The Learning with Rank Errors problem and an application to symmetric authentication

Slim Bettaieb*, Loïc Bidoux*, Yann Connan*†, Philippe Gaborit† and Adrien Hauteville†
*Worldline, ZI Rue de la pointe, 59113 Seclin, France
†University of Limoges, XLIM-DMI, 123, Av. Albert Thomas, 87060 Limoges, France

Abstract—**In this paper, we introduce a new hard problem opening up the construction for new quantum resistant cryptographic schemes. The latter is called Learning Rank with Errors (LRE) and can be seen as an adaptation of the LPN problem to the rank metric setting. In addition, we describe $\mathrm{HB}_{\mathrm{LRE}}$, an HB-like authentication protocol that constitutes an application of the aforementioned problem. We also prove that $\mathrm{HB}_{\mathrm{LRE}}$ is secure against passive attacks and compare its parameters to those of the initial HB scheme.**

I. Introduction

In light of the threat caused by several quantum algorithms to cryptography [1] [2], the NIST [3] has recently started to prepare its future transition to post-quantum cryptography. This confirms that identifying and studying quantum-resistant problems is of great importance for the cryptography community. Learning Parity with Noise (LPN) is one of these problems and has been shown to have many cryptographic applications in [4], [5]. This problem is equivalent to the problem of decoding random linear codes, does not succumb to known quantum algorithms attacks and has been proved NP-Hard [6]. LPN can be described as follows: let $\mathbf{s} \in \{0,1\}^k$ be a random secret of length k, find $\mathbf{s}$ given a number of samples $(\mathbf{a_i}\ ;\ (\mathbf{s}\cdot\mathbf{a_i}) \oplus \nu_i)$ with $\mathbf{a_i} \in \{0,1\}^k$ and $\nu_i \in \{0,1\}$ some noise following a Bernouilli distribution. The LPN problem is based on the Hamming metric, the historical metric used in code-based cryptography. In this paper, we are studying LPN along with the rank metric, an other metric used in code-based cryptography that has been less studied than the Hamming one but presents some promising properties.

In addition, we are interested by one of the applications of the LPN problem that has been studied extensively namely symmetric authentication protocols. The first LPN-based authentication protocol is due to Hopper and Blum [7]. The HB protocol is relatively simple and particularly well suited for lightweight cryptography. Suppose a prover P wants to authenticate to a verifier V with whom he shares a secret key $\mathbf{s} \in \{0,1\}^k$. The protocol is parametrized by a noise parameter $\eta \in [0, \frac{1}{2}]$, a number of instances $n \in \mathbb{N}$ and a threshold t such that $\eta n \leqslant t$. The verifier V generates $\mathbf{a} \in \{0;1\}^k$ and sends it to P. Then P generates $\nu \in \{0,1\}$ such that $\nu = 1$ with probability η, and sends $z = (\mathbf{s}\cdot\mathbf{a}) \oplus \nu$ to the prover. Next, the verifier V checks if z is equal to $\mathbf{s}\cdot\mathbf{a}$. This process is repeated n times and the verifier accepts the prover after n instances only if the number of unsuccessful iterations is at most t. Additional details can be found in [7], [8] and [9].

Contributions.

In this paper, we introduce Learning with Rank Errors (LRE), a hard problem based on the rank metric, resistant to known quantum attacks, that is analog to the LPN problem. Moreover, we also propose $\mathrm{HB}_{\mathrm{LRE}}$, an HB-like authentication protocol that relies on the hardness of the aforementioned LRE problem.

Paper organisation. In section II, we describe the LRE problem and discuss some of its properties (uniqueness of the solution, random self reducibility and pseudorandomness), then we give a description of the attacks on this problem with their complexity. In section III, we present the $\mathrm{HB}_{\mathrm{LRE}}$ protocol which adapts the HB protocol to the rank metric setting. Next, we discuss different relevant security models for $\mathrm{HB}_{\mathrm{LRE}}$ and prove that it is secure against passive attacks. We also compare the parameters of HB and $\mathrm{HB}_{\mathrm{LRE}}$ with respect to a given level of security. Finally, a concluding section sums up these results and suggests some perspectives.

II. The LRE problem

The Learning with Rank Errors problem is based on the rank metric and the use of $\mathbb{F}_{q^m}$ linearity and as such presents some interesting properties. On one hand, the complexity of pratical attacks is quadrically exponential in the size of parameters in the rank setting whereas it is simply exponential in the size of parameters in the Hamming case. On the other hand, the use of $\mathbb{F}_{q^m}$ linearity permits to construct protocols and schemes with reduced communication cost and key size.

