $|I| = t$, if $\beta_i(n) := \mathbf{Prob}_{\rho \in_U \Re; X \in_U \mathcal{S}_{n,k-1,t}(I)}[\mathcal{B}(i,\rho,X) = 1]$ with $i \in \{0,\ldots,n\}$ it holds that*
1. For all $i \notin I$ $|\beta_{i-1}(n) - \beta_i(n)|$ is negligible.
2. There exists an $i_0 \in I$ such that $|\beta_{i_0-1}(n) - \beta_{i_0}(n)|$ is non-negligible and $i_0 \leq n-k+1$.

Proof. For simplicity we assume that $\mathcal{D}_1$ is the uniform distribution. The proof is similar in both cases (see below for comments in the case $\mathcal{D}$ is not uniform). Regarding the success probability $\alpha(n)$ of $\mathcal{A}$ we have that for all I with $|I| = t$, and for all PPT $\mathcal{A}'$ the probability distance below is non-negligible in n:

$$|\ \mathbf{Prob}_{r \in_U \mathcal{R}; X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{A}(r,X) = g(s_X(w))] - \mathbf{Prob}_{r' \in_U \mathcal{R}'; u \in_U \mathbb{F}}[\mathcal{A}'(r') = g(u)]\ |$$

Let $\mathcal{B}$ be the following PPT that operates on $\mathcal{S}_{n,k-1,t}(I)$ with random input string $\rho := \langle u,y,r\rangle \in_U \Re := \mathbb{F} \times \mathbb{F}^n \times \mathcal{R}$ (in the case $\mathcal{D}$ is not uniform, u is not part of the random input of $\mathcal{B}$ but rather it is sampled using the PPT that samples $\mathcal{D}_1$). Given some $X \in \mathcal{S}_{n,k-1,t}(I) := \{\langle z_i, y_i\rangle\}_{i=1}^n$. The set of pairs $X^* := \{\langle z_i, (z_i - w)y_i + u\rangle\}_{i=1}^n$ is computed. Note that X^* is a random instance of $\mathcal{S}_{n,k,t}(I)$ (similarly if u was distributed according to some non-uniform distribution $\mathcal{D}_1$, then X would follow the corresponding distribution $\mathcal{D}_1$ over $\mathcal{S}_{n,k,t}$). Subsequently the y-part of the first i pairs of X^* is randomized by substituting them with the first i values of the given string $y \in \mathbb{F}^n$. The resulting partially randomized instance is denoted by X_i^*. Then $\mathcal{A}$ is simulated on input (r, X_i^*). If $\mathcal{A}$ returns $g(u)$ (i.e. $\mathcal{A}$ is correct) then $\mathcal{B}$ returns 1 (0 otherwise).

It is easy to see that $\beta_0(n) = \alpha(n)$. When $i = n-k+1$, $\mathcal{B}$ completely randomizes the first $n-k+1$ positions of the y-part of the constructed $\mathcal{S}_{n,k,t}(I)$ instance. Consider a PPT $\mathcal{A}'$ that first samples a random $Y \in \mathcal{S}_n := (\mathbb{F})_n \times \mathbb{F}^n$ (where $(\mathbb{F})_n$ denotes the set of all n-tuples over $\mathbb{F}$ without repetitions) and then simulates $\mathcal{A}$ on Y. It holds that,

$$\alpha'(n) := \mathbf{Prob}_{u \in_U \mathbb{F}}[\mathcal{A}(\cdot) = g(u)] = \mathbf{Prob}_{r \in_U \mathcal{R}; Y \in_U \mathcal{S}_n; u \in_U \mathbb{F}}[\mathcal{A}(r,Y) = g(u)]$$

Let $C' := \{\langle r,Y,u\rangle \mid \mathcal{A}(r,Y) = g(u); Y \in \mathcal{S}_n\}$, it holds that: $\alpha'(n) = \frac{\#C'}{\#\mathcal{R} \times \mathcal{S}_n \times \mathbb{F}}$. We want to compare the probability $\beta_{n-k+1}(n)$ to $\alpha'(n)$. Define the mapping $J(i,u,y,X) := \langle X_i^*, u\rangle$, where X_i^* is defined as in the description of $\mathcal{B}$. Given a certain $\langle Y,u\rangle$ for a $Y \in \mathcal{S}_n$ we want to compute how many pre-images of the form $\langle y,X\rangle$ has, under the mapping $J(n-k+1,u,\cdot,\cdot)$. Let $h := |I \cap \{n-k+2,\ldots,n\}|$; obviously $h \leq k-1$. Fix h values of the polynomial-solution of X to the corresponding y-positions of Y and $k-1-h$ values of the non-polynomial values of X to the corresponding positions of Y. This leaves a total of $|\mathbb{F}|^{n-t+k-1}$ choices for the pre-images of $\langle Y,u\rangle$. It follows that:

$$\beta_{n-k+1}(n) = \frac{|\mathbb{F}|^{n-t+k-1}\#C'}{\#\mathcal{R} \times \mathbb{F} \times \mathbb{F}^n \times \mathcal{S}_{n,k-1,t}(I)} = \frac{|\mathbb{F}|^{n-t+k-1}\#C'}{\#\mathcal{R} \cdot |\mathbb{F}| \cdot (|\mathbb{F}|)_n \cdot |\mathbb{F}|^{2n-t+k-1}} =$$

$$= \frac{\#C'}{\#\mathcal{R} \cdot |\mathbb{F}| \cdot (|\mathbb{F}|)_n \cdot |\mathbb{F}|^n} = \alpha'(n)$$

From the assumption of the theorem it is immediate that $|\alpha(n) - \alpha'(n)|$ is non-negligible and as a result we conclude that $|\beta_0(n) - \beta_{n-k+1}(n)|$ is non-negligible in n. It follows easily that for some $i_0 \in \{1, \ldots, n-k+1\}$ it should be the case that $|\beta_{i_0-1}(n) - \beta_{i_0}(n)|$ is non-negligible (by the triangular inequality). It remains to show that it cannot be the case that $i_0 \notin I$.

In particular we will show that for any $i \notin I$ it holds that $|\beta_{i-1}(n) - \beta_i(n)|$ is negligible.

Let $C_i := \{\langle r, y, X, u\rangle \mid \mathcal{A}(r, X_i^*) = g(u); X_i^* = J(i, X, y, u); X \in \mathcal{S}_{n,k-1,t}(I)\}$. It follows that

$$\mathbf{Prob}_{r \in_U \mathcal{R}; y \in_U \mathbb{F}^n; X \in \mathcal{S}_{n,k,t}(I); u \in_U \mathbb{F}}[\mathcal{B}(i, u|y|r, X) = 1] = \frac{\#C_i}{\#\mathcal{R} \times \mathbb{F}^n \times \mathcal{S}_{n,k-1,t}(I) \times \mathbb{F}}$$

Suppose $i \notin I$. Next we will compare the number of elements of $\#C_i$ and $\#C_{i-1}$.

Let $\langle r, y^-, X^-, u\rangle$ be an element of $\mathcal{R} \times \mathbb{F}^n \times \mathcal{S}_{n,k-1,t}(I) \times \mathbb{F}$ with the i-th position of y and the i-th y-position of X left "blank." Define $V_{r,y^-,X^-,u} := \{v \mid \mathcal{A}(r, y^-/v/X^-) = g(u)\}$; here $y^-/v/X^-$ denotes the set of pairs $\{\langle z_i, y_i'\rangle\}_{i=1}^n$ such that up to $i-1$ y_i' agrees with y, $y_i' = v$ and from $i+1$ and on $y_i' = (z_i - w)y_i + u$ (where $X^- = \{\langle z_i, y_i\rangle\}_{i=1}^n$). Any $\langle r, y^-, X^-, u\rangle$ together with some $v \in V_{r,y^-,X^-,u}$ can be extended to:

- $|\mathbb{F}|$ tuples $\langle r, y^-_{[v']}, X^-_{[v]}, u\rangle$ that belong in C_{i-1}; the number of tuples stems from the free choice of $v' \in \mathbb{F}$.
- $|\mathbb{F}|$ tuples $\langle r, y^-_{[v]}, X^-_{[v']}, u\rangle$ that belong in C_i; the number of tuples stems from the free choice of $v' \in \mathbb{F}$.

It follows that

$$\#C_i = |\mathbb{F}| \sum_{\langle r, y^-, X^-, u\rangle} \#V_{r,y^-,X^-,u} = \#C_{i-1}$$

and as a result $\beta_{i-1}(n) = \beta_i(n)$. ■

The proof of this Lemma is a crucial contribution. It exhibits the two main proof-techniques used throughout; one technique involves controlling portions of the instance's solution, whereas the other technique involves a "walking argument" over the points of the instance. Now observe that if $\mathcal{A}_1(i, r, X) := \mathcal{B}(i, r, X)$ and $\mathcal{A}_2(i, r, X) := \mathcal{B}(i-1, r, X)$, it follows easily that $\mathcal{A}_1, \mathcal{A}_2$ is a gap-predicate-pair. As a result,

Theorem 3.3 *Suppose that there is a poly-time computable $g : \mathbb{F} \to R$ and a probability distribution $\mathcal{D}_1$ for which* PR$[n, k, t]$ *leaks partial information. Then the DPR-Assumption fails for parameters* $[n, k-1, t]$.

