Algorithms for the ECDLP

Steven Galbraith @elliptickiwi

University of Auckland, New Zealand

Outline

- ECDLP
- Baby-step-giant-step
- Summation polynomials and the ECDLP
- Symmetries
- Open problems

Joint work with many people, including Gebregiyorgis, Wang, Zhang, Gaudry.

Please interrupt me and ask questions.

AsiaCrypt 2015

November 29-December 3, 2015
Auckland, New Zealand
www.auckland.ac.nz/asiacrypt2015

Elliptic Curve Discrete Logarithm Problem

- **Elliptic curve discrete logarithm problem (ECDLP)**: Given $P, Q \in E(\mathbb{F}_q)$ to find an integer a, if it exists, such that $Q = aP$.

Elliptic Curve Discrete Logarithm Problem

- Let E be an elliptic curve over a finite field $\mathbb{F}_q$.
  $q = p^n$ and p is prime.
- Suppose $P \in E(\mathbb{F}_q)$ has prime order $r > 2\sqrt{q}$.
- **ECDLP**:
  Given $P, Q \in E(\mathbb{F}_q)$, both of order r, find a such that $Q = aP$.
- **Multiple-ECDLP**:
  Given $Q_1, \ldots, Q_L \in E(\mathbb{F}_q)$ to compute $a_1, \ldots, a_L$ such that $Q_i = a_i P$ for all $1 \le i \le L$.
- Well-known that Pollard rho solves ECDLP in $(1.25 + o(1))\sqrt{r}$ group operations.
- One can solve multiple-ECDLP in $O(\sqrt{rL})$ group operations.

Baby-step-giant-step algorithm

Textbook Baby-step-giant-step (BSGS)

- Let P have order r and $Q = aP$.
  Let $M = \lceil \sqrt{r} \rceil$. Then $a = a_0 + Ma_1$ with $0 \leq a_0, a_1 < M$.
- Compute sorted list of "baby steps" (aP, a) for $0 \leq a < M$.
  Let $P' = MP$. Compute "giant steps" $Q - bP'$ for $b = 0, 1, 2, \ldots$ until get a match.
- The worst-case running time is $2\sqrt{r}$ group operations, and average case is $1.5\sqrt{r}$ group operations.
- Pollard showed the average-case running time can be reduced to $(4/3)\sqrt{r}$ group operations by "interleaving" the baby-steps and giant-steps.
- What more needs to be said?

Best possible baby-step-giant-step?

- What is the best possible running time for a generic algorithm for the ECDLP?
- Suppose the algorithm computes a list of k group elements (depending on P, Q) such that any collision solves the ECDLP.
- There are $\binom{k}{2} \approx k^2/2$ possible pairs.
- So the ECDLP is solved with probability roughly $k^2/2r$.
- The expected running time is therefore roughly

$$\sum_{k=1}^{\sqrt{2r}} (1 - k^2/(2r)) \approx \int_0^{\sqrt{2r}} (1 - x^2/2r)dx = \tfrac{2\sqrt{2r}}{3} \approx 0.948\sqrt{r}.$$

- Can we achieve this?

Best possible baby-step-giant-step?

- Chateauneuf, Ling and Stinson considered this problem in a model where a computation $uP + vQ$ for $u, v \in \mathbb{Z}$ is counted as the basic operation.
- Bernstein and Lange compute three lists: one of baby-steps and two lists of giant-steps, walking in "opposite directions".
- A collision between any two lists solves the ECDLP.
- After k group operations we have three lists of size $k/3$ and success probability is approximately

$$\binom{3}{2}\frac{(k/3)^2}{r} = \frac{k^2}{3r}.$$

Since $k^2/4 < k^2/3 < k^2/2$ we expect the method to be better than the basic two-list algorithm but not to match the "ideal" lower bound.

