Extracting bits from coordinates of a point of an elliptic curve

Nicolas Gürel

Abstract

In the classic Diffie-Hellman protocol based on a generic group $\mathbb{G}$, Alice and Bob agree on a common secret K_{AB} (master secret) which is indistinguishable from another element of $\mathbb{G}$ but not from a random bits-string of the same length. In this paper, we present a new deterministic method to extract bits from K_{AB} when $\mathbb{G}$ is an elliptic curve defined over a quadratic extension of a finite field. In the last section, we show that it is also possible to extract a few bits when $\mathbb{G}$ is an elliptic curve defined over a prime field.

Keywords: Elliptic curve cryptosystem, key derivation, pseudo-random extractors.

Introduction

Let $\mathbb{G} = \langle g \rangle$ be a cryptographic cyclic group of order q and ℓ be the bit size of q. In the classic Diffie-Hellman key exchange [6], Alice and Bob agree on a common secret K_{AB} (master secret) which is indistinguishable from another element of $\mathbb{G}$ under the decisional Diffie-Hellman assumption (DDH). In most cryptographic protocols, the secret key K_s, to encrypt and authenticate data, has to be indistinguishable from a random bit-string with a uniform distribution. This imply that K_{AB} can not be directly used as a session key. For instance, if $\mathbb{G}$ is a subgroup of the multiplicative group of a finite field $\mathbb{F}_p$, or if $\mathbb{G}$ is an elliptic curve defined over a finite field $\mathbb{F}_p$, there is no easy ℓ-bit long representation for group elements, and therefore the representation involves some redundancy. Although this redundancy is difficult to remove, it is trivial to detect if a bit string corresponds to a representation of an element of $\mathbb{G}$. Hence the indistinguishability must be guaranteed with an additional device.

The classical solution is to use a hash function we have in our toolbox to map the element K_{AB} into an ℓ-bit string. Then the indistinguishability can not be proved under the standard DDH assumption. The Random Oracle or some other technical assumption has to be added. Some general results in that direction can be found for example in [8, 7].

More recently Chevassut, Fouque, Gaudry and Pointcheval in [3] pointed out that working with elliptic curves can imply some nice solutions to the randomness extraction problem (their TAU method is provably secure assuming only DDH on elliptic curves). In this paper we propose two other methods to extract random bits from a point on an elliptic curve, that are also provable under DDH only.

Our first extractor, $\mathcal{H}$, works for elliptic curves defined over a quadratic extension of a prime finite field. It simply takes the first $\mathbb{F}_p$-coefficient of the abscissa of the point. The proof that this extractor has the wanted properties relies on a counting theorem that amounts to applying the Hasse Weil bound on some curve included in the Weil restriction of the elliptic curve.

The second extractor works for elliptic curves defined over a prime field. It takes the first k-bits of the binary representation of the abscissa of a point for k small enough (strictly less than $\ell/2$ if p is an ℓ-bits prime number). In this case, the proof is a consequence of the Polya-Vinogranov's inequality.

In the next section, we start by quickly recalling the TAU method and then we propose our new extractors. The core of the paper will be devoted to the proof that our extractor $\mathcal{H}$ is a good extractor. In section 4, we discuss a few practical applications. The last section describe our second method in detail.

1 Using elliptic curves for the extraction of bits

Let us fix some classical definitions related to probability.

Indistinguishability. Let X_ℓ and Y_ℓ be two distributions over $\{0,1\}^\ell$. We say that the *statistical distance* between X_ℓ and Y_ℓ is bounded by $\Delta(\ell)$ if

$$\sum_{x\in\{0,1\}^\ell} \left| \Pr_{K\in_R X_\ell}[K=x] - \Pr_{K\in_R Y_\ell}[K=x] \right| \leq \Delta(\ell).$$

We say that X_ℓ and Y_ℓ are *statistically indistinguishable* if for any polynomial P we have asymptotically

$$\Delta(\ell) < \frac{1}{P(\ell)}.$$

1.1 Adding conditions on the curve for an efficient extractor

In the following methods, the uniform distribution in the field $\mathbb{F}_p$ and the uniform distribution in $\{0,1\}^\ell$ have to be statistically indistinguishable. It is possible only if p is a prime number of the form $2^\ell - \varepsilon$ where ε is less than $2^{\ell/2}$. In this section we fix p a prime number of this form.

