Multivariate Cryptography: Design of Selected Schemes

Olivier Billet

April 30, 2007

Multivariate Crypto: Selected Schemes

- preliminaries
- hard mathematical problems:
  - MQ–Multivariate Quadratic system
  - IP–Isomorphism of Polynomials
  - MinRank
- selected asymmetric schemes covered in this talk:
  - C^*, SFLASH, HFE, PMI
  - Birational Permutations, OV, UOV, Rainbow
- multivariate scheme in a group setting:
  - traceable block cipher
- multivariate schemes for symmetric cryptography:
  - QUAD stream cipher

Multivariate Polynomials

- multivariate polynomials are just polynomials in several variables
- we are interested in polynomials over finite fields
- every function over $\mathrm{GF}(q^n)$ can be seen as a univariate polynomial:

$$p(x) = \sum_{0 \leqslant i < q^n} a_i x^i$$

- when viewing $\mathrm{GF}(q^n)$ as an extension of degree n over $\mathrm{GF}(q)$:

$$x = \boxed{x_1}\,\boxed{x_2}\,\boxed{x_3}\,\cdots\,\boxed{x_{n-1}}\,\boxed{x_n} \qquad \mathrm{GF}(q^n) \simeq \mathrm{GF}(q)[z]/(z^n + \cdots)$$

thus p can be written as a set of multivariate polynomials over $\mathrm{GF}(q)$:

$$p = (p_1, \ldots, p_n) \quad \text{where} \quad p_i(x_1, \ldots, x_n) = \sum_{\alpha \in \mathbf{N}^n} a_\alpha \prod_{1 \leqslant i \leqslant n} x_i^{\alpha_i}$$

Quadratic Multivariate Polynomials

- a generic multivariate polynomial of degree d in n unknowns defined over $\mathrm{GF}(q)$ is handled with complexity $O\left(\binom{n+d}{d}\right)$
- efficiency reasons ask for quadratic polynomials!
- over $\mathrm{GF}(q)$, the Frobenius mapping $x \longmapsto x^q$ is $\mathrm{GF}(q)$-linear
- hence, any polynomial over $\mathrm{GF}(q^n)$ of the form

$$p(x) = \sum_{0 \leqslant i,j < n} a_{i,j} x^{q^i + q^j}, \qquad a_{i,j} \in \mathrm{GF}(q^n)$$

can be expressed as

$$p(x) = (q_1(x_1, \ldots, x_n), \ldots, q_n(x_1, \ldots, x_n))$$

where each q_i is a multivariate quadratic polynomial:

$$q_i(x_1, \ldots, x_n) = \sum_{1 \leqslant i \leqslant j \leqslant n} b_{i,j} x_i x_j, \qquad b_{i,j} \in \mathrm{GF}(q)$$

MQ: Multivariate Quadratic Systems

$$k = 1, \ldots, m \qquad \sum_{1 \leqslant i \leqslant j \leqslant n} \alpha_{i,j}^{(k)} x_i x_j + \sum_{1 \leqslant i \leqslant n} \beta_i^{(k)} x_i + \delta^{(k)} = y_k$$

- NP-complete problem via reduction from 3-SAT
  thought to be hard on the average
- easy for $m = 1$ (multivariate polynomial roots)
- easy for $m = O(n^2)$ (linearisation)
- $(x_1, \ldots, x_n) \longmapsto (y_1, \ldots, y_m)$ as one way function?
- Bardet, Faugère, and Salvy 2004
- complexity for generic overdefined systems over $\mathrm{GF}(2)$

IP: Isomorphism of Polynomials

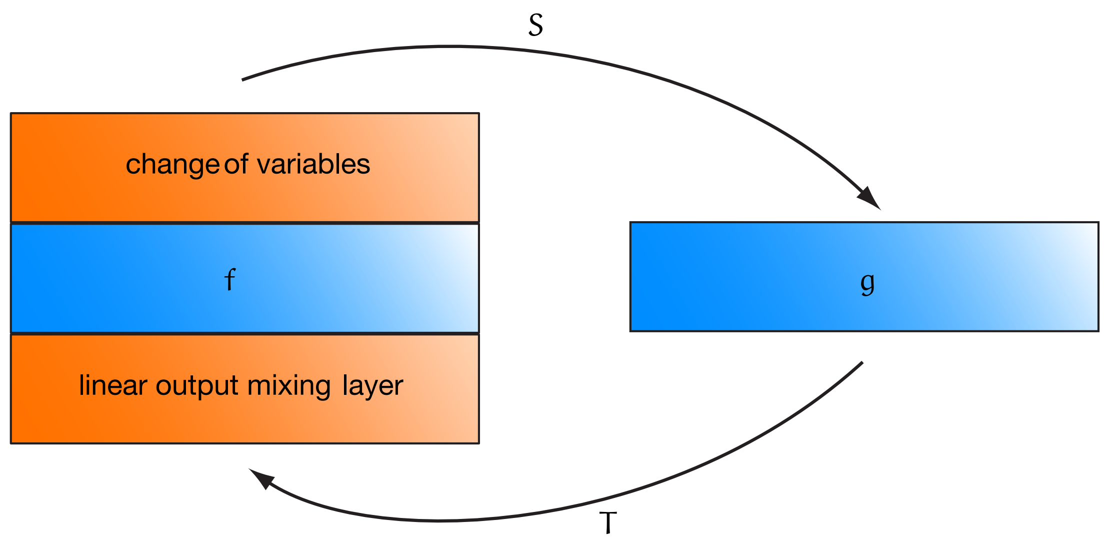
- introduced by Patarin in 1996
- $F = (f_i)_{1 \leqslant i \leqslant m}$ and $G = (g_i)_{1 \leqslant i \leqslant m}$ systems of multivariate equations
- IP with one secret

$$F \overset{\mathsf{IP}}{\sim} G \iff \exists s \in \mathsf{GL}(n, \mathsf{GF}(q))\ \forall i \in [\![1, m]\!] \quad f_i(x_1, \ldots, x_n) = (g_i \circ s)(x_1, \ldots, x_n)$$

- IP with two secrets

$$F \overset{\mathsf{IP}}{\approx} G \iff \exists (s, t) \in \mathsf{GL}(n, \mathsf{GF}(q))^2\ \forall k \in [\![1, m]\!] \quad \sum_{1 \leqslant i \leqslant m} t_{k,i}\, f_i(x_1, \ldots, x_n) = (g_k \circ s)(x_1, \ldots, x_n)$$

IP: Isomorphism of Polynomials



- Patarin, Goubin, and Courtois 1998
  - IP with one secret is at least as hard as GI
  - decisional IP with two secrets is not NP-complete unless the polynomial heriarchy collapses
- Geiselmann, Meier, and Steinwandt 2003
  Levy-dit-Vehel and Perret 2003, Perret 2005, Faugère and Perret 2006

MinRank

- given a set $\{M_1, \ldots, M_m\}$ of $n \times n$ matrices defined over GF(q), find a linear combination of the M_i having a small rank,

$$\text{Rank}\left(\sum_{i=1}^{m} \lambda_i M_i\right) \leqslant r$$

- decisional problem is NP-complete for varying r but polynomial when r fixed
- leads to powerful cryptanalysis of some multivariate cryptosystems
- naïve algorithm: $O(q^m r^\omega)$

MinRank: Algorithms

- Goubin and Courtois 2000
- assume $M = \lambda_1 M_1 + \cdots + \lambda_m M_m$ has rank lower than r
- randomly choose $\left\lceil \frac{m}{n} \right\rceil$ vectors $x_1, \ldots, x_{\left\lceil \frac{m}{n} \right\rceil}$ and hope they lie in the kernel of M
- happens with proba. greater than $q^{r\left\lceil \frac{m}{n} \right\rceil}$ and then the following holds:

$$\forall j \in \{1, \ldots, \lceil m/n \rceil\} \quad 0 = \left(\sum_{i=1}^{m} \lambda_i M_i \right) x_j = \sum_{i=1}^{m} \lambda_i \left(M_i x_j \right)$$

