The Cryptographic Power of Random Selection

Matthias Krause and Matthias Hamann

Theoretical Computer Science
University of Mannheim
Mannheim, Germany

Abstract. The principle of random selection and the principle of adding biased noise are new paradigms used in several recent papers for constructing lightweight RFID authentication protocols. The cryptographic power of adding biased noise can be characterized by the hardness of the intensively studied Learning Parity with Noise (LPN) Problem. In analogy to this, we identify a corresponding learning problem for random selection and study its complexity. Given L secret linear functions $f_1, \ldots, f_L : \{0,1\}^n \longrightarrow \{0,1\}^a$, $RandomSelect\,(L, n, a)$ denotes the problem of learning $f_1, \ldots, f_L$ from values $(u, f_l\,(u))$, where the secret indices $l \in \{1, \ldots, L\}$ and the inputs $u \in \{0,1\}^n$ are randomly chosen by an oracle. We take an algebraic attack approach to design a nontrivial learning algorithm for this problem, where the running time is dominated by the time needed to solve full-rank systems of linear equations over $O\left(n^L\right)$ unknowns. In addition to the mathematical findings relating correctness and average running time of the suggested algorithm, we also provide an experimental assessment of our results.

Keywords: Lightweight Cryptography, Algebraic Attacks, Algorithmic Learning, Foundations and Complexity Theory

1 Introduction

The very limited computational resources available in technical devices like RFID (radio frequency identification) tags implied an intensive search for lightweight authentication protocols in recent years. Standard block encryption functions like Triple-DES or AES seem to be not suited for such protocols largely because the amount of hardware to implement and the energy consumption to perform these operations is too high (see, e.g., [7] or [17] for more information on this topic).

This situation initiated two lines of research. The first resulted in proposals for new lightweight block encryption functions like PRESENT [4], KATAN and KTANTAN [10] by use of which standard block cipher-based authentication protocols can be made lightweight, too. A second line, and this line we follow in the paper, is to look for new cryptographic paradigms which allow for designing new symmetric lightweight authentication protocols. The two main suggestions discussed so far in the relevant literature are the principle of random selection and the principle of adding biased noise.

The principle of adding biased noise to the output of a linear basis function underlies the HB-protocol, originally proposed by Hopper and Blum [16] and later improved to HB$^+$ by Juels and Weis [17], as well as its variants HB$^\#$ and Trusted-HB (see [13] and [6], respectively). The protocols of the HB-family are provably secure against passive attacks with respect to the Learning Parity with Noise Conjecture but the problem to design HB-like protocols which are secure against active adversaries seems to be still unsolved (see, e.g., [14], [20], [12]).

The principle of random selection underlies, e.g., the CKK-protocols of Cichoń, Klonowski, and Kutyłowski [7] as well as the F_f-protocols in [3] and the Linear Protocols in [18]. It can be described as follows.

Suppose that the verifier Alice and the prover Bob run a challenge-response authentication protocol which uses a lightweight symmetric encryption operation $E : \{0,1\}^n \times \mathcal{K} \longrightarrow \{0,1\}^m$ of block length n, where $\mathcal{K}$ denotes an appropriate key space. Suppose further that E is weak in the sense that a passive adversary can efficiently compute the secret key $K \in \mathcal{K}$ from samples of the form $(u, E_K(u))$. This is obviously the case if E is linear.

Random selection denotes a method for compensating the weakness of E by using the following mode of operation. Instead of holding a single $K \in \mathcal{K}$, Alice and Bob share a collection $K_1, \ldots, K_L$ of keys from $\mathcal{K}$ as their common secret information, where $L > 1$ is a small constant. Upon receiving a challenge $u \in \{0,1\}^n$ from Alice, Bob chooses a random index $l \in \{1, \ldots, L\}$ and outputs the response $y = E(u, K_l)$. The verification of y with respect to u can be efficiently done by computing $E_{K_l}^{-1}(y)$ for all $l = 1, \ldots, L$.

The main problem this paper is devoted to is to determine the level of security which can be reached by applying this principle of random selection.

Note that the protocols introduced in [7], [3], and [18] are based on random selection of $GF(2)$-linear functions. The choice of linear basis functions is motivated by the fact that they can be implemented efficiently in hardware and have desirable pseudo-random properties with respect to a wide range of important statistical tests.

It is quite obvious that, with respect to passive adversaries, the security of protocols which use random selection of linear functions can be bounded from above by the complexity of the following learning problem referred to as $\mathit{RandomSelect}\,(L, n, a)$: Learn $GF(2)$-linear functions $f_1, \ldots, f_L : \{0,1\}^n \longrightarrow \{0,1\}^a$ from values $(u, f_l\,(u))$, where the secret indices $l \in \{1, \ldots, L\}$ and the inputs $u \in \{0,1\}^n$ are randomly chosen by an oracle. In order to illustrate this notion, we sketch in appendix B how an efficient learning algorithm for $\mathit{RandomSelect}\,(L, n, a)$ can be used for attacking the linear $(n, k, L)^+$-protocol described by Krause and Stegemann [18].

In this paper, we present an algebraic attack approach for solving the above learning problem $\mathit{RandomSelect}\,(L, n, a)$. The running time of our algorithm is dominated by the effort necessary to solve a full-rank system of linear equations of $O(n^L)$ unknowns over the field $GF(2^a)$. Note that trivial approaches for solving $\mathit{RandomSelect}\,(L, n, a)$ lead to a running time exponential in n.

In recent years, people from cryptography as well as from complexity and coding theory devoted much interest to the solution of learning problems around linear structures. Prominent examples in the context of lightweight cryptography are the works by Goldreich and Levin [15], Regev [21], and Arora and Ge [2]. But all these results are rather connected to the Learning Parity with Noise Problem. To the best of our knowledge, there are currently no nontrivial results with respect to the particular problem of learning randomly selected linear functions, which is studied in the present paper.

We are strongly convinced that the complexity of *RandomSelect* also defines a lower bound on the security achievable by protocols using random selection of linear functions, e.g., the improved $(n, k, L)^{++}$-protocol in [18]. Thus, the running time of our algorithm hints at how the parameters n, k, and L should be chosen in order to achieve an acceptable level of cryptographic security. Note that choosing $n = 128$ and $L = 8$ or $n = 256$ and $L = 4$, solving $RandomSelect\,(L, n, a)$ by means of our algorithm implies solving a system of around 2^{28} unknowns, which should be classified as sufficiently difficult in many practical situations.

The paper is organized as follows. In sections 2, 3, and 4, our learning algorithm, which conducts an algebraic attack in the spirit of [22], will be described in full detail. We represent the L linear basis functions as assignments A to a collection $X = \left(x_i^l\right)_{i=1,\ldots,n,l=1,\ldots,L}$ of variables taking values from the field $K = GF(2^a)$. We will then see that each example $(u, f_l\,(u))$ induces a degree-L equation of a certain type in the X-variables, which allows for reducing the learning problem $RandomSelect\,(L, n, a)$ to the problem of solving a system of degree-L equations over K. While, in general, the latter problem is known to be NP-hard, we can show an efficient way to solve this special kind of systems.

One specific problem of our approach is that, due to inherent symmetries of the degree-L equations, we can never reach a system which has full linear rank with respect to the corresponding monomials. In fact, this is the main difference between our learning algorithm and the well-known algebraic attack approaches for cryptanalyzing LFSR-based keystream generators (see, e.g., [19], [8], [9], [1]).

We circumvent this problem by identifying an appropriate set $T(n, L)$ of basis polynomials of degree at most L which allow to express the degree-L equations as linear equations over $T(n, L)$. The choice of $T(n, L)$ will be justified by Theorem 2 saying that if $|K| \geq L$, then the system of linear equations over $T(n, L)$ induced by all possible examples has full rank $|T(n, L)|$. (Note that according to Theorem 1, this is not true if $|K| < L$.) Our experiments, which are presented in section 5, indicate that if $|K| \geq L$, then with probability close to one, the number of examples needed to get a full rank system over $T(n, L)$ exceeds $|T(n, L)|$ only by a small constant factor. This implies that the effort to compute the unique *weak* solution $t\,(A) = (t_*\,(A))_{t_* \in T(n,L)}$ corresponding to the *strong* solution A equals the time needed to solve a system of $|T(n, L)|$ linear equations over K.

But in contrast to the algebraic attacks in [19], [8], [9], [1], we still have to solve another nontrivial problem, namely, to compute the *strong* solution A, which identifies the secret functions $f_1, \ldots, f_L$, from the unique weak so-

lution. An efficient way to do this will complete our learning algorithm for $RandomSelect\,(L,n,a)$ in section 4. Finally, we also provide an experimental evaluation of our estimates using the computer algebra system Magma [5] in section 5 and conclude this paper with a discussion of the obtained results as well as an outlook on potentially fruitful future work in section 6.

2 The Approach

We fix positive integers n, a, L and secret $GF(2)$-linear functions $f_1, \ldots, f_L : \{0,1\}^n \longrightarrow \{0,1\}^a$. The learner seeks to deduce specifications of $f_1, \ldots, f_L$ from an oracle which outputs in each round an example $(u,w) \in \{0,1\}^n \times \{0,1\}^a$ in the following way. The oracle chooses independently and uniformly a random input $u \in_U \{0,1\}^n$, then chooses secretly a random index $l \in_U [L]$[①], computes $w = f_l\,(u)$ and outputs (u,w).