Proof. The proof is immediate from lemma 3.2 and the definition of the DPR assumption. ■

In the rest of the section we present special cases of the above Theorem which appear frequently in cryptographic settings. Let us assume that the distribution $\mathcal{D}_1$ is uniform. Let $g : \mathbb{F} \to R$ be a poly-time computable function. Define $\mathbb{F}_a = \{u \mid g(u) = a; u \in \mathbb{F}\}$ for any $a \in R$. We say that g is *balanced* if for all $a \in R$ and all polynomials q it holds that $|\frac{|\mathbb{F}_a|}{|\mathbb{F}|} - \frac{1}{|R|}| < \frac{1}{q(\log|\mathbb{F}|)}$ (for sufficiently large $|\mathbb{F}|$). The balanced property means that any image

under g corresponds to roughly the same number of pre-images. This is a very general condition that applies to individual bits of elements of $\mathbb{F}$ as well as to various length bit-sequences of elements of $\mathbb{F}$.

Naturally, guessing an unknown value of a balanced function with a uniformly distributed pre-image cannot be done with probability significantly greater than $1/|R|$:

Fact 3.4 *Let $g : \mathbb{F} \to R$ be balanced, poly-time computable and let n be polynomially related to $\log|\mathbb{F}|$. Then, for any* PPT *in n, $\mathcal{A}'$, if $\alpha'(n) := \mathbf{Prob}_{r' \in_U \mathcal{R}'; u \in_U \mathbb{F}}[\mathcal{A}'(r') = g(u)]$ it holds that $|\ \alpha'(n) - \frac{1}{|R|}\ |$ is negligible in $\log|\mathbb{F}|$.*

Proof. Let $\mathcal{R}'_a := \{r' \mid \mathcal{A}'(r') = a\}$ for any $a \in R$. Note that it holds that $\cup_{a \in R} \mathcal{R}'_a = \mathcal{R}'$. Let q be any polynomial; now because g is balanced:

$$\alpha'(n) = \frac{\sum_{a \in R} |\mathbb{F}_a||\mathcal{R}'_a|}{|\mathbb{F}||\mathcal{R}'|} < \frac{\sum_{a \in R} |\mathcal{R}'_a|}{|\mathcal{R}'|}\left(\frac{1}{|R|} + \frac{1}{q(\log|\mathbb{F}|)}\right) = \frac{1}{|R|} + \frac{1}{q(\log|\mathbb{F}|)}$$

and

$$\alpha'(n) = \frac{\sum_{a \in R} |\mathbb{F}_a||\mathcal{R}'_a|}{|\mathbb{F}||\mathcal{R}'|} > \frac{\sum_{a \in R} |\mathcal{R}'_a|}{|\mathcal{R}'|}\left(\frac{1}{|R|} - \frac{1}{q(\log|\mathbb{F}|)}\right) = \frac{1}{|R|} - \frac{1}{q(\log|\mathbb{F}|)}$$

consequently $|\alpha'(n) - \frac{1}{|R|}|$ is negligible in $\log|\mathbb{F}|$. ■

The corollary of fact 3.4 and theorem 3.3 is the following:

Corollary 3.5 *For any balanced $g : \mathbb{F} \to R$, the success of any PPT $\mathcal{A}$ that given $X \in \mathcal{S}_{n,k,t}$, computes the value $g(s_X(w))$ is only by a negligible fraction different than $1/|R|$ unless the* DPR*-Assumption fails for parameters $[n, k-1, t]$.*

More specifically we can give the following examples of balanced predicates/functions that are hard to compute given a PR$[n, k, t]$-instance:

Proposition 3.6 *The following problems are hard under the* DPR$[n, k-1, t]$*:*

1. *Let* $\mathrm{BIT}_l(a)$ *denote the l-th LSB of $a \in \mathbb{F}$. Given $X \in \mathcal{S}_{n,k,t}$ predict* $\mathrm{BIT}_l(s_X(w))$ *with non-negligible advantage where l represents any bit, except the* $\log\log|\mathbb{F}|$ *most significant — in particular l as a function of* $\log|\mathbb{F}|$ *should satisfy that for any $c \in \mathbb{N}$, $l < \log|\mathbb{F}| - c\log\log|\mathbb{F}|$ for sufficiently large* $\log|\mathbb{F}|$*.*

2. *Let* $\mathrm{BITS}_l(a)$ *denote the sequence of the l least significant bits of $a \in \mathbb{F}$. Given $X \in \mathcal{S}_{n,k,t}$ predict* $\mathrm{BITS}_l(s_X(w))$ *with probability $\frac{1}{2^l} + \alpha(n)$ where $\alpha(n)$ is non-negligible.*

3. *Let* $\mathrm{QR}(a)$ *be 1 iff $a \in \mathbb{F}$ is a quadratic residue, and assume $\mathbb{F}$ is of prime order. Given $X \in \mathcal{S}_{n,k,t}$ predict* $\mathrm{QR}(s_X(w))$ *with non-negligible advantage.*

Proof. (1) Let H_v denote the number of elements of $\mathbb{F}$ that their l-th LSB is v (where $v \in \{0,1\}$). We want to show that $\frac{|\mathbb{F}_0| - |\mathbb{F}_1|}{|\mathbb{F}|}$ is negligible in $\log|\mathbb{F}|$. Let $f := |\mathbb{F}| \bmod 2^l$. If is easy to see that $|\mathbb{F}_0| - |\mathbb{F}_1| = f$ if $f \leq 2^{l-1}$ and that $|\mathbb{F}_0| - |\mathbb{F}_1| = 2^l - f$ if $f > 2^{l-1}$. At any rate we would like to show that $\frac{2^{l-1}}{|\mathbb{F}|}$ is negligible in $\log|\mathbb{F}|$, which is easy to see under the condition of the theorem.

(2) For any bitstring $b \in \{0,1\}^l$ (where $l = 1, \ldots, \lfloor \log |\mathbb{F}| \rfloor$) it holds that $|\mathbb{F}_b|$ is either (a) $\lfloor \frac{|\mathbb{F}|}{2^l} \rfloor$ or (b) $\lfloor \frac{|\mathbb{F}|}{2^l} \rfloor + 1$. Case (a): $|\frac{|\mathbb{F}_b|}{|\mathbb{F}|} - \frac{1}{|R|}| = |\frac{\lfloor |\mathbb{F}|/2^l \rfloor}{|\mathbb{F}|} - \frac{1}{2^l}|$ which is easy to see that is negligible in $\log |\mathbb{F}|$. Case (b) is similar.

(3) Straightforward as we assume that $\mathbb{F}$ is a field of prime order. ■

We note that the exclusion of the $\log \log |\mathbb{F}|$ most significant bits from the item (1) above is independent of our treatment as depending on the order of the field they may be easy to guess, and as a result BIT_l might not be balanced. Note that if the finite field is chosen appropriately all bits of $s_X(w)$ will be hard: e.g. if we restrict to finite fields $\mathbb{F}$ such that there is a $c \in \mathbb{N}$: $|\mathbb{F}| - 2^{\lfloor \log |\mathbb{F}| \rfloor} \leq (\log |\mathbb{F}|)^c$ then all bits will be hard (e.g. a field of numbers modulo a Mersenne prime):

Corollary 3.7 *Under the* DPR*-Assumption with parameters* $[n, k-1, t]$*, predicting any bit in a point of the graph of the solution polynomial of a* $\mathrm{PR}[n,k,t]$ *instance is hard.*

A natural question to ask at this point is whether simultaneously more than one point of the polynomial solution enjoys the hardness of extraction properties showed in theorem 3.3. In particular we can extend the definition of leaking partial information to many points at the same time as follows:

Definition 3.8 *Fix some* $w_1, \ldots, w_h \in \mathbb{F}$ *with* $h \in \{1, \ldots, k-1\}$*. We say that* $\mathrm{PR}[n,k,t]$ *leaks no partial information for* h *points simultaneously if for all poly-time computable* $g : \mathbb{F}^h \to R$ *and all polynomial-time samplable probability distributions* $\mathcal{D}_h$ *over* $\mathbb{F}^h$ *it holds: for all* PPT $\mathcal{A}$ *there exists a* PPT $\mathcal{A}'$ *such that the following is negligible in* n*:*

$$|\ \mathbf{Prob}_{r \in_U \mathcal{R}; X \in_{(\mathcal{D}_h^{w_1,\ldots,w_h})} \mathcal{S}_{n,k,t}}[\mathcal{A}(r, X) = g(s_X(w_1), \ldots, s_X(w_h))]$$

$$- \mathbf{Prob}_{r' \in_U \mathcal{R}'; \mathbf{u} \in_{\mathcal{D}_h} \mathbb{F}^h}[\mathcal{A}'(r') = g(\mathbf{u})]\ |$$

