Decomposition Attack for ECDLP

Igor Semaev

CTCrypt2016, June 7, 2016

Elliptic curves

- F_q finite field of size $q = p^n$, p is called characteristic
- Cubic $Y^2 = X^3 + aX + b$, characteristic $p \neq 2, 3$
- or $Y^2 + XY = X^3 + aX^2 + b$, characteristic $= 2$
- $E(F_q)$ is points $P = (x, y)$ satisfy the cubic and ∞
- For $P = (x, y)$, denote
  - $\bar{P} = (x, -y)$, characteristic $p \neq 2$
  - $\bar{P} = (x, y + x)$, characteristic $p = 2$

Group Operation

- To sum points $P_1, P_2 \in E$ do
- $P_1 \neq P_2$ draw line and get $P_3 \in E$. Then $P_1 + P_2 = \bar{P}_3$
- $P_1 = P_2$ draw tangent, get $P_3 \in E$, then $P_1 + P_2 = \bar{P}_3$
- $E(F_q)$ finite commutative group with ∞ neutral element

Discrete Logarithm Problem

- Given $P, Q \in E(F_q)$ find integer x such that
- $Q = xP$(in multiplicative notation $Q = P^x$)
- Hard problem except few easy cases
- Miller and Koblitz independently suggested first elliptic curve based protocols(like Diffie-Hellman etc)

Easy cases

- supersingular curves, like $Y^2 = X^3 - X$, over F_p and $p \equiv 3(mod 4)$, Menezes-Okamoto-Vanstone(1990), Semaev(1990)
- anomalous curves, where $|E(F_p)| = p$, Semaev(1995), Satoh-Araki(1997), Smart(1997)

NIST Curves

- over F_{2^n}, where $n = 163, 233, 283, 409, 571$:
- Koblitz curve $Y^2 + XY = X^3 + aX^2 + 1, a = 0, 1$
- "Pseudo-random" curve $Y^2 + XY = X^3 + X^2 + b$

- over F_p, where $\log_2 p \approx 192, 224, 521, 256, 384, 521$:
- "Pseudo-random" curve $Y^2 = X^3 - 3X + b$

Pollard's ρ

- Basic idea: iterate a pseudorandom function f on $E(F_q)$:
- $x_{i+1} = f(x_i)$, check for a collision $x_{2i} = x_i$
- Results in the problem solution
- Runs in time $q^{1/2}$, memory size negligible
- Fully parallelizable, Oorschot-Wiener(1996)
- Improved by Teske(1998), Gallant-Lambert-Vanstone(1999), etc
- Asymptotic bound remained

Logarithms in F_p

- Given $a, b \in F_p$, find integer x such that $b \equiv a^x \mod p$
- Index Calculus, Kraitchik(1924):
- B parameter, $2, 3, .., q < B$ small primes
- for random y compute $d \equiv a^y b \mod p$
- try to factor $d = 2^{l_2} 3^{l_3} .. q^{l_q}$, if yes
- $y + x \equiv l_2 \log_a 2 + l_3 \log_a 3 + \ldots + l_q \log_a q \mod p - 1$
- Collect $\approx B / \ln B$ such equations. Resolve and find x.

Logarithms in F_p. Asymptotic

- Take $B \approx exp(0.5\sqrt{\ln p \ln \ln p})$
- Algorithm runtime

$$\approx exp(1.5\sqrt{\ln p \ln \ln p})$$

- Far better than Pollard's $p^{1/2}$
- Was largely improved
- Best in Matyukhin(2003)
- Can that work for elliptic curves?

Decomposition

- In F_p $d \equiv a^y b \mod p$, where $b \equiv a^x$
- decompose $d = 2^{l_2} 3^{l_3} .. q^{l_q}$
- to the factor base $\{2, 3, \ldots, q\}$
- In elliptic curves, $R = yP + Q$, where $Q = xP$
- decompose $R = P_1 + P_2 + \ldots + P_m$
- $P_i \in V$ factor base?

