CRYPTANALYSIS OF THE
HIDDEN FIELD EQUATIONS (HFE)
PUBLIC KEY CRYPTOSYSTEM

ADI SHAMIR
APPLIED MATH DEPT
THE WEIZMANN INSTITUTE
ISRAEL

JOINT WORK WITH:

AVIAD KIPNIS
NDS TECHNOLOGIES
ISRAEL
**Motivation:**

The RSA Scheme:

Based on the single univariate equation

\[ x^e = c \pmod{n} \]

Disadvantages:

- **Security:** Based on factoring
- **Efficiency:** Slow for large \( n \)

A Multivariate Scheme:

Based on many multivariate equations over a (small) finite field:

\[ p_0(x_0, x_1, \ldots, x_{n-1}) = c_0 \]
\[ p_1(x_0, x_1, \ldots, x_{n-1}) = c_1 \]
\[ \vdots \]
\[ p_{m-1}(x_0, x_1, \ldots, x_{n-1}) = c_{m-1} \]

**Encryption:** \( x \rightarrow c \)

**Decryption:** \( c \rightarrow x \)

Usually \( m = n \)
ADVANTAGES:

- The general problem is NP-complete (even for quadratic equations over the two element field $F_2$).
- The operations are very fast.

DISADVANTAGES:

- The public key is large.
- A long history of failures.
PATARIN'S HFE ENCRYPTION SCHEME

A SMALL FIELD \( F \) \( \leftrightarrow \) AN EXTENSION FIELD \( K \) OF DEGREE \( n \) OVER \( \mathbb{F}_q \)

\((x_0, x_1, \ldots, x_{n-1}) \in F^n \leftrightarrow x = \sum_{i=0}^{n-1} x_i w_i \in K\)

MANY MULTIVARIATE EQUATIONS OVER \( F \) \( \leftrightarrow \) A SINGLE UNIVARIATE EQUATION OVER \( K \)

QUADRATIC EQUATIONS \( \leftrightarrow \) A SPECIAL TYPE OF EQUATION

SOLVING EQUATIONS IS NP-COMPLETE \( \leftrightarrow \) BERLEKAMP'S PROB. POLYNOMIAL TIME ALG.

THE BASIC IDEA:

LET \( |F| = q \). THEN \( K \) HAS CHARACTERISTIC \( q \):

\[(a + b)^q = a^q + b^q + \sum_{i=1}^{q-1} \binom{q}{i} a^{q-i} b^i\]

\[(a + b)^q = a^q + b^q \Rightarrow (x)^q \text{ IS A LINEAR OPERATION} \]

\[\Rightarrow (x)^{q^i} \text{ IS A LINEAR OPERATION} \Rightarrow \]

\[x^{q^i + q^j} = x^{q^i} \cdot x^{q^j} \text{ IS A QUADRATIC MAPPING} \]

THE SECRET KEY IN PATARIN'S SCHEME:

\[P(x) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} P_{ij} x^{q^i + q^j} \text{ OF MAX DEGREE } d \text{ OVER THE LARGE FIELD } K\]
THE KEY GENERATION PROCESS:

- Choose parameters
  (Typical: $|F|=q=2$, $|K|=q^m=2^{128}$, $d=8,000$)

- Choose secret univariate polynomial:
  \[ P(x) = \sum_{i=0}^{12} \sum_{j=0}^{12} P_{ij} x^i + q^j \]

- Represent this polynomial as a system of \textit{m} quadratic polynomials in \textit{n} variables over \( F \):
  \[ P_0(x_0,\ldots,x_{n-1}), P_1(x_0,\ldots,x_{n-1}), \ldots, P_{n-1}(x_0,\ldots,x_{n-1}) \]

- Hide the special structure with an input linear transformation \( S \) and an output linear transformation \( T \).

- Publish the resultant system of \textit{m} quadratic polynomials in \textit{n} variables over the small field \( F \).
UNIVARIATE REPRESENTATIONS OF SYSTEMS OF MULTIVARIATE POLYNOMIALS

**Lemma:** \( x \mapsto \sum_{i=0}^{n-1} a_i x^i \) represents a linear mapping over \( n \)-tuples of values in \( F \).

**Lemma:** Any linear mapping over \( n \)-tuples of values in \( F \) has such a representation for some \( a_0, a_1, \ldots, a_{n-1} \in K \).