- solving the resulting system in λ takes $O(m^\omega)$
- overall complexity $O\left(q^{r\left\lceil \frac{m}{n} \right\rceil} m^\omega \right)$

Asymmetric

Multivariate

Constructions

Scheme C^*

- Matsumoto and Imai 1985

- n unknowns over the finite field $GF(q)$
- uses an embedding

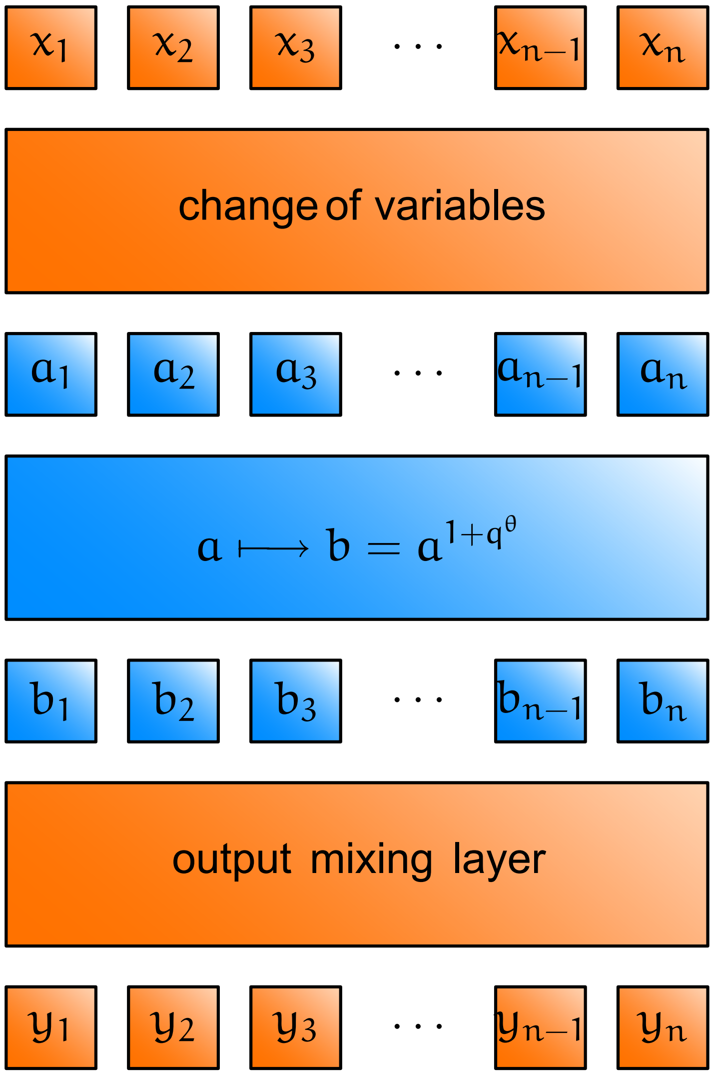
$$\Phi : GF(q)^n \longrightarrow GF(q^n)$$

- the internal mapping is $a \mapsto a^{1+q^\theta}$
- this internal mapping is $GF(q)$-quadratic
- public key can be described by:

$$y_k = \sum_{1 \leqslant i \leqslant j \leqslant n} \alpha_{i,j}^{[k]} x_i x_j$$

- decryption

Cryptanalysis of C^*



- in 1995, Patarin noticed:

$$b = a^{q^\theta+1} \iff b^{q^\theta-1} = a^{q^{2\theta}-1}$$

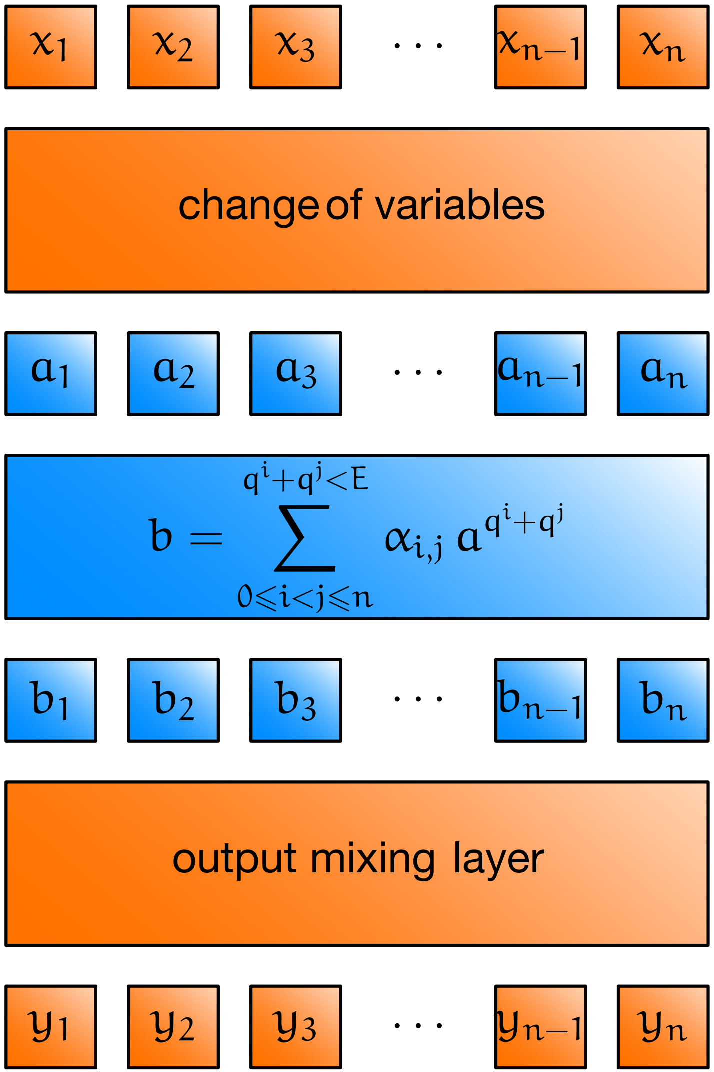
- multiplying this equation by ab gives

$$ab^{q^\theta} = a^{q^{2\theta}}b$$

- with many plaintext/ciphertext pairs interpolate these bilinear equations
- then fix y to the value of some ciphertext
- solve for x in the linear system you get

- underlying IP problem resists [FP06]

