(pseudo-)Random walks in cryptography

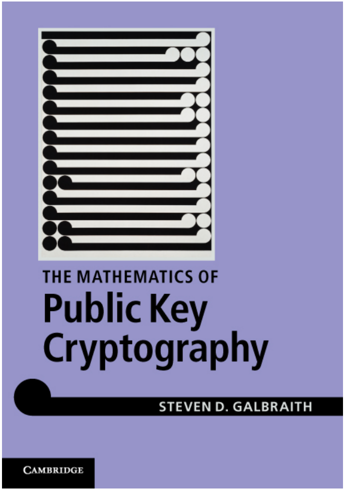
Steven Galbraith
Auckland University, NZ.

ECC 2012, Queretaro, Mexico

Joint work with: John Pollard, Raminder Ruprai, Mark Holmes, Wei Liu, Anton Stolbunov, Chang-An Zhao, Christina Delfs, . . .

Thanks to: Tanja Lange , Pierrick Gaudry and David Kohel.

Advertisement



Almost everything you need to know about ECC.

Goals of this talk

- Survey Pollard's ideas of using pseudo-random walks to solve the DLP.
- Present some recent results on the discrete logarithm problem in an interval and other variants of the DLP.
- Present a generalisation of the birthday paradox.
- Discuss some computational problems in isogeny graphs.

Please interrupt me and ask questions.

Discrete logarithm problem

- Let G be a group of prime order r, e.g., an elliptic curve over a finite field.
  I will write all groups multiplicatively in this talk.
- Let $g \in G$.
  The discrete logarithm problem (DLP) is: Given $h \in G$ to find a such that $h = g^a$.
- Exhaustive search: $O(r)$ group operations.
- Baby-step-giant-step: $O(\sqrt{r})$ group operations and storage of $O(\sqrt{r})$ group elements.
- Pollard rho (1970s): expected $O(\sqrt{r})$ group operations.
  Using distinguished points, following van Oorschot and Wiener (1990s), $(1 + o(1))\sqrt{\pi r/2} \approx 1.25\sqrt{r}$ group operations.

Pollard row

Random walk algorithms

- DLP: Given g, h, find a such that $h = g^a$.
- Idea: Suppose we can find integers a_1, a_2, b_1, b_2 such that

  $$g^{a_1} h^{b_1} = g^{a_2} h^{b_2}.$$

  We hope to solve the DLP as $a = (a_2 - a_1)(b_1 - b_2)^{-1} \pmod{r}$.
- The idea is to generate pseudorandom sequences of values

  $$x_i = g^{a_i} h^{b_i}$$

  such that (a_i, b_i) are also known.
- A **collision** is when $x_i = x_j$.
- **Pollard's big idea**: use pseudorandom walks where the next step only depends on the current position.
  Hence, if $x_i = x_j$ then $x_{i+1} = x_{j+1}$.
- It follows that one can detect collisions without storing all points, by only storing "distinguished" points.

Two basic types of walk

- DLP: Given g, h, find a such that $h = g^a$.
- **First type**: Elements of walks are

$$x_i = g^{a_i} h^{b_i}$$

  where a_i and b_i are "pseudorandom".
  Any collision $x_i = x_j$ is potentially useful. Such algorithms are analysed using the birthday paradox.
- **Second type**: Walks are either **tame** $x_i = g^{a_i}$ or **wild** $y_j = hg^{b_j}$.
- A collision $x_i = y_j$ allows to solve the DLP as $h = g^{a_i - b_j}$.
  Collisions $x_i = x_j$ or $y_i = y_j$ are useless.
  Such algorithms are analysed using several notions in probability theory.

Birthday paradox

- Suppose we sample uniformly at random from a set of size N. The expected number of trials until an element is sampled twice is $\sqrt{\pi N/2}$.
- When $N = 365$ this expected number is ≈ 23.94.
- Now sample uniformly at random from a set of size N and record each element in one of two lists.
  The expected number of trials until an element appears in both lists is $\sqrt{\pi N}$.
- The expected number of people in a room before there is a boy and a girl with the same birthday is ≈ 33.86.