**Proof:** By a counting argument:

\[
\#\text{linear mappings} = \# n \times n \text{ matrices} = q^{n^2}
\]

\[
\#\text{such polynomials} = \# (a_0, \ldots, a_{n-1}) = (q^n)^n
\]

\[P_1(x) = P_2(x) \Rightarrow P_1(x) - P_2(x) \text{ has degree} \leq n-1, q^n \]

**Remark:** Once we know that such a representation exists, its coefficients can be found by interpolation.
UNIVARIATE REPRESENTATIONS OF
ARBITRARY SYSTEMS OF POLYNOMIALS

EXAMPLE:

\[(x_1, x_2, x_3) \rightarrow (x_1 x_3 + x_2^2, 3x_1^2 + 2x_3^2, x_1 x_2 + x_2 x_3)\]

POSSIBLE PROBLEMS:

- The univariate representation may not exist.
- The univariate representation may be exponentially large.
- The univariate representation may be hard to compute.
GENERALIZATION TO ARBITRARY SYSTEMS OF MULTIVARIATE POLYNOMIAL

THEOREM: FOR ANY \( p_0(x_0, \ldots, x_{n-1}), \ldots, p_{n-1}(x_0, \ldots, x_{n-1}) \), there are coefficients \( a_0, a_1, \ldots, a_{q^{n-1}} \in \mathbb{K} \) such that \( y = \sum_{i=0}^{q^{n-1}} a_i x^i \), where
\[
x = \sum_{i=0}^{n-1} x_i w_i, \quad y = \sum_{i=0}^{n-1} y_i w_i, \quad \forall \gamma \in p_i(x_0, \ldots, x_{n-1})
\]

PROOF: WLOG, assume that \( w_0 = 1 \).
- \((x_0, \ldots, x_{n-1}) \mapsto (x_0, 0, \ldots, 0)\) is linear, so it has a univariate representation.
- \((x_0, \ldots, x_{n-1}) \mapsto (x_i x_j, 0, \ldots, 0)\) can be represented by the product of the two univariate polynomials.
- \((x_0, \ldots, x_{n-1}) \mapsto (p_k(x_0, \ldots, x_{n-1}), 0, \ldots, 0)\) by generalization.
- \((x_0, \ldots, x_{n-1}) \mapsto (0, 0, \ldots, p_k(x_0, \ldots, x_{n-1}), \ldots, 0, 0)\) can be obtained by multiplying all coefficients by \( w_k \).
- \((x_0, \ldots, x_{n-1}) \mapsto (p_0(x_0, \ldots, x_{n-1}), \ldots, p_{n-1}(x_0, \ldots, x_{n-1}))\) can be obtained by summing all the univariate representation.
Corollary: given a collection of $m$ homogeneous multivariate polynomials of degree $d$ in $m$ variables over $F$, the only powers of $x$ which appear in their univariate representation are sums of exactly $d$ (not necessarily distinct) powers of $x^i$: $x^{g_1} + x^{g_2} + \ldots + x^{g_d}$.

Corollary: if $d$ is a constant, the polynomial is sparse, and can be constructed in poly time (in $m$).

Corollary: solving a single univariate polynomial equation over a finite field

- in random polynomial time if the polynomial is represented by a list of all its coefficients
- NP-complete if the polynomial is represented by a list of its non-zero coefficients.
HOW TO REPRESENT THE PUBLIC KEY

\[ G_0(x_1, \ldots, x_{m-1}), G_1(x_0, x_1, \ldots, x_{m-1}), \ldots, G_m(x_0, \ldots, x_{m-1}) \]

Each \( G_i \) is a quadratic function

- Bad Representation:

\[
\begin{bmatrix}
   x_0 & x_1 & \ldots & x_{m-1}
\end{bmatrix}
\begin{bmatrix}
   x_0 \\
   x_1 \\
   \vdots \\
   x_{m-1}
\end{bmatrix}
\]

- Good Representation:

Combine the published \( G_i \) over \( F \) into a single \( G \) over \( K \) of the form:

\[ G(x) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} g_{ij} x_i^j + q^j \]

then:

\[
\begin{bmatrix}
   x_0^p & x_1^p & \ldots & x_{n-1}^p \\
   x_0^p & x_1^p & \ldots & x_{n-1}^p \\
   \vdots & \vdots & \ddots & \vdots \\
   x_0^p & x_1^p & \ldots & x_{n-1}^p
\end{bmatrix}
\begin{bmatrix}
   x_0^g \\
   x_1^g \\
   \vdots \\
   x_{n-1}^g
\end{bmatrix}
\]
THE EFFECT OF $S$ AND $T$ ON $G(x)$

