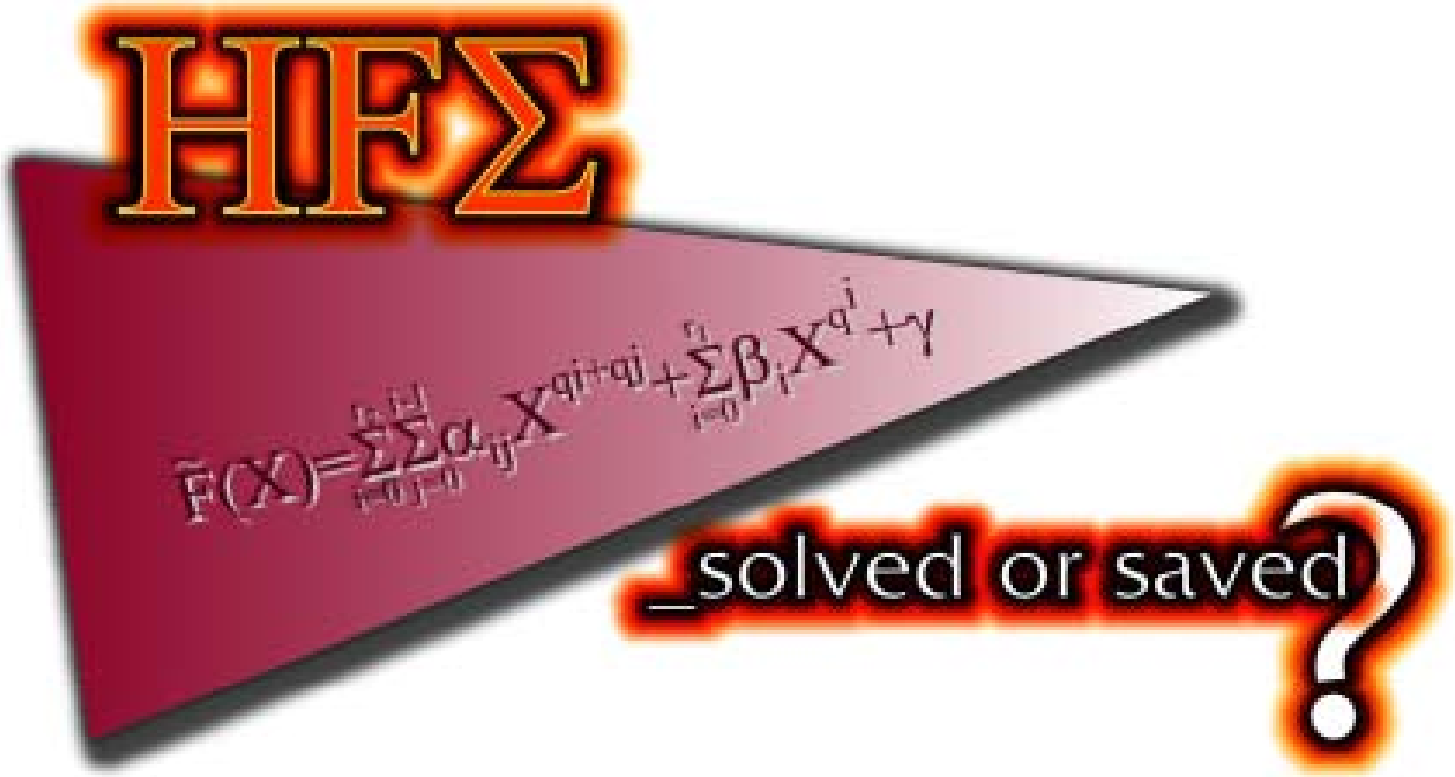
TECHNISCHE UNIVERSITÄT DARMSTADT




# Oberseminar WS07/08, CDC TU Darmstadt

Fabian Werner

# Dude, what do you want?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Invention of F4/F5
- → HFE considered to be insecure

- show that (with a little modification), **HFE can defeat F4/F5!**

# HFE – what's in a name?

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE = abbreviation for **H**idden **F**ield **E**quation
- Multivariate Public Key Cryptosystem

$q = \text{characteristic of small field}$

$n = \text{number of variables}$

$k = GF(q)$

$k[x_1, \ldots, x_n]$

# HFE – what's in a name?

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE = abbreviation for **H**idden **F**ield **E**quation
- Multivariate Public Key Cryptosystem

$q = \text{characteristic of small field}$

$n = \text{number of variables}$

$k = GF(q)$

$k[x_1, \ldots, x_n]$

$$K = GF(q^n)$$

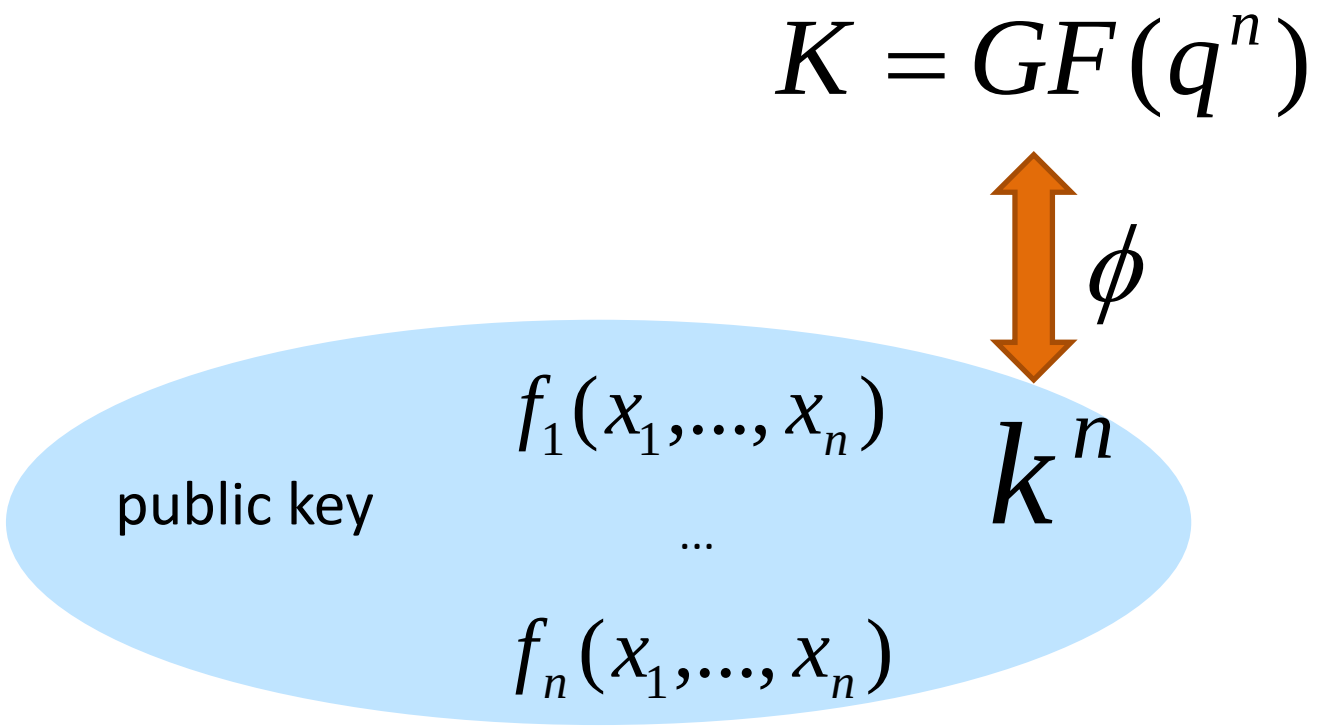
$$\updownarrow \phi$$

$$k^n$$

# HFE – what's in a name?

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE = abbreviation for **H**idden **F**ield **E**quation
- Multivariate Public Key Cryptosystem



# HFE – what's in a name?

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE = abbreviation for **H**idden **F**ield **E**quation
- Multivariate Public Key Cryptosystem

$\tilde{F}(X) \in K[X]$ private key

$K = GF(q^n)$

$\phi$

$k^n$

public key

$f_1(x_1, \ldots, x_n)$

...

