Fuzzy Authentication using Rank Distance

Alessandro Neri, Joachim Rosenthal and Davide Schipani*

March 9, 2017

Abstract

Fuzzy authentication allows authentication based on the fuzzy matching of two objects, for example based on the similarity of two strings in the Hamming metric, or on the similiarity of two sets in the set difference metric. Aim of this paper is to show other models and algorithms of secure fuzzy authentication, which can be performed using the rank metric. A few schemes are presented which can then be applied in different scenarios and applications.

1 Introduction

Recent years have seen a lot of research around the problem of authentication using approximate matching under a certain metric of similarity, while still enabling a secure storage of sensible authentication data. The typical, but not the only scenario, where such a system is needed, is in the use of biometric features, like fingerprints, for authentication purposes.

Several models have been proposed that may be more appropriate for different applications. For example the fuzzy commitment scheme [8] models data as bit strings and compares strings in the Hamming metric; the fuzzy vault [7] models data as sets of elements and compares sets in the set difference metric.

In this paper we present fuzzy authentication schemes using the rank metric by generalizing the schemes mentioned above for other model scenarios and highlighting possible applications. The structure of the paper is the following. Section 2 recalls some mathematical concepts and definitions concerning rank metric codes and linearized polynomials. Section 3 presents the fuzzy commitment scheme in the rank distance, a model whereby the tolerance needed in the authentication is not based on the number of different bits between two strings but on the similarity of two matrices, more precisely on the rank of their difference. Section 4 is devoted to a fuzzy vault scheme using linearized polynomials, which relates the set difference with the rank metric. The scheme is an alternative to the standard fuzzy vault based on Reed-Solomon decoding. Section 5 gives hints on possible applications and model scenarios of the schemes presented in the previous sections.

*Institute of Mathematics, University of Zurich, Switzerland

2 Rank Metric Codes and Linearized Polynomials

Let q be a prime power and let $\mathbb{F}_q$ denote the finite field with q elements. Recall that $\mathbb{F}_{q^m}$ is isomorphic (as a vector space over $\mathbb{F}_q$) to the vector space $\mathbb{F}_q^m$. One then easily obtains the isomorphic description of matrices over the base field $\mathbb{F}_q$ as vectors over the extension field, i.e. $\mathbb{F}_q^{m\times n} \cong \mathbb{F}_{q^m}^n$.

Definition 1. *The* rank distance d_R *on* $\mathbb{F}_q^{m\times n}$ *is defined by*

$$d_R(X,Y) := \mathrm{rk}(X-Y), \quad X,Y \in \mathbb{F}_q^{m\times n}.$$

In the same way it is possible to define the rank distance between two elements $\boldsymbol{x},\boldsymbol{y} \in \mathbb{F}_{q^m}^n$ *as the rank of the difference of the respective matrix representations in* $\mathbb{F}_q^{m\times n}$.

A *rank metric code* $\mathcal{C}$ is a subset of $\mathbb{F}_q^{m\times n}$ (or $\mathbb{F}_{q^m}^n$) equipped with the rank distance. The *minimum distance* of a rank metric code $\mathcal{C}$ is the quantity

$$d_R(\mathcal{C}) := \min\left\{d_R(u,v) \mid u,v \in \mathcal{C}, u \neq v\right\}.$$

We can define special classes of rank metric codes introducing linearity. An $\mathbb{F}_{q^m}$-linear rank metric code of dimension k is a rank metric code that is also a k-dimensional subspace of the $\mathbb{F}_{q^m}$-vector space $\mathbb{F}_{q^m}^n$. An $\mathbb{F}_q$-linear rank metric code of dimension k' is a rank metric code that is also a k'-dimensional subspace of the $\mathbb{F}_q$-vector space $\mathbb{F}_{q^m}^n \cong \mathbb{F}_q^{m\times n}$.

Observe that an $\mathbb{F}_{q^m}$-linear rank metric code of dimension k is also an $\mathbb{F}_q$-linear code of dimension mk.

We will use the notation $[n,k,d]$-code for a k-dimensional $\mathbb{F}_{q^m}$-linear code with minimum distance d, and $[nm,k',d']$-code for a k'-dimensional $\mathbb{F}_q$-linear code with minimum distance d'.