THE ORIGINAL LOW DEGREE $P(x)$ HAS THE FORM

$$
\begin{bmatrix}
    x^9, x^{9'}, \ldots, x^{127}
\end{bmatrix}
\begin{bmatrix}
    x^9, x^{9'} \\
    x^{127}
\end{bmatrix}
= x^t P x^t
$$

ONLY 1% IS NONZERO

THE REPRESENTATION OF THE PUBLIC KEY:

$$
\begin{bmatrix}
    x^9, x^{9'}, \ldots, x^{127}
\end{bmatrix}
\begin{bmatrix}
    x^9, x^{9'} \\
    x^{127}
\end{bmatrix}
= x G x^t
$$

RANDOM LOOKING BUT KNOWN

$S$ AND $T$ CAN BE REPRESENTED AS UNIVARIATE POLYNOMIALS, AND THUS:

$$
G(x) = T(P(S(x))) \quad \text{OVER } K
$$

OR:

$$
T^{-1}(G(x)) = P(S(x)) \quad \text{OVER } K
$$

WHERE:

$$
S(x) = \sum_{i=0}^{n-1} s_i x^i, \quad T^{-1}(x) = \sum_{i=0}^{n-1} t_i x^i \quad \text{OVER } K
$$
A long evaluation yields:

\[ T^{-1}(G(x)) = \sum_{k=0}^{n-1} t_k G^{*k} \]

\[ P(S(x)) = WPW^t \]

where: \( G^{*k} \) is obtained from \( G \) by:
- raising each entry to the power \( q^k \)
- rotating both rows and columns (cyclically) by \( k \) steps

and: 
\[ W = [w_{ij}]_{n \times m} \quad \text{where} \quad w_{ij} = (s_{j-i})^{q^i} \]

This defines the fundamental equation:

\[ G' = WPW^t \quad \text{over} \quad K \left( \begin{array}{c} \text{equation} \\
\end{array} \right) \]

Each entry is a linear combination in the unknown coefficients \( t_k \) (m unknowns).
Each entry in the \( q^i \) power of some unknown coefficient \( s_{j-i} \) (99x known)
Zero except at top left 13x13 corner (m unknowns).
RECOVERING $T$

In the fundamental equation

$$G' = WPW^t$$

$P$ has low rank (13) so $G'$ has low rank.

Problem: Given a matrix of linear forms, find an assignment which makes the matrix low rank.

Idea: Express the condition as a large number of (quadratic) equations in a small number of variables.

Example: Find $t_1, t_2, t_3 \in K$ making $\text{rank}(G') = 1$.

$$G' = \begin{bmatrix}
t_1 + 3t_3 & 2t_1 + t_2 & t_1 + t_2 + t_3 \\
2t_2 + t_3 & t_2 + 3t_3 & t_1 + 4t_2 \\
t_2 + 2t_3 & t_1 + 2t_2 + t_3 & 3t_2
\end{bmatrix}_{3 \times 3}$$
EQUATIONS EXPRESSING LOW RANK:

$G'_{n \times n}$ HAS RANK $\pi \ll n \Rightarrow$

$G'$ HAS LEFT KERNEL OF DIM $n-\pi$:

$\forall G' = 0 \Rightarrow$

$G'$ HAS LEFT KERNEL OF SIZE $(q^{n-\pi})$

$\Rightarrow \exists v \in \ker(G')$ WITH ANY CHOICE OF FIRST $n-\pi$ COORDINATES IN $K$ (WITH REASONABLE PROBABILITY) $\Rightarrow$

\[
\begin{array}{c|c|c}
(n-\pi) \times m & m \times n & (n-\pi) \times n \\
\hline
\text{KERNEL} & \text{MATRX} & \text{ZEROES} \\
\text{VCTRS} & \text{KERN} & \text{WHT} \\
\text{ARBITRARILY SPECIFIED} & \text{COLUMNS} & \text{WITH m UNCVRNLY VALUES} \\
\text{n-\pi} & \text{PL} & \text{O} \\
\end{array}
\]

WE GET A SYSTEM OF QUADRATIC EQUATIONS WITH

# UNKNOWNS = $\pi(m-\pi) + m \approx \pi \cdot n$

# EQUATIONS = $(n-\pi) \cdot m \approx n^2$

GREATLY OVERDETERMINED WHEN $n=\text{CONST}$, $n \to \infty$
SOLVING OVERDEFINED SYSTEMS OF $e m^2$ QUADRATIC EQUATIONS IN $m$ VAR.