Summation Polynomials

- Semaev(2004): $R = P_1 + P_2 + \ldots + P_m$ is equivalent to
- explicitly constructed polynomial equation(equation system)

$$S_{m+1}(R_x, x_1, \ldots, x_m) = 0$$

- $R = (R_x, R_y)$ and $P_i = (x_i, y_i)$
- Over F_{2^n}, naturally $\deg x_i \leq n/m$
- Over F_p, naturally $x_i \leq p^{n/m}$
- **Point Decomposition Problem:**
- R random point with known logarithm, find such x_i
- Asymptotically, to overcome Pollard's a fast resolution for $m = 5$ requires

Summation Polynomials

- Notation $P_i = (x_i, y_i)$. For some y_i
- $P_1 + P_2 + \ldots + P_m = \infty \Leftrightarrow S_m(x_1, \ldots, x_m) = 0$
- Characteristic 2 curve

$$Y^2 + XY = X^3 + aX^2 + 1, a = 0, 1$$

- $S_2(x_1, x_2) = x_1 + x_2$
- $S_3(x_1, x_2, x_3) = (x_1x_2 + x_2x_3 + x_1x_3)^2 + x_1x_2x_3 + 1$
- $S_4(x_1, x_2, x_3, x_4) = \mathrm{Res}_X(S_3(x_1, x_2, X), S_3(x_3, x_4, X))$
- $S_5(x_1, x_2, x_3, x_4, x_5) = \mathrm{Res}_X(S_4(x_1, x_2, x_3, X), S_3(x_4, x_5, X))$
- $\ldots$

Development

- Gaudry(2004), Diem(2004) in case $E(F_{q^n})$:
- Solve the system from

$$S_{m+1}(R_x, x_1, \ldots, x_m) = 0, \quad \deg x_i \leq n/m$$

  with a Gröbner basis algorithm
- Asymptotical improvements over Pollard's for $E(F_{q^n})$, where
- n fixed, q grows, final bound is exponential
- n, q grow simultaneously in some way, final bound is sub-exponential in q^n
- also Joux-Vitse(2012)

Gröbner basis algorithms

- Solve

$$f_1(x_1, \ldots, x_n) = 0,$$
$$f_2(x_1, \ldots, x_n) = 0,$$
$$\ldots$$
$$f_m(x_1, \ldots, x_n) = 0,$$

  for $x_i \in F_q$ by computing Gröbner basis for $< f_1, f_2, \ldots, f_m >$.
- Lazard(1983): construct matrices with rows
- coefficient vectors of gf_j, $\deg gf_j \leq d$
- compute Row Echelon Form
- construct back polynomials from rows: a Gröbner basis for some d
- An efficient variant F4, in Faugère(1999), in MAGMA

Complexity

- regularity degree d_{reg} maximum degree before the basis is computed
- complexity $\approx n^{d_{reg}\omega}, \quad 2 \leq \omega \leq 3$
- Bardet et al(2003): for semi-regular polynomials $f_1, \ldots, f_m$
- d_{reg} only depends on $n, m, \deg f_i$
- For quadratic($\deg f_i \leq 2$) semi-regular polynomials over F_2

m, n	d_{reg}	complexity
$m = \alpha n$	βn	exponential
$n = o(m)$	$o(n)$	sub-exponential
$m = \alpha n^2$	β	polynomial

- E.g. $m > \frac{n^2}{6}$ then $d_{reg} = 3$.
- Conjecture(Bardet et al): Probability that a random equation system is semi-regular tends to 1

