Notes on ECC

ZHOU, Junda

Preliminary

Group

A set $\mathbb{G}$, with a binary operation, $\circ$, is a ***group*** if the following conditions holds:

Closure

Existence of Identity e

Existence of Inverse a^{-1}

Associativity

A group is abelian if the following holds:

Commutativity

For finite group $\mathbb{G}$, denote the ***order*** of the group by $|\mathbb{G}|$. $\forall g \in \mathbb{G}$, $g^{|\mathbb{G}|} = e$.

The ***order*** of $g \in \mathbb{G}$ is the smallest positive integer i s.t. $g^i = e$. Furthermore, $i \Big| |G|$.

If $\mathbb{G}$ is a group of prime order p, then $\mathbb{G}$ is cyclic. Furthermore, all elements of $\mathbb{G}$ except the identity are generators of $\mathbb{G}$.

$\mathbb{Z}_N \overset{def}{=} \{0, \ldots, N-1\}$ is an abelian group under addition modulo N.

$\mathbb{Z}_N^* \overset{def}{=} \{b \in \{1, \ldots, N-1\} | gcd(b, N) = 1\}$ is an abelian group under multiplication modulo N.

For prime p, $\mathbb{Z}_p^*$ is a cyclic group of order $p-1$.

For prime p, $y \in \mathbb{Z}_p^*$ is a ***quadratic residue*** if $\exists x \in \mathbb{Z}_p^*$ s.t. $x^2 = y \, mod \, p$. Otherwise, y is a ***quadratic non-residue***.

For prime $p > 2$, half of the elements in $\mathbb{Z}_p^*$ are quadratic residue and every quadratic residue has exactly two square roots.

For prime $p > 5$, $x, y \in \mathbb{Z}_p^*$, the product xy is

$$\begin{cases} \text{quadratic residue} & \text{both x, y are quadratic residues or quadratic non-residues} \\ \text{quadratic non-residue} & \text{otherwise} \end{cases}.$$

Elliptic curve over $\mathbb{Z}_p$

Weierstrass Curve

Equation $E : y^2 = x^3 + ax + b \; mod \; p$ where $a, b \in \mathbb{Z}_p$, prime $p \geq 5, 4a^3 + 27b^2 \neq 0 \; mod \; p$.

$4a^3 + 27b^2 \neq 0 \; mod \; p$ ensures that E has no repeated roots.

$E(\mathbb{Z}_p) \overset{def}{=} \{(x, y) | x, y \in \mathbb{Z}_p, y^2 = x^3 + ax + b \; mod \; p\} \cup \{\mathcal{O}\}$, where $\mathcal{O}$ is the ***point of infinity***.

$\mathcal{O}$ is the identity.

Let $P_1, P_2 \neq \mathcal{O}$ be points in E, with $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$.

1. If $x_1 \neq x_2$, then $P_1 + P_2 = (x_3, y_3)$with
$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \; mod \; p \\ x &= m^2 - x_1 - x_2 \; mod \; p \\ y &= (2x_1 + x_2)m - m^3 - y_1 \; mod \; p \end{aligned}$$
2. If $x_1 = x_2$ and $y_1 \neq y_2$, then $P_1 + P_2 = \mathcal{O}$
3. If $P_1 = P_2$ and $y_1 = 0$, then $P_1 + P_2 = 2P_1 = \mathcal{O}$
4. If $P_1 = P_2$ and $y_1 \neq 0$, then $P_1 + P_2 = 2P_1 = (x_3, y_3)$with
$$\begin{aligned} m &= \frac{3x_1^2 + a}{2y_1} \; mod \; p \\ x &= m^2 - 2x_1 \; mod \; p \\ y &= 3x_1 m - m^3 - y_1 \; mod \; p \end{aligned}$$

$E(\mathbb{Z}_p)$ with the above point addition forms an abelian group, called the ***elliptic curve group*** of E.