A. Definitions

Definition 1 (Learning with Rank Errors (LRE) search problem)**.** *Let χ be an error distribution over $\mathbb{F}_{q^m}^l$. Let $\mathbf{s} \xleftarrow{\$} \mathbb{F}_{q^m}^n$ be a fixed secret vector. Let $\mathcal{O}_{\mathrm{LRE}}$ be an oracle which generates samples of the form $(\mathbf{A}, \mathbf{sA}+\mathbf{e})$ where $\mathbf{A} \xleftarrow{\$} \mathbb{F}_{q^m}^{n\times l}$ and $\mathbf{e} \xleftarrow{\$} \chi$. The $\mathrm{LRE}(q,m,n,l,\chi)$ search problem consists to find $\mathbf{s}$ given access to the oracle $\mathcal{O}_{\mathrm{LRE}}$.*

If $\mathbf{e}$ is chosen uniformly at random amongst the errors of weight at most r, we denote the problem $\mathrm{LRE}(q,m,n,l,r)$. Moreover, if $q=2$, we denote the problem $\mathrm{LRE}(m,n,l,r)$.

Definition 2 (Decision LRE (DLRE) problem)**.** *Let χ be an error distribution over $\mathbb{F}_{q^m}^l$ and $\mathbf{s} \xleftarrow{\$} \mathbb{F}_{q^m}^n$ be a fixed vector. Let $\mathcal{O}_0$ be an oracle which produces samples of* $\mathrm{LRE}(q,m,n,l,\chi)$ *and $\mathcal{O}_1$ be an oracle which produces samples of the form*

$(\mathbf{A}, \mathbf{u})$ *with* $\mathbf{u} \xleftarrow{\$} \mathbb{F}_{q^m}^l$. *Let* $b \xleftarrow{\$} \{0;1\}$ *and* $\mathcal{O} = \mathcal{O}_b$. *The* LRE$(q,m,n,l,\chi)$ *decisional problem consists in deciding whether* $b = 0$ *or* $b = 1$ *given only access to* $\mathcal{O}$.

B. Properties

In this subsection, we show that the LRE search problem has a unique solution given enough samples and is random self reducible. These two properties are straightforward. We assume that LRE is also pseudorandom and discuss on this assumption.

Proposition 1 (Uniqueness)**.** *The* $LRE(m,n,l,r)$ *search problem has a unique solution with a very high probability if the number* N *of samples is greater than* $\frac{mn}{(m-r)(l-r)}$.

Proof. The approach of this proof is the same as for the calculation of the Gilbert-Varshamov bound in the rank metric, only the form of the error is modified to correspond to the LRE search problem.

Let $\mathbf{A} = (\mathbf{A}_1|\dots|\mathbf{A}_N)$ and $\mathbf{e} = (\mathbf{e}_1|\dots|\mathbf{e}_N)$ be respectively the concatenation of the matrices and the errors of the N samples. $\mathbf{A}$ is a matrix of size $n \times Nl$ and $\mathbf{e}$ is a vector of length Nl. Since $N > \frac{mn}{(m-r)(l-r)}$, then $Nl > n$. So $\mathbf{A}$ is the generator matrix of an $[Nl, n]$ code over $\mathbb{F}_{q^m}$. Let $\boldsymbol{H}$ be a parity-check matrix of this code, i.e. a $(Nl-n) \times n$ full-rank matrix over $\mathbb{F}_{q^m}$ such that $\boldsymbol{A}\boldsymbol{H}^T = 0$. We have:

$$\boldsymbol{e}\boldsymbol{H}^T = (\boldsymbol{y}_1|\dots|\boldsymbol{y}_N)\boldsymbol{H}^T \in \mathbb{F}_{q^m}^{Nl-n}$$

where $\boldsymbol{y}_i = \boldsymbol{s}\boldsymbol{A}_i + \boldsymbol{e}_i$.
The vector $\boldsymbol{e}$ is the unique solution of the form $(\boldsymbol{e}_1|\dots|\boldsymbol{e}_N)$ of this equation with a very high probability if the number of errors of this form is inferior to $q^{m(Nl-n)}$. This number is equal to $\left(\prod_{i=0}^{r-1} \frac{q^m-q^i)(q^l-q^i)}{q^r-q^i}\right)^N = \Theta\left(q^{Nr(m+l-r)}\right)$. Hence

$$\begin{aligned} & q^{Nr(m+l-r)} && < q^{m(Nl-n)} \\ \Leftrightarrow\ & Nr(m+l-r) && < m(Nl-n) \\ \Leftrightarrow\ & mn && < N(ml-rm-r(l-r)) \\ \Leftrightarrow\ & \frac{mn}{(m-r)(l-r)} && < N \end{aligned}$$

□

Proposition 2 (Random Self Reducibility)**.** *Assume there is a PPT algorithm* $\mathcal{A}$ *that solves the LRE(q,m,n,l,*χ*) search problem with a non negligible probability. Then there exists a PPT algorithm* $\mathcal{A}'$ *which find* $\boldsymbol{s} \in \mathbb{F}_{q^m}^n$ *with a non negligible probability given only access to samples of the form* $(\boldsymbol{A}, \boldsymbol{s}\boldsymbol{A} + \boldsymbol{e}), \boldsymbol{A} \xleftarrow{\$} \mathbb{F}_{q^m}^{n\times l}, \boldsymbol{e} \xleftarrow{\$} \chi$ *for any* $\boldsymbol{s} \in \mathbb{F}_{q^m}^n$.