First Fall Degree vs Regularity Degree

- Gröbner basis algorithm repeatedly computes $\sum_i g_i f_i$
- First fall degree d_{ff} smallest d such that

$$\max_i \{\deg g_i + \deg f_i\} = d$$

- but $\deg \sum_i g_i f_i < d$. First Fall Degree Assumption $d_{reg} \approx d_{ff}$
- **********
- Trivially $f_i f_i = f_i$ over F_2. So if $d_{reg} \approx d_{ff}$ than P=NP
- Nevertheless for some equation systems like HFE $d_{ff} = d_{reg}$
- for some others like AES $d_{ff} < d_{reg}$
- d_{ff} is relatively easy to compute, d_{reg} not

Development

- Faugère et al(2012) for

$$S_{m+1}(R_x, x_1, \ldots, x_m) = 0, x_i \in V :$$

- in char 2, experimentally d_{reg} is lower than expected
- Petit-Quisquater(2012): $d_{ff} \leq m^2 + 1$
- If $d_{reg} \approx d_{ff}$, then
- Resolution is sub-exponential
- Supported by experiments in Shantz-Teske(2013)
- $\Rightarrow$ Conjectured subexponential bound for binary ECDLP

Development

- Semaev(2015) rewrite relation

$$P_1 + P_2 + \cdots + P_{m+1} = \infty \Leftrightarrow S_{m+1}(x_1, \ldots, x_m, x_{m+1}) = 0$$

- as

$$\begin{array}{rl} P_1 + P_2 & = (u_1, *) \\ P_1 + P_2 + P_3 & = (u_2, *) \\ \cdots & \\ P_1 + \cdots + P_m + P_{m+1} & = \infty, \end{array} \quad \Leftrightarrow \quad \begin{array}{rl} S_3(x_1, x_2, u_1) & = 0 \\ S_3(u_1, x_3, u_2) & = 0 \\ \cdots & \\ S_3(u_{m-2}, x_m, x_{m+1}) & = 0, \end{array}$$

- In characteristic 2, the equations of degree ≤ 3
- $x_{m+1} \leftarrow R_x$, solve in $\deg x_i \leq n/m$ and $\deg u_j < n$
- #variables= #equations $= n(m-1)$
- with Gröbner basis algorithm

Experiments in Semaev(2015)

- Strikingly faster than in Petit -Quisquater and Shantz-Teske:
- $n = 21, m = 3$, solve

  $S_4(R_x, x_1, x_2, x_3) = 0$, $\deg x_i < 7$ (subspace of dim $= 7$) in $F_{2^{21}}$

- 21 six degree equations in 21 variables: 6910 sec and 27 GB, degree $d_{reg} = 7$
- equivalent system of 42 cubic equations

$$\begin{aligned} S_3(x_1, x_2, u) &= 0 \\ S_3(u, x_3, R_x) &= 0 \end{aligned}$$

  in 42 variables: 133 sec and 2.5 GB, degree $d_{reg} = 4$
- Similar results independently in Karabina(2015)

Regularity Degree

- In all experiments cubic equations and $d_{reg} \leq 4$
- Maximum number of variables was 60 and still $d_{reg} = 4$
- For semi-regular(generic case) cubic equations in 60 variables $d_{reg} = 15$
- Idea: even if d_{reg} grows with n but very slowly

Conjectured bound for binary ECDLP

- For m proportional to $\sqrt{n/\ln n}$
- Decomposition $\Leftrightarrow$ solving cubic equations
- in $n(m-1) = \sqrt{n^3/\ln n}$ variables
- If(assumption) $d_{reg} = o(\sqrt{n/\ln n})$
- then ECDLP over F_{2^n} is solved in

$$2^{c\sqrt{n \ln n}}$$

  for $c \approx 1.62...$
- Some FIPS curves are broken if $d_{reg} = 4$(extreme case)

Does d_{reg} grow?

