J. Cryptology (1991) 4: 207–213

Journal of Cryptology

Elliptic Curve Implementation of Zero-Knowledge Blobs[1]

Neal Koblitz
Department of Mathematics GN-50, University of Washington, Seattle, WA 98195, U.S.A.

Abstract. In [2] the authors show how to construct the building blocks for perfect zero-knowledge proofs called "blobs" using the discrete log problem. Contrary to what they remark on p. 73 of [2], we argue that the Mordell group of an elliptic curve is *more* suitable than the multiplicative group of a finite field for the construction of a hard cryptographic suite of problems.

Key words. Discrete logarithm, Elliptic curve, Zero knowledge, Blob, Hard problem, Primality.

1. Introduction

In [3] a certain basic building block in the construction of zero-knowledge protocols was introduced. This concept was later termed a "blob." A "blob" is simply an encryption of a bit. If a system of blobs satisfies certain properties, then it can be used for perfect zero-knowledge proofs. For details and formal definitions see [2] and [4].

The implementation of blobs can be based on various cryptographic assumptions. In [2] the authors show how to do this with an assumption about intractability of the discrete logarithm. After discussing the discrete log in the multiplicative group of a prime finite field $\mathbf{F}_p$, they describe the generalization to discrete log in any suitable family of finite groups. Here their main theorem states that perfect zero-knowledge interactive arguments for a wide class of problems follow from what they call Cryptographic Assumption Π^+ ("there is a hard cryptographic constructive suite of discrete-logarithm problems"). However, on p. 73 they comment that "because Mordell groups [of elliptic curves] need not be cyclic, it seems unlikely that Cryptographic Assumption Π^+ holds in this case."

The purpose of this paper is to argue that, on the contrary, a family of Mordell groups of suitably chosen elliptic curves is *more* likely to give a hard cryptographic suite—even in a strengthened sense of "hard"—than is a family of groups $\mathbf{F}_p^*$. The advantages of Mordell groups arise because of:

(1) the much greater choice available (by varying the coefficients of the defining equation over each $\mathbf{F}_p$);

[1] Date received: October 17, 1990. Date revised: June 12, 1991.

(2) the resulting possibility of limiting ourselves to groups of prime order; and
(3) the absence of any known discrete log algorithm (except in very special cases) which is not fully exponential.

2. Elliptic Curves and Exponentially Hard Problems

Basic information on elliptic curves can be found, for example, in [6] and [13]. In addition, in [7] we studied the particular problem of finding elliptic curves over a finite field having a prime number of points.

In the formalism of Section 5 of [2], the *elements* of the prime-order elliptic curve discrete log problem are strings $\langle\langle p, X_0, Y_0, X_1, Y_1, a, N\rangle, x\rangle$ such that p is a prime number; $X_0, Y_0, X_1, Y_1, a \in \mathbf{F}_p$; $4a^3 + 27b^2 \neq 0$, where b has been set equal to $Y_0^2 - X_0^3 - aX_0$ (this means that $E: Y^2 = X^3 + aX + b$ is an elliptic curve containing $P_0 = (X_0, Y_0)$); $P_1 = (X_1, Y_1)$ is another point of E; $N = \#E$ is a prime number; and $P_1 = xP_0$. An *instance* of the problem is a string of the form $\langle p, X_0, Y_0, X_1, Y_1, a, N\rangle$ with the above properties, and a *solution* is the unique integer x modulo N such that $P_1 = xP_0$.

In [2] the notion of a *hard* suite of problems is defined, where, roughly speaking, the term "hard" means "harder than polynomial." We introduce a stronger notion, that of an *exponentially hard* suite of problems.

Definition 1. A family $\mathscr{C} = \{C_m\}$ of randomized circuits, where C_m is the circuit for input of length m, is *exponential* if there exist a positive constant c and an m_0 such that the size of the circuit C_m is $> e^{cm}$ for $m > m_0$. A family $\mathscr{C}$ that is not exponential is said to be *subexponential.*

Following [2], we then make the following:

Definition 2. A suite of problems X is *exponentially hard* if, for every subexponential family $\mathscr{C} = \{C_m\}$ of randomized circuits, there exist a positive constant c and an m_0 such that, for all $m > m_0$,

$$\text{Prob}\{\alpha \in X_m\text{: Prob}\{\alpha \text{ is solved by } C_m\} > e^{-cm}\}$$

is less than e^{-cm}.