It is easy to see that *RandomSelect* can be efficiently solved in the case $L = 1$ by collecting examples $\left(u^1, w_1\right), \ldots, \left(u^m, w_m\right)$ until $\left\{u^1, \ldots, u^m\right\}$ contains a basis of $GF(2)^n$. The expected number of iterations until the above goal is reached can be approximated by $n + 1.61$ (see, e.g., the appendix in [11]).

We will now treat the case $L > 1$, which immediately yields a sharp rise in difficulty. First we need to introduce the notion of a *pure basis*.

Definition 1. *Let us call a set* $\mathcal{V} = \left\{\left(u^1, w_1\right), \ldots, \left(u^n, w_n\right)\right\}$ *of* n *examples a pure basis, if* $\left\{u^1, \ldots, u^n\right\}$ *is a basis of* $GF(2)^n$ *and there exists an index* $l \in [L]$ *such that* $w_i = f_l\left(u^i\right)$ *is satisfied for all* $i = 1, \ldots, n$.

Recalling our preliminary findings, we can easily infer that for $m \in Ln + \Omega\,(1)$, a set of m random examples contains such a pure basis with high probability. Moreover, note that for a given set $\tilde{\mathcal{V}} = \left\{\left(\tilde{u}^1, \tilde{w}_1\right), \ldots, \left(\tilde{u}^n, \tilde{w}_n\right)\right\}$ the pure basis property can be tested efficiently. The respective strategy makes use of the fact that in case of a random example (u,w), where $u = \bigoplus_{i \in I} \tilde{u}^i$ and $I \subseteq [n]$[②], the probability p that $w = \bigoplus_{i \in I} \tilde{w}_i$ holds is approximately L^{-1} if $\tilde{\mathcal{V}}$ is pure and at most $(2 \cdot L)^{-1}$ otherwise. The latter estimate is based on the trivial observation that if $\tilde{\mathcal{V}}$ is not a pure basis, it contains at least one tuple $\left(\tilde{u}^j, \tilde{w}_j\right)$, $j \in [n]$, which would have to be exchanged to make the set pure. As $j \in I$ holds true for half of all possible (but valid) examples, the probability that $w = \bigoplus_{i \in I} \tilde{w}_i$ is fulfilled although $\tilde{\mathcal{V}}$ is not pure can be bounded from above by $(2 \cdot L)^{-1}$.

However, it seems to be nontrivial to extract a pure basis from a set of $m \in Ln + \Omega\,(1)$ examples. Exhaustive search among all subsets of size n yields

[①] For a positive integer N, we denote by $[N]$ the set $\{1, \ldots, N\}$.

[②] Let $B = \left\{v^1, \ldots, v^n\right\}$ denote a basis spanning the vector space V. It is a simple algebraic fact that every vector $v \in V$ has a unique representation $I \subseteq [n]$ over B, i.e., $v = \bigoplus_{i \in I} v^i$.

a running time exponential in n. This can be shown easily by applying Stirling's formula[3] to the corresponding binomial coefficient $\binom{m}{n}$.

We exhibit the following alternative idea for solving $RandomSelect\,(L, n, a)$ for $L > 1$. Let $e^1, \ldots, e^n$ denote the standard basis of the $GF(2)$-vector space $\{0,1\}^n$ and keep in mind that $\{0,1\}^n = GF(2)^n \subseteq K^n$, where K denotes the field $GF(2^a)$. For all $i = 1, \ldots, n$ and $l = 1, \ldots, L$ let us denote by x_i^l a variable over K representing $f_l\left(e^i\right)$. Analogously, let A denote the $(n \times L)$-matrix with coefficients in K completely defined by $A_{i,l} = f_l\left(e^i\right)$. Henceforth, we will refer to A as a *strong solution* of our learning problem, thereby indicating the fact that its coefficients fully characterize the underlying secret $GF(2)$-linear functions $f_1, \ldots, f_L$.

Observing an example (u, w), where $u = \bigoplus_{i \in I} e^i$, the only thing we know is that there is some index $l \in [L]$ such that $w = \bigoplus_{i \in I} A_{i,l}$. This is equivalent to the statement that A is a solution of the following degree-L equation in the x_i^l-variables.

$$\left(\bigoplus_{i \in I} x_i^1 \oplus w\right) \cdot \ldots \cdot \left(\bigoplus_{i \in I} x_i^L \oplus w\right) = 0. \tag{1}$$

Note that equation (1) can be rewritten as

$$\bigoplus_{J \subseteq I, 1 \leq |J| \leq L'} \bigoplus_{j=|J|}^{L} w^{L-j} t_{J,j} = w^L, \tag{2}$$

$L' = \min\{L, |I|\}$, where the basis polynomials $t_{J,j}$ are defined as

$$t_{J,j} = \bigoplus_{g, |dom(g)|=j, im(g)=J} m_g$$

for all $J \subseteq [n]$, $1 \leq |J| \leq L$, and all j, $|J| \leq j \leq L$. The corresponding monomials m_g are in turn defined as

$$m_g = \prod_{l \in dom(g)} x_{g(l)}^l$$

for all partial mappings g from $[L]$ to $[n]$, where $dom\,(g)$ denotes the domain of g and $im\,(g)$ denotes its image.

Let $T\,(n, L) = \{t_{J,j} \mid J \subseteq [n], 1 \leq |J| \leq L, |J| \leq j \leq L\}$ denote the set of all basis polynomials $t_{J,j}$ which may appear as part of equation (2). Moreover, we define

$$\Phi\,(a, b) = \sum_{i=0}^{b} \binom{a}{i}$$

[3] Stirling's formula is an approximation for large factorials and commonly written $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

for integers $0 \leq b \leq a$ and write

$$\begin{aligned} |T(n,L)| &= \sum_{j=1}^{L} \binom{n}{j} (L-j+1) \\ &= (L+1)(\Phi(n,L)-1) - \sum_{j=1}^{L} n\binom{n-1}{j-1} \\ &= (L+1)(\Phi(n,L)-1) - n\Phi(n-1,L-1). \end{aligned} \tag{3}$$

Consequently, each set of examples $\mathcal{V} = \{(u^1, w_1), \ldots, (u^m, w_m)\}$ yields a system of m degree-L equations in the x_i^l-variables, which can be written as m K-linear equations in the $t_{J,j}$-variables. In particular, the strong solution $A \in K^{n \times L}$ satisfies the relation

$$M(\mathcal{V}) \circ t(A) = W(\mathcal{V}), \tag{4}$$

where

- $K^{n \times L}$ denotes the set of all $(n \times L)$-matrices with coefficients from K,
- $M(\mathcal{V})$ is an $(m \times |T(n,L)|)$-matrix built by the m linear equations of type (2) corresponding to the examples in $\mathcal{V}$,
- $W(\mathcal{V}) \in K^m$ is defined by $W(\mathcal{V})_i = w_i^L$[④] for all $i = 1, \ldots, m$,
- $t(A) \in K^{T(n,L)}$ is defined by $t(A) = (t_{J,j}(A))_{J \subseteq [n], 1 \leq |J| \leq L, |J| \leq j \leq L}$.

Note that in section 3, we will treat the special structure of $M(\mathcal{V})$ in further detail. Independently, it is a basic fact from linear algebra that if $M(\mathcal{V})$ has full column rank, then the linear system (4) has the unique solution $t(A)$, which we will call the *weak solution*.

Our learning algorithm proceeds as follows:

(1) Grow a set of examples $\mathcal{V}$ until $M(\mathcal{V})$ has full column rank $|T(n,L)|$.
(2) Compute the unique solution $t(A)$ of system (4), i.e., the weak solution of our learning problem, by using an appropriate algorithm which solves systems of linear equations over K.
(3) Compute the strong solution A from $t(A)$.

We discuss the correctness and running time of steps (1) and (2) in section 3 and an approach for step (3) in section 4.

[④] Keep in mind that, unlike for the previously introduced K-variables $x_s^1, \ldots, x_s^L$, $s \in [n]$, the superscripted L in case of w_i^L is not an index but an exponent. See, e.g., equation (2).

3 On Computing a Weak Solution

Let n and L be arbitrarily fixed such that $2 \leq L \leq n$ holds. Moreover, let $\mathcal{V} \subseteq \{0,1\}^n \times K$ denote a given set of examples obtained through linear functions $f_1, \ldots, f_L : \{0,1\}^n \longrightarrow K$, where $K = GF(2^a)$. By definition, for each tuple $(u,w) \in \mathcal{V}$, where $u = \bigoplus_{i \in I} e^i$ and $I \subseteq [n]$ denotes the unique representation of u over the standard basis $e^1, \ldots, e^n$ of $\{0,1\}^n$, the relation $w = f_{l'}(u) = \bigoplus_{i \in I} f_{l'}\left(e^i\right)$ is bound to hold for some $l' \in [L]$. We denote by $K^{\min} \subseteq K$ the subfield of K generated by all values $f_l\left(e^i\right)$, where $l \in [L]$ and $i \in [n]$. Note that $w \in K^{\min}$ for all examples (u,w) induced by $f_1, \ldots, f_l$.

