

Finite Fields and Symmetric Cryptography

Andrea Caranti

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DEGLI STUDI DI TRENTO,
VIA SOMMARIVE 14, I-38050 POVO (TRENTO), ITALY

E-mail address: `caranti@science.unitn.it`

URL: `http://science.unitn.it/~caranti/`

Contents

Introduction	5
Why these notes are not in Italian	5
Goal	5
Notes from the author to the author	5
Chapter 1. Finite Fields	7
1.1. Background	7
1.2. Simple extension	7
1.3. Galois theory	8
1.4. Irreducible polynomials	9
1.5. Linear Functions	9
Chapter 2. The Moebius function	13
2.1. Definition	13
2.2. Moebius inversion	13
2.3. Irreducible polynomials	14
Chapter 3. Linear and affine functions	17
3.1. Linear and affine transformations	17
3.2. The dihedral groups	17
3.3. Cryptanalysis of linear transformations	18
3.4. The probability of generating a finite vector space	18
3.5. Cryptanalysis of affine transformations	19
Chapter 4. AES	21
4.1. Rijndael and AES	21
4.2. Generalities	21
4.3. S-boxes	22
Chapter 5. Nonlinearity	23
5.1. Correlation	23
5.2. Distance from affine functions	23
5.3. A scalar product	24
5.4. Parities	24
5.5. Parseval's identity	26
5.6. Inversion in a finite field	26
Chapter 6. Truncated differential cryptanalysis of AES	29
6.1. What we are trying to avoid	29

6.2. Notation and statement	29
6.3. AES	30
6.4. Proof	31
6.5. Additive subgroups of finite fields containing their inverses	32
6.6. Equations of degree two in characteristic two	33
Chapter 7. Codes	35
7.1. The singleton bound	35
7.2. Circulant matrices	35
7.3. Circulant matrices in characteristic two	35
Bibliography	37

Introduction

Why these notes are not in Italian

I started writing these notes for a course on “Campi finiti e crittografia simmetrica” I gave in Trento in 2005/06. Sandro Mattarei and I have written extensive notes of other courses, but this is the first time I write them up in English, and not in Italian. These notes will thus turn handy in 2009/10, when the start with the new *Laurea Magistrale* in English.

I did not do much in 2005/07, but in 2007/08 I wrote some more, and I plan to do the same in 2008/09, when I am writing this. This means that at time these notes might look disconnected, when I have written A but not B or, God forbid, B but not A.

Goal

The goal of the course is to give an overview of the basic theory of finite fields, and show the role they play in the construction of Rijndael/AES [DR02].

Notes from the author to the author

This is a *Things to do list*

- In Moebius, mention polynomials when speaking of the convolution product.
- Write up the singleton bound.
- Perhaps give the referee’s proof of you-know-what?

CHAPTER 1

Finite Fields

1.1. Background

A finite field E has order p^n , for a suitable prime p . Here p is the characteristic of E (and thus E contains the field $F = \mathbf{F}_p$ with p elements), and $|E : F| = n$.

Write $q = p^n$. The elements of E are roots of $f = x^q - x \in F[x]$.

Given p and n , a field of order q can be constructed as the splitting field of f over F . This is unique up to (F -)isomorphisms. Write $\mathbf{GF}(q)$ for it.

1.2. Simple extension

One proves, as in [Ser73], that a finite subgroup of the multiplicative group of a field is cyclic.

We first show that

$$(1.2.1) \quad n = \sum_{d|n} \varphi(d),$$

where φ is the Euler function.

One method is to *define* a function φ via these formulas, and show that it coincides with Euler's φ . This approach is similar to that for the Moebius function (see Chapter 2).

In another approach, we count the number $X(d)$ of elements of order d in a cyclic group $\langle a \rangle$ of order n . We first prove that the order of a^k (with $0 \leq k < n$, say) is

$$\frac{n}{(n, k)}.$$

This follows from the fact that if $(a^k)^t = 1$, then $n \mid kt$, so $n/(n, k) \mid t$ and $t = s \cdot n/(n, k)$. The smallest such t for $s = 1$ is $n/(n, k)$, and in fact $(a^k)^{n/(n, k)} = (a^{n/(n, k)})^k = 1$.

So the order of a^k is n when $(n, k) = 1$. Thus $X(n) = \varphi(n)$.

Now we prove the following strong converse of Lagrange's theorem for $\langle a \rangle$.

THEOREMA 1.2.1. *A cyclic group $\langle a \rangle$ of order n has exactly one subgroup of order d for $d \mid n$, and this subgroup is cyclic.*

PROOF. Clearly $\langle a^{n/d} \rangle$ is such a subgroup, as the element $a^{n/d}$ has order d . Now let H be a subgroup of $\langle a \rangle$ of order d . Let $a^k \in H$, so that $(a^k)^d = 1$. It follows that $n \mid kd$, so that $n/(n, d) \mid k$, and thus a^k is a power of $a^{n/d}$, and thus $a^k \in \langle a^{n/d} \rangle$, and $H \subseteq \langle a^{n/d} \rangle$. Because they have the same order d , they are thus equal. \square

Now to see what is $X(d)$, we note that if $x \in \langle a \rangle$ has order d , then $\langle x \rangle = \langle a^{n/d} \rangle$ by the theorem above, so $X(d)$ is the number of elements of order d in a cyclic group $\langle a^{n/d} \rangle$ of order d , and $X(d) = \varphi(d)$ as above.

Since clearly $n = \sum_{d|n} X(d)$ in *any* finite group of order n , we get (1.2.1).

THEOREMA 1.2.2. *Let G be a group of finite order n , with the property that if d divides n , then*

$$(1.2.2) \quad |\{a \in G : a^d = 1\}| \leq d.$$

then G is cyclic.

Note that (1.2.2) holds if G is a finite multiplicative subgroup of L^* , where L is a field. This is because the left-hand side is the set of the roots in L of the polynomial $x^d - 1$, of degree d .

PROOF. In fact, let $X(d)$ be the number of elements of G of order d , for $d | n$. It might be that $X(d) = 0$. If $X(d) \neq 0$, and $a \in G$ has order d , then the d elements of $\langle a \rangle$ are exactly the elements of the set in (1.2.2). Any $b \in G$ of order d must therefore lie in $\langle a \rangle$, as $b^d = 1$. Thus there are $\varphi(d)$ elements of order d in G here. We have

$$n = \sum_{d|n} \varphi(d) \leq \sum_{d|n} X(d) = n.$$

Therefore equality holds, $X(d) = \varphi(d)$ for all $d | n$, in particular $X(n) = \varphi(n) \neq 0$. \square

Going back to finite fields, it follows that E is a simple extension of F , that is, $E = F[\alpha]$ for some $\alpha \in E$. This follows from the fact E^* is cyclic, $E^* = \langle \alpha \rangle$, for some α .

Note that the minimal polynomial of α over F has degree n , as $|F[\alpha] : F| = |E : F| = n$, and it is irreducible. This over $F = \mathbf{F}_p$ there are irreducible polynomials of any degree. (Note that irreducible polynomials over \mathbf{R} have degree 1 or 2 only.)

Note that give a finite field E of order $q = p^n$, there might be elements α for which $E = F[\alpha]$, but that are not *primitive*, that is, they do not have order $q - 1$, or $\langle \alpha \rangle < E^*$, as the following example shows.