Theorem 1 (Singleton-like Bound)**.** *Let* $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ *be a rank metric code. Then*

$$|\mathcal{C}| \leq \min\left\{q^{m(n-d+1)}, q^{n(m-d+1)}\right\}.$$

Proof. See [5], or [14, Theorem 1]. □

Definition 2. *Codes attaining the Singleton-like bound are called* Maximum Rank Distance (MRD) Codes.

When $n \leq m$ a class of codes attaining the Singleton-like bound was first proposed in [5] and then generalized in [9]. These codes are $\mathbb{F}_{q^m}$-linear rank metric codes. Let $(v_1,\ldots,v_n) \in \mathbb{F}_{q^m}^n$ be a vector, we denote the $k \times n$ *s-Moore matrix* by

$$M_{s,k}(v_1,\ldots,v_n) := \begin{pmatrix} v_1 & v_2 & \cdots & v_n \\ v_1^{[s]} & v_2^{[s]} & \cdots & v_n^{[s]} \\ \vdots & & & \vdots \\ v_1^{[s(k-1)]} & v_2^{[s(k-1)]} & \cdots & v_n^{[s(k-1)]} \end{pmatrix},$$

where $[i] := q^i$.

Definition 3. *Let $g_1, \ldots, g_n \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$ and let s be coprime to m. We define a* generalized Gabidulin code *$\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ of dimension k as the linear block code with generator matrix $M_{s,k}(g_1, \ldots, g_n)$. Using the isomorphic matrix representation we can interpret $\mathcal{C}$ as a matrix code in $\mathbb{F}_q^{m \times n}$.*

These codes are optimum for rank distance, since they are $[n, k, n-k+1]$-codes. Moreover, for this class of codes there exist polynomial-time decoding algorithms decoding up to their error-correcting capability $t = \left\lfloor \frac{n-k}{2} \right\rfloor$, [10, 13, 17].

Observe that when $s = 1$, this definition of Gabidulin codes is the q-analog of Reed-Solomon codes with the Hamming distance. Here, a set of distinct elements is replaced by a set of linearly independent elements, and the power g_i^j is replaced by the Frobenius power $g_i^{[j]}$. Reed-Solomon codes can also be seen as evaluation of polynomials of degree less than k in n distinct points. We can give a q-analog of this interpretation for Gabidulin codes, as evaluation of linearized polynomials in n linearly independent elements.

Definition 4. *A* linearized polynomial *over $\mathbb{F}_{q^m}$ is a polynomial $f(x) \in \mathbb{F}_{q^m}[x]/(x^{q^m} - x)$ of the form*

$$\sum_{i=0}^{m-1} f_i x^{[i]}.$$

We denote by $\mathcal{L}_m(\mathbb{F}_{q^m})$ the space of linearized polynomials over $\mathbb{F}_{q^m}$.

Let $\mathcal{G}_{k,s} \subseteq \mathcal{L}_m(\mathbb{F}_{q^m})$ be the set defined as

$$\mathcal{G}_{k,s} := \left\{ f_0 x + f_1 x^{[s]} + \ldots + f_{k-1} x^{[s(k-1)]} \mid f_i \in \mathbb{F}_{q^m} \right\}.$$

Proposition 1. *Let $g_1, \ldots, g_n \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$ and let s be an integer coprime to m. Let moreover $\mathcal{C}$ be the Generalized Gabidulin code whose generator matrix is $M_{s,k}(g_1, \ldots, g_n)$. Then*

$$\mathcal{C} = \{(f(g_1), f(g_2), \ldots, f(g_n)) \mid f \in \mathcal{G}_{k,s}\}.$$

From now on we will write $\mathcal{G}_{k,s}(g_1, \ldots, g_n)$ for such a code.

For many years Gabidulin codes have been the only known MRD codes over $\mathbb{F}_{q^m}$. Recently some construction of non-Gabidulin MRD codes have been discovered [2, 3], but many of these codes are not linear over $\mathbb{F}_{q^m}$. Some constructions of linear non-Gabidulin MRD codes can be found in [6] and as a special class of the codes presented in [16].

Although there are few known constructions of MRD codes, it was shown in [12] that most linear rank metric codes are MRD and that the Gabidulin codes are only a small franction among the MRD codes.