In the following, we show that our learning algorithm is precluded from succeeding if the secret linear functions $f_1, \ldots, f_L$ happen to be of a certain type or if K itself lacks in size.

Theorem 1 *If $\left|K^{\min}\right| < L$, then the columns of $M(\mathcal{V})$ are linearly dependent for any set $\mathcal{V}$ of examples, i.e., a unique weak solution does not exist.*

Proof: Let n, K, L, and $f_1, \ldots, f_L$ be arbitrarily fixed such that $2 \leq \left|K^{\min}\right| < L \leq n$ holds and let $\mathcal{V}$ denote a corresponding set of examples. Obviously, for each tuple $(u,w) \in \mathcal{V}$, where $u = \bigoplus_{i \in I} e^i$ and $I \subseteq [n]$, the two cases $1 \in I$ and $1 \notin I$ can be differentiated.

If $1 \in I$ holds, then it follows straightforwardly from equation (2) that the coefficient with coordinates (u,w) and $t_{\{1\},(L-1)}$ in $M(\mathcal{V})$ equals $w^{L-(L-1)} = w^1$. Analogously, the coefficient with coordinates (u,w) and $t_{\{1\},(L-|K^{\min}|)}$ in $M(\mathcal{V})$ equals $w^{L-\left(L-|K^{\min}|\right)} = w^{|K^{\min}|}$. Note that $t_{\{1\},(L-|K^{\min}|)}$ is a valid (and different) basis polynomial as

$$|\{1\}| = 1 \leq \left(L - \left|K^{\min}\right|\right) \leq (L-2) < (L-1) < L$$

holds for $2 \leq |K^{\min}| < L$. As $K^{\min} \subseteq K$ is a finite field of characteristic 2, we can apply Lagrange's theorem and straightforwardly conclude that the relation $z^1 = z^{\left|K^{\min}\right|}$ holds for all $z \in K^{\min}$ (including $0 \in K^{\min}$). Hence, if $1 \in I$ holds for an example (u,w), then in the corresponding row of $M(\mathcal{V})$ the two coefficients indexed by $t_{\{1\},(L-1)}$ and $t_{\{1\},(L-|K^{\min}|)}$ are always equal.

If $1 \notin I$ holds for an example (u,w), then the coefficient with coordinates (u,w) and $t_{\{1\},(L-1)}$ in $M(\mathcal{V})$ as well as the coefficient with coordinates (u,w) and $t_{\{1\},(L-|K^{\min}|)}$ in $M(\mathcal{V})$ equals 0.

Consequently, if $\left|K^{\min}\right| < L$ holds, then the column of $M(\mathcal{V})$ indexed by $t_{\{1\},(L-1)}$ equals the column indexed by $t_{\{1\},(L-|K|)}$ for any set $\mathcal{V}$ of examples, i.e., $M(\mathcal{V})$ can never achieve full column rank. □

Corollary 1 *If K is chosen such that $|K| < L$, then the columns of $M(\mathcal{V})$ are linearly dependent for any set $\mathcal{V}$ of examples, i.e., a unique weak solution does not exist.* □

While we are now aware of a lower bound for the size of K, it yet remains to prove that step (1) of our learning algorithm is, in fact, correct.

This will be achieved by introducing the $((2^n |K|) \times |T(n,L)|)$-matrix $M^* = M(\{0,1\}^n \times K)$, which clearly corresponds to the set of *all* possible examples, and showing that M^* has full column rank $|T(n,L)|$ if $L \leq |K|$ holds.

However, be careful not to misinterpret this finding, which is presented below in the form of Theorem 2. The fact that M^* has full column rank $|T(n,L)|$ by no means implies that, eventually, this will also hold for $M(\mathcal{V})$ if only the corresponding set of observations $\mathcal{V}$ is large enough. In particular, the experimental results summarized in section 5 (see, e.g., table 1) show that there are cases in which the rank of $M(\mathcal{V})$ is always smaller than $|T(n,L)|$, even if $L \leq |K|$ is satisfied and $\mathcal{V}$ equals the set $\{(u, f_l(u)) \mid u \in \{0,1\}^n, l \in [L]\} \subseteq \{0,1\}^n \times K$[5] of all possible *valid* examples.

Still, as a counterpart of Theorem 1, the following theorem proves the possibility of existence of a unique weak solution for arbitrary parameters n and L satisfying $2 \leq L \leq n$. In other words, choosing $T(n,L)$ to be the set of basis polynomials does not necessarily lead to systems of linear equations which cannot be solved uniquely.

Theorem 2 *Let n and L be arbitrarily fixed such that $2 \leq L \leq n$ holds. If K satisfies $L \leq |K|$, then M^* has full column rank $|T(n,L)|$.*

Proof: We denote by $\mathcal{Z}(n)$ the set of monomials $z_0^{d_0} \cdot \ldots \cdot z_n^{d_n}$, where $0 \leq d_i \leq |K| - 1$ for $i = 0, \ldots, n$. Obviously, the total number of such monomials is $|\mathcal{Z}(n)| = |K|^{n+1}$. Let us recall the aforementioned fact that the relation $z^1 = z^{|K|}$ holds for all $z \in K$ (including $0 \in K$). This straightforwardly implies that each monomial in the variables $z_0, \ldots, z_n$ is (as a function from K^{n+1} to K) equivalent to a monomial in $\mathcal{Z}(n)$. Let $\mu_{J,j}$ denote the monomial $\mu_{J,j} = z_0^{L-j} \prod_{r \in J} z_r$ for all $J \subseteq [n]$ and j, $0 \leq j \leq L$. The following lemma can be easily verified:

Lemma 2.1 *For all $J \subseteq [n]$, $1 \leq |J| \leq L$, and j, $|J| \leq j \leq L$, and examples $(u,w) \in \{0,1\}^n \times K$, it holds that $\mu_{J,j}(w,u)$ equals the coefficient in M^* which has the coordinates (u,w) and $t_{J,j}$.* □

For $i = 1, \ldots, |K|$, we denote by k_i the i-th element of the finite field K. Moreover, we suppose the convention that $0^0 = 1$ in K. Let (u,w) be an example defined as above and keep in mind that we are treating the case $L \leq |K|$. It should be observed that the coefficients in the corresponding equation of type (2) are given by w^{L-j}, where $1 \leq j \leq L$. Thus, the set of possible exponents $\{L - j \mid 1 \leq j \leq L\}$ is bounded from above by $(L-1) < L \leq |K|$. It follows straightforwardly from Lemma 2.1 that the (distinct) columns of M^* are columns of the matrix $W \otimes B^{\otimes n}$, where

$$W = \left(k_i^j\right)_{i=1,\ldots,|K|, j=0,\ldots,|K|-1} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

[5] It can be seen easily that for random linear functions $f_1, \ldots, f_L$, the relation $\{(u, f_l(u)) \mid u \in \{0,1\}^n, l \in [L]\} \neq \{0,1\}^n \times K$ will always hold if $L < |K|$ and is still very likely to hold if $L = |K|$.

As W and B are regular, $W \otimes B^{\otimes n}$ is regular, too. This, in turn, implies that the columns of M^* are linearly independent, thus proving Theorem 2. $\square$

We will see in section 4 that for $|K| \in O\left(dnL^4\right)$, the strong solution can be reconstructed from the weak solution in time $n^{O(L)}$ with error probability at most d^{-1}. Furthermore, section 5 will feature an experimental assessment of the number of random (valid) observations needed until $M\left(\mathcal{V}\right)$ achieves full column rank $|T\left(n, L\right)|$ for various combinations of n, L, and K (see table 2).

4 On Computing a Strong Solution from the Unique Weak Solution

Let n, K, L, and $f_1, \ldots, f_L$ be defined as before. Remember that the goal of our learning algorithm is to compute a strong solution fully characterized by the L sets $\left\{\left(e^i, f_l\left(e^i\right)\right) \mid i \in [n]\right\}, l = 1, \ldots, L$, where e^i denotes the i-th element of the standard basis of $GF(2)^n$ and $f_l\left(e^i\right) = x_i^l \in K$. Obviously, this information can equivalently be expressed as a matrix $A \in K^{n \times L}$ defined by $A_{i,\cdot} = \left(x_i^1, \ldots, x_i^L\right)$ for all $i = 1, \ldots, n$.

