蒙哥马利曲线与扭曲爱德华曲线

longcpp

longcpp9@gmail.com

August 5, 2019

ECDSA 签名算法 (基于 secp256k1 或者 secp256r1 曲线) 是目前主流的数字签名算法. 然而在具体应用 ECDSA 签名算法时, 稍有不慎就会引发诸多问题[1], 这包括签名过程中对随机数需求会在随机数选取不当的情况下引发多种安全问题, 也包括签名值 (r, s) 的可锻造性在特殊应用场景下 (例如区块链场景) 引发的干扰, 还包括公钥压缩时总需要额外的一个字节来表示一个比特的信息造成的存储空间浪费. 随着各种问题的出现, 也有了应对措施, 例如 RFC 6979[2] 中将签名过程中随机数的随机选取变更为通过私钥和待签名消息进行确定性行派生过程能够以避开与随机数选取相关的安全问题; 而通过在应用层约束 s 的取值范围, 则可以规避可锻造性的问题; 同样通过在应用层的逻辑约束, 例如利用素数域上二次剩余[3] 的性质, 也可以将压缩公钥的表示从 33 个字节压缩为 32 个字节. 在实现方面, 尤其是在椭圆曲线点群的加法运算, 由于 secp256k1 和 secp256r1 曲线上椭圆曲线点群加法的不完备性 (需要判定多种边界条件) 使得签名过程的常量时间实现愈发困难. 通过相应技术手段同样可以达到常量实现时间, 但也相应增加了实现的难度与代码复杂度, 同时不可避免的会对执行速度产生影响. 虽然 secp256k1 和 secp256r1 在构造曲线和选取参数时纳入了工程实现的考量, 例如 secp256k1 自带的自同态映射[4]能够加速验签操作以及 secp256r1 所采用的蒙哥马利友好的 (Montgomery Friendly)[5] 底层素数域的特征 p. 但是一个很自然的问题是, 数字签名算法的运行效率可否做到更快? 借助工程手段

[1] longcpp. ECDSA 签名机制在区块链领域中的应用. 2019. https://github.com/longcpp/CryptoInAction/tree/master/ecdsa-blockchain-dangers

[2] RFC 6979. Deterministic Usage of the Digital Signature Algorithm (DSA) and Elliptic Curve Digital Signature Algorithm (ECDSA). https://tools.ietf.org/html/rfc6979

[3] Pieter Wuille. Schnorr Signatures for secp256k1. 2018. https://github.com/sipa/bips/blob/bip-schnorr/bip-schnorr.mediawiki

[4] longcpp. 基于 secp256k1 的自同态映射加速 ECDSA 验签. 2019. https://github.com/longcpp/CryptoInAction/tree/master/secp256k1-endomorphism

[5] Gueron, Shay, and Vlad Krasnov. Fast prime field elliptic-curve cryptography with 256-bit primes. Journal of Cryptographic Engineering 5, no. 2 (2015): 141-151. https://eprint.iacr.org/2013/816.pdf

以及 SIMD 指令集的应用, 可以逐步提升执行效率. 然而更好安全性与更高的执行效率的诉求, 或许无法通过这种小步迭代和缝缝补补方式得到满足.

同时解决前述的应用安全, 实现安全以及执行效率的问题, 要求在工程手段之外更为深度的改进, 一个自然的方向是重新构建椭圆曲线以及签名机制以便在多个层次上同时改进: 改进底层算术运算加速中层点群运算, 中层点群运算适配上层协议, 并同时考虑 ECDSA 签名机制的问题与局限性加以避免. EdDSA (Edwards-curve Digital Signature Algorithm) 签名机制是这个研究方向上的成果. EdDSA 签名机制是 Bernstein 等人[6] 在 2012 年设计的基于爱德华曲线 (Edwards Curves) 的数字签名算法. EdDSA 签名机制是 Schnorr 签名机制的一个变种, 其设计初衷是在不牺牲安全性的前提下提升签名/验签速度, 并同时解决前述的 ECDSA 在应用方面存在的一些问题.

广泛使用的 EdDSA 签名机制是基于哈希函数 SHA-512 和椭圆曲线 Edwards25519 的 Ed25519 签名机制. 扭曲爱德华曲线 Edwards25519 双向有理等价于蒙哥马利曲线 Curve25519, 提供大约 128 比特的安全强度 (与 secp256k1 和 secp256r1 安全强度一致). Curve25519 是 Bernstein[7] 在 2005 年为了提升 ECDH 密钥交换协议 (Elliptic Curve Diffie-Hellman Key Agreement) 效率而提出的蒙哥马利曲线, 并同时提供了高速软件实现, 相关文献和代码实现参见 Curve25519 的网站[8]. 值得注意的是, 在 2005 年的论文中 Curve25519 实际上用来指代 ECDH 密钥交换协议, 然而后来多使用 Curve25519 指代底层的椭圆曲线, 造成了讨论时候的困难. 因此 Bernstein 在邮件中给出了名词约定规范[9], 用 Curve25519 指代底层的椭圆曲线, 用 X25519 指代基于 Curve25519 的 ECDH 密钥协议, 本文也遵循这种命名规范.

Curve25519 是基于素数域 $\mathbb{F}_q$, $q = 2^{255} - 19$ 上的蒙哥马利曲线 (Montgomery Curve), 曲线方程为 $y^2 = x^3 + 486662x^2 + x$. Curve25519 曲线双向有理等价于 (Birational Equivalent) 扭曲爱德华曲线 (Twisted Edwards Curves) Edwards25519: $x^2 + y^2 = 1 + (121665/121666)x^2y^2$, 而这条扭曲爱德华曲线则同构于 (Isomorphic) 爱德华曲线 (Edwards Curves) untwisted-Edwards25519: $-x^2 + y^2 = 1 - (121665/121666)x^2y^2$. 为什么 X25519 直接构建在 Curve25519 之上, 而 Ed25519 构建在 Edwards25519 之上, 并

[6] Bernstein, Daniel J., Niels Duif, Tanja Lange, Peter Schwabe, and Bo-Yin Yang. High-speed high-security signatures. Journal of Cryptographic Engineering 2, no. 2 (2012): 77-89. https://ed25519.cr.yp.to/ed25519-20110926.pdf

[7] Bernstein, Daniel J. Curve25519: new Diffie-Hellman speed records. In International Workshop on Public Key Cryptography, pp. 207-228. Springer, Berlin, Heidelberg, 2006. https://cr.yp.to/ecdh/curve25519-20060209.pdf

[8] Daniel J. Bernstein. A state-of-the-art Diffie-Hellman function. https://cr.yp.to/ecdh.html

[9] Daniel J. Bernstein. [Cfrg] 25519 naming. https://mailarchive.ietf.org/arch/msg/cfrg/-9LEdnzVrE5RORux3Oo_oDDRksU

且 Curve25519 和 Twisted-Edwards25519 是双向有理等价的. 这是因为 ECDH 协议和 EdDSA 协议计算过程中重度依赖的点群运算不同, 这是为更好的适配的上层协议而刻意选择的中层的椭圆曲线点的表示的结果. 在继续深入技术原理之前, 我们先看下 Ed25519 和 X25519 在工业界的应用情况.