TAU method of [3]. Let $\mathbb{E}$ be an elliptic curve defined over $\mathbb{F}_p$ that is

$$\mathbb{E} = \big\{(x,y)\in\mathbb{F}_p\times\mathbb{F}_p : y^2 = x^3+ax+b\big\}\cup\{\infty_{\mathbb{E}}\},$$

where a and b are both in $\mathbb{F}_p$ and where $\infty_{\mathbb{E}}$ denote the point at infinity. Let c be a quadratic non-residue of $\mathbb{F}_p$.

Let x_0 be an element of $\mathbb{F}_p$ and consider the element $z = x_0^3 + ax_0 + b$. There are two cases:

- z is a quadratic residue in $\mathbb{F}_p$ thus there exists y_0 such that $y_0^2 = z$ and $(x_0,y_0)\in\mathbb{E}$;
- z/c is a quadratic residue in $\mathbb{F}_p$ thus there exists y_0 such that $cy_0^2 = z$ and (x_0,y_0) is a point of an elliptic curve defined by

$$\widetilde{\mathbb{E}} = \big\{(x,y)\in\mathbb{F}_p\times\mathbb{F}_p : cy^2 = x^3+ax+b\big\}\cup\{\infty_{\widetilde{\mathbb{E}}}\},$$

called *the quadratic twist* of $\mathbb{E}$.

We define $\mathcal{H}_{\mathsf{TAU}} : \mathbb{E} \cup \widetilde{\mathbb{E}} \to \mathbb{F}_p$ by $\mathcal{H}_{\mathsf{TAU}}(P) = [P]_{\mathsf{abs}}$, the abscissa of P. It is obvious that for any $x \in \mathbb{F}_p$ we have $\#\mathcal{H}_{\mathsf{TAU}}^{-1}(x) = 2$ and from that result we can deduce that the distribution of the image by $\mathcal{H}_{\mathsf{TAU}}$ of a random point in $(\mathbb{E} \cup \widetilde{\mathbb{E}})$ and the uniform distribution in $\{0,1\}^\ell$ are statistically indistinguishable. In practice, the choice of $\mathbb{E}$ is a little bit more subtle because the discrete logarithm has to be hard in both $\mathbb{E}$ and $\widetilde{\mathbb{E}}$. In [3], they explain how to construct this kind of curve using the theory of Complex Multiplication or the SEA algorithm. Notice that the standard curve defined over the finite field P-384 as describe in [5] (see section 4 for more details), can be used for the TAU method.

Definition of the extractor $\mathcal{H}$. Let $\mathbb{E}$ be an elliptic curve defined over a finite field $\mathbb{F}_{p^2} = \mathbb{F}_p[t]/(t^2 - c)$ where c is a quadratic non-residue of $\mathbb{F}_p$. The field $\mathbb{F}_{p^2}$ can be considered as a $\mathbb{F}_p$-vector space equipped with the natural basis $\{1, t\}_{\mathbb{F}_p}$. This implies that if $z \in \mathbb{F}_{p^2}$ then there exist two elements z_0, z_1 in $\mathbb{F}_p$ such that $z = z_0 + tz_1$. We show that if $P = (x_0 + tx_1, y_0 + ty_1) \in_R \mathbb{E}$ then the function $\mathcal{H}(P) = x_0$ gives a good randomness extractor. More exactly, theorem 1 of section 2 gives some explicit bounds to $\#\mathcal{H}^{-1}(x)$ for x in $\mathbb{F}_p$ and we use this result to prove, in section 3, that the distribution of the image by $\mathcal{H}$ of a random point in $\mathbb{E}$ and the uniform distribution in $\{0,1\}^\ell$ are statistically indistinguishable. In section 4, we give some examples of curves which can be used in this context.