Proof. Let $\mathcal{A}$ be a PPT algorithm $\mathcal{A}$ that solves the LRE(q,m,n,l,χ) search problem with a non negligible probability p. Let $\boldsymbol{s}' \xleftarrow{\$} \mathbb{F}_{q^m}^n$, $\boldsymbol{s}'' = \boldsymbol{s} + \boldsymbol{s}'$ and $\boldsymbol{y} = \boldsymbol{s}\boldsymbol{A} + \boldsymbol{e} + \boldsymbol{s}'\boldsymbol{A} = \boldsymbol{s}''\boldsymbol{A} + \boldsymbol{e}$. Since $\boldsymbol{s}'$ is distributed uniformly at random, $\boldsymbol{s}''$ is also distributed uniformly at random. Thus, given access to the samples $(\boldsymbol{A}, \boldsymbol{y})$, the algorithm $\mathcal{A}$ outputs $\boldsymbol{s}''$ with probability p. If it is the case, we compute $\boldsymbol{s}'' - \boldsymbol{s}'$ to retrieve $\boldsymbol{s}$ with probability p. □

Unfortunately, there is no known search-to-decision reduction for the LRE problem.

Assumption 1 (Pseudorandomness)**.** *DLRE(q,m,n,l,r) is pseudorandom, that-is-to-say there is no PPT distinguisher with a non negligible advantage between the distribution* $(\boldsymbol{A}, \boldsymbol{s}\boldsymbol{A}+\boldsymbol{e})$ *and* $(\boldsymbol{A}, \boldsymbol{u})$ *where* $\boldsymbol{A}$ *is a random matrix of size* $n \times l$ *over* $\mathbb{F}_{q^m}$, $\boldsymbol{e}$ *a random vector of length* l *and weight* r *and* $\boldsymbol{u}$ *a random vector of length* l.

Discussion on this assumption: Even if there is no reduction, the DLRE problem is linked to the decisional Exact-LPN, we can transform samples of the Exact-LPN problem into samples of a very close distribution of LRE. Let us recall this problem: given the distributions $\mathcal{D}_{ELPN} = (\boldsymbol{A}, \boldsymbol{s}\boldsymbol{A} + \boldsymbol{e})$ and $\mathcal{D}_U = (\boldsymbol{A}, \boldsymbol{u})$ where $\boldsymbol{A} \xleftarrow{\$} \mathbb{F}_2^{n\times l}$, $\boldsymbol{e}$ is a random vector of $\mathbb{F}_2$ of Hamming weight of r and $\boldsymbol{u} \xleftarrow{\$} \mathbb{F}_2^l$, is it possible to distinguish these two distributions with a non negligible advantage. We use a transformation inspired by [10] to embed these distributions into the rank metric. Let $m > 8l$ and $\alpha_1, \dots, \alpha_l$ l $\mathbb{F}_2$-independent elements of $\mathbb{F}_{2^m}$.

$$\text{Let } \psi : \begin{array}{ccc} \mathbb{F}_2^l & \to & \mathbb{F}_{2^m}^l \\ (x_1,\dots,x_l) & \mapsto & (\alpha_1 x_1, \dots, \alpha_l x_l) \end{array}$$

ψ transforms a vector of $\mathbb{F}_2$ of Hamming weight r into a vector of $\mathbb{F}_{2^m}$ of rank weight r. ψ can be extended to the matrices of $\mathbb{F}_2^{n\times l}$. Let $\boldsymbol{M}$ be an $l \times l$ matrix over $\mathbb{F}_{2^m}$ whose coefficients generate a $\mathbb{F}_2$-subspace of $\mathbb{F}_{2^m}$ of dimension 2. The distribution of samples of the form $(\psi(\boldsymbol{A})\boldsymbol{M}, \psi(\boldsymbol{s}\boldsymbol{A} + \boldsymbol{e})\boldsymbol{M})$ is close to the distribution of $LRE(m,n,l,2r)$. Indeed, we have $\psi(\boldsymbol{s}\boldsymbol{A} + \boldsymbol{e})\boldsymbol{M} = \boldsymbol{s}\psi(\boldsymbol{A})\boldsymbol{M} + \psi(\boldsymbol{e})\boldsymbol{M}$ and $\psi(\boldsymbol{e})\boldsymbol{M}$ is a vector of weight $2r$ with very high probability. However the matrix $\psi(\boldsymbol{A})\boldsymbol{M}$ is not random.

C. Attacks on the LRE problem

In this subsection we present the different attacks on the LRE problems. There are two families of attacks:

- the algebraic attacks consists in solving a multivariate system whose unknowns are $\boldsymbol{s}$ and the vectors $\boldsymbol{e}_i$ with Groebner bases. Their advantage is that they ignore the size of the field q, that's why they are generally more efficient when $q > 2$. However their complexity (in time and in space) depends on the number of unknowns of the system. In our case, each sample $(\boldsymbol{A}, \boldsymbol{s}\boldsymbol{A} + \boldsymbol{e})$ adds the coordinates of $\boldsymbol{e}$ as unknowns. Thus these attacks cannot be used to solve efficiently the LRE problem.
- the combinatorial attacks try to guess some coordinates of the errors $\boldsymbol{e}_i$ or the subspace E_i generated by the coordinates of $\boldsymbol{e}_i$ in order to solve a linear system. They are the adaptation of decoding algorithms of rank metric linear codes. Since these algorithms need to guess some information about the errors, they are better when q is small. They can efficiently use the multiplicity of the samples, that is why they are the best known attacks against the LRE problem.