HFE: Hidden Field Equation

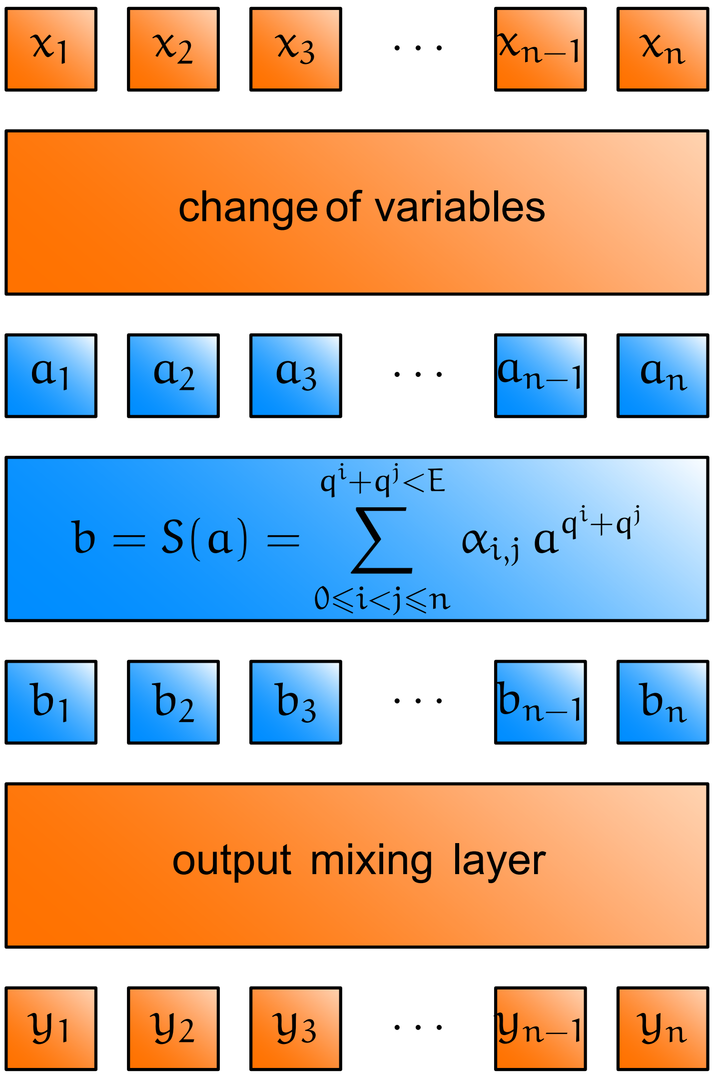


- Patarin 1996
- generalizing the internal transformation to:

$$\sum_{0 \leqslant i < j \leqslant n}^{q^i+q^j<D} \alpha_{i,j}\, a^{q^i+q^j}$$

- still quadratic but thwarts Patarin's attack
- usual univariate polynomial solving (like Berlekamp's algorithm) allows a legitimate user to invert the polynomial provided degree is bounded $q^i + q^j < D$
- polynomial needs not be a bijection (but then redundancy is necessary)

HFE: Cryptanalysis



- small rank attack Kipnis and Shamir 1999
  but unknown complexity

- Gröbner basis Faugère and Joux 2003

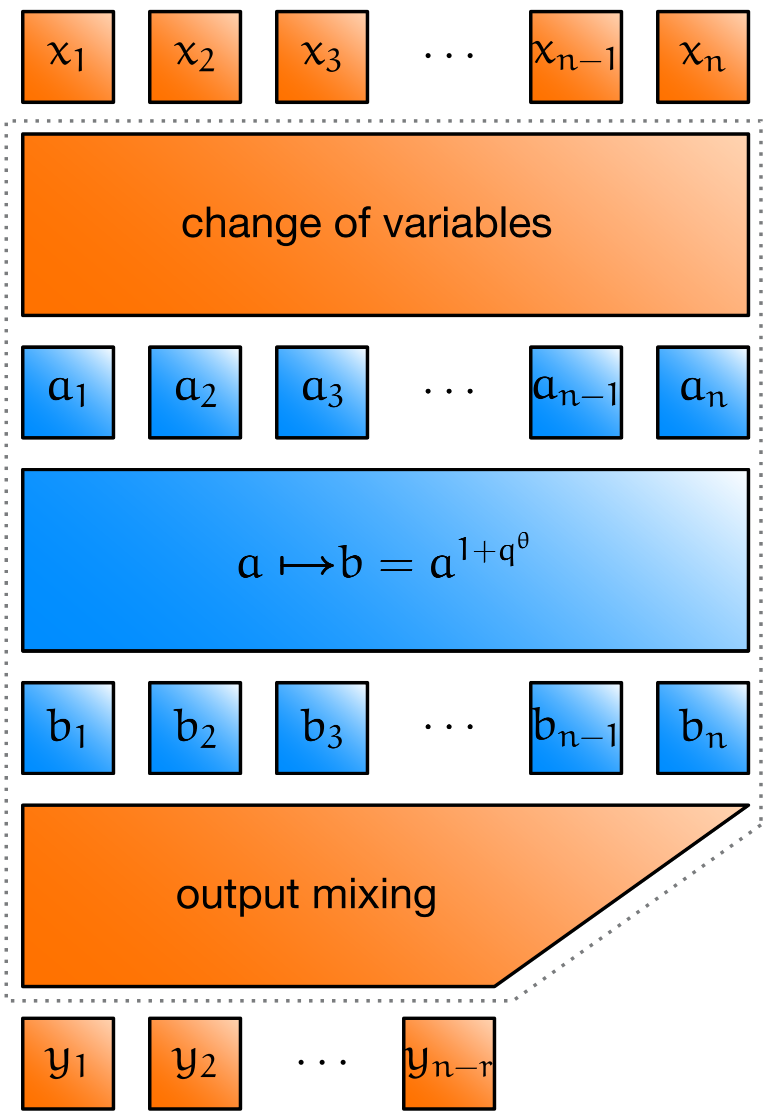
  80 polynomials in 80 binary unknowns

$$\forall (d_i, j) \quad x^{d_i} S(x)^{q^j}$$

  d_i has q-Hamming weight lower than H

- HFE inversion is quasi-polynomial
  Granboulan, Joux, and Stern 2006
- actually Courtois 2001: [KS99] + [CSV93]
  key recovery on HFE is quasi-polynomial

SFLASH: Signature Scheme



- Patarin, Goubin, and Courtois in 1998
- thwarts C^* attack
- Gröbner basis: now several solutions
- 2003 highly efficient implementation Akkar, Courtois, Goubin, and Duteuil
- SFLASHv1 broken
  - Gilbert and Minier
  - linear functions are in subfield
- SFLASHv2 broken (as well as v3)
  - Dubois, Fouque, Shamir, and Stern
  - $\mathrm{GF}(q) = \mathrm{GF}(2^7)$, $n = 37$, $r = 11$

Introducing Randomness

- $+$: Patarin 1998
  - randomly choose a small number of polynomials in the n unknowns $\rho_1(x_1,\ldots,x_n),\ldots,\ \rho_c(x_1,\ldots,x_n)$
  - public key $q_1,\ldots,\ q_m$ now becomes

$$q_1+\sum_{i=1}^{c}\lambda_i^{(1)}\rho_i,\ \ldots,\ q_m+\sum_{i=1}^{c}\lambda_i^{(m)}\rho_i$$

- PMI: Ding 2004 broken by Fouque, Granboulan, and Stern 2005
  - randomly choose m polynomials in a small number of unknowns $\rho_1(x_1,\ldots,x_c),\ldots,\ \rho_m(x_1,\ldots,x_c)$
  - public key $q_1,\ldots,\ q_m$ now becomes

$$q_1+\rho_1,\ \ldots,\ q_m+\rho_m$$

Birational Permutations

- Shamir 1993
- internal transformation F

$$\begin{cases} f_1(x_1) = x_1, \\ f_2(x_1, x_2) = l_2(x_1) \cdot x_2 + q_2(x_1), \\ f_3(x_1, x_2, x_3) = l_3(x_1, x_2) \cdot x_3 + q_3(x_1, x_2), \\ \vdots = \ddots \\ f_n(x_1, x_2, \dots, x_n) = l_n(x_1, x_2, \dots, x_{n-1}) \cdot x_n + q_n(x_1, x_2, \dots, x_{n-1}), \end{cases}$$

- public key: $G = T \circ F \circ S$

Birational Permutations: Cryptanalysis

- Coppersmith, Stern, and Vaudenay 1993

- a public polynomial has the form: $g_k = \delta_k f_n + \sum_{2 \leqslant i < n} t_{k,i} \cdot f_i \circ s$
- $f_n(x_1, x_2, \ldots, x_n) = l_n(x_1, x_2, \ldots, x_{n-1}) \cdot x_n + q_n(x_1, x_2, \ldots, x_{n-1})$
- how to remove the contribution of f_n? use rank reduction!