- Solve $S_3(x_1, x_2, R_x) = 0, \deg x_i \leq n/2$ in $E(F_{2^n})$
- already for $n = 45$, $d_{reg} = 5$, not 4 (reported by Kosters)
- the computation took > 10 hours and $120GB$
- Try hybrid method?
- (studied by Shantz-Teske for $n \leq 40$)
- For $n = 45$, guess 5 bits in each x_i, 2^{10} guesses
- Run a Gröbner basis algorithm for each guess
- Overall < 2 min and 0.2 GB, degree $d_{reg} = 4$

Asymptotic Decomposition Complexity

- Hybrid method to evaluate for large n
- E.g. solve $S_3(x_1, x_2, R_x) = 0, \deg x_i \leq n/2$ in $E(F_{2^n})$
- $k = k(n)$ optimal number of guessed bits
- Choose k, run over $100 < 2^k$ random guesses, note runtime, extrapolate to all guesses
- Find local minimum in k
- Probably gives global minimum $k(n)$ and minimum runtime
- or an upper bound

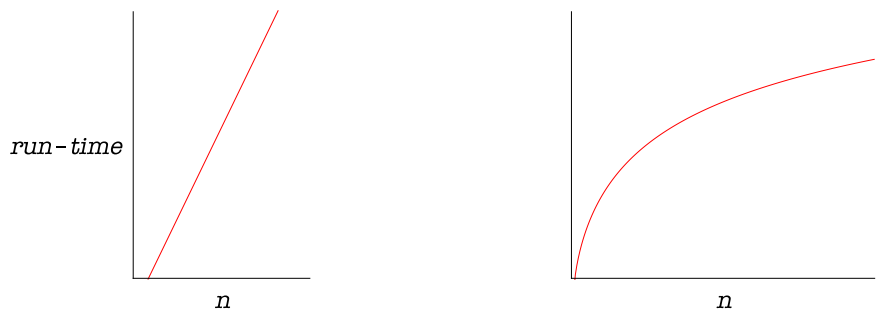
Complexity Evaluation

- Local minimum of runtime as a function of k

n	k	sec	d_{reg}
46	10	122	4
48	10	92	4
50	10	650	4
52	12	420	4
54	14	1510	4
56	14	2005	4
58	14	2595	4
60	14	5364	4
62	16	20345	4
64	16	14185	4
66	18	85747	4
68	18	65850	4
70	20	296694	4
72	20	467350	4

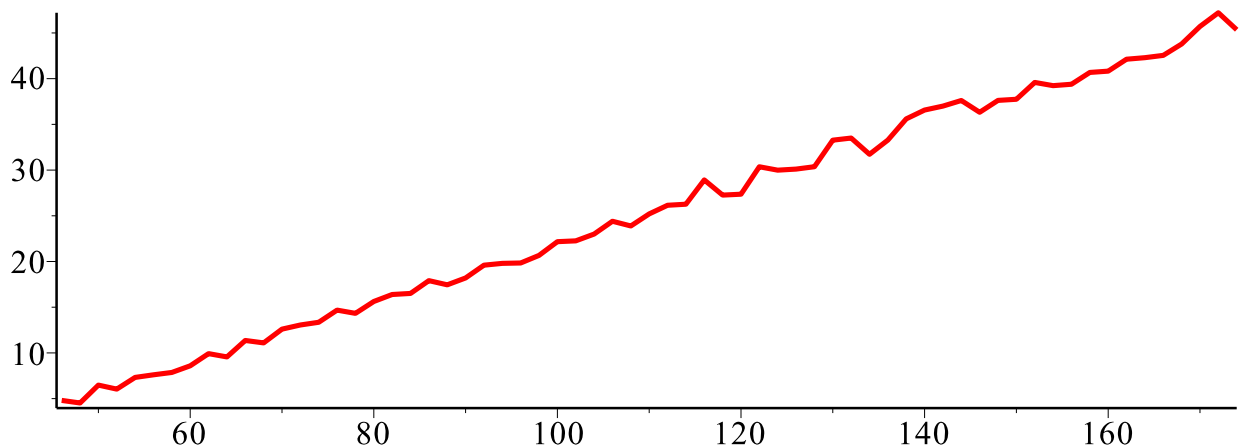
Decomposition Complexity

- Continue till $n = 174$, underground field size 2^{174}
- Draw the curve in log-scale
- Exponential or sub-exponential as a function in n?
- Typical curves



