A distinguisher for high-rate McEliece Cryptosystems

J.C. Faugère (INRIA, SALSA project),
Valérie Gauthier (Math. dep. Tech. Univ. of Denmark),
A. Otmani (Université Caen- INRIA, SECRET project),
L. Perret (INRIA, SALSA project),
J.-P. Tillich (INRIA, SECRET project)

May 12th, 2011
1. (Generalized) McEliece Cryptosystem $\text{McE}(\mathcal{K}_{n,k,t})$

$C$ a $q$-ary, length $n$, dimension $k$, $t$-error correcting code

- Public key: $G$ a $k \times n$ generator matrix of $C$ in $\mathcal{K}(n, k, t)$
- Secret key: $\Psi$ a $t$-error correcting procedure for $C$
- Encryption: $x \rightarrow xG + e$ with $e$ of Hamming weight $t$
- Decryption: $y \rightarrow \Psi(y)G^{-1}$ with $G^{-1}$ a right inverse of $G$. 
Introduction

Alternant codes/Goppa codes

\[ \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q_m}^n \text{ with } x_i \neq x_j \text{ if } i \neq j \]

\[ \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{F}_{q_m}^n \text{ with } y_i \neq 0 \]

For any \( r < n \), let
\[
\mathbf{H}_r(\mathbf{x}, \mathbf{y}) \overset{\text{def}}{=} \begin{pmatrix}
    y_1 & y_2 & \cdots & y_n \\
    y_1x_1 & y_2x_2 & \cdots & y_nx_n \\
    \vdots & \vdots & \ddots & \vdots \\
    y_1x_1^{r-1} & y_2x_2^{r-1} & \cdots & y_nx_n^{r-1}
\end{pmatrix}
\]

**Definition 1.** An **alternant code** is the kernel of an \( \mathbf{H} \) of this type

\[
\mathcal{A}_r(\mathbf{x}, \mathbf{y}) = \{ \mathbf{v} \in \mathbb{F}_{q}^n | \mathbf{H}_r(\mathbf{x}, \mathbf{y})\mathbf{v}^T = \mathbf{0} \}.
\]

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

Proposition 1. [decoding alternant codes] \( r/2 \) 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 \( (q = 2) \), \( r \) errors can be decoded in polynomial time, if \( x \) and \( \Gamma \) are known and if \( \Gamma \) has only simple roots.

More generally a factor \( \frac{q}{q-1} \) can be gained (exploited for instance in wild McEliece [Bernstein-Lange-Peters 2010]) by a suitable choice of \( \Gamma \).
(public key) 2. Distinguisher problem

\( \mathcal{K}^{\text{Goppa}}(n, k, t) \) the ensemble of generator matrices of \( t \)-error correcting Goppa codes of length \( n \), dimension \( k \)

\( \mathcal{K}^{\text{alt}}(n, k) \) the ensemble of generator matrices of alternant codes of length \( n \), dimension \( k \)

\( \mathcal{K}^{\text{lin}}(n, k) \) the ensemble of generator matrices of linear codes of length \( n \) and dimension \( k \).

Can we distinguish between the cases

(i) \( G \in \mathcal{K}^{\text{Goppa}}(n, k, t) \)

(ii) \( G \in \mathcal{K}^{\text{alt}}(n, k) \)

(iii) \( G \in \mathcal{K}^{\text{lin}}(n, k) \) ?
Niederreiter Nied ($\mathcal{K}_{n,k,t}$)

$C$ a $q$-ary, length $n$, dimension $k$, $t$-error correcting code.

- **Public key:** $H$ a $(n - k) \times n$ parity check matrix of $C$, $H \in \mathcal{K}_{n,k,t}$
- **Secret key:** $\Psi$ a $t$-error correcting procedure for $C$
- **Encryption:** $e \rightarrow eH^T$ with $e$ of Hamming weight $t$
- **Decryption:** To decipher $s$, choose any $y$ of syndrome $s$, i.e. such that $s = yH^T$, and output $y - \Psi(y)$. 

