Ideals over a Non-Commutative Ring and their Application in Cryptology

E. M. Gabidulin, A. V. Paramonov and O. V. Tretjakov
Moscow Institute of Physics and Technology
141700 Dolgoprudnii
Moscow Region, USSR

Abstract: A new modification of the McEliece public-key cryptosystem is proposed that employs the so-called maximum-rank-distance (MRD) codes in place of Goppa codes and that hides the generator matrix of the MRD code by addition of a randomly-chosen matrix. A short review of the mathematical background required for the construction of MRD codes is given. The cryptanalytic work function for the modified McEliece system is shown to be much greater than that of the original system. Extensions of the rank metric are also considered.

1. INTRODUCTION

The purpose of this paper is to show that error-correcting codes for the rank metric, as introduced recently in [1], can advantageously be used to replace codes for the usual Hamming metric in McEliece's public-key cryptosystem [2]. The next section develops the necessary mathematical background for describing codes over the rank metric. The codes themselves are derived in Section 3. Section 4 describes how these codes are used in the McEliece system and quantifies the increase in security compared to the system based on codes for the Hamming metric.

2. MATHEMATICAL PRELIMINARIES

A *linearized polynomial* with coefficients in the finite field $GF(q^N)$ is a polynomial of the form

$$F(z) = \sum_{i=0}^{n} f_i z^{[i]}$$

where here and hereafter "[i]" in an exponent is shorthand for "q^i". The largest i such that $f_i \neq 0$ will be called the *norm* of the polynomial. By way of convention, the norm of the linearized polynomial 0 is taken to be $-\infty$. We write $R_N[z]$ to denote the set of all linearized polynomials with

D.W. Davies (Ed.): Advances in Cryptology - EUROCRYPT '91, LNCS 547, pp. 482-489, 1991.

coefficients in $GF(q^N)$. Addition and multiplication in $R_N[z]$ are defined by

$$F(z) + G(z) = \sum_{i=0}^{n} (f_i + g_i)\, z^{[i]}$$

and by

$$F(z) * G(z) = \sum_{i=0}^{n} \Big(\sum_{k+s=i} f_s\, g_k^{[s]} \Big)\, z^{[i]},$$

respectively, It is important to note that multiplication in $R_N[z]$ is *not* commutative. The set $R_N[z]$ under these two operations is a non-commutative ring whose multiplicative identity is the polynomial z.

The polynomial $z^{[N]} - z$ commutes under multiplication with every polynomial F(z) in $R_N[z]$. Moreover, if G(z) and H(z) are polynomials in $R_N[z]$ with leading coefficients 1 and such that $G(z) * H(z) = z^{[N]} - z$, then G(z) and H(z) also commute under multiplication, i.e., $G(z) * H(z) = H(z) * G(z)$. Thus, one can speak unambiguously of the divisors of $z^{[N]} - z$. We will write $L_N[z]$ to denote the *quotient ring* $R_N[z]/(z^{[N]} - z)$, i. e., the ring whose operations are addition and multiplication modulo the polynomial $z^{[N]} - z$. The ring $L_N[z]$ is a non-commutive ring with q^m elements where $m = N^2$. Every left (or right) ideal in this ring is a principal ideal generated by a polynomial G(z) that divides $z^{[N]} - z$. The elements of this left ideal are all the polynomials in $R_N[z]$ of the form $F(z) * G(z)$ with F(z) in the ring $L_N[z]$.

3. IDEALS ON $L_N[z]$ AS ERROR-CORRECTING CODES

We will consider left ideals of $L_N[z]$ as *codes* over the "large" field $GF(q^N)$. Instead of the Hamming metric that is most frequently used to study the error-correcting properties of codes, we will instead use a family of metrics induced by the so-called *rank metric*. This metric was introduced in [1], where a complete theory of codes with maximal rank distance (MRD) was given, including encoding and decoding techniques.

