CRYPTANALYSIS OF THE OIL AND VINEGAR SIGNATURE SCHEME

AVIAD KIPNIS — NDS TECHNOLOGIES, ISRAEL

ADI SHAMIR — THE WEIZMANN INST', ISRAEL

UNIVARIATE CRYPTOGRAPHY

EXAMPLE: RSA/RABIN

$$x^2 = V \pmod{n}$$

- EVALUATION IS EASY
- SOLUTION IS DIFFICULT
- CAN BE USED FOR ENCRYPTION OR SIGNATURE SCHEMES
- RESISTED CRYPTANALYSIS FOR 20 YEARS

BIVARIATE CRYPTOGRAPHY

EXAMPLE: ONG-SCHNORR-SHAMIR (OSS)

$$\underset{\text{PUBLIC KEY}}{k} = \underset{\text{SECRET KEY}}{s^2} \pmod{\underset{\substack{\text{SECRET FACTORIZATION} \\ n=p\cdot q}}{n}}$$

MESSAGE: V SIGNATURE: x, y

VERIFICATION CONDITION:

$$x^2 - ky^2 = v \pmod{n}$$

SIGNATURE GENERATION:

$$x^2 - s^2y^2 = v \pmod{n}$$

$$\underset{\overset{\shortparallel}{r}}{(x+sy)}\underset{\overset{\shortparallel}{v/r}}{(x-sy)} = v \pmod{n}$$

$$\begin{matrix} x+sy=r \\ x-sy=v/r \end{matrix} \Rightarrow \begin{matrix} x=\frac{1}{2}(r+v/r) \\ y=\frac{1}{2s}(r-v/r) \end{matrix}$$

BROKEN BY POLLARD (TWO DIFFERENT ATTACKS ON QUADRATIC AND CUBIC VARIANTS).

MULTIVARIATE CRYPTOGRAPHY

- BASED ON SYSTEMS OF SEVERAL EQUATIONS IN SEVERAL VARIABLES
- MOST COMMON CASE: QUADRATIC EQ'S:

$$\sum_{i,j} c_{ij}^1 x_i x_j = v_1 \pmod{n}$$
$$\vdots$$
$$\sum_{i,j} c_{ij}^t x_i x_j = v_t \pmod{n}$$

TWO BASIC APPROACHES:

- FEW VARIABLES, LARGE DOMAIN
- MANY VARIABLES, SMALL DOMAIN

THE FIRST CASE: (FEW VARIABLES)

- SUSCEPTIBLE TO POLLARD-LIKE ATTACKS WHEN TOO MANY SOLUTIONS EXIST
- SUSCEPTIBLE TO GROEBNER-BASE ATTACKS WHEN TOO MANY EQUATIONS EXIST
- CAN BE PROVEN EQUIVALENT TO FACTORING WHEN ALGEBRAIC SYMMETRIES EXIST AND #SOLUTIONS IS BOUNDED
- IS NOT COMPUTATIONALLY ADVANTAGEOUS IN THIS CASE.

THE COMPLEXITY OF SOLVING SYSTEMS OF MANY ALGEBRAIC EQUATIONS IN MANY VARIABLES:

FRAENKEL AND YESHA:

IT IS NP-COMPLETE EVEN WHEN:

- ALL THE EQUATIONS ARE QUADRATIC.
- THE DOMAIN IS GF(2).

THE SECOND CASE: (MANY VARIABLES)

- MANY SCHEMES PROPOSED IN LAST 10 YEARS
- ALMOST ALL OF THEM WERE BROKEN
- PARTICULARLY SUSCEPTIBLE TO LOW RANK ATTACKS:

GIVEN THE QUADRATIC FORM:

$$\sum_{i,j} c_{ij} x_i x_j$$

WE CAN (USUALLY) CHANGE IT VIA A LINEAR CHANGE OF VARIABLES TO:

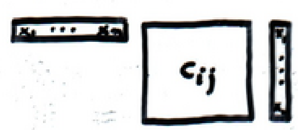
$$\sum_i d_i \bar{x}_i^2$$

AND THE RANK OF THE FORM IS THE NUMBER OF NON-ZERO d_i'S IN THIS REPRESENTATION.