By choosing the appropriate parameters for the DPR assumption it is possible to show hardness of partial information extraction even in this extended setting:

Theorem 3.9 *Suppose that there is a poly-time computable* $g : \mathbb{F}^h \to R$ *and a probability distribution* $\mathcal{D}_h$ *for which* $\mathrm{PR}[n,k,t]$ *leaks partial information for* h *points simultaneously. Then the DPR-Assumption fails for parameters* $[n, k-h, t]$*.*

Proof. The proof of the theorem is a straightforward multidimensional extension of the proof of lemma 3.2. ■

4 Pseudorandomness

In this section we will show that distinguishing instances of $\mathrm{PR}[n,k,t]$ from random elements of $\mathcal{S}_n := (\mathbb{F})_n \times \mathbb{F}^n$ is hard under the DPR-Assumption (which essentially amounts to saying that instances of $\mathrm{PR}[n,k,t]$ are pseudorandom under the DPR). We start with a standard definition:

Definition 4.1 *Let $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ be a family of sets, such that $\mathcal{F}_n$ contains* all *possible choices of elements of size n. Two families of sets with $A_n, B_n \subseteq \mathcal{F}_n$ are* (polynomial-time, computationally) indistinguishable *if for any* PPT *predicate $\mathcal{A}$,*

$$| \mathbf{Prob}_{r\in_U\mathcal{R};X\in_U A_n}[\mathcal{A}(r,X)=1] - \mathbf{Prob}_{r\in_U\mathcal{R};X\in_U B_n}[\mathcal{A}(r,X)=1] |$$

is negligible in n. If on the other hand there is an $\mathcal{A}$ for which the probability above is non-negligible in n, we will say that $\mathcal{A}$ is a distinguisher *for A_n, B_n. A family of sets A_n is called* pseudorandom *if it is indistinguishable from $\mathcal{F}_n$.*

Note that for this section we consider $B_n = \mathcal{F}_n := \mathcal{S}_n = (\mathbb{F})_n \times \mathbb{F}^m$ and $A_n := \mathcal{S}_{n,k,t}$ (the set of PR$[n,k,t]$ instances). Let $\mathcal{A}$ be a distinguisher for $\mathcal{S}_{n,k,t}$ and $\mathcal{S}_n$. Because of lemma 2.9 it holds that the particular choice of the index-solution set I is independent of the distinguishing probability, i.e. for all $I \subseteq \{1,\ldots,n\}$, $|I| = t$, it holds that the following is non-negligible in n:

$$| \mathbf{Prob}_{r\in_U\mathcal{R};X\in_U\mathcal{S}_{n,k,t}(I)}[\mathcal{A}(r,X)=1] - \mathbf{Prob}_{r\in_U\mathcal{R};X\in_U\mathcal{S}_n}[\mathcal{A}(r,X)=1] |$$

In other words lemma 2.9 suggests that any distinguisher between $\mathcal{S}_{n,k,t}$ and $\mathcal{S}_n$ also serves as a distinguisher between $\mathcal{S}_{n,k,t}(I)$ and $\mathcal{S}_n$ for all subsets I.

The core of the pseudorandomness proof is the next lemma that given such distinguisher it shows how to extract a parameterized over $\{0,\ldots,n\}$ predicate $\mathcal{B}$ that its behavior is sensitive to some choice of the parameter that belongs in the index-solution-set of the given instance.

Lemma 4.2 *Let $\mathcal{A}$ be a* PPT *predicate s.t. for all $I \subseteq \{1,\ldots,n\}$ with $|I| = t$, $\mathcal{A}$ is a distinguisher for $\mathcal{S}_{n,k,t}(I)$ and $\mathcal{S}_n$. Then there exists an* PPT *$\mathcal{B}$, for which it holds that for all $I \subseteq \{1,\ldots,n\}$ with $|I| = t$, there exists a $i_0 \in I$ with $i_0 \leq n-k$, such that if*

$$\beta_i(n) := \mathbf{Prob}_{X\in_U\mathcal{S}_{n,k,t}(I);\rho\in_U\Re}[\mathcal{B}(i,\rho,X)=1] \quad \text{for} \quad i\in\{0,\ldots,n\}$$

it holds that $|\beta_{i-1}(n) - \beta_i(n)|$ is negligible for any $i \notin I$ and non-negligible for i_0.

Proof. Let $\mathcal{R}$ be the set of random strings used by the distinguisher $\mathcal{A}$. $\mathcal{B}$ is the following algorithm: given y, r, i, X where $i \in \{0,\ldots,n\}$ and $y \in \mathbb{F}^m$ substitute the first i y-positions of $X = \{\langle z_i, y_i\rangle\}_{i=1}^n$ by the first i values of y; denote this partially randomized instance by $J(i,X,y)$. Then $\mathcal{B}$ simulates $\mathcal{A}$ on input r and $J(i,X,y)$. Note that the randomness used by $\mathcal{B}$ is $\rho := \langle r, y\rangle \in \Re$ where $\Re := \mathcal{R}\times\mathbb{F}^m$.

Define the probabilities $\alpha_1(n) := \mathbf{Prob}_{r\in_U\mathcal{R};X\in_U\mathcal{S}_{n,k,t}(I)}[\mathcal{A}(r,X)=1]$ and $\alpha_2(n) := \mathbf{Prob}_{r\in_U\mathcal{R};X\in_U\mathcal{S}_n}[\mathcal{A}(r,X)=1]$. Define the following sets:

- $C_i := \{\langle r,y,X\rangle \mid \mathcal{A}(r,Y)=1; r\in\mathcal{R}; y\in\mathbb{F}^m; X\in\mathcal{S}_{n,k,t}(I); Y = J(i,X,y)\}$
- $V_1 := \{\langle r,X\rangle \mid \mathcal{A}(r,X)=1; r\in\mathcal{R}; X\in\mathcal{S}_{n,k,t}(I)\}$
- $V_2 := \{\langle r,X\rangle \mid \mathcal{A}(r,X)=1; r\in\mathcal{R}; X\in\mathcal{S}_n\}$

It is easy to see that $\beta_i(n) = \frac{\#C_i}{\#\mathcal{S}_{n,k,t}(I)\times\mathcal{R}\times\mathbb{F}^n}$; $\alpha_1(n) = \frac{\#V_1}{\#\mathcal{S}_{n,k,t}(I)\times\mathcal{R}}$ and that $\alpha_2(n) = \frac{\#V_2}{\#\mathcal{S}_n\times\mathcal{R}}$. Moreover from the lemma's hypothesis we know that $|\ \alpha_1(n) - \alpha_2(n)\ |$ is non-negligible.

Consider C_0; it is immediate that $\#C_0 = |\mathbb{F}|^n \#V_1$ and as a result $\beta_0(n) = \alpha_1(n)$.

Consider C_{n-k}; let $h := |I \cap \{n-k+1, \ldots, n\}|$, obviously it holds that $h \in \{0, \ldots, k\}$. Let $Y := \{\langle z_i, y_i \rangle\}_{i=1}^{n} \in \mathcal{S}_n$. It is not difficult to show that Y has $|\mathbb{F}|^{n-t+k}$ pre-images under $J(n-k, \cdot, \cdot)$. It follows that,

$$\beta_{n-k}(n) = \frac{\#C_{n-k}}{\#\mathcal{S}_{n,k,t}(I) \times \mathcal{R} \times \mathbb{F}^n} = \frac{|\mathbb{F}|^{n-t+k} \#V_2}{\#\mathcal{S}_{n,k,t}(I) \times \mathcal{R} \times \mathbb{F}^n} = \alpha_2(n)$$

We conclude that $|\beta_0(n) - \beta_{n-k}(n)|$ is non-negligible. This means that there has to be an $i_0 \in \{1, \ldots, n-k\}$ such that $|\beta_{i_0-1}(n) - \beta_{i_0}(n)|$ is non-negligible (using the triangular inequality).

To complete the proof we show that when $i \notin I$ it holds that $|\beta_{i-1}(n) - \beta_i(n)|$ is negligible.

Fix $i \notin I$. Let y^- denote a $\mathbb{F}^n$ vector with its i-th position "blank" (so essentially a $\mathbb{F}^{n-1}$ vector); in a similar manner define X^- to be an instance of $\mathcal{S}_{n,k,t}(I)$ with its i-th y-position "blank". Denote by $y^-_{[v]}$ the $\mathbb{F}^n$ vector that has v "filled" in its i-th position. Similarly define $X^-_{[v]}$.