EXAMPLE 1.2.3. Let $p = 2$, so that $F = \mathbf{F}_2 = \{0, 1\}$. The polynomial $x^4 + x^3 + x^2 + x + 1 \in F[x]$ is irreducible. Thus α is one of its roots, $F[\alpha]$ is a field E of order 2^4 . But $\alpha^5 = 1$, as it follows from the identity $x^5 - 1 = (x - 1) \cdot (x^4 + x^3 + x^2 + x + 1)$.

1.3. Galois theory

Let $q = p^n$, and E and F as usual. Let $f = x^q - x \in F[x]$.

Since E/F is the splitting field of f , which has distinct roots, E/F is a Galois extension. Its Galois group G is cyclic, generated by the Frobenius morphism

$$\sigma(a) = a^p.$$

The subgroups of G are of the form $\langle \sigma^d \rangle$, for $d | n$. Such a subgroup is associated, via the Galois correspondence, to a subfield K such that $|K : F| =$

$|G| / |\langle \sigma^d \rangle| = d$, so that K is a field of order p^d . In fact by the Galois correspondence $K = \{a \in E : \sigma^d(a) = a\} = \{a \in E : a^{p^d} = a\} - a = 0$.

It follows that $\mathbf{GF}(p^d) \subseteq \mathbf{GF}(p^n)$ if and only if d divides n . This follows also (without any appeal to Galois theory) from the elementary fact that the polynomial $x^{p^d} - x$ divides $x^{p^n} - x$ iff $d \mid n$.

1.4. Irreducible polynomials

We have seen above that over $F = \mathbf{F}_p$ (field) there are irreducible polynomials of all degrees. This is true over any finite field E . In fact, let E have order p^n , and fix k . Consider the field L of order p^{nk} . It contains E , by what we have just seen, and $L = F[\alpha] = E[\alpha]$ for some α . Now

$$nk = |L : F| = |L : E| \cdot |E : F| = |E[\alpha] : E| \cdot n.$$

It follows that $|E[\alpha] : E| = k$, so that the minimal polynomial $f \in E[x]$ of α over E has degree k .

Using Galois theory, one can show that $f = x^q - x \in F[x]$ (for $q = p^n$) is the product of all irreducible polynomials in $F[x]$ of degree dividing n . In fact, it is a matter of degrees that if an irreducible polynomial divides $x^q - x$, then its degree divides n . Conversely, let $g \in F[x]$ be an irreducible polynomial of degree $d \mid n$. Let α be a root of g . Then $L = F[\alpha]$ is a field of order p^d , and as such is contained in E , the field of order p^n , which is the splitting field of f . Since E/F is a Galois extension, and contains a root α of the irreducible polynomial $g \in F[x]$, all the roots of g are in E , and they are distinct. Thus $g \mid x^q - x$, as required.

See Chapter 2 for a formula for the number of irreducible polynomials.

1.5. Linear Functions

Let $E = \mathbf{GF}(p^n)$, where p is a prime, and $F = \mathbf{GF}(p)$. There are p^{n^2} functions $E \rightarrow E$ which are linear (over F). The elements of the Galois group, that is the maps

$$\varphi^i : x \mapsto x^{p^i},$$

for $0 \leq i < n$ are all linear.

LEMMA 1.5.1 (Dedekind). *Let E be a field. If $\varphi_1, \dots, \varphi_m$ are distinct isomorphisms $E \rightarrow E$, then they are linearly independent over E .*

PROOF. Suppose $\sum_{i=1}^m a_i \varphi_i = 0$. May assume all $a_i \neq 0$ (so say $a_1 = 1$), and there is no nontrivial relation with some of the $a_i = 0$. Since $\varphi_1 \neq \varphi_2$, let x_0 be

such that $\varphi_1(x_0) \neq \varphi_2(x_0)$. We have, for all $x \in E$,

$$\begin{aligned} \sum_{i=1}^m a_i \varphi_i(x) &= 0 \\ \sum_{i=1}^m a_i \varphi_i(x_0 x) &= \sum_{i=1}^m a_i \varphi_i(x_0) \varphi(x) = 0 \\ \sum_{i=1}^m a_i \varphi_1(x_0) \varphi_i(x) &= 0, \end{aligned}$$

where the last one is obtained multiplying by $\varphi_1(x_0)$ the first one. Subtracting the last two, we get

$$\sum_{i=2}^m a_i (\varphi_1(x_0) - \varphi_i(x_0)) \varphi_i(x) = 0.$$

The first coefficient is zero, but the second one is not, as $a_i(\varphi_1(x_0) - \varphi_2(x_0)) \neq 0$. This contradicts our initial assumption. \square

Thus the φ^i are independent, and thus their linear combinations

$$\sum_{i=0}^{n-1} a_i \varphi^i : x \mapsto a_0 x + a_1 x^p + \cdots + a_{n-1} x^{p^{n-1}},$$

with $a_i \in E$ are distinct linear maps. And there are $|E|^n = p^{n^2}$ of them, so they are all the linear maps.

Which of these map $E \rightarrow F$? By Galois theory, we need

$$\begin{aligned} a_0 x + a_1 x^p + a_2 x^{p^2} + \cdots + a_{n-1} x^{p^{n-1}} &= \\ &= (a_0 x + a_1 x^p + a_2 x^{p^2} + \cdots + a_{n-1} x^{p^{n-1}})^p = \\ &= a_{n-1}^p x + a_0^p x^p + a_1^p x^{p^2} + \cdots + a_{n-2}^p x^{p^{n-1}}. \end{aligned}$$

for all x . Thus $a_1 = a_0^p$, $a_2 = a_1^p = a_0^{p^2}$, and thus $a_i = a_0^{p^i}$ so that the linear functions $E \rightarrow F$ are of the form

$$\sum_{i=0}^{n-1} a^{p^i} \varphi^i : x \mapsto ax + a^p x^p + \cdots + a^{p^{n-1}} x^{p^{n-1}} = \text{tr}(ax)$$

for $a \in E$. In fact there are $|E^*| = |E|$ of them.

Here $\text{tr} : E \rightarrow F$ is the trace map,

$$\text{tr}(x) = \sum_{g \in \text{Gal}(E/F)} g(x),$$

itself a linear function $E \rightarrow F$.

In another approach, start with the trace. By Dedekind's Lemma, it is nonzero, that is, it does not have constant value 0. In particular, if $c \neq 0$, then cx takes all the values in E as $x \in E$, so $x \rightarrow \text{tr}(cx)$ is also nonzero. It follows that all functions $E \rightarrow F$ given by $x \mapsto \text{tr}(ax)$ are distinct, for $a \in E$. This is because if

for all $x \in E$ we have $\text{tr}(ax) = \text{tr}(bx)$, that is, $\text{tr}((a - b)x) = 0$, and so $a - b = 0$. But these are precisely $|E|$ functions, and thus all of them.

CHAPTER 2

The Moebius function

2.1. Definition

Let \mathbf{N}^* be the set of positive integers. define a function $\mu : \mathbf{N}^* \rightarrow \mathbf{Z}$ via

$$(2.1.1) \quad \sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

- μ is uniquely determined by the previous formula, that is, there is a *unique* function μ that satisfies the formula for all n .

Clearly $\mu(1) = 1$. Proceed by induction on n . If the values $\mu(d)$ are known, for $d < n$, then the value of $\mu(n)$ is uniquely determined by (2.1.1).