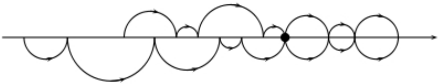
Birthday paradox

- Suppose we sample uniformly at random from a set of size N. The expected number of trials until an element is sampled twice is $\sqrt{\pi N/2}$.
- When $N = 365$ this expected number is ≈ 23.94.
- Now sample uniformly at random from a set of size N and record each element in one of two lists.
  The expected number of trials until an element appears in both lists is $\sqrt{\pi N}$.
- The expected number of people in a room before there is a boy and a girl with the same birthday is ≈ 33.86.
- Puzzle: In my hotel there is a meeting of the "boys born in January" club, and a meeting of the "random girls" club. How many of each should I put in a room until I expect a boy and girl with the same birthday?

The DLP in an Interval

- Given g, h, N find a, if it exists, such that $h = g^a$ and $0 \leq a < N$.
- This problem arises in practice: pseudorandom generator by Gennaro, decryption in the Boneh-Goh-Nissim scheme, analysis of the static/strong Diffie-Hellman problem, etc.
- Pollard kangaroo method using distinguished points (van Oorschot and Wiener 1996/1999) solves problem in average case expected $(2 + o(1))\sqrt{N}$ group operations.
- Important: kangaroo method is not analysed using the birthday paradox.
  Instead, steps in the walk are "short", meaning $x_i = g^{a_i}$ and $x_{i+1} = g^{a_{i+1}}$ is such that $a_{i+1} \approx a_i + m$.

Pollard kangaroo algorithm



- The tame kangaroo starts in the middle of the interval.
- Steps in the walk are, on average, distance m.
- The rear kangaroo "catches up" with the starting point of the front kangaroo in average time $N/(4m)$.
- There is now approximately one footprint by the front kangaroo in every interval of length m, so we expect the rear kangaroo to land on a footstep of the front kangaroo after m steps.
- Running time $2(N/(4m) + m + 1/\theta)$.
- Taking $m = \sqrt{N}/2$ and $1/\theta = o(\sqrt{N})$ gives running time $(2 + o(1))\sqrt{N}$ group operations.

Improved DLP in an Interval (G.-Pollard-Ruprai, to appear)

Two ways to improve:

- Three (actually, four) kangaroos method in $\approx 1.71\sqrt{N}$ group operations.
  Idea is to start wild kangaroos at both h and h^{-1}.
  Walks are now of three types: $x_i = g^{a_i}, \; y_j = hg^{b_j}$ or $z_k = h^{-1}g^{c_k}$.
  A collision between walks of any two types solves the DLP.
  (Assume group order odd.)

Improved DLP in an Interval (G.-Pollard-Ruprai, to appear)

Two ways to improve:

- Three (actually, four) kangaroos method in $\approx 1.71\sqrt{N}$ group operations.
  Idea is to start wild kangaroos at both h and h^{-1}.
  Walks are now of three types: $x_i = g^{a_i}, \; y_j = hg^{b_j}$ or $z_k = h^{-1}g^{c_k}$.
  A collision between walks of any two types solves the DLP. (Assume group order odd.)
- Gaudry-Schost algorithm (cockroaches) in $\approx 1.66\sqrt{N}$ group operations.
- Paper available on my webpage.

Gaudry-Schost Algorithm

- A way to tackle constrained problems using a variant of the birthday paradox.
- One has a "tame set" T and a "wild set" W and seeks collisions in $T \cap W$.
- The random walks are "cockroaches": staying in an appropriate-sized neighbourhood of the starting point.
- Basic idea for DLP in an interval:

  $$T = \{g^x : 0 \leq x < N\}, \quad \text{and} \quad W = \{hg^x : -N/2 < x < N/2\}.$$

  Then $N/2 \leq \#(T \cap W) \leq N$.
- We model the cockroaches as pseudorandom sampling from $T \cap W$ and apply a variant of the birthday paradox. This gives an average case expected running time of $(2.08 + o(1))\sqrt{N}$ group operations – worse than Pollard kangaroo.
- There are some inconvenient aspects: e.g., boundaries of T and W.