\[
x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 + x_1x_3 + x_2x_3 = 3
\]
\[
2x_1^2 + x_2^2 + 3x_3^2 + x_1x_2 + 2x_1x_3 + 3x_2x_3 = 1
\]
\[
\vdots
\]
\[
4x_1^2 + 2x_2^2 + x_3^2 + 3x_1x_2 + 4x_1x_3 + 2x_2x_3 = 2
\]

THE LINEARIZATION TECHNIQUE:
- DEFINE A NEW SET OF $\frac{m(m+1)}{2} \approx \frac{1}{2} m^2$ VARS

\[y_{ij} = x_i x_j\]

- SOLVE THE $e m^2$ LINEAR EQUATIONS IN
  THE NEW $\frac{1}{2} m^2$ VARIABLES

- THE SOLUTION IS LIKELY TO BE UNIQUE WHEN $e > \frac{1}{2}$

UNFORTUNATELY, IN OUR CASE:

$m = n \cdot m$ VARS, $e m^2 = n^2$ EQUATIONS $\Rightarrow$

$e = \frac{4}{n^2}$ (for $n=13$, $e=\frac{1}{169}$), so we get

only 1% of the required number of EQUATIONS IN THE LINEARIZATION METHOD.
THE NEW RELINEARIZATION METHOD

GIVEN: \( m^2 \) QUADRATIC EQUATIONS IN \( m \) VARIABLES \( x_i \)

LINEARIZE: DEFINE \( \frac{1}{2} m^2 \) NEW VARIABLES \( y_{ij} \)

SOLVE: EXPRESS EACH \( y_{ij} \) AS A PARAMETER

LINEAR EXPRESSION IN \( (\frac{1}{2} - \varepsilon)^m \) NEW \( z_k \)

ADD NEW QUADRATIC EQUATIONS IN THE \( y_{ij} \)

\[
(x_a x_b)(x_c x_d) = (x_a x_c)(x_b x_d) = (x_a x_d)(x_b x_c) \Rightarrow
\]

\[
y_{ab} \cdot y_{cd} = y_{ac} \cdot y_{bd} = y_{ad} \cdot y_{bc}
\]

NUMBER OF NEW EQUATIONS: \( (m^4/4!) \cdot 2 \approx \frac{m^4}{12} \)

IN TERMS OF \( (\frac{1}{2} - \varepsilon)^m \) NEW PARAMETERS \( z_k \)

RELINEARIZE THE NEW SYSTEM VIA \( \mathbf{w}_{kl} = \mathbf{z}_k \cdot \mathbf{z}_l \)

CAN BE SOLVED UNIQUELY WHEN

\[
\frac{m^4}{12} \geq [(\frac{1}{2} - \varepsilon)^m]^2 / 2
\]

OR: \( (\frac{1}{2} - \varepsilon)^2 \leq \frac{1}{6} \Rightarrow \varepsilon \geq \frac{1}{2} - \frac{1}{\sqrt{6}} \approx 0.1 \)

SO WE NEED ONLY ONE FIFTH OF PREVIOUS # OF EQUATIONS
COMPLETING THE ATTACK

-Problem: There are \( n \) solutions for \( T \), and relinearization can only provide their linear span.

Solution: Add random equations to reduce the number of solutions to \( \approx 1 \).
- Do it over \( F \) rather than over \( K \).

-Problem: How to recover \( S \) from the fundamental equation \( G' = WPW^t \).

Solution: The equations become linear over \( F \).
- There are \( n^2 m^2 \) linear equations in \( m^2 \) new variables over \( F \).

-Problem: The attack is polynomial but infeasible.

Solution: Optimize!
any solution of this type is satisfactory.