- μ is multiplicative, that is $\mu(nm) = \mu(n) \cdot \mu(m)$ if $(n, m) = 1$.

Let $(n, m) = 1$. We may clearly assume $n, m > 1$. Assume by induction that $\mu(x)$ is multiplicative for $x < nm$. Now every divisor d of nm can be written as $d = ef$, where $e | n$ e $f | m$, thus $(e, f) = 1$, so that we have

$$\begin{aligned} -\mu(nm) &= \sum_{d|nm, d \neq nm} \mu(d) \\ &= \sum_{e, f, e|n, f|m, ef \neq nm} \mu(e)\mu(f) \\ &= \sum_{e, e|n} \mu(e) \sum_{f, f|m} \mu(f) - \mu(n)\mu(m) \\ &= -\mu(n)\mu(m) \end{aligned}$$

as $n, m > 1$.

- Show that

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \text{ is divisible by the square of a prime,} \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes.} \end{cases}$$

Proceeding by induction on k , one shows the formula to hold for $n = p^k$, where p is a prime. Then use the previous item μ is called the *Moebius function*.

2.2. Moebius inversion

We have the important

THEOREMA 2.2.1 (Moebius Inversion). *Let $f : \mathbf{N}^* \rightarrow \mathbf{C}$ be a function, and let*

$$g(n) = \sum_{d|n} f(d).$$

Then

$$f(n) = \sum_{d|n} \mu(n/d)g(d).$$

We compute

$$\sum_{d|n} \mu(n/d)g(d) = \sum_{d,e,de=n} \mu(e)g(d) = \sum_{d,e,de=n} \mu(e) \sum_{k|d} f(k).$$

What's the coefficient on $f(k)$ in the previous formula, for a given k ? It's

$$\sum_{d,e,de=n,k|d} \mu(e) = \sum_{e,e|n/k} \mu(e) = \begin{cases} 1 & \text{if } k = n, \\ 0 & \text{if } k \neq n. \end{cases}$$

A probably more enlightening approach requires to introduce the *convolution* of two functions f, g as $f * g(n) = \sum_{d,e,de=n} f(d)g(e)$. One shows it is commutative and associative. Define the constant function 1, which evaluates to 1 for all n , and the Dirac delta as

$$(2.2.1) \quad \delta_a(n) = \begin{cases} 1 & \text{if } n = a, \\ 0 & \text{if } n \neq a. \end{cases}$$

One shows

- $\delta_1 = 1 * \mu$ (this is the formula (2.1.1) that defines μ),
- $\delta_a * \delta_b = \delta_{ab}$,
- $\delta_1 * f = f$.

Finally, $\mu * (f * 1) = f * (\mu * 1) = f * \delta_1 = f$.

2.3. Irreducible polynomials

What is the number of irreducible polynomials of degree n over $F = \mathbf{F}_p$? Write $q = p^n$, and let E be the field with q elements.

We first count the *primitive* ones, that is, those whose roots have order $q - 1$. There are $\varphi(q - 1)$ such roots, and thus $\varphi(q - 1)/n$ such polynomials.

Now we note that $x^q - x$ is the product of all irreducible polynomials in $F[x]$ of degree d dividing n . Write $P(d)$ for the product of the irreducible polynomials of degree d . Then

$$(2.3.1) \quad x^q - x = \prod_{d|n} P(d).$$

Using a multiplicative version of Theorem 2.2.1, we obtain

$$P(n) = \prod_{d|n} (x^{p^d} - x)^{\mu(n/d)}.$$

For example,

$$P(6) = \frac{(x^{p^6} - x) \cdot (x^p - x)}{(x^{p^2} - x) \cdot (x^{p^3} - x)}.$$

The meaning is that we first take out from $x^{p^6} - x$ the irreducible polynomials of degree dividing 2 and 3, but then we have removed those of degree 1 twice, so we need to compensate.

As for the number of polynomials, write $Q(n)$ for the number of irreducible polynomials in $F[x]$ of degree n . Then (2.3.1) says

$$p^n = \prod_{d|n} Q(d),$$

and this by Theorem 2.2.1

$$Q(n) = \prod_{d|n} \mu(n/d) Q(d).$$

CHAPTER 3

Linear and affine functions

To be added: Very much to be written

3.1. Linear and affine transformations

To be added: Includes the one-dimensional case, the representation in matrices (notice how many parameters we need).

Let $V = V(n, F)$ be a vector space of dimension d over the field F . If $F = \mathbf{GF}(q)$, we write $V = V(n, q)$. We also think usually of $V = F^n$, if a basis is chosen implicitly or explicitly.

A(n F -)linear map $V \rightarrow V$ is the usual thing, while $f : V \rightarrow V$ is *affine* if $xf = xa + b$, where a is linear and $b \in V$. (Note that we write maps on the right here.) We write $f = f_{a,b}$. Thus $f_{1,0}$ is the identity map. Note that two such maps compose as

$$xf_{a,b} \circ f_{c,d} = (xa + b)f_{c,d} = xac + bc + d = xf_{ac, bc+d}.$$

Note that $f_{c,d}$ is the inverse of $f_{a,b}$ if $ac = 1$, so that a is invertible and $c = a^{-1}$, and $d = -ba^{-1}$.

3.2. The dihedral groups

Let A be either \mathbf{Z} or $\mathbf{Z}/n\mathbf{Z}$. For $a \in A$ invertible and $b \in B$, define $f_{a,b} : A \rightarrow A$ by $xf_{a,b} = xa + b$. Clearly these maps form a group. Consider the subset

$$\{ f_{a,b} : a \in \{1, -1\}, b \in A \}.$$

This is clearly a group, the *dihedral* group.

When $A = \mathbf{Z}$, this is the group of congruences of \mathbf{Z} . In fact for $a = 1$ we have the translations, while for $a = -1$ we have the reflections.

Note first that $f_{-1,0} : x \mapsto -x$ is the reflection around 0.

If $b = 2c$ is even, then

$$xf_{-1,b} = -x + 2c = -(x - c) + c,$$

so $f_{-1,b}$ is a reflection around the point c on \mathbf{Z} . If $b = 2c + 1$ is odd, then

$$xf_{-1,b} = -x + 2c + 1 = -(x - (c + \frac{1}{2})) + c + \frac{1}{2},$$

so it is a reflection around the non-integer $c + 1/2$.

For instance, in the first case, note that

$$f_{-1,b} = f_{1,c}^{-1} \circ f_{-1,0} \circ f_{1,c}.$$

So $f_{-1,b}$ is obtained by a change of coordinates (conjugation) from $f_{-1,0}$.

If $A = \mathbf{Z}/n\mathbf{Z}$, it is the group of congruences of a regular n -gon, whose vertices are labelled consecutively $0, 1, \dots, n-1$, where we mean classe modulo n . For $a = 1$ we have rotations, while for -1 we have reflections. If n is odd we have only one type of reflections, with respect to an axis that goes through a vertex and its opposed side. This is because 2 is invertible in A , so that $b = 2c$ for some c , and

$$xf_{-1,b} = -x + 2c = -(x - c) + c$$

is the reflection through the axis passing through the vertex c and the opposing side. (Note $cf_{-1,b} = c$) If $n = 2m$ is even, we have two cases. If $b = 2c$, then $f_{-1,b}$ is as above, a reflection through the axis going through the vertices c and $m + c$. (In fact we have also $(m + c)f_{-1,b} = -(m + c - c) + c = -m + c = m + c$) If $b = 2c + 1$, then $f_{-1,b}$ does not fix any vertex (check) and it is a reflection through an axis going through the middle points of two opposite sides.