1.2 The general case

Let p be a prime number and $\mathbb{E}$ be an elliptic curve defined over a finite field $\mathbb{F}_p$. We show in the last section that we can extract a few bits from the abscissa of an point $P = (x, y) \in \mathbb{E}$. More exactly, we denote by $[x]_k$ the first k-bits of the binary representation of x and we define $\mathcal{H}_k(P) = [x]_k$. We give an explicit bound, dependent on k and p, to the statistical distance between the distribution of the image by $\mathcal{H}_k$ of a random point in $\mathbb{E}$ and the uniform distribution in $\{0,1\}^k$. We discus the efficiency of this method through an example.

2 The main theorem

In this section we prove the following result:

Theorem 1 *Let $\mathbb{E}$ be a curve defined over $\mathbb{F}_{q^2}$ by an affine equation $Y^2 = X^3 + aX + b$. Fix a polynomial representation $\mathbb{F}_{q^2} \cong \mathbb{F}_q[t]/(t^2 - c)$ where c is a quadratic non-residue of $\mathbb{F}_q$ and, for all x in $\mathbb{F}_{q^2}$, write $x = x_0 + tx_1$.*

We define the function $\mathcal{H}$ from $\mathbb{E}$ to $\mathbb{F}_q$ by $\mathcal{H}(x, y) = x_0$ and $\mathcal{H}(\mathcal{O}_{\mathbb{E}}) = 0$. Then for all $z \in \mathbb{F}_q^*$,

$$|\#\mathcal{H}^{-1}(z) - (q+1)| \leq 20\sqrt{q} + 14$$

and

$$m \leq \#\mathcal{H}^{-1}(0) \leq M$$

where $m = \min(2(q+1) - 4\sqrt{q}, (q+1) - 20\sqrt{q} - 14)$ and $M = \max(2(q+1) + 4\sqrt{q}, (q+1) + 20\sqrt{q} + 14)$.

Proof. Let A be the Weil restriction of $\mathbb{E}$ to $\mathbb{F}_q$. This is an abelian variety of dimension 2 defined over $\mathbb{F}_q$. For each value of z in $\mathbb{F}_q$, the points of $\mathcal{H}^{-1}(z)$ form a curve $\mathcal{C}_z$ in A. Using the substitutions $a = a_0 + ta_1$, $b = b_0 + tb_1$... and carrying out the computations symbolically, we obtain two explicit equations defining $\mathcal{C}_z$ in $\mathbb{A}^3_{\mathbb{F}_q}$

$$\begin{cases} P_1 &= Y_0^2 + cY_1^2 - 3czX_1^2 - ca_1X_1 - z^3 - b_0 - a_0z \\ P_2 &= 2Y_0Y_1 - cX_1^3 - (a_0 + 3z^2)X_1 - a_1z - b_1 \end{cases}$$

where Y_0, X_1, Y_1 are indeterminates. We note $f_1(X_1) = -(3czX_1^2 + ca_1X_1 + z^3 + b_0 + a_0z)$ and $f_2(X_1) = -(cX_1^3 + (a_0 + 3z^2)X_1 + a_1z + b_1)$. Let $\mathcal{C}'$ be the plane curve defined by the affine equation $R = \mathrm{Res}_{Y_0}(Y_0^2 + cY_1^2 + f_1, 2Y_0Y_1 + f_2) = 4cY_1^4 + 4f_1Y_1^2 + f_2^2$. Using a Gröbner basis computation, one can show that $\mathcal{C}'$ is the projection of $\mathcal{C}_z$ with respect to Y_0 but we do not need this result and we already know that $\mathcal{C}'$ contains the projection of $\mathcal{C}_z$. The end of the proof is organised as follows:

- if $\mathcal{C}'$ is not irreducible, study the geometry to reduce to the irreducible case ;
- if $\mathcal{C}'$ is irreducible
  1. find an explicit constant α such that $|\#\mathcal{C}_z - \#\mathcal{C}'| \le \alpha$;
  2. find an explicit constant β such that $|\#\mathcal{C}' - N| \le \beta$ where N is the number of points on the desingularised projective model of $\mathcal{C}'$;
  3. by Weil's theorem, the number of points N is bounded by
  $$|N - (q + 1)| \le 2g\sqrt{q}$$
  where g is the genus of C';
  4. put everything together to conclude.