In the case of the discrete logarithm problem on an elliptic curve, the one general purpose algorithm we have is the Shanks giant-step–baby-step algorithm, which, unlike index calculus-type algorithms, does not depend on the structure of the group G in which we are working. The running time of this algorithm is slightly more than the square root of the largest prime factor of $\#G$. Thus, if $m = \log \#G$ denotes the size of G and if $\#G$ is "almost prime" (i.e., the product of a prime number and a small integer), then we have the running time $e^{(0.5+o(1))m}$, i.e., the algorithm is fully exponential. It is also clear that there is exponentially small probability of hitting upon an instance of the discrete log problem which the algorithm can solve in faster than exponential time, and random reruns of the algorithm also have an exponen-

tially small probability of finding the discrete logarithm. In other words, the prime-order elliptic curve discrete log problem is exponentially hard if we have only the giant-step–baby-step algorithm.

However, very recently a second algorithm was developed by Menezes, Okamoto, and Vanstone (MOV) [9]. Using the Weil pairing on an elliptic curve E, MOV were able to imbed the group of $\mathbf{F}_p$-points on E in the multiplicative group of the field $\mathbf{F}_{p^k}$ for some integer k. That reduces the discrete log problem on E to the discrete log problem in $\mathbf{F}_{p^k}^*$. Now in $\mathbf{F}_{p^k}$ we can hope to use a version of the number field sieve to obtain an algorithm with running time

$$\exp((c + o(1))((\log p^k)^{1/3})((\log \log p^k)^{2/3})). \tag{1}$$

See [5] for the case $k = 1$; the algorithm has not yet been extended to $k > 1$, but we adopt the "optimistic" supposition that the above time estimate is the complexity of the discrete log in $\mathbf{F}_{p^k}^*$ for $k > 1$ as well. Since the time for the MOV reduction from E to $\mathbf{F}_{p^k}^*$ is much less than (1), we also take (1) as the complexity of the discrete log on E using MOV reduction followed by number field sieve.

Note that k must be small for this second algorithm to be subexponential. Namely, a necessary condition is that $k \leq \log^2 p$, since otherwise $(\log p^k)^{1/3} > \log p$, and the running time (1) is fully exponential in $\log \#E \sim \log p$. There is a very special class of elliptic curves—called *supersingular* curves—for which k is small. However, a randomly generated elliptic curve has an exponentially small probability of being supersingular; and, as we shall see below, for most randomly generated elliptic curves we must necessarily have $k > \log^2 p$.

Without having to examine the details of the MOV reduction, we immediately see that a necessary condition for E to be imbedded in $\mathbf{F}_{p^k}^*$ is that the order $N = \#E$ divide $p^k - 1$, i.e., that k be a multiple of the order of the element p in the multiplicative group modulo N. It is shown below that for two different large primes p and N, it is highly unlikely that the order of p modulo N is less than $\log^2 p$. Thus, in our selection of prime-order elliptic curves, we can easily avoid the rare cases when MOV reduction leads to a subexponential solution of the discrete log problem. This informal argument is made precise in the proof of the main theorem (see Lemma 2).

In conclusion, the prime-order elliptic curve discrete logarithm is exponentially hard if the order of p modulo $N = \#E$ is $> \log^2 p$, assuming the best algorithms available at present (i.e., giant-step–baby-step and MOV reduction plus number field sieve).

3. The Cryptographic Property

In [2] a *cryptographic* suite of problems is defined to mean a suite of problems where each problem is accompanied by a certificate that demonstrates that it is a problem of the type claimed. It is the cryptographic property which leads to perfect (rather than "almost perfect") zero-knowledge interactive arguments.

In [2] the use of the discrete log in $\mathbf{F}_p^*$ is complicated by the need to give a certificate that the base g is really a generator (or at least an element of large enough order). This entails factoring $p - 1$, or else choosing special types of primes p.

In our situation this complication disappears, because we allow only groups of prime order. Thus, the only thing that needs a certificate is the primality of p and of $N = \#E$, and by using the method of Adleman and Huang [1] this can be provided in probabilistic polynomial time. At first it might seem that requiring the groups to have prime order is asking too much. However, our main theorem, proved in the next section, says that, because of all the flexibility we have in choosing p and the equation of the curve, a suite of Mordell groups all of prime order can be constructed in probabilistic polynomial time.

4. A Probabilistic Polynomial-Time Construction

We use the following procedure to generating a suite of prime-order elliptic curve discrete log problems:

(1) Find a random prime p of suitable size, using the method of Adleman and Huang [1] to generate a certificate of primality.
(2) Generate random X_0, Y_0, $a \in \mathbf{F}_p$, set $b = Y_0^2 - X_0^3 - aX_0$, and check that $4a^3 + 27b^2 \neq 0$ (repeat step (2) if $4a^3 + 27b^2 = 0$).
(3) Use Schoof's polynomial-time algorithm [12] to compute $N = \#E$, where E is the elliptic curve $Y^2 = X^3 + aX + b$.
(4) If N is composite, go back to step (2); if N is prime, generate an Adleman–Huang certificate for it.
(5) Check that $p^j \not\equiv 1 \bmod N$ for $1 \leq j \leq \log^2 p$ (go back to step (2) if p has order $\leq \log^2 p$ modulo N).
(6) Choose a random positive integer $x < N$, and set $P_1 = xP_0$, where $P_0 = (X_0, Y_0)$, $P_1 = (X_1, Y_1)$.