2005 年就提出的 Curve25519 以及 X25519 起初并没有得到广泛的重视和应用, 然而受 Dual_EC_DRBG 事件[10] 的影响, 工业界有了很多关于 NIST 推荐的密码算法标准的质疑. 也因此, 设计规则完全透明, 没有版权保护并且效率更高的 X25519 和 Ed25519 得到重视. RFC 7748[11] 中描述了椭圆曲线 Curve25519 和 Curve448, 以及基于这两条曲线的 ECDH 协议规范: X25519 和 X448. RFC 8032[12] 则给出了 EdDSA (Edwards-curve Digital Signature Algorithm) 签名机制的规范, 并给出了基于两个椭圆曲线 Edwards25519 和 Edwards448 的 EdDSA 算法的具体实例化: Ed25519, Ed25519ph, Ed25519ctx, Ed448, Ed448ph. 更多的 RFC 规则进一步给出了在特定场景下使用 X25519 或者 Ed25519 的规范. RFC 8031[13] 中给出了在 IKEv2 使用 Curve25519 和 Curve448 进行临时密钥交换的规范. RFC 8080[14] 中给出了在 DNSSEC 中使用 Ed25519 和 Ed448 的规范. RFC 8410[15] 中为算法 Ed25519, Ed448, X25519, X448 定义了用于 PKI 体系的 X.509 证书的标识符. RFC 8420[16] 中给出了在 IKEv2 中使用 EdDSA 时 Ed25519 和 Ed448 的 ASN.1 Objects. RFC 8446[17] 中 TLS1.3 的算法套件中包含了 Ed25519 和 Ed448. RFC 8463[18] 为 DomainKeys Identified Mail (DKIM) (RFC 6376) 添加了新的签名算法 Ed25519-SHA256.

[10] Bernstein, D.J., Lange, T. and Niederhagen, R., 2016. Dual EC: A standardized back door. In The New Codebreakers (pp. 256-281). Springer, Berlin, Heidelberg. https://projectbullrun.org/dual-ec/documents/dual-ec-20150731.pdf

[11] RFC 7748. https://tools.ietf.org/html/rfc7748. https://tools.ietf.org/html/rfc7748

[12] RFC 8032. Edwards-Curve Digital Signature Algorithm (EdDSA). https://tools.ietf.org/html/rfc8032

[13] RFC 8031. Curve25519 and Curve448 for the Internet Key Exchange Protocol Version 2 (IKEv2) Key Agreement. https://tools.ietf.org/html/rfc8031

[14] RFC 8080. Edwards-Curve Digital Security Algorithm (EdDSA) for DNSSEC. https://tools.ietf.org/html/rfc8080

[15] RFC 8410. Algorithm Identifiers for Ed25519, Ed448, X25519, and X448 for Use in the Internet X.509 Public Key Infrastructure. https://tools.ietf.org/html/rfc8410

[16] RFC 8420. Using the Edwards-Curve Digital Signature Algorithm (EdDSA) in the Internet Key Exchange Protocol Version 2 (IKEv2). https://tools.ietf.org/html/rfc8420

[17] RFC 8446. The Transport Layer Security (TLS) Protocol Version 1.3. https://tools.ietf.org/html/rfc8446

[18] RFC 8463. A New Cryptographic Signature Method for DomainKeys Identified Mail (DKIM). https://tools.ietf.org/html/rfc8463

1 蒙哥马利曲线与爱德华曲线

相比 secp256k1/secp256r1 的 Short Weierstrass 形式的椭圆曲线表示 $y^2 = x^3 + ax + b$, 蒙哥马利曲线 $Y^2 = X^3 + AX^2 + X$ 与爱德华曲线 (扭曲爱德华曲线) $x^2 + y^2 = 1 + dx^2y^2$ $(-X^2 + Y^2 = 1 - dX^2Y^2)$ 较为陌生. Short Weierstrass, 蒙哥马利曲线以及爱德华曲线都可以通过符号代换与广义 Weierstrass 曲线 $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ 相互转换. X25519 和 Ed25519 的做依赖的点的运算也都可以转换成为 Weierstrass 曲线上的点运算, 然而使用特定的曲线形式, 对于高效安全的 X25519 或者 Ed25519 大有裨益. 以 twist-Edwards25519 为例, 其上的点的加法运算是完备 (Complete):

$$(x_1, y_1) + (x_2, y_2) = \left(\frac{x_1y_2 + y_1x_2}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - x_1x_2}{1 - dx_1x_2y_1y_2} \right),$$

并且单位元为点 $(0, 1)$. 习惯了 Short Weierstrass 形式下椭圆曲线点加运算的各种边界条件判断, 上面的完备的点加运算, 简洁优雅的让人有些意外. 更值得注意的是, 为了构造椭圆曲线的加法点群, 这里无需引入一个假想的无穷远点来满足群的条件. 接下来我们介绍如何从椭圆曲线的一种形式转换成另一种形式[19], 以理解蒙哥马利曲线和扭曲爱德华曲线形式的采用为 X25519 密钥交换和 Ed25519 签名机制带来的益处.