(Hasse bound) $p + 1 - 2\sqrt{p} < |E(\mathbb{Z}_p)| < p + 1 + 2\sqrt{p}$.

Choosing curve

n shouldn't be product of small numbers; otherwise solving ECDLP is becomes much easier.

Treating doubling different from the general one may leak critical information if addition and doubling are distinguishable.

Clear origin of a, b; otherwise that might be some yet-unknown weakness.

Weak curves

Anomalous curve: $|E(\mathbb{Z}_p)| = p$

Super-singular curve: $|E(\mathbb{Z}_p)| = p + 1$

$|E(\mathbb{Z}_p)|$ divides $p^k - 1$ for "small" k

Common curves

NIST curves

drawback: the origin of b is unknown to public

Curve25519

$y^2 = x^3 + 486662x^2 + x, p = 2^{255} - 19$, base point $x = 9$

shorter key, faster computation, unified addition law

Montgomery Curve

Equation $M : By^2 = x^3 + Ax^2 + x \; mod \; p$ where $A, B \in \mathbb{Z}_p$, prime $p \geq 5, B(A^2 - 4) \neq 0$.

$M(\mathbb{Z}_p) \overset{def}{=} \{(x, y) | x, y \in \mathbb{Z}_p, By^2 = x^3 + Ax^2 + x \; mod \; p\} \cup \{\mathcal{O}\}$, where $\mathcal{O}$ is the ***point of infinity***.

Let $P_1, P_2 \neq \mathcal{O}$ be points in M, with $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$.

1. If $x_1 \neq x_2$, then $P_1 + P_2 = (x_3, y_3)$ with

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \; mod \; p \\ x &= Bm^2 - A - x_1 - x_2 \; mod \; p \\ y &= (2x_1 + x_2 + A)m - Bm^3 - y_1 \; mod \; p \end{aligned}$$

2. If $x_1 = x_2$ and $y_1 \neq y_2$, then $P_1 + P_2 = \mathcal{O}$
3. If $P_1 = P_2$ and $y_1 = 0$, then $P_1 + P_2 = 2P_1 = \mathcal{O}$
4. If $P_1 = P_2$ and $y_1 \neq 0$, then $P_1 + P_2 = 2P_1 = (x_3, y_3)$ with

$$\begin{aligned} m &= \frac{3x_1^2 + 2Ax_1 + 1}{2By_1} \; mod \; p \\ x &= Bm^2 - A - 2x_1 \; mod \; p \\ y &= (3x_1 + A)m - Bm^3 - y_1 \; mod \; p \end{aligned}$$

All Montgomery curves can be transferred to Weierstrass curve, the converse is not true.

Double and Add

Input: $P = (x_P, y_p), d = (d_0, \ldots, d_{n-1})$

Output: $R = dP$

```
Let $R = \mathcal{O}$.
for $j = n - 1$ down to $0$ do
    $R = 2R$
    if $d_j = 1$ then $R = R + P$
return $R$
```

x-coordinate Ladder

Input: $P = (x_P, *), d = (d_0, \ldots, d_{n-1})$

Output: $R = (x_R, *) = dP$

```
Let $R_0 = \mathcal{O}, R_1 = P$.
for $j = n - 1$ down to $0$ do
```

```
    $R_{1-d_j} = R_{1-d_j} + R_{d_j}, R_{d_j} = 2R_{d_j}$
  return $R_0$
```

Only two point operations are used:

doubling, and

adding two points with a difference of P.

For Weierstrass Curve [2]

1. If $x_1 \neq x_2$, then $P_1 + P_2 = (x_3, y_3)$ with $x_3(x_1 - x_2)^2 = -2y_1y_2 + 2b + (a + x_1x_2)(x_1 + x_2)$.
   Let $P_1 - P_2 = (x_4, y_4)$, then $x_4(x_1 - x_2)^2 = 2y_1y_2 + 2b + (a + x_1x_2)(x_1 + x_2)$.
   Combine these two equations, we have $x_3x_4(x_1 - x_2)^2 = [-4b(x_1 + x_2) + (x_1x_2 - a)^2](x_1 - x_2)^2$.
2. If $P_1 = P_2$ and $y_1 \neq 0$, then $P_1 + P_2 = 2P_1 = (x_3, y_3)$ with $4x_3(x_1^3 + ax_1 + b) = (x_1 - a)^2 - 8bx_1$.