Theorem. *The above procedure constructs a cryptographic suite of discrete logarithm problems in bounded probabilistic polynomial time. Assuming time complexity for elliptic curve discrete log as determined by the fastest algorithms known at present (i.e., (1) giant-step–baby-step and (2) MOV reduction followed by number field sieve), the suite of problems is exponentially hard.*

Proof. In Section 2 we saw that the suite of problems is exponentially hard, assuming no fundamental breakthrough in the available algorithms. To show that the procedure is probabilistic polynomial time, we need some lemmas.

Lemma 1. *Let S_M denote the set of points (x, y) in the region $M/2 \leq x \leq M$, $x - \sqrt{x} \leq y \leq x + \sqrt{x}$ such that both x and y are prime numbers. Then, for some effectively computable positive constant c_1 and for large M,*

$$\#S_M \geq c_1 \frac{M^{3/2}}{\log^2 M}. \tag{2}$$

Proof of Lemma. A. Odlyzko pointed out to me that this is elementary. Namely, without loss of generality we may assume that M is of the form $M = 4m^2$. Divide the interval $(\frac{1}{2}M, M]$ into the $2m$ subintervals $((i-1)m, im]$, $2m < i \leq 4m$. Let

$A = \pi(M) - \pi(\frac{1}{2}M)$ be the number of primes in $(\frac{1}{2}M, M]$, and let a_i be the number of primes in the ith subinterval. By the Prime Number Theorem, $M/3 \log M < A < M/\log M$ for M large. Any ordered pair (x, y) of primes in the same subinterval gives a point in the set S_M, since $\sqrt{x} > m > |y - x|$. This leads to the following lower bound for $\#S_M$: $\sum a_i^2$; but the minimum of $\sum a_i^2$ subject to the condition $\sum a_i = A$ is attained when all the a_i are equal, i.e., $a_i = A/2m$. This leads to the lower bound

$$2m\left(\frac{A}{2m}\right)^2 = M^{-1/2}A^2 > \frac{M^{3/2}}{9 \log^2 M}. \qquad \square$$

Lemma 2. *Let $\tilde{S}_M$ be the subset of S_M consisting of points such that the order of x modulo y is $\leq \log^2 x$. Then $\#\tilde{S}_M < 2M \log^3 M$.*

Proof of Lemma. For a fixed prime y, let $S_{M,y}$ denote the set $\{x \in \mathbf{Z} | \frac{1}{2}y < x < \frac{3}{2}y$, order of x modulo y is $\leq \log^2 M\}$. Then, since at most j residue classes mod y have order j, it follows that $\#S_{M,y} \leq \sum_{1 \leq j \leq \log^2 M} j < \log^4 M$. Clearly,

$$\begin{aligned} \#\tilde{S}_M &\leq \#\{(x, y) \in \mathbf{Z}^2 | 0 \leq x \leq M, \tfrac{2}{3}x < y < 2x, \\ &\qquad y \text{ is prime, order of } x \bmod y \text{ is } \leq \log^2 x\} \\ &\leq \sum_{\text{primes } y \leq 2M} \#S_{M,y} < \pi(2M) \log^4 M < 2M \log^3 M. \end{aligned} \qquad \square$$

Lemma 3. *For p a prime, let $S(p)$ denote the set of integers y in the interval $p - \sqrt{p} \leq y \leq p + \sqrt{p}$ such that y is prime and the order of p modulo y is $> \log^2 p$. Then, for some effectively computable positive constant c_2 and for M large,*

$$\sum_{M/2 \leq p \leq M, p \text{ prime}} \#S(p) \geq c_2 \frac{M^{3/2}}{\log^2 M}.$$

This lemma follows immediately from Lemmas 1 and 2.

The next lemma, which is due to Lenstra, tells us that the number of points on elliptic curves over $\mathbf{F}_p$ is nearly uniformly distributed in the interval of size $\sqrt{p}$ around $p + 1$.

Lemma 4 (Proposition 1.16(a) of [8]). *There exists an effectively computable positive constant c_3 such that, for each prime number $p > 3$ and for any subset S of*

$$\{s \in \mathbf{Z} | |s - (p + 1)| \leq \sqrt{p}\},$$

the number of triples $\langle X_0, Y_0, a \rangle \in \mathbf{F}_p^3$ such that

$$4a^3 + 27b^2 \neq 0, \qquad \text{where} \quad b = Y_0^2 - X_0^3 - aX_0,$$

and the elliptic curve $E: Y^2 = X^3 + aX + b$ *satisfies* $\#E \in S$

is at least

$$c_3(\#S - 2)\frac{p^{5/2}}{\log p}.$$

We now return to the proof of the theorem. We want a lower bound for the number of four-tuples $\langle p, X_0, Y_0, a\rangle$ such that

(i) $p \le M$ is a prime,
(ii) $Y^2 = X^3 + aX + b$, where $b = Y_0^2 - X_0^3 - aX_0$, is an elliptic curve E over $\mathbf{F}_p$,
(iii) $N = \#E$ is prime, and
(iv) p has order $> \log^2 p$ modulo N.