1.1 从广义 Weierstrass 约化到 Short Weierstrass

一种构造椭圆曲线的方式是将其定义为满足 Weierstrass 方程的点的结合

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1}$$

如果系数 a_1, a_2, a_3, a_4, a_6 取自域 (Field) $\mathbb{F}$, 则 E 就是定义在 $\mathbb{F}$ 上的. 注意所有的有限域都可以写成 $\mathbb{F}_p^m$ 的形式, 其中 m 为任意正整数, 而 p 是有限域 $\mathbb{F}_p^m$ 的特征 (Characteristic), 记为 $char(\mathbb{F}_p^m) = p$. 当 $p \neq 2$ 并且 $p \neq 3$ 时, 可以将广义 Weierstrass 形式简化成为 Short Weierstrass 形式, 随着推算的进行, 我们会看到排除 $p \neq 2$ 以及 $p \neq 3$ 情况的原因. 本文中, 默认都是在有限域 $\mathbb{F}_p$ 上的椭圆曲线, 别处不再复述.

方程 (1) 等号左边, 可以看成关于 y 的一元二次多项式, 这意味着总可以找到 λ 满足

$$(y + \lambda)^2 - \lambda^2 = y^2 + a_1xy + a_3y,$$

由于 λ 的值与 y 无关, 则可以通过符号代换, 用 $\gamma = y + \lambda$ 替换 y 从而将方程 (1) 中的变量 y 消除. 从上式中可以得到 $2y\lambda = a_1xy + a_3y$, 也即 $\lambda = (a_1x + a_3)/2$. 由于当

[19] 关于曲线变化的介绍主要参考了 Bassam El Khoury Seguias 的博客 Elliptic Curve Groups –Crypto Theoretical Minimum. https://delfr.com/prerequisites/elliptic-curve-cryptography/

$char(\mathbb{F}) = 2$ 时, $2 \equiv 0 \mod 2$, 也即在 $\mathbb{F}$ 中 2 的逆不存在, 因此我们要求 $char(\mathbb{F}) \neq 2$. 将 λ 带入方程 (1) 中得到:

$$\begin{aligned} & \gamma^2 - \frac{a_1 x + a_3}{2}^2 = x^3 + a_2 x^2 + a_4 x + a_6 \\ \iff & \gamma^2 = \frac{a_1 x}{2}^2 + \frac{a_3}{2}^2 + \frac{a_1 a_3 x}{2} + x^3 + a_2 x^2 + a_4 x + a_6 \\ \iff & \gamma^2 = x^3 + (a_2 + \frac{a_1^2}{4}) x^2 + (a_4 + \frac{a_1 a_3}{2}) x + (a_6 + \frac{a_3^2}{4}) \end{aligned}$$

继续用 $a_2' = a_2 + \frac{a_1^2}{4}, a_4' = a_4 + \frac{a_1 a_3}{2}, a_6' = a_6 + \frac{a_3^2}{4}$ 进行符号代换, 得到:

$$E : \gamma^2 = x^3 + a_2' x^2 + a_4' x + a_6'$$

继续化简上述方程的等式右边将 x^2 消除掉即可得到期望的 Short Weierstrass 形式. 用 $\chi = x + \nu$ 代换 x, 其中 ν 是按照特意选择的值, 在带入上面方程后可以消除掉平方项. 计算过程如下:

$$\begin{aligned} & \gamma^2 = (\chi - \nu)^3 + a_2'(\chi - \nu)^2 + a_4'(\chi - \nu) + a_6' \\ \iff & \gamma^2 = \chi^3 - 3\nu\chi^2 + 3\nu^2\chi - \nu^3 + a_2'\chi^2 - 2a_2'\nu\chi + a_2'\nu^2 + a_4'\chi - a_4'\nu + a_6' \\ \iff & \gamma^2 = \chi^3 + (a_2' - 3\nu)\chi^2 + (3\nu^2 - 2a_2'\nu + a_4')\chi - \nu^3 + a_2'\nu^2 - a_4'\nu + a_6' \end{aligned}$$

我们希望消除掉 χ^2, 则需要给 ν 添加约束 $a_2' - 3\nu = 0$, 意味着需要 $\nu = a_2'/3$. 当 $\mathbb{F}$ 的特征为 3 的时候, 在 $\mathbb{F}$ 中 3 没有逆元, 因此我们要求 $char(\mathbb{F}) \neq 3$. 由此我们得到

$$\begin{aligned} & \gamma^2 = \chi^3 + (a_2' - 3\nu)\chi^2 + (3\nu^2 - 2a_2'\nu + a_4')\chi - \nu^3 + a_2'\nu^2 - a_4'\nu + a_6' \\ \iff & \gamma^2 = \chi^3 + \left(a_4' - \frac{(a_2')^2}{3}\right)\chi + \left(a_6' + \frac{2(a_2')^3}{27} - \frac{a_2' a_4'}{3}\right) \end{aligned}$$

用符号 x 代换 χ, 用 y 代换 γ, 并记

$$a = a_4' - \frac{(a_2')^2}{3}, \ b = a_6' + \frac{2(a_2')^3}{27} - \frac{a_2' a_4'}{3}$$

就得到了有限域 $\mathbb{F}$ 上的 Short Weierstrass 形式

$$E : y^2 = x^3 + ax + b, \text{ 其中 } a, b \in \mathbb{F}, char(\mathbb{F}) \neq 2 \text{ 且 } char(\mathbb{F}) \neq 3.$$

接下来考虑在 $E : y^2 = x^3 + ax + b$ (也可记为 $E : f(x, y) = y^2 - x^3 + ax + b = 0$) 上构建椭圆曲线点群所需的条件. E 上两个点的加法运算规则依赖过两点的直线与曲线的另一个交点. 当两个点为同一个点时, 则是该点的切线与曲线的另一个交点. 由此需要在点群中的每个点都是可微的 (Differentiable). 也因此我们想要避开包含奇点 (Singularity)

的曲线. 接下来考察椭圆曲线在何种情况下会包含奇点. 椭圆曲线 $f(x,y)=0$ 上一个点 (x_P,y_P) 是奇点的充分必要条件为在该点的偏导数为 0, 也即:

$$\begin{aligned} f(x_P,y_P)=0 &\iff y_P^2-x_P^3-ax_P-b=0,\\ f_x(x_P,y_P)=0 &\iff -3x_P^2-a=0,\\ f_y(x_P,y_P)=0 &\iff 2y_P=0 \end{aligned}$$