The first attack is an adaptation of the GRS algorithm [11]. Let $(\boldsymbol{A}_i, \boldsymbol{s}\boldsymbol{A}_i + \boldsymbol{e}_i)_{1\le i\le N}$ be the N samples of the LRE problem. Let $E_1, \dots, E_N$ be the supports of $\boldsymbol{e}_1, \dots, \boldsymbol{e}_N$, i.e. the $\mathbb{F}_q$-subspaces generated by the coordinates of $\boldsymbol{e}_i$. The idea is to find N $\mathbb{F}_q$-subspaces $F_1, \dots, F_N$ of dimension r' such that $E_i \subset F_i$ for all $1 \leq i \leq N$. In this case, we can express the coordinates e_{ij} of $\boldsymbol{e}_i$ in a basis of F_i, which gives us Nlr' unknowns over $\mathbb{F}_q$. If $Nl > n$ let $\boldsymbol{H}$ be a parity-check matrix of the code generated by $\mathbf{A} = (\mathbf{A}_1 | \dots | \mathbf{A}_N)$. We have

$$\boldsymbol{e}\boldsymbol{H}^T = (\boldsymbol{y}_1 | \dots | \boldsymbol{y}_N)\boldsymbol{H}^T \in \mathbb{F}_{q^m}^{Nl-n}$$

where $\boldsymbol{y}_i = \boldsymbol{s}\boldsymbol{A}_i + \boldsymbol{e}_i$, which gives us $m(Nl-n)$ equations over $\mathbb{F}_q$. These equations have only one solution with non negligible probability if $Nlr' \leq m(Nl-n)$. The probability that the F_i's have been correctly chosen is $\left(\begin{bmatrix} r' \\ r \end{bmatrix}_q \Big/ \begin{bmatrix} m \\ r \end{bmatrix}_q\right)^N = \mathcal{O}\big(q^{Nr(m-r')}\big)$. By taking $r' = m - \left\lceil \frac{mn}{Nl} \right\rceil$, we obtain a final complexity of $\mathcal{O}\big(m^3(Nl-n)^3 q^{Nr\left\lceil \frac{mn}{Nl} \right\rceil}\big)$.

We can improve the complexity of this attack by using the algebraic structure over $\mathbb{F}_{q^m}$. Let consider the matrix $\boldsymbol{A}' = \left(\begin{array}{c} \boldsymbol{A} \\ \hline \boldsymbol{y} \end{array}\right)$ which generates an $[Nl, n+1]$ $\mathbb{F}_{q^m}$ linear code $\mathcal{C}'$. By linearity, the vector $\boldsymbol{e}$ belongs to $\mathcal{C}'$ and also all multiples $\alpha\boldsymbol{e}, \alpha \in \mathbb{F}_{q^m}$ since $\alpha(-\boldsymbol{s}|1)\boldsymbol{A}' = \alpha(-\boldsymbol{s}\boldsymbol{A} + \boldsymbol{y}) = \alpha\boldsymbol{e}$. Let $\boldsymbol{H}'$ be a parity-check matrix of $\mathcal{C}'$. Each multiples of $\boldsymbol{e}$ is a solution of $\boldsymbol{H}'\boldsymbol{x}^T = 0$ of unknown $\boldsymbol{x}$. The idea is the same as before, we want to find N $\mathbb{F}_q$-subspaces $F_1, \dots, F_N$ of dimension r' to express the coordinates of $\boldsymbol{x}$ in a basis of F_i, block by block. Since we search for any multiple of $\boldsymbol{e}$, the probability to choose correctly the F_i is higher than previously, but due to the lack of space we cannot compute it here. Then if we know a multiple $\alpha\boldsymbol{e}$, we can compute a multiple $\alpha\boldsymbol{s}$ by solving the linear system $(\alpha\boldsymbol{s}| - \alpha)\boldsymbol{A}' = -\alpha\boldsymbol{e}$ and then retrieve $\boldsymbol{s}$.

In average, the complexity of this attack is

$$C_1 = \mathcal{O}\big(m^3(Nl-n)^3 q^{Nr\left\lceil \frac{m(n+1)}{Nl} \right\rceil - m}\big)$$

operations in $\mathbb{F}_q$.