3 Fuzzy Commitment Scheme with the Rank Distance

In 1999 Juels and Wattenberg [8] proposed a fuzzy commitment scheme to allow fuzzy authentication with secure storage of biometric data in binary form. In [15] the authors revisited the scheme in the setting of an arbitrary finite field by focusing on implementations and security concerns. In [1] they proposed a dual version of

the scheme, called fuzzy syndrome hashing, featuring some advantages in terms of security and use of iterative decoding. In [4] they presented scenarios involving burst error correction and higher dimensional data.

Here we are going to describe a new fuzzy commitment scheme using the rank metric. In a following section about applications, we will describe a few scenarios where this scheme can be applied.

In our authentication model, we wish to consider two vectors $b, b' \in \mathbb{F}_{q^m}^n$ (or, equivalently, their matrix representations $B, B' \in \mathbb{F}_q^{m\times n}$) as belonging to the same person or entity as long as their rank distance is less than a certain predetermined threshold. And for security concerns we do not want to store vectors (or matrices) unencrypted.

Suppose now that we have a rank metric code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ whose minimum distance is $d = 2t + 1$ and assume there exists an efficient algorithm for decoding up to t errors.

Let $h : \mathbb{F}_q^{m\times n} \to \mathbb{F}_q^{m\times r}$ be a collision resistant hash function, i.e. such that it is not feasible to compute an $u \in h^{-1}(v)$ for any $v \in \mathbb{F}_q^{m\times r}$. Observe that a hash function $h' : \mathbb{F}_{q^m}^n \to \mathbb{F}_{q^m}^r$ can be defined starting from h, as the diagram

$$\begin{array}{ccc} \mathbb{F}_q^{m\times n} & \xrightarrow{h} & \mathbb{F}_q^{m\times r} \\ \updownarrow & & \updownarrow \\ \mathbb{F}_{q^m}^n & \xrightarrow{h'} & \mathbb{F}_{q^m}^r \end{array}$$

shows.

As in the standard fuzzy commitment scheme, we select at random a codeword $c_b \in \mathcal{C}$ and we store the tuple

$$(l, h(c_b))$$

where $l = b - c_b$.

This scheme is essentially the analogue of the standard fuzzy commitment with the difference that we use rank metric codes instead of Hamming codes. Analogously as in [15] one can show the following result.

Theorem 2. *If $b \in \mathbb{F}_{q^m}^n$ can be chosen uniformly over the entire ambient space $\mathbb{F}_{q^m}^n$, then computing $b \in \mathbb{F}_{q^m}^n$ from the stored data $(l, h(c_b))$ is computationally equivalent to invert the restricted hash function*

$$h_{|\mathcal{C}} : \mathcal{C} \longrightarrow \mathbb{F}_{q^m}^r.$$

4 A Linearized Polynomial Fuzzy Vault Scheme

The polynomial fuzzy vault (PFV) scheme was introduced in [7] and allows fuzzy authentication in the set-difference metric. In [11] the authors proposed a fuzzy vault scheme using codes in another metric, relating the set difference with the subspace distance on the set of Grassmanians. The PFV scheme can also be generalized in a natural way using linearized polynomials and codes over the rank metric as follows.

First, we make the following assumption about the set of features used for authentication, both the set initially used to build the vault and the one submitted later for authentication.

Assumption 1. *Assume that the set of features (A or W in the following) is given by n $\mathbb{F}_q$-linearly independent elements in $\mathbb{F}_{q^n}$, i.e. it is an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^n}$.*

This is usually not a restrictive assumption given the follwing result:

Lemma 1. *If the features are chosen with uniform distribution, then Assumption 1 is satisfied with probability*

$$\prod_{i=0}^{n-1} \frac{(q^n - q^i)}{(q^n - i)} \geq \prod_{i=0}^{n-1} (1 - q^{i-n}).$$

Proof. The number of $\mathbb{F}_q$-basis of $\mathbb{F}_{q^n}$ is $\frac{\prod_{i=0}^{n-1}(q^n - q^i)}{n!}$, while the number of subsets of $\mathbb{F}_{q^n}$ with cardinality n is $\binom{q^n}{n}$. □

Now, let $\ell < n$ be two positive integers and let $0 < s < n$ be another integer coprime with n. Let $(k_0, \ldots, k_{\ell-1}) \in \mathbb{F}_{q^n}^{\ell}$ the secret key and $\kappa(x) = k_0 x + k_1 x^{[s]} + \ldots + k_{\ell-1} x^{[s(\ell-1)]} \in \mathcal{L}_n(\mathbb{F}_{q^n})$ be the corresponding linearized polynomial. Consider a set of features $A = \{g_1, \ldots, g_n\} \subseteq \mathbb{F}_{q^n}$ given by n $\mathbb{F}_q$-linearly independent elements. Choose a random map $\lambda : \mathbb{F}_{q^n} \longrightarrow \mathbb{F}_{q^n}$ such that $\lambda(x) \neq \kappa(x)$ for all $x \in B$, where $B = \mathbb{F}_{q^n} \setminus A$.