则有 $y_P=0$, $x_P^3+ax_P+b=0$ 以及 $3x_P^2+a=0$. 也即点 (x_P,y_P) 是奇点的充分必要条件 $x_P^3+ax_P+b=0$ 以及 $3x_P^2+a=0$. 注意到 x_P 同时满足是 x^3+ax+b 以及其导数 $3x^2+a$ 的根, 则 x_P 是 x^3+ax+b 的二重根, 假设另一个根为 α, 则有

$$\begin{aligned} x^3+ax+b&=(x-x_P)^2(x-\alpha)=(x^2-2x_Px+x_P^2)(x-\alpha)\\ &=x^3-(2x_P+\alpha)x^2+(x_P^2+2x_P\alpha)x-x_P^2\alpha \end{aligned}$$

要使等式成立, 则有以下条件:

$$\begin{aligned} 2x_P+\alpha=0 &\iff \alpha=-2x_P,\\ a=x_P^2+2x_P\alpha &\iff a=x_P^2+2x_P(-2x_P)=-3x_P^2\\ b=-x_P^2\alpha &\iff b=-x_P^2(-2x_P)=2x_P^3 \end{aligned}$$

由于 $a^3=-27x_P^6$ 并且 $b^2=4x_P^6$, 合并两个条件就有

$$a=-3x_P^2, b=2x_P^3 \iff \Delta=4a^3+27b^2=0.$$

则定义在 $\mathbb{F}$ 上的椭圆曲线 $E: f(x,y)=y^2-x^3-ax-b=0, a,b\in\mathbb{F}, char(\mathbb{F})\neq 2, char(\mathbb{F})\neq 3$ 存在奇点的充分必要条件为 $\Delta=4a^3+27b^2=0$. 然而还存在一种情况, 当过一个点的切线斜率的垂直于横坐标时, 则过该点的曲线没有第三个交点, 为了处理这种情况, 引入了无穷远点 $\mathcal{O}$ 来处理该特殊情况. $\mathcal{O}$ 也在点群中扮演了单位元的角色. 由此, 定义在有限域 $\mathbb{F}_p$ 上的基于 Short Weierstrass 形式的椭圆曲线点群可记为 $(E_{a,b}^W(\mathbb{F}_p),+^W)$:

$$\begin{aligned} E_{a,b}^W(\mathbb{F}_p)=\{(x,y)\in\mathbb{F}_p^2 \mid y^2\equiv x^3+ax+b \mod p,\\ a,b\in\mathbb{F}_p, \Delta=4a^3+27b^2\neq 0,\ p\notin\{2,3\}\}\cup\{\mathcal{O}^W\}, \end{aligned}$$

而 $+^W$ 表示对应 Short Weierstrass 形式的椭圆曲线点群上的加法运算. 对于 $E_{a,b}^W(\mathbb{F}_p)$ 中的两个点 $P=(x_1,y_1), Q=(x_2,y_2)$, 则运算 $+^W$ 按照如下规则计算:

1. $-\mathcal{O}^W=\mathcal{O}^W$, $-P=(x_1,-y_1)$, $\mathcal{O}^W+^W P=P$
2. 如果 $Q=-P$, 则 $P+^W Q=\mathcal{O}^W$

3. 如果 $Q \neq -P$, 则 $P +^W Q = (x_3, y_3)$, 其中

$$\begin{cases} x_3 &= \lambda^2 - x_1 - x_2 \\ y_3 &= -y_1 + \lambda(x_1 - x_3) \end{cases}, \lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1}, & x_2 \neq x_1 \\ \frac{3x_1^2 + a}{2y_1}, & x_2 = x_1 \end{cases}.$$

1.2 蒙哥马利曲线 Curve25519

蒙哥马利曲线是 Montgomery 在 1987 年为了加速 Lenstra 的 ECM 大数分解算法而提出椭圆曲线形式[20]. 2017 年 Costello 等人总结了蒙哥马利曲线及其上的算术运算[21]. 定义在有限域 $\mathbb{F}_p, p > 2$ 上的蒙哥马利曲线点群可以表示为 $(E^M_{A,B}(\mathbb{F}_p), +^M)$:

$$E^M_{A,B}(\mathbb{F}_p) = \{(x, y) \in \mathbb{F}_p^2 \mid By^2 \equiv x^3 + Ax^2 + x \mod p, \\ p > 2, A, B \in \mathbb{F}_p, B(A^2 - 4) \neq 0\} \cup \{\mathcal{O}^M\},$$

接下来展示蒙哥马利形式椭圆曲线点群 $(E^M_{A,B}(\mathbb{F}_p), +^M)$ 与 Short Weierstrass 形式椭圆曲线点群 $(E^W_{a,b}(\mathbb{F}_p), +^W)$ 之间的转换.

$$E^W_{a,b}(\mathbb{F}_p) = \{(x, y) \in \mathbb{F}_p^2 \mid y^2 \equiv x^3 + ax + b \mod p, \\ a, b \in \mathbb{F}_p, \Delta = 4a^3 + 27b^2 \neq 0,\ p \notin \{2, 3\}\} \cup \{\mathcal{O}^W\},$$

从 $E^M_{A,B}(\mathbb{F}_p)$ 到 $E^W_{a,b}(\mathbb{F}_p)$ 的映射必须将每一个点 $(u, v) \in E^M_{A,B}(\mathbb{F}_p)$ 映射到 $E^W_{a,b}(\mathbb{F}_p)$ 中的一个点 (x, y). 先考虑非无穷远点的情况. 进行符号代换 $u = xB - \frac{A}{3}, v = yB$

$$\begin{aligned} & B\left(y^2B^2\right) = \left(xB - \frac{A}{3}\right)^3 + A\left(xB - \frac{A}{3}\right)^2 + \left(xB - \frac{A}{3}\right) \\ \iff & 27y^2B^3 = (3xB - A)^3 + 3A(3xB - A)^2 + 9(3xB - A) \\ \iff & 27B^3y^2 = 27B^3x^3 - 9A^2Bx + 27Bx + 2A^3 - 9A \end{aligned}$$

Short Weierstrass 形式要求 y^2 的系数为 1, 则等式两边除以 $27B^3$, 这也同时要求 $B \neq 0 \mod p$ (以及 B 在 $\mathbb{F}_p$ 有逆元), 则有:

$$\begin{aligned} & y^2 = x^3 - \frac{A^2}{3B^2}x + \frac{1}{B^2}x + \frac{2A^3 - 9A}{27B^3} \\ \iff & y^2 = x^3 + \frac{3 - A^2}{3B^2}x + \frac{2A^3 - 9A}{27B^3} \end{aligned}$$