Baby-step-giant-step

(Joint work with Ping Wang and Fangguo Zhang)

Algorithm	Average-case	Worst-case
Textbook BSGS	1.5	2.0
Textbook BSGS for av. case	1.414	2.121
Interleaving BSGS	1.333	2.0
Bernstein-Lange grumpy giants	1.2?	2.9?
Pollard rho	1.253	∞
BSGS with negation	1.0	1.5
Interleaving BSGS with negation	0.943	1.414
Grumpy giants with negation	0.9?	≤ 2.7?
Pollard rho using negation	0.886	∞
Block method	0.38	0.57
Grumpy giants with blocks	0.36?	≤ 1?
Pollard rho with Montgomery trick	0.47	∞

Summation polynomials and the ECDLP

Igor Semaev

Index calculus concept for ECDLP

- Let $P, Q \in E(\mathbb{F}_{q^n})$ be an ECDLP instance.
- Define a suitable factor base $\mathcal{F} \subseteq E(\mathbb{F}_{q^n})$.
- Generate random points $R = aP + bQ$ and try to write

$$R = P_1 + P_2 + \cdots + P_m$$

  where $P_1, \dots, P_m \in \mathcal{F}$.
- Each successful **point decomposition** is called a **relation**.
- When enough relations have been computed one can solve the ECDLP using sparse linear algebra.
- Left kernel version: Relations depend on ECDLP instance.
- Right kernel version: Further DLP step required.

Point decomposition

- We wish to solve

$$R = P_1 + P_2 + \cdots + P_m \qquad (*)$$

where $P_1, \ldots, P_m \in \mathcal{F}$.

- The right hand side is a rational function in the variables $x_i, y_i \in \mathbb{F}_{q^n}$ such that $P_i = (x_i, y_i) \in E(\mathbb{F}_{q^n})$.
- Hence, solving the equation (*) reduces to solving a system of polynomial equations in $2m$ variables in $\mathbb{F}_{q^n}$.
- It is natural to choose $\mathcal{F}$ to reduce the number of variables.
- Gaudry and Diem used Weil restriction to provide a natural definition for $\mathcal{F}$ that reduces the number of variables while increasing the number of equations.

Semaev's summation polynomials

- Semaev defines, for fixed elliptic curve E, polynomials $\mathrm{Sum}_{m+1}(x_1, \ldots, x_{m+1})$ such that: If points $R, P_1, \ldots, P_m \in E(\mathbb{F}_{q^n})$ satisfy

$$R = P_1 + P_2 + \cdots + P_m \qquad (*)$$

  then $\mathrm{Sum}_{m+1}(x(P_1), x(P_2), \ldots, x(P_m), x(R)) = 0$.
- Converse true up to choice of signs.
- Semaev explains how to compute these polynomials and proves they are symmetric and have degree 2^{m-1} in each variable.

Factor base

- Let $V \subseteq \mathbb{F}_{q^n}$ be an $\mathbb{F}_q$-vector space of dimension ℓ.
- Define $\mathcal{F} = \{P \in E(\mathbb{F}_{q^n}) : x(P) \in V\}$.
  Then $\#\mathcal{F} \approx q^\ell$.
- We expect approximately $\#\mathcal{F}^m/m! \approx q^{\ell m}/m!$ points of the form $P_1 + \cdots + P_m$ for $P_i \in \mathcal{F}$.
- Hence, a relation (*) exists with probability approximately $q^{\ell m}/(m! q^n)$.
- Computing a relation using Semaev's polynomials and Weil restriction with respect to $\mathbb{F}_{q^n}/\mathbb{F}_q$ requires solving a system with ℓm variables and n equations.

Point decomposition revisited

- The rational function

$$R = P_1 + P_2 + \cdots + P_m \qquad (*)$$

  where $P_i = (x_i, y_i) \in E(\mathbb{F}_{q^n})$ has $2m$ variables and the degree is determined by the elliptic curve group law and the degree of the defining equations $y_i^2 = f(x_i)$ of the elliptic curve.
- Semaev's approach is to minimize number of variables at the expense of exponential degree.
- Other choices of coordinates can lead to lower degree but more variables.
- **Problem:** Determine the optimal tradeoff of number of variables versus degree for point decomposition algorithms?
- The "splitting trick" can be viewed as a different tradeoff.