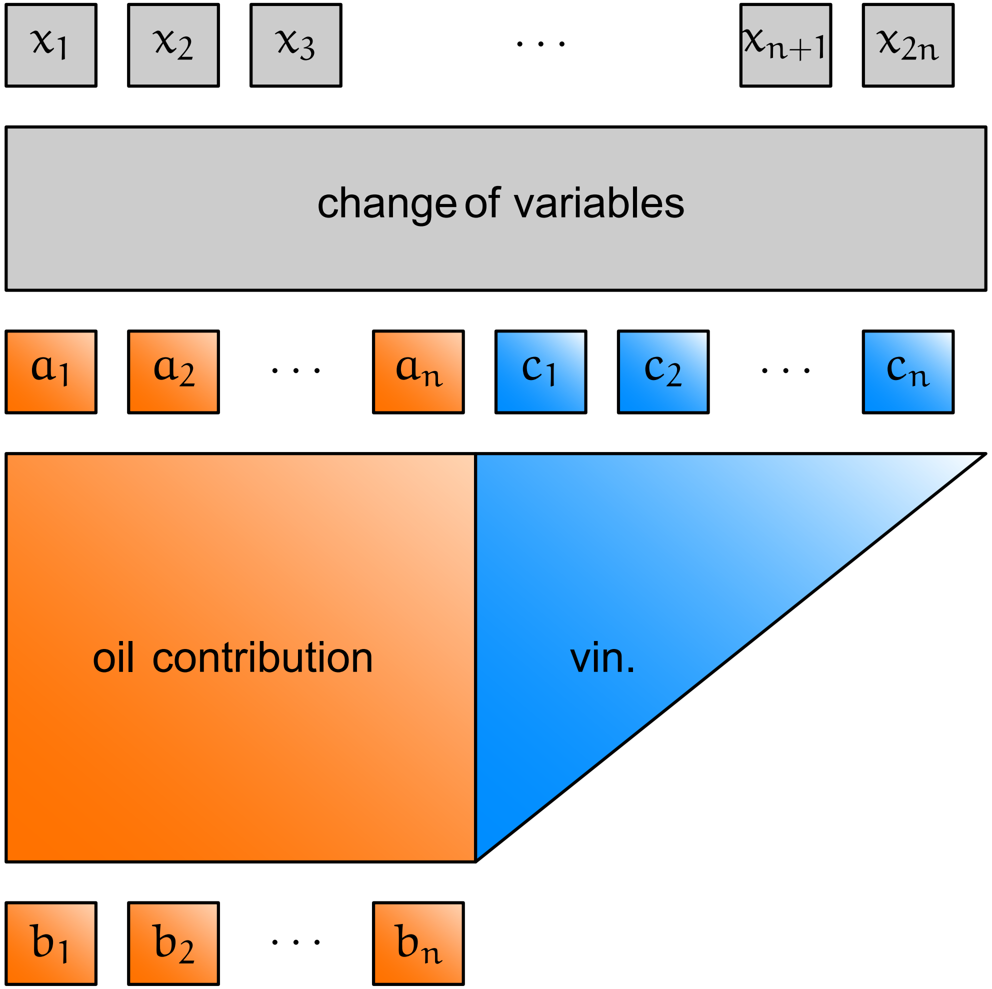
$$\begin{pmatrix} \boxed{Q_\lambda} & \boxed{u_\lambda} \\ \boxed{u_\lambda^T} & 0 \end{pmatrix}$$

- $\det(g_i - \lambda g_j) = 0$ holds when $\lambda = \delta_i / \delta_j$
- reveals λ as double root of the above determinant

Oil and Vinegar

- Patarin in 1997
- output variables of the internal transformation have the special form:

$$b_k = \sum_{1 \leqslant i \leqslant j \leqslant n} \alpha_{i,j}^{[k]} a_i c_j + \sum_{1 \leqslant i \leqslant j \leqslant n} \beta_{i,j}^{[k]} c_i c_j + \sum_{1 \leqslant l \leqslant n} \gamma_l^{[k]} c_l + \delta_l^{[k]} a_l + \eta^{[k]}$$

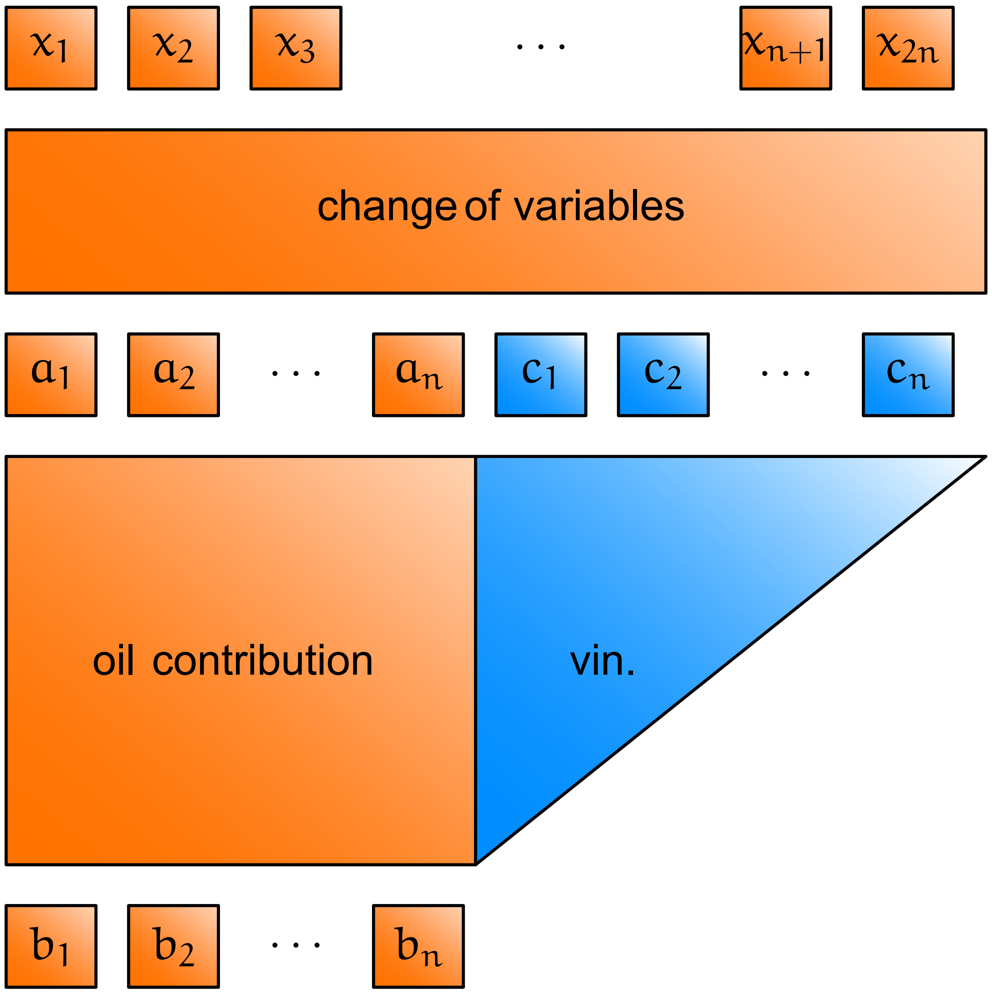


Oil and Vinegar: Public Key

- public system of equations after applying the secret change of variables

- output variables take the form:

$$b_k = \sum_{1 \leqslant i \leqslant j \leqslant 2n} \alpha_{i,j}^{[k]} x_i x_j + \sum_{1 \leqslant l \leqslant 2n} \gamma_l^{[k]} x_l + \eta^{[k]}$$