Let $V_{r,y^-,X^-} := \{v \mid \mathcal{A}(r, y^-/v/X^-) = 1\}$, where the notation $y^-/v/X^-$ stands for an element Y of $\mathcal{S}_n$ s.t. its y-part is comprised of the first $i-1$ elements of y^-, followed by v, followed by the $n-i$ final elements of the y-part of X^-, and $z(Y) = z(X^-)$ (where $z(\cdot)$ denotes the z-elements of a PR instance). Any $v \in V_{r,y^-,X^-}$ together with $\langle r, y^-, X^- \rangle$ can be extended to:

- $|\mathbb{F}|$ tuples $\langle r, y^-_{[v']}, X^-_{[v]} \rangle \in C_{i-1}$ — the fact that there are $|\mathbb{F}|$ tuples follows from the free choice of v'.
- $|\mathbb{F}|$ tuples $\langle r, y^-_{[v]}, X^-_{[v']} \rangle \in C_i$, (recall: $i \notin I$) — the fact that there are $|\mathbb{F}|$ tuples follows from the free choice of v'.

It follows:

$$\#C_i = |\mathbb{F}| \sum_{\langle r, y^-, X^- \rangle} \#V_{r,y^-,X^-} = \#C_{i-1}$$

as a result $\beta_{i-1}(n) = \beta_i(n)$. ■

Now observe that if $\mathcal{A}_1(i, r, X) := \mathcal{B}(i, r, X)$ and $\mathcal{A}_2(i, r, X) := \mathcal{B}(i-1, r, X)$, it follows easily that $\mathcal{A}_1, \mathcal{A}_2$ is a gap-predicate-pair. As a result,

Theorem 4.3 *Under the* DPR-*Assumption for* $[n, k, t]$*, the set of instances* $\mathcal{S}_{n,k,t}$ *is pseudo-random.*

5 Applications

5.1 One-Way Function with Built-in Semantic Security

In this section we present a one-way function based on polynomial reconstruction that acts as a "secure envelope" under the DPR-Assumption and can be used to build commitment schemes. Note that there are generic ways [Gol90, Na91, HILL99] for obtaining such cryptographic primitives based on the results we presented in sections 3 and 4, however describing a direct construction with improved concealment properties is interesting in its own right for efficiency and applicability purposes.

Definition 5.1 *A function $f : A_n \rightarrow B_n$ is one-way if f is polynomial-time computable and for any* PPT $\mathcal{A}'$ *it holds that* $\mathbf{Prob}_{r \in_U \mathcal{R}; a \in_U A_n}[\mathcal{A}'(f(a)) \in f^{-1}(f(a))]$ *is negligible in n.*

Fix some parameters $[n, k, t]$. The probabilistic function $F_{n,k,t} : \mathbb{F}^k \rightarrow \mathcal{S}_{n,k,t}$ operates as follows: given $\mathbf{x} \in \mathbb{F}^k$, it samples a random element $Y := \{\langle z_i, y_i \rangle\}_{i=1}^n$ of $\mathcal{S}_{n,k,t}$ so that (i) Y has a solution s_X that satisfies $s_X(0) = (\mathbf{x})_0, \ldots, s_X(k-1) = (\mathbf{x})_{k-1}$, and (ii) $\{z_1, \ldots, z_n\} \cap \{0, \ldots, k-1\} = \emptyset$.

We note here that $F_{n,k,t}$ is not an injection as it could be the case that $F_{n,k,t}(\mathbf{x}) = F_{n,k,t}(\mathbf{x}')$ for $\mathbf{x} \neq \mathbf{x}'$. This happens when the randomness selected to engulf the polynomial derived from $\mathbf{x}$ happens to correspond to several points of the graph of the polynomial defined by the vector $\mathbf{x}'$. Nevertheless this means that the PR instance generated by $F_{n,k,t}$ has two distinct solutions something that happens with negligible probability as shown in lemma 2.8 (given that $\log |\mathbb{F}| \geq 2n$ and $t > k$). As a result we consider $F_{n,k,t}$ to be an injection for all purposes of definition 5.1. Nevertheless it is important to point out that some user of $F_{n,k,t}$ may deliberately embed more than polynomial-solution into the output of the function $F_{n,k,t}$. As a result $F_{n,k,t}$ thought of as an encryption function enjoys a natural "ambiguous commitment" property.

Theorem 5.2 *Under* DPR$[n, k, t]$ *the function $F_{n,k,t}$ is a one-way function.*

Proof. Suppose that there is $\mathcal{A}$ with $\mathbf{Prob}[\mathcal{A}(F_{n,k,t}(\mathbf{x})) = \mathbf{x})]$ non-negligible, where the probability is taken over all $\mathbf{x} \in \mathbb{F}^k$ and the internal coin tosses of $\mathcal{A}$ and $F_{n,k,t}$. Obviously it holds that $\mathcal{A}$ solves the PR with non-negligible probability. Let $\mathcal{A}'$ be a PPT that first permutes the pairs on the input (instance of PR) and then simulates $\mathcal{A}$ on the permuted pairs. It is easy to show (cf. lemma 2.9) that for all $I \subseteq \{1, \ldots, n\}$, $|I| = t$, $\mathbf{Prob}_{X \in \mathcal{S}_{n,k,t}(I)}[\mathcal{A}'(X) = s_X]$ is non-negligible in n.

Now we show how to use $\mathcal{A}'$ to construct a gap-predicate-pair. Let $\mathcal{B}$ be a PPT that given $X \in \mathcal{S}_{n,k,t}(I)$ and $i \in \{0, \ldots, n\}$ it does the following: first it randomizes the first i y-positions of X and then simulates $\mathcal{A}'$ on this instance. If $\mathcal{A}'$ returns the correct answer (something that is checkable in polynomial-time — a proposed solution for a PR instance can be verified in poly-time), $\mathcal{B}$ returns 1, otherwise $\mathcal{B}$ returns 0.

It is easy to verify that $\mathbf{Prob}_{X \in \mathcal{S}_{n,k,t}(I)}[\mathcal{B}(0, X) = 1]$ is non-negligible function in n, whereas $\mathbf{Prob}_{X \in \mathcal{S}_{n,k,t}(I)}[\mathcal{B}(n-k, X) = 1]$ is negligible function in n since $\mathcal{A}'$ cannot predict a polynomial which has been completely randomized (cf. lemma 4.2). It follows that

$$|\mathbf{Prob}_{X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{B}(0, X) = 1] - \mathbf{Prob}_{X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{B}(n-k, X) = 1]|$$

is non-negligible in n and by the triangular inequality it follows that for some $i_0 \in \{1, \ldots, n-k\}$ it holds that

$$|\mathbf{Prob}_{X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{B}(i_0 - 1, X) = 1] - \mathbf{Prob}_{X \in_U \mathcal{S}_{n,k,t}(I)}[\mathcal{B}(i_0, X) = 1]|$$

is non-negligible in n. Using a similar argument as in proof of lemma 4.2 it can be shown that i_0 should be an element of I. It follows that $\mathcal{A}_1(i, \cdot) := \mathcal{B}(i-1, \cdot)$ and $\mathcal{A}_2(i, \cdot) := \mathcal{B}(i, \cdot)$ constitute a gap-predicate-pair and as a result the Decisional-PR assumption is violated. ■

Based on the results of section 3, we draw the following corollary:

Corollary 5.3 *Under* DPR$[n, k-1, t]$, $F_{n,k,t}$ *is a one-way function so that: if* $g : \mathbb{F} \to R$ *is some computable function, an adversary given* $V_x := F_{n,k,t}(\langle x, r_1, \ldots, r_{k-1}\rangle)$, *with* $r_1, \ldots, r_{k-1}$ *are selected at random over* $\mathbb{F}$, *gains no advantage in computing* $g(x)$ *even if* x *follows an adversarially chosen probability distribution.*

The above corollary suggests that V_x is a "secure envelope" for the value x. In fact it is possible to increase the ratio of concealed information as the following theorem reveals:

Theorem 5.4 *Let* $h \in \{1, \ldots, k-1\}$. *Under* DPR$[n, k-h, t]$, *if* $V_{\mathbf{x}} := F_{n,k,t}(\langle x_0, \ldots, x_{h-1}, r_1, \ldots, r_{k-h}\rangle)$, *where* $r_1, \ldots, r_{k-h}$ *are random elements of* $\mathbb{F}$, *then* $V_{\mathbf{x}}$ *leaks no partial information about* $\mathbf{x} := \langle x_0, \ldots, x_{h-1}\rangle$ *even if these values follow an adversarially chosen probability distribution over* $\mathbb{F}^h$.