PROPERTIES OF THE RANK:

- RANDOM QUADRATIC FORMS USUALLY HAVE HIGH RANK.
- IF $\sum_{ij} c_{ij} x_i x_j$ HAS LOW RANK, THEN IT IS IDENTICALLY ZERO ON A LARGE LINEAR SUBSPACE.

MATRIX REPRESENTATION OF QUADRATIC FORMS: $x^T A x$:



- c_{ij} IS TYPICALLY SYMMETRIC: $c_{ij} = c_{ji}$
- IN NORMAL FORM, A IS DIAGONAL.

EXAMPLE: AN UNPUBLISHED PK ENCRYPTION SCHEME:

[CRUCIALLY DEPENDS ON PROPERTIES OF $GF(2)$]

CONSIDER THE EQUATIONS:

$$y_1 y_2 = v_1 \pmod 2$$
$$y_3 y_4 = v_2 \pmod 2$$
$$\vdots$$
$$y_{4m-1} y_{4m} = v_{2m} \pmod 2$$

UNDER A LINEAR TRANSFORMATION

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{4m} \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \pmod 2$$

AND A LINEAR MIXING OF OUTPUTS

$$\begin{bmatrix} Q_1 \\ \vdots \\ Q_{2m} \end{bmatrix} = \begin{bmatrix} & B & \end{bmatrix} \begin{bmatrix} L_1 \cdot L_2 \\ L_3 \cdot L_4 \\ \vdots \\ L_{4m-1} L_{4m} \end{bmatrix}$$

THE PUBLIC KEY:

$2m$ QUADRATIC FORMS $Q_i(x_1 \cdots x_m)$ IN m VARIABLES x_j OVER $GF(2)$.

ENCRYPTION OF BINARY CLEARTEXT

$m = m_1 m_2 \cdots m_n$: EVALUATE THE $2m$ $Q_1 \cdots Q_{2m}$ UNDER $x_1 = m_1 \cdots x_n = m_n$, GIVING BINARY RESULTS $v_1, \cdots, v_{2m}$.

DECRYPTION OF BINARY CIPHERTEXT

$V = v_1 v_2 \cdots v_{2m}$:

- SEEMS TO BE DIFFICULT, SINCE THE Q_i'S LOOK RANDOM, BUT:
- APPLY B^{-1} TO CHANGE EQUATIONS TO:

$$L_1(\underline{x}) \cdot L_2(\underline{x}) = v_1'$$
$$L_3(\underline{x}) \cdot L_4(\underline{x}) = v_2'$$
$$\vdots$$
$$L_{4m-1}(\underline{x}) \cdot L_{4m}(\underline{x}) = v_{2m}'$$

ABOUT ONE QUARTER OF THESE ARE 1, GIVING TWO LINEAR EQUATIONS

$(0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0,\ 1 \cdot 1 = 1)$

- SOLVE THE RESULTANT SYSTEM OF $\sim m$ LINEAR EQUATIONS IN m VARIABLES

(UNPUBLISHED) LOW-RANK ATTACK DUE TO COPPERSMITH, STERN:

$$Q = z_1 Q_1(\underline{x}) + z_2 Q_2(\underline{x}) + \cdots + z_{2m} Q_{2m}(\underline{x}) = L_1(\underline{x}) \cdot L_2(\underline{x})$$