$f_n(x_1, \ldots, x_n)$

# HFE – what's in a name?

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE = abbreviation for **H**idden **F**ield **E**quation
- Multivariate Public Key Cryptosystem

$\tilde{F}(X) \in K[X]$ private key

$K = GF(q^n)$

"translation"

$$\phi(a_0 + a_1 t + \ldots + a_{n-1} t^{n-1}) = (a_0, \ldots, a_n)$$

$\phi$

$k^n$

$f_1(x_1, \ldots, x_n)$

public key

...

$f_n(x_1, \ldots, x_n)$

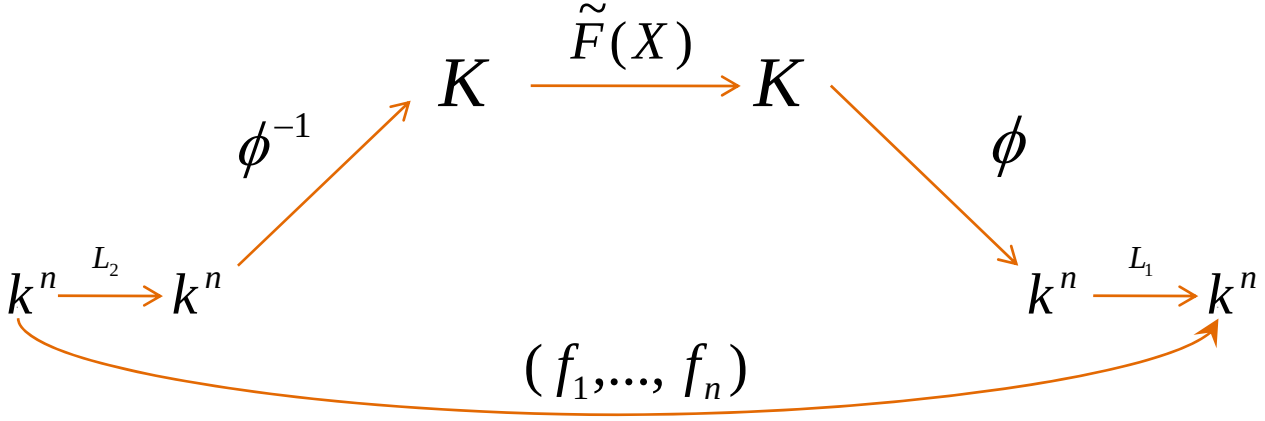
# PK, encryption

$$(f_1,\ldots,f_n) = L_1 \circ \phi \circ \tilde{F} \circ \phi^{-1} \circ L_2(x_1,\ldots,x_n)$$

PK creation

# PK, encryption

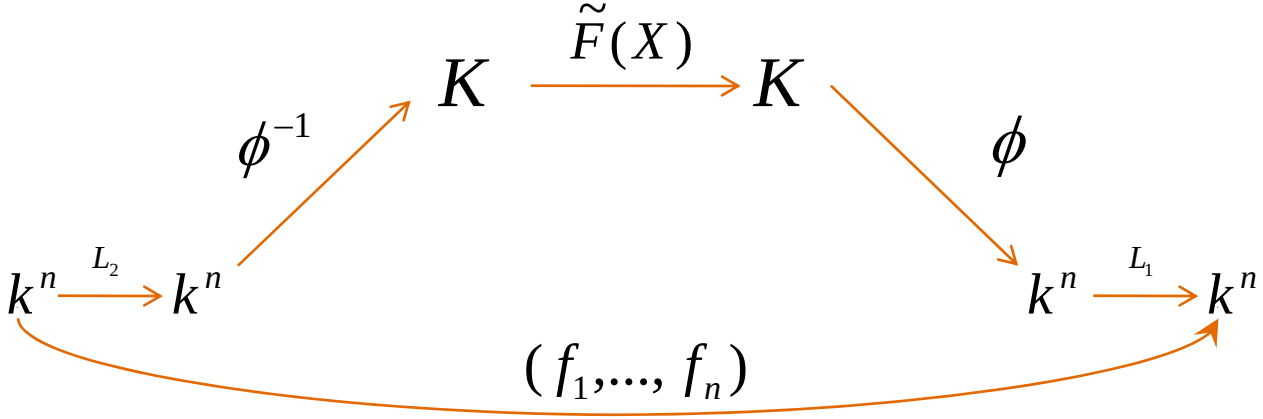
$$(f_1,\ldots,f_n) = L_1 \circ \phi \circ \tilde{F} \circ \phi^{-1} \circ L_2(x_1,\ldots,x_n)$$




PK creation

# PK, encryption

$$(f_1,\ldots,f_n) = L_1 \circ \phi \circ \tilde{F} \circ \phi^{-1} \circ L_2(x_1,\ldots,x_n)$$




PK creation

$$c_1 = f_1(p_1,\ldots,p_n)$$

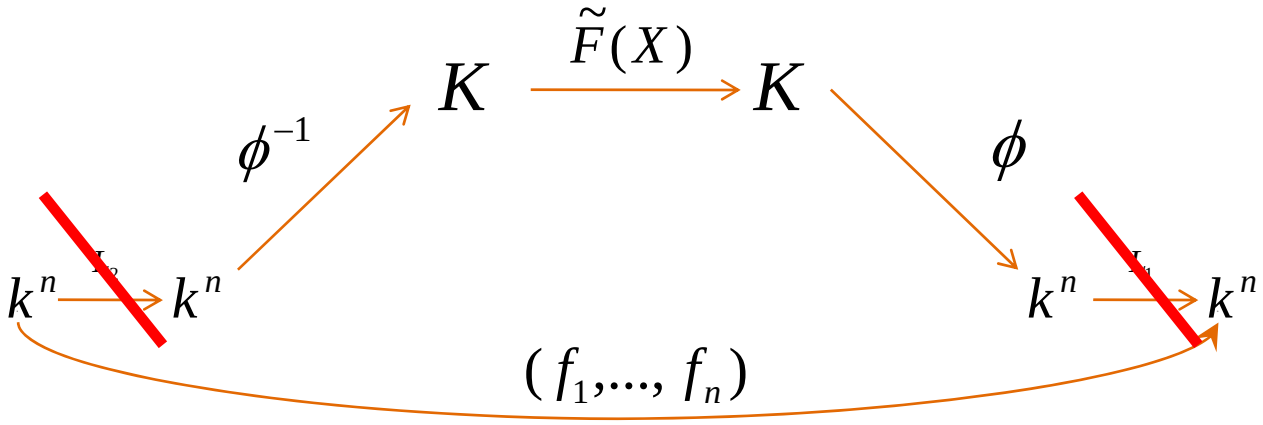
$$\ldots$$

$$c_n = f_n(p_1,\ldots,p_n)$$

encryption

# PK, encryption

$$(f_1,\ldots,f_n) = L_1 \circ \phi \circ \tilde{F} \circ \phi^{-1} \circ L_2(x_1,\ldots,x_n)$$




PK creation

$$c_1 = f_1(p_1,\ldots,p_n)$$

$$\ldots$$

$$c_n = f_n(p_1,\ldots,p_n)$$

encryption

# The private key

TECHNISCHE UNIVERSITÄT DARMSTADT

- $X$ will look like $x_0 + tx_1 + t^2 x_2 + \ldots + t^{n-1} x_{n-1}$
- private key:

$$\tilde{F}(X) = \sum_{i=0}^{r_2} \sum_{j=0}^{i} \alpha_{ij} X^{q^i} X^{q^j} + \sum_{i=0}^{r_1} \beta_i X^{q^i} + \gamma$$

- why?