Hence, we have to solve the following problem: Compute the matrix $A \in K^{n \times L}$ from the information $t\left(A\right)$, where

$$t\left(A\right) = \left(t_{J,j}\left(A\right)\right)_{J \subseteq [n], 1 \leq |J| \leq L, |J| \leq j \leq L}$$

is the unique weak solution determined previously. But before we lay out how (and under which conditions) a strong solution A can be found, we need to introduce the following two definitions along with an important theorem linking them:

Definition 2. *Let for all vectors $x \in K^L$ the signature $sgt\left(x\right)$ of x be defined as $sgt\left(x\right) = \left(|x|_k\right)_{k \in K}$, where $|x|_k$ denotes the number of components of x which equal k.*

Furthermore, consider the following new family of polynomials:

Definition 3. *a) For all $L \geq 1$ and $j \geq 0$ let the simple symmetric polynomial s_j over the variables $x_1, \ldots, x_L$ be defined by $s_0 = 1$ and*

$$s_j = \bigoplus_{S \subseteq [L], |S| = j} m_S,$$

where $m_S = \prod_{i \in S} x_i$ for all $S \subseteq [L]$. Moreover, we denote

$$s\left(x\right) = \left(s_0\left(x\right), s_1\left(x\right), \ldots, s_L\left(x\right)\right)$$

for all $x \in K^L$.

b) Let n, L, $1 \leq L \leq n$, hold as well as j, $0 \leq j \leq L$, and $J \subseteq [n]$. The symmetric polynomial $s_{J,j} : K^{n \times L} \longrightarrow K$ is defined by

$$s_{J,j}(A) = s_j\left(\bigoplus_{i \in J} A_{i,\cdot}\right)$$

for all matrices $A \in K^{n \times L}$. Moreover, we denote

$$s_J(A) = (s_{J,0}(A), \ldots, s_{J,L}(A)).$$

The concept of signatures introduced in Definition 2 and the family of simple symmetric polynomials described in Definition 3 will now be connected by the following theorem:

Theorem 3 *For all $L \geq 1$ and $x, x' \in K^L$ it holds that $s(x) = s(x')$ if and only if $sgt(x) = sgt(x')$.*

Proof: See appendix A.

Building on this result, we can then prove the following proposition, which is of vital importance for computing the strong solution A on the basis of the corresponding weak solution $t(A)$:

Theorem 4 *Let $A \in K^{n \times L}$ and $t(A)$ be defined as before. For each subset $I \subseteq [n]$ of rows of A, the signature of the sum of these rows, i.e., $sgt\left(\bigoplus_{i \in I} A_{i,\cdot}\right)$, can be computed by solely using information derived from $t(A)$, in particular, without knowing the underlying matrix A itself.*

Proof: We first observe that the s-polynomials can be written as linear combinations of the t-polynomials. Trivially, the relation $t_{\{i\},j} = s_{\{i\},j}$ holds for all $i \in [n]$ and j, $1 \leq j \leq L$. Moreover, for all $I \subseteq [n]$, $|I| > 1$, it holds that

$$s_{I,j} = \bigoplus_{Q \subseteq I, 1 \leq |Q| \leq j} \left(\bigoplus_{g:[L] \longrightarrow [n], |dom(g)|=j, im(g)=Q} m_g \right) = \bigoplus_{Q \subseteq I, 1 \leq |Q| \leq j} t_{Q,j}. \quad (5)$$

Note that for all $J \subseteq [n]$ and j, $|J| \leq j \leq L$, relation (5) implies

$$t_{J,j} = s_{J,j} \oplus \bigoplus_{Q \subset J} t_{Q,j}. \quad (6)$$

By an inductive argument, we obtain from relation (6) that the converse is also true, i.e., the t-polynomials can be written as linear combinations of the s-polynomials.

We have seen so far that given $t(A)$, we are able to compute $s_{I,j}$ for all j, $1 \leq j \leq L$, and each subset $I \subseteq [n]$ of rows of A. Recall

$$s_{I,j}(A) = s_j\left(\bigoplus_{i \in I} A_{i,\cdot}\right) \quad \text{and} \quad s_I(A) = (s_{I,0}(A), \ldots, s_{I,L}(A))$$

from Definition 3 and let $x \in K^L$ be defined by $x = \bigoplus_{i \in I} A_{i,\cdot}$. It can be easily seen that $s_I(A) = s(x)$ holds.

In conjunction with Theorem 3, this straightforwardly implies the validity of Theorem 4. □

Naturally, it remains to assess the degree of usefulness of this information when it comes to reconstructing the strong solution $A \in K^{n \times L}$. In the following, we will prove that if K is large enough, then with high probability, A can be completely (up to column permutations) and efficiently derived from the signatures of all single rows of A and the signatures of all sums of pairs of rows of A:

Theorem 5 *Let $K = GF(2^a)$ fulfill $|K| \geq \frac{1}{4} \cdot d \cdot n \cdot L^4$, i.e., $a \geq \log(n) + \log(d) + 4\log(L) - 2$. Then, for a random matrix $A \in_U K^{n \times L}$, the following is true with a probability of approximately at least $\left(1 - \frac{1}{d}\right)$: A can be completely reconstructed from the signatures $sgt(A_{i,\cdot})$, $1 \leq i \leq n$, and $sgt(A_{i,\cdot} \oplus A_{j,\cdot})$, $1 \leq i < j \leq n$.*

Proof: See appendix A.

As we have seen now that, under certain conditions, it is possible to fully reconstruct the strong solution A by solely resorting to information obtained from the weak solution $t(A)$, we can proceed to actually describe a conceivable approach for step (3) of the learning algorithm:

We choose a constant error parameter d and an exponent a, i.e., $K = GF(2^a)$, in such a way that Theorem 5 can be applied. Note that $L \leq n$ and $|K| \in n^{O(1)}$. In a pre-computation, we generate two databases DB_1 and DB_2 of size $n^{O(L)}$. While DB_1 acts as a lookup table with regard to the one-to-one relation between $s(x)$ and $sgt(x)$ for all $x \in K^L$, we use DB_2 to store all triples of signatures $S, S', \tilde{S}$ for which there is exactly one solution pair $x, y \in K^L$ fulfilling $sgt(x) = S$ and $sgt(y) = S'$ as well as $sgt(x \oplus y) = \tilde{S}$.

Given $t(A)$, i.e., the previously determined weak solution, we then compute $sgt(A_{i,\cdot})$ for all i, $1 \leq i \leq n$, and $sgt(A_{i,\cdot} \oplus A_{j,\cdot})$ for all i, j, $1 \leq i < j \leq n$, in time $n^{O(1)}$ by using DB_1 and relation (5), which can be found in the proof of Theorem 4. According to Theorem 5, it is now possible to reconstruct A by the help of database DB_2 with probability at least $1 - \frac{1}{d}$.

5 Experimental Results

To showcase the detailed workings of our learning algorithm as well as to evaluate its efficiency at a practical level, we created a complete implementation using the computer algebra system Magma. In case of success, it takes approximately 90 seconds on standard PC hardware (Intel i7, 2.66 GHz, with 6 GB RAM) to compute the unique strong solution on the basis of a set of 10,000 randomly generated examples for $n = 10$, $a = 3$ (i.e., $K = GF(2^a)$), and $L = 5$. Relating to this, we performed various simulations in order to assess the corresponding

Parameters			Performed Iterations				
			Rank of $M(\mathcal{V}) < \lvert T(n,L)\rvert$		Rank of $M(\mathcal{V}) = \lvert T(n,L)\rvert$		Total
n	K	L	Number	Ratio	Number	Ratio	Number
4	$GF\left(2^2\right)$	2	37	0.37 %	9,963	99.63 %	10,000
4	$GF\left(2^2\right)$	3	823	8.23 %	9,177	91.77 %	10,000
4	$GF\left(2^2\right)$	4	7,588	75.88 %	2,412	24.12 %	10,000
5	$GF\left(2^2\right)$	4	4,556	45.56 %	5,444	54.44 %	10,000
5	$GF\left(2^2\right)$	5	10,000	100.00 %	0	0.00 %	10,000
6	$GF\left(2^3\right)$	4	0	0.00 %	1,000	100.00 %	1,000
8	$GF\left(2^3\right)$	4	0	0.00 %	1,000	100.00 %	1,000
8	$GF\left(2^3\right)$	6	0	0.00 %	100	100.00 %	100
8	$GF\left(2^3\right)$	7	0	0.00 %	100	100.00 %	100
8	$GF\left(2^3\right)$	8	0	0.00 %	100	100.00 %	100
9	$GF\left(2^3\right)$	8	0	0.00 %	10	100.00 %	10
9	$GF\left(2^3\right)$	9	10	100.00 %	0	0.00 %	10

Table 1. An estimate of the rank of $M(\mathcal{V})$ on the basis of all possible valid observations for up to 10,000 randomly generated instances of $RandomSelect(L,n,a)$. For each choice of parameters, $|T(n,L)|$ denotes number of columns of $M(\mathcal{V})$ as defined in section 2 and listed in table 2.

probabilities, which were already discussed in sections 3 and 4 from a theoretical point of view.

The experimental results summarized in table 1 clearly suggest that if $|K|$ is only slightly larger than the number L of secret linear functions, then in all likelihood, $M(\mathcal{V})$ will eventually reach full (column) rank $|T(n,L)|$, thus allowing for the computation of a unique weak solution. Moreover, in accordance with Corollary 1, the columns of $M(\mathcal{V})$ were always linearly dependent in the case of $n=5$, $K=GF\left(2^2\right)$ and $L=5$, i.e., $|K|=4<5=L$. A further analysis of the underlying data revealed in addition that, for arbitrary combinations of n, K, and L, the matrix $M(\mathcal{V})$ never reached full column rank if at least two of the corresponding L random linear functions $f_1,\ldots,f_L$ were identical during an iteration of our experiments. Note that, on the basis of the current implementation, it was not possible to continue table 1 for larger parameter sizes because, e.g., in the case of $n=8$, $K=GF\left(2^3\right)$ and $L=7$, performing as few as 100 iterations already took more than 85 minutes on the previously described computer system.