Proof. The proof follows closely the arguments of lemma 3.2. Let $\mathcal{A}$ be a PPT and $\mathcal{D}_h$ a probability distribution over $\mathbb{F}^h$ for which the commitment of some values $\mathbf{x} := \langle x_0, \ldots, x_h\rangle$ leaks some partial information: i.e. for some poly-time computable function $g : \mathbb{F}^h \to R$, $\mathcal{A}$ computes the value $g(\mathbf{x})$ with non-negligible advantage. As a result and due to lemma 2.9 we can formulate the success probability of $\mathcal{A}$ as follows: for all $I \subseteq \{1, \ldots, n\}$, $|I| = t$,

$$\alpha(n) := \mathbf{Prob}_{r \in_U \mathcal{R}; X \in_{\mathcal{D}_h} \mathcal{S}_{n,k,t}(I)}[\mathcal{A}(r, X) = g(\langle s_X(0), \ldots, s_X(h-1)\rangle)]$$

The proof follows directly from theorem 3.9. ■

An interesting property of the above "secure envelope" is that the hidden value $\mathbf{x}$ can be superpolynomial size in the security parameter n. This is because the size of $\mathbf{x}$ is proportional to $\log|\mathbb{F}|$ which can be selected to be superpolynomial in the security parameter n without affecting the security of the primitive. The "secure envelope" properties of $F_{n,k,t}$ suggest that the PR-based one-way function can be used directly in the design of commitment schemes. More details about the use of $F_{n,k,t}$ in commitment schemes are presented in the next section.

5.2 Value Commitment

A value commitment scheme involves two players A and B that act in two phases: the commitment phase where A commits to some private input $\mathbf{x}$. The output of this phase denoted by $V_{\mathbf{x}}$ is transmitted to player B. The decommitment or "open" phase where A transmits the decommitment witness U to player B. Player B applies U on $V_{\mathbf{x}}$ (a process that reveals $\mathbf{x}$) and either accepts or rejects the commitment. A commitment scheme should be (i) binding: player A should not be able to "open" $V_{\mathbf{x}}$ to a value $\mathbf{x}' \neq \mathbf{x}$; (ii) hiding: player B should not be able to extract any partial information about $\mathbf{x}$ given $V_{\mathbf{x}}$. A commitment scheme is called "non-interactive", if no interaction is required from the two players (the communication flow is only from player A to player B).

We point here that using generic techniques ([Na91]) it is possible to derive a PR-based commitment scheme based on our pseudorandomness results of section 4. Nevertheless such generic techniques are typically expensive to implement and it is of interest to pursue more direct designs.

Theorem 5.4 suggests that the function $F_{n,k,t}$ can be used to commit to an element $\mathbf{x} \in \mathbb{F}^h$ by publishing $V_{\mathbf{x}}$ as the commitment value. The decommitment witness is defined to be the

index-solution-set I of $V_{\mathbf{x}}$. This scheme is non-interactive and hiding under the DPR$[n, k-h, t]$. Nevertheless the scheme is not binding for the commiter since player A might embed more than one solution in the instance $V_{\mathbf{x}}$ and open one of them at her choice; as a result the scheme applies only to the "honest commiter" case. By coupling the PR-based non-binding commitment with a binding commitment scheme we derive a scheme with a unique property:

Commitment with Sublinear Decommitment Witness. Typically in commitment schemes the size of the decommitment witness is of the same size as the committed value (or larger). For example in Pedersen's non-interactive scheme [Ped91], that is based on the discrete-logarithm assumption, the commitment to some $x < Q$ is a value $g^r h^x$ that belongs to $\mathbf{Z}_P^*$ (where $P = 2Q+1$ with P, Q large primes, and $g, h \in \mathbf{Z}_P^*$ public parameters which are quadratic residues modulo P) and the decommitment information is $\langle r, x \rangle$ (note that $r < Q$ is selected at random). Clearly the size of the decommitment witness is linear in the size of the committed information. In many settings it is of great interest to minimize the size of the decommitment information for private storage space saving.

In the case of PR-based commitment, we can use an alternative commitment scheme with which player A commits to the index-solution-set I of $V_{\mathbf{x}}$. The combined scheme becomes binding. Because of the fact that the size of the committed value $\mathbf{x}$ (which is proportional to $\log|\mathbb{F}|$) can be much larger (even superpolynomially) compared to the size of the index-solution-set (which is n) this turns a binding/hiding commitment to a bitstring of small length (n) to a binding/hiding commitment of a large value of length $\log|\mathbb{F}|$. Note that this does not compromise security since $\min\{\binom{n}{t}, \binom{n}{k}\}$ (which is the number of steps required for a brute-force attack against PR) can be chosen to be superpolynomial in $\log|\mathbb{F}|$ even if $\log|\mathbb{F}|$ is superpolynomial in n.

Let us instantiate the above using Pedersen's commitment scheme: suppose we want to commit to a value x of size b bits. Using Pedersen's commitment the decommitment witness would be of size $\mathcal{O}(b)$. Instead, we commit to x using the PR-based commitment over a finite field $\mathbb{F}$ with $\log|\mathbb{F}| > b$ by sending the value $F^I_{n,k,t}(x, r_1, \ldots, r_{k-1})$ (where I is the index-solution-set of the output of the PR-based one-way function); additionally we commit to v_I (which stands for a value that describes the set I) by sending $g^r h^{v_I}$. The decommitment information is $\langle r, v_I \rangle$ and is of size $\mathcal{O}(n)$. To achieve sublinear decommitment witness size we select the parameter n to be sublinear in the parameter b.

Proposition 5.5 *The combined commitment scheme described above is hiding, binding and non-interactive under the* DPR$[n, k-1, t]$ *over a finite field* $\mathbb{F}$*, and the Discrete-Logarithm Assumption over a multiplicative group of element size* n*, and can be used to commit to values of size* $\log|\mathbb{F}| > n$ *with decommitment witness information of size* $\mathcal{O}(n)$*.*

Proof. The proof is straightforward from the properties of the Pedersen's commitment scheme and corollary 5.3. ■

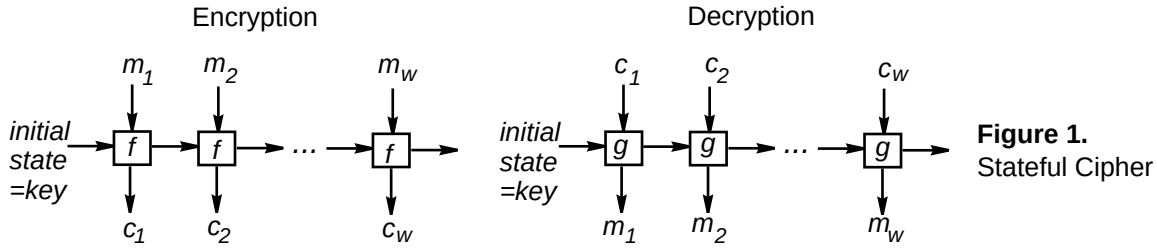
Corollary 5.6 *The combined commitment scheme supports sublinear decommitment witness size since* n *can be selected sublinear to* $\log|\mathbb{F}|$ *without affecting the security of the scheme (which depends solely on the security parameter* n*).*

5.3 A Secure Stateful-Cipher

A cipher design involves two parties, who share some common random input (the key). The goal of a cipher design is the secure transmission of a sequence of messages. Suppose that I denotes the shared randomness between the sender and the receiver. A cipher is defined by two probabilistic functions $f_I : \mathcal{K} \times \mathbb{P} \to \mathcal{K} \times \mathbb{C}$ and $g_I : \mathcal{K} \times \mathbb{C} \to \mathcal{K} \times \mathbb{P}$. The spaces $\mathcal{K}, \mathbb{P}, \mathbb{C}$ denote the state-space, plaintext-space and ciphertext-space respectively. The functions f, g have the property that if $f_I(s, m) = (s', c)$ (encryption) it holds that $g_I(s, c) = (s', m)$ (decryption); note that s' (given by both f, g) is the state that succeds the state s.

Stream-ciphers use public state sequences of the form $\langle 0, 1, 2, 3, \ldots \rangle$. The reader is referred to [Lub96] for more details on stream ciphers and how they can be built based on pseudorandom number generators. Block-ciphers encrypt messages of size equal to some fixed security parameter which are called blocks. Such ciphers are typically at the same state throughout and this state is considered to be secret (it coincides with the secret shared random key). The reader is referred to [Gol98] for further details on block-ciphers and generic constructions.

If a cipher, which operates on blocks, employs a "secret state-sequence update" and uses the shared randomness (the key) only as the initial state of the state-sequence, it is called a *stateful* cipher, see figure 1; (note that in a stateful cipher we suppress the subscript I from the functions f, g).



In the remaining of this section we introduce a stateful cipher that is based on PR and possesses unique properties.