In the case $A = \mathbf{Z}$, note that $f_{-1,0}$ and $f_{-1,1}$ are *involutions* (they have period two), while their product $f_{-1,0} \circ f_{-1,1} = f_{1,1}$ is a translation by 1 , and thus has infinite period. In the case $A = \mathbf{Z}/n\mathbf{Z}$, the composition is a rotation of $2\pi/n$, so it has period n .

3.3. Cryptanalysis of linear transformations

Our space is $V = V(d, q)$.

Suppose the encryption is done with a linear function $f = f_{a,0}$, and we the possibility of a *chosen plaintext* attack. This means we, the attacker, can choose which plaintexts to encrypt. Clearly we choose a basis e_i of V , which has d elements, and this allows us to reconstruct a , as $e_i f$ is the i -th row of a , regarded as a matrix.

In a *given plaintext* attack, we only observe $x_i f$, where the values of x_1, x_2, \dots are random. For this, we need the results of the next section.

3.4. The probability of generating a finite vector space

Let $V = V(n, q)$. Let v_i be a random sequence of elements of V . Eventually we will get a basis out of it. Here's a proof one can no doubt improve upon. Let $\varepsilon > 0$. The probability that v_1 is 0 is $1/q^n$. So the probability that after t_1 attempts we are always getting zero is $1/q^{nt_1}$. This is less than ε when

$$(3.4.1) \quad t_1 > -\frac{\log_q(\varepsilon)}{n}.$$

So with these many attempts we have a probability $1 - \varepsilon$ to have hit a nonzero vector. The probability that the next vector is dependent on the first one we have thus obtained is q/q^n . So we get that the probability of failing after t_2 further attempts is less than ε when

$$(3.4.2) \quad t_2 > -\frac{\log_q(\varepsilon)}{n-1}.$$

The result is that after

$$(3.4.3) \quad t_1 + t_2 + \dots + t_n > -\log_q(\varepsilon) \left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{2} + 1 \right)$$

attempts we have obtained a basis with probability at least $(1 - \varepsilon)^n > 1 - n\varepsilon$. Note that

$$\gamma = \lim_{n \rightarrow \infty} \left(\left(\sum_{k=1}^n \frac{1}{k} \right) - \log(n) \right)$$

is the *Euler-Mascheroni* constant ≈ 0.5772 .

3.5. Cryptanalysis of affine transformations

Here $f = f_{a,b}$.

In a chosen plaintext situation, we first compute $0f_{a,b} = b$, and then we are in the case of a linear function.

In the given plaintext situation, we start with a *differential cryptanalysis* approach. That is, given $u, v \in V$, we compute $uf - vf = (u - v)f_{a,0}$, where now $f_{a,0}$ is linear. So once we have enough vectors so that their pairwise differences form a basis for V , we are have found $f_{a,0}$, and thus we find easily b too.

There is only one thing to be noticed. Even if we have $n + 1$ vectors v_i , such that every subset of n vectors is a basis, it might be that their differences is not a basis. This happens when there is a relation of the form

$$\sum_{i=1}^{n+1} a_i v_i, \quad \text{with} \quad \sum_{i=1}^{n+1} a_i = 0.$$

To be added: to be polished

CHAPTER 4

AES

I have just started writing this chapter, so it is very tentative, and contains notes to myself for further development. Blanket reference is [DR02].

4.1. Rijndael and AES

Rijndael is the name of the cryptosystem Joan Daemen and Vincent Rijmen proposed for the Advanced Encryption System (AES) competition. A version of Rijndael was then adopted as AES.

To be added: Spell the differences

We will be discussing mainly AES, spelling out the differences to Rijndael occasionally.

4.2. Generalities

AES is a *symmetric* (or *private key*) cryptosystem. That is, the two parties who use it to communicate have to share a secret (private) key beforehand. A public key cryptosystem like RSA might be used to exchange the private key; from then on, AES is used.

AES is a *block* cryptosystem. A message is split up in blocks of 128 bits, and every block is processed separately.

AES must run also in situations in which memory and processing power are limited, such as on a smart card. To minimize the length of the code, AES is an *iterated* cryptosystem. That is, a simple function (which would not offer any security by itself, but which has a short code) is iterated several times, until security is achieved. (Think of the way we shuffle a deck of cards: we repeat several times a simple shuffling operation, until the cards are well mixed [Dia98].) This simple function is called a *round*.

Thus, when we submit a plaintext block to AES, it undergoes several transformations, until we get the final ciphertext. Each intermediate state in the process is called - ehm - the *state*. The plaintext, the ciphertext and all intermediate states are all elements of a vector space $V = V(2, 128)$ of dimension 128 over the field $F = \mathbf{F}_2$ with two elements.

To be added: Simple examples to illustrate?

In turn, every round is composed of several simpler functions. One of them involves the *round key*. Starting from the *master key*, one different key for each round is calculated. In round i , the round key k_i is simply used by addition, that is, the state x is mapped to $x + k_i$.

4.3. S-boxes

All components of AES but one are *affine* functions. (In cryptography, one often says *linear*, when a mathematician would say affine.) This one has to be a permutation γ of V , a set with 2^{128} elements. How does one write down such a beast? The idea is to split

$$V = \bigoplus_{i=1}^{16} V_i,$$

where each V_i has dimension 8. The permutation γ is then defined as

$$v\gamma = \sum_{i=1}^{16} v_i\gamma_i,$$

where $v = \sum_{i=1}^{16} v_i$, with $v_i \in V_i$, and γ_i is a permutation of V_i . Each V_i is taken to be the additive group of $\mathbf{GF}(2^8)$, and then one takes

$$u\gamma_i = \begin{cases} 0 & \text{if } u = 0, \\ u^{-1} & \text{otherwise.} \end{cases}$$

This function is chosen (see [Nyb94]) because it is very much non-linear, see Chapter 5.

To be added: To be continued

CHAPTER 5

Nonlinearity

5.1. Correlation

A rielaboration of [DR02, Chap. 7] in terms of group theory.

Let $V = \mathbf{F}_2^n$. Suppose you have two binary Boolean functions $f, g : V \rightarrow \mathbf{F}_2$. Their Hamming distance $d(f, g)$ is simply the number of points on which they differ. On average, you will expect them to coincide on about half of the points. That is, if you take two random Boolean functions, the bets are they will have the same value on half of the points, that is, the probability

$$1 - \frac{d(f, g)}{2^n}$$

that they coincide is likely to be $1/2$.

So you measure their *correlation* as

$$C(f, g) = 2 \cdot \left(1 - \frac{d(f, g)}{2^n}\right) - 1 = 1 - \frac{d(f, g)}{2^{n-1}}.$$

(We have used the linear function $\lambda : \mathbf{R} \rightarrow \mathbf{R}$ given by $\lambda(x) = 2x - 1$, which maps the interval $[0, 1]$ onto $[-1, 1]$.) So correlation 0 means they behave as your favourite pair of average random Boolean functions. Correlation 1 means $f = g$. Correlation -1 means that $g(x) = 1 + f(x)$ for all x , or g evaluates to 0 on the points where f evaluates to 1 and viceversa. Any correlation different from 0 means that the knowledge of f tells you some information about g as well.