We will write F^N to denote the N-dimensional vector space of N-tuples over the "large" field $GF(q^N)$. Let $\mathbf{x} = (x_0, x_1, \ldots x_{N-1})$ be a vector in F^N. Then the *rank norm* of **x**, denoted r(**x**), is defined to be the maximum number of components of **x** that are linearly independent when $GF(q^N)$ itself is considered as an N-dimensional vector space over the "small" field GF(q). The *rank distance* between **x** and **y**, denoted d(**x**, **y**), is then defined as r(**x** - **y**). It is easy to show that the minimum rank distance d of an (N, k) linear code over $GF(q^N)$ satisfies $d \leq N - k + 1$, a code for which $d = N - k + 1$ is called a *maximum-rank-distance* (MRD) code.

The vector $\tilde{\mathbf{x}} = (x_{N-1}^{[s]}, x_0^{[s]}, \ldots x_{N-2}^{[s]})$ will be called the *[s]-cyclic shift* of **x**. Note that $\tilde{\mathbf{x}}$ is obtained from **x** first by a right cyclic shift of its components followed by raising these components to the q^s power. A code $\mathcal{M}$ will be call an *[s]-cyclic code* if the [s]-cyclic shift of a code word is always itself a code word. Note that **x** and $\tilde{\mathbf{x}}$ have the same rank norm. The main construction of [1]-cyclic codes is given by the following theorem.

Theorem 1 [1]: Let γ be an element of $GF(q^N)$ such that $\gamma = \gamma^{[0]}, \gamma^{[1]}, \ldots, \gamma^{[N-1]}$ are linearly independent over the "small" field GF(q) [or, equivalently, such that these elements form a so-called *normal basis* for $GF(q^N)$], then the linear code $\mathcal{M}$ over $GF(q^N)$, with parity-check matrix

$$H = \begin{bmatrix} \gamma^{[0]} & \gamma^{[1]} & \ldots & \gamma^{[N-1]} \\ \gamma^{[1]} & \gamma^{[2]} & \ldots & \gamma^{[0]} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma^{[d-2]} & \gamma^{[d-1]} & \ldots & \gamma^{[d-3]} \end{bmatrix}$$

is a [1]-cyclic MRD code of length N over $GF(q^N)$ with minimum rank distance d and dimension $k = N - d + 1$.

We now wish to treat code words and other N-tuples over the "large" field $F = GF(q^N)$ as $N \times N$ matrices over the "small" field GF(q) so as to pave the way for consideration of "errors" in the digits lying in the "small" field. To do this, we suppose that a normal basis $\gamma^{[0]}, \gamma^{[1]}, \ldots, \gamma^{[N-1]}$ for F has been fixed, and we associate the vector **x** with the matrix

$$\mathbf{X} = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0:N-1} \\ x_{10} & x_{11} & \cdots & x_{1:N-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-1:0} & x_{N-1:1} & \cdots & x_{N-1:N-1} \end{bmatrix}$$

whose entries are elements of GF(q) determined from the components of **x** according to the representation of these components in the normal basis, i.e.,

$$x_i = \sum_{i=0}^{N-1} x_{ji} \gamma^{[i]}.$$

In this manner, single errors **e** in the rank metric correspond to matrices **E** with rank 1 and have the form

$$\mathbf{E} = \mathbf{C}\mathbf{J}\mathbf{D},$$

where **C**, **J** and **D** are $N \times N$ matrices over GF(q) such that **J** has N identical non-zero columns, **C** is nonsingular, and **D** is a non-zero diagonal matrix.