5.3.1 Description of the PR-Cipher

Let $[n, \frac{k-1}{2}, t]$ with $k \leq t$ be sound parameters for the PR problem. We work in a finite field $\mathbb{F}$ with $\log |\mathbb{F}| \geq 3n$. The state-space $\mathcal{K}$ is defined to be the set of n-bitstrings with Hamming weight t. For some $s \in \mathcal{K}$ we define I_s to be the corresponding subset of $\{1, \ldots, n\}$, and v_s be the corresponding integer that has s as its binary representation. We denote by $V_{\mathcal{K}}$ the set of all numbers that their binary representation belongs in $\mathcal{K}$. Let $\mathbb{P} := \mathbb{F}^{\frac{k-1}{2}}$ and $\mathbb{C} := (\mathbb{F})_n \times \mathbb{F}^n$. The shared randomness between the two parties is a random $s_0 \in \mathcal{K}$, that is the initial state of the cipher. The encryption function of the cipher is defined as follows

$$f(s, \mathbf{m}) := F^{I_s}_{n,k,t}(\langle s', (\mathbf{m})_1, \ldots, (\mathbf{m})_{\frac{k-1}{2}}, r_1, \ldots, r_{\frac{k-1}{2}} \rangle)$$

where $F^{I_s}_{n,k,t}$ is the PR-based one-way function of section 5.1 so that index-solution-set of the output of $F_{n,k,t}$ is set to I_s; $r_1, \ldots, r_{\frac{k-1}{2}}$ are random elements of $\mathbb{F}$, and s' is a random element of $V_{\mathcal{K}}$. The decryption function g is defined as follows: given $\langle s, C \rangle \in \mathcal{K} \times \mathbb{C}$, the polynomial p

that corresponds to the pairs of C whose index is in I_s is interpolated. The decrypted message is set to be $\langle p(1), \ldots, p(\frac{k-1}{2}) \rangle$ and the next state is set to the binary representation of $p(0)$.

5.3.2 Semantic-Security

A semantic-security breaking adversary $\mathcal{A}$ for a stateful cipher is a PPT that takes the following steps: (i) queries a polynomial number of times the encryption-mechanism, (ii) generates two messages M_1, M_2 and obtains the ciphertext that corresponds to the encryption of M_b where b is selected at random from $\{1, 2\}$, (iii) queries the encryption-mechanism a polynomial number of times. Finally the adversary predicts the value of b with probability substantially better than $1/2$. This is illustrated in figure 2. A cipher is said to be semantically secure if any semantic-security breaking adversary predicts b with negligible advantage in the security parameter n. For more details regarding semantically secure symmetric encryption, see [Lub96, KY00].

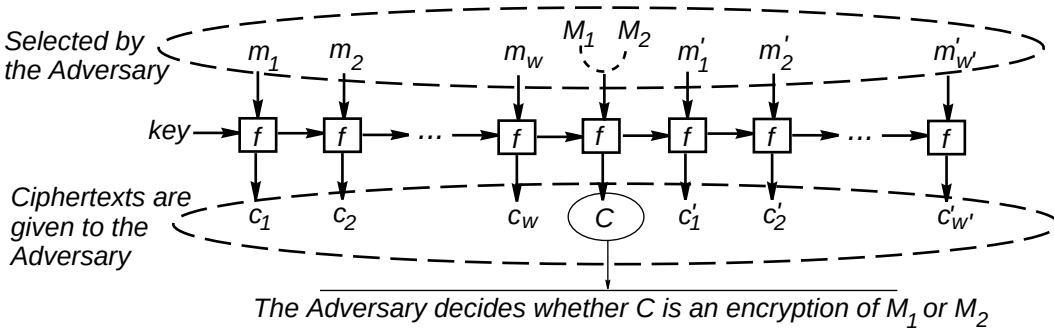


Figure 2. Semantic Security Adversary

More formally semantic security in the context of stateful ciphers is defined as follows:

Definition 5.7 *Let $\mathcal{O}^b$, with $b \in \{1, 2\}$ be an encryption oracle for a stateful cipher initialized to a random initial state that accepts two kinds of input: (i) a plaintext, where $\mathcal{O}^b$ returns its encryption under the current state, (ii) a pair of plaintexts M_1, M_2, where $\mathcal{O}^b$ returns the encryption of M_b (such input is allowed only once). A semantic security breaking adversary is a* PPT *$\mathcal{A}$ that given oracle access to $\mathcal{O}^b$ it predicts b with probability substantially better than $1/2$, i.e. the distance*

$$| \mathbf{Prob}_{b \in_U \{1,2\}}[\mathcal{A}^{\mathcal{O}^b}(1^n) = b] - \frac{1}{2} |$$

is non-negligible in n, where the probability is taken over all internal coin-tosses of $\mathcal{O}^b$ and $\mathcal{A}$ and all possible initial states for the cipher. If, for a certain cipher, there do not exist semantic security breaking adversaries then we say that the cipher is semantically secure.

We remark that the two kinds of input to the encryption oracle define three stages of adversarial action, namely (i) querying the encryption oracle a number of times, (ii) submitting the "challenge" (the pair of plaintexts of which the adversary receives the encryption of one of the two at random), and (iii) querying the encryption oracle a number of times before guessing which of the two plaintexts of the challenge was encrypted. We proceed to show that the PR-Cipher is semantically secure under the Decisional PR-Assumption, specifically:

Theorem 5.8 *The PR-Cipher is semantically secure under* DPR$[n, \frac{k-1}{2}, t]$.

Proof. We start with a definition: we denote by $\mathcal{L}^{(u)}_{n,k,t}[\mathbf{m}_1,\ldots,\mathbf{m}_u]$ the output of an encryption oracle of the PR-cipher when accessed by an semantic security adversary. In other words it is the space of sequences of $\mathcal{S}_{n,k,t}$ instances $X_1,\ldots,X_u$ so that $\mathbf{m}_j := \langle s_{X_j}(1),\ldots,s_{X_j}(\frac{k-1}{2})\rangle$ and so that the binary representation of $s_{X_j}(0)$ corresponds to the characteristic string of $I(X_{j+1})$, for $j=1,\ldots,u-1$. For two families of sets A_n and B_n we write $A_n \approx B_n$ if they are polynomial-time indistinguishable (see definition 4.1).

Claim 1. For any $u \geq 1$, $L^{(u)}_{n,k,t}[\mathbf{m}_1,\ldots,\mathbf{m}_u] \approx (S_n) \times L^{(u-1)}_{n,k,t}[\mathbf{m}_2,\ldots,\mathbf{m}_u]$ unless the DPR with parameters $[n,\frac{k-1}{2},t]$ fails.

Proof. Suppose the two families are distinguishable by some adversary $\mathcal{A}$ with non-negligible advantage. We will show how to use the adversary to violate the DPR with parameters $[n,\frac{k-1}{2},t]$.

Adaptive Encryption Sampler. The input is $X \in \mathcal{S}_{n,\frac{k-1}{2},t}(I)$ so that $X := \{\langle z_i, y_i\rangle\}_{i=1}^{n}$ and $z_i \notin \{0,\ldots,k-1\}$, and a *sequence* of messages $\mathbf{m}_1,\ldots,\mathbf{m}_u$ (submitted one by one). Let $p'(x)$ be a random polynomial of degree less than k so that (i) $p'(0)$ is a random element so that $p'(0) \leq 2^n$ and the hamming weight of $p'(0)$ is t, and (ii) $p'(i) = (\mathbf{m}_1)_i$ for $i=1,\ldots,\frac{k-1}{2}$. Consider the instance $X_{m_1} := \{\langle z_i, z_i(z_i-1)\ldots(z_i-\frac{k-1}{2})y_i + p'(z_i)\}$. Define I_2 to be the subset of $\{1,\ldots,n\}$ so that its characteristic string is identical to the binary representation of $p'(0)$. Next we sample X_{m_2} so that (i) $\langle s_{X_{m_2}}(1),\ldots,s_{X_{m_2}}(\frac{k-1}{2})\rangle = \mathbf{m}_2$, and (ii) the characteristic string of $I(X_{m_2})$ is identical to the binary representation of $s_{X_{m_1}}(0)$. Continuing in a similar manner we construct adaptively the instances $X_{m_3},\ldots,X_{m_u}$. It is clear that this series of samples is uniformly distributed over $L^{(u)}_{n,k,t}[\mathbf{m}_1,\ldots,\mathbf{m}_u]$.

Now suppose that the above sampling method is also given a parameter $i \in \{0,\ldots,n-\frac{k-1}{2}\}$ and the sampler randomizes the first i positions of X_{m_1}.

Now consider the predicates: $\mathcal{A}_1$ that simulates $\mathcal{A}$ using the adaptive encryption sampler to simulate the encryption oracle with parameter i, and $\mathcal{A}_2$ that simulates $\mathcal{A}$ using the adaptive encryption sampler to simulate the encryption oracle with parameter $i-1$. Following similar arguments as in the proof of lemma 3.2 one can see that $\mathcal{A}_1, \mathcal{A}_2$ constitute a gap-predicate-pair with parameters $[n,\frac{k-1}{2},t]$. ■

Claim 2. $L^{(u)}_{n,k,t}[\mathbf{m}_1,\ldots,\mathbf{m}_u] \approx (S_n)^u$ unless the DPR with parameters $[n,\frac{k-1}{2},t]$ fails.