# The private key

- $X$ will look like $x_0 + tx_1 + t^2 x_2 + \ldots + t^{n-1} x_{n-1}$
- private key:

$$\tilde{F}(X) = \sum_{i=0}^{r_2} \sum_{j=0}^{i} \alpha_{ij} X^{q^i} X^{q^j} + \sum_{i=0}^{r_1} \beta_i X^{q^i} + \gamma$$

- why?

$$X^{q^i} \text{ is linear, } X^{q^i} X^{q^j} \text{ is quadratic,...}$$

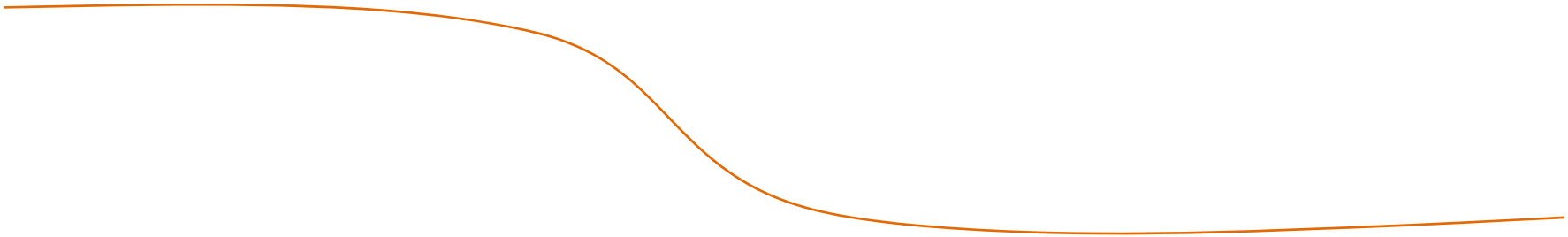
$$\rightarrow x_i + x_j + c \qquad \rightarrow x_i^2 + x_j x_k + c$$