We can reformulate the above as

$$d(f, g) = 2^{n-1}(1 - C(f, g)).$$

5.2. Distance from affine functions

We will be interested in the (minimum) distance of such a function f from the set of affine Boolean functions. Now an affine, nonlinear function $V \rightarrow \mathbf{F}_2$ is of the form $x \mapsto g(x) + 1$, with g linear. We have $d(f, g + 1) = 2^n - d(f, g) = 2^n - 2^{n-1} + 2^{n-1}C(f, g) = 2^{n-1}(1 + C(f, g))$. So the distance of f from affine functions is

$$\begin{aligned} \min \{ 2^{n-1}(1 \pm C(f, g)) : g \text{ linear} \} &= \\ &= \begin{cases} \{ 2^{n-1}(1 + \min C(f, g)) : g \text{ linear} \} & \text{if } C(f, g) \geq 0 \\ \{ 2^{n-1}(1 - \max C(f, g)) : g \text{ linear} \} & \text{if } C(f, g) < 0 \end{cases} \end{aligned}$$

Since $\min C(f, g) = -\max(-C(f, g))$, we get that the distance of f from affine functions is

$$(5.2.1) \quad 2^{n-1}(1 - \max \{ |C(f, g)| : g \text{ linear} \}).$$

5.3. A scalar product

If f is a Boolean function (that is, a function with values 0, 1), we may represent it via a \mathbf{C} -valued function $\hat{f}(x) = (-1)^{f(x)}$, that is, the function $\hat{f} : V \rightarrow \mathbf{C}$ such that

$$\hat{f}(x) = \begin{cases} 1 & \text{if } f(x) = 0, \\ -1 & \text{if } f(x) = 1. \end{cases}$$

(We will see soon why we might want to do that.) Clearly

$$\widehat{f+g}(x) = (-1)^{f(x)+g(x)} = (-1)^{f(x)}(-1)^{g(x)} = \hat{f}(x) \cdot \hat{g}(x).$$

We now define a scalar (better: Hermitian) product among functions $\varphi, \psi : V \rightarrow \mathbf{C}$ via

$$\langle \varphi, \psi \rangle = \frac{1}{|V|} \sum_{x \in V} \varphi(x) \overline{\psi(x)}.$$

This is bilinear (well, to be more precise, in the second variable...), and visibly nondegenerate: just note that

$$\langle \varphi, \varphi \rangle = \frac{1}{2^n} \sum_{x \in V} |\varphi(x)|^2 > 0,$$

if $\varphi \neq 0$. Consider now Boolean functions f, g . If $K = \{x \in V : f(x) \neq g(x)\}$, so that $d(f, g) = |K|$, we have

$$\begin{aligned} 2^n \langle \hat{f}, \hat{g} \rangle &= \sum_{x \in V} \hat{f}(x) \hat{g}(x) = \sum_{x \in V} (-1)^{f(x)+g(x)} \\ &= \sum_{x \in K} (-1) + \sum_{x \in V \setminus K} 1 = -d(f, g) + (2^n - d(f, g)) \\ &= 2^n - 2d(f, g) = 2^n C(f, g). \end{aligned}$$

In other words

$$C(f, g) = \langle \hat{f}, \hat{g} \rangle.$$

In particular the norm $\langle \hat{f}, \hat{f} \rangle = 1$ for Boolean functions f .

5.4. Parities

The dual space V^* of V is the vector space of all linear maps $V \mapsto \mathbf{F}_2$. A choice of a basis v_i on V gives a dual basis $v_i^* \in V^*$ (where $v_i^*(v_j) = \delta(i, j)$). We will write elements of V and V^* as vectors in \mathbf{F}_2^n , with respect to such a pair of bases.

If $w \in V^*$ (regarded, as we said, as an element of \mathbf{F}_2^n), as a linear function $V \rightarrow \mathbf{F}_2$ acts as $x \mapsto x \cdot w$ (where “ \cdot ” represent the row-by-row product of x and

w , where $x \in V$ is also regarded as an element of \mathbf{F}_2^n . Their hats are the functions called *parities*

$$\begin{aligned}\pi_w : V &\rightarrow \{1, -1\} \\ x &\mapsto (-1)^{x \cdot w}.\end{aligned}$$

Note that $\pi_x(y) = \pi_y(x)$.

We claim that every binary Boolean function (in the hat form), that is, every function $V \rightarrow \mathbf{C}$, can be written as a linear combination of these. Note that the space of functions $V \rightarrow \mathbf{C}$ is a vector space over \mathbf{C} of dimension 2^n over \mathbf{C} .

This is a special case of a rather more general fact from the theory of group representation theory and discrete Fourier transforms. Let G be a finite abelian (commutative) group. Then its (*linear*) *characters* are the group morphisms from G into the multiplicative group \mathbf{C}^* of the nonzero complex numbers.

If G is our V above, such a morphism φ satisfies

$$\varphi(x + y) = \varphi(x)\varphi(y),$$

thus we have for all $v \in V$

$$1 = \varphi(0) = \varphi(2v) = \varphi(v + v) = \varphi(v)\varphi(v) = \varphi(v)^2,$$

so that $\varphi(v) \in \{1, -1\}$. And in fact these characters are exactly the parities above. This is because a parity is a character, and conversely, if φ is a character, then we can write $\varphi(x) = (-1)^{\psi(x)}$ for a unique $\psi : V \rightarrow \mathbf{F}_2$, and ψ is linear, as

$$(-1)^{\psi(x+y)} = \varphi(x + y) = \varphi(x)\varphi(y) = (-1)^{\psi(x)}(-1)^{\psi(y)} = (-1)^{\psi(x)+\psi(y)}.$$

Note now that two distinct parities are orthogonal, as we have

$$\begin{aligned}\langle \pi_v, \pi_w \rangle &= \frac{1}{2^n} \sum_{x \in V} (-1)^{x \cdot v} (-1)^{x \cdot w} \\ &= \frac{1}{2^n} \sum_{x \in V} (-1)^{x \cdot (v+w)} \\ &= \delta(v, w).\end{aligned}$$

In fact, if $v = w$ all the summands are 1. If $u = v + w \neq 0$, the sum is zero instead. This is because $x \mapsto x \cdot u$ is a linear map $V \rightarrow \mathbf{F}_2$. It has value 0 on its kernel, which is a subspace of V of codimension one, and thus with $2^n/2$ elements, and value 1 on the remaining $2^n/2$ elements.

If these parities are orthogonal, they must be linearly independent, and thus they are a basis of the space of \mathbf{C} -valued functions on V . Let us find how to write any (hat) function in terms of the parities.

The coefficients will be of course just the scalar products of a \hat{f} with the parities, that is, the coefficient of \hat{f} with respect to the parity π_w will be

$$F(w) = C(f, \pi_w) = \left\langle \hat{f}, \pi_w \right\rangle = \frac{1}{2^n} \sum_{x \in V} (-1)^{f(x)+x \cdot w}.$$

This function $F : V^* \rightarrow \mathbf{C}$ is called the Walsh-Hadamard transform of f , or equivalently \hat{f} . It is a special case of a Fourier transform.