Gaudry-Schost Algorithm

- There is great flexibility in the choice of sets T and W, and how one samples from them.
- Combining the ideas of using smaller sets with the “four kangaroos” idea gives an algorithm for the DLP in an interval with average case expected time of around $1.66\sqrt{N}$ group operations.
- In principle, this beats the $1.71\sqrt{N}$ coming from the four kangaroo algorithm.

Equivalence Classes

- Many groups have efficiently computable automorphisms ψ. For example, the map $\psi : g \mapsto g^{-1}$ is easy for elliptic curves and the torus T_2.
- Gallant-Lambert-Vanstone/Wiener-Zuccherato solve the DLP by defining a random walk on G/ψ (sets of orbits in the group G under ψ).
- For Pollard rho, using equivalence classes with respect to inversion "should" speed-up the algorithm by a factor of $\sqrt{2}$.

Equivalence Classes

- Many groups have efficiently computable automorphisms ψ. For example, the map $\psi : g \mapsto g^{-1}$ is easy for elliptic curves and the torus T_2.
- Gallant-Lambert-Vanstone/Wiener-Zuccherato solve the DLP by defining a random walk on G/ψ (sets of orbits in the group G under ψ).
- For Pollard rho, using equivalence classes with respect to inversion "should" speed-up the algorithm by a factor of $\sqrt{2}$.
- A catch: lots of nasty little cycles in the pseudorandom walks. Handling these adds some overhead.
  See Bernstein, Lange, Schwabe (PKC 2011).
- All recent records for the ECDLP have **not** used equivalence classes (except for the case of Koblitz curves).

DLP in an Interval Using Equivalence Classes

- It seems impossible to combine Pollard's kangaroo algorithm with equivalence classes.
- Consider DLP in an interval: g, h, N to find a, if it exists, such that $h = g^a$ and $-N/2 \leq a \leq N/2$.
- Use Gaudry-Schost method: Define the tame set, of equivalence classes,

$$T = \{\{g^x, g^{-x}\} : 0 \leq x \leq N/2\}$$

and

$$W = \{\{hg^x, (hg^x)^{-1}\} = \{g^{a+x}, g^{-(a+x)}\} : -N/2 \leq x \leq N/2\}.$$

- We have $\#T \approx N/2$, but $\#W$ and $\#(T \cap W)$ depend on a.
- Further subtlety: if one chooses a unique representative for each class, one finds that sampling elements hg^x where x is chosen uniformly in $-N/2 \leq x \leq N/2$ does not necessarily correspond to sampling uniformly in W.

A Generalisation of the Birthday Paradox

(Building on work of Selivanov.)
Theorem (G.-Holmes) Let $C \in \mathbb{N}$. Balls will be chosen of colour $1 \le c \le C$ with probability q_c.
Let $N \in \mathbb{N}$ (we consider this as variable). A ball of colour $1 \le c \le C$ is assigned to urn $1 \le a \le N$ with probability $\mathrm{q}_{c,a}$.
Suppose that some technical conditions hold.
Let

$$A_N = \sum_{c=1}^{C} \mathrm{q}_c \left(\sum_{c'=1, c' \neq c}^{C} \mathrm{q}_{c'} \left(\sum_{a=1}^{N} \mathrm{q}_{c,a} \mathrm{q}_{c',a} \right) \right).$$

Then the expected number of trials before an urn contains balls of at least 2 different colours is

$$\sqrt{\frac{\pi}{2A_N}} \quad + \quad O(N^{1/4})$$

as $N \to \infty$.

DLP in an Interval Using Equivalence Classes

Theorem: (G.-Ruprai; PKC 2010) There is an algorithm to solve the DLP in an interval of size N in groups with fast inversion that requires (ignoring the troubles with cycles) average expected $(1.36 + o(1))\sqrt{N}$ group operations.

Possibly can be slightly improved?