# Example

TECHNISCHE UNIVERSITÄT DARMSTADT

$q = 2,\ n = 3$

$K = GF(2^3)$



$k = GF(2)$

# Example

TECHNISCHE UNIVERSITÄT DARMSTADT

$q = 2,\ n = 3$

$K = GF(2^3)$

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

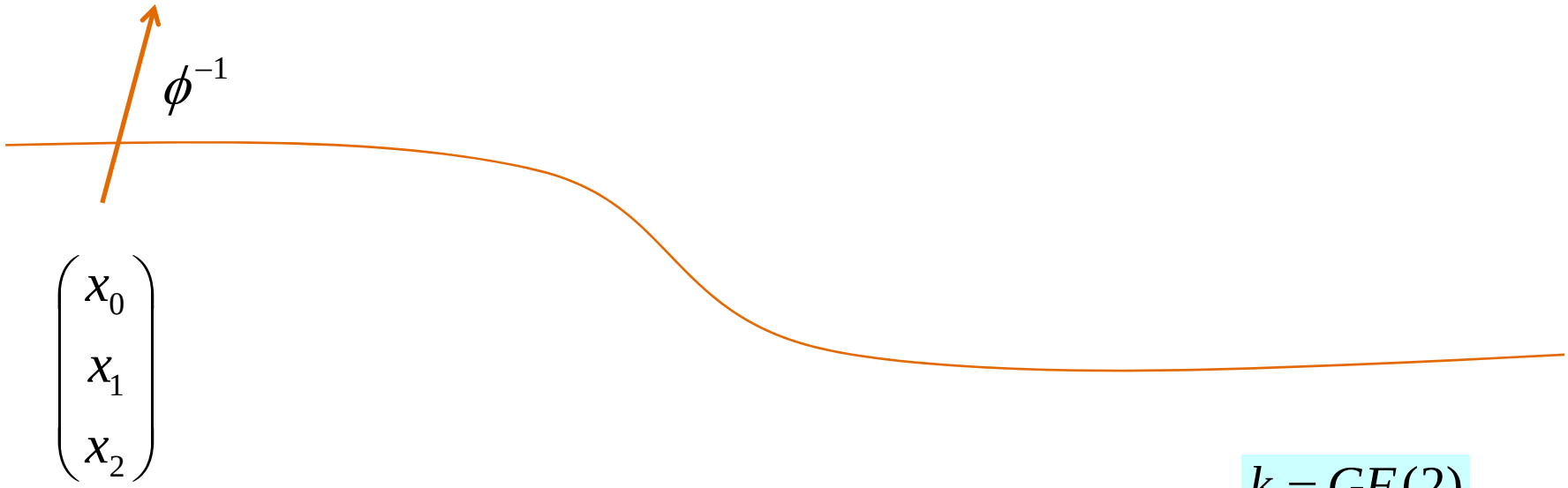
$k = GF(2)$

# Example

$q = 2,\ n = 3$

$K = GF(2^3)$

$$\vec{x} = x_0 + tx_1 + t^2 x_2$$




$k = GF(2)$

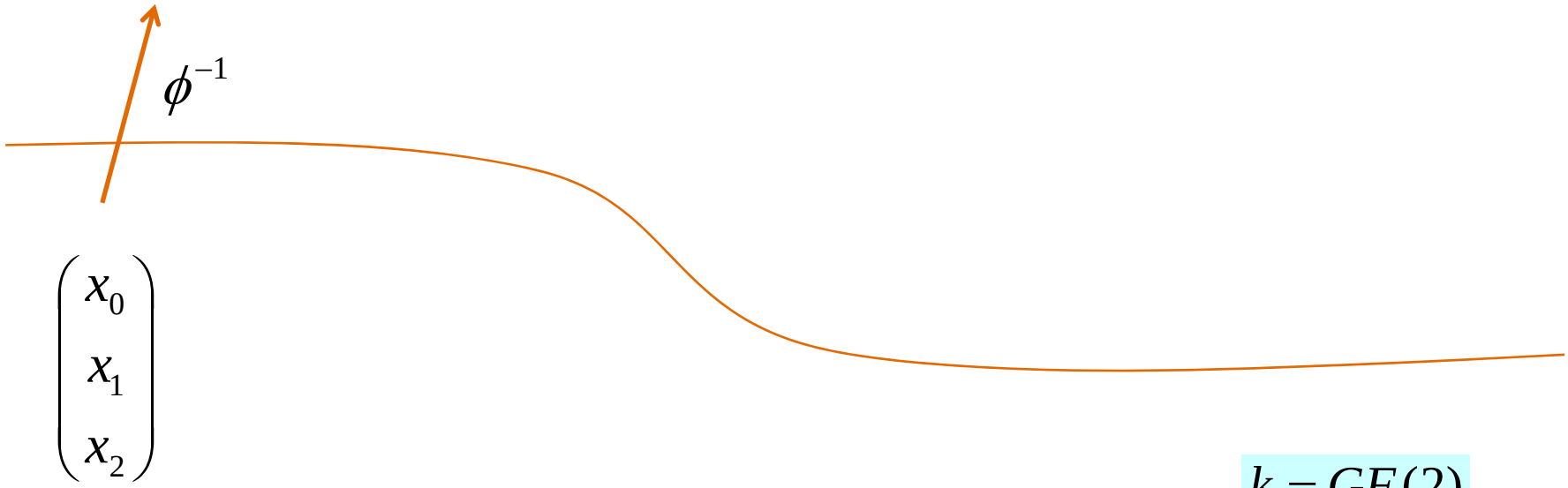
# Example

$q = 2,\ n = 3$

$$\tilde{F}(X) = (t+1)X^3 + (t)X$$

$K = GF(2^3)$

$$\vec{x} = x_0 + tx_1 + t^2 x_2$$




$k = GF(2)$

# Example

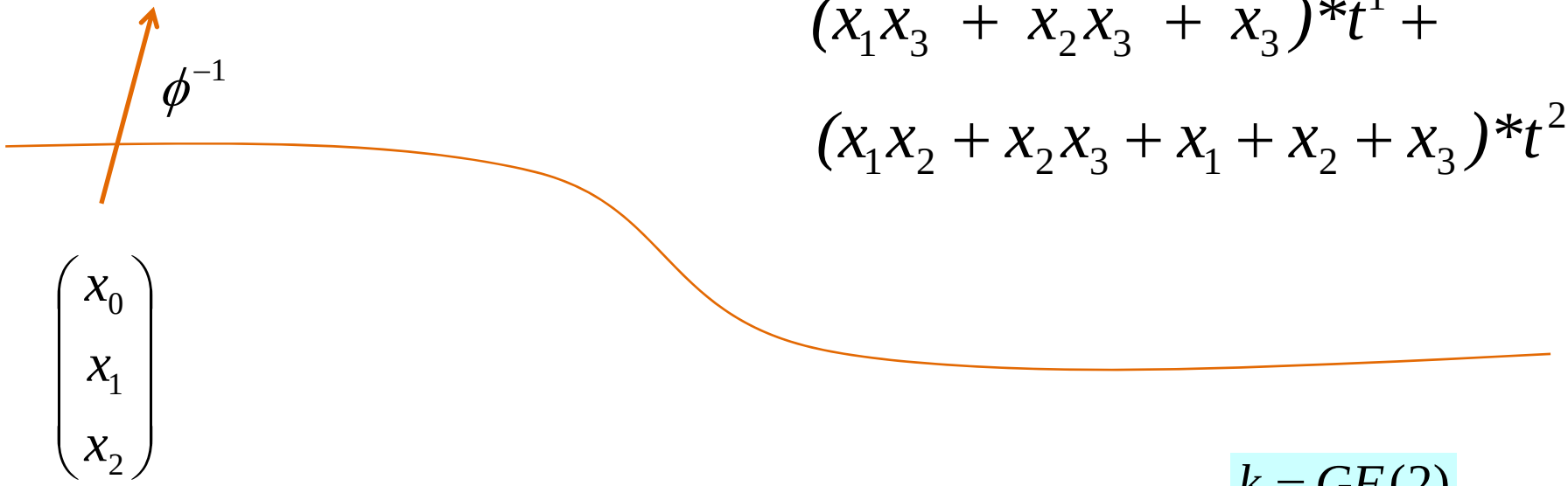
$q = 2,\ n = 3$

$$\tilde{F}(X) = (t+1)X^3 + (t)X$$

$K = GF(2^3)$

$$\vec{x} = x_0 + tx_1 + t^2 x_2$$

$$\tilde{F}(\vec{x}) = (x_1x_3 + x_3)*t^0 + (x_1x_3 + x_2x_3 + x_3)*t^1 + (x_1x_2 + x_2x_3 + x_1 + x_2 + x_3)*t^2$$




$k = GF(2)$

# Example

$q = 2,\ n = 3$

$$\tilde{F}(X) = (t+1)X^3 + (t)X$$

$K = GF(2^3)$

$$\vec{x} = x_0 + tx_1 + t^2 x_2$$

$$\tilde{F}(\vec{x}) = (x_1x_3 + x_3)*t^0 + (x_1x_3 + x_2x_3 + x_3)*t^1 + (x_1x_2 + x_2x_3 + x_1 + x_2 + x_3)*t^2$$

$\phi^{-1}$

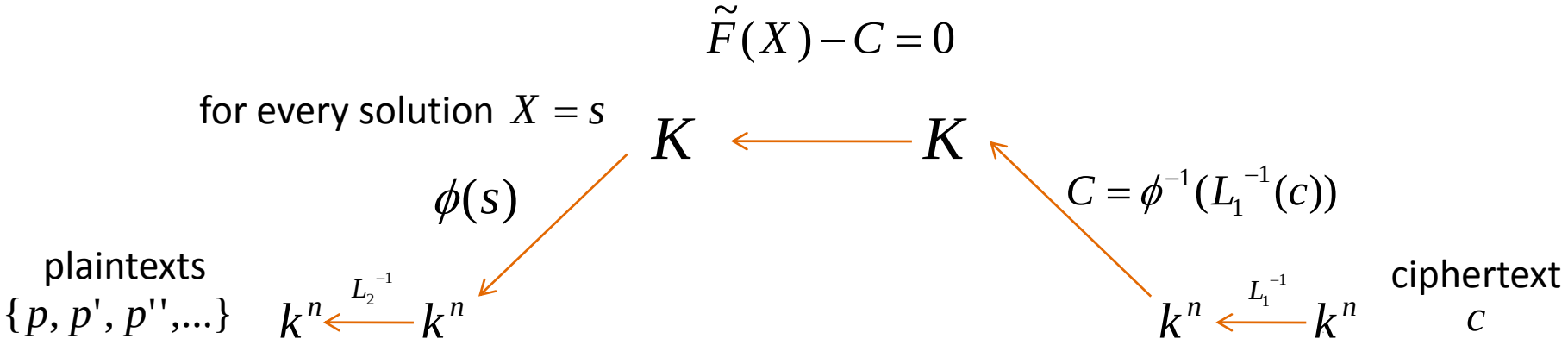
$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