- IN Q, THE COEFFICIENT OF EACH TERM $x_i x_j$ IS A LINEAR FORM IN THE NEW $z_1 \cdots z_{2m}$.
- THE LINEAR SUBSPACE WHICH MAKES $L_1(\underline{x}) = 0$ CONTAINS HALF THE $\underline{x}$ SPACE, AND MAKES Q IDENTICALLY ZERO.
- LET $k \approx \log_2 2m$. CHOOSE k RANDOM VECTORS IN $\underline{x}$ SPACE. ALL OF THEM ARE ON THIS LINEAR SUBSPACE WITH PROB $(\frac{1}{2})^k \approx \frac{1}{2m}$.
  - FOR EACH LINEAR COMBINATION OF THESE GUESSED $\underline{x}$ VECTORS, THE CONDITION $Q = 0$ GIVES LINEAR EQUATION IN THE UNKNOWN $z_1 \cdots z_{2m}$.
  - THE GUESSING YIELDS $2m$ LINEAR EQUATIONS IN $2m$ z_i VARIABLES, WHICH WE CAN SOLVE.

THE OIL AND VINEGAR SCHEME

(PATARIN, 1996)

A SIMPLE WAY TO CONSTRUCT SOLVABLE, MEDIUM RANK SYSTEM OF QUADRATIC FORMS.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{2k} \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{2k} \end{bmatrix} \quad \text{OVER } GF(q)$$

$$\forall i = 1..m \; F_i = [y_1 \cdots y_{2k}] \begin{array}{c} \overbrace{\quad}^{m} \overbrace{\quad}^{m} \\ \left[\begin{array}{c|c} 0 & B_1 \\ \hline B_2 & B_3 \end{array}\right] \end{array} \begin{bmatrix} y_1 \\ \vdots \\ y_{2k} \end{bmatrix} \begin{array}{l} \}^{m} \\ \}^{m} \end{array} \quad \text{OVER } GF(q)$$

$$F_i(\underline{y}) = \underline{y}^t F_i \underline{y}, \quad G_i(\underline{x}) = \underline{x}^t (A^t F_i A) \underline{x}$$

PUBLISH THE COEFFICIENTS OF ALL THE $G_i(\underline{x})$ [THERE IS NO NEED TO MIX THEM LINEARLY] AS THE PUBLIC SIGNATURE VERIFICATION KEY.

THE MESSAGE: $M=(m_1,\cdots,m_k)$

THE SIGNATURE: $\underline{X}=(x_1,\cdots,x_{2k})$

VERIFICATION: $\forall i=1,\cdots,k:\ G_i(\underline{x})=m_i$

GENERATION: CAN BE DONE EFFICIENTLY IF THE SECRET KEY **A** IS KNOWN:

$$G_i(\underline{x})=m_i \Longleftrightarrow F_i(\underline{y})=m_i$$

$$F_i=\left[\begin{array}{c:c} 0 & * \\ \hdashline * & * \end{array}\right] \Rightarrow F_i=\sum_{i,j} c_{ij}\, y_i y_j$$

DOES NOT CONTAIN $y_i \cdot y_j$ BOTH FROM FIRST HALF $i,j=1,\cdots,k$.

DEFINITION: $y_1 \cdots y_k$ ARE OIL VARIABLES

$y_{k+1} \cdots y_{2k}$ ARE VINEGAR VAR'S.

F_i HAS ONLY OIL·VINEGAR, VINEGAR·OIL, VINEGAR·VINEGAR OCCURRENCES, SO CHOOSE ARBITRARY VINEGAR VALUES, AND SOLVE THE LINEAR SYSTEM IN OIL VAR'S.

MAIN PROBLEM: CAN YOU SEPARATE THE OIL AND VINEGAR VARIABLES IN THE QUADRATIC FORMS $G_1(\underline{x})\cdots G_{2n}(x)$?

THE OIL SUBSPACE:

IN $\underline{Y}$ SPACE: ALL VECTORS OF FORM

$$(\underbrace{*,*,\cdots,*}_{n},\underbrace{0,0,\cdots,0}_{n})$$

IN $\underline{X}$ SPACE: THE PREIMAGE BY A OF THE OIL $\underline{Y}$ SPACE.