The idea of the second algorithm is to try and obtain a null error on some coordinates of $\boldsymbol{y}$ by linear combinations over $\mathbb{F}_q$. Indeed, let $\boldsymbol{y}_i = \boldsymbol{s}\boldsymbol{A}_i + \boldsymbol{e}_i$ be a sample of LRE. There exists a matrix $\boldsymbol{M}_i \in \mathbb{F}_q^{l \times l}$ such that $\boldsymbol{e}_i\boldsymbol{M}_i = (e'_1, \dots, e'_r, 0, \dots 0)$ where $e'_1, \dots e'_r$ generate the support of $\boldsymbol{e}$. Now let us consider the vector $\boldsymbol{y}_i\boldsymbol{M}_i = \boldsymbol{s}\boldsymbol{A}_i\boldsymbol{M}_i + \boldsymbol{e}_i\boldsymbol{M}_i$. We can split the vector $\boldsymbol{y}_i\boldsymbol{M}_i$ into two vectors of length r and $l-r$ and the matrix $\boldsymbol{A}_i\boldsymbol{M}_i$ into two matrices of size $n \times r$ and $n \times (l-r)$ such that $\boldsymbol{y}_i\boldsymbol{M}_i = (\boldsymbol{y}'_{i1}|\boldsymbol{y}'_{i2})$ and $\boldsymbol{A}_i\boldsymbol{M}_i = (\boldsymbol{A}'_{i1}|\boldsymbol{A}'_{i2})$. Since the $l-r$ last coordinates of $\boldsymbol{e}_i\boldsymbol{M}_i$ are null, we have $\boldsymbol{y}'_{i2} = \boldsymbol{s}\boldsymbol{A}'_{i2}$. This gives us $(l-r)$ equations of $\boldsymbol{s}$ without errors. If we have N samples such that $N(l-r) \geq n$, we can get enough equations to compute $\boldsymbol{s}$ by inverting an $n \times n$ matrix over $\mathbb{F}_{q^m}$.

Let us now describe the attack. For each vector $\boldsymbol{y}_i$, we choose a random invertible matrix $\boldsymbol{M}_i \in \mathbb{F}_q^{l \times l}$ and compute $\boldsymbol{y}_i\boldsymbol{M}_i$. Then we suppose that we obtain an error $\boldsymbol{e}_i\boldsymbol{M}_i$ with $l-r$ null coordinates. The probability to obtain one null coordinate is equal to q^{-r} for each coordinate of $\boldsymbol{e}_i\boldsymbol{M}_i$ is a random element of the support of $\boldsymbol{e}_i$. Since we need n null coordinates, the probability to choose correctly the matrices $\boldsymbol{M}_i$ is q^{-nr}. The cost of the linear algebra is dominated by the inversion of an $n \times n$ matrix over $\mathbb{F}_{q^m}$, hence this attack has a complexity of

$$C_2 = \mathcal{O}\big(n^3 m \log m \log\log m q^{nr}\big)$$

operations in $\mathbb{F}_q$.

Asymptotically, in the typical case $q = 2$ and $l = m = n$, we have

$$\log_2 C_1 = \tilde{\mathcal{O}}\big(rn + r - n\big)$$
$$\log_2 C_2 = \tilde{\mathcal{O}}\big(rn\big)$$

III. HB-like protocol based on LRE

HB [7] is a authentication protocol based on LPN. As a symmetric scheme based on a hard problem that is very practical for resource-constraints devices, HB have received extensive attention. In this section, we introduce $\mathrm{HB}_{\mathrm{LRE}}$, an HB variant that relies on the LRE problem.

A. The $\mathrm{HB}_{\mathrm{LRE}}$ protocol

Following [7], a symmetric key authentication protocol is defined by a pair of algorithms (P, V) sharing a secret s. The interaction between the prover P and the verifier V on inputs x and y is denoted as $\langle P(x), V(y) \rangle$. Given such an interaction, the verifier should accept the prover (denoted $\langle P, V \rangle$ = accept) whenever $x = y$ and reject it otherwise.

Definition 3. *A symmetric authentication protocol is a pair of probabilistic polynomial-time algorithms (P, V) sharing a secret s such that (i) for all inputs s, $\langle P(s), V(s) \rangle$ =* accept *with probability 1 and (ii) for each pair $x \neq y$, $\langle P(x), V(y) \rangle$ =* accept *with negligible probability.*

In the remaining of this document, $\omega(\mathbf{v})$ denotes the rank weight of a vector $\mathbf{v} \in \mathbb{F}_{q^m}^n$. By considering $\mathbb{F}_{q^m}$ as a vector space of dimension m over $\mathbb{F}_q$, one can represent the vector $\mathbf{v}$ as a matrix $\mathbf{E}(\mathbf{v}) \in \mathbb{F}_q^{m \times n}$. We refer the reader to [12] for additional details on the rank metric. The rank weight of $\mathbf{v}$ is defined as the rank of the aforementioned matrix $\mathbf{E}(\mathbf{v})$ namely $\omega(\mathbf{v}) = \mathsf{rank}(\mathbf{E}(\mathbf{v}))$. Furthermore, the set of words of weight at most r in $\mathbb{F}_{q^m}^n$ is denoted by $\mathcal{B}_r^n(\mathbb{F}_{q^m}) = \{\mathbf{v} \in \mathbb{F}_{q^m}^n \mid \omega(\mathbf{v}) \leq r\}$. Elements of $\mathcal{B}_r^n(\mathbb{F}_{q^m})$ are generated by randomly choosing r elements in $\mathbb{F}_{q^m}$ and constructing n random linear combinations of these r elements.