The plane curve $\mathcal{C}'$ is irreducible if and only if the discriminant D of $\widetilde{R} := 4cY_1^2 + 4f_1Y_1 + f_2^2$ with respect to Y_1 is not a square in $\mathbb{F}_{q^\infty}[X_1]$. We have $D = 16(f_1 + tf_2)(f_1 - tf_2)$ which is a square if and only if $f_1 = 0$ or $f_2 = 0$. (More precisely, we use the equation of the curve $(Y_0 + tY_1)^2 = f_1 + tf_2 = (z + tX_1 - \lambda_1)(z + tX_1 - \lambda_2)(z + tX_1 - \lambda_3)$ in $\mathbb{F}_{q^\infty}$,this is the equation of an elliptic curve thus the λ's are different.) The leading coefficient of f_2 is $-c \neq 0$ and $f_1 = 0$ implies in particular that $z = 0$. It thus follows that if $z \neq 0$ then $\mathcal{C}'$ is irreducible.

Case $z \neq 0$: let (x_1, y_1) be a point on $\mathcal{C}'$. By equation P_2, if $y_1 \neq 0$ then there exists $y_0 \in \mathbb{F}_q$, which is unique, such that (y_0, y_1, x_1) is a point on $\mathcal{C}_z$. If $y_1 = 0$ then $f_2(x_1) = 0$ and there are at most two points on the fiber of (x_1, y_1) for at most 3 elements $x_1 \in \mathbb{F}_q$. Using projective coordinates, there are at most two points on the fiber of the point at infinity of $\mathcal{C}'$. We have

$$|\#\mathcal{C}_z - \#\mathcal{C}'| \le 4.$$

The curve $\mathcal{C}'$ is irreducible of degree 6. The genus of $\mathcal{C}'$ is bounded by its arithmetic genus $(6 - 1)(6 - 2)/2 = 10$. There are at most 10 singular points thus

$$|\#\mathcal{C}' - N| \le 10$$

where N verifies

$$|N - (q + 1)| \le 20\sqrt{q}.$$

Case $z = 0$ and $f_1 \neq 0$: the curve $\mathcal{C}'$ is irreducible and we can do the same computation that in the previous case.

Case $z = 0$ and $f_1 = 0$: $\mathcal{C}_z$ is the union of two elliptic curves $\mathbb{E}_1$ and $\mathbb{E}_2$ of equations $Y_1^2 \pm 1/(2t)f_2$. By Weil's theorem we have

$$|\#C_0 - 2(q+1)| \leq 4\sqrt{q}.$$

□

3 The function $\mathcal{H}$ is a good randomness extractor

In this section, we assume that p verifies the following condition: there exist two integers ℓ and ε such that $\varepsilon \in \left[1, 2^{\ell/2}\right]$ and $p = 2^\ell - \varepsilon$. We show that the distribution of the secret key K, if we take it as the image by $\mathcal{H}$ of a random point of $\mathbb{E}$, is statistically almost uniformally distributed on $\{0,1\}^\ell$ under the Elliptic Curve Decisional Diffie-Hellman assumption.

First, we compute the statistical distance between the distribution of K and the uniform distribution $\mathcal{U}_p$ in $\mathbb{F}_p$ (lemma 1) then between $\mathcal{U}_p$ and the uniform distribution $\mathcal{U}_\ell$ in $\{0,1\}^\ell$ (lemma 2). Notice that we identify $\mathbb{F}_p$ with the set $\{0, 1, \ldots, p-1\}$.