[20] Montgomery, Peter L. Speeding the Pollard and elliptic curve methods of factorization. Mathematics of computation 48, no. 177 (1987): 243-264. https://www.ams.org/journals/mcom/1987-48-177/S0025-5718-1987-0866113-7/S0025--5718-1987-0866113-7.pdf

[21] Costello, Craig, and Benjamin Smith. Montgomery curves and their arithmetic. Journal of Cryptographic Engineering 8, no. 3 (2018): 227-240. https://arxiv.org/pdf/1703.01863.pdf

再进行符号代换 $a = \frac{3-A^2}{3B^2}$ 和 $b = \frac{2A^3-9A}{27B^3}$ 即可得到 Short Weierstrass 形式. 注意 Short Weierstrass 形式要求 $\Delta = 4a^3 + 27b^2 \neq 0$, 等价于要求

$$4\left(\frac{3-A^2}{3B^2}\right)^3 + 27\left(\frac{2A^3-9A}{27B^3}\right)^2 \neq 0$$
$$\iff 4\left(3-A^2\right)^3 + \left(2A^3-9A\right)^2 \iff A^2 \neq 4.$$

在上述推算过程中, 有两个约束 $B \neq 0 \mod p$ 以及 $A^2 \neq 4 \mod p$, 可以通过乘法运算合并成 $B(A^2-4) \neq 0 \mod p$, 也即蒙哥马利曲线对参数 A, B 的约束. 前述的映射中将每一个蒙哥马利形式椭圆曲线上的点映射到了 short-Weierstrass 形式椭圆曲线上的点, 但是没有一个点能够映射到 $\mathcal{O}^W$. 因此需要引入 $\mathcal{O}^M$ 并将其映射到 $\mathcal{O}^W$. 因此我们有了如下定义的从 $E^M_{A,B}(\mathbb{F}_p)$ 到 $E^W_{a,b}(\mathbb{F}_p)$ 的映射 ϕ:

$$\phi : E^M_{A,B}(\mathbb{F}_p) \to E^W_{a,b}(\mathbb{F}_p)$$
$$\phi(u,v) \to (x,y) = \left(\tfrac{3u+A}{3B}, \tfrac{v}{B}\right), \text{ if } (u,v) \neq \mathcal{O}^M$$
$$\phi(\mathcal{O}^M) \to \mathcal{O}^W$$

同样的, 也可以将 short-Weierstrass 形式可以转换为蒙哥马利形式:

$$\phi^{-1} : E^W_{a,b}(\mathbb{F}_p) \to E^M_{A,B}(\mathbb{F}_p)$$
$$\phi^{-1}(x,y) \to (u,v) = \left(xB - \tfrac{A}{3}, yB\right), \text{ if } (x,y) \neq \mathcal{O}^W$$
$$\phi^{-1}(\mathcal{O}^W) \to \mathcal{O}^M$$

另外可以注意到, 除了需要对无穷远点进行特殊处理之外, ϕ 和 ϕ^{-1} 都是 $\mathbb{F}_p$ 上的有理映射 (Rational Map), Short-Weierstrass 形式和蒙哥马利形式之间的这种双射关系称为双向有理等价 (Birational Equivalence).

取 $A = 486662, B = 1, p = 2^{255} - 19$ 就得到了蒙哥马利曲线 Curve25519:

$$y^2 = x^3 + 486662x^2 + x, \ B(A^2-4) \neq 0 \mod (2^{255}-19).$$

根据前面的讨论, 取 $a = \frac{3-A^2}{3B^2}$ 和 $b = \frac{2A^3-9A}{27B^3}$ 可以得到与之双向有理等价的 short-Weierstrass 形式的曲线方程:

$$y^2 = x^3 + ax + b, \Delta = 4a^3 + 27b^2 \neq 0 \mod (2^{255}-19),$$

根据 Listing 1, 得到 a, b, Δ 的具体值如下

$$\begin{aligned} a &= \texttt{0x2aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa984914a144} \\ b &= \texttt{0x7b425ed097b425ed097b425ed097b425ed097b425ed097b4260b5e9c7710c864} \\ \Delta &= \texttt{0x7fffffffffffffffffffffffffffffffffffffffffffffffffffffffc8db3de3cd} \end{aligned}$$

并且 $\Delta \neq 0 \mod (2^{255}-19)$ 满足约束条件. 与 secp256k1 和 secp256r1 的 a, b 取值相比较, 这样的 a, b 的值非常不利于高效的工程实现.

Listing 1: Curve25519 曲线的 short-Weierstrass 形式的曲线参数

```
sage: fp = GF(2^255 - 19)
sage: A = fp(486662)
sage: B = fp(1)
sage: hex(int(B * (A^2 - 4)))
'0x3724c21c20'
sage: a = (3 - A^2) * (3 * B^2)^(-1)
sage: b = (2 * A^3 - 9 * A) * (27 * B^3)^-1
sage: hex(int(a))
'0x2aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa984914a144L'
sage: hex(int(b))
'0x7b425ed097b425ed097b425ed097b425ed097b425ed097b4260b5e9c7710c864L'
sage: delta = 4 * a^3 + 27 * b^2
sage: hex(int(delta))
'0x7fffffffffffffffffffffffffffffffffffffffffffffffffffffffffffc8db3de3cdL'
```

X25519 是定义在蒙哥马利曲线 Curve25519 上的 ECDH 密钥交换协议, 其特殊之处是由于蒙哥马利曲线的采用, 点群的加法运算可以仅利用点的横坐标构建. 利用蒙哥马利阶梯算法 (Montgomery Ladder) 算法可以加速运算, 而利用蒙哥马利的 double-and-add 算法可以更容易实现常量时间运算. 下一篇中再具体介绍 X25519 密钥交换协议.