A probabilistic model of an attacker

A \((T, \epsilon)\) adversary \(A\) for \(\text{Nied}(\mathcal{K}_{n,k,t})\) is a program which runs in time at most \(T\) and is such that

\[
\text{Prob}_{H,e}(A(H, eH^T) = e| H \in \mathcal{K}_{n,k,t}) \geq \epsilon
\]

Most attacks actually deal with an adversary for \(\text{Nied}(\mathcal{K}^{\text{lin}}(n, k))\) instead of \(\text{Nied}(\mathcal{K}^{\text{Goppa}}(n, k, t))\).
How the distinguisher appears

\[
\text{Adv} \triangleq \text{Prob}(\mathcal{A}(H, eH^T) = e \mid H \in \mathcal{K}_n^{\text{Goppa}}) - \text{Prob}(\mathcal{A}(H, eH^T) = e \mid H \in \mathcal{K}_n^{\text{lin}})
\]

**Distinguisher \( D \):**

- **input** \( H \in \mathbb{F}_q^{(n-k)\times n} \)
- **Step 1**: pick a random \( e \in \mathbb{F}_q^n \) of weight \( t \)
- **Step 2**: if \( \mathcal{A}(H, eH^T) = e \) then return 1, else return 0.

Advantage of \( D = \|\text{Adv}\| \).
Either a decoding algorithm on linear codes or a distinguisher for Goppa codes

Proposition 3. If \( \exists (T, \epsilon) \)-adversary against \( \text{Nied}(\mathcal{K}_{n,k,t}^{Goppa}) \), then there exists either

(i) a \( (T, \epsilon/2) \)-adversary against \( \text{Nied}(\mathcal{K}_{n,k}^{\text{lin}}) \) (i.e. a decoder for general linear codes working in time \( T \) with success probability at \( \geq \epsilon/2 \)).

(ii) A distinguisher between \( H \in \mathcal{K}_{n,k,t}^{Goppa} \) and \( H \in \mathcal{K}_{n,k}^{\text{lin}} \) working in time \( T + O(n^2) \) and with advantage at least \( \epsilon/2 \).
3. Algebraic approach for attacking the McEliece cryptosystem

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

What we also know: $\exists x, y \in \mathbb{F}_{q^m}^n$ s.t.

$$H_r(x, y)G^T = 0. \quad (1)$$

What we want to find: 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

\( H_r(x, y)G^T = 0 \) translates to

\[
\begin{align*}
&g_{1,1}Y_1 + \cdots + g_{1,n}Y_n = 0 \\
&\vdots \\
&g_{k,1}Y_1 + \cdots + g_{k,n}Y_n = 0 \\
&g_{1,1}X_1 + \cdots + g_{1,n}X_n = 0 \\
&\vdots \\
&g_{k,1}X_1 + \cdots + g_{k,n}X_n = 0 \\
&g_{1,1}X_1^{r-1} + \cdots + g_{1,n}X_n^{r-1} = 0 \\
&\vdots \\
&g_{k,1}X_1^{r-1} + \cdots + g_{k,n}X_n^{r-1} = 0
\end{align*}
\]

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

**Proposition 4.** 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-Barreto 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]
4. A naive attack

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

\[
G = \begin{pmatrix}
1 & \cdots & 0 \\
0 & \cdots & 1 \\
\end{pmatrix}
\]
Step 1 – expressing the $Y_i X_i^d$’s in terms of the $Y_j X_j^d$’s for $j \in \{k + 1, \ldots, n\}$.

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

$$\begin{cases} 
Y_i &= \sum_{j=k+1}^{n} p_{i,j} Y_j \\
Y_i X_i &= \sum_{j=k+1}^{n} p_{i,j} Y_j X_j \\
\vdots \\
Y_i X_i^{r-1} &= \sum_{j=k+1}^{n} p_{i,j} Y_j X_j^{r-1}
\end{cases}$$

(3)

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

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

\( (4) \)

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

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

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

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

\(\binom{n-k}{2} \approx \frac{m^2 r^2}{2}\) unknowns

\(k = n - mr\) equations

\(\Rightarrow\) reveals \(Z_{jj'}\) when \(n - mr \geq \frac{m^2 r^2}{2}\)?