Following the classical PFV scheme, we define the sets

$$\begin{aligned} \mathcal{P}_{auth} &= \{(x, \kappa(x)) \mid x \in A\}, \\ \mathcal{P}_{chaff} &= \{(x, \lambda(x)) \mid x \in B\}, \\ \mathcal{V} &= \mathcal{P}_{auth} \cup \mathcal{P}_{chaff}. \end{aligned}$$

$\mathcal{P}_{auth}$ is called *set of authentic points*, $\mathcal{P}_{chaff}$ is the *set of chaff points*, and $\mathcal{V}$ is called *set of vault points*.

The last ingredients of the fuzzy vault scheme are the code

$$\mathcal{C} = \mathcal{G}_{\ell,s}(g_1, \ldots, g_n)$$

and an error correction decoding algorithm for $\mathcal{C}$.

For our constructions of the Linearized Polynomial Fuzzy Vault (LPFV), it is convenient to consider a Gabidulin code as a code whose codewords consist of evaluations of a linearized polynomial $f \in \mathcal{G}_{\ell,s}$ over any set of n linearly independent elements in $\mathbb{F}_{q^n}$. Concretely, we think of a codeword as a set of pairs $\{(g_i, y_i)\}_{i=1}^n$, where $g_i \in \mathbb{F}_{q^n}$, are linearly independent over $\mathbb{F}_q$, and $y_i = f(g_i)$, for a linearized polynomial $f \in \mathcal{G}_{\ell,s}$. In this framework, suppose that a witness attempts to gain access to the key, and submits a set of features $W \subset \mathbb{F}_{q^n}$.

Given Assumption 1, if $Z \subseteq \mathcal{V}$ is the subset of vault points (x, y) with $x \in W$, we can consider the $\mathbb{F}_q$-linear map

$$L_Z : \mathbb{F}_{q^n} \longrightarrow \mathbb{F}_{q^n}$$

such that $L_Z(x) = y$ for all $(x, y) \in Z$. Now, think of the received word c' as consisting of the set of pairs $\{(g_i, L_Z(g_i))\}_{i=1}^n$, for $g_i \in A$. The secret codeword of the LPFV scheme is instead c, given by the set of pairs $\{(g_i, \kappa(g_i))\}_{i=1}^n$. With this notation it is easy to see that

$$d_R(c, c') = \mathrm{rk}(\kappa - L_Z).$$

The following results relate the rank distance with the set difference, showing that the rank metric can be a good approximation of the set-difference metric. Let $d_\Delta(A, W) := |(A \backslash W) \cup (W \backslash A)|$ denote the set-difference between A and W.

Proposition 2. *In the setting of the LPFV scheme, suppose that the values $\lambda(x)$, for $x \in B$ are chosen at random uniformly and independently in $\mathbb{F}_{q^n} \setminus \{\kappa(x)\}$. Then*

1. $2d_R(c, c') \leq d_\Delta(A, W)$.
2. *Let $0 \leq u \leq n$ be an integer. Then*

$$\Pr\{2d_R(c, c') = d_\Delta(A, W) \mid |A \cap W| = u\} = \prod_{i=0}^{n-u-1} \frac{(q^n - q^i)}{(q^n - 1)} = 1 + O(q^{-u-1}).$$

Proof. 1. Let W be the set of features submitted, and let $u = |A \cap W|$. Then we have $d_\Delta(A, W) = 2n - 2|A \cap W| = 2n - 2u$. Consider now the $\mathbb{F}_q$-linear map $L_Z : \mathbb{F}_{q^n} \longrightarrow \mathbb{F}_{q^n}$ such that $L_Z(x) = y$ for $(x, y) \in Z$. The set of first coordinates of Z is an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^n}$ and the linear map $\kappa - L_Z$ is 0 on $A \cap W$. Therefore

$$d_R(c, c') = \mathrm{rk}(\kappa - L_Z) \leq n - u = \frac{d_\Delta(A, W)}{2}.$$