1.3 扭曲爱德华曲线 Edwards25519

爱德华曲线是 Harold M. Edwards 在 2007 年提出的椭圆曲线形式[22], 定义在有限域 $\mathbb{F}_p, p > 2$ 上的爱德华曲线 $E_d^{E'}$ 方程为:

$$x^2 + y^2 = 1 + dx^2y^2, d \in \mathbb{F}_p, d \notin \{0, 1\}$$

如果 $(x_1, y_1), (x_2, y_2)$ 是爱德华曲线上的两个点, 则 $E_d^{E'}$ 上的加法运算 $+^{E'}$ 规则如下:

$$(x_1, y_1) +^{E'} (x_2, y_2) = \left(\frac{x_1y_2 + y_1x_2}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - x_1x_2}{1 - dx_1x_2y_1y_2} \right),$$

其中单位元为 $(0, 1)$, 点 (x, y) 的逆元 $-(x, y) = (-x, y)$. 值得注意的是, 构建爱德华曲线上的加法点群无需引入无穷远点. Weierstrass 形式曲线上的加法运算规则与过两个点的直线和椭圆曲线的第三个交点紧密相关, 但是对爱德华曲线来说, 构建的规则更为复杂. 另外,如果 d 是 $\mathbb{F}_p$ 上的二次非剩余, 则上述点加运算规则是完备的 (Complete), 无需像 $+^W$ 一样需要特殊处理无穷远点, 两个点相同, 两个互为逆元等情况.

[22] Edwards, Harold. A normal form for elliptic curves. Bulletin of the American mathematical society 44, no. 3 (2007): 393-422. https://www.ams.org/journals/bull/2007-44-03/S0273-0979-07-01153-6/S0273-0979-07-01153-6.pdf

2008 年 Bernstein 等人指出有限域上只有一小部分椭圆曲线能够表示为爱德华曲线, 并进一步提出了更为广义的扭曲爱德华曲线 (Twisted Edwards Curves)[23]. 有限域 $\mathbb{F}_p, p > 2$ 上的蒙哥马利曲线集合与扭曲爱德华曲线集合相同, 也即这两类曲线之间是双向有理等价的. 有限域 $\mathbb{F}_p, p > 2$ 上的扭曲爱德华曲线 $E^E_{a,d}$ 的方程为:

$$ax^2 + y^2 = 1 + dx^2y^2, a, d \in \mathbb{F}_p, ad(a-d) \neq 0.$$

扭曲爱德华曲线中的参数 $a = 1$ 时即为爱德华曲线. 如果 $(x_1, y_1), (x_2, y_2)$ 是扭曲爱德华曲线上的两个点, 则 $E^E_{a,d}$ 上的加法运算 $+^E$ 规则如下:

$$(x_1, y_1) +^E (x_2, y_2) = \left(\frac{x_1y_2 + y_1x_2}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - ax_1x_2}{1 - dx_1x_2y_1y_2} \right),$$

其中, 单位元为 $(0, 1)$, 点 (x, y) 的逆元 $-(x, y) = (-x, y)$, 与爱德华曲线类似这里无需引入无穷远点. 如果 a 是 $\mathbb{F}_p$ 上的二次剩余而 d 不是, 则上述扭曲爱德华曲线上点的加法运算 $+^E$ 是完备的. 用 $(E^E_{a,d}(\mathbb{F}_p), +^E)$ 表示基于扭曲爱德华曲线的加法点群, 其中

$$\begin{aligned} E^E_{a,d}(\mathbb{F}_p) = \{(x, y) \in \mathbb{F}_p^2 \mid \quad & ax^2 + y^2 \equiv 1 + dx^2y^2, a, d \in \mathbb{F}_p, ad(a-d) \neq 0 \\ & a \text{ 为二次剩余}, d \text{ 为二次非剩余}\} \end{aligned}$$

接下来我们展示扭曲爱德华曲线与蒙哥马利曲线之间的双向有理等价关系. 首先展示如何通过符号变换与代换将扭曲爱德华曲线映射到蒙哥马利曲线. 我们沿用之前蒙哥马利曲线表示 $E^M_{A,B}(\mathbb{F}_p)$, 并首先展示如何将 $E^E_{a,d}(\mathbb{F}_p)$ 上的点 (x, y) 映射到 $E^M_{A,B}(\mathbb{F}_p)$ 上的点 (u, v). 首先令 $x = u/v, y = (u-1)/(u+1)$ 并进行符号带换, 这种代换会将几个特殊的点排除在外, 这包括 $E^M_{A,B}(\mathbb{F}_p)$ 中的无穷远点 $\mathcal{O}^M$, $v \equiv 0 \mod p$ 的点, 以及 $u \equiv -1 \mod p$ 点. 根据 $E^E_{a,d}(\mathbb{F}_p)$ 的方程, $v \equiv 0 \mod p$ 意味着:

$$u(u^2 + Au + 1) = Bv^2 = 0 \iff u = 0 \text{ 或者 } u^2 + Au + 1 = 0.$$

为了将尽可能多的点纳入映射当中, 要求 $u^2 + Au + 1 \equiv 0 \mod p$ 无解. 这就等价于要求判别式 $A^2 - 4$ 是 $\mathbb{F}_p$ 中的二次非剩余. 在这个约束条件下, 由于 $v \equiv 0 \mod p$ 被排除点仅有 $(u, v) = (0, 0)$. 接下来考虑 $u \equiv -1 \mod p$ 的点, 根据蒙哥马利曲线方程 $B \neq 0 \mod p$ 的前提条件:

$$Bv^2 = A - 2 \iff v^2 = (A-2)/B,$$

同样为了减少被排除的点的个数,要求 $(A-2)/B$ 是 $\mathbb{F}_p$ 中的二次非剩余. 在这个条件下, 没有点会因为 $u \equiv -1 \mod p$ 被排除. 在前面的论述中, 被排除的仅有 $E^M_{A,B}(\mathbb{F}_p)$ 中的无

[23] Bernstein, Daniel J., Peter Birkner, Marc Joye, Tanja Lange, and Christiane Peters. Twisted edwards curves. In International Conference on Cryptology in Africa, pp. 389-405. Springer, Berlin, Heidelberg, 2008. https://cr.yp.to/newelliptic/twisted-20080313.pdf

穷远点 $\mathcal{O}^M$ 和 $(0,0)$. $E^M_{A,B}(\mathbb{F}_p)$ 中剩下点都有 $u \neq 0 \mod p$, 而 $x = u/v$ 意味着也同时排除了 $E^E_{a,d}(\mathbb{F}_p)$ 中 $x = 0$ 的点, 这包括 $(0,1)$ 和 $(0,-1)$.