Using Lemmas 3 and 4, we find that the number of such four-tuples is at least

$$\begin{aligned} & c_3 \sum_{M/2 \le p \le M,\, p \text{ prime}} \max(0,\ \#S(p) - 2)\frac{p^{5/2}}{\log p} \\ & \ge c_3 2^{-5/2} \frac{M^{5/2}}{\log M}\left(\left(\sum_{M/2 \le p \le M} \#S(p)\right) - 2\pi(M)\right) \\ & \ge c_3 2^{-5/2} \frac{M^{5/2}}{\log M}\left(c_2 \frac{M^{3/2}}{\log^2 M} - 2\frac{M}{\log M}\right) \\ & \ge c_4 \frac{M^4}{\log^3 M} \end{aligned} \tag{3}$$

for M large, where c_4 is an effectively computable positive constant.

To show that the procedure at the beginning of the section is probabilistic polynomial time, it suffices to show that, in running through the set of four-tuples of positive integers $\langle p, X_0, Y_0, a\rangle$ in the range $p \le M$, $X_0 \le p$, $Y_0 \le p$, $a \le p$, the probability that such a four-tuple satisfies conditions (i)–(iv) above is bounded from below by a reciprocal power of $\log M$. However, since the number of four-tuples is asymptotic to $M^4/4$, it follows from (3) that this probability is

$$\ge \frac{4c_4}{\log^3 M}.$$

This completes the proof of the theorem. □

Remarks. 1. For a fixed prime p, at present we cannot prove anything about the number of primes in the interval $[p - \sqrt{p}, p + \sqrt{p}]$—even that there are any. Thus, a subtlety in the proof of our theorem is that it is important that we are simultaneously ranging through different p and different $N = \#E(\mathbf{F}_p)$ (see Lemma 1).

2. In an actual implementation, we would probably want to make modifications which, while sacrificing theoretical certainty, make the exponentially hard cryptographic suite of problems "constructive" in the practical sense of the word:

(1) Agree to be satisfied with a probabilistic primality test for p and for $N = \#E$ rather than requiring an Adleman–Huang certificate.
(2) Work with elliptic curves E for which $N = \#E$ is "almost prime," i.e., the product $N = N_0 N_1$ of a small factor N_0 and a prime N_1, where we consider the discrete log problem in the group generated by a point of order N_1 (or equivalently, in the image group $N_0 E$ of E under multiplication by N_0).

References

[1] L. Adleman and M. Huang, Recognizing primes in random polynomial time, *Proc. 19th ACM Symp. on Theory of Computing*, 1987, pp. 462–469.
[2] J. F. Boyar, S. A. Kurtz, and M. W. Krentel, A discrete logarithm implementation of perfect zero-knowledge blobs, *J. Cryptology*, **2** (1990), 63–76.
[3] G. Brassard and C. Crépeau, Nontransitive transfer of confidence: a perfect zero-knowledge interactive protocol for SAT and beyond, *Proc. 27th IEEE Symp. on Foundations of Computer Science*, 1986, pp. 188–195.
[4] G. Brassard, D. Chaum, and C. Crépeau, Minimum disclosure proofs of knowledge, *J. Comput. System Sci.*, **37** (1988), 156–189.
[5] D. M. Gordon, Discrete logarithms in $GF(p)$ using the number field sieve, Preprint.
[6] N. Koblitz, Elliptic curve cryptosystems, *Math. Comp.*, **48** (1987), 203–209.
[7] N. Koblitz, Primality of the number of points on an elliptic curve over a finite field, *Pacific J. Math.*, **131** (1988), 157–165.
[8] H. W. Lenstra, Factoring integers with elliptic curves, *Ann. of Math.*, **126** (1987), 649–673.
[9] A. Menezes, T. Okamoto, and S. A. Vanstone, Reducing elliptic curve logarithms to logarithms in a finite field, Preprint.
[10] V. Miller, Short programs for functions on curves, Unpublished manuscript, 1986.
[11] V. Miller, Uses of elliptic curves in cryptography, *Advances in Cryptology: Proceedings of Crypto '85*, Springer-Verlag, Berlin, 1986, pp. 417–426.
[12] R. Schoof, Elliptic curves over finite fields and the computation of square roots mod p, *Math. Comp.*, **44** (1985), 483–494.
[13] J. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986.