Proof. Suppose that there is a distinguisher $\mathcal{A}$ between the two families (the "extreme hybrids"). Then by the triangular inequality $\mathcal{A}$ can distinguish between two "neighboring hybrids" i.e. $(\mathcal{S}_n)^v \times L^{(u-v)}_{n,k,t}[\mathbf{m}_{u-v},\ldots,\mathbf{m}_u] \not\approx (\mathcal{S}_n)^{v+1} \times L^{(u-v-1)}_{n,k,t}[\mathbf{m}_{u-v-1},\ldots,\mathbf{m}_u]$ for some $v \in \{0,\ldots,u-1\}$. Based on claim 1 this contradicts the DPR with parameters $[n,\frac{k-1}{2},t]$. ■

(*Proof* of theorem 5.8) Suppose now that $\mathcal{A}$ is a semantic security breaking adversary for the PR-cipher. Consider a predicate $\mathcal{B}$ that operates as follows:

$\mathcal{B}$ receives as input $i \in \{0,\ldots,n-\frac{k-1}{2}\}$ and $X \in \mathcal{S}_{n,\frac{k-1}{2},t}$ and communicates with the adversary $\mathcal{A}$ (refer to figure 2 that presents the operation of the adversary). In the first w queries to the adversary, $\mathcal{B}$ replies by random samples of $\mathcal{S}_n$. The adversary cannot detect the difference as the results of claim 2 reveal. When the adversary submits M_1, M_2, $\mathcal{B}$ selects $b \in \{1,2\}$ at random and using X, samples an encryption of M_b denoted by X_{M_b} using the technique described in the adaptive encryption sampler of the proof of claim 1. Subsequently $\mathcal{B}$ randomizes the first i positions of X_{M_b}. The remaining w' queries of $\mathcal{A}$ are answered by proper encryptions of the messages it submits (something that is possible for $\mathcal{B}$ since it resets the the key of the cipher in the construction of X_{M_b}). Finally $\mathcal{B}$ returns 1 if the adversary

guesses b correctly or 0 otherwise.

Define the predicate $\mathcal{A}_1 := \mathcal{B}$, and let $\mathcal{A}_2$ be the predicate that simulates $\mathcal{B}$ on input $i-1$ and X. Following similar arguments as in the proof of lemma 3.2 one can see that $\mathcal{A}_1, \mathcal{A}_2$ constitute a gap-predicate-pair for the parameters $[n, \frac{k-1}{2}, t]$ and as a result the DPR is violated. ■

5.3.3 Forward Secrecy

A cipher is said to satisfy forward secrecy if in the case of a total security breach at some point of its operation (i.e. the internal state is revealed) the adversary is unable to extract any information about the previously communicated messages.

This is formalized by two chosen plaintext security adversaries who are submitting adaptively messages to the encryption oracle. The encryption oracle flips a coin and answers by encrypting the plaintexts submitted by one of the two adversaries (the same adversary throughout). At some point the internal state of the system is revealed to the adversaries. Forward secrecy is violated if the adversaries can tell with probability significantly better than one half whose messages the encryption oracle was returning. More formally,

Definition 5.9 *Let $\mathcal{O}^b_{\text{fs}}$, with $b \in \{1,2\}$ be an encryption oracle for a stateful cipher initialized to a random initial state that accepts two kinds of input: (i) a pair of plaintexts $\mathbf{m}_1, \mathbf{m}_2$, where $\mathcal{O}^b_{\text{fs}}$ returns the encryption of $\mathbf{m}_b$ under the current state, (ii) a termination message, where $\mathcal{O}^b_{\text{fs}}$ returns the current internal state; no more queries are accepted by $\mathcal{O}^b_{\text{fs}}$ after the termination message is submitted. A forward secrecy breaking adversary is a* PPT *$\mathcal{A}$ that given oracle access to $\mathcal{O}^b_{\text{fs}}$ it predicts b with probability substantially better than $1/2$, i.e. the distance*

$$| \mathbf{Prob}_{b \in_U \{1,2\}}[\mathcal{A}^{\mathcal{O}^b_{\text{fs}}}(1^n) = b] - \frac{1}{2} |$$

is non-negligible in n, where the probability is taken over all internal coin-tosses of $\mathcal{O}^b_{\text{fs}}$ and $\mathcal{A}$ and all possible initial states for the cipher. If, for a certain cipher, there do not exist forward secrecy breaking adversaries then we say that the cipher satisfies forward secrecy.

The following theorem summarizes the properties of the PR-Cipher:

Theorem 5.10 *The PR-Cipher satisfies forward secrecy under* DPR$[n, \frac{k-1}{2}, t]$.

Proof. We denote by $\mathcal{L}^{(u)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_1\\ \mathbf{m}^1_1\end{smallmatrix}, \ldots, \begin{smallmatrix}\mathbf{m}^0_u\\ \mathbf{m}^1_u\end{smallmatrix}]$ the output of an encryption oracle of the stateful cipher when accessed by the two chosen plaintext adversaries that are part of the forward security attack. In other words it is the space of sequences of $\mathcal{S}_{n,k,t}$ instances $X_1, \ldots, X_u$ so that $\langle s_{X_j}(1), \ldots, s_{X_j}(\frac{k-1}{2})\rangle = \mathbf{m}^b_j$ for all $j = 1, \ldots, u$ where b is a random coin toss; the binary representation of $s_{X_j}(0)$ corresponds to the characteristic string of $I(X_{j+1})$, for $j = 1, \ldots, u-1$.

Claim 3. For any $u \geq 1$, $\mathcal{L}^{(u)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_1\\ \mathbf{m}^1_1\end{smallmatrix}, \ldots, \begin{smallmatrix}\mathbf{m}^0_u\\ \mathbf{m}^1_u\end{smallmatrix}] \approx (S_n) \times \mathcal{L}^{(u-1)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_2\\ \mathbf{m}^1_2\end{smallmatrix}, \ldots, \begin{smallmatrix}\mathbf{m}^0_u\\ \mathbf{m}^1_u\end{smallmatrix}]$ unless the DPR with parameters $[n, \frac{k-1}{2}, t]$ fails.

The arguments of the proof of claim 3 are similar to those of the proof of claim 1 of theorem 5.8.

Claim 4. $\mathcal{L}^{(u)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_1\\ \mathbf{m}^1_1\end{smallmatrix}, \ldots, \begin{smallmatrix}\mathbf{m}^0_u\\ \mathbf{m}^1_u\end{smallmatrix}] \approx (S_n)^u$ unless the DPR with parameters $[n, \frac{k-1}{2}, t]$ fails.

Again, this is shown using the same argument as in the proof of claim 2 of theorem 5.8.

Now the result follows easily since: the output of the encryption oracle is indistinguishable for the choice of $b \in \{1,2\}$ provided that the DPR with parameters $[n, \frac{k-1}{2}, t]$ holds. This implies in a straightforward manner that the adversary cannot predict b with probability significantly better than $1/2$. ■

5.3.4 Computational Perfect Secrecy

A generic chosen plaintext adversary for a stateful cipher is defined as follows:

Definition 5.11 *Let $\mathcal{O}$ be an encryption oracle for a stateful cipher that is initialized to a random initial state; given a plaintext, $\mathcal{O}$ returns its encryption under the current state. A generic chosen plaintext adversary is a* PPT *$\mathcal{A}$ that is given oracle access to $\mathcal{O}$.*

For some stateful-cipher we consider the following two attacks that can be launched by a generic chosen plaintext adversary: (i) "existential" where the generic chosen plaintext adversary is allowed to query the encryption oracle a number of times and then is asked to decrypt the next ciphertext (which encrypts a random secret message) (ii) "universal" where a generic chosen plaintext adversary is allowed to query the encryption oracle a number of times and then is asked to recover the state of the cipher (something that allows the recovery of all future messages from that point on).

It is clear that for any cipher an existential attack reduces to a universal attack. Nevertheless it is not at all apparent if the opposite direction in the reduction holds.

Definition 5.12 *A stateful-cipher for which it holds that a generic chosen plaintext adversary launching an existential attack implies the existence of a generic chosen plaintext adversary launching a universal attack is said to satisfy* computational perfect secrecy.

The equivalence of attacks that recover the message to attacks that recover the key has been postulated by Shannon as "perfect secrecy." Blum and Goldwasser [BG84] designed a factoring based public-key system where they reduced semantic security of a message to breaking the key (i.e. factoring the composite). They coined the notion of "computational perfect secrecy," a variant of which we define above.