Table 2 features additional statistical data with respect to the number of examples needed (in case of success) until the matrix $M(\mathcal{V})$ reaches full column rank $|T(n,L)|$. Please note that, in contrast to the experiments underlying table 1, such examples $(u,f_l(u))$ are generated iteratively and independently choosing random pairs $u\in_U\{0,1\}^n$ and $l\in_U[L]$, i.e., they are not processed in their canonical order but observed randomly (and also repeatedly) to simulate a practical passive attack. While we have seen previously that for most choices of n, K and L, the matrix $M(\mathcal{V})$ is highly likely to eventually reach full column rank, the experimental results summarized in table 2, most no-

Parameters				Number of Random Examples until Rank $(M(\mathcal{V})) = \lvert T(n,L)\rvert$							
n	K	L	$\lvert T(n,L)\rvert$	Avg.	Max.	Min.	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
4	$GF\left(2^2\right)$	1	4	5.5	18	4	4	4	5	6	8
4	$GF\left(2^2\right)$	2	14	24.4	93	14	18	20	23	27	32
4	$GF\left(2^2\right)$	3	28	71.8	273	33	51	58	67	81	99
4	$GF\left(2^2\right)$	4	43	226.2	701	95	147	175	211	261	317
5	$GF\left(2^2\right)$	4	75	218.5	591	140	176	192	211	237	263
6	$GF\left(2^3\right)$	4	124	201.6	318	162	184	192	200	211	220
8	$GF\left(2^3\right)$	4	298	378.7	419	345	365	371	378	386	393
8	$GF\left(2^3\right)$	6	762	1401.6	1565	1302	1342	1364	1405	1427	1458
8	$GF\left(2^3\right)$	7	1016	2489.7	2731	2275	2369	2417	2477	2547	2645
8	$GF\left(2^3\right)$	8	1271	5255.3	7565	4302	4706	4931	5227	5557	5706
9	$GF\left(2^3\right)$	8	2295	6266.1	6553	6027	6078	6136	6199	6415	6504

Table 2. An estimate of the number of randomly generated examples $(u, f_l(u))$ which have to be processed (in case of success) until the matrix $M(\mathcal{V})$ reaches full column rank $\lvert T(n,L)\rvert$. Given a probability p, we denote by Q_p the p-quantile of the respective sample.

tably the observed p-quantiles, strongly suggest that our learning algorithm for $RandomSelect(L,n,a)$ will also be able to efficiently construct a corresponding LES which allows for computing a unique weak solution.

Parameters			Performed Iterations (i.e., randomly chosen $A \in_U K^{n\times L}$)				
			A **not** $sgt(2)$-identifiable		A **was** $sgt(2)$-identifiable		Total
n	K	L	Number	Ratio	Number	Ratio	Number
4	$GF\left(2^2\right)$	2	0	0.00 %	10,000	100.00 %	10,000
4	$GF\left(2^2\right)$	3	69	0.69 %	9,931	99.31 %	10,000
4	$GF\left(2^2\right)$	4	343	3.43 %	9,657	96.57 %	10,000
6	$GF\left(2^3\right)$	4	0	0.00 %	10,000	100.00 %	10,000
8	$GF\left(2^3\right)$	4	0	0.00 %	10,000	100.00 %	10,000
8	$GF\left(2^3\right)$	6	0	0.00 %	1,000	100.00 %	1,000
8	$GF\left(2^3\right)$	7	0	0.00 %	1,000	100.00 %	1,000
8	$GF\left(2^3\right)$	8	0	0.00 %	100	100.00 %	100
9	$GF\left(2^3\right)$	8	0	0.00 %	100	100.00 %	100

Table 3. An estimate of the ratio of $sgt(2)$-identifiable $(n \times L)$-matrices over K.

It remains to clear up the question, to what extent Theorem 5 reflects reality concerning the probability of a random $(n \times L)$-matrix over K being $sgt(2)$-identifiable (see Definitions 5.1 and 5.2 in the proof of Theorem 5), which is necessary and sufficient for the success of step (3) of our learning algorithm. Our corresponding simulations yielded table 3, which immediately suggests that even for much smaller values of $\lvert K\rvert$ than those called for in Theorem 5, a strong solution $A \in_U K^{n\times L}$ can be completely reconstructed from the signatures $sgt(A_{i,\cdot})$,

$1 \leq i \leq n$, and $sgt\,(A_{i,\cdot} \oplus A_{j,\cdot})$, $1 \leq i < j \leq n$. In conjunction with the experimental results concerning the rank of $M\,(\mathcal{V})$, this, in turn, implies that our learning algorithm will efficiently lead to success in the vast majority of cases.

6 Discussion

The running time of our learning algorithm for $RandomSelect\,(L, n, a)$ is dominated by the complexity of solving a system of linear equations with $|T(n, L)|$ unknowns. Our hardness conjecture is that this complexity also constitutes a lower bound to the complexity of $RandomSelect\,(L, n, a)$ itself, which would imply acceptable cryptographic security for parameter choices like $n = 128$ and $L = 8$ or $n = 256$ and $L = 6$. The experimental results summarized in the previous section clearly support this view. Consequently, employing the principle of random selection to design new symmetric lightweight authentication protocols might result in feasible alternatives to current HB-based cryptographic schemes.

A problem of independent interest is to determine the complexity of reconstructing an $sgt\,(r)$-identifiable matrix A from the signatures of all sums of at most r rows of A. Note that this problem is wedded to determining the complexity of $RandomSelect\,(L, n, a)$ with respect to an *active* learner, who is able to receive examples (u, w) for inputs u of his choice, where $w = f_l\,(u)$ and $l \in_U [L]$ is randomly chosen by the oracle. It is easy to see that such learners can efficiently compute $sgt\,(f_1\,(u), \ldots, f_L\,(u))$ by repeatedly asking for u. As the approach for reconstructing A which was outlined in section 4 needs a data structure of size exponential in L, it would be interesting to know if there are corresponding algorithms of time and space costs polynomial in L.

From a theoretical point of view, another open problem is to determine the probability that a random $(n \times L)$-matrix over K is $sgt\,(r)$-identifiable for some r, $2 \leq r \leq L$. Based on the results of our computer experiments, it appears more than likely that the lower bound derived in Theorem 5 is far from being in line with reality and that identifiable matrices occur with much higher probability for fields K of significantly smaller size.

References

1. F. Armknecht and M. Krause. Algebraic attacks on combiners with memory. In *Proceedings of Crypto 2003*, volume 2729 of *LNCS*, pages 162–176. Springer, 2003.
2. S. Arora and R. Ge. New algorithms for learning in presence of errors. Submitted, 2010. `http://www.cs.princeton.edu/~rongge/LPSN.pdf`.
3. E.-O. Blass, A. Kurmus, R. Molva, G. Noubir, and A. Shikfa. The F_f-family of protocols for RFID-privacy and authentication. In *5th Workshop on RFID Security, RFIDSec'09*, 2009.
4. A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. H. Vikkelsoe. PRESENT: An ultra-lightweight block cipher. In *Proceedings of Cryptographic Hardware and Embedded Systems (CHES) 2007*, volume 4727 of *LNCS*, pages 450–466. Springer, 2007.

5. W. Bosma, J. Cannon, and C. Playoust. The Magma algebra system. I. The user language. *J. Symbolic Comput.*, 24(3-4):235–265, 1997.
6. J. Bringer and H. Chabanne. Trusted-HB: A low cost version of HB^+ secure against a man-in-the-middle attack. *IEEE Trans. Inform. Theor.*, 54:4339–4342, 2008.
7. J. Cichoń, M. Klonowski, and M. Kutyłowski. Privacy protection for RFID with hidden subset identifiers. In *Proceedings of Pervasive 2008*, volume 5013 of *LNCS*, pages 298–314. Springer, 2008.
8. N. Courtois. Fast algebraic attacks on stream ciphers with linear feedback. In *Proceedings of Crypto 2003*, volume 2729 of *LNCS*, pages 176–194. Springer, 2003.
9. N. Courtois and W. Meier. Algebraic attacks on stream ciphers with linear feedback. In *Proceedings of Eurocrypt 2003*, volume 2656 of *LNCS*, pages 345–359. Springer, 2003.
10. C. De Cannière, O. Dunkelman, and M. Knežević. KATAN and KTANTAN – A family of small and efficient hardware-oriented block ciphers. In *Proceedings of the 11th International Workshop on Cryptographic Hardware and Embedded Systems (CHES) 2009*, volume 5747 of *LNCS*, pages 272–288. Springer, 2009.
11. Z. Gołębiéwski, K. Majcher, and F. Zagórski. Attacks on CKK family of RFID authentication protocols. In *Proceedings Adhoc-now 2008*, volume 5198 of *LNCS*, pages 241–250. Springer, 2008.
12. D. Frumkin and A. Shamir. Untrusted-HB: Security vulnerabilities of Trusted-HB. Cryptology ePrint Archive, Report 2009/044, 2009. `http://eprint.iacr.org`.
13. H. Gilbert, M. J. B. Robshaw, and Y. Seurin. $HB^{\#}$: Increasing the security and efficiency of HB^+. In *Proceedings of Eurocrypt 2008*, volume 4965 of *LNCS*, pages 361–378, 2008.
14. H. Gilbert, M. J. B. Robshaw, and H. Sibert. Active attack against HB^+: A provable secure lightweight authentication protocol. *Electronic Letters*, 41:1169–1170, 2005.
15. O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing (STOC)*, pages 25–32. ACM Press, 1989.
16. N. J. Hopper and M. Blum. Secure human identification protocols. In *Proceedings of Asiacrypt 2001*, volume 2248 of *LNCS*, pages 52–66. Springer, 2001.
17. A. Juels and S. A. Weis. Authenticating pervasive devices with human protocols. In *Proceedings of Crypto 2005*, volume 3621 of *LNCS*, pages 293–308. Springer, 2005.
18. M. Krause and D. Stegemann. More on the security of linear RFID authentication protocols. In *Proceedings of SAC 2009*, volume 5867 of *LNCS*, pages 182–196. Springer, 2009.
19. W. Meier, E. Pasalic, and C. Carlet. Algebraic attacks and decomposition of boolean functions. In *Proceedings of Eurocrypt 2004*, volume 3027 of *LNCS*, pages 474–491. Springer, 2004.
20. K. Ouafi, R. Overbeck, and S. Vaudenay. On the security of $HB^{\#}$ against a man-in-the-middle attack. In *Proceedings of Asiacrypt 2008*, volume 5350 of *LNCS*, pages 108–124. Springer, 2008.
21. O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (STOC)*, pages 84–93. ACM Press, 2005.
22. A. Shamir, J. Patarin, N. Courtois, and A. Klimov. Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In *Proceedings of Eurocrypt 2000*, volume 1807 of *LNCS*, pages 474–491. Springer, 2000.