\(\Rightarrow\) This happens for the Courtois-Finiasz-Sendrier scheme, ex: \(n = 2^{21}, r = 10, m = 21\) which has to choose small values of \(r\).
Definition 2. Assume that the public key $G$ of the McEliece cryptosystem is in systematic form $(I_k \mid P)$

The linearized system associated to $G$ is

$$
\begin{align*}
\sum_{j=k+1}^{n} \sum_{j'>j} p_1,j p_1,j' Z_{jj'} &= 0 \\
\sum_{j=k+1}^{n} \sum_{j'>j} p_2,j p_2,j' Z_{jj'} &= 0 \\
\vdots \\
\sum_{j=k+1}^{n} \sum_{j'>j} p_k,j p_k,j' Z_{jj'} &= 0
\end{align*}
$$

The dimension of the solution space is denoted by $D$. 
Algebraic Distinguisher

Solving this system requires that

- Number of equations $k$ is greater than the number of unknowns $\binom{n-k}{2}$
- rank is (almost) equal to the number of unknowns

If $G$ is random then one would expect that the rank is $\min\left\{k, \binom{n-k}{2}\right\}$

$$\implies D = \max\left\{0, \binom{n-k}{2} - k\right\}$$

But for several structured (Goppa, alternant) codes rank $< \min\left\{k, \binom{n-k}{2}\right\}$

and this defect can be quantified
**Example** \( q = 2 \) and \( m = 14 \)

<table>
<thead>
<tr>
<th>( r )</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>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>((n-k))</td>
<td>861</td>
<td>1540</td>
<td>2415</td>
<td>3486</td>
<td>4753</td>
<td>6216</td>
<td>7875</td>
<td>9730</td>
<td>11781</td>
<td>14028</td>
<td>16471</td>
<td>19110</td>
</tr>
<tr>
<td>( k )</td>
<td>16342</td>
<td>16328</td>
<td>16314</td>
<td>16300</td>
<td>16286</td>
<td>16272</td>
<td>16258</td>
<td>16244</td>
<td>16230</td>
<td>16216</td>
<td>16202</td>
<td>16188</td>
</tr>
<tr>
<td>( D_{\text{rand}} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>269</td>
<td>2922</td>
</tr>
<tr>
<td>( D_{\text{alternant}} )</td>
<td>42</td>
<td>126</td>
<td>308</td>
<td>560</td>
<td>882</td>
<td>1274</td>
<td>1848</td>
<td>2520</td>
<td>3290</td>
<td>4158</td>
<td>5124</td>
<td>6188</td>
</tr>
<tr>
<td>( D_{\text{Goppa}} )</td>
<td>252</td>
<td>532</td>
<td>980</td>
<td>1554</td>
<td>2254</td>
<td>3080</td>
<td>4158</td>
<td>5390</td>
<td>6776</td>
<td>8316</td>
<td>10010</td>
<td>11858</td>
</tr>
</tbody>
</table>
Example $q = 2$ and $m = 14$

<table>
<thead>
<tr>
<th>$r$</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\binom{n-k}{2}$</td>
<td>21945</td>
<td>24976</td>
<td>28203</td>
<td>31626</td>
<td>35245</td>
<td>39060</td>
<td>43071</td>
<td>47278</td>
<td>51681</td>
<td>56280</td>
<td>61075</td>
<td>66066</td>
<td>71253</td>
</tr>
<tr>
<td>$k$</td>
<td>16174</td>
<td>16160</td>
<td>16146</td>
<td>16132</td>
<td>16118</td>
<td>16104</td>
<td>16090</td>
<td>16076</td>
<td>16062</td>
<td>16048</td>
<td>16034</td>
<td>16020</td>
<td>16006</td>
</tr>
<tr>
<td>$D_{\text{rand}}$</td>
<td>5771</td>
<td>8816</td>
<td>12057</td>
<td>15494</td>
<td>19127</td>
<td>22956</td>
<td>26981</td>
<td>31202</td>
<td>35619</td>
<td>40232</td>
<td>45041</td>
<td>50046</td>
<td>55247</td>
</tr>
<tr>
<td>$D_{\text{alternant}}$</td>
<td>7350</td>
<td>8816</td>
<td>12057</td>
<td>15494</td>
<td>19127</td>
<td>22956</td>
<td>26981</td>
<td>31202</td>
<td>35619</td>
<td>40232</td>
<td>45041</td>
<td>50046</td>
<td>55247</td>
</tr>
<tr>
<td>$D_{\text{Goppa}}$</td>
<td>13860</td>
<td>16016</td>
<td>18564</td>
<td>21294</td>
<td>24206</td>
<td>27300</td>
<td>30576</td>
<td>34034</td>
<td>37674</td>
<td>41496</td>
<td>45500</td>
<td>50046</td>
<td>55247</td>
</tr>
</tbody>
</table>
Alternant Case

Let $\ell \overset{\text{def}}{=} \lfloor \log_q(r - 1) \rfloor$. 