Using symmetries

- Since $\mathrm{Sum}_{m+1}(x_1, \ldots, x_m, x(R))$ is invariant under action by symmetric group S_m, one can write it in terms of elementary symmetric polynomials σ_j.
- This leads to a system of equations of lower degree.
- Faugère, Gaudry, Huot and Renault (J. Crypto., 2014) solved system using Gröbner basis with respect to grevlex order (F4 or F5 algorithm) and then FGLM (Faugère, Gianni, Lazard and Mora) change of ordering algorithm.
- Three benefits:
  - Degree of polynomials is lower, so F4/F5 is more efficient;
  - Fewer solutions so FGLM algorithm faster;
  - Can reduce the factor base and speed-up the linear algebra.

Using symmetries

- Faugère, Gaudry, Huot and Renault also use invariants under a larger group, coming from action of symmetric group and points of small order.
- This gives improvement to both point decomposition and linear algebra.
  (But don't forget extra step of translating solutions back to original coordinates.)
- Faugère, Huot, Joux, Renault and Vitse (EUROCRYPT 2014) discuss computing summation polynomials using invariant variables.
  They compute the 8-th summation polynomial.

Larger values for n

- Most work done for $E(\mathbb{F}_{q^n})$ where q is medium/large and n is fairly small.
- Then $\mathcal{F}$ is defined using the $\mathbb{F}_q$-vector space $V = \mathbb{F}_q \subseteq \mathbb{F}_{q^n}$, and Weil restriction is defined with respect to $\mathbb{F}_{q^n}/\mathbb{F}_q$.
  Important fact: $x_1, x_2 \in \mathbb{F}_q$ implies $x_1 x_2 \in \mathbb{F}_q$.
  So if $x_1, \dots, x_m \in V$ then $\sigma_j(x_1, \dots, x_m) \in V$.

Larger values for n

- Most work done for $E(\mathbb{F}_{q^n})$ where q is medium/large and n is fairly small.
- Then $\mathcal{F}$ is defined using the $\mathbb{F}_q$-vector space $V = \mathbb{F}_q \subseteq \mathbb{F}_{q^n}$, and Weil restriction is defined with respect to $\mathbb{F}_{q^n}/\mathbb{F}_q$.
  Important fact: $x_1, x_2 \in \mathbb{F}_q$ implies $x_1 x_2 \in \mathbb{F}_q$.
  So if $x_1, \dots, x_m \in V$ then $\sigma_j(x_1, \dots, x_m) \in V$.
- We consider the case $E(\mathbb{F}_{2^n})$ where n is prime.
- Huang, Petit, Shinohara and Takagi (IWSEC 2013) study large extension degrees.
- Define factor base using $\mathbb{F}_2$-vector space $V \subset \mathbb{F}_{2^n}$.
  Important fact: $x_1, x_2 \in V \subset \mathbb{F}_{2^n}$ does not imply $x_1 x_2 \in V$.

Huang, Petit, Shinohara and Takagi

- Fix polynomial basis $\{1, \theta, \theta^2, \ldots, \theta^{n-1}\}$ for $\mathbb{F}_{2^n}/\mathbb{F}_2$.
- Choose V to have basis $\{1, \theta, \ldots, \theta^{\ell-1}\}$.
- Then if x_1, x_2 lie in V then $x_1 x_2$ lies in space $V^{(2)}$ with basis $\{1, \theta, \ldots, \theta^{2(\ell-1)}\}$.
- Hence have $\sigma_1 \in V$, $\sigma_2 \in V^{(2)}$, $\sigma_3 \in V^{(3)}$ and so on.
- From some point onwards we have $V^{(j)} = \mathbb{F}_{2^n}$.
- Biggest example in their paper: $n = 53$, $m = 3$, $\ell = 6$, computation takes around 30 seconds.

Joint work with Shishay Gebregiyorgis

"Summation Polynomial Algorithms for Elliptic Curves in Characteristic Two", INDOCRYPT 2014.

- New choice of invariant variables for binary Edwards models.
- Factor base that "breaks symmetry" and hence significantly increases the probability that relations exist.
- Experiment with SAT solvers rather than Gröbner basis methods for solving polynomial systems over $\mathbb{F}_2$.