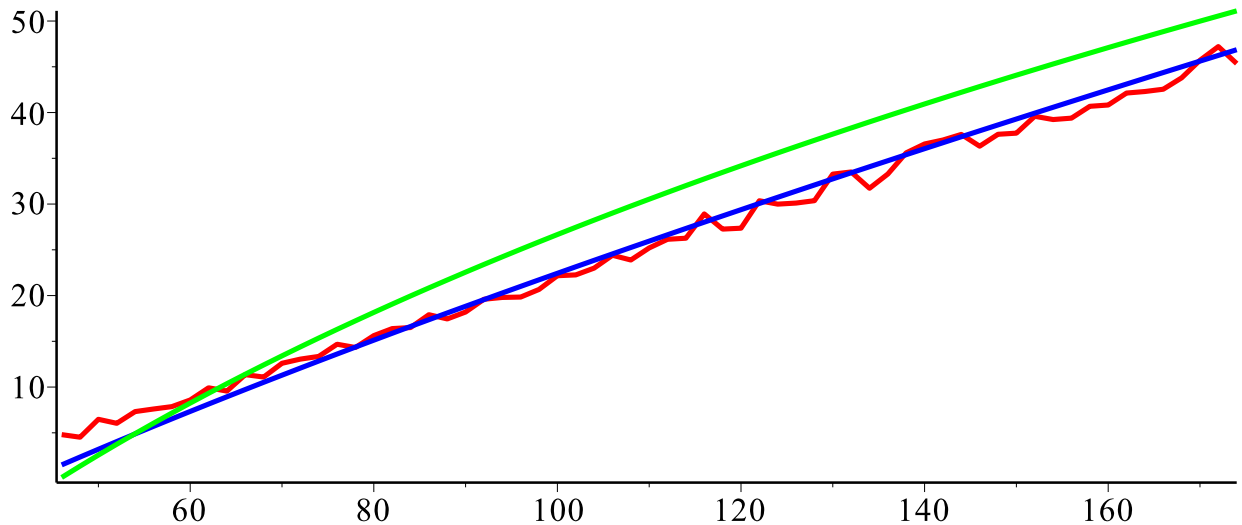
Decomposition Complexity

- Complexity curve(in red) in log-scale
- F4 takes a few sec for each guess of $\approx n/3 + o(n)$ bits
- $n = 174$ local min at $k = 62$,
- F4 took 11 sec, 2GB, degree $d_{reg} = 4$ for 174 eqns in 112 var,
- Time(sec) $< 2^{0.4n}$ for $n \leq 174$
- Overcomes brute-force $2^{0.5n}$?



Decomposition Complexity

- blue curve $1.5n^{3/4} - 25$
- green curve $\ln(n)^{2.75} - 40$, both $o(n)$
- sub-exponential complexity $2^{o(n)}$? Requires more data to decide
- Local minima not necessarily global minima
- So red curve might be more gradual



Decomposition Complexity

- Similarly, complexity of solving
  $S_{m+1}(x_1, \ldots, x_m, R_x) = 0, \quad \deg x_i \leq n/m$
- By splitting $S_{m+1} = 0$ into $m - 2$ equations $S_3 = 0$
- Run experiments with Hybrid method
- In progress