2. Since the $\lambda(x)$, for $x \in B$, are chosen at random uniformly and independently in $\mathbb{F}_{q^n} \setminus \{\kappa(x)\}$, then the values $(L_Z - \kappa)(x)$, for $x \in W \setminus (A \cap W)$ are chosen at random uniformly and independently in $\mathbb{F}_{q^n} \setminus \{0\}$. Furthermore, the condition $2d_R(c, c') = d_\Delta(A, W)$ is equivalent to the condition that the values $(L_z - \kappa)(x)$, for $x \in W \setminus (A \cap W)$ are linearly independent. Hence,

$$\Pr\{2d_R(c, c') = d_\Delta(A, W) \mid |A \cap W| = u\} = \frac{\left|\left\{A \in \mathbb{F}_q^{n \times (n-u)} \mid \mathrm{rk}(A) = n - u\right\}\right|}{(q^n - 1)^{(n-u)}}.$$

□

Theorem 3. *Under the same hypothesis of Proposition 2, the following statements hold.*

1. *If $d_\Delta(A, W) \leq 2\left\lfloor \frac{n-\ell}{2} \right\rfloor$, then the vault recovers the key $\kappa(x)$.*
2. $$\Pr\{2d_R(c, c') = d_\Delta(A, W)\} = 1 + O(q^{-1}).$$

Proof. 1. By Proposition 2 we have $2d_R(c, c') \leq d_\Delta(A, W) \leq 2\left\lfloor \frac{n-\ell}{2} \right\rfloor$. Therefore we are within the error-correction capability and we can correctly obtain the codeword c, and hence the key $\kappa(x)$.

2. We can write $\Pr\{2d_R(c,c') = d_\Delta(A,W)\}$ as

$$\begin{aligned}
&\sum_{u=0}^{n} \Pr\{2d_R(c,c') = d_\Delta(A,W) \mid |A \cap W| = u\} \Pr\{|A \cap W| = u\} \\
=&\sum_{u=0}^{n} \left(1 + O(q^{-u-1})\right) \Pr\{|A \cap W| = u\} \\
=&\sum_{u=0}^{n} \left(1 + O(q^{-1})\right) \Pr\{|A \cap W| = u\} \\
=&\left(1 + O(q^{-1})\right) \sum_{u=0}^{n} \Pr\{|A \cap W| = u\} \\
=&1 + O(q^{-1}).
\end{aligned}$$

□

Remark. Probabilistic results in Proposition 2 and Theorem 3 do not depend on the probability distribution of the choice of the features. We are only assuming that our construction of the Linearized Polynomial Fuzzy Vault is made by choosing at random uniformly and independently the values $\lambda(x)$ for $x \in B$.

4.1 Generalization of the LPFV Scheme

In our construction of the LPFV we considered Gabidulin codes of length n over $\mathbb{F}_{q^n}$. The motivation is that given a set of features W satisfying Assumption 1, the map $L_Z : \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$ is uniquely determined, and hence also the received word c'.

We can generalize our LPFV considering Gabidulin codes of length n over the field $\mathbb{F}_{q^m}$, where $n < m$, but we need to define the map L_Z in a suitable way.

Before explaining how to construct L_Z, we can observe that an analogue of Lemma 1 holds and it can be proved in the same way, but in this case the probability that the set of features is made of linearly independent elements is equal to

$$\prod_{i=0}^{n-1} \frac{(q^m - q^i)}{(q^m - i)} = 1 + O(q^{-1-m+n}).$$

Now, let $\mathcal{W}$ and $\mathcal{A}$ be the $\mathbb{F}_q$-subspaces of $\mathbb{F}_{q^m}$ spanned respectively by W and A. First, we can observe that, in order to build the received word c' as the set $\{(g_i, L_Z(g_i))\}_{i=1}^n$, we only need to define map L_Z on $\mathcal{A}$. We propose the following construction.

We first define the application L_Z on W as $L_Z(x) = y$ for all $(x,y) \in Z$. Then complete W to a basis B of $\mathcal{A} + \mathcal{W}$, by adding the elements g_i in increasing order with respect to the indices i. For those g_i, we set $L_Z(g_i) = \kappa(g_i) + \alpha^{q^i}$, where $\alpha \in \mathbb{F}_{q^m}$ and $\left\{\alpha^{q^i}\right\}_{i=0}^{m-1}$ is a normal basis of $\mathbb{F}_{q^m}$ as an $\mathbb{F}_q$-vector space.