THE VINEGAR SUBSPACE:

IN $\underline{Y}$ SPACE: ALL VECTORS OF FORM

$$(\underbrace{0,0,\cdots,0}_{n},\underbrace{*,*,\cdots,*}_{n})$$

IN $\underline{X}$ SPACE: THE PREIMAGE BY A OF THE VINEGAR $\underline{Y}$ SPACE.

EACH SPACE IS THE DIRECT SUM OF ITS OIL AND VINEGAR SUBSPACES.

CLAIM: ALL THE PUBLISHED QUADRATIC FORMS $G_1(\underline{x})\cdots G_{2n}(\underline{x})$ ARE IDENTICALLY ZERO ON THE OIL SUBSPACE OF $\underline{X}$.

PROOF: IN $F_i(\underline{Y})=\sum_{ij} c_{ij}\, y_i y_j$

EACH TERM HAS AT LEAST ONE INDEX IN SECOND HALF, WHICH IS 0 IN THE OIL SUBSPACE.

OBSERVATION: THE ZEROES OF EACH $G_i(\underline{x})$ ARE USUALLY A SUPERSET OF THE OIL SUBSPACE. THEIR INTERSECTION IS USUALLY EXACTLY THE OIL SUBSPACE.

BASIC IDEA: CONSIDER THE MATRIX OF COEFFICIENTS F_i OR G_i BOTH AS A QUADRATIC FORM AND AS A LINEAR MAPPING

$$[x_1\, x_2 \cdots x_{2k}] \begin{bmatrix} \boxed{\quad G_i \quad} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{2n} \end{bmatrix} = \text{VALUE}$$

OR

$$\begin{bmatrix} \boxed{\quad G_i \quad} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{2k} \end{bmatrix} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_{2k}' \end{bmatrix}$$

CLAIM: IF F_i IS REGULAR, THEN THE LINEAR MAPPING IT REPRESENTS MAPS THE OIL SPACE ONTO THE VINEGAR SPACE, AND F_i^{-1} MAPS THE VINEGAR SPACE ONTO THE OIL SPACE.

PROOF: $\begin{bmatrix} 0 & * \\ * & * \end{bmatrix}\begin{bmatrix} * \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ * \end{bmatrix}$

PROBLEM: A LINEAR CHANGE OF VARIABLES $\underline{y}=A\underline{x}$ CHANGES THE QUADRATIC FORM VIA $F \to A^tFA$:

$$\underline{x}^t F \underline{x} = (A\underline{x})^t F (A\underline{x}) = \underline{x}^t (A^t F A)\underline{x}$$

BUT CHANGES THE LINEAR MAPPING VIA $F \to A^{-1}FA$:

$$\underline{z} = F\underline{y} \to (A\underline{w}) = F(A\underline{x}) \to \underline{w} = (A^{-1}FA)\underline{x}$$

COROLLARY:

G DOES NOT MAP THE OIL $\underline{x}$ SPACE TO THE VINEGAR $\underline{x}$ SPACE (AS A LINEAR MAPPING)

BASIC IDEA: CONSIDER EXPRESSIONS OF THE FORM $G_i^{-1}G_j$ AND $F_i^{-1}F_j$

- F_j MAPS THE OIL SPACE TO THE VINEGAR SPACE, F_i^{-1} MAPS IT BACK, SO $F_i^{-1}F_j$ MAPS THE OIL SPACE TO ITSELF
- $G_j = A^tF_jA$, $G_i = A^tF_iA$, $G_i^{-1} = A^{-1}F_i^{-1}(A^t)^{-1}$

$$G_i^{-1}G_j = \left[A^{-1}F_i^{-1}(A^t)^{-1}\right]\left[A^tF_jA\right] = A^{-1}F_i^{-1}F_jA$$

COROLLARY: $G_i^{-1}G_j$ MAPS THE OIL $\underline{x}$ SPACE TO ITSELF AS A LINEAR MAPPING.

REMARK: REQUIRES EXTRA CARE OVER FIELDS OF CHARACTERISTIC 2.