$\phi$

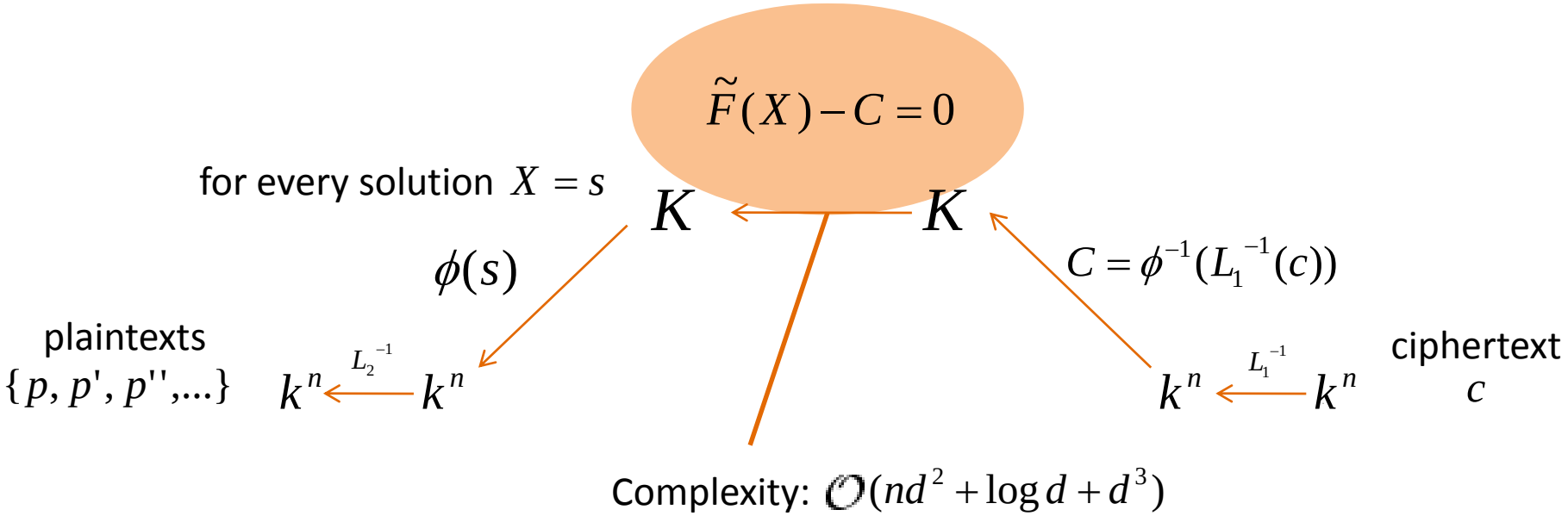
$$\begin{pmatrix} x_1x_3 + x_3 \\ x_1x_3 + x_2x_3 + x_3 \\ x_1x_2 + x_2x_3 + x_1 + x_2 + x_3 \end{pmatrix}$$

$k = GF(2)$

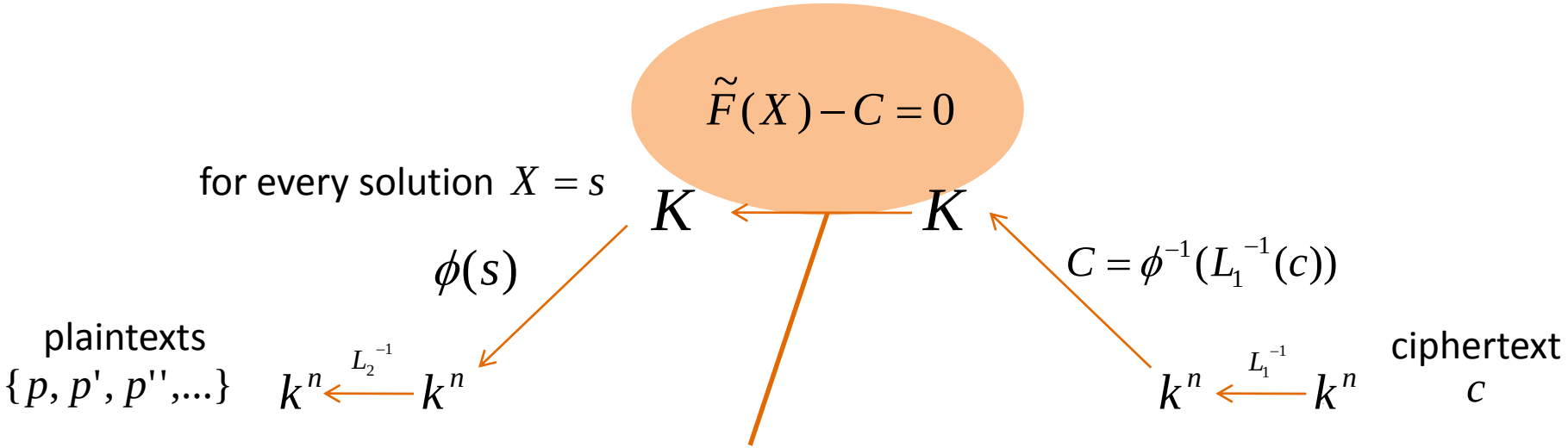
# Decryption


$\tilde{F}(X) - C = 0$
for every solution $X = s$
$K$
$K$
$\phi(s)$
$C = \phi^{-1}(L_1^{-1}(c))$
plaintexts
$\{p, p', p'', \ldots\}$
$k^n$
$L_2^{-1}$
$k^n$
$k^n$
$L_1^{-1}$
$k^n$
ciphertext
$c$

# Decryption


$\tilde{F}(X) - C = 0$
for every solution $X = s$
$K$
$K$
$\phi(s)$
$C = \phi^{-1}(L_1^{-1}(c))$
plaintexts
$\{p, p', p'', \ldots\}$
$k^n$
$L_2^{-1}$
$k^n$
$k^n$
$L_1^{-1}$
$k^n$
ciphertext
$c$
Complexity: $\mathcal{O}(nd^2 + \log d + d^3)$

# Decryption

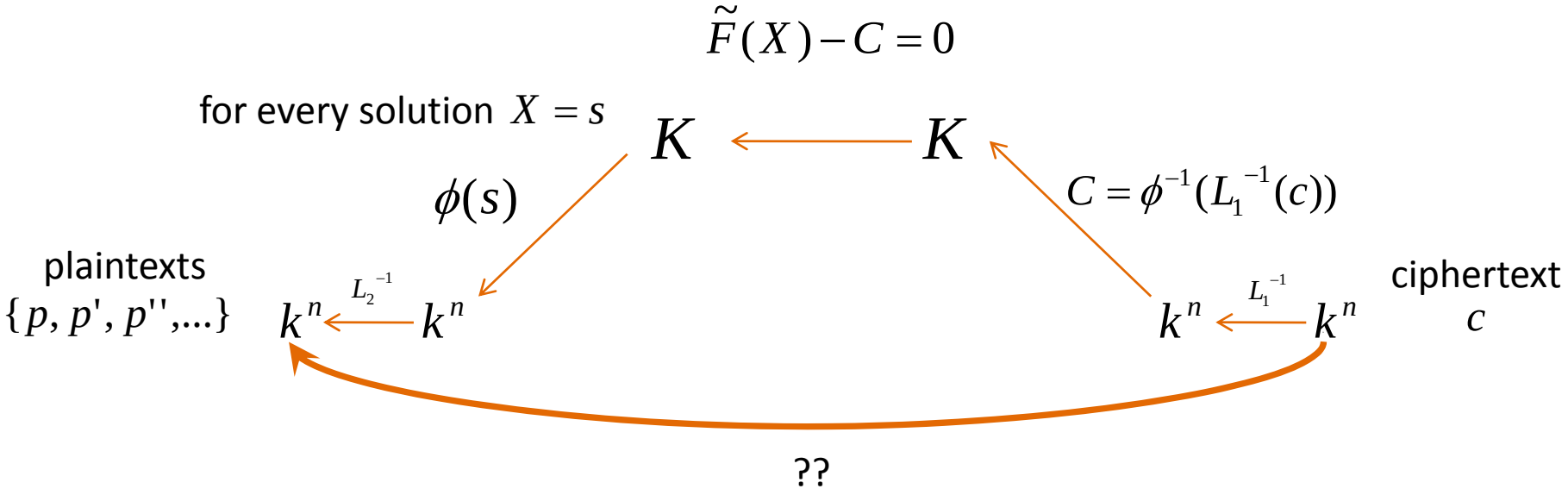



Complexity: $\mathcal{O}(nd^2 + \log d + d^3)$

where $d = \deg \tilde{F}(X)$

→efficiency strongly depends on n,d

# Decryption


$\tilde{F}(X) - C = 0$
for every solution $X = s$
$K$
$K$
$\phi(s)$
$C = \phi^{-1}(L_1^{-1}(c))$
plaintexts
$\{p, p', p'', \ldots\}$
$k^n$
$L_2^{-1}$
$k^n$
$k^n$
$L_1^{-1}$
$k^n$
ciphertext
$c$
??