Higher Dimensional Versions

- The original Gaudry-Schost algorithm was introduced for solving problems like: Given g_1, g_2, h, N_1, N_2 find (a_1, a_2) if it exists such that
$$h = g_1^{a_1} g_2^{a_2}$$
and
$$0 \leq a_1 \leq N_1, \quad 0 \leq a_2 \leq N_2.$$
- Applications: point counting on hyperelliptic curves; Brands protocol; Cramer, Gennaro and Schoenmakers; DLP coming from GLV method.
- Let $N = N_1 N_2$. G.-Ruprai (Cirencester 2009): one can solve the 2-dim DLP in $(2.36 + o(1))\sqrt{N}$ group operations.
- Wei Liu considers equivalence classes of size 4 that naturally arise in ECDLP instances arising from the GLV method. She showed this 2-dimensional ECDLP (corresponding to a set of N possible DLP instances) can be solved in $(1.03 + o(1))\sqrt{N}$ group operations.

Isogenies

- Let E_1 and E_2 be elliptic curves over $\mathbb{F}_q$. An isogeny is a morphism $\phi : E_1 \rightarrow E_2$ of elliptic curves that is a group homomorphism.
- Isogenies transfer the ECDLP from $E_1(\mathbb{F}_q)$ to $E_2(\mathbb{F}_q)$.
- If $\#E_1(\mathbb{F}_q) = \#E_2(\mathbb{F}_q)$ then there is an isogeny (defined over $\mathbb{F}_q$) from E_1 to E_2.
- A natural problem is: Given E_1 and E_2 as above, to find an isogeny between them.
- The isogeny problem is the problem of finding collisions in the Charles-Goren-Lauter hash function.
  (In this case E_1 and E_2 are supersingular.)
- There are many other applications of computing isogenies: See Drew Sutherland's talk and his survey paper in the ANTS 2012 conference proceedings.

Ordinary case

- Let E be an elliptic curve over $\mathbb{F}_q$ and let L be a set of primes. Define a graph X whose vertices are labelled by all $j(E')$ such that $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$ and with edges corresponding to ℓ-isogenies for $\ell \in L$.

  Let N be the number of vertices in X.
- Building heavily on work of Kohel, I gave an algorithm in 1999 to compute an isogeny from E_1 to E_2 (for arbitrary $j(E_1), j(E_2) \in X$).

  The algorithm grows two trees out from E_1 and E_2. In the graph theory literature it is called "bi-directional search" and is due to Pohl in 1969.
- Ignoring steep volcanos, it runs in time and space $O(\sqrt{N})$ field operations/elements.

Ordinary case

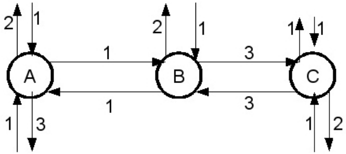
- For ordinary elliptic curves there is a low memory algorithm for this problem due to G.-Hess-Smart.
  This algorithm uses random walks and distinguished points. It is analogous to the kangaroo algorithm as there are two types of walks. The probability analysis uses the birthday paradox. The running time is $O(\sqrt{N})$ field operations.
- Anton Stolbunov has noted a bug in the description of that algorithm, and a way to greatly improve the constant in the running time. This work will appear in a joint paper.
- One feature is that the algorithm first generates a "long" isogeny chain that can be "smoothed" using notions from index calculus algorithms for ideal class groups.
  Also see Bröker-Charles-Lauter and Jao-Soukharev.
  Bisson and Sutherland give an exponential time method (also using random walks) to generate a very smooth isogeny. See "A Pollard-type algorithm for finding short product representations in finite groups"