Let $\mathcal{D}$ be the distribution of K which is

$$\mathcal{D} = \{P \in_R \mathbb{E} : K = \mathcal{H}(P)\}.$$

Lemma 1 *The distribution $\mathcal{D}$ is statistically indistinguishable to the uniform distribution $\mathcal{U}_p$ in $\mathbb{F}_p$:*

$$\delta = \sum_{x \in \mathbb{F}_p} \left| \Pr_{K \in_R \mathcal{U}_p}[K = x] - \Pr_{K \in_R \mathcal{D}}[K = x] \right| \leq \frac{21\sqrt{2}}{\sqrt{2^\ell}}.$$

Proof. Let x be an element of $\mathbb{F}_p$, for the uniform distribution $\mathcal{U}_p$ the probability is given by $\Pr_{K \in_R \mathcal{U}_p}[K = x] = 1/p$. For the distribution $\mathcal{D}$ we have

$$\Pr_{K \in_R \mathcal{D}}[K = x] = \Pr_{P \in_R \mathbb{E}}[\mathcal{H}(P) = x] = \frac{\#\mathcal{H}^{-1}(x)}{\#\mathbb{E}}.$$

Using the explicit bounds of theorem 1 and Weil's theorem for $\#\mathbb{E}$ we obtain

$$\frac{p + 1 - (20\sqrt{p} + 14)}{p^2 + 1 + 2p} \leq \Pr_{K \in_R \mathcal{D}}[K = x] \leq \frac{p + 1 + (20\sqrt{p} + 14)}{p^2 + 1 - 2p}$$

if $x \neq 0$, and

$$\frac{2(p+1) - 4\sqrt{p}}{p^2 + 1 + 2p} \leq \Pr_{K \in_R \mathcal{D}}[K = 0] \leq \frac{2(p+1) + 4\sqrt{p}}{p^2 + 1 - 2p}.$$

Thus for δ we deduce the desired inequality

$$\begin{aligned}
\delta &\leq \max_{i \in \{0,1\}} \left(\left| \frac{1}{p} - \frac{2(p+1) + (-1)^i 4\sqrt{p}}{p^2 + 1 + (-1)^{i+1} 2p} \right| + p \left| \frac{1}{p} - \frac{p + 1 + (-1)^i (20\sqrt{p} + 14)}{p^2 + 1 + (-1)^{i+1} 2p} \right| \right) \\
&\leq \frac{20p^2\sqrt{p} + 18p^2 + 4p\sqrt{p} + 3p - 1}{p\,(p^2 - 2p + 1)} \leq \frac{21}{\sqrt{p}} \leq \frac{21\sqrt{2}}{\sqrt{2^\ell}},
\end{aligned}$$

for p large enough (e.g. $p > 500$). □

Lemma 2 *Let us recall that $\mathcal{U}_p$ denote the uniform distribution on the space $\mathbb{F}_p$ and $\mathcal{U}_\ell$ the uniform distribution on the space $\{0,1\}^\ell = \{0,1,\ldots,2^\ell - 1\}$. Then the statistical distance between $\mathcal{U}_p$ and $\mathcal{U}_\ell$ is bounded by $2/\sqrt{2^\ell}$.*

Proof. It is exactly the proof of lemma 9 in [3]. Let δ' be the statistical distance between $\mathcal{U}_p$ and $\mathcal{U}_\ell$ then