# Breaking HFE

TECHNISCHE UNIVERSITÄT DARMSTADT

- "breaking": given ciphertext $c$ and public key $f_1,\ldots, f_n$, recover the plaintext $p$
- hence, solve

$$\begin{array}{l} 0 = f_1(p_1,\ldots,p_n) - c_1 \\ \ldots \\ 0 = f_n(p_1,\ldots,p_n) - c_n \end{array} \Leftrightarrow \begin{array}{l} c_1 = f_1(p_1,\ldots,p_n) \\ \ldots \\ c_n = f_n(p_1,\ldots,p_n) \end{array} \Leftrightarrow \tilde{F}(X) - \phi^{-1}(c) = 0$$

- this is exactly, what "Gröbner Basis algorithms" (respectively the Faugère $F_4/F_5$-algorithm) do
- Question: how efficient?

# Breaking HFE (2)

TECHNISCHE UNIVERSITÄT DARMSTADT

- long story cut short:
  - in 2002, Jean-Charles Faugère was able to break HFE for n=80 (and D = 96) in approx. 96 hours.
  - 2004: Allan Steel, MAGMA implementation of $\mathrm{F}_4$: 23 hours
- claim of Faugère/Joux: complexity is $\mathcal{O}(n^{\log(D)})$
- so HFE was considered to be insecure

- important: base field was assumed to be $k = GF(2)$

# Field equations ( $x_i^q \equiv x_i$ )

TECHNISCHE UNIVERSITÄT DARMSTADT

- Assume we calculate a Groebner Basis for $\langle f_1 - c_1, \ldots, f_n - c_n \rangle$
- Equivalently: find solutions to $\tilde{F}(X) - \phi^{-1}(c) = 0$
- GB-Algorithms proven fact: keep every solution

# Field equations ( $x_i^q \equiv x_i$ )

TECHNISCHE UNIVERSITÄT DARMSTADT

- Assume we calculate a Groebner Basis for $\langle f_1 - c_1, \ldots, f_n - c_n \rangle$
- Equivalently: find solutions to $\tilde{F}(X) - \phi^{-1}(c) = 0$
- GB-Algorithms proven fact: keep every solution

- → Every solution: may contain some of $AC(K)$

- We are interested in solutions over $K$ only!

# Field equations ( $x_i^q \equiv x_i$ )

TECHNISCHE UNIVERSITÄT DARMSTADT

- Assume we calculate a Groebner Basis for $\langle f_1 - c_1, \ldots, f_n - c_n \rangle$
- Equivalently: find solutions to $\tilde{F}(X) - \phi^{-1}(c) = 0$
- GB-Algorithms proven fact: keep every solution

- → Every solution: may contain some of $AC(K)$

- We are interested in solutions over $K$ only!

- Algorithm is told to search for solutions over $K$ by adding field equations
  → GB is computed for $\langle f_1 - c_1, \ldots, f_n - c_n, \; x_1^q - x_1, \ldots, x_n^q - x_n \rangle$

# Saving HFE

TECHNISCHE UNIVERSITÄT DARMSTADT

- changing the characteristic of the base field $k$

  → what's the big difference?

- GF(2) - Field equations easy to use
- GF(11) - Field equations **difficult** to use, because

# Saving HFE

TECHNISCHE UNIVERSITÄT DARMSTADT

- changing the characteristic of the base field $k$

  → what's the big difference?

- GF(2) - Field equations easy to use
- GF(11) - Field equations **difficult** to use, because

- Field equations: $x_i^{11} \equiv x_i$
  - basic algebra: given $n, d$, amount of monomials is $\binom{n+d}{d}$
  - for n=32, d=11 this already exceeds $2^{32}$ (4GB)

    → polynomials are not storable anymore

    → time/memory usage increase exponential

# Leaving out the field equations

TECHNISCHE UNIVERSITÄT DARMSTADT

- say D = 11 + 11 for example
- say the GB of $\langle f_1 - c_1, \ldots, f_n - c_n \rangle$ is $\{g_1, \ldots, g_t\}$
- #solutions of $\{g_1, \ldots, g_t\}$ = #solutions of $\tilde{F}(X) - C$

# Leaving out the field equations

TECHNISCHE UNIVERSITÄT DARMSTADT

- say D = 11 + 11 for example
- say the GB of $\langle f_1 - c_1, \ldots, f_n - c_n \rangle$ is $\{g_1, \ldots, g_t\}$
- #solutions of $\{g_1, \ldots, g_t\}$ = #solutions of $\tilde{F}(X) - C$ = 22 (its degree)
- **includes solutions at infinity → AC(K)**
  - → "distribution of degrees" has to match
  - → either $g_t$ has degree 22 or degrees of $\{g_1, \ldots, g_t\}$ do exactly match the prime factorization of 22= 2*11
  - → at least one $g_i$ has to have degree 11
    - → too much time/memory again!

# Solutions what?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Solutions at infinity:
- Solutions that live in AC(K) but not in K itself!

- Assume: find roots of $F(X) = X^2 + X + 1$

- No solution in GF($7^n$) (F is irreducible here)

# Solutions what?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Solutions at infinity:
- Solutions that live in AC(K) but not in K itself!

- Assume: find roots of $F(X) = X^2 + X + 1$

- No solution in GF($7^n$) (F is irreducible here)

- Another solution by doing level 2 extension of GF($7^n$)
  - define the irreducible polynomial to be $g(e) = e^2 + e + 1$
  - 2 solutions in GF(($7^n$)$^2$) = GF($7^{2n}$)
  - e and –(e+1)

# Solutions what?

TECHNISCHE UNIVERSITÄT DARMSTADT

- There **are** 2 solutions!

- → Groebner Basis **has** to keep them

LIVE DEMO:
SOLUTIONS AT INFINITY

# Summarization

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE can defeat F4/F5 when using GF(11) for example, because
  - degree of Gröbner Basis goes up
  - most of the choices: polynomials in the GB get too big

# Summarization

TECHNISCHE UNIVERSITÄT DARMSTADT

- HFE can defeat F4/F5 when using GF(11) for example, because
  - degree of Gröbner Basis goes up
  - most of the choices: polynomials in the GB get too big

- curious question: why doesnt it suffice to set k=GF($2^p$)?

?