In this way, our map is uniquely determined on $\mathcal{A} + \mathcal{W}$, and in particular on $\mathcal{A}$. Let again c be the codeword given by the set of pairs $\{(g_i, \kappa(g_i))\}_{i=1}^n$. With this notation it is easy to see that

$$d_R(c,c') = \mathrm{rk}(\kappa - L_Z)_{|_{\mathcal{A}}} \leq \mathrm{rk}(\kappa - L_Z).$$

The following results are the analogues of Proposition 2 and Theorem 3, and they relate the rank distance of c and c' with the set difference of A and W.

Proposition 3. *In the setting of the generalized LPFV scheme, suppose that the values $\lambda(x)$, for $x \in B$ are chosen at random uniformly and independently in $\mathbb{F}_{q^m} \setminus \{\kappa(x)\}$.*

1. *Let the subspace distance be $d_S(\mathcal{A}, \mathcal{W}) := \dim_{\mathbb{F}_q}(\mathcal{A}) + \dim_{\mathbb{F}_q}(\mathcal{W}) - 2\dim_{\mathbb{F}_q}(\mathcal{A} \cap \mathcal{W})$. Then*

$$d_S(\mathcal{A}, \mathcal{W}) \leq 2d_R(c, c') \leq d_S(\mathcal{A}, \mathcal{W}) + 2\,\mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}\cap\mathcal{W}} \leq d_\Delta(A, W).$$

2. *Let $0 \leq u \leq v \leq n$ be two integers. Then*

$$\Pr\left\{2d_R(c, c') = d_\Delta(A, W) \mid |A \cap W| = u, \dim(\mathcal{A} \cap \mathcal{W}) = v\right\} = \prod_{i=n-v}^{n-u-1} \frac{(q^m - q^i)}{q^m - 1}.$$

Proof. 1. Following the construction of the map L_Z, we can write the subspace $\mathcal{A}$ as the direct sum of $\mathcal{A} \cap \mathcal{W}$ and the subspace $\widehat{\mathcal{A}}$, where $\widehat{\mathcal{A}} = \langle g_i \mid i \in I \rangle$ and $I \subset \{1, \ldots, n\}$ with $|I| = n - \dim_{\mathbb{F}_q}(\mathcal{A} \cap \mathcal{W})$. Therefore we can write

$$\mathrm{rk}(\kappa - L_Z)_{|\widehat{\mathcal{A}}} \leq \mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}} \leq \mathrm{rk}(\kappa - L_Z)_{|\widehat{\mathcal{A}}} + \mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}\cap\mathcal{W}}. \quad (1)$$

Let $r = \dim_{\mathbb{F}_q}(\widehat{\mathcal{A}})$. By definition of the L_Z, we have

$$\mathrm{rk}(\kappa - L_Z)_{|\widehat{\mathcal{A}}} = \mathrm{rk}(\alpha^{q^{i_1}}, \ldots, \alpha^{q^{i_r}}).$$

By construction $\{\alpha^{q^i}\}_{i=0}^{m-1}$ is a normal basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$, and hence we can conclude that

$$\mathrm{rk}(\kappa - L_Z)_{|\widehat{\mathcal{A}}} = r = \dim_{\mathbb{F}_q}(\widehat{\mathcal{A}}) = n - \dim_{\mathbb{F}_q}(\mathcal{A} \cap \mathcal{W}) = \frac{d_S(\mathcal{A}, \mathcal{W})}{2}.$$

Substituting this equation in (1) we obtain the first two inequalities.
For the last inequality we notice that the map $(\kappa - L_Z)_{|\mathcal{A}\cap\mathcal{W}}$ is 0 on $|A \cap W|$, and therefore

$$\mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}\cap\mathcal{W}} \leq \dim_{\mathbb{F}_q}(\mathcal{A} \cap \mathcal{W}) - |A \cap W|.$$

Hence we can conclude that

$$d_S(\mathcal{A}, \mathcal{W}) + 2\,\mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}\cap\mathcal{W}} \leq 2n - 2|A \cap W| = d_\Delta(A, W).$$