We now introduce the $\mathrm{HB}_{\mathrm{LRE}}$ protocol (see figure 1) which is an HB-like protocol based on the LRE problem rather than the LPN one. This protocol constitutes one application of the LRE problem and illustrates how the latter can be used in order to construct cryptographic schemes. Interestingly, one should note that contrary to the initial HB protocol, $\mathrm{HB}_{\mathrm{LRE}}$ does not suffer from false negative issues.

Many variants of the initial HB protocol have been proposed: see [8], [13], [14] for instance and refer to [15] for a survey on LPN-based HB protocols. Each of these scheme increases the security or the efficiency of HB in various ways. Most of these improvements could also be adapted to the LRE setting. We defer such extensions of $\mathrm{HB}_{\mathrm{LRE}}$ to future work.

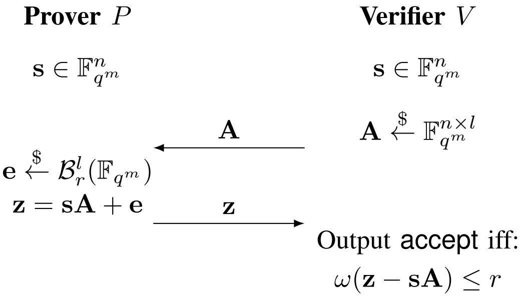
Prover P
Verifier V
$\mathbf{s} \in \mathbb{F}_{q^m}^n$
$\mathbf{s} \in \mathbb{F}_{q^m}^n$
$\mathbf{A}$
$\mathbf{A} \xleftarrow{\$} \mathbb{F}_{q^m}^{n \times l}$
$\mathbf{e} \xleftarrow{\$} \mathcal{B}_r^l(\mathbb{F}_{q^m})$
$\mathbf{z} = \mathbf{s}\mathbf{A} + \mathbf{e}$
$\mathbf{z}$
Output accept iff:
$\omega(\mathbf{z} - \mathbf{s}\mathbf{A}) \leq r$

Fig. 1. The HB_{LRE} protocol

B. Security models

The main security notion for symmetric authentication protocols is security against impersonation. This notion is defined using three security models: passive, active and man-in-the-middle [16]. Those models can be formalized as a security game involving two major steps: a learning phase and an impersonation phase. Depending on the security model, the adversary is given access to additional oracles, and thus is given more power.

Within the passive attack model, the adversary is only allowed to obtain transcripts of exchanges between the prover and the verifier during the learning step. In the second step, the adversary tries to impersonate a prover interacting with a honest verifier. The active attack model is a stronger model, in which the attacker is given access to transcripts as in the passive model but can also interact with a honest prover only (in the detection model) or with both honest prover and verifier (in the prevention model). In the impersonation phase, the adversary only interacts with the verifier. The man-in-the-middle attack model is the strongest security model. In fact, the attacker is given access to protocol transcripts and is allowed to interact with the prover and the verifier during both the learning phase and the impersonation phase.

We prove the security of HB_{LRE} in the passive attack model. Informally, if an adversary is able to break the security of our scheme in this model, then he can be used to solve the LRE problem. Following [9], we give a formal description of the security game of the passive attack model:

- **Setup:** Generate a shared secret key ($\mathbf{s} \xleftarrow{\$} \mathbb{F}_{q^m}^n$) and give it to the prover P and the verifier V.
- **Learning step:** The adversary $\mathcal{A}$ is given access to a polynomial number of transcripts of protocol executions. We denote these transcripts by $\mathsf{trans}(\langle P, V \rangle)$.
- **Impersonation step:** The adversary $\mathcal{A}$ plays the role of the prover and interacts with the verifier V. We denote by $\langle \mathcal{A}, V \rangle = \mathsf{accept}$ an execution of the protocol in which the verifier V authenticates the adversary $\mathcal{A}$.

We say that the adversary wins the game if the verifier V accepts $\mathcal{A}$ at the end of the interaction.

Definition 4 (Passive Attacks)**.** *An authentication protocol is secure against passive attacks if for all efficient adversaries $\mathcal{A}$, the advantage $Adv_{\mathcal{A},\text{HB}_{\text{LRE}}}^{passive}$ is negligible where:*

$$Adv_{\mathcal{A},\text{HB}_{\text{LRE}}}^{passive} = \mathsf{Pr}\left[\langle \mathcal{A}, V \rangle = \mathsf{accept} \,|\, \mathbf{s} \xleftarrow{\$} \mathbb{F}_{q^m}^n; \mathsf{trans}(\langle P, V \rangle)\right]$$

C. Security against passive attacks

Let $A_{s,r} = \{(\mathbf{A}, \mathbf{s}\mathbf{A} + \mathbf{e}) \mid \mathbf{A} \xleftarrow{\$} \mathbb{F}_{q^m}^{l \times n}; \mathbf{e} \xleftarrow{\$} \mathcal{B}_r^l(\mathbb{F}_{q^m})\}$ denotes the set of all LRE instances and U denotes the uniform distribution over $\mathbb{F}_{q^m}^{l \times n} \times \mathbb{F}_{q^m}^l$.