# Summarization

TECHNISCHE UNIVERSITÄT DARMSTADT

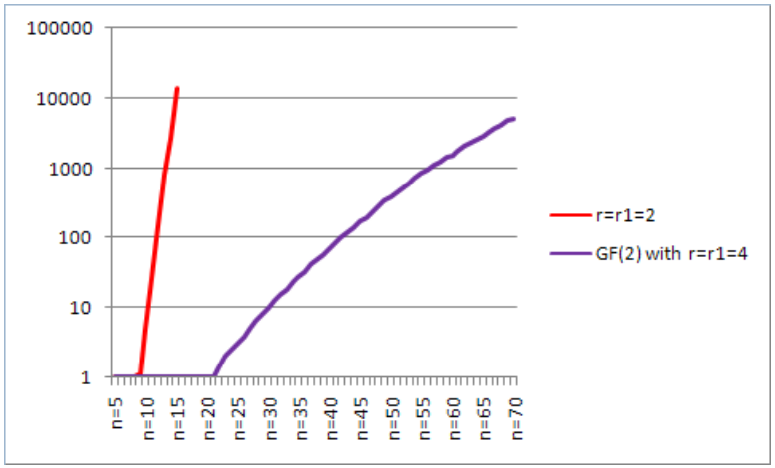
- HFE can defeat F4/F5 when using GF(11) for example, because
  - degree of Gröbner Basis goes up
  - most of the choices: polynomials in the GB get too big

- curious question: why doesnt it suffice to set k=GF($2^p$)?
  - →increase k in terms of cardinatlity is not enougth!
  - say n is fixed: K = GF(($2^p$)$^n$)
  - but this is the same as K = GF($2^{pn}$)
  - →could been solved by F4/F5 again!

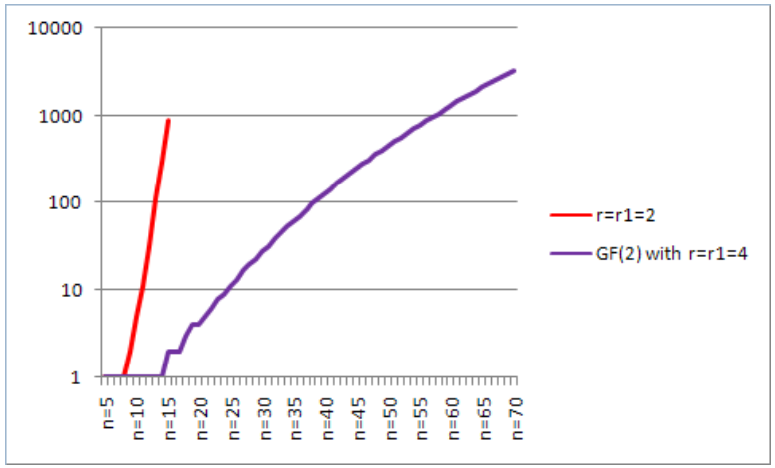
# Experimental results: GF(11) vs. GF(2)

TECHNISCHE UNIVERSITÄT DARMSTADT

We used the MAGMA implementation of the $F_4$ - algorithm:



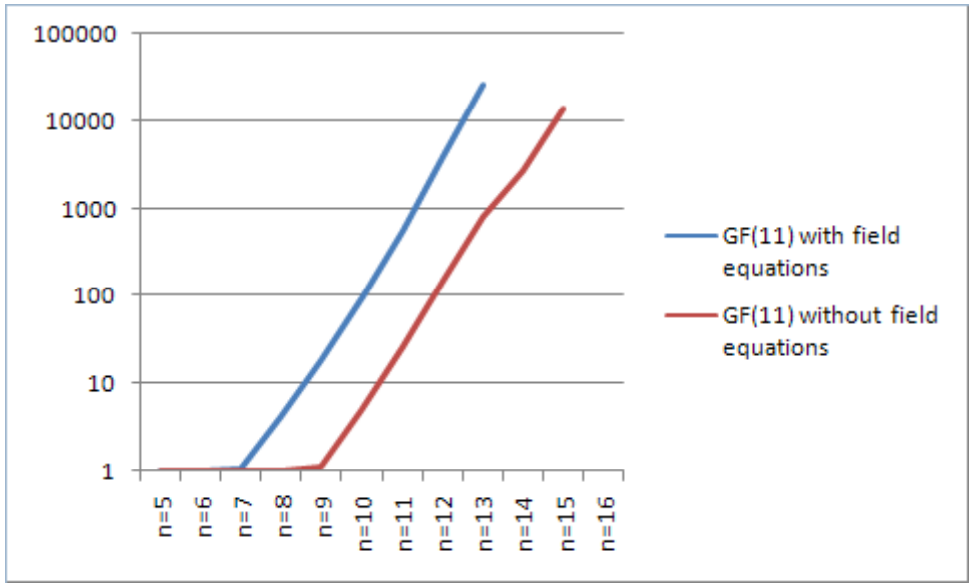

Time




Memory usage

# Field equations can hurt!

TECHNISCHE UNIVERSITÄT DARMSTADT



# $\deg(\tilde{F}(X)-C)$ equals amount of solutions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Along the way i have claimed that

$$\deg(\tilde{F}(X)-C) = \#\text{solutions to } \tilde{F}(X)-C=0$$

- Only one case where this is not true:
  - factorization of $\tilde{F}(X)-C$ contains factor with multiplicity > 1

# $\deg(\tilde{F}(X)-C)$ equals amount of solutions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Along the way i have claimed that

$$\deg(\tilde{F}(X)-C) = \#\text{solutions to } \tilde{F}(X)-C=0$$

- Only one case where this is not true:
  - factorization of $\tilde{F}(X)-C$ contains factor with multiplicity > 1

**→LIVE DEMO**

# $\deg(\tilde{F}(X)-C)$ equals amount of solutions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Along the way i have claimed that

$$\deg(\tilde{F}(X)-C) = \#\text{solutions to } \tilde{F}(X)-C=0$$

- Only one case where this is not true:
  - factorization of $\tilde{F}(X)-C$ contains factor with multiplicity > 1

**→LIVE DEMO**

- randomly picked polynomial in K: high chance that it is irreducible

# IPHFE

TECHNISCHE UNIVERSITÄT DARMSTADT

- usual private key:

$$\tilde{F}(X) = \sum_{i=0}^{r_2} \sum_{j=0}^{i} \alpha_{ij} X^{q^i} X^{q^j} + \sum_{i=0}^{r_1} \beta_i X^{q^i} + \gamma$$

# IPHFE

- usual private key:

$$\tilde{F}(X) = \sum_{i=0}^{r_2} \sum_{j=0}^{i} \alpha_{ij} X^{q^i} X^{q^j} + \sum_{i=0}^{r_1} \beta_i X^{q^i} + \gamma$$

- IPHFE private key:

$$\tilde{F}(X) = \sum_{i=0}^{r_2} \sum_{j=0}^{i} \alpha_{ij} X^{q^i} X^{q^j} + \sum_{i=0}^{r_1} \beta_i X^{q^i} + \gamma$$

$$+ \sum_{i=0}^{r1} \sum_{j=0}^{n-1} \lambda_{ij} X^{q^i} Z^{q^j} + \sum_{i=0}^{n-1} \sum_{j=0}^{i} \mu_{ij} Z^{q^i} Z^{q^j}$$

$$+ \sum_{i=0}^{n-1} \vartheta_i Z^{q^i}$$

# What is Z?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Z(X) is a k-linear map of dimension IPr
- how to find?

$$Z = z_0 + z_1 t + \ldots + z_{\mathrm{IPr}} t^{\mathrm{IPr}}$$

- $z_i$'s are symbols (freeze our basis vectors in one single symbol)

# What is Z?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Z(X) is a k-linear map of dimension IPr
- how to find?

$$Z = z_0 + z_1 t + \ldots + z_{\mathrm{IPr}} t^{\mathrm{IPr}}$$

- $z_i$'s are symbols (freeze our basis vectors in one single symbol)
- to be substituted with random basis vectors
- i.e. random linear combination of $x_i$'s
- we assume that they are linearly independent
- high prob. when amount of variables is feasibly high