Oil and Vinegar: Cryptanalysis

- Kipnis and Shamir 1998
- S the secret change of base
- G_i the bilinear matrix associated to the i-th output polynomial
- bilinear matrix F_i of i-th internal polynomial:

$$F_i = \begin{pmatrix} 0 & M^{[i]}_{h,v} \\ M^{[i]}_{v,h} & M^{[i]}_{h,h} \end{pmatrix}$$

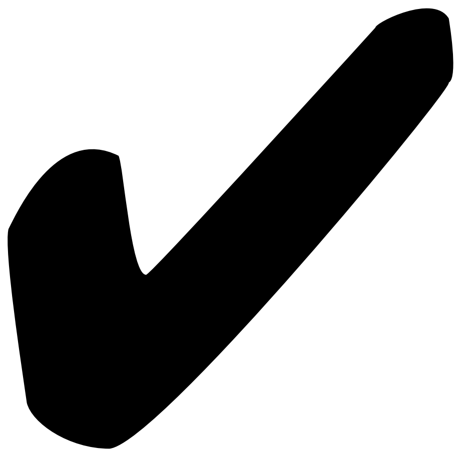
  when F_j invertible, $F_i\,F_j^{-1}$ fixes the space of oil variables
- $G_i = {}^{\mathsf{T}}S\,F_i\,S$ and when G_j invertible
  matrix $G_j^{-1}\,G_i = S^{-1}\,F_j^{-1}\,F_i\,S$ fixes the preimage through S
  of the vector space of oil variables

Unbalanced Oil and Vinegar

- Kipnis, Patarin, and Goubin 1998
- n oil variables and m vinegar variables
  - weak for $m < n$ and $m \sim n$
  - weak for $m = O(n^2)$
  - no known attack for $m = c \cdot n$, c small provided q^m is large enough ($> 2^{80}$)
- Gröbner basis?
  - do not like multiple solutions
  - Courtois, Daum, and Felke 2003

- unbalanced version still not broken

Rainbow Layers of UOV



- Ding and Schmidt 2005
- original version

  27 equations

  33 unknowns

  over GF(2^8)

  broken by [BG06]
- main threat: rank attacks

  attacks still exponential

Rainbow Internal Transformation

✔

- over GF(2^8), dimensions are 11, 5, 5, 6, 6

Multivariate

Traitor Tracing

Scheme

Traitor Tracing Schemes

- Chor, Fiat, and Naor 1994

  key generation, encryption, decryption, tracing

- each of the N users gets a key K_i
  - allows to decrypt broadcasted content
  - uniquely identifies at least one of them

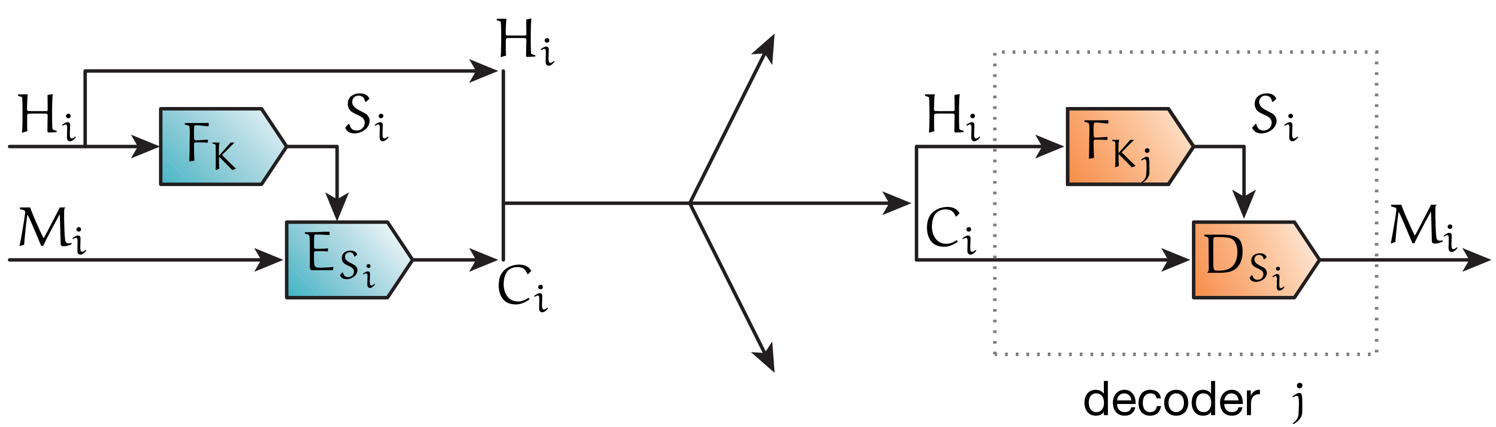
- no coalition of at most k traitors can build a pirate decoder while hiding identities of all the traitors

A Traceable Block Cipher

- [BG 03]
- F_K should be a secure encryption scheme
- key generation of F_{K_j} should verify

$$F_K \equiv F_{K_1} \equiv \cdots \equiv F_{K_j} \equiv \cdots \equiv F_{K_N}$$

- should resist k-coalitions
- how to work with control words



TBC: Principle

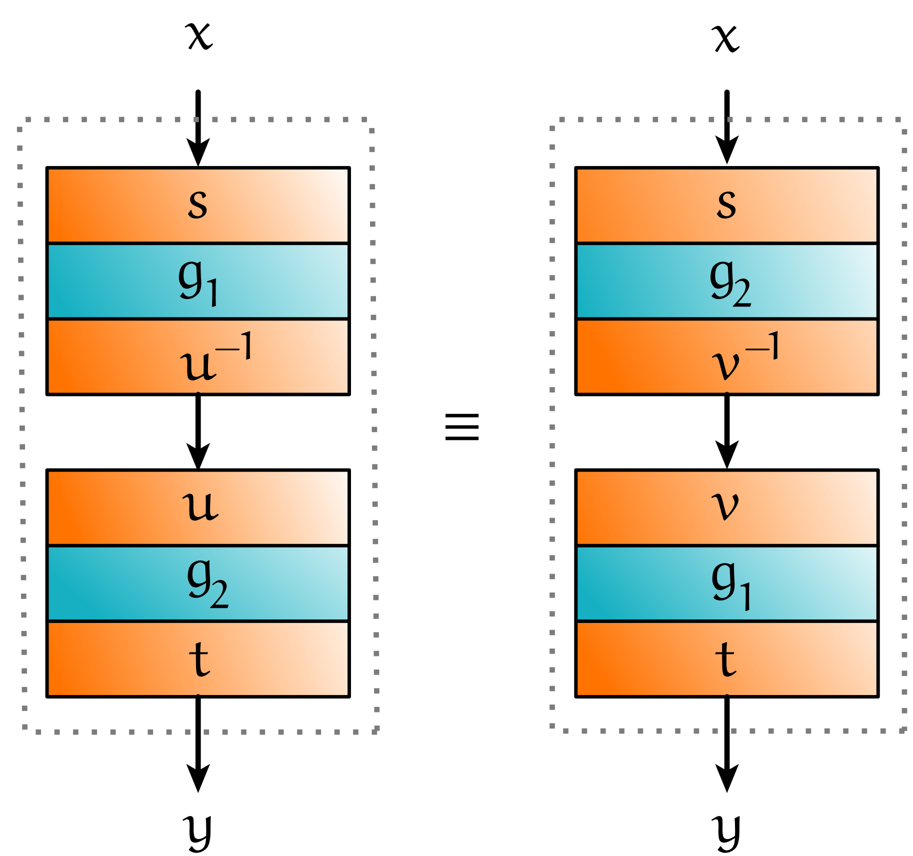
- assuming g_1 and g_2 commute you'll get equivalent descriptions

- this is true when choosing:

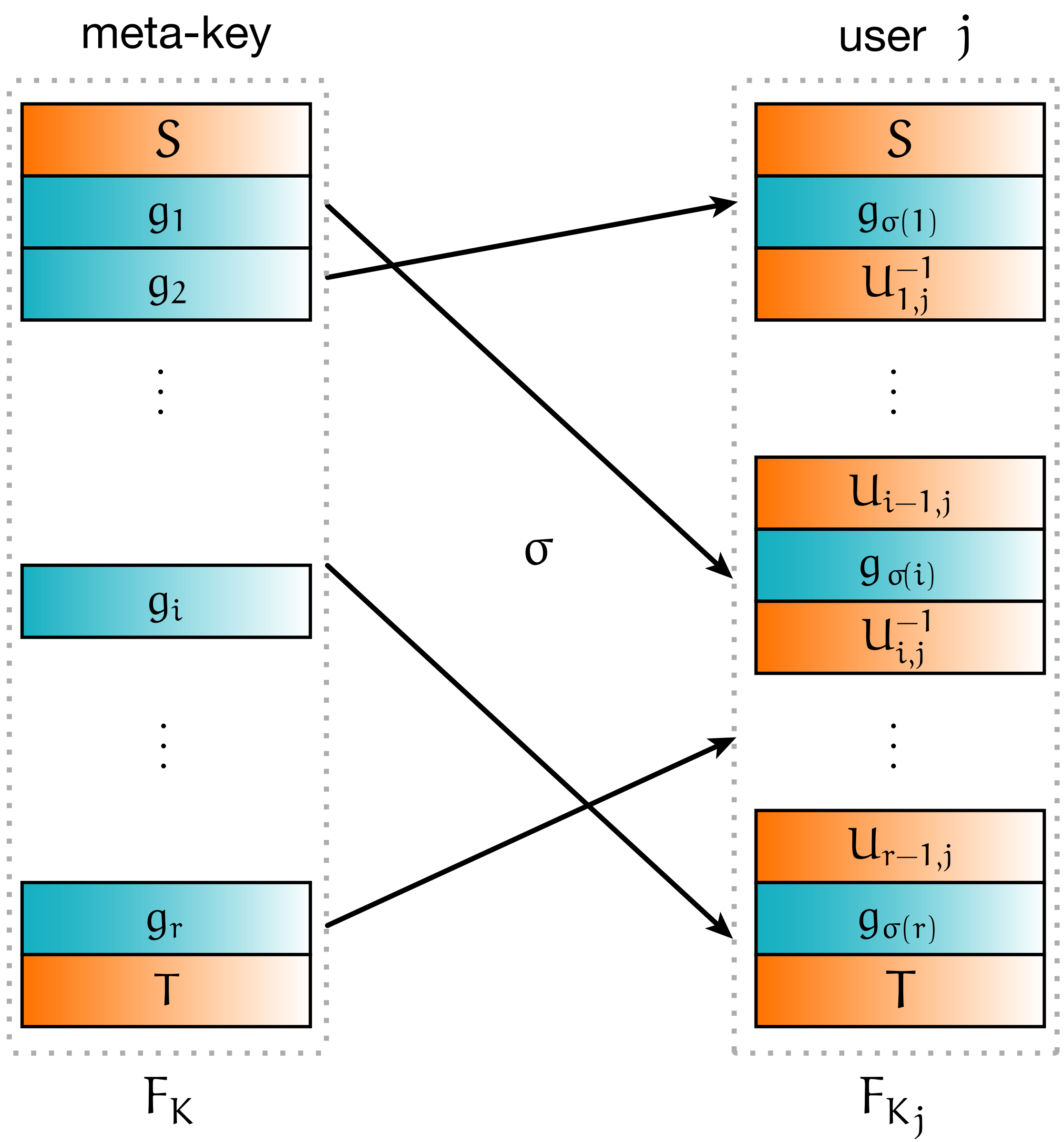
$$g_\theta : a \mapsto b = a^{1+q^{\theta_1}+\cdots+q^{\theta_{d-1}}}$$

- $d > 2$ is enough, even though C^* is invertible (cf. Patarin's attack)

- the interesting hard problem here is the IP problem



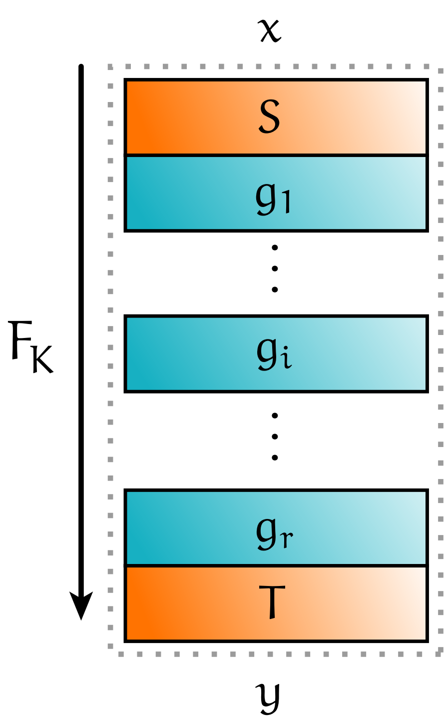
TBC: Key Generation



TBC: Choosing Parameters

from plaintext/ciphertext pairs
an attacker should not be able

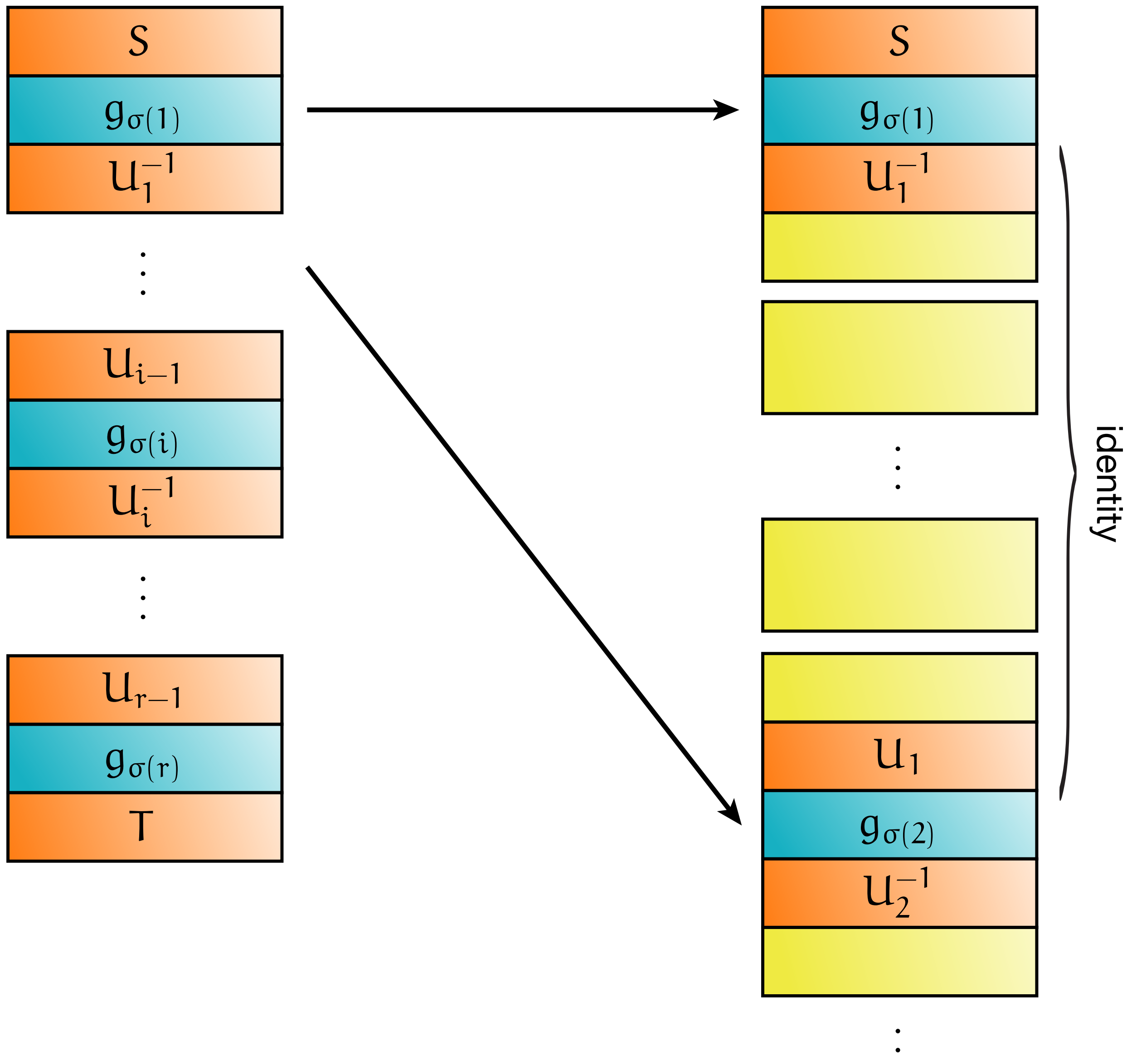
- to determine or interpolate F_K
- to distinguish F_K from a PRP