排除 $E^M_{A,B}(\mathbb{F}_p)$ 中的 $\mathcal{O}^M$, $(0,0)$ 和 $E^E_{a,d}(\mathbb{F}_p)$ 中的 $(0,1)$, $(0,-1)$ 之后, 将上述符号代换带入 $E^E_{a,d}(\mathbb{F}_p)$ 方程:

$$ax^2 + y^2 = 1 + dx^2y^2 \quad \Longleftrightarrow \quad a\left(\frac{u}{v}\right)^2 + \left(\frac{u-1}{u+1}\right)^2 = 1 + d\left(\frac{u}{v}\right)^2\left(\frac{u-1}{u+1}\right)^2$$

由于 $v \neq 0 \mod p$ 并且 $u \neq -1 \mod p$, 等式两边同时乘以 $v^2(u+1)^2$ 可以得到:

$$\begin{aligned} & au^2(u+1)^2 + v^2(u-1)^2 \equiv v^2(u+1)^2 + du^2(u-1)^2 \mod p \\ \Longleftrightarrow\ & (a-d)u^4 + (2a-2d)u^3 + (a-d)u^2 - 4v^2u \equiv 0 \mod p \end{aligned}$$

由于 $u \neq 0 \mod p$, 方程两侧同时乘以 u^{-1}, 得到

$$4v^2 \equiv (a-d)u^3 + (2a-2d)u^2 + (a-d)u \mod p$$

注意到 $E^M_{A,B}(\mathbb{F}_p)$ 的方程中 x^3 的系数为 1, 上式中为了使 u^3 的系数为 1, 我们要求 $a-d \neq 0 \mod p$, 则方程两侧同时乘以 $(a-d)^{-1}$, 得到

$$\left(\frac{4}{a-d}\right)v^2 \equiv u^3 + \left(\frac{2a+2d}{a-d}\right)u^2 + u \mod p$$

令 $B = 4/(a-d), A = (2a+2d)/(a-d)$ 并用 x 代换 u, y 代换 v 即可得到蒙哥马利形式曲线方程. 注意到蒙哥马利形式方程要求 $B(A^2-4) \neq 0 \mod p$. 由于 $B = 4/(a-d)$, 显然有 $B \neq 0 \mod p$. 从而仅需要

$$\begin{aligned} A^2 - 4 \neq 0 \mod p \quad & \Longleftrightarrow \left(\frac{2a+2d}{a-d}\right)^2 - 4 \neq 0 \mod p \\ & \Longleftrightarrow (a+d)^2 \neq (a-d)^2 \mod p \\ & \Longleftrightarrow ad \neq 0 \mod p \end{aligned}$$

至此完成了从 $(x,y) \in E^E_{a,d}(\mathbb{F}_p)\backslash\{(0,1),(0,-1)\}$ 到 $(u,v) \in E^M_{A,B}(\mathbb{F}_p)\backslash\{\mathcal{O}^M,(0,0)\}$ 映射:

$$(u,v) = \psi(x,y) = \left(\frac{1+y}{1-y}, \frac{1+y}{x(1-y)}\right).$$

这个过程中引入了几个限制条件: 要求 $(A-2)/B$ 是 $\mathbb{F}_p$ 中的二次非剩余等价于要求

$$\left(\frac{2a+2d}{a-d} - 2\right) / \frac{4}{a-d} = d$$

为 $\mathbb{F}_p$ 上的二次非剩余. 要求 A^2-4 是 $\mathbb{F}_p$ 中的二次非剩余等价于要求

$$\left(\frac{2a+2d}{a-d}\right)^2 - 4 = \frac{16ad}{(a-d)^2}$$

为 $\mathbb{F}_p$ 上的二次非剩余. 由于 16 和 $(a-d)^2$ 显然是 $\mathbb{F}_p$ 中的二次剩余, 并且两个二次剩余的乘积仍为二次剩余, 而两个二次非剩余的乘积也有可能是二次剩余, 因此为了确保 $(16ad)/(a-d)^2$ 为二次非剩余, 要求 a 和 d 中一个为二次剩余, 一个为二次非剩余. 由于 d 已经被限定为二次非剩余, 则只能要求 a 为二次剩余. 另外的两个约束条件 $a \neq d \mod p$ 和 $ad \neq 0 \mod p$ 可以写为 $ad(a-d) \neq 0 \mod p$. 这与 $E^E_{a,d}(\mathbb{F}_p)$ 中关于参数 a 和 d 的约束一致, 或者说 $E^E_{a,d}(\mathbb{F}_p)$ 关于 a 和 d 的约束保证了我们可以计算映射 ψ. 最后 $E^E_{a,d}(\mathbb{F}_p)$ 和 $E^M_{A,B}(\mathbb{F}_p)$ 中都恰好有两个元素没被映射 ψ 处理, 为 $(0,-1),(0,1)$ 定义特殊的 ψ 映射到 $(0,0),\mathcal{O}^M$ 也就完成了从 $E^E_{a,d}(\mathbb{F}_p)$ 到 $E^M_{A,B}(\mathbb{F}_p)$ 的完整映射:

$$\begin{gathered}\psi : E^E_{a,d}(\mathbb{F}_p) \to E^M_{A,B}(\mathbb{F}_p) \\ (x,y) \to (u,v) \equiv \psi(x,y) = \left(\tfrac{1+y}{1-y}, \tfrac{1+y}{x(1-y)}\right), (x,y) \notin \{(0,1),(0,-1)\} \\ (0,-1) \to \psi(0,-1) = (0,0) \\ (0,1) \to \psi(0,1) = \mathcal{O}^M\end{gathered}$$

考察逆映射 $(u,v) \in E^M_{A,B}(\mathbb{F}_p) \to (x,y) \in E^E_{a,d}(\mathbb{F}_p)$, 下面再次给出 $E^M_{A,B}(\mathbb{F}_p)$, $E^E_{a,d}(\mathbb{F}_p)$ 的定义:

$$\begin{aligned}E^M_{A,B}(\mathbb{F}_p) = \{(x,y) \in \mathbb{F}_p^2 \mid & By^2 \equiv x^3 + Ax^2 + x \mod p, \\ & p > 2, A, B \in \mathbb{F}_p, B(A^2-4) \neq 0\} \cup \{\mathcal{O}^M\},\end{aligned}$$

$$\begin{aligned}E^E_{a,d}(\mathbb{F}_p) = \{(x,y) \in \mathbb{F}_p^2 \mid & \quad ax^2 + y^2 \equiv 1 + dx^2y^2, a, d \in \mathbb{F}_p, ad(a-d) \neq 0, \\ & \quad a \text{ 为二次剩余}, d \text{ 为二次非剩余}\}.\end{aligned}$$