A The Proofs of Theorems 3 and 5

A.1 The Proof of Theorem 3

Theorem 3 *For all $L \geq 1$ and $x, x' \in K^L$ it holds that $s(x) = s(x')$ if and only if $sgt(x) = sgt(x')$.*

Proof: The *if*-direction of Theorem 3 follows directly from the definitions of $sgt(x)$ and $s(x)$.

We will prove the *only-if*-direction of Theorem 3 by induction on L. The case $L = 1$ is obvious. Let us fix an arbitrary $L > 1$ and let us suppose that the following is true for all $L' < L$ and all $x, x' \in K^{L'}$: if $s(x) = s(x')$ then $sgt(x) = sgt(x')$.

Lemma 3.1 *For all $x \in K^{L-1}$, $k \in K$ and j, $1 \leq j \leq L$, the following is true: $s_j(x, k) = s_j(x) \oplus k \cdot s_{j-1}(x)$.* □

Henceforth, for all $k \in K$, $r \geq 1$ and $x \in K^r$, we write $k \in x$ if some component of x equals k.

Lemma 3.2 *For all $y, y' \in K^L$, the following is true: if $s(y) = s(y')$ and there is some $k \in K$ with $k \in y$ and $k \in y'$, then $sgt(y) = sgt(y')$.*

Proof of Lemma 3.2: Suppose, w.l.o.g., that $y = (x, k)$ and $y' = (x', k)$, where $x, x' \in K^{L-1}$. It suffices to prove that $s(x) = s(x')$ as this implies by induction hypothesis that $sgt(x) = sgt(x')$ and, consequently, $sgt(y) = sgt(y')$. We prove $s(x) = s(x')$ by showing via induction on j that $s_j(x) = s_j(x')$ holds for all j, $0 \leq j \leq L$. The case $j = 0$ follows straightforwardly from Definition 3. Let us fix some $j > 0$ and suppose that $s_r(x) = s_r(x')$ holds for all non-negative integers $r < j$. As $s_j(y) = s_j(y')$ is satisfied, it follows from Lemma 3.1, in conjunction with the induction hypothesis, that

$$s_j(x) + k \cdot s_{j-1}(x) = s_j(x') + k \cdot s_{j-1}(x') = s_j(x') + k \cdot s_{j-1}(x)$$

and, consequently, $s_j(x) = s_j(x')$ holds. □

Lemma 3.3 *For all $y, y' \in K^L$, the following is true: if $s(y) = s(y')$ and $0 \in y$, then $sgt(y) = sgt(y')$.*

Proof of Lemma 3.3: Trivially, $s(y) = s(y')$ implies that $s_L(y) = s_L(y')$. As $0 \in y$, it follows directly from Definition 3 that $s_L(y) = 0$, which, in turn, implies that $s_L(y') = 0$ and, consequently, $0 \in y'$. The proof now follows from Lemma 3.2. □

Finally, we have to consider the last remaining case of $y, y' \in K^L$ given by

- $s(y) = s(y')$,
- $0 \notin y$ and $0 \notin y'$,

– $Y \cap Y' = \emptyset$,

where Y (Y') denotes the set of components of y (y').

In order to proceed, we need the following technical definition, accompanied by two technical lemmas:

Definition 3.1 *For all $r \geq 1$ and $x \in K^r$, we denote*

$$S(x) = \bigoplus_{j=0}^{r} s_j(x).$$

Lemma 3.4 *For all $r \geq 1$ and $x \in K^r$, the following is true: $S(x) = 0$ if and only if $1 \in x$.*

Proof of Lemma 3.4: We prove the lemma by induction on r. The case $r = 1$ can be easily verified. Let us fix some $r > 1$ and suppose that for all q, $1 \leq q \leq (r-1)$, and all $z \in K^q$, the following is true: $S(z) = 0$ if and only if $1 \in z$. Furthermore, let us fix some $x \in K^r$ satisfying $S(x) = 0$ and suppose that $x = (z, k)$ for $z \in K^{r-1}$ and $k \in K$. In conjunction with Lemma 3.1, this yields the following equation:

$$\begin{aligned} 0 = S(x) = 1 \oplus \bigoplus_{j=1}^{r} s_j(x) &= 1 \oplus \bigoplus_{j=1}^{r} (s_j(z) \oplus k \cdot s_{j-1}(z)) \\ &= 1 \oplus \bigoplus_{j=1}^{r-1} s_j(z) \oplus k \cdot \bigoplus_{j=1}^{r} s_{j-1}(z) \\ &= \bigoplus_{j=0}^{r-1} s_j(z) \oplus k \cdot \bigoplus_{j=0}^{r-1} s_j(z) \\ &= (1 \oplus k) \cdot S(z) \end{aligned}$$

Consequently, either $k = 1$ or, by induction hypothesis, $1 \in z$. □

Lemma 3.5 *For all $r \geq 1$ and $x, x' \in K^r$, the following is true: if $s(x) = s(x')$, then $s(k \cdot x) = s(k \cdot x')$.*

Proof of Lemma 3.5: This follows straightforwardly from the simple fact that $s_j(k \cdot x) = k^j \cdot s_j(x)$ holds for all j, $0 \leq j \leq r$. □

The finding given below will complete the proof of Theorem 3:

Lemma 3.6 *For all $y, y' \in K^L$, the following is true: if $0 \notin y$ as well as $0 \notin y'$ and the sets of components of y and y' are disjoint, then $s(y) \neq s(y')$.*

Proof of Lemma 3.6: Due to Lemma 3.5, we can, w.l.o.g., suppose that $y_L = 1$. Let us denote $d = y'_L$, $y = (x, 1)$, $y' = (x', d)$ and keep in mind that $d \notin \{0, 1\}$. We will prove this lemma by contradiction. Hence, let us assume that $s(y) = s(y')$ holds.

By applying Lemma 3.1 to the assumption, we can deduce that

$$s_j(x) \oplus 1 \cdot s_{j-1}(x) = s_j(x') \oplus d \cdot s_{j-1}(x')$$

holds for all $j = 1, \ldots, L$. This implies

$$\begin{aligned} & s_1(x) \oplus 1 = s_1(x') \oplus d \\ \Leftrightarrow \quad & s_1(x') = s_1(x) \oplus d \oplus 1. \end{aligned}$$

Analogously, we obtain

$$s_2(x) \oplus s_1(x) = s_2(x') \oplus d \cdot s_1(x'),$$

i.e.,

$$\begin{aligned} s_2(x') &= s_2(x) \oplus s_1(x) \oplus d(s_1(x) \oplus d \oplus 1) \\ &= s_2(x) \oplus (d \oplus 1) s_1(x) \oplus d(d \oplus 1). \end{aligned}$$

Iterating this, one can easily show that

$$s_j(x') = s_j(x) \oplus \bigoplus_{r=1}^{j} \left(d^{r-1}(d \oplus 1) s_{j-r}(x)\right) \tag{7}$$

holds for all j, $1 \leq j \leq (L-1)$. In conjunction with the fact that

$$1 \cdot s_{L-1}(x) = s_L(y) = s_L(y') = d \cdot s_{L-1}(x')$$

in the case of $j = L$, relation (7) implies

$$\begin{aligned} d^{-1} s_{L-1}(x) &= s_{L-1}(x) \oplus \bigoplus_{r=1}^{L-1} \left(d^{r-1}(d \oplus 1) s_{L-1-r}(x)\right) \\ \Leftrightarrow \quad 0 &= d^{-1}(1 \oplus d) s_{L-1}(x) \oplus \bigoplus_{r=1}^{L-1} \left(d^{r-1}(d \oplus 1) s_{L-1-r}(x)\right) \\ \Leftrightarrow \quad 0 &= \bigoplus_{r=0}^{L-1} \left(d^{r-1}(d \oplus 1) s_{L-1-r}(x)\right). \end{aligned}$$