Supersingular case

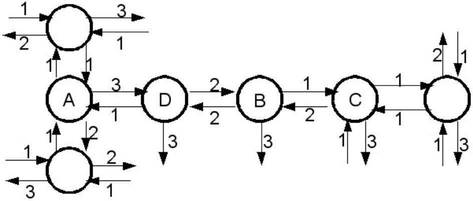
- An elliptic curve E over $\mathbb{F}_{p^n}$ is supersingular if $\#E(\mathbb{F}_{p^n}) \equiv 1 \pmod{p}$.
- If E is supersingular then $j(E) \in \mathbb{F}_{p^2}$.
- Let L be a set of primes, e.g., $L = \{2\}$.
  Define the supersingular isogeny graph X to have vertices $j(E)$ where $E/\mathbb{F}_{p^2}$ is supersingular and edges being isogenies of degree ℓ (over $\overline{\mathbb{F}}_p$) for $\ell \in L$.
  This is an expander graph.
- The isogeny graph has $N \approx p/12$ vertices.
  When $L = \{2\}$ it is a 3-regular graph.
- The basic bi-directional search solves the isogeny problem in time and space $O(\sqrt{N}) = O(\sqrt{p})$ field operations/elements.
- One can consider a low-storage algorithm for the problem, that uses random walks and the birthday paradox.
  However, when $L = \{2\}$ there are serious difficulties with small cycles in the random walks.

Supersingular case

Chang-An Zhao and I have studied how to design random walks that avoid small cycles.



Four cycle



Supersingular case over $\mathbb{F}_p$

- Let L be a set of primes.
  Define the supersingular isogeny graph X to have vertices E where $E/\mathbb{F}_p$ is supersingular and edges being isogenies of degree ℓ (over $\overline{\mathbb{F}}_p$) for $\ell \in L$.
- If $E/\mathbb{F}_p$ is supersingular then $\sqrt{-p} \in \mathrm{End}(E)$.
- The theory of complex multiplication implies that E is the reduction of an elliptic curve E' in characteristic zero with $\mathrm{End}(E')$ equal to either $\mathbb{Z}[\sqrt{-p}]$ or $\mathbb{Z}[(1+\sqrt{-p})/2]$.
- Hence the isogeny graph has a volcano structure with floor of size $h(-4p)$ corresponding to ideal classes in $\mathbb{Z}[\sqrt{-p}]$ and, if $p \equiv 3 \pmod 4$, a crater of size $h(-p)$ corresponding to ideal classes in $\mathbb{Z}[(1+\sqrt{-p})/2]$.
- Each j-invariant appears twice, as the non-trivial quadratic twist of a curve with $p+1$ points also has $p+1$ points.

Supersingular case over $\mathbb{F}_p$

- The isogeny graph has $N = O(\sqrt{p}\log(p))$ vertices.
- A natural task is to solve the isogeny problem in this graph in time $O(\sqrt{N})$.
- The basic bi-directional search solves the isogeny problem in time and space $\tilde{O}(\sqrt{N}) = \tilde{O}(p^{1/4})$ field operations/elements.
- One can consider a low-storage algorithm for the problem, that uses random walks and the birthday paradox, analogous to the algorithm for the ordinary isogeny graph.

  One can also use the concept of smoothing.
- This is work in progress with Christina Delfs.

Conclusion

- Pseudorandom walks enable low-storage and distributed algorithm for various computational problems related to ECC.
- The Gaudry-Schost algorithm is a really useful idea that perhaps can be used to solve some currently unsolved problems.
- We have used the Gaudry-Schost idea to combine the DLP in an interval (or higher-dimensional box) with equivalence classes.
- We have developed a powerful generalisation of the birthday paradox that will be a useful tool when working with the Gaudry-Schost algorithm.

Thank you very much

Volcanos and pyramids . . .

Open Question

Low Hamming weight DLP: Given g, h, w to find $0 \leq a < r$, if it exists, such that $h = g^a$ and the Hamming weight of a is $\leq w$. Let $N = \binom{\lfloor \log_2(r) \rfloor + 1}{w}$.

- Time/memory tradeoff in $O(\sqrt{N})$ group operations and storage.
- Van Oorschot and Wiener have a general method to transform a time/memory tradeoff into a low memory algorithm, but it only gives an algorithm with expected $cN^{3/4}$ group operations.
- It is an open problem to give a low memory algorithm for this problem with expected running time of $O(N^{1/2})$ group operations.