A distinguisher for high-rate McEliece Cryptosystems

J.C. Faugère (INRIA, SALSA project),
A. Otmani (Université Caen- INRIA, SECRET project),
L. Perret (INRIA, SALSA project),
J.-P. Tillich (INRIA, SECRET project)

May 28th, 2010
1. Algebraic approach for attacking the McEliece cryptosystem

- $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ with $x_i \neq x_j$ if $i \neq j$
- $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{F}_{q^m}^n$ with $y_i \neq 0$

For any $t < n$, let $H = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ y_1x_1 & y_2x_2 & \cdots & y_nx_n \\ \vdots & \vdots & \ddots & \vdots \\ y_1x_1^{t-1} & y_2x_2^{t-1} & \cdots & y_nx_n^{t-1} \end{pmatrix}$.

**Definition 1.** An **alternant** code is the kernel of an $H$ of this type

$$A_t(\mathbf{x}, \mathbf{y}) = \{ \mathbf{v} \in \mathbb{F}_q^n | H\mathbf{v}^T = \mathbf{0} \}.$$

**Goppa code:** $\exists \Gamma$, polynomial of degree $t$ such that $y_i = \Gamma(x_i)^{-1}$. 
Decoding Alternant and Goppa codes

Proposition 1. [decoding alternant codes] \( t/2 \) errors can be decoded in polynomial time as long as \( x \) and \( y \) are known.

Proposition 2. [The special case of binary Goppa codes] In the case of a binary Goppa code \((q = 2)\), \( t \) errors can be decoded in polynomial time, if \( x \) and \( \Gamma \) are known.
The problem

What is known: a basis of the code → rows of a generator matrix $G = (g_{i,j})$ of size $k \times n$.

What we also know:

$$HG^T = 0.$$  \hspace{1cm} (1)

What we want to find: $H$

Find in the case of an alternant code $x, y$, and in the special case of a binary Goppa code $x$ and $\Gamma$. 
The algebraic system

\( HG^T = 0 \) translates to

\[
\begin{align*}
&g_{1,1}Y_1 + \cdots + g_{1,n}Y_n = 0 \\
&\vdots \\
&g_{k,1}Y_1 + \cdots + g_{k,n}Y_n = 0 \\
&g_{1,1}Y_1X_1 + \cdots + g_{1,n}Y_nX_n = 0 \\
&\vdots \\
&g_{k,1}Y_1X_1 + \cdots + g_{k,n}Y_nX_n = 0 \\
&g_{1,1}Y_1X_1^{t-1} + \cdots + g_{1,n}Y_nX_n^{t-1} = 0 \\
&\vdots \\
&g_{k,1}Y_1X_1^{t-1} + \cdots + g_{k,n}Y_nX_n^{t-1} = 0
\end{align*}
\]

where the \( g_{i,j} \)'s are known coefficients in \( \mathbb{F}_q \) and \( k \geq n - tm \).
Freedom of choice in (2)

Proposition 3. Theoretically, the system has $2n$ unknowns but we can take arbitrary values for one $Y_i$ and for three $X_i$'s (as long as these values are different).
Applications

When the number of unknowns is small, ex:

- Berger-Cayrel-Gaborit-Otmani proposal at AfricaCrypt’09 based on quasi-cyclic alternant codes
- Misoczki-Baretto at SAC’09 variant based on quasi-dyadic Goppa codes

⇒ algebraic system can be solved by (dedicated) Grobner basis techniques.

▶ breaks all parameters proposed in these articles ([Faugère-Otmani-Perret-Tillich; Eurocrypt 2010] with the exception of binary dyadic codes. Related to [Leander-Gauthier Umana; SCC2010]}
2. A naive attack

W.l.o.g. we can assume that $G$ is systematic in its $k$ first positions.

$$G = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$
Step 1 – expressing the $Y_iX_i^d$'s in terms of the $Y_jX_j^d$'s for $j \in \{k + 1, \ldots, n\}$.

$P = (p_{ij})_{1 \leq i \leq k, k + 1 \leq j \leq n}$. We can rewrite (2) as

\[
\begin{align*}
Y_i &= \sum_{j=k+1}^{n} p_{i,j} Y_j \\
Y_iX_i &= \sum_{j=k+1}^{n} p_{i,j} Y_j X_j \\
\phantom{Y_iX_i} &\vphantom{=} \cdots \\
Y_iX_i^{t-1} &= \sum_{j=k+1}^{n} p_{i,j} Y_j X_j^{t-1}
\end{align*}
\]

for all $i \in \{1, \ldots, k\}$. 
Step 2.– Exploiting $Y_i(Y_i X_i^2) = (Y_i X_i)^2$

\[
\begin{align*}
    Y_i & = \sum_{j=k+1}^{n} p_{i,j} Y_j \\
    Y_i X_i & = \sum_{j=k+1}^{n} p_{i,j} Y_j X_j \\
    Y_i X_i^2 & = \sum_{j=k+1}^{n} p_{i,j} Y_j X_j^2 \\
\end{align*}
\]

\[\Rightarrow \left( \sum_{j=k+1}^{n} p_{i,j} Y_j \right) \left( \sum_{j=k+1}^{n} p_{i,j} Y_j X_j^2 \right) = \left( \sum_{j=k+1}^{n} p_{i,j} Y_j X_j \right)^2\]

\[\Rightarrow \sum_{j=k+1}^{n} \sum_{j' > j} p_{i,j} p_{i,j'} \left( Y_j Y_{j'} X_{j'}^2 + Y_{j'} Y_j X_j^2 \right) = 0\]
Step 3. – Linearization

\[ Z_{jj'} \overset{\text{def}}{=} Y_j Y_{j'} X_j^2 + Y_{j'} Y_j X_j^2 \]

\[ \sum_{j=k+1}^{n} \sum_{j'>j} p_{i,j} p_{i,j'} Z_{jj'} = 0. \]

- \( (n-k) \approx \frac{m^2 t^2}{2} \) unknowns
- \( k = n - mt \) equations

⇒ reveals \( Z_{jj'} \) when \( n - mt \geq \frac{m^2 t^2}{2} \) ?

