MATEMATISKA INSTITUTIONEN
STOCKHOLMS UNIVERSITET
Avd. Matematik
Examinator: Jonas Bergström
Lecturer: Tuomas Tajakka

Tentamensskrivning i
Mathematics of Cryptography, 7,5 hp
14 March 2023
14.00-19.00

There are 2 pages and 8 problems with total score of 85 points. The score from the exam is added to the score from the homework assignments. Grades are then given by the following intervals:

A: 100-92 p   B: 91-84 p   C: 83-76p   D: 75-68 p   E: 67-60 p

Remember to justify your answers carefully. No calculators or computers may be used.

1. Define the following terms:

a) symmetric key cryptosystem 2 p

Solution: a *symmetric key cryptosystem* is a cryptosystem in which both parties know a secret key k which is used for both encryption and decryption.

b) chosen plaintext attack 2 p

Solution: a *chosen plaintext attack* is an attack on a cryptosystem in which the adversary chooses messages $m_1, \ldots, m_n$, obtains the encrypted messages $e(m_1), \ldots, e(m_n)$, and from this tries to deduce a way to decrypt a general cyphertext.

c) cryptographic hash function 2 p

Solution: a *cryptographic hash function* is a function which sends a document to a binary string. It is typically required to be fast to compute, hard to invert, and it should be hard to find two documents with the same hash.

d) encoding scheme 2 p

Solution: an *encoding scheme* is a method of converting one sort of data into another sort of data, e.g. converting text to numbers.

e) big-$\mathcal{O}$ notation 3 p

Solution: let $f, g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be functions. Then we say $f = \mathcal{O}(g)$ if there exist $C, N \in \mathbb{R}$ such that $f(x) \leq Cg(x)$ for all $x > N$.

2. a) State Fermat's little theorem. 2 p

Solution: Let p be a prime number and $a \in \mathbb{N}$. Then

$$a^{p-1} \equiv \begin{cases} 1 \pmod{p} & p \nmid a; \\ 0 \pmod{p} & p \mid a. \end{cases}$$

b) Use Fermat's little theorem and the fast powering algorithm to find the multiplicative inverse of 5 in $\mathbb{F}_{13}$. Show all steps. 4 p

Solution: We apply Fermat's little theorem with $a = 5$ and $p = 13$ to see that $5^{-1} \equiv 5^{p-2} = 5^{11} \pmod{13}$.
Writing $11 = 2^0 + 2^1 + 2^3$, we compute

$$\begin{aligned} 5^{2^0} &\equiv 5 \pmod{13}; \\ 5^{2^1} &\equiv 5^2 \equiv 25 \equiv -1 \pmod{13}; \\ 5^{2^2} &\equiv (-1)^2 \equiv 1 \pmod{13}; \\ 5^{2^3} &\equiv 1^2 \equiv 1 \pmod{13}. \end{aligned}$$

Hence we calculate

$$5^{-1} \equiv 5^{2^0+2^1+2^3} \equiv 5 \cdot (-1) \cdot 1 \equiv -5 \equiv 8 \pmod{13}.$$

c) In general, how many multiplications does the fast powering algorithm require? 4 p

Solution: To compute $a^n \pmod{p}$, the fast powering algorithm requires at most $2\log_2(n)$ multiplications: by successively squaring, one can compute $a^{2^{\lfloor \log_2(n) \rfloor}}$ in $\lfloor \log_2(n) \rfloor$ multiplications. To get $a^n \pmod{p}$, one has to then multiply at most $\lceil \log_2(n) \rceil$ of these values together, which requires at most another $\lfloor \log_2(n) \rfloor$ multiplications.

3. a) What do we mean by the discrete logarithm problem in a finite group G? 2 p

Solution: The discrete logarithm problem in a finite group G means the problem of finding $x \in \mathbb{Z}$ satisfying $g^x = h$ for given $g, h \in G$.

b) Consider the following invertible matrices with coefficients in $\mathbb{F}_7$:

$$g = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad h = \begin{pmatrix} 1 & 0 \\ 6 & 1 \end{pmatrix}.$$

Implement Shank's