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

as long as $\binom{n-k}{2} - D_{\text{alternant}} < k$. 
Let $\ell$ the unique integer such that $q^\ell - 2q^{\ell-1} + q^{\ell-2} < r \leq q^{\ell+1} - 2q^\ell + q^{\ell-1}$

$$D_{\text{Goppa}} = \begin{cases} \frac{1}{2}m(r - 1)(r - 2) = D_{\text{alternant}} & \text{for } r < q - 1 \\ \frac{1}{2}mr((2\ell + 1)r - 2q^\ell + 2q^{\ell-1} - 1) & \text{for } r \geq q - 1 \end{cases}$$

as long as $\binom{n-k}{2} - D_{\text{Goppa}} < k$. 
Example $q = 2$ and $m = 14$

<table>
<thead>
<tr>
<th>$r$</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>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(n-k)$</td>
<td>861</td>
<td>1540</td>
<td>2415</td>
<td>3486</td>
<td>4753</td>
<td>6216</td>
<td>7875</td>
<td>9730</td>
<td>11781</td>
<td>14028</td>
<td>16471</td>
<td>19110</td>
</tr>
<tr>
<td>$k$</td>
<td>16342</td>
<td>16328</td>
<td>16314</td>
<td>16300</td>
<td>16286</td>
<td>16272</td>
<td>16258</td>
<td>16244</td>
<td>16230</td>
<td>16216</td>
<td>16202</td>
<td>16188</td>
</tr>
<tr>
<td>$D_{\text{rand}}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>269</td>
</tr>
<tr>
<td>$D_{\text{alternant}}$</td>
<td>42</td>
<td>126</td>
<td>308</td>
<td>560</td>
<td>882</td>
<td>1274</td>
<td>1848</td>
<td>2520</td>
<td>3290</td>
<td>4158</td>
<td>5124</td>
<td>6188</td>
</tr>
<tr>
<td>$T_{\text{alternant}}$</td>
<td>42</td>
<td>126</td>
<td>308</td>
<td>560</td>
<td>882</td>
<td>1274</td>
<td>1848</td>
<td>2520</td>
<td>3290</td>
<td>4158</td>
<td>5124</td>
<td>6188</td>
</tr>
<tr>
<td>$D_{\text{Goppa}}$</td>
<td>252</td>
<td>532</td>
<td>980</td>
<td>1554</td>
<td>2254</td>
<td>3080</td>
<td>4158</td>
<td>5390</td>
<td>6776</td>
<td>8316</td>
<td>10010</td>
<td>11858</td>
</tr>
<tr>
<td>$T_{\text{Goppa}}$</td>
<td>252</td>
<td>532</td>
<td>980</td>
<td>1554</td>
<td>2254</td>
<td>3080</td>
<td>4158</td>
<td>5390</td>
<td>6776</td>
<td>8316</td>
<td>10010</td>
<td>11858</td>
</tr>
</tbody>
</table>
Example $q = 2$ and $m = 14$