# Example

- say we selected

$$\tilde{F}(X) = t * X^3 + Z^3$$

- and we select IPr = 2
- → 2 variables: $z_1$ and $z_2$

# Example

- say we selected

$$\tilde{F}(X) = t * X^3 + Z^3$$

- and we select IPr = 2
- → 2 variables: $z_1$ and $z_2$

$$Z = z_1 + t * z_2$$

- →final private key is

$$\tilde{F}(X) = t * X^3 + (z_1 + t * z_2)^3 = \ldots$$

# Example

- say we selected

$$\tilde{F}(X) = t * X^3 + Z^3$$

- and we select IPr = 2
- → 2 variables: $z_1$ and $z_2$

$$Z = z_1 + t * z_2$$

- →final private key is

$$\tilde{F}(X) = t * X^3 + (z_1 + t * z_2)^3 = \ldots$$

- public key is created by substituting

$$z_1 = x_1 + x_3$$

$$z_2 = x_2 + 1$$

# Decryption

- search through the whole image space of Z(X)
- example: given ciphertext c=(0,1,0)
- → „guess“ $z_1=0$, $z_2=0$

$$\tilde{F}(X) = t * X^3 + (0 + t * 0)^3 = t * X^3$$

$$factors(t * X^3) = [t, X]$$

- X=0 is a solution →$x_1=0$, $x_2=0$, $x_3=0$
- BUT: does $z_1=0$, $z_2=0$ still hold?
- NO, because

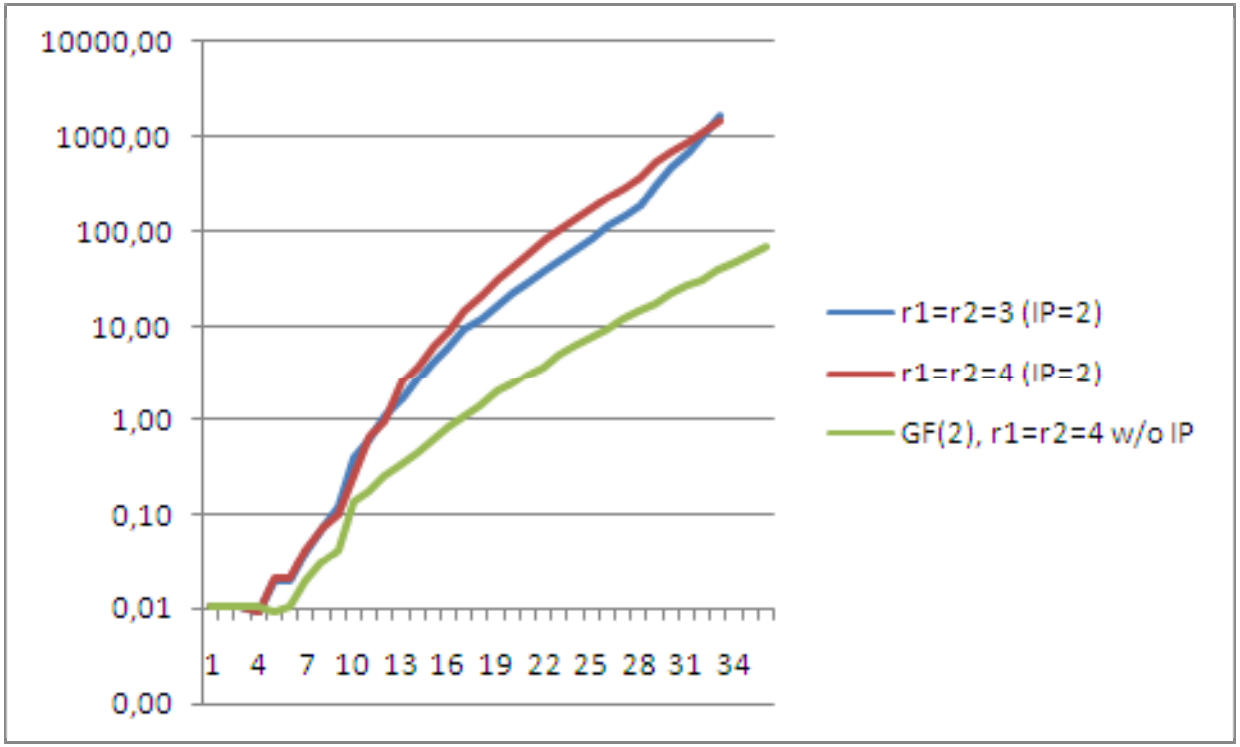
$$z_1 = x_1 + x_3 = 0 + 0 = 0$$

$$z_2 = x_2 + 1 = 0 + 1 \neq 0$$

# IPHFE: philosophy

TECHNISCHE UNIVERSITÄT DARMSTADT

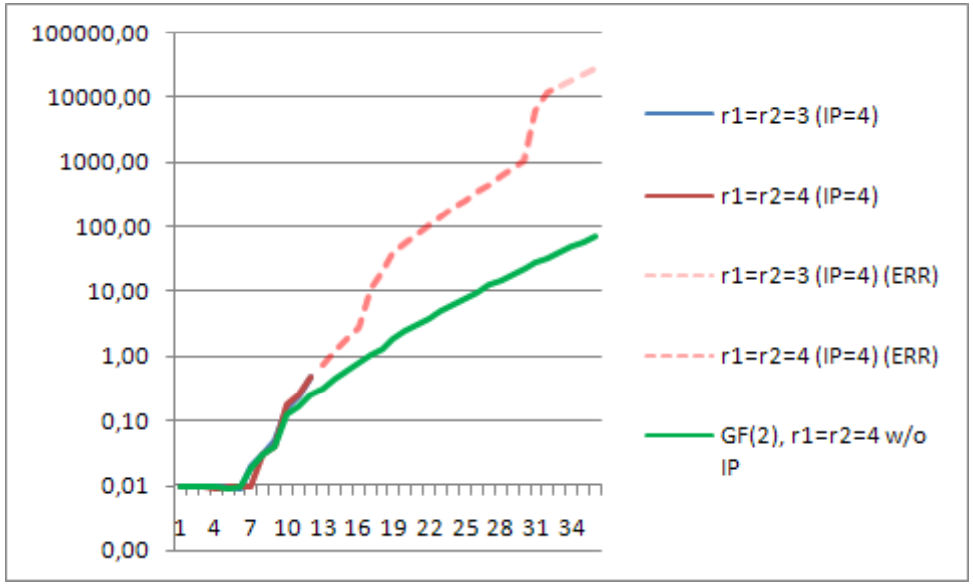
- external perturbation has failed
- Hope to save HFE by adding internal perturbation
- result (not good):



# Another aspect of IPHFE

TECHNISCHE UNIVERSITÄT DARMSTADT

- F4/F5 were able to recover plaintext so fast because:
- they ignored polynomials with degree > 4 during computation
- →IPHFE should confuse
- →GB-algorithms should not be able to do this anymore
- result (unclear):



# Conclusion

TECHNISCHE UNIVERSITÄT DARMSTADT

- Message: HFE can defeat GB-attack when selecting a prime nr

  q>2!

- Internal perturbation is not a good stand-alone adaption

# THANK YOU FOR LISTENING!

# Questions? Remarks?