4. APPLICATION TO THE McELIECE PUBLIC-KEY CRYPTOSYSTEM

In [2], McEliece introduced a public-key cryptosystem based on algebraic codes that can be described as follows. The cryptographer chooses a $k \times n$ generator matrix $\mathbf{G}$ for a t-error-correcting binary Goppa code, for which a fast decoding algorithm is known, and chooses also a $k \times k$ nonsingular "scrambling" matrix $\mathbf{S}$ and an $n \times n$ permutation matrix $\mathbf{P}$. He then computes the matrix $\mathbf{K} = \mathbf{S}\,\mathbf{G}\,\mathbf{P}$, which is also the generator matrix of an (n, k) linear t-error-correcting code, but one for which no fast decoding algorithm is known [3] if $\mathbf{S}$, $\mathbf{G}$, and $\mathbf{P}$ are not individually known. He then publishes $\mathbf{K}$ as his public encryption key. When someone wishes to send him a message, that person fetches $\mathbf{K}$ from the public directory, then encrypts his k-bit message $\mathbf{m}$ as

$$\mathbf{c} = \mathbf{m}\,\mathbf{K} + \mathbf{e}$$

where $\mathbf{e}$ is a randomly chosen pattern of t or fewer errors. The legitimate receiver, i. e., the cryptographer, upon receipt of $\mathbf{c}$ first computes $\mathbf{c}\,\mathbf{P}^{-1} = (\mathbf{m}\,\mathbf{S})\,\mathbf{G} + \mathbf{e}\,\mathbf{P}^{-1}$. He then applies his fast decoding algorithm to this vector to obtain $\mathbf{m}\,\mathbf{S}$, and finally recovers the message $\mathbf{m}$ as $(\mathbf{m}\,\mathbf{S})\,\mathbf{S}^{-1}$.

The cryptanalyst's *work function* for breaking this scheme by the attack considered by McEliece [2] is

$$W \approx \beta\, k^3 \binom{n}{k} / \binom{n-t}{k},$$

where $\beta\, k^3$ is the number of computations required to invert a non-singular $k \times k$ matrix; $\beta = 1$ will be used in all examples hereafter. For the parameters suggested by McEliece ($n = 1024$, $k = 524$, $t = 50$), this gives $W \approx 2^{80.7}$. Adams and Meijer [4] determined that the value of t that maximizes W for $n = 1024$ was $t = 37$, which gives $k = 654$ and $W \approx 2^{84.1}$.
Lee and Brickell [5] improved the attack; against their attack the best choice is $t = 38$ which gives $W \approx 2^{73.4}$.

The main disadvantages of McEliece's public-key cryptosystem are its large public key (about 2^{19} bits for McEliece's original parameters), its expansion of the plaintext by a factor n/k (about 2), and the existence of a systematic attack for the cryptanalyst. We will see that these disadvantages can be overcome at least partially by the use of codes for the rank metric and its induced metrics.

To adapt McEliece's scheme to MRD codes requires some modification. First, because there are no distinguished coordinates in an MRD code, there is no point to using the permutation matrix $\mathbf{P}$ when $\mathbf{G}$ is the generator matrix of a MRD code. However, the matrix $\mathbf{S}\,\mathbf{G}$ is still the

generator matrix of another MRD code so this structure must be hidden in another way to make the decoding problem "hard". We suggest hiding this structure by adding a matrix $\alpha^T \mathbf{e}_g$ to $\mathbf{S}\,\mathbf{G}$ where α is a non-zero k-tuple and $\mathbf{e}_g$ is a non-zero n-tuple over GF(q^n) such that $r(\mathbf{e}_g) \leq t_g < t$, where t_g is a design parameter and t is the rank-error-correcting capability of the code.

The modified McEliece cryptosystem works as follows. The cryptographer chooses a $k \times n$ generator matric $\mathbf{G}$ for a t-rank-error-correcting code, and chooses also a $k \times k$ nonsingular "scrambling" matrix $\mathbf{S}$ together with a matrix $\alpha^T \mathbf{e}_g$ as described above. He then computes the matrix $\mathbf{K} = \mathbf{S}\,\mathbf{G} + \alpha^T \mathbf{e}_g$. He then publishes $\mathbf{K}$ as his public encryption key. When someone wishes to send him a message, that person fetches $\mathbf{K}$ from the public directory, then encrypts his k-bit message $\mathbf{m}$ as

$$\mathbf{c} = \mathbf{m}\,\mathbf{K} + \mathbf{e}_e$$

where $\mathbf{e}_e$ is a randomly chosen pattern of $t_e = t - t_g$ or fewer rank errors. The legitimate receiver applies his fast decoding algorithm to this $\mathbf{c}$ to remove the error pattern $\mathbf{m}\,\alpha^T \mathbf{e}_g + \mathbf{e}_e$ (which has rank weight at most t) to obtain $\mathbf{m}\,\mathbf{S}$, and finally recovers the message $\mathbf{m}$ as $(\mathbf{m}\,\mathbf{S})\,\mathbf{S}^{-1}$.