<table>
<thead>
<tr>
<th>$r$</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\binom{n-k}{2}$</td>
<td>21945</td>
<td>24976</td>
<td>28203</td>
<td>31626</td>
<td>35245</td>
<td>39060</td>
<td>43071</td>
<td>47278</td>
<td>51681</td>
<td>56280</td>
<td>61075</td>
<td>66066</td>
<td>71253</td>
</tr>
<tr>
<td>$k$</td>
<td>16174</td>
<td>16160</td>
<td>16146</td>
<td>16132</td>
<td>16118</td>
<td>16104</td>
<td>16090</td>
<td>16076</td>
<td>16062</td>
<td>16048</td>
<td>16034</td>
<td>16020</td>
<td>16006</td>
</tr>
<tr>
<td>$D_{\text{rand}}$</td>
<td>5771</td>
<td>8816</td>
<td>12057</td>
<td>15494</td>
<td>19127</td>
<td>22956</td>
<td>26981</td>
<td>31202</td>
<td>35619</td>
<td>40232</td>
<td>45041</td>
<td>50046</td>
<td>55247</td>
</tr>
<tr>
<td>$D_{\text{alternant}}$</td>
<td>7350</td>
<td>8816</td>
<td>12057</td>
<td>15494</td>
<td>19127</td>
<td>22956</td>
<td>26981</td>
<td>31202</td>
<td>35619</td>
<td>40232</td>
<td>45041</td>
<td>50046</td>
<td>55247</td>
</tr>
<tr>
<td>$T_{\text{alternant}}$</td>
<td>7350</td>
<td>8610</td>
<td>10192</td>
<td>11900</td>
<td>13734</td>
<td>15694</td>
<td>17780</td>
<td>19992</td>
<td>22330</td>
<td>24794</td>
<td>27384</td>
<td>30100</td>
<td>32942</td>
</tr>
<tr>
<td>$D_{\text{Goppa}}$</td>
<td>13860</td>
<td>16016</td>
<td>18564</td>
<td>21294</td>
<td>24206</td>
<td>27300</td>
<td>30576</td>
<td>34034</td>
<td>37674</td>
<td>41496</td>
<td>45500</td>
<td>50046</td>
<td>55247</td>
</tr>
<tr>
<td>$T_{\text{Goppa}}$</td>
<td>13860</td>
<td>16016</td>
<td>18564</td>
<td>21294</td>
<td>24206</td>
<td>27300</td>
<td>30576</td>
<td>34034</td>
<td>37674</td>
<td>41496</td>
<td>45500</td>
<td>49686</td>
<td>54054</td>
</tr>
</tbody>
</table>
Simplified Formulas for binary Goppa Codes

Let $\ell \overset{\text{def}}{=} \lceil \log_2 r \rceil + 1$.

$$D_{\text{Goppa}} = \frac{1}{2}mr \left( (2\ell + 1)r - 2^{\ell} - 1 \right)$$

as long as $\binom{mr}{2} - D_{\text{Goppa}} < n - mr$. 

Binary Goppa Codes

In particular, assuming that $n = 2^m$, the binary Goppa code distinguishing problem is solved for any $r < r_{\text{max}}$

<table>
<thead>
<tr>
<th>$m$</th>
<th>8</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>
<th>22</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{max}}$</td>
<td>5</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>
<td>290</td>
<td>400</td>
</tr>
</tbody>
</table>

- $m = 13$ and $r = 19$ corresponds to a 90-bit security McEliece public key.
- All CFS parameters fits in the range of validity of the algebraic distinguisher.
5. Explanation

- Formulas obtained through experimentations for random codes, alternant codes and irreducible Goppa codes over fields of size $q \in \{2, 4, 8, 16\}$.

- We have an explanation for alternant codes and binary Goppa codes by guessing a basis of the solution vector space over $\mathbb{F}_q$.

- It does not provide a proof.
Explanation for Alternant Codes – Step I

- Note that the entries of the system are in $\mathbb{F}_q$ and solutions are sought in $\mathbb{F}_{q^m}$.
- Let us view $\mathbb{F}_{q^m}$ as a $\mathbb{F}_q$-vector space of dimension $m$, and let $\pi_i : \mathbb{F}_{q^m} \to \mathbb{F}_q$ be the function giving the $i$-th coordinate.
- Hence, if a vector $\mathbf{v}$ with $v_j \in \mathbb{F}_{q^m}$ is a solution then $\pi_i(\mathbf{v}) = \left( \pi_i(v_j) \right)_j$ whose entries are in $\mathbb{F}_q$ is also a solution.