Theorem 1. *If there exists an adversary $\mathcal{A}$ eavesdropping on at most k executions of the* HB_{LRE} *protocol, running in time t and achieving $Adv_{\mathcal{A},\text{HB}_{\text{LRE}}}^{passive} = \delta$, then there exists an algorithm D making $(k+1)$ oracle queries, running in $\mathcal{O}(t)$ such that:*

$$\left|\mathsf{Pr}[D^{A_{s,r}}(1^n) = 1 \mid \mathbf{s} \xleftarrow{\$} \mathbb{F}_{q^m}^n] - \mathsf{Pr}\left[D^U(1^n) = 1\right]\right| \geq \delta - p$$

with $p = \left(\sum_{i=0}^{2r} \prod_{j=0}^{i-1}(q^l - q^j)\begin{bmatrix} m \\ i \end{bmatrix}_q\right) / q^{ml}$ where $\begin{bmatrix} m \\ i \end{bmatrix}_q$ denotes the Gaussian binomial coefficient i.e. the number of subspaces of dimension i in a vector space of dimension m over a finite field with q elements.

Following the proof ideas of theorem 2 from [9], we show how to construct an efficient algorithm D that can distinguish between instances of $A_{s,r}$ and U using the adversary $\mathcal{A}$.

Proof. D is given access to an oracle returning $(\mathbf{A}, \mathbf{z})$ and proceeds as follows:

1) D runs the learning phase of $\mathcal{A}$. Each time $\mathcal{A}$ requests to view a transcript of the protocol, D obtains $(\mathbf{A}, \mathbf{z})$ from its oracle and returns it to $\mathcal{A}$.
2) When $\mathcal{A}$ is ready for the impersonation phase, D obtains a sample $(\bar{\mathbf{A}}, \bar{\mathbf{z}})$ from its oracle. D sends the challenge $\bar{\mathbf{A}}$ to $\mathcal{A}$ and receives in return a response $\mathbf{z}'$.
3) D outputs 1 if and only if $\omega(\bar{\mathbf{z}} - \mathbf{z}') \leq 2r$.

Two cases have to be considered depending on which oracle D is interacting with:

- **If** D **is interacting with** U: The probability that D outputs 1 is the same as the probability that a randomly chosen vector of length l ends up being inside the ball of radius $2r$ centered in $\mathbf{0}$ (see [12]):

  $$p = \frac{|\mathcal{B}_{q^m}^l(\mathbf{0}, 2r)|}{|\mathbb{F}_{q^m}^l|} = \frac{\sum_{i=0}^{2r}\left(\prod_{j=0}^{i-1}(q^l - q^j)\begin{bmatrix} m \\ i \end{bmatrix}_q\right)}{q^{ml}}$$

  This probability is negligible as long as the parameters q, m, r, l and n are chosen correctly.
- **If** D **is interacting with** $A_{s,r}$: Let $\mathbf{z}^* = \mathbf{s}\mathbf{A}$ denotes the response of a prover without any error. We have $\mathsf{Pr}[\omega(\bar{\mathbf{z}} - \mathbf{z}^*) \leq r] = 1$ since $\bar{\mathbf{z}}$ is distributed exactly as the answer of an honest prover and $\mathsf{Pr}[\omega(\mathbf{z}' - \mathbf{z}^*) \leq r] = \delta$ since $\mathcal{A}$ successfully impersonates the prover in those cases. It follows that D outputs 1 with probability at least δ since $\mathsf{Pr}[\omega(\bar{\mathbf{z}} - \mathbf{z}') \leq 2r] \geq \mathsf{Pr}[\omega(\bar{\mathbf{z}} - \mathbf{z}^*) \leq r \,\cap\, \omega(\mathbf{z}^* - \mathbf{z}') \leq r)] = \mathsf{Pr}[(\omega(\mathbf{z}^* - \mathbf{z}') \leq r)] = \delta$.

Thus, D is able to distinguish between instances from $A_{s,r}$ and U with probability at least $\delta - p$. □

In the case of $l = m = n$, we have $p \approx q^{2r(2n-2r)-n^2} = q^{4r(n-r)-n^2}$.

D. Parameters comparison between HB and $\mathrm{HB}_{\mathrm{LRE}}$

In order to achieve a security level of 80 bits, previous works (see [17]–[22]) suggest to use a $(592, 1/8)$ LPN instance, namely a 592 bits long secret key along with an error rate of $1/8$. One should nonetheless note that larger instances might need to be considered according to recent work [23]. Given an error rate of $1/8$, if one want completeness errors (respectively soundness errors) to occur with probability less than 2^{-40} (respectively 2^{-80}), it should instanciate the HB protocol with 441 rounds [17]. Therefore, in the initial HB protocol, 32 Kio have to be exchanged between the prover and the verifier to achieve a security level of 80 bits. One has to consider a $(18, 18, 18, 6)$ LRE instance in order to achieve the same security level. These parameters correspond to 324 bits long secret keys and require only 648 bits to be exchanged during an execution of the protocol. For 128 bits of security, one may use a $(20, 20, 20, 7)$ LRE instance which correspond to 400 bits long secret keys and 800 bits of exchanged data.