There are two possible attacks on this modified scheme. The first is similar to that in [4] and [5] for the original scheme. The difference is that, with high probability, there will be no subset of k code coordinates that is error-free, which means that the cryptanalyst must search through all error patterns of rank t_e or less in some selected set of k coordinates. The number of k-tuples $\mathbf{e}$ over GF(q^n) with $r(\mathbf{e}) \leq t_e$ is much greater than the number of such $\mathbf{e}$ with Hamming weight at most t_e when $t_e < n/2$. Thus, the complexity of this attacking algorithm for the rank metric is much greater than for the Hamming metric. The work function for the rank metric is

$$W \approx k^3\, L(k, t_e)$$

where $L(k, t_e)$ is the number of k-tuples $\mathbf{e}$ over GF(q^n) with $r(\mathbf{e}) = t_e$. This number is given by

$$L(k, i) \approx \begin{bmatrix} k \\ i \end{bmatrix} (q^n - 1)(q^n - q) \ldots (q^n - q^{i-1}),$$

where

$$\begin{bmatrix} k \\ i \end{bmatrix} = \frac{(q^k - 1)(q^k - q) \ldots (q^k - q^{i-1})}{(q^i - 1)(q^i - q) \ldots (q^i - q^{i-1})}$$

is the number of i-dimensional subspaces of a k-dimensional vector space over GF(q^n).

The second attack is for the cryptanalyst to try to determine from the public key **K** a generator matrix for the code for which he can find a decoding algorithm. The complexity of the best attack of this kind that we have been able to formulate is

$$W \approx L(k, t_g)\,(q^n - 1)^k\, n^3.$$

The complexity of this second attack is much greater than that of the first, but it must be remembered that the second attack needs to be carried out only once to solve any number of cryptograms whereas the first attack must be carried out for each cryptogram to be solved.

As a numerical example, consider the case where $n = 20, k = 12$ and $t_e = 3$. The code rate is then $k/n = 3/5$, i. e., the plaintext is expanded by a factor of 5/3. The size of the public key is $n^2 k = 4800 \approx 2^{12}$ bits. The work functions for the first and second attack are about 2^{100} and 2^{290}, respectively. [Note that the operations counted are in $GF(2^n)$ rather than in GF(2).] All parameters are substantially better than for the original McEliece system that uses codes based on the Hamming metric.

5. EXTENSIONS OF THE RANK METRIC

The rank metric matches [s]-cyclic shifting for any s in the sense that a vector x and its [s]-cyclic shift have the same rank norm. But it is also possible to introduce a new set of metrics that apply for specific values of s, as we shall now do.

Consider now the mapping φ_s (where $0 \le s < N$) defined by

$$\varphi_s(x) = (x_0^{[0]}, x_1^{[s]}, x_2^{[2s]}, \ldots, x_{N-1}^{[(N-1)s]}).$$

This mapping is a bijection on F^N but is nonlinear for $s \neq 0$. We will call the metric D_s defined by

$$D_s(x, y) = r(\varphi_s^{-1}(x - y))$$

the *metric on F^N induced by φ_s*. Single errors in the induced metric D_s have the same structure **C J D** as for the rank metric except that the matrix **J** now may have any non-zero first column. One now considers this first column as representing an element of $GF(q^N)$, say $x_0^{[0]}$, and forms subsequent elements $x_0^{[s]}, x_0^{[2s]}, \ldots, x_0^{[(N-1)s]}$, then returns to the matrix representation. If one uses a normal-basis representation, then each subsequent column of the maxtrix is simply the cyclic shift by s positions of the components of the previous column.