Multiplying this by $\left(d^{-(L-2)}\left(d \oplus 1\right)^{-1}\right)$, where $d \notin \{0,1\}$, yields

$$\begin{aligned} & 0 = \bigoplus_{r=0}^{L-1} \left(d^{-((1-r)+(L-2))} s_{L-1-r}\left(x\right)\right) \\ \Leftrightarrow \quad & 0 = \bigoplus_{r=0}^{L-1} \left(d^{-((L-1)-r)} s_{(L-1)-r}\left(x\right)\right) \\ \Leftrightarrow \quad & 0 = \bigoplus_{j=0}^{L-1} \left(d^{-j} s_j\left(x\right)\right) \\ \Leftrightarrow \quad & 0 = S\left(d^{-1} \cdot x\right). \end{aligned}$$

In conjunction with Lemma 3.4, this implies that $1 \in \left(d^{-1} \cdot x\right)$, which, in turn, means that $d \in x$. Consequently, $d \in y$ and also $d \in y'$, which violates the condition that the sets of components of y and y' are disjoint. Hence, the assumption $s\left(y\right) = s\left(y'\right)$ must be false. □

To conclude, Theorem 3 now follows straightforwardly from Lemma 3.2, Lemma 3.3 and Lemma 3.6. □

A.2 The Proof of Theorem 5

Theorem 5 *Let $K = GF\left(2^a\right)$ fulfill $|K| \geq \frac{1}{4} \cdot d \cdot n \cdot L^4$, i.e., $a \geq \log\left(n\right) + \log\left(d\right) + 4\log\left(L\right) - 2$. Then, for a random matrix $A \in_U K^{n \times L}$, the following is true with a probability of approximately at least $\left(1 - \frac{1}{d}\right)$: A can be completely reconstructed from the signatures $sgt\left(A_{i,\cdot}\right)$, $1 \leq i \leq n$, and $sgt\left(A_{i,\cdot} \oplus A_{j,\cdot}\right)$, $1 \leq i < j \leq n$.*

Proof: In order to prove the theorem, we first need the following definition:

Definition 5.1 *a) Two matrices $A, B \in K^{n \times L}$ are called column-equivalent if A can be obtained from B by permuting the columns.*
b) Two matrices $A, B \in K^{n \times L}$ are called $sgt\left(r\right)$-equivalent if

$$sgt\left(\bigoplus_{i \in I} A_{i,\cdot}\right) = sgt\left(\bigoplus_{i \in I} B_{i,\cdot}\right)$$

holds for all $I \subseteq [n]$, $1 \leq |I| \leq r$.

Clearly, if the matrices A and B are column-equivalent, then they are also $sgt\left(r\right)$-equivalent for all r, $1 \leq r \leq L$. The converse is not necessarily true as can be seen from the following example, where $A, B \in GF\left(8\right)^{2 \times 3}$:

$$A = \begin{bmatrix} 1+z & 1+z^2 & 0 \\ z^2 & 1 & z \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1+z & 1+z^2 & 0 \\ 1 & z & z^2 \end{bmatrix}.$$

This crucial observation leads to the following definition:

Definition 5.2 *A matrix $A \in K^{n\times L}$ is called $sgt\,(r)$-identifiable if $sgt\,(r)$-equivalence to A implies column-equivalence to A.*

We show Theorem 5 by proving a lower bound on the probability of a random matrix $A \in_U K^{n\times L}$ being $sgt\,(2)$-identifiable. In order to do so, we will now introduce a sufficient condition for the $sgt\,(2)$-identifiability of an $(n \times L)$-matrix over K and further show that with high probability, it is fulfilled if $|K|$ is large enough.

Definition 5.3 *a) For all $L \geq 1$ and vectors $x \in K^L$, we denote by $\{x\}$ the set of all $k \in K$ occurring in x, i.e., $\{x\} = \{k \in K \mid |x|_k > 0\}$.*
b) For all subsets $M \subseteq K$, we denote by $\Delta\,(M)$ the set of differences generated by M, i.e., $\Delta\,(M) = \{k \oplus k' \mid k \neq k' \in M\}$.
c) Two subsets $M, M' \subseteq K$ are called diff.disjoint if $\Delta\,(M) \cap \Delta\,(M') = \emptyset$.
d) A matrix $A \in K^{n\times L}$ is called strongly diff.disjoint if there is some $i \in [n]$ such that $|\{A_{i,\cdot}\}| = L$ and, for all $j \in [n] \setminus \{i\}$, $\{A_{i,\cdot}\}$ and $\{A_{j,\cdot}\}$ are diff.disjoint.

In addition, we need the following technical lemma:

Lemma 5.1 *Let $M, M' \subseteq K$ be two given subsets which are diff.disjoint. For all $m_1, m_2 \in M$ and $m'_1, m'_2 \in M'$, the following is true: if $m_1 \oplus m'_1 = m_2 \oplus m'_2$ then $m_1 = m_2$ and $m'_1 = m'_2$.*

Proof of Lemma 5.1: Trivially, $m_1 \oplus m'_1 = m_2 \oplus m'_2$ can be transformed into $m_1 \oplus m_2 = m'_1 \oplus m'_2$. The latter relation would obviously violate the condition of M and M' being diff.disjoint if $m_1 \neq m_2$ (and thus $m'_1 \neq m'_2$) held. □

The following lemma states a sufficient condition for the $sgt\,(2)$-identifiability of an $(n \times L)$-matrix over K:

Lemma 5.2 *If a matrix $A \in K^{n\times L}$ is strongly diff.disjoint, then A is also $sgt\,(2)$-identifiable.*

Proof of Lemma 5.2: Let us consider a strongly diff.disjoint matrix $A \in K^{n\times L}$ and suppose that, w.l.o.g., $|\{A_{1,\cdot}\}| = L$ holds (i.e., the first row of A contains the maximum number L of different elements). Furthermore, let us fix some matrix $B \in K^{n\times L}$ which is $sgt\,(2)$-equivalent to A. In order to prove the lemma, we have to show that A and B are column-equivalent.

As A and B are $sgt\,(2)$-equivalent, we know that $sgt\,(A_{1,\cdot}) = sgt\,(B_{1,\cdot})$ holds, implying the existence of some column-permutation $\rho \in \mathcal{S}_L$ such that $A_{1,\cdot} = \rho\,(B_{1,\cdot})$. Now let us fix some arbitrary j, $1 < j \leq n$. From $sgt\,(A_{j,\cdot})$, we learn which elements occur in row $B_{j,\cdot}$ and from $sgt\,(A_{1,\cdot} \oplus A_{j,\cdot})$, we learn which elements occur in $B_{1,\cdot} \oplus B_{j,\cdot}$. As $\{B_{1,\cdot}\}$ and $\{B_{j,\cdot}\}$ are diff.disjoint, Lemma 5.1 implies that for each element occurring in $B_{1,\cdot} \oplus B_{j,\cdot}$, there is exactly one possibility of writing it as the sum of an element from $B_{1,\cdot}$ and an element from $B_{j,\cdot}$. Moreover, these two elements have to be in the same column of B.

Due to this and the fact that all components of $B_{1,\cdot}$ are different, the positions of all elements occurring in $B_{j,\cdot}$ are uniquely determined. In particular, the aforementioned column-permutation ρ not only satisfies $A_{1,\cdot} = \rho(B_{1,\cdot})$ but also $A_{j,\cdot} = \rho(B_{j,\cdot})$ for all $1 < j \leq n$. Clearly, this proves the column-equivalence of A and B, thus implying the correctness of the lemma. $\square$

Consequently, Theorem 5 can be shown by proving an appropriate lower bound on the probability of a random matrix $A \in_U K^{n \times L}$ being strongly diff.disjoint. Our argument will be based on the following lemma:

Lemma 5.3 *Given a subset $M \subseteq K$ such that $|M| = L$ holds and a sequence $x = (x_1, \ldots, x_L)$ chosen randomly from K^L with respect to the uniform distribution, the lower bound on the probability of $\{x\}$ being diff.disjoint from M can be approximated by $1 - \frac{L^4}{4|K|}$.*

Proof of Lemma 5.3: In the course of this proof, we will make use of the approximation $\prod_{i=1}^{q-1}\left(1 - \frac{i}{p}\right) \approx e^{-\frac{q^2}{2p}}$, commonly found in the context of the well-known birthday paradox, as well as the approximations $e^{-\frac{1}{x}} \approx \left(1 - \frac{1}{x}\right)$ and $\left(1 - \frac{1}{x}\right)^x \approx e^{-1}$. In order to obtain a sequence $x \in K^L$ whose set of components is diff.disjoint from M, for all i, $1 < i \leq L$, the element x_i needs to be chosen in such a way that

$$\Delta(M) \cap \{x_i \oplus x_1, \ldots, x_i \oplus x_{i-1}\} = \emptyset.$$

The probability of this being fulfilled for a randomly (i.e., independently and uniformly) chosen x_i can be bounded from below by $\frac{|K| - (i-1) \cdot |\Delta(M)|}{|K|}$. Hence, in case of a random sequence $x \in K^L$, the probability of $\{x\}$ being diff.disjoint from M is at least

$$\begin{aligned}
&\frac{(|K| - |\Delta(M)|) \cdot (|K| - 2|\Delta(M)|) \cdot \ldots \cdot (|K| - (L-1)|\Delta(M)|)}{|K|^{L-1}} \\
&= \left(1 - \frac{1}{|K| / |\Delta(M)|}\right) \cdot \left(1 - \frac{2}{|K| / |\Delta(M)|}\right) \cdot \ldots \cdot \left(1 - \frac{L-1}{|K| / |\Delta(M)|}\right) \\
&= \prod_{i=1}^{L-1} \left(1 - \frac{i}{|K| / |\Delta(M)|}\right).
\end{aligned}$$