2. Let $u = |A \cap W|, v = \dim(\mathcal{A} \cap \mathcal{W})$. Then we can write

$$W = \{u_1, \ldots, u_{n-v}, w_{n-v+1}, \ldots, w_{n-u}, g_{j_1}, \ldots, g_{j_u}\},$$

where $u_i \notin \mathcal{A}$ for $i = 1, \ldots, n - v$ and $w_i \in \mathcal{A} \setminus A$ for $i = n - v + 1, \ldots, n - u$. Therefore $2\,\mathrm{rk}(\kappa - L_Z)_{|\widehat{\mathcal{A}}} = 2n - 2v$, and the condition

$$\mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}} = \mathrm{rk}(\kappa - L_Z)_{|\widehat{\mathcal{A}}} + \mathrm{rk}(\kappa - L_Z)_{|\mathcal{A}\cap\mathcal{W}} = n - u$$

is equivalent to the condition

$$\mathrm{rk}(\alpha^{q^{i_1}}, \ldots, \alpha^{q^{i_{n-v}}}, (\kappa - L_Z)(w_{n-v+1}), \ldots, (\kappa - L_Z)(w_{n-u})) = n - u.$$

By hypothesis the values $(L_Z - \kappa)(w_i)$, for $i = n - v + 1, \ldots, n - u$ are chosen at random uniformly and independently in $\mathbb{F}_{q^m} \setminus \{0\}$, and we can conclude that the probability we are looking for is equal to

$$\frac{\left|\left\{A \in \mathbb{F}_q^{m\times(v-u)} \mid \mathrm{rk}(A \mid X) = n - u\right\}\right|}{(q^m - 1)^{(v-u)}},$$

where X is the matrix representation over $\mathbb{F}_q$ of the vector $(\alpha^{q^{i_1}}, \ldots, \alpha^{q^{i_{n-v}}})$. Since

$$\left|\left\{A \in \mathbb{F}_q^{m\times(v-u)} \mid \mathrm{rk}(A \mid X) = n - u\right\}\right| = \prod_{i=n-v}^{n-u-1} (q^m - q^i),$$

this concludes the proof.

□

Theorem 4. *Under the same hypothesis of Proposition 3, the following statements hold.*

1. *If $d_\Delta(A, W) \leq 2\left\lfloor \frac{n-\ell}{2} \right\rfloor$, then the vault recovers the key $\kappa(x)$.*
2. $$\Pr\{2d_R(c, c') = d_\Delta(A, W)\} = 1 + O(q^{-1-m+n}).$$

Proof. 1. The proof is essentially the same as the proof of Theorem 3.1, using Proposition 3.1.

2. In order to simplify the notation we introduce the events $D_u = \{|A \cap W| = u\}$, $E_v = \left\{\dim_{\mathbb{F}_q}(\mathcal{A} \cap \mathcal{W}) = v\right\}$ for $0 \leq u, v \leq n$, and $X = \{2d_R(c, c') = d_\Delta(A, W)\}$. Then we have

$$\begin{aligned}
\Pr\{X\} &= \sum_{0\leq u\leq v\leq n} \Pr\{X \mid D_u \cap E_v\} \Pr\{D_u \cap E_v\} \\
&= \sum_{0\leq u\leq v\leq n} \left(1 + O(q^{-1-m-u+n})\right) \Pr\{D_u \cap E_v\} \\
&= \sum_{0\leq u\leq v\leq n} \left(1 + O(q^{-1-m+n})\right) \Pr\{D_u \cap E_v\} \\
&= \left(1 + O(q^{-1-m+n})\right) \sum_{0\leq u\leq v\leq n} \Pr\{D_u \cap E_v\} \\
&= 1 + O(q^{-1-m+n}).
\end{aligned}$$

□

Remark. Suppose one wants to use a generalized LPFV scheme with n genuine features, and suppose moreover that a field $\mathbb{F}_q$ and an extension field $\mathbb{F}_{q^m}$, with $m \geq n$, are given. By Theorem 4.2 we can see that the bigger is m the better is the approximation of the set difference with the rank distance. On the other hand, increasing m implies an increase of the computational cost of the operations. Then one can choose the best m based on the application and the particular requirements of the context.

5 Applications

The schemes presented above can be applied in several scenarios for different purposes. In this section we would like to give just a few examples.