- This happens for the Courtois-Finiasz-Sendrier scheme, ex: \( n = 2^{21}, t = 10, m = 21 \) which has to choose small values of \( t \).
Naive attack

This approach always fails...

$D_{\text{alternant}}$, resp. $D_{\text{Goppa}}$ dimension of the linear solution space when $G$ is the generator matrix of an alternant code, resp. Goppa code.

**Experimental fact 1.** Let $D_{\text{rand}} \overset{\text{def}}{=} \left( \frac{mt}{2} \right) - k$, with high probability

$$D_{\text{alternant}} = \max \left( D_{\text{rand}}, \frac{m(t-1)}{2} \left\{ (2\ell + 1)t - 2q^{\ell+1-1} \right\} \right)$$

for $\ell \overset{\text{def}}{=} \left\lfloor \log_q (t-1) \right\rfloor$

$$D_{\text{Goppa}} = D_{\text{alternant}} = \max \left( D_{\text{rand}}, \frac{m(t-1)(t-2)}{2} \right) \text{ for } t < q - 1$$

$$D_{\text{Goppa}} = \max \left( D_{\text{rand}}, \frac{mt}{2} \left\{ (2\ell + 1)t - 2q^\ell + 2q^{\ell-1} - 1 \right\} \right),$$

for $t \geq q - 1$ and with $\ell$ s.t. $q^\ell - 2q^{\ell-1} + q^{\ell-2} < t \leq q^{\ell+1} - 2q^\ell + q^{\ell-1}$
Naive attack

Table 1: $q = 2$ and $m = 10$

<table>
<thead>
<tr>
<th>$t$</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\binom{mt}{2}$</td>
<td>435</td>
<td>780</td>
<td>1225</td>
<td>1770</td>
<td>2415</td>
<td>3160</td>
<td>4005</td>
<td>4950</td>
<td>5995</td>
</tr>
<tr>
<td>$k$</td>
<td>994</td>
<td>984</td>
<td>974</td>
<td>964</td>
<td>954</td>
<td>944</td>
<td>934</td>
<td>924</td>
<td>914</td>
</tr>
<tr>
<td>$D_{rand}$</td>
<td>0</td>
<td>0</td>
<td>251</td>
<td>806</td>
<td>1461</td>
<td>2216</td>
<td>3071</td>
<td>4026</td>
<td>5081</td>
</tr>
<tr>
<td>$D_{alternant}$</td>
<td>30</td>
<td>90</td>
<td>251</td>
<td>806</td>
<td>1461</td>
<td>2216</td>
<td>3071</td>
<td>4026</td>
<td>5081</td>
</tr>
<tr>
<td>$T_{alternant}$</td>
<td>30</td>
<td>90</td>
<td>220</td>
<td>400</td>
<td>630</td>
<td>910</td>
<td>1320</td>
<td>1800</td>
<td>2350</td>
</tr>
<tr>
<td>$D_{Goppa}$</td>
<td>180</td>
<td>380</td>
<td>700</td>
<td>1110</td>
<td>1610</td>
<td>2216</td>
<td>3071</td>
<td>4026</td>
<td>5081</td>
</tr>
<tr>
<td>$T_{Goppa}$</td>
<td>180</td>
<td>380</td>
<td>700</td>
<td>1110</td>
<td>1610</td>
<td>2200</td>
<td>2970</td>
<td>3850</td>
<td>4840</td>
</tr>
</tbody>
</table>
3. A Distinguisher

\[ D_{\text{Goppa}} \geq D_{\text{alternant}} \geq D_{\text{rand}} \]

Table 2: \( t_{\text{min}} = \) smallest degree of the Goppa polynomial \( \Gamma \) for which we can not distinguish a binary Goppa code from a random binary linear code when \( n = 2^m \).

<table>
<thead>
<tr>
<th>( m )</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_{\text{min}} )</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>16</td>
<td>20</td>
<td>26</td>
<td>34</td>
<td>47</td>
<td>62</td>
<td>85</td>
<td>114</td>
<td>157</td>
<td>213</td>
</tr>
</tbody>
</table>
An explanation for the distinguisher

We have used

\[ Y_i Y_i X_i^2 = (Y_i X_i)^2 \]

Any identity of the form

\[ Y_i X_i^a Y_i X_i^b = Y_i X_i^c Y_i X_i^d \]

with \( a, b, c, d \in \{0, 1, \ldots, t - 1\} \) such that \( a + b = c + d \) would do the same job:

\[
Z_{jj'}^{a,b,c,d} \overset{\text{def}}{=} Y_j X_j^a Y_{j'} X_{j'}^b + Y_{j'} X_{j'}^a Y_j X_j^b + Y_j X_j^c Y_{j'} X_{j'}^d + Y_{j'} X_{j'}^c Y_j X_j^d
\]

\[
\sum_{j=k+1}^{n} \sum_{j'>j} p_{i,j} p_{i,j'} Z_{jj'}^{a,b,c,d} = 0
\]
Conclusion

- Combinatorial explanation of the distinguisher in the alternant case. Partial combinatorial explanation in the Goppa case.

- A slightly better distinguisher can be obtained by taking the subcode of codewords of even weights.

- Distinguisher $\Rightarrow$ attack?

- Approach requires $\frac{k}{n}$ very close to 1. Should very high rates be avoided in a McEliece like scheme?