Binary Edwards model

- Let $d_1, d_2 \in \mathbb{F}_{2^n}$, $d_1 \neq 0$ and $d_2 \neq d_1^2 + d_1$.
  Binary Edwards model of an elliptic curve is

$$E \ : \ d_1(x+y) + d_2(x^2+y^2) = xy + xy(x+y) + x^2y^2.$$

- Identity is $(0,0)$ and $-(x,y) = (y,x)$.
- $T_2 = (1,1)$ has order 2 and if $P = (x,y) \in E$, then $P + T_2 = (x+1, y+1)$.
- Hence $t(P) = x(P) + y(P)$ is invariant under $P \mapsto -P$ and $P \mapsto P + T_2$.
- We define the factor base using the (degree 4) function t:

$$\mathcal{F} \ = \ \{P \in E(\mathbb{F}_{2^n}) : t(P) \in V\}.$$

Binary Edwards model

- When $d_1 = d_2$ then $T_4 = (1, 0)$ is a point on E, and it has order 4.
- $t(P + T_4) = t(P) + 1$.
- If $R = P_1 + \cdots + P_m$ then

  $$R = (P_1 + u_1 T_4) + (P_2 + u_2 T_4) + \cdots + (P_m + u_m T_4)$$

  where $u_1 + u_2 + \cdots + u_m \equiv 0 \pmod 4$.
- Hence the summation polynomial is fixed under the action of $(\mathbb{Z}/4\mathbb{Z})^{m-1} \rtimes S_m$ on the polynomial ring.
- Note that $t(P)(t(P) + 1)$ is invariant under addition by T_4.
- Hence the summation polynomial can be expressed in terms of coordinates that are elementary symmetric polynomials in $t_i(t_i + 1)$.

Binary Edwards model

- By the same ideas as already discussed, can reduce point decomposition to solving a system of polynomial equations over $\mathbb{F}_2$.
- Our choice of coordinates provides a small improvement over previous work, but we are still limited to very small examples, such as $m = 3$ or $m = 4$.
- SAT solvers are an interesting alternative to Gröbner basis methods and can be faster in some situations.
- None of our computations are even close to beating Pollard rho for $E(\mathbb{F}_{2^n})$ where n is prime.

Larger group actions?

- Can we use larger groups than $(\mathbb{Z}/4\mathbb{Z})^{m-1} \rtimes S_m$?
- To do this we'd need either a totally new idea, or to have addition by a torsion point which is "linear" with respect to the coordinate system.
- It seems unlikely that one can work with larger groups.
- See Kohel's paper at INDOCRYPT 2012.
- Related question: Can one exploit large symmetry groups when using the splitting/unrolling trick?

Breaking symmetry

- As noted, the symmetric group S_m acts on $R = P_1 + \cdots + P_m$ and hence acts on $\text{Sum}_{m+1}(x(R), x_1, \ldots, x_m)$.
- Good news: We can write this polynomial in terms of elementary symmetric variables and this lowers the degree.
- Bad news: Probability of a relation has $1/m!$ term.
- Counterintuitive: we can evaluate the symmetric variables at combinations of non-symmetric variables.
- So get benefit of lower degree polynomial equations without the additional $m!$ factor in the running time.

Breaking symmetry

- Precisely: Let $V \subseteq \mathbb{F}_{2^n}$ be a vector space of dimension ℓ.
- Instead of one set $\mathcal{F} = \{P \in E(\mathbb{F}_{2^n}) : x(P) \in V\}$ we define m sets $\mathcal{F}_i = \{P \in E(\mathbb{F}_{2^n}) : x(P) \in V + v_i\}$ where $v_i \in \mathbb{F}_{2^n}$ are elements of a certain form so that the sets $V + v_i$ are all distinct.
- Suppose V has basis $\{1, \theta, \dots, \theta^{\ell-1}\}$.
- Let $v_1 = 0$, $v_1 = \theta^{\ell}$, $v_2 = \theta^{\ell+1}$, $v_3 = \theta^{\ell} + \theta^{\ell+1}$ etc.
- Then $V + v_i$ are distinct and yet only need a couple more variables to represent the combination.
- Hence, we have $\sigma_1 \in V'$ spanned by $\{1, \theta, \dots, \theta^{\ell+1}\}$, $\sigma_2 \in V''$ spanned by $\{1, \theta, \dots, \theta^{2(\ell+1)}\}$ etc.
- Care needed to pull solutions in the σ_j back to solutions in x_j.
- Diem and Matsuo have also used different factor bases $\mathcal{F}_i$.