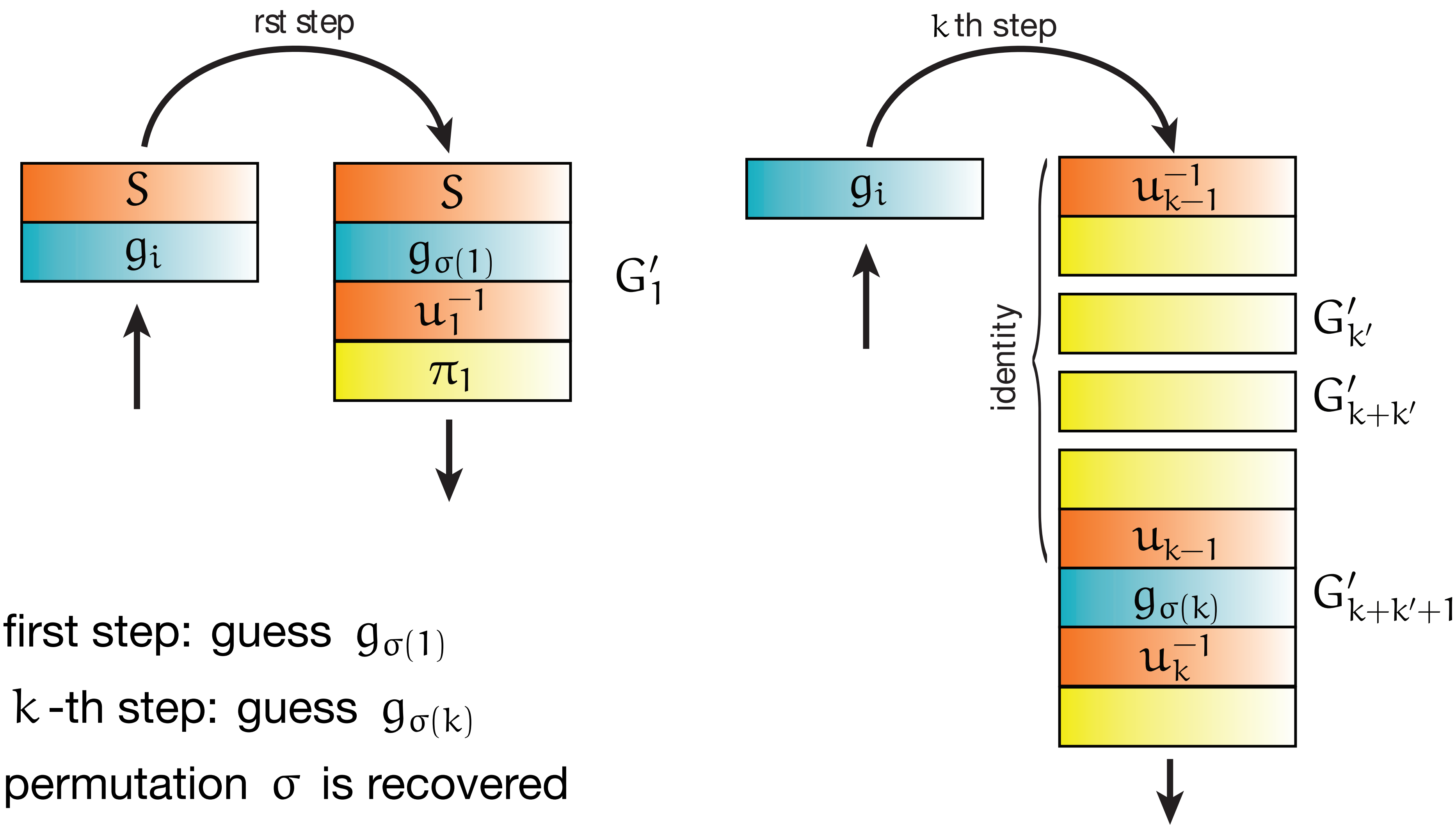
sample parameters:

- GF(2^9), $n = 19$, $d = 3$, $r = 33$
- there are about 1330 distinct monomials
- each block $G_{i,j}$ requires 26790 multiplications
- the whole description fits in 916 Ko

TBC: Single Traitor's Strategy



TBC: Tracing a Single Traitor



- first step: guess $g_{\sigma(1)}$
- k -th step: guess $g_{\sigma(k)}$
- permutation σ is recovered

Multivariate Symmetric Cryptography

MQ and Hash Functions

- multivariate quadratic systems provide a one-way primitive
- there is no need to embed a trapdoor here
- why not use it as a compression function?

- answer:
  assuming q is multivariate quadratic compression function:

$$q(x+\delta) = q(x) \Longleftrightarrow q(x+\delta) - q(x) = 0$$

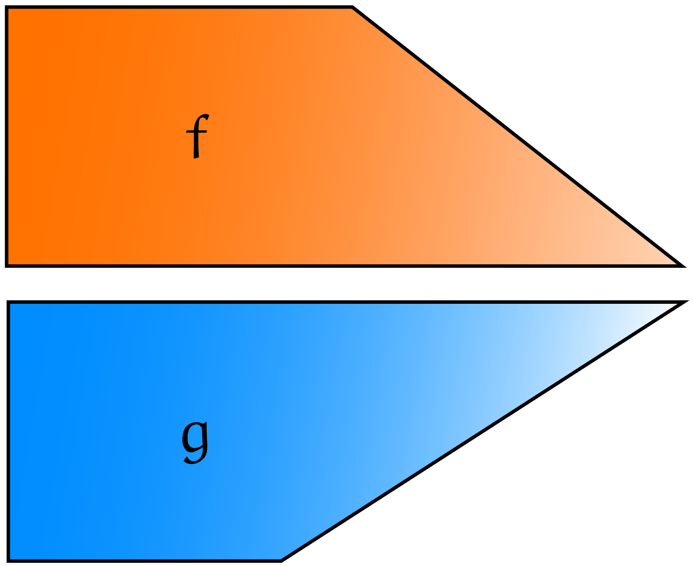
but $q(x+\delta) - q(x) = q(\delta) + B(x,\delta)$ where B is a bilinear function,
so drawing a random δ

$$q(x+\delta) = q(x) \Longleftrightarrow q(\delta) + B(x,\delta) = 0$$

which is linear with respect to x

MQ-Hash

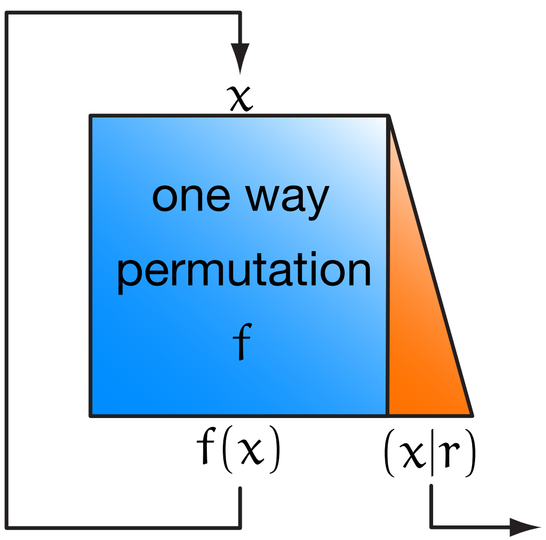
- is there any way
  to work around this issue?
- [BRP 07] propose using
  a one-way function
  as a preprocessing
- however, preprocessing
  must be collision free!