By applying the above-mentioned approximations, we obtain that the lower bound on the probability of $\{x\}$ being diff.disjoint from M is around

$$e^{-\frac{L^2}{2(|K|/|\Delta(M)|)}} \approx 1 - \frac{L^2}{2(|K| / |\Delta(M)|)} = 1 - \frac{L^2 \cdot |\Delta(M)|}{2|K|}.$$

As $|\Delta(M)| \leq \binom{L}{2}$, an even coarser approximation can be given by

$$1 - \frac{L^2 \cdot \binom{L}{2}}{2|K|} = 1 - \frac{L^2 \cdot \frac{L \cdot (L-1)}{2}}{2|K|} > 1 - \frac{L^4}{4|K|},$$

which proves the lemma. □

Now let $A \in_U K^{n \times L}$ be a random $(n \times L)$-matrix over K. Similarly to the argument in the previous proof, we can learn that with probability around $1 - \frac{L^2}{2|K|}$, the first row of A contains L different coefficients. In this particular case, it follows straightforwardly from Lemma 5.3 that for all j, $2 \leq j \leq n$, the lower bound on the probability of $\{A_{1,\cdot}\}$ and $\{A_{j,\cdot}\}$ being diff.disjoint can be approximated by $1 - \frac{L^4}{4|K|}$.

Consequently, if K satisfies

$$\frac{L^2}{2\,|K|} \leq \frac{L^4}{4\,|K|} \leq \frac{1}{dn}, \tag{8}$$

then due to the implication

$$\left(1 - \frac{1}{dn}\right)^n \leq \left(1 - \frac{L^2}{2\,|K|}\right) \cdot \left(1 - \frac{L^4}{4\,|K|}\right)^{n-1},$$

in conjunction with Lemmata 5.2 and 5.3, the probability that A is $sgt\,(2)$-identifiable can be bounded from below by approximately

$$\left(1 - \frac{1}{dn}\right)^n \approx e^{-\frac{1}{d}} \approx 1 - \frac{1}{d}.$$

Observing that relation (8) holds if

$$|K| \geq \frac{1}{4} \cdot (dn) \cdot L^4,$$

i.e.,

$$a \geq \log\,(n) + \log\,(d) + 4\log\,(L) - 2,$$

completes the proof of Theorem 5. □

B On attacking the $(n, k, L)^+$-protocol by solving *RandomSelect* (L, n, a)

The following outline of an attack on the $(n, k, L)^+$-protocol by Krause and Stegemann [18] is meant to exemplify the immediate connection between the previously introduced learning problem *RandomSelect* (L, n, a) and the security of this whole new class of lightweight authentication protocols. Similar to the basic communication mode described in the introduction, the $(n, k, L)^+$-protocol is based on L n-dimensional, injective linear functions $F_1, \ldots, F_L : GF(2)^n \longrightarrow GF(2)^{n+k}$ (i.e., the secret key) and works as follows.

Each instance is initiated by the verifier Alice, who chooses a random vector $a \in_U GF(2)^{n/2}$ and sends it to Bob, who then randomly (i.e., independently and

uniformly) chooses $l \in_U [L]$ along with an additional value $b \in_U GF(2)^{n/2}$, in order to compute his response $w = F_l(a, b)$. Finally, Alice accepts $w \in GF(2)^{n+k}$ if there is some $l \in [L]$ with $w \in V_l$ and the prefix of length $n/2$ of $F_l^{-1}(w)$ equals a, where V_l denotes the n-dimensional linear subspace of $GF(2)^{n+k}$ corresponding to the image of F_l.

This leads straightforwardly to a problem called *Learning Unions of L Linear Subspaces* (LULS), where an oracle holds the specifications of L secret n-dimensional linear subspaces $V_1, \ldots, V_L$ of $GF(2)^{n+k}$, from which it randomly chooses examples $v \in_U V_l$ for $l \in_U [L]$ and sends them to the learner. Knowing only n and k, he seeks to deduce the specifications of $V_1, \ldots, V_L$ from a sufficiently large set $\{w_1, \ldots, w_s\} \subseteq \bigcup_{l=1}^{L} V_l$ of such observations. It is easy to see that this corresponds to a passive key recovery attack against (n, k, L)-type protocols. Note that there is a number of exhaustive search strategies to solve this problem, e.g., the generic exponential time algorithm called search-for-a-basis heuristic, which was presented in the appendix of [18].

It should be noted that an attacker who is able to solve the LULS problem needs to perform additional steps to fully break the $(n, k, L)^+$-protocol as impersonating the prover requires to send responses $w \in GF(2)^{n+k}$ which not only fulfill $w \in \bigcup_{l=1}^{L} V_l$ but also depend on some random nonce $a \in GF(2)^{n/2}$ provided by the verifier. However, having successfully obtained the specifications of the secret subspaces $V_1, \ldots, V_L$ allows in turn for generating a specification of the image of $F_l(a, \cdot)$ for each $l \in [L]$ by repeatedly sending an arbitrary but fixed (i.e., selected by the attacker) $a \in GF(2)^{n/2}$ to the prover. Remember that, although the prover chooses a random $l \in_U [L]$ each time he computes a response w based on some fixed a, an attacker who has determined $V_1, \ldots, V_L$ will know which subspace the vector w actually belongs to. Krause and Stegemann pointed out that this strategy allows for efficiently constructing specifications of linear functions $G_1, \ldots, G_L : GF(2)^n \longrightarrow GF(2)^{n+k}$ and bijective linear functions $g_1, \ldots, g_L : GF(2)^{n/2} \longrightarrow GF(2)^{n/2}$ such that

$$F_l(a, b) = G_l(a, g_l(b))$$

for all $l \in [L]$ and $a, b \in GF(2)^{n/2}$ [18]. Hence, the efficiently obtained specifications of the functions $((G_1, \ldots, G_L), (g_1, \ldots, g_L))$ are equivalent to the actual secret key $(F_1, \ldots, F_L)$. However, keep in mind that the running time of this attack is dominated by the effort needed to solve the LULS problem first and that $RandomSelect(L, n, a)$ in fact refers to a special case of the LULS problem, which assumes that the secret subspaces have the form

$$V_l = \{(v, f_l(v)) \mid v \in GF(2)^n\} \subseteq GF(2)^{n+k}$$

for all $l \in [L]$ and secret $GF(2)$-linear functions $f_1, \ldots, f_L : GF(2)^n \longrightarrow GF(2)^k$. This is true with probability $p(n) \approx 0.2887$ as, given an arbitrary $((n + k) \times n)$-matrix A over $GF(2)$, the general case $V = \{A \circ v \mid v \in GF(2)^n\}$ can be written in the special form iff the first n rows of A are linearly independent (see, e.g., [11]).

In order to solve this special problem efficiently, we suggest the following approach, which makes use of our learning algorithm for $RandomSelect\,(L, n, a)$ and works by

- determining an appropriate number $a \in O\,(\log\,(n))$ which, w.l.o.g., divides k (i.e., $k = \gamma \cdot a$ for some $\gamma \in \mathbb{N}$),
- identifying vectors $w \in \{0,1\}^k$ with vectors $w = (w_1, \ldots, w_\gamma) \in GF\,(2^a)^\gamma$ and functions $f : \{0,1\}^n \longrightarrow \{0,1\}^k$ with γ-tuples $\left(f^1, \ldots, f^\gamma\right)$ of component functions $f^1, \ldots, f^\gamma : \{0,1\}^n \longrightarrow GF\,(2^a)$ based on the following rule: $f^i\,(u) = w_i$ for all $i = 1, \ldots, \gamma$ if and only if $f\,(u) = (w_1, \ldots, w_\gamma)$,
- learning $f_1, \ldots, f_L : \{0,1\}^n \longrightarrow \{0,1\}^k$ by learning each of the corresponding sets of component functions $f_1^i, \ldots, f_L^i : \{0,1\}^n \longrightarrow GF\,(2^a)$ in time $n^{O(L)}$ for $i = 1, \ldots, \gamma$.

Clearly, for efficiency reasons, a should be as small as possible. However, in section 4 we show that a needs to exceed a certain threshold, which can be bounded from above by $O\,(\log\,(n))$, to enable our learning algorithm to find a unique solution with high probability.

Please note that, throughout this paper, a is assumed to be fixed as we develop a learning algorithm for sets of secret $GF(2)$-linear functions $f_1, \ldots, f_L : \{0,1\}^n \longrightarrow K$, where $K = GF\,(2^a)$. In particular, for the sake of simplicity, we write $f_1, \ldots, f_L$ while actually referring to a set of component functions as explained above.