Constant time, good side channel attacks but slow.

For Montgomery Curve [3]

1. If $x_1 \neq x_2$, then $P_1 + P_2 = (x_3, y_3)$ with $x_3(x_1 - x_2)^2 = \frac{B(x_2y_1 - x_1y_2)^2}{x_1x_2}$.
   Let $P_1 - P_2 = (x_4, y_4)$, then $x_4(x_1 - x_2)^2 = \frac{B(x_2y_1 + x_1y_2)^2}{x_1x_2}$.
   Combine these two equations, we have $x_3x_4(x_1 - x_2)^2 = (x_1x_2 - 1)^2$.
2. If $P_1 = P_2$ and $y_1 \neq 0$, then $P_1 + P_2 = 2P_1 = (x_3, y_3)$ with $4x_1x_3(x_1 + Ax_1 + 1) = (x_1 - 1)^2$.

Point representation

Projective Coordinate, Jacobian Coordinate

ECDH

Common

Elliptic curve E

Prime-order n subgroup of $E(\mathbb{Z}_p)$ for prime p

Base point (or generator point) G

Alice private key k_a, public key $P_a = k_aG$

Bob private key k_b, public key $P_b = k_bG$

The ***shared secret*** $P = k_aP_b = k_bP_a$. Then P is used to derive symmetric keys for authentication and encryption.

Possible attack

If the received point hasn't been validated to be a point on the legitimate curve E with the correct order n, the following attacks might happen.

Small subgroup attack

Target

Curve that the cofactor $h = \frac{|E(\mathbb{Z}_p)|}{n}$ is large, and

Users don't check the received point Q has the correct order.

How the attack works

The attacker send a point Q on the legitimate curve but with smaller order, usually equals to the cofactor h.

Shared secret mQ, also with smaller order, is used for key derivation.

After enumerating all the possibilities and verifying the data, $m\ mod\ order(Q)$ is revealed to the attacker.

Invalid curve attack

Target

Weierstrass Curve, Double and Add implementation, and

Users don't check whether the received point Q is on the legitimate curve.

How the attack works:

The attacker send a point Q on another curve $y^2 = x^3 + ax^2 + b'$ and with smaller order.

Similarly, $m\ mod\ order(Q)$ is revealed to the attacker.

Choosing Q for different prime orders, by Chinese Remainder Theorem, m can be revealed.

This is feasible as the parameter b is not involved the point addition in Weierstrass Curve.

Twist curve attack

Preliminary

Weierstrass curve $E : y^2 = x^3 + ax + b\ mod\ p$ and $E' : uy^2 = x^3 + ax + b\ mod\ p$.

If u is a quadratic residue, than $E(\mathbb{Z}_p), E'(\mathbb{Z}_p)$ are isomorphic and they consider to have same level of security. We call $E'(\mathbb{Z}_p)$ a ***quadratic twist***.

If u is a quadratic non-residue, than $|E(\mathbb{Z}_p)| + |E'(\mathbb{Z}_p)| = 2p + 2$. For x s.t. $R.H.S \neq 0$, it either corresponds to two points on E, or two points on E'. We call $E'(\mathbb{Z}_p)$ a ***nontrivial quadratic twist***. A nontrivial quadratic twist might ease the difficulty of ECDLP.

Target

Single-coordinate ladder implementation, and

There exists weak twists curve, and

Users don't check the received x-coordinate corresponds to points on the legitimate curve.

Restraint

Requires the use of ephemeral-static key exchange to have enough time to solve ECDLP.