DEFINITION: A LINEAR SUBSPACE U IS AN EIGENSPACE OF A MATRIX M IF $MU \subseteq U$. IT IS A COMMON EIGENSPACE OF $M_1 \cdots M_t$ IF $\forall i \; M_iU \subseteq U$.

COROLLARY: LET T BE THE CLOSURE OF ALL THE MATRICES OF THE FORM $G_i^{-1}G_j$ UNDER ADDITION, MULTIPLICATION, AND MULT' BY A CONSTANT. THEN THE OIL SUBSPACE IS A COMMON EIGENSPACE OF ALL THE MATRICES IN T.

EFFICIENT ALGORITHMS FOR FINDING COMMON EIGENSPACES

DEFINITION: LET $P(x)$ BE THE CHARACTERISTIC POLYNOMIAL OF $n \times n$ MATRIX B.

BY CALEY-HAMILTON THEOREM: $P(B)=0$.

LEMMA: FOR ANY POLYNOMIAL $P'(x)$, $\mathrm{kernel}(P'(B))$ IS AN EIGENSPACE OF B.

PROOF: $\underline{z} \in \mathrm{kernel}(P'(B)) \to P'(B)\underline{z} = \underline{0}$.

B COMMUTES WITH ITS POWERS, AND THUS WITH ANY POLYNOMIAL IN B. SO:

$$P'(B) \cdot B\underline{z} = B \cdot P'(B)\underline{z} = 0$$

SO $B\underline{z} \in \mathrm{kernel}(P'(B))$

THE CONVERSE IS NOT TRUE: $B=I$

THE ONLY SINGULAR POLYNOMIAL IN B IS THE 0 MATRIX WITH FULL SPACE AS KERNEL.

THEOREM: IF THE CHARACTERISTIC POLYNOMIAL OF B IS IRREDUCIBLE, THEN THE ONLY EIGENSPACES OF B ARE $\{0\}$ AND THE WHOLE SPACE.

PROOF: FOR EACH VECTOR AND LINEAR SUBSPACE, THERE IS A MINIMAL POLYNOMIAL IN B MAPPING IT TO $\underline{0}$. THIS POLYNOMIAL IS A DIVISOR OF THE CHARACTERISTIC POLYNOMIAL OF B.

LET $0 \neq \underline{z} \in$ EIGENSPACE V. $P(x)$ IS IRREDUCIBLE $\rightarrow$ MINPOLY OF $\underline{z}$ IS $P(x)$ ITSELF. THUS $\underline{z}, B\underline{z}, B^2\underline{z} \cdots B^{n-1}\underline{z}$ ARE n LINEARLY INDEPENDENT VECTORS, WHICH ARE ALL IN THE EIGENSPACE V. SO V IS FULL DIMENTIONAL, I.E., THE WHOLE SPACE.

OBSERVATION: $F_i^{-1} F_j = \left[\begin{array}{c|c} * & * \\ \hline 0 & * \end{array}\right]$

SO ITS CHARACTERISTIC POLYNOMIALS ARE ALWAYS THE PRODUCTS OF TWO HALF DEGREE POLYNOMIALS.

(THIS REMAINS UNCHANGED BY SIMILARITY TRANSFORMATION, WHEN $F_i^{-1} F_j$ IS MAPPED TO $G_i^{-1} G_j$).

NEXT SIMPLEST CASE:

THE CHARACTERISTIC POLYNOMIAL $P(x)$ OF B FACTORS INTO TWO IRREDUCIBLE POLY'S: $P(x) = P_1(x) \cdot P_2(x)$

LET $B_1 = P_1(B)$, $B_2 = P_2(B)$, $K_1 = \ker(B_1)$, $K_2 = \ker(B_2)$. THEN:

$K_1 \cap K_2 = \{\phi\}$, $\dim(K_1) + \dim(K_2) = n$

THE ONLY EIGENSPACES OF B ARE $\{0\}$, K_1, K_2, WHOLE SPACE.