Using Frobenius?

- Is there any way to speed up index calculus algorithms based on summation polynomials by using Galois action e.g., when $E/\mathbb{F}_2$ but ECDLP is in $E(\mathbb{F}_{2^n})$?
- Let $\pi : E \to E$ be the 2-power Frobenius map.
- If $R = P_1 + \cdots + P_m$ then $\pi(R) = \pi(P_1) + \cdots + \pi(P_m)$.
- Alternatively,

$$\begin{aligned} & \mathrm{Sum}_{m+1}(x(R), x_1, \ldots, x_m) = 0 \\ \Leftrightarrow \quad & \mathrm{Sum}_{m+1}(\pi(x(R)), \pi(x_1), \ldots, \pi(x_m)) = 0. \end{aligned}$$

- What is this good for?
- Gorla and Massierer represent ECDLP instance using the trace zero variety and perform index calculus using summation polynomials directly on the trace zero variety.

Ideas from finite field DLP

- Recently there was tremendous progress on the DLP in finite fields of small characteristic.
- Key equation:

$$x^q - x = \prod_{\alpha \in \mathbb{F}_q} (x - \alpha) \qquad (1)$$

  in $\mathbb{F}_q[x]$.
- Two more key ideas:
  - The automorphisms of the projective line are given by Möbius transformations.
    So substitute $(ax+b)/(cx+d)$ for $a, b, c, d \in \mathbb{F}_{p^d}$ into this equation.
  - Can represent Frobenius x^q as a low-degree rational function $u(x)/v(x)$.
- Can we find an elliptic curve version of this?

Ideas from finite field DLP

- One natural approach is to use the double cover $E \to \mathbb{P}^1$ given by $(x, y) \mapsto x$.
- Key equation can be interpreted as

$$\operatorname{div}(x^q - x) \;=\; \sum_{P \in E(\overline{\mathbb{F}}_q), x(P) \in \mathbb{F}_q} (P) \;-\; 2q(\infty).$$

- But this is un-interesting from a group-theoretic viewpoint, as the right hand side just represents the fact that $x(P) = x(-P)$ and $P + (-P) = \infty$.
- Also:
  - $\operatorname{Aut}(E)$ very small compared with $\operatorname{Aut}(\mathbb{P}^1)$.
  - No idea how to do the individual DLP descent stage.
- Other ideas?

Other Ideas

- Palash Sarkar and Shashank Singh, "A Simple Method for Obtaining Relations Among Factor Basis Elements for Special Hyperelliptic Curves", eprint 2015/179.
- Obtains some relations easily by intersecting lines with the curve equation.
- Section 4.2 of the paper gives an example for ECDLP over $\mathbb{F}_{p^7}$ where the method finds triples P_1, P_2, P_3 of points in the factor base ($x(P) \in \mathbb{F}_p$) such that $P_1 + P_2 + P_3 = \infty$. Note that $y(P_1) = y(P_2) = y(P_3)$.
- The paper claims only about $p/12$ such relations can be found.

Conclusion

- There are still open questions about the baby-step-giant-step algorithm.
- Symmetries are useful when computing with summation polynomials, but further progress seems limited.
- SAT solvers are an interesting alternative to Gröbner basis methods and can be faster in some situations.
- ECDLP in $E(\mathbb{F}_{2^n})$ for prime $n > 160$ seems to be completely immune to point decomposition attacks, even just trying to get a cube-root algorithm (case $m = 4$).

Thank You

See you at Asiacrypt!

2015 AsiaCrypt