IV. Conclusion

In this paper we have introduced LRE, a problem similar to the LPN problem that relies on the rank metric as well as its decisional version. We have studied some of its properties namely the uniqueness of its solution and its self-reducibility. We have discussed on its pseudorandomness and given some arguments in its favor. In order to provide a first application to the aforementioned problem, we have also described the $\mathrm{HB}_{\mathrm{LRE}}$ protocol, an HB-like protocol based on the LRE problem. A natural perspective of this work would be to design and study other cryptographic schemes relying on the difficulty of LRE. Furthermore, our initial $\mathrm{HB}_{\mathrm{LRE}}$ protocol could be extended following the body of work realized on the initial HB protocol (see [8], [13]–[15]) in order to improve both its security and its performances. $\mathrm{HB}_{\mathrm{LRE}}$ provides better parameters than HB but is less computationally efficient, therefore it would also be interesting to compare it with other symmetric authentication protocols.

References

[1] P. W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM review*, vol. 41, no. 2, pp. 303–332, 1999.

[2] L. K. Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. ACM, 1996, pp. 212–219.

[3] NIST, “Submission requirements and evaluation criteria for the post-quantum cryptography standardization process,” 2016.

[4] A. Blum, M. Furst, M. Kearns, and R. J. Lipton, “Cryptographic primitives based on hard learning problems,” in *Annual International Cryptology Conference*. Springer, 1993, pp. 278–291.

[5] K. Pietrzak, “Cryptography from learning parity with noise.” in *SOFSEM*, vol. 7147. Springer, 2012, pp. 99–114.

[6] E. Berlekamp, R. McEliece, and H. Van Tilborg, “On the inherent intractability of certain coding problems (corresp.),” *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 384–386, 1978.

[7] N. J. Hopper and M. Blum, “Secure human identification protocols,” in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2001, pp. 52–66.

[8] A. Juels, S. A. Weis *et al.*, “Authenticating pervasive devices with human protocols,” in *Crypto*, vol. 3621. Springer, 2005, pp. 293–308.

[9] J. Katz, J. S. Shin, and A. Smith, “Parallel and concurrent security of the HB and HB+ protocols,” *Journal of cryptology*, vol. 23, no. 3, pp. 402–421, 2010.

[10] P. Gaborit and G. Zémor, “On the hardness of the decoding and the minimum distance problems for rank codes,” *IEEE Trans. Information Theory*, vol. 62(12), pp. 7245–7252, 2016.

[11] P. Gaborit, O. Ruatta, and J. Schrek, “On the complexity of the rank syndrome decoding problem,” *IEEE Trans. Information Theory*, vol. 62, no. 2, pp. 1006–1019, 2016.

[12] P. Loidreau, “Properties of codes in rank metric,” *arXiv preprint cs/0610057*, 2006.

[13] H. Gilbert, M. J. Robshaw, and Y. Seurin, “HB#: Increasing the security and efficiency of HB+,” *Lecture Notes in Computer Science*, vol. 4965, pp. 361–378, 2008.

[14] E. Kiltz, K. Pietrzak, D. Venturi, D. Cash, and A. Jain, “Efficient authentication from hard learning problems,” *Journal of Cryptology*, vol. 30, no. 4, pp. 1238–1275, 2017.

[15] A. Karrothu, R. Scholar, and J. Norman, “An analysis of LPN based HB protocols,” in *Advanced Computing (ICoAC), 2016 Eighth International Conference on*. IEEE, 2017, pp. 138–145.

[16] V. Lyubashevsky and D. Masny, “Man-in-the-middle secure authentication schemes from LPN and weak PRFs,” in *Advances in Cryptology–CRYPTO 2013*. Springer, 2013, pp. 308–325.

[17] É. Levieil and P.-A. Fouque, “An improved LPN algorithm,” in *SCN*, vol. 4116. Springer, 2006, pp. 348–359.

[18] D. J. Bernstein and T. Lange, “Never trust a bunny.” *RFIDSec*, vol. 7739, pp. 137–148, 2012.

[19] Q. Guo, T. Johansson, and C. Löndahl, “Solving LPN using covering codes,” in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2014, pp. 1–20.

[20] B. Zhang, L. Jiao, and M. Wang, “Faster algorithms for solving LPN,” in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2016, pp. 168–195.

[21] S. M. Bogos and S. Vaudenay, “Observations on the LPN solving algorithm from eurocrypt" 16,” Tech. Rep., 2016.

[22] B. Zhang and X. Gong, “New algorithms for solving LPN,” Cryptology ePrint Archive, Report 2017/780, 2017, https://eprint.iacr.org/2017/780.

[23] A. Esser, R. Kübler, and A. May, “LPN decoded.” *IACR Cryptology ePrint Archive*, vol. 2017, p. 78, 2017.