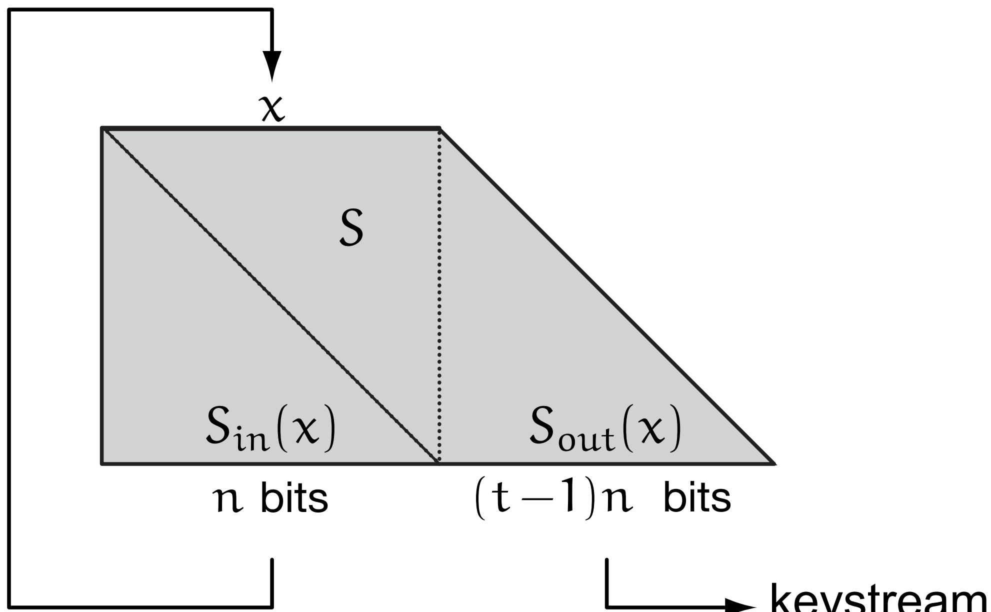
- this design shares some ideas with [AHV 98]
- security proof for collision resistance?
- efficiency issues

PRNG from One Way Functions

- seminal work by [BM84, Y82, GL89, ILLH99]
- constructions rely on various assumptions:
  - discrete logarithm [BM84]
  - RSA assumption [ACGS84]
  - quadratic residuosity [BBS86]
  - subset sum problem [IN96]
  - syndrome decoding problem [FS96]
  - and others . . .
- most constructions are impractical
- usually extracts $O(\log n)$ linear bits per iteration



QUAD: A Multivariate Stream Cipher



- Berbain, Gilbert, and Patarin 2006
- aims to use the MQ problem to build one way functions
- how much can be extracted from the state? IV setup? sizes, efficiency?

QUAD: Keystream Generation

- internal state $x = (x_1, \ldots, x_n) \in \mathrm{GF}(q)^n$
- iteration of a set $S = (Q_1, \ldots, Q_{tn})$
  of tn quadratic multivariate polynomials in n unknowns

✔

- at each iteration:
  - compute and output $S_{\mathrm{out}}(x)$ as keystream bits
  - compute $S_{\mathrm{in}}(x)$ and use it to update x

QUAD: Key and IV Setup

- uses two publicly known systems S_0 and S_1

✔

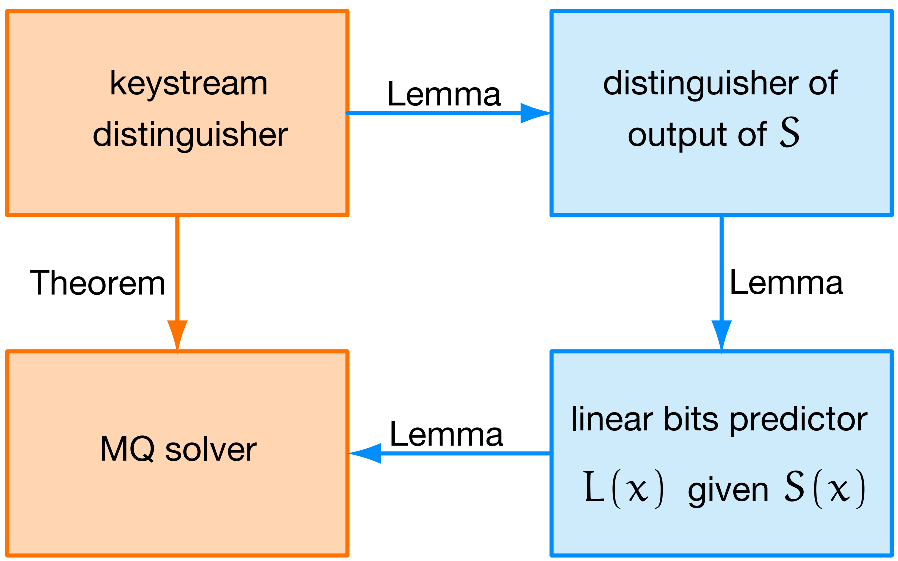
- x initialized with the key K (padded to n bits)
- for each bit IV_i:
  - compute $S_0(x)$ and $S_1(x)$
  - update x with $S_{IV_i}(x)$
- runup: clock the cipher n times without outputting any keystream

QUAD: Performances

- version recommended by the authors: $n = 160$, $m = 320$ over GF(2)
- Software performances [BBG06]
  - over GF(2) : 2081 cycles/byte
  - bigger fields might reveal bad tradeoffs in practice [YCBC07]
- Hardware performances [ABBG07]
  - compact implementation: 3694 GE, 9.5 kbps
    QUAD virtually fits any RFID !!!
  - best size/throughput ratio: 10184 GE, 3.3Mbps

QUAD: Idea of the Proof

Theorem: any distinguisher of a $L = \lambda(t-1)n$-bit keystream sequence running in time T with prob. ϵ over all quadratic systems S and over all initial state values x can be converted into an MQ solver running in time $T' = O(\frac{n^2\lambda^2}{\epsilon^2}T)$ with probability $\frac{\epsilon}{2^3\lambda}$



Openings

- we presented a selection of multivariate schemes
- most of them were cryptanalysed through their algebraic structure
- still a lot of things to understand (HFE^{--}, UOV)
- successful attacks against multivariate cryptosystems suggested new ways to attack symmetric systems such as AES or stream ciphers
- Gröbner basis via $\mathrm{F4}$, $\mathrm{F5/2}$ revealed bad algebraic properties
- multivariate symmetric schemes are promising (they don't need to embed a trapdoor)

special thanks to

Ryad Benadjila, Côme Berbain,

Henri Gilbert, and Yannick Seurin

Questions?