If a code $\mathfrak{M}$ is optimal for the rank metric, then the *image code* $\varphi_s(\mathfrak{M})$ is optimal for the induced metric D_s. Moreover, encoding and decoding

schemes for the original code in the rank metric are easily adapted to the image code in the induced metric.

It was shown in [1] that any left ideal $\mathfrak{I}$ of $L_N[z]$ is a [1]-cyclic MRD code. The following theorem is an immediate consequence.

Theorem 2: Let $\mathfrak{I}$ be a left ideal of $L_N[z]$ and let $\mathfrak{I}_s = \varphi_s(\mathfrak{I})$. Then $\mathfrak{I}_s$ is an [s+1]-cyclic codes with maximal minimum distance for the induced metric D_s.

The choice $s = N - 1$ in Theorem 2 gives nonlinear codes over GF(q) that are cyclic in the usual sense and that have maximal minimum distance in the induced metric D_s.

The public encryption key when the metric D_s is used in McEliece's cryptosystem is the matrix

$$K_s = S\,\phi_s(G) + \alpha^T \phi_s(e_g)$$

where, as above, G is the $k \times n$ generator matrix for a t-rank-error-correcting code, e_g is a vector of rank norm t_g, and $\phi_s(G)$ is the matrix obtained by applying the mapping ϕ_s to each row of G. The k-bit message m is encrypted as

$$c = m\,K_s + \phi_s(e)$$

where e is a randomly chosen vector of rank norm at most $t - t_g$.

The use of these induced norms in the McEliece system allows one to increase the number of possible public encryption keys compared to the case where only the rank metric is used. For a fixed $s \neq 0$, the system based on the D_s metric is equivalent to that based on the rank metric. We are currently investigating whether it would be possible to increase the security of the system by somehow making s part of the private key only.

REMARK

We point out that MRD codes can also be useful in implementing perfect local randomizers. Maurer and Massey's bound [6] on the degree δ of perfect local randomizers obtained with maximum-distance-separable (MDS) codes for the Hamming metric also applies to MRD codes, since MRD codes are also MDS. However, this bound can sometimes be improved for MRD codes.

REFERENCES

[1] E. M. Gabidulin, "Theory of Codes with Maximum Rank Distance", *Problems of InformationTransmission* , vol. 21, no. 1, pp. 1-12, July, 1985 (Russian Original, January-March, 1985).

[2] R. J. McEliece, "A Public-Key Cryptosystem Based on Algebraic Coding Theory", pp. 114-116 in *DSN Progress Report 42-44*, Jet Propulsion Lab., Pasadena, CA, January-February, 1978.

[3] E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg, "On the Inherent Intractability of Certain Coding Problems", *IEEE Trans. Inf. Th.*, vol. IT-24, pp. 384-386, May 1978.

[4] C. M. Adams and H. Meijer, "Security-Related Comments Regarding McEliece's Public-Key Cryptosystem", pp. 224-228 in *Advances in Cryptology--CRYPTO '87* (Ed. C. Pomerance), Lecture Notes in Computer Sci. No. 293. Heidelberg and New York: Springer-Verlag, 1988.

[5] P. J. Lee and E. F. Brickell, "An Observation on the Security of the McEliece Public-Key Cryptosystem", pp. 275-280 in *Advances in Cryptology--EUROCRYPT '88* (Ed. C. Günther), Lecture Notes in Computer Sci. No. 330. Heidelberg and New York: Springer-Verlag, 1988.

[6] U. M. Maurer and J. L. Massey, "Perfect Local Randomness in Pseudo-Random Sequences, pp. 100-112 in *Advances in Cryptology--CRYPTO '89* (Ed. G. Brassard), Lecture Notes in Computer Sci. No. 435. Heidelberg and New York: Springer-Verlag, 1990.