How this attack works

Attackers send an x-coordinate that lies on a weak twist curve. By solving ECDLP on the weak twist the attacker can learn the user's secret key.

This attack is feasible as x-coordinate ladder is valid for both point on the original curve and its twists.

A remark was given here [4]. (It was originally presented in projective coordinate, may not correct under affined coordinate.)

> x-coordinate ladder is valid for points $(x, y) \in E(\mathbb{Z}_p)$ with $y \in \mathbb{Z}_{p^2}$, instead of simply $y \in \mathbb{Z}_p$.
>
> $$\begin{aligned} S &= \{\mathcal{O}\} \cup \{(x, y) \in E(\mathbb{Z}_{p^2}), x \in \mathbb{Z}_p, y \in \mathbb{Z}_{p^2}\} \\ &= \{\mathcal{O}\} \cup \{(x, 0) \in E(\mathbb{Z}_{p^2}), x \in \mathbb{Z}_p\} \cup \{(x, y) \in E(\mathbb{Z}_{p^2}), x \in \mathbb{Z}_p, y \in \mathbb{Z}_p^*\} \cup \{(x, y) \in E(\mathbb{Z}_{p^2}), x \in \mathbb{Z}_p, y \in \mathbb{Z}_{p^2} \setminus \mathbb{Z}_p\} \\ &= \{\mathcal{O}\} \cup S_0 \cup S_1 \cup S_2 \end{aligned}$$
>
> We have $E(\mathbb{Z}_p) = \{\mathcal{O}\} \cup S_0 \cup S_1$, and $E'(\mathbb{Z}_p) = \{\mathcal{O}\} \cup S_0 \cup S_2$.

Question: Why it holds? Any relations between quadratic (non-)/residues in $\mathbb{Z}_p$ and $\mathbb{Z}_{p^2}$?

Curve downgrade attack

Target

Connections between clients and servers, and

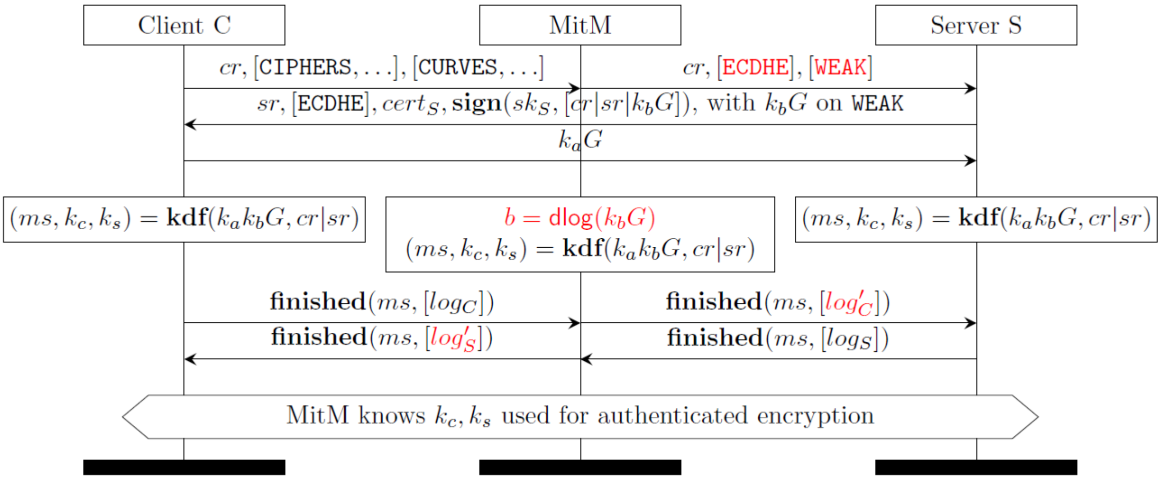
The negotiation on elliptic curve group is not authenticated and verified.