$$\begin{aligned}
\delta' &= \sum_{x\in\{0,1\}^\ell} \left| \Pr_{X\in_R\mathcal{U}_\ell}[X=x] - \Pr_{X\in_R\mathcal{U}_p}[X=x] \right| \\
&= \sum_{\substack{x\in\{0,1\}^\ell \\ x<p}} \left| \Pr_{X\in_R\mathcal{U}_\ell}[X=x] - \Pr_{X\in_R\mathcal{U}_p}[X=x] \right| + \sum_{\substack{x\in\{0,1\}^\ell \\ x\geq p}} \left| \Pr_{X\in_R\mathcal{U}_\ell}[X=x] - \Pr_{X\in_R\mathcal{U}_p}[X=x] \right| \\
&= \sum_{\substack{x\in\{0,1\}^\ell \\ x<p}} \left| \frac{1}{2^\ell} - \frac{1}{p} \right| + \sum_{\substack{x\in\{0,1\}^\ell \\ x\geq p}} \left| \frac{1}{2^\ell} - 0 \right| = p \times \left| \frac{1}{2^\ell} - \frac{1}{p} \right| + (2^\ell - p) \times \frac{1}{2^\ell} \\
&\leq \frac{2(2^\ell - p)}{2^\ell} \leq \frac{2\varepsilon}{2^\ell} \leq \frac{2}{\sqrt{2^\ell}}.
\end{aligned}$$

□

A simple application of lemma 1 and 2 gives us the following corollary.

Corollary 1 *When $0 < p - 2^\ell \leq 2^{\ell/2}$, an upper bound of the statistical distance between the uniform distribution on $\mathcal{U}_\ell$ and the $\mathcal{H}$ technique is given by $\frac{2}{\sqrt{2^\ell}} + \frac{21\sqrt{2}}{\sqrt{2^\ell}}$.*

4 Choice of the curve and practical consequences

Using a prime of the form $2^\ell - \varepsilon$ (see section 3) is not so restrictive. Such a prime is useful in practice because it allows a faster arithmetic than a more general prime. Noticed also that prime fields over which are defined most of the curves proposed in standards [4, 5] have this special property.

Using a curve defined over $\mathbb{F}_{p^2}$. In this case, if one needs a 80-bits key, one has to construct a cryptographically secure curve over $\mathbb{F}_{p^2}$ with p a prime near 2^{80}. It is well known how to construct such a curve using the SEA algorithm (for instance, this is implemented in MAGMA [1]).

Using a standard curve. It is also possible to use a recommended curve, $\mathbb{E}_{\mathsf{P}\text{-}N}$, described in [5] but considered over $\mathbb{F}_{p^2}$.

More exactly, the curve $\mathbb{E}_{\mathsf{P}\text{-}N}$, $N \in \{192, 384, 512\}$, is defined by its affine equation

$$\mathbb{E}_{\mathsf{P}\text{-}N} = \left\{(x,y) \in \mathbb{F}_p \times \mathbb{F}_p : y^2 = x^3 - 3x + b\right\} \cup \{\infty_{\mathbb{E}_{\mathsf{P}\text{-}N}}\},$$

where $p = 2^N - \varepsilon$ is a prime number with properties describe in section 3, b is an element in $\mathbb{F}_p$, $\infty_{\mathbb{E}_{\mathsf{P}\text{-}N}}$ the point at infinity and a generator $\mathbb{E}_{\mathsf{P}\text{-}N} = \langle G_1 \rangle$. We are going to use the curve $\mathbb{E}$ defined over $\mathbb{F}_{p^2}$ by

$$\mathbb{E} = \left\{(x,y) \in \mathbb{F}_{p^2} \times \mathbb{F}_{p^2} : y^2 = x^3 - 3x + b\right\} \cup \{\infty_{\mathbb{E}}\}.$$