$\implies$ Any solution with entries over $\mathbb{F}_{q^m}$ would potentially provide a basis of $m$ solutions with entries over $\mathbb{F}_q$. 
Explanation for Alternant Codes – Step II

- We have used \( Y_iY_iX_i^2 = (Y_iX_i)^2 \) which leads to:

\[
\forall i \in \{1, \ldots, k\}, \quad \sum_{j=k+1}^{n} \sum_{j' > j} p_{i,j}p_{i,j'}Y_jY_{j'} (X_j^2 + X_{j'}^2) = 0
\]

- But we can use any relation \( Y_iX_i^aY_iX_i^b = Y_iX_i^cY_iX_i^d \) with \( a, b, c, d \) in \( \{0, \ldots, r-1\} \) such that \( a + b = c + d \)

\[
\sum_{j=k+1}^{n} \sum_{j' > j} p_{i,j}p_{i,j'}Y_jY_{j'}(X_j^aX_{j'}^b + X_j^bX_{j'}^a + X_j^cX_{j'}^d + X_j^dX_{j'}^c) = 0
\]
For $r \geq q$, the automorphism $x \mapsto x^{q\ell}$ for any $0 \leq \ell \leq m - 1$ can be used.

\[ \forall e \in \{0, \ldots, r - 1\}, \quad Y_i X_i^e = \sum_{j=k+1}^{n} p_{ij} Y_j X_j^e \implies Y_i^q X_i^{eq} = \sum_{j=k+1}^{n} p_{ij} Y_j^q X_j^{eq} \]

We therefore can use the same trick, for instance \( Y_i(Y_i X_i)^q = Y_i^q Y_i X_i^q \),

\[ \sum_{j=k+1}^{n} \sum_{j' > j} p_{i,j} p_{i,j'} \left( Y_j Y_{j'}^q X_{j'}^q + Y_{j'} Y_j^q X_j^q + Y_j^q Y_{j'} X_{j'}^q + Y_{j'}^q Y_j X_j^q \right) = 0. \]
Explanation for Alternant Codes

However the equations obtained \((Y_i X_i^a Y_i X_i^b)^q = (Y_i X_i^c Y_i X_i^d)^q\) do not provide new solutions after decomposition over \(\mathbb{F}_q\) since they are linearly dependent of those obtained from \(Y_i X_i^a Y_i X_i^b = Y_i X_i^c Y_i X_i^d\).

Hence, we only consider equations obtained from integers \(a, b, c, d, \ell\) such that \(a + bq^\ell = c + dq^\ell\)

\[
Y_i X_i^a (Y_i X_i^b)^q^\ell = Y_i X_i^c (Y_i X_i^d)^q^\ell
\]

\(Z_{a,b,c,d,\ell} \overset{\text{def}}{=} \left( Y_j X_j^a Y_j^q^\ell X_j^{bq^\ell} + Y_j X_j^a Y_j^q^\ell X_j^{bq^\ell} + Y_j X_j^c Y_j^q^\ell X_j^{dq^\ell} + Y_j X_j^c Y_j^q^\ell X_j^{dq^\ell} \right)_{1 \leq j < j' \leq n-k} \)
Let us assume that $d > b$ and set $\delta \overset{\text{def}}{=} d - b$ and then $a = c + q^\ell \delta$

$$\implies Z_{a,b,c,d,\ell} = Z_{c+q^\ell \delta,b,c+b+\delta,\ell}$$

Let $B_r$ be the set $Z_{c+q^\ell \delta,b,c+b+\delta,\ell}$ obtained with $\delta = 1$ and satisfying:

$$\begin{cases} 
0 \leq b \leq r - 2 \text{ and } 0 \leq c \leq r - 1 - q^\ell & \text{if } 1 \leq \ell \leq \lceil \log_q(r - 1) \rceil \\
0 \leq b < c \leq r - 2 & \text{if } \ell = 0.
\end{cases}$$

**Proposition 5.**

- Any $Z_{c+q^\ell \delta,b,c+b+\delta,\ell}$ belongs to the $\mathbb{F}_{q^m}$-vector space generated by $B_r$

- The cardinality of $B_r$ with $r \geq 3$ is equal to $D/m$. 
For random choices of $x_i$'s and $y_i$'s defining the alternant code, the set
\[ \{ \pi_i(Z) \mid Z \in \mathcal{B}_r \text{ and } 1 \leq i \leq m \} \]
forms a basis of the vector space that is solution to the linearized system.
Conclusion

- Large dimension comes from the many different ways of combining the equations together yielding the same linearized system.
- What happens for random generator is proven now.
- Binary Goppa codes can also be explained but no explanation for non-binary Goppa codes.
- The most difficult task is identifying a basis of the vector space of solutions.
- 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?