Reciprocally, we have

$$\begin{aligned}
C(F, \pi_y) &= \langle F, \pi_y \rangle = \frac{1}{2^n} \sum_{w \in V} F(w) \pi_y(w) \\
&= \frac{1}{2^n} \sum_{w \in V} F(w) (-1)^{y \cdot w} \\
&= \frac{1}{2^{2n}} \sum_{w \in V} \left((-1)^{y \cdot w} \sum_{x \in V} (-1)^{f(x) + x \cdot w} \right) \\
&= \frac{1}{2^{2n}} \sum_{x \in V} \left((-1)^{f(x)} \sum_{w \in V} (-1)^{(x+y) \cdot w} \right) \\
&= \frac{1}{2^n} \sum_{x \in V} ((-1)^{f(x)} \langle \pi_x, \pi_y \rangle) \\
&= \frac{1}{2^n} \sum_{x \in V} ((-1)^{f(x)} \delta(x, y)) = \frac{1}{2^n} \hat{f}(y).
\end{aligned}$$

So indeed

$$\hat{f}(y) = \sum_{w \in V} F(w) \pi_y(w)$$

a function is determined by its correlations (scalar products) with the parities.

5.5. Parseval's identity

We note Parseval's identity, which will turn handy later

$$\begin{aligned}
1 &= \langle \hat{f}, \hat{f} \rangle = \left\langle \sum_{v \in V} F(v) \pi_v, \sum_{w \in V} F(w) \pi_w \right\rangle = \\
&= \sum_{v, w \in V} F(v) F(w) \delta(v, w) = \sum_{v \in V} F(v)^2.
\end{aligned}$$

Now $F(v)$ is the correlation of f with the linear function π_v . So if $M = \max \{ |C(f, g)| : g \text{ linear} \}$, we obtain $1 \leq |V| M^2$, that is

$$(5.5.1) \quad M \geq \frac{1}{2^{n/2}}.$$

5.6. Inversion in a finite field

Now suppose $V = E = \mathbf{GF}(2^n)$. So each linear function $E \rightarrow \mathbf{F}_2$ is of the form $x \mapsto \text{tr}(ax)$ for a suitable $a \in E$.

What's the correlation of a (hat) function \hat{f} with such a linear function? It is

$$C(f, \text{tr}(a \cdot)) = \left\langle \hat{f}, \widehat{\text{tr}(a \cdot)} \right\rangle = \frac{1}{2^n} \sum_{x \in V} (-1)^{f(x) + \text{tr}(ax)}.$$

Consider the case when $f(x)$ is a component of inversion, that is, $f(x) = \text{tr}(bx^{-1})$ for some $b \in E^*$. Rescaling, we have to evaluate a *Kloosterman sum*

$$\sum_{x \in E^*} (-1)^{\text{tr}(x+bx^{-1})}$$

Using for instance [CU57] we get

$$\left| \sum_{x \in E^*} (-1)^{\text{tr}(x+bx^{-1})} \right| \leq 2 \cdot 2^{n/2},$$

that is,

$$|C(f, \text{tr}(a \cdot))| \leq \frac{2}{2^{n/2}},$$

which is within a factor 2 of the bound (5.5.1), and thus the distance of inversion from the affine functions is given, according to (5.2.1), by

$$2^{n-1}(1 - \max \{ |C(f, g)| : g \text{ linear} \}) \geq 2^{n-1}(1 - \frac{2}{2^{n/2}}) = 2^{n-1} - 2^{n/2}.$$

Truncated differential cryptanalysis of AES

Warning! This part is taken from [CDVSV04], and needs some adjustments.

6.1. What we are trying to avoid

Suppose $T : V \rightarrow V$ is a cryptographic transformation. (Of course $T = T_k$ depends on the key.) Here V has dimension n (say even) over \mathbf{F}_2 . Brute force would require searching through all 2^n elements of V .

But suppose there is a subspace W of V , of dimension $n/2$, such that if $x, y \in V$, and $x - y \in W$, then $T(x) - T(y) \in W$. This means that T sends a coset $x + W$ of W into another such coset. In fact if $y = x + w \in x + W$, so that $w \in W$, we have $y - x \in W$, so $T(y) - T(x) \in W$, and $T(y) \in T(x) + W$. In other words $T(x + W) \subseteq T(x) + W$, and thus $T(x + W) = T(x) + W$, because T is a bijection, and the two sets have the same number $|W|$ of elements.

So the cryptanalyst builds a quick membership test (sifting) for elements of W , and a set R , of size $2^{n/2}$, such that

$$\{x + W : x \in V\} = \{r + W : r \in R\}.$$

Given a ciphertext c_0 , the cryptanalyst searches through R until he finds $r_0 \in R$ such that $c_0 \in T(r_0) + W$, that is, $c_0 - T(r_0) \in W$. So $c_0 \in T(r_0 + W)$. The cryptanalyst now searches through W until he finds $w_0 \in W$ such that $c_0 = T(r_0 + w_0)$. So $p_0 = r_0 + w_0$ is the plaintext, and it has taken $2 \cdot 2^{n/2}$ attempts to find it.

(*Note.* Some arguments are more general, valid also over any (infinite) field and for subspaces of arbitrary dimension.)

6.2. Notation and statement

Recall

LEMMA 6.2.1. $\mathbf{GF}(p^n) \subseteq \mathbf{GF}(p^m)$ if and only if n divides m .

We are staying close to the notation of [DR02]. We assume $\rho = \gamma\lambda$, where γ and λ are permutations. Here γ is a bricklayer transformation, consisting of a number of S-boxes. The message space V is written as a direct sum

$$V = V_1 \oplus \cdots \oplus V_{n_t},$$

where each V_i has the same dimension m over $\mathbf{GF}(2)$. For $v \in V$, we will write $v = v_1 + \cdots + v_{n_t}$, where $v_i \in V_i$. Also, we consider the projections $\pi_i : V \rightarrow V_i$, which map $v \mapsto v_i$. We have

$$v\gamma = v_1\gamma_1 \oplus \cdots \oplus v_{n_t}\gamma_{n_t},$$

where the γ_i are S-boxes, which we allow to be different for each V_i .

λ is a linear mixing layer.

In AES the S-boxes are all equal, and consist of inversion in the field $\mathbf{GF}(2^8)$ with 2^8 elements (see later in this paragraph), followed by an affine transformation. The latter map thus consists of a linear transformation, followed by a translation. When interpreting AES in our scheme, we take advantage of the well-known possibility of moving the linear part of the affine transformation to the linear mixing layer, and incorporating the translation in the key addition (see for instance [MR02]). Thus in our scheme for AES we have $m = 8$, we identify each V_i with $\mathbf{GF}(2^8)$, and we take $x\gamma_i = x^{2^8-2}$, so that γ_i maps nonzero elements to their inverses, and zero to zero. As usual, we abuse notation and write $x\gamma_i = x^{-1}$. Note, however, that with this convention $xx^{-1} = 1$ only for $x \neq 0$.

Our result, for a key-alternating block cipher as described earlier in this section, is the following.