Restraints:

Requires the use of ephemeral-static key exchange to have enough time to solve ECDLP.

How this attack works

Man in the middle replaces the unprotected messages from client to the server, forcing them to work on a weak elliptic curve. By solving ECDLP on the weak curve, the attacker can learn the secret key. The attacker can then learn the shared secret. See below an example of this attack in early version of TLS.


Client C
MitM
Server S
cr, [CIPHERS, ...], [CURVES, ...]
cr, [ECDHE], [WEAK]
sr, [ECDHE], $cert_S$, **sign**(sk_S, [$cr|sr|k_bG$]), with k_bG on WEAK
k_aG
$(ms, k_c, k_s) =$ **kdf**$(k_ak_bG, cr|sr)$
$b =$ dlog(k_bG)
$(ms, k_c, k_s) =$ **kdf**$(k_ak_bG, cr|sr)$
$(ms, k_c, k_s) =$ **kdf**$(k_ak_bG, cr|sr)$
finished$(ms, [log_C])$
finished$(ms, [log'_C])$
finished$(ms, [log'_S])$
finished$(ms, [log_S])$
MitM knows k_c, k_s used for authenticated encryption

CurveSwap[1] allows a man-in-the-middle to trigger a downgrade attack to force a connection to use the weakest elliptic curve that both parties support.

CurveSwap is a vulnerability in the TLS protocol itself, and affects TLS 1.0, 1.1 and 1.2.

Partial solution

In the above versions of TLS, the **premaster secret** is computed from the **entire transcript of the handshake**, so in the case of an attempted parameter downgrade attack of this form, the attacker would be forced to man in the middle the entire connection instead of merely downgrading it.

Total solution

In TLS 1.3, the server sends a **certificate verify message** that includes a **signature of the entire handshake transcript hash**. In order to downgrade the connection, the attacker would need to forge this signature.

ECDSA

Common

Elliptic curve $E : y^2 = x^3 + ax + b \, mod \, p$

Prime-order n subgroup of $E(\mathbb{Z}_p)$ for prime p

Base point (or generator point) G

Signer

Private key d, Public key dG

Hash value of the message after modulo n, h

Pick k from $1, \ldots, n-1$ randomly, compute $kG = (x, y)$.

Let $r = x \, mod \, p, \;\; s = \frac{h+rd}{k} \, mod \, p$.

The signature is (r, s).

Verifier

Compute $Q = hs^{-1}G + rs^{-1}P = (x', y')$.

Verify $r \overset{?}{=} x' \, mod \, p$.

Possible attack

Bad randomness in k

If k is predictable, the private key d can be calculated from s.

If k is used repeatedly, r is also repeated. k can be calculated from $s_1,\ s_2$, then d is also revealed.

Side channel attack (**SDA**) on the computation time consumption.

Reference

[1] L. Valenta, N. Sullivan, A. Sanso and N. Heninger, "In Search of CurveSwap: Measuring Elliptic Curve Implementations in the Wild," 2018 IEEE European Symposium on Security and Privacy (EuroS&P), 2018, pp. 384-398, doi: 10.1109/EuroSP.2018.00034.

[2] É. Brier and M. Joye, “Weierstraß Elliptic Curves and Side-Channel Attacks,” in *Public Key Cryptography*, 2002, pp. 335–345, doi: 10.1007/3-540-45664-3_24.

[3] P. L. Montgomery, “Speeding the Pollard and Elliptic Curve Methods of Factorization,” *Mathematics of Computation*, vol. 48, no. 177, pp. 243–243, Jan. 1987, doi: 10.1090/s0025-5718-1987-0866113-7.

[4] P. Fouque, R. Lercier, D. Réal and F. Valette, "Fault Attack on Elliptic Curve Montgomery Ladder Implementation," 2008 5th Workshop on Fault Diagnosis and Tolerance in Cryptography, 2008, pp. 92-98, doi: 10.1109/FDTC.2008.15.