根据上述的映射 ψ, 当 $x \neq 0 \mod p, y \neq 1 \mod p$ 时, 用 $u = \frac{1+y}{1-y}, v = \frac{1+y}{x(1-y)}$ 进行代换得到:

$$B\left(\tfrac{1+y}{x(1-y)}\right)^2 \equiv \left(\tfrac{1+y}{1-y}\right)^3 + A\left(\tfrac{1+y}{1-y}\right)^2 + \left(\tfrac{1+y}{1-y}\right) \mod p,$$

注意 $x \equiv 0 \mod p$, 有 $y^2 \equiv 1 \mod p$, 这同时剔除了 $E^E_{a,d}(\mathbb{F}_p)$ 中的两个点 $(0,1),(0,-1)$. 当 $y = \pm 1 \mod p$ 时, 有 $ax^2 \equiv dx^2 \mod p$, 只要 $a \neq d$, 就不会剔除外的点. 根据 x, y 取值的约束, 则有

$$u = \frac{1+y}{1-y} = -1 - \frac{2}{y-1}, v = \frac{1+y}{x(1-y)}, x \neq 0, y \neq \pm 1 \implies u \neq 0, u \neq -1, v \neq 0$$

$u = 0$ 时, 由于 $B \neq 0 \mod p$, 根据曲线方程有

$$Bv^2 = 0 \mod p \iff v \equiv 0 \mod p,$$

也即对于 u 的约束剔除了 $E^M_{A,B}(\mathbb{F}_p)$ 中的点 $(0,0)$. $u = -1$ 时, 根据曲线方程有

$$By^2 = A - 2 \iff y^2 = \frac{A-2}{B},$$

要求 $(A-2)/B$ 为二次非剩余, 则上述方程无解, 也即 $u \neq -1$ 的条件没有剔除额外的点. $v = 0 \mod p$ 时, 根据曲线方程有

$$u(u^2 + Au + 1) \equiv 0 \mod p \iff u \equiv 0 \mod p \text{ 或 } u^2 + Au + 1 \equiv 0 \mod p,$$

为了尽可能少的剔除点,要求 $A^2 - 4$ 是 $\mathbb{F}_p$ 中的二次非剩余, 也即 $u^2 + Au + 1$ 在 $\mathbb{F}_p$ 中没有根, 则上述情形中只能是 $u \equiv 0 \mod p$, 则关于 v 的约束条件没有剔除额外的点. 上面代换隐式剔除的另一个点是 $E^M_{A,B}(\mathbb{F}_p)$ 中的无穷远点 $\mathcal{O}^M$. 上述方程简化之后得到:

$$\begin{aligned} B(1+y)^2(1-y) &= x^2(1+y)\left((1+y)^2 + A(1+y)(1-y) + (1-y)^2\right) \\ \iff B - By^2 &= (2-A)x^2y^2 + (2+A)x^2 \\ \iff (2+A)x^2 + By^2 &= B - (2-A)x^2y^2 \\ \iff \tfrac{2+A}{B}x^2 + y^2 &= 1 - \tfrac{2-A}{B}x^2y^2 \end{aligned}$$

用 $a = (A+2)/B, d = (A-2)/B$ 进行代换, 可以得到扭曲爱德华曲线形式. 由于 $A^2-4 \neq 0 \mod p$, 则有

$$ad = \left(\frac{A+2}{B}\right)\left(\frac{A-2}{B}\right) = \frac{A^2-4}{B^2} \neq 0 \mod p.$$

并且

$$a - d = \frac{4}{B} \neq 0 \mod p.$$

因此, $E^E_{a,d}(\mathbb{F}_p)$ 中的条件 $ad(a-d) \neq 0$ 得到满足. 接下来考察 $E^E_{a,d}(\mathbb{F}_p)$ 对于 a 为二次剩余和 d 为二次非剩余的约束. 根据之前的讨论, 可知两个约束条件 $(A-2)/B$ 为二次非剩余和 A^2-4 是的二次非剩余, 转换成 a, d 的约束记为要求 a 为二次剩余,d 为二次非剩余, 与 $E^E_{a,d}(\mathbb{F}_p)$ 中的要求一致. 由此, 得到了以下映射:

$$\begin{aligned} \psi^{-1} &: E^M_{A,B}(\mathbb{F}_p) \to E^E_{a,d}(\mathbb{F}_p) \\ (u,v) \to (x,y) \equiv \psi^{-1}(u,v) &= \left(\tfrac{u}{v}, \tfrac{u-1}{u+1}\right), \text{ if } (u,v) \neq (0,0) \\ (0,0) \to \psi^{-1}(0,0) &= (0,-1) \\ \mathcal{O}^M \to \psi^{-1}(\mathcal{O}^M) &= (0,1) \end{aligned}$$

Edwards25519 是定义在有限域 $\mathbb{F}_p$, $p = 2^{255} - 19$ 上选用参数 $a = -1$, $d = \frac{121665}{121666}$ 的扭曲爱德华曲线: $-x^2 + y^2 \equiv 1 - (121665/121666)x^2y^2 \mod p$. Ed25519 签名机制是定义在 Edwards25519 曲线上的变种 Schnorr 签名. RFC 8032 中总结到, EdDSA 签名机制的优势在于: 在多种计算平台上都能达到较高的性能; 签名过程中不需要唯一的随机数, 能够避免随机数引发的安全问题; 对于侧信道攻击等具有更好的免疫效果; 公钥和签名值都较小 (Ed25519 公钥为 32 个字节, 签名值为 64 字节); 计算公式是完备 (Complete), 无需对不相信的点执行点的验证操作; EdDSA 能抵抗碰撞, 底层哈希函数的碰撞不会破坏 EdDSA 签名机制 (PureEdDSA). 下一篇中逐条解释每一项特性.