THEOREM 6.2.2. *Suppose the following hold:*

- (1) $0\gamma = 0$ and $\gamma^2 = 1$, the identity transformation.
- (2) *There is $1 \leq r < m/2$ such that for all i*
 - *for all $0 \neq v \in V_i$, the image of the map $V_i \rightarrow V_i$, which maps $x \mapsto (x + v)\gamma_i + x\gamma_i$, has size greater than 2^{m-r-1} , and*
 - *there is no subspace of V_i , invariant under γ_i , of codimension less than or equal to $2r$.*
- (3) *No sum of some of the V_i (except $\{0\}$ and V) is invariant under λ .*

Then there is no subspace $U \neq \{0\}, V$ of U such that if an inout difference is in U , so is the corresponding output difference.

6.3. AES

We note immediately

LEMMA 6.3.1. *AES satisfies the hypotheses of Theorem 6.2.2.*

PROOF OF LEMMA 6.3.1. Condition (1) is clearly satisfied.

So is (3), by the construction of the mixing layer. In fact, suppose $U \neq \{0\}$ is a subspace of V which is invariant under λ . Suppose, without loss of generality, that $U \supseteq V_1$. Because of **MixColumns** [DR02, 3.4.3], U contains the whole first column of the state. Now the action of **ShiftRows** [DR02, 3.4.2] and **MixColumns** on the first column shows that U contains four whole columns, and considering (if the state has more than four columns) once more the action of **ShiftRows** and **MixColumns** one sees that $U = V$.

The first part of Condition (2) is also well-known to be satisfied, with $r = 1$ (see [Ny94] but also [DR06]). We recall the short proof for convenience. For $a \neq 0$, the map $\mathbf{GF}(2^8) \rightarrow \mathbf{GF}(2^8)$, which maps $x \mapsto (x + a)^{-1} + x^{-1}$, has image of size $2^7 - 1$. In fact, if $b \neq a^{-1}$, the equation

$$(6.3.1) \quad (x + a)^{-1} + x^{-1} = b$$

has at most two solutions. Clearly $x = 0, a$ are not solutions, so we can multiply by $x(x + a)$ obtaining the equation

$$(6.3.2) \quad x^2 + ax + ab^{-1} = 0,$$

which has at most two solutions. If $b = a^{-1}$, equation (6.3.1) has four solutions. Two of them are $x = 0, a$. Two more come from (6.3.2), which becomes

$$x^2 + ax + a^2 = a^2 \cdot ((x/a)^2 + x/a + 1) = 0.$$

By Lemma 6.2.1, $\mathbf{GF}(2^8)$ contains $\mathbf{GF}(4) = \{0, 1, c, c^2\}$, where c, c^2 are the roots of $y^2 + y + 1 = 0$. Thus when $b = a^{-1}$ equation (6.3.1) has the four solutions $0, a, ac, ac^2$. It follows that the image of the map $x \mapsto (x + a)^{-1} + x^{-1}$ has size

$$\frac{2^8 - 4}{2} + \frac{4}{4} = 2^7 - 1,$$

as claimed.

As to the second part of Condition (2), one could just use GAP [GAP05] to verify that the only nonzero subspaces of $\mathbf{GF}(2^8)$ which are invariant under inversion are the subfields. According to Lemma 6.2.1, the largest proper one is thus $\mathbf{GF}(2^4)$, of codimension $4 > 2 = 2r$. However, this follows from the more general Theorem 6.5.1, which we give in the Appendix. \square

6.4. Proof

PROOF OF THEOREM 6.2.2. Suppose, by way of contradiction, that there is a subspace $U \neq \{0\}$, V of V such that if $v, v + u \in V$ are two messages whose difference u lies in the subspace U , then the output difference also lies in U , that is

$$(v + u)\rho + v\rho \in U.$$

Since λ is linear, we have

FACT 1. For all $u \in U$ and $v \in V$ we have

$$(6.4.1) \quad (v + u)\gamma + v\gamma \in U\lambda^{-1} = W,$$

where W is also a linear subspace of V , with $\dim(W) = \dim(U)$.

Setting $v = 0$ in (6.4.1), and because of Condition (1), we obtain

FACT 2. $U\gamma = W$ and $W\gamma = U$.

Now if $U \neq \{0\}$, we will have $U\pi_i \neq \{0\}$ for some i . We prove some increasingly stronger facts under this hypothesis.

FACT 3. Suppose $U\pi_i \neq \{0\}$ for some i . Then $W \cap V_i \neq \{0\}$.

Let $u \in U$, with $u_i \neq 0$. Take any $0 \neq v_i \in V_i$. Then $(u + v_i)\gamma + v_i\gamma \in W$, and also $u\gamma \in W$, by Fact 2. It follows that $u\gamma + (u + v_i)\gamma + v_i\gamma \in W$. The latter vector has all nonzero components but for the one in V_i , which is $u_i\gamma_i + (u_i + v_i)\gamma_i + v_i\gamma_i \in W \cap V_i$. If the latter vector is zero for all $v_i \in V_i$, then the image of the map $V_i \rightarrow V_i$, which maps $v_i \mapsto (v_i + u_i)\gamma_i + v_i\gamma_i$, is $\{u_i\gamma_i\}$, of size 1. This contradicts the first part of Condition (2).

Clearly $(W \cap V_i)\gamma = U \cap V_i$. It follows

FACT 4. Suppose $U\pi_i \neq \{0\}$ for some i . Then $U \cap V_i \neq \{0\}$.

Finally we obtain

FACT 5. Suppose $U\pi_i \neq \{0\}$ for some i . Then $U \supseteq V_i$.

According to Fact 4, there is $0 \neq u_i \in U \cap V_i$. By the first part of Condition (2) the map $V_i \rightarrow V_i$, which maps $x \mapsto (x + u_i)\gamma_i + x\gamma_i$, has image of size $> 2^{m-r-1}$. Since this image is contained in the linear subspace $W \cap V_i$, it follows that the latter has size at least 2^{m-r} , that is, codimension at most r in V_i . The same holds for $U \cap V_i = (W \cap V_i)\gamma$. Thus the linear subspace $U \cap W \cap V_i$ has codimension at most $2r$ in V_i . In particular, it is different from $\{0\}$, as $m > 2r$. From Fact 2 it follows that $U \cap W \cap V_i$ is invariant under γ . By the second part of Condition (2) we have $U \cap W \cap V_i = V_i$, so that $U \supseteq V_i$ as claimed.

From Fact 5 we obtain immediately

FACT 6. U is a direct sum of some of the V_i , and $W = U$

The second part follows from the fact that $W = U\gamma$, and $V_i\gamma = V_i$ for all i .

Since $U = W\lambda$ by (6.4.1), we obtain $U = U\lambda$, with $U \neq \{0\}, V$. This contradicts Condition (3), and completes the proof. \square

The proof of Theorem 6.2.2 can be adapted to prove a slightly more general statement, in which Conditions (1) and (2) are replaced with

(1') $0\gamma = 0$ and $\gamma^s = 1$, for some $s > 1$.

(2') There is $1 \leq r < m/s$ such that for all i

- for all $0 \neq v \in V_i$, the image of the map $V_i \rightarrow V_i$, which maps $x \mapsto (x + v)\gamma_i + x\gamma_i$, has size greater than 2^{m-r-1} , and
- there is no proper subspace of V_i , invariant under γ_i , of codimension less than or equal to sr .