If $\#\mathbb{E}_{\mathsf{P}\text{-}N} = p+1-t = q_1$, for an integer t, then it is clear that $\#\mathbb{E} = (p+1-t)(p+1+t) = q$ (consequence of the Weil's theorem). For this choice of N, $\mathbb{E}$ is cyclic but $q_2 = p+1+t$ is not necessarily almost prime (in particular if $N = 192$ or 512) so that the discrete logarithm problem in $\mathbb{E}$ can be easier than in $\mathbb{E}_{\mathsf{P}\text{-}N}$. We denote by G_2 an element of $\mathbb{E}$ of order q_2 (there exists such an element as soon as $\mathbb{E}$ is cyclic and $q_2|q$) and we define $G = G_1 + G_2$. We have to modify a little the Diffie-Hellman part of the protocol. For instance, Alice has to pick up a random integer n_A in $\mathbb{Z}/q_1\mathbb{Z}$ and a random integer r_A in $\mathbb{Z}/q_2\mathbb{Z}$ which can be considered as a public information as soon as q_2 is smooth. She computes N_A such that $n_A = N_A \mod q_1$ and $r_A = N_A \mod q_2$ and sends $N_A \cdot G = n_A \cdot G_1 + r_A \cdot G_2$. Bob does the same kind of computations and they obtain the master key $K_{AB} = (N_A N_B) \cdot G = (n_A n_B) \cdot G_1 + (r_A r_B) \cdot G_2$ which can be considered as a random element of $\mathbb{E}$ (if n_A, n_B, r_A and r_B are randomly chosen in there respective set). The security of K_{AB} is based on the discrete logarithm of $\mathbb{E}_{\mathsf{P}\text{-}N}$ and $K = \mathcal{H}(K_{AB})$ can be used as a pseudo-random bit string of length N.

5 Using an elliptic curve defined over a prime field

In this section, we give a pseudo-random extractor which is working for any secure elliptic curves defined over a prime field. This method seems to be the more attractive for cryptographic applications but more than half of the entropy is lost for a weak randomness quality compare to the $\mathcal{H}$ method.

Notations and tools. Let p be a prime number and let $\mathbb{F}_p$ be a prime field with p elements which we identify with the set $\{0, 1, \ldots, p-1\}$. If $x \in \mathbb{F}_p$, we denote by $[x]_k$ the k-first less significant bits of the binary representation of x.

Let $\{S_i\}_i$ be a family of sets in $\mathbb{F}_p$ defined by:

$$\begin{aligned} S_0 &= \{x \in \mathbb{F}_p : [x]_k = 0\}, \\ S_1 &= \{x \in \mathbb{F}_p : [x]_k = 1\}, \\ &\vdots \\ S_{2^k} &= \{x \in \mathbb{F}_p : [x]_k = 2^k - 1\}. \end{aligned}$$

We write $p = s2^k + r$ where $0 < r < 2^k$. We have $\#S_i = s$ or $s - 1$.

Let f be an irreducible squarefree polynomial of degree 3 defined over $\mathbb{F}_p$ and we denote by $\left(\frac{\cdot}{p}\right)$ the Legendre symbol. We recall that the Weil's bound is gives

$$\left| \sum_{x \in \mathbb{F}_p} \left(\frac{f(x)}{p} \right) \right| \leq 2\sqrt{p},$$

and if N is a positive integer smaller than p the Polya-Vinogranov's bound (see [2]) gives

$$\left| \sum_{x=0}^{N-1} \left(\frac{f(x)}{p} \right) \right| \leq 3\sqrt{p} \log p.$$

A good random-extractor. Let $\mathbb{E}$ be an elliptic curve defined over $\mathbb{F}_p$ given by

$$\mathbb{E} = \{(x, y) \in \mathbb{F}_p \times \mathbb{F}_p : y^2 = x^3 + ax + b\} \cup \{\infty_{\mathbb{E}}\},$$

where $\infty_{\mathbb{E}}$ denote the point at infinity.

We define the function $\mathcal{H}_k$ from $\mathbb{E}$ to $\{0,1\}^k$ by $\mathcal{H}_k(x, y) = [x]_k$. For $z \in \{0,1\}^k$ we have

$$\left|\#\mathcal{H}_k^{-1}(z) - s\right| \leq 3\sqrt{p}\log p.$$

using the Polya-Vinogranov's inequality.