One scenario for the fuzzy commitment scheme in the rank metric is the following. Suppose B is the matrix used to create the stored tuple and imagine it as an image. It may happen for some reason that B gets somehow damaged in a way that a few rows (or columns) are erased or anyway not the same as before. One can then authenticate with the new matrix B' as long as not too many rows (or columns) are different. In another situations the matrix B may be slightly changed into B' by having all elements increased by a common error, and again the difference between the two matrices is a matrix of low rank, exactly 1 in this case.

Another scenario involves a multi-factor authentication problem. Suppose that in order to perform authentication one needs a large number of conditions fulfilled, namely imagine a matrix with a large number of columns whereby condition number i is fulfilled whenever column number i equals a predetermined vector v_i. If you want to allow authentication as long as a certain big enough number of conditions are satisfied, then the fuzzy commitment scheme in the rank metric can be used. Indeed having two matrices A and A' that both satisfy a certain condition corresponds to a zero column in the difference $A - A'$ which directly affects the rank distance between the two.

Applications for the linearized polynomial fuzzy vault scheme overlap with those of the standard fuzzy vault, i.e. we are considering authentication based on the set difference metric. It may be preferable to use the linearized version and decoding in the rank metric for certain choices and combinations of parameters which are usually dependent on the application. Also, the use of linear maps may be preferred for certain implementations.

6 Acknowledgments

The authors were supported by Swiss National Science Foundation grant n.169510.

References

[1] M. Baldi, M. Bianchi, F. Chiaraluce, J. Rosenthal, and D. Schipani. On fuzzy syndrome hashing with LDPC coding. In *4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, pages 1–5. ACM, 2011.

[2] A. Cossidente, G. Marino, and F. Pavese. Non-linear maximum rank distance codes. *Designs, Codes and Cryptography*, 79(3):597–609, 2016.

[3] J. de la Cruz, M. Kiermaier, A. Wassermann, and W. Willems. Algebraic structures of MRD codes. *Advances in Mathematics of Communications*, (10):499–510, 2016.

[4] F. Fontein, K. Marshall, J. Rosenthal, D. Schipani, and A.-L. Trautmann. On burst error correction and storage security of noisy data. In *20th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 1–7, 2012.

[5] E. M. Gabidulin. Theory of codes with maximum rank distance. *Problemy Peredachi Informatsii*, 21(1):3–16, 1985.

[6] A. Horlemann-Trautmann and K. Marshall. New criteria for MRD and Gabidulin codes and some rank-metric code constructions. *arXiv:1507.08641, to appear in Advances in Mathematics of Communications*, 2016.

[7] A. Juels and M. Sudan. A fuzzy vault scheme. *Des. Codes Cryptography*, 38(2):237–257, 2006.

[8] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *6th ACM conference on Computer and communications security*, CCS '99, pages 28–36, 1999.

[9] A. Kshevetskiy and E. Gabidulin. The new construction of rank codes. In *International Symposium on Information Theory (ISIT), 2005*, pages 2105–2108, 2005.

[10] P. Loidreau. A Welch–Berlekamp like algorithm for decoding Gabidulin codes. In *Coding and cryptography*, pages 36–45. Springer, 2006.

[11] K. Marshall, D. Schipani, A.-L. Trautmann, and J. Rosenthal. Subspace fuzzy vault. In *Physical and Data-Link Security Techniques for Future Communication Systems*, pages 163–172. Springer, 2016.

[12] A. Neri, A.-L. Horlemann-Trautmann, T. Randrianarisoa, and J. Rosenthal. On the genericity of maximum rank distance and Gabidulin codes. *arXiv:1605.05972*, 2016.

[13] G. Richter and S. Plass. Error and erasure decoding of rank-codes with a modified Berlekamp-Massey algorithm. *ITG FACHBERICHT*, pages 203–210, 2004.

[14] R. Roth. Maximum-rank array codes and their application to crisscross error correction. *IEEE Transactions on Information Theory*, 37(2):328–336, 1991.

[15] D. Schipani and J. Rosenthal. Coding solutions for the secure biometric storage problem. In *Information Theory Workshop (ITW), 2010*, pages 1–4, 2010.

[16] J. Sheekey. A new family of linear maximum rank distance codes. *Advances in Mathematics of Communications*, (10):475–488, 2016.

[17] D. Silva and F. Kschischang. Fast encoding and decoding of Gabidulin codes. *International Symposium on Information Theory (ISIT), 2009*, pages 2858–2862, 2009.