Theorem 5.13 *The PR-Cipher satisfies computational perfect secrecy.*

Proof. Suppose that it is possible to launch an existential attack with u queries to the encryption mechanism. We show how to launch a universal attack: first we make $(u+1)$-queries to the encryption mechanism so we have the plaintext-ciphertext pairs $\langle M_1, C_1\rangle, \ldots, \langle M_{u+1}, C_{u+1}\rangle$ where $M_1, \ldots, M_u$ are chosen following the query algorithm of the existential attack algorithm and M_{u+1} is chosen at random. Suppose that $C_{u+1} := \{\langle z_i, y_i\rangle\}_{i=1}^{n}$. We compute $X' := \{\langle z_i + 1, y_i\rangle\}_{i=1}^{n}$, and we feed X' to the existential attack algorithm to obtain the "message" $\langle a_1, \ldots, a_{\frac{k-1}{2}}\rangle$ with probability of success α. Observe that $s_{X'}(x) = s_X(x-1)$ and as a result $a_1 = s_{X'}(1) = s_X(0)$. It follows that the binary representation of a_1 is the characteristic string of the next key (for the $(u+2)$-th encryption of the cipher). As a result the universal attack reduces to an existential attack with the same probability of success. ■

5.3.5 Superpolynomial Message-Size

A cryptosystem that has this property allows the plaintext size to be superpolynomial in the key-size, or in other words, it allows the key-size to be substantially shorter (inverse-super-polynomial) in the size of messages.

This property allows much saving in the storage of the shared key which can be an expensive resource in many settings. Additionally, it can be particularly useful in settings where we want to extract a key from a small amount of information (such as key-extraction from biometric data, see e.g. [MRW99]).

In the PR-Cipher the plaintext size is $\frac{k-1}{2}\lfloor \log |\mathbb{F}| \rfloor$ and can be superpolynomial in the security parameter since $\log |\mathbb{F}|$ can be chosen to be superpolynomial in the security parameter n without affecting the security of the cryptosystem. This is because a brute-force attack against PR requires $\min\{\binom{n}{t}, \binom{n}{k}\}$ steps worst-case and this quantity can be selected to be superpolynomial in $\log |\mathbb{F}|$ even if $\log |\mathbb{F}|$ is superpolynomial in n.

5.3.6 Error-Correcting Decryption

A cryptosystem is said to allow error-correcting decryption if the decryption procedure is able to correct errors that are introduced during the transmission (possibly by an adversary). This combines the decryption operation with the error-correction operation (that is important to apply independently in any setting where two parties communicate).

A cryptosystem that transmits plaintext blocks of size d is called d'-error-correcting if up to d' corrupted blocks can be corrected for each transmitted ciphertext. The PR-cipher (which transmits plaintext blocks of size $\frac{k-1}{2}$ over the underling finite field $\mathbb{F}$ in each ciphertext) is $\frac{t-k}{2}$-error-correcting since the interpolation step during decryption can be substituted by the [BW86] polynomial-reconstruction algorithm that can withstand up to $\frac{t-k}{2}$ errors (in the worst-case).

5.3.7 Key-Equivalence

A symmetric cryptosystem is said to satisfy the key-equivalence property if there are no families of keys of measurable size that are susceptible to attacks that do not apply to the key-space in general. By "measurable-size" we mean that the ratio of the size of the family of keys over the key-space size is a non-negligible function. More formally,

Definition 5.14 *Let $\mathcal{K}_n$ denote the key-space of a cipher, where n denotes the security parameter. Let $\mathcal{A}$ be a* PPT *(thought of as a generic adversary) that takes as input a sequence of ciphertexts s_κ as transmitted over the public channel by the sender to the receiver who share a secret-key κ. The cipher satisfies the key-equivalence property if there exists a* PPT *$\mathcal{A}'$ s.t. for any family of keys $\mathcal{K}'_n \subseteq \mathcal{K}_n$ of measurable size: $(\#\mathcal{K}'_n/\#\mathcal{K}_n)$ is non-negligible in n, it holds that for all v in the range of $\mathcal{A}$, the distance*

$$| \mathbf{Prob}_{\kappa \in_U \mathcal{K}'_n}[\mathcal{A}(s_\kappa) = v] - \mathbf{Prob}_{\kappa \in_U \mathcal{K}_n}[\mathcal{A}'(s_\kappa) = v] |$$

is negligible in n, where the probability is taken over the coin-tosses of $\mathcal{A}, \mathcal{A}'$ and the coin tosses of the sender who generates the sequence of ciphertexts. Intuitively this suggests that an attack of any kind against the cipher over a certain family of keys, can be generalized to an attack against the cipher over the whole key-space. Note that v is possibly a function of s_κ.

The key-equivalence property is an important security aspect for a symmetric cryptosystem as it suggests that there are no "weak" keys.

Proposition 5.15 *The PR-Based Stateful Cipher satisfies the key-equivalence property.*

Proof. This can be seen easily as a corollary of lemma 2.9 and the fact that the key-space for the PR-based stateful cipher is defined to be the set of all subsets of $\{1, \ldots, n\}$ of size t. ■

References

[Ber68] Elwyn R. Berlekamp, *Algebraic Coding Theory.* McGraw-Hill, 1968.

[BW86] Elwyn R. Berlekamp and L. Welch, *Error Correction of Algebraic Block Codes.* U.S. Patent, Number 4,633,470, 1986.

[BN00] Daniel Bleichenbacher and Phong Nguyen, *Noisy Polynomial Interpolation and Noisy Chinese Remaindering,* In Advances in Cryptology — Eurocrypt 2000, Lecture Notes in Computer Science, Springer-Verlag, vol. 1807, pp. 53–69, May 2000.

[BG84] Manuel Blum and Shafi Goldwasser, *An Efficient Probabilistic Public-Key Encryption Scheme Which Hides All Partial Information*, In Advances in Cryptology — Crypto 1984, Lecture Notes in Computer Science, Springer-Verlag, vol. 196, pp. 289–302, 1985.

[Gol90] Oded Goldreich, *A note on computational indistinguishability*, Information Processing Letters, vol. 34, no. 6, pp. 277–281, 1990.

[Gol98] Oded Goldreich, *Foundations of Cryptography: Fragments of a Book*, manuscript 1998, available at `http://www.wisdom.weizmann.ac.il/ oded/frag.html`

[GSR95] Oded Goldreich, Madhu Sudan and Ronitt Rubinfeld, *Learning Polynomials with Queries: The Highly Noisy Case,* in the Proceedings of the 36th Annual Symposium on Foundations of Computer Science, IEEE Computer Society, pp. 294–303, 1995.

[GM84] Shafi Goldwasser and Silvio Micali, *Probabilistic encryption*, Journal of Computer and System Sciences, vol. 28(2), pp. 270-299, April 1984.

[GS98] Venkatesan Guruswami and Madhu Sudan, *Improved Decoding of Reed-Solomon and Algebraic-Geometric Codes.* In the Proceedings of the 39th Annual Symposium on Foundations of Computer Science, IEEE Computer Society, pp. 28–39, 1998.

[HILL99] Johan Hastad, Russel Impagliazzo, Leonid Levin and Michael Luby, *Construction of a pseudo-random generator from any one-way function*, SIAM J. Comput. 28(4):1364-1396, 1999.

[KY00] Jonathan Katz and Moti Yung, *Complete Characterization of Security Notions for Probabilistic Private-key Encryption*, in the Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, ACM, pp. 245–254, 2000.

[Lub96] Michael Luby, *Pseudorandomness and Cryptographic Applications*, Princeton University Press, 1996.

[MRW99] Fabian Monrose, Michael K. Reiter, and Suzanne Wetzel, *Password Hardening based on Keystroke Dynamics.* in the Proceedings of the 6th ACM Conference on Computer and Communications Security, ACM, pp. 73–82, 1999.

[MS77] F. J. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes.* North Holland, Amsterdam, 1977.

[Na91] Moni Naor, *Bit Commitment Using Pseudorandomness*, Journal of Cryptology 4(2): pp. 151-158, 1991.

[NP99] Moni Naor and Benny Pinkas, *Oblivious Transfer and Polynomial Evaluation.* in the Proceedings of the 31st Annual ACM Symposium on Theory of Computing, ACM, pp. 245–254, 1999. (Full Version *Oblivious Polynomial Evaluation*, available at `http://www.wisdom.weizmann.ac.il/ naor/onpub.html`.)

[NR97] Moni Naor and Omer Reingold, *Number-theoretic Constructions of Efficient Pseudo-random Functions*, In the Proceedings of the 38th Annual Symposium on Foundations of Computer Science, IEEE Computer Society, pp. 458–467, 1997.

[Ped91] Torben P. Pedersen, *Non-Interactive and information-theoretic secure verifiable secret-sharing*, In Advances in Cryptology — Crypto 1991, Lecture Notes in Computer Science, Springer-Verlag, vol. 576, pp. 129–140, 1992.

[Sud97] Madhu Sudan, *Decoding of Reed Solomon Codes beyond the Error-Correction Bound.* Journal of Complexity 13(1), pp. 180–193, 1997.

[Yao82] Andrew C. Yao, *Theory and applications of trapdoor functions*, in the Proceedings of the 23rd Annual Symposium on Foundations of Computer Science, IEEE Computer Society, pp. 80–91, 1982.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'