A Appendix: A R Linearization Example

We demonstrate the complete relinearization technique on a toy example of 5 random quadratic equations in three variables \( x_1, x_2, x_3 \) modulo 7:

\[
\begin{align*}
3x_1x_1 + 5x_1x_2 + 5x_1x_3 + 2x_2x_2 + 6x_2x_3 + 4x_3x_3 &= 5 \\
6x_1x_1 + 1x_1x_2 + 4x_1x_3 + 4x_2x_2 + 5x_2x_3 + 1x_3x_3 &= 6 \\
5x_1x_1 + 2x_1x_2 + 6x_1x_3 + 2x_2x_2 + 3x_2x_3 + 2x_3x_3 &= 5 \\
2x_1x_1 + 0x_1x_2 + 1x_1x_3 + 6x_2x_2 + 5x_2x_3 + 5x_3x_3 &= 0 \\
4x_1x_1 + 6x_1x_2 + 2x_1x_3 + 5x_2x_2 + 1x_2x_3 + 4x_3x_3 &= 0
\end{align*}
\]

After replacing each \( x_ix_j \) by \( y_{ij} \), we solve the system of 5 equations in 6 variables to obtain a parametric solution in a single variable \( z \):

\[
y_{11} = 2 + 5z, \quad y_{12} = z, \quad y_{13} = 3 + 2z, \quad y_{22} = 6 + 4z, \quad y_{23} = 6 + z, \quad y_{33} = 5 + 3z
\]

This single parameter family contains 7 possible solutions, but only two of them also solve the original quadratic system. To filter out the parasitic solutions, we impose the additional constraints: \( y_{11}y_{23} = y_{12}y_{13}, \quad y_{12}y_{23} = y_{13}y_{22}, \quad y_{12}y_{33} = y_{13}y_{23} \). Substituting the parametric expression for each \( y_{ij} \), we get:

\[
(2+5z)(6+z) = z(3+2z), \quad z(6+z) = (3+2z)(6+4z), \quad z(5+3z) = (3+2z)(6+z)
\]

These equations can be simplified to:

\[
3z^2 + z + 5 = 0, \quad 0z^2 + 4z + 4 = 0, \quad 1z^2 + 4z + 3 = 0
\]

The relinearization step introduces two new variables \( z_1 = z \) and \( z_2 = z^2 \), and treats them as unrelated variables. We have three linear equations in
these two new variables, and their unique solution is $x_1 = 6, x_2 = 1$. Working backwards we find that $y_{13} = 4, y_{22} = 2, y_{33} = 2$, and by extracting their square roots modulo 7 we find that $x_1 = \pm 2, x_2 = \pm 3, x_3 = \pm 3$. Finally, we use the values $y_{12} = 6$ and $y_{23} = 5$ to combine these roots in just two possible ways to obtain $x_1 = 2, x_2 = 3, x_3 = 4$ and $x_1 = 5, x_2 = 4, x_3 = 3$, which solve the original quadratic system.

A more interesting example, which we do not describe in full in this extended abstract, consists of 5 randomly generated homogeneous quadratic equations in 4 variables. Note that this is barely larger than the minimum number of equations required to make the solution well defined. The number of linearized variables $y_{ij} = x_i x_j$ for $1 \leq i \leq j \leq 4$ is 10, and the solution of the system of 5 linear equations in these 10 variables can be defined by affine expressions in 5 new parameters $z_i$. There are 20 equations which can be derived from fundamentally different ways of parenthesizing products of 4 $x_i$ variables:

$$y_{12}y_{34} = y_{13}y_{24} = y_{14}y_{23}$$

$$y_{11}y_{23} = y_{12}y_{13}, \quad y_{11}y_{24} = y_{12}y_{14}, \quad y_{11}y_{34} = y_{13}y_{14}$$

$$y_{22}y_{13} = y_{12}y_{23}, \quad y_{22}y_{14} = y_{12}y_{24}, \quad y_{22}y_{34} = y_{23}y_{24}$$

$$y_{33}y_{12} = y_{13}y_{23}, \quad y_{33}y_{14} = y_{13}y_{24}, \quad y_{33}y_{24} = y_{23}y_{24}$$

$$y_{44}y_{12} = y_{14}y_{24}, \quad y_{44}y_{13} = y_{14}y_{23}, \quad y_{44}y_{23} = y_{24}y_{23}$$

$$y_{11}y_{22} = y_{12}y_{12}, \quad y_{11}y_{33} = y_{13}y_{13}, \quad y_{11}y_{44} = y_{14}y_{14}$$

$$y_{22}y_{33} = y_{23}y_{23}, \quad y_{22}y_{44} = y_{24}y_{24}, \quad y_{33}y_{44} = y_{34}y_{34}$$

When we substitute the affine expressions in the 5 new $z_i$ parameters and relinearize it, we get 20 linear equations in the 5 $z_i$ and their 15 products $z_i z_j$ for $1 \leq i \leq j \leq 5$, which is just big enough to make the solution unique (up to $\pm$ sign) with reasonable probability. □

References