Courtois(2016) results

- Solve $S_3(x_1, x_2, R_x) = 0, \deg x_i \leq n/2$ in $E(F_{2^n})$
- Brute force: $2^{n/2}$ operations
- Event: x_1, x_2 have a known common factor f of degree $n/6$
- then x_1, x_2 easy to find
- Courtois: take random R_x(with known logarithm)
- wait the event to happen, search over f. Expected runtime $2^{n/3}$
- Also

$$S_4(x_1, x_2, x_4, R_x) = 0, \deg x_i \leq n/3 \quad \text{in} \quad E(F_{2^n})$$

in time $2^{n/3}$

Reduce d_{reg} further

- Get very over-defined equation system
- by introducing more new variables like in Kipnis-Shamir(1999) relinearization
- $S_3(x, y, R_x) = (xy + (x + y)R_x)^2 + xyR_x + 1$

$$= (U + VR_x)^2 + UR_x + 1 = 0 \quad (1)$$

- where $U = xy$ and $V = x + y$
- System of n linear equations in $\frac{3n}{2}$ bit variables of U, V
- Any f such that $f(U, V) = 0$ (10 quadratic polynomials)
- is added to linear equations (1)

Put the idea to extremes

- Introduce $N \approx \frac{n^2}{8}$ new variables

$$\begin{aligned} z_{ij} &= x_i y_j + x_j y_i, \\ z_{ii} &= x_i y_i, \\ zi &= x_i + y_i \end{aligned} \tag{2}$$

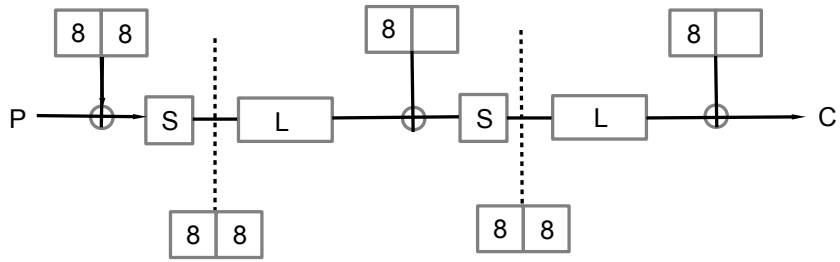
- Write $U = xy, V = x + y$ as linear combinations of (2)
- So $(U + VR_x)^2 + UR_x + 1 = 0$ linear equations in new variables
- Get $M \approx \frac{n^4}{16 \times 48}$ linearly-independent quadratic equations.
- Add quadratic equations to the linear
- Run a Gröbner basis algorithm or the hybrid method

Complexity

- Recall that for semi-regular equation system:
- $d_{reg} = 3$ for $M > \frac{N^2}{6}$ (just the case)
- $\Rightarrow$ Polynomial solution in time $N^{3\omega} = O(n^{6\omega})$, where $\omega \leq 3$
- By experiments, regularity degree really drops
- to $d_{reg} = 3$ for $n = 30$
- **Problem:** Is that true for larger n?
- *****
- Similar works for $S_{m+1}(x_1, \ldots, x_m, R_x) = 0$
- after splitting to $S_3 = 0$

AES vs PDP

- Scaled AES: find 16-bit key(48 effective variables, quadratic equations, 2100 sec with F4, degree $d_{reg} = 4$)



- Decompose $R = P_1 + P_2 + P_3 + P_4$ in $E(F_{2^{16}})$ by solving

$$\begin{aligned} S_3(x_1, x_2, u_1) &= 0, \deg x_i < 4, \deg u_i < 16 \\ S_3(x_3, u_1, u_2) &= 0 \\ S_3(x_4, u_2, R_x) &= 0 \end{aligned}$$

48 variables, cubic equations, 375 sec with F4, degree $d_{reg} = 4$

Resolution methods

- Gröbner basis algorithm or hybrid method
- SAT-solvers, tried by Galbraith-Gebregiyorgis(2014)
- MRHS by Raddum-Semaev(2008)

Conclusions

- Recent results and solving technics for decomposition attacks are surveyed
- Complexity evaluation for large parameters by hybrid method is introduced
- Construction of very over-defined equations for PDP by introducing even more new variables is described
- Solving AES vs PDP is discussed