6.5. Additive subgroups of finite fields containing their inverses

We are grateful to Sandro Mattarei (see [Mat05], and also [GGSZ04], for more general results) for the following

TEOREMA 6.5.1. *Let F be a field of characteristic two. Suppose $U \neq 0$ is an additive subgroup of F which contains the inverses of each of its nonzero elements. Then U is a subfield of F .*

PROOF. Hua's identity, valid in any associative (but not necessarily commutative) ring A , shows

$$(6.5.1) \quad a + ((a - b^{-1})^{-1} - a^{-1})^{-1} = aba$$

for $a, b \in A$, with $a, b, ab - 1$ invertible.

First of all, $1 \in U$. This is because U has even order, and each element different from 0, 1 is distinct from its inverse.

Now (6.5.1) for $b = 1$, and $a \in U \setminus \{0, 1\}$ shows that for $a \in U$, also $a^2 \in U$. (This is clearly valid also for $a = 0, 1$.) It follows that any $c \in U$ can be represented in the form $c = a^2$ for some $a \in U$. Now (6.5.1) shows that U is closed under products, so that U is a subring, and thus a subfield, of F . \square

6.6. Equations of degree two in characteristic two

Suppose we have the equation $x^2 + ux + v = 0$ over a finite field F of characteristic two, and order 2^n . Clearly the usual formula for the solutions does not work, as we cannot divide by two, and in fact completing the square does not work here: if we substitute $x + a$ to x , we get $x^2 + ux + v + a^2 + au$, with no gain.

Now if $u = 0$ we get one solution, as the function $x \mapsto x^2$ is an isomorphism here. So assume $u \neq 0$. Substitute ux to x to get $u^2x^2 + u^2x + v = 0$, and divide by u^2 to get $x^2 + x + vu^{-2} = 0$. The function $x \mapsto x^2 + x$ is a substitute in characteristic two of the usual “square root” function. It is linear over \mathbf{F}_2 . Its kernel is $\{0, 1\}$, so its image has size $|F|/2$. If $b = a^2 + a$ is in the image, then $b^2 = a^4 + a^2, \dots, b^{2^{n-1}} = a^{2^n} + a^{2^{n-1}} = a + a^{2^{n-1}}$, so that $\text{tr}(b) = b + b^2 + \dots + b^{2^{n-1}} = 0$. It follows that the image of $x \mapsto x^2 + x$ is the kernel of the trace function $\text{tr} : F \rightarrow \mathbf{F}_2$, which has indeed order $|F|/2$. We obtain

TEOREMA 6.6.1. *Consider the equation $x^2 + ux + v = 0$ over the finite field F of characteristic two.*

- (1) *If $u = 0$, the equation has one double solution.*
- (2) *If $u \neq 0$, the equation has two distinct solutions in F if and only if $\text{tr}(vu^{-2}) = 0$.*

CHAPTER 7

Codes

7.1. The singleton bound

To be written.

7.2. Circulant matrices

Let $A = \mathbf{C}[x]/(x^n - 1)$. Every element of A can be represented uniquely as the class of a polynomial of degree $< n$, so that A has basis $1, x, \dots, x^{n-1}$ over \mathbf{C} . Let $c = c_0 + c_1x + \dots + c_{n-1}x^{n-1} \in A$, and consider the linear function $\varphi_c : A \rightarrow A$ that maps $a \mapsto ac$. With respect to the standard basis, the matrix of c is *circulant*

$$\begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{n-3} & c_{n-2} & c_{n-1} \\ c_{n-1} & c_0 & c_1 & c_2 & \dots & c_{n-3} & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & c_1 & c_2 & \dots & c_{n-3} \\ & & & \ddots & & & \\ c_1 & c_2 & c_3 & \dots & c_{n-2} & c_{n-1} & c_0 \end{bmatrix}$$

This is better understood if A is mapped onto $B = \mathbf{C}^n$ via $a \mapsto [a(\omega^i)]_{0 \leq i < n}$, where ω is a primitive n -th root of 1. (This is well defined on A , because the ω^i are precisely the roots of $x^n - 1$.) Since $a \cdot c(z) = a(z) \cdot c(z)$, with respect to the canonical basis of B multiplication by c becomes multiplication by the scalar matrix which has $c(\omega^i)$ on the diagonal. So these are the eigenvalues of φ_c , we see that φ_c is invertible iff no ω^i is a root of c , etc.

7.3. Circulant matrices in characteristic two

Here part of the arguments fail. Nevertheless we consider $A = F[x]/(x^4 + 1)$, where $F = \mathbf{GF}(2^8)$, and $c = (\alpha + 1)x^3 + x^2 + x + \alpha$, where α is a root of $m = x^8 + x^4 + x^3 + x + 1$, the Rijndael polynomial. We have $c(1) = 1$, so φ_c is invertible with inverse φ_{c^3} , as $c = (x + 1)q + 1$ for some q , $c^3c = c^4 = (x^4 + 1)q^4 + 1 \equiv 1 \pmod{x^4 + 1}$. We compute

$$c^3 = (\alpha^3 + \alpha + 1)x^3 + (\alpha^3 + \alpha^2 + 1)x^2 + (\alpha^3 + 1)x + \alpha^3 + \alpha^2 + \alpha,$$

and find its four coefficients are nonzero.

Bibliography

- [CDVSV04] A Caranti, F. Dalla Volta, M. Sala, and F. Villani, *Primitivity for groups generated by the round functions of iterated block cryptosystems*, Working paper, 2004.
- [CU57] L. Carlitz and S. Uchiyama, *Bounds for exponential sums*, Duke Math. J. **24** (1957), 37–41. MR MR0082517 (18,563c)
- [Dia98] Persi Diaconis, *From shuffling cards to walking around the building: an introduction to modern Markov chain theory*, Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998), no. Extra Vol. I, 1998, pp. 187–204 (electronic). MR MR1648031 (99e:60154)
- [DR02] Joan Daemen and Vincent Rijmen, *The design of Rijndael*, Information Security and Cryptography, Springer-Verlag, Berlin, 2002, AES—the advanced encryption standard. MR MR1986943 (2006b:94025)
- [DR06] ———, *Two-round AES differentials*, IACR e-print eprint.iacr.org/2006/039.pdf, 2006.
- [GAP05] The GAP Group, *GAP – Groups, Algorithms, and Programming, Version 4.4*, 2005, (<http://www.gap-system.org>).
- [GGSZ04] D. Goldstein, R. Guralnick, L. Small, and E. Zelmanov, *Inversion invariant additive subgroups of division rings*, Pacific J. Math. (2004), to appear.
- [Mat05] Sandro Mattarei, *Inversion invariant additive subgroups of division rings*, Israel J. Math. (2005), to appear.
- [MR02] Sean Murphy and Matthew J.B. Robshaw, *Essential algebraic structure within the AES*, Advances in Cryptology - CRYPTO 2002 (M. Yung, ed.), Lecture Notes in Computer Science, vol. 2442, Springer, Berlin/Heidelberg, 2002, pp. 1–16.
- [Nyb94] Kaisa Nyberg, *Differentially uniform mappings for cryptography*, Advances in Cryptology — EUROCRYPT '93 (Lofthus, 1993), Lecture Notes in Comput. Sci., vol. 765, Springer, Berlin, 1994, pp. 55–64. MR MR1290329 (95e:94039)
- [Ser73] J-P Serre, *A course in arithmetic*, Graduate Texts in Math., vol. 7, Springer, New York, 1973.