Let $\mathcal{D}_k$ be the distribution of the image by $\mathcal{H}_k$ of a random point in $\mathbb{E}$ and let us compute δ_k, the statistical distances between $\mathcal{D}_k$ and the uniform distribution in $\{0,1\}^k$:

$$\begin{aligned}
\delta_k &= \sum_{x \in \{0,1\}^k} \left| \Pr_{X \in_R \{0,1\}^k}[X = x] - \Pr_{P \in_R \mathbb{E}}[\mathcal{H}_k(P) = x] \right| = \sum_{x \in \{0,1\}^k} \left| \frac{1}{2^k} - \frac{\#\mathcal{H}_k^{-1}(x)}{\#\mathbb{E}} \right| \\
&\leq \max_{i=0,1} \left| 1 - \frac{2^k s + (-1)^i 2^k 3\sqrt{p}\log(p)}{p + 1 + (-1)^{i+1} 2\sqrt{p}} \right| = \max_{i=0,1} \left| 1 - \frac{p - r + (-1)^i 2^k 3\sqrt{p}\log(p)}{p + 1 + (-1)^{i+1} 2\sqrt{p}} \right| \\
&\leq \frac{2^k 3\sqrt{p}\log(p) + 2\sqrt{p} - 1}{p - 2\sqrt{p} + 1} = Q_k(p).
\end{aligned}$$

Formally, if we take α such that $2^k = p^\alpha/(3\log p)$ then we have

$$Q_k(p) = \frac{1}{p^{1/2-\alpha}} + \frac{2}{\sqrt{p}} + O\left(\frac{1}{\sqrt{p}}\right)$$

and, by definition, the distribution $\mathcal{D}_k$ and the uniform distribution in $\{0,1\}^k$ are statistically indistinguishable as soon as α is small enough compare to $1/2$.

Example. Let p a 200-bits prime number, for 50 pseudo-random bits, the statistical distance δ_{50} is bounded by 2^{-42} (for various k : $\delta_{60} < 2^{-32}$, $\delta_{70} < 2^{-22} \ldots$).

Notice that, for the same cost during the Diffie-Hellman part of the protocol, if we work with a 100-bits prime over a quadratic extension then we have 100 pseudo-random bits and the statistical distance δ is bounded by 2^{-93}.

Conclusion

We have constructed a deterministic randomness extractor $\mathcal{H}$ which can be used in any elliptic curve based protocol. The main condition to use $\mathcal{H}$ is that the ground field, over which the curve is defined, has to be a quadratic extension. It is easy to construct a secure curve on a quadratic extension and if we really want a curve defined over a prime field (for instance to use a standard curve), it suffices to work in a quadratic extension of this field.

We have also constructed a deterministic randomness extractor $\mathcal{H}_k$ which can be used without any conditions on the elliptic curve. In this case, the random bits extractor is less efficient than $\mathcal{H}$.

References

[1] W. Bosma and J. Cannon. *Handbook of Magma functions*, 1997. Available at `http://www.maths.usyd.edu.au:8000/u/magma/`.

[2] D. A. Burgess. On Dirichlet characters of polynomials. *Proc. London Math. Soc. (3)*, 13:537–548, 1963.

[3] O. Chevassut, P.-A. Fouque, P. Gaudry, and D. Pointcheval. Key derivation and randomness extraction. Cryptology ePrint Archive, Report 2005/061, 2005. Available at `http://eprint.iacr.org/`.

[4] Certicom Corp. SEC 1: Elliptic curve cryptography. Technical report, Standards for Efficient Cryptography Group, 2000. Available at `http://www.secg.org/collateral/sec1_final.pdf`.

[5] Certicom Corp. SEC 2: Recommended elliptic curve domain parameters. Technical report, Standards for Efficient Cryptography Group, 2000. Available at `http://www.secg.org/collateral/sec2_final.pdf`.

[6] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Trans. Inform. Theory*, IT–22(6):644–654, November 1976.

[7] Y. Dodis, R. Gennaro, J. Håstad, H. Krawczyk, and T. Rabin. Randomness extraction and key derivation using the CBC, cascade and HMAC modes. In Matthew K. Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 494–510. Springer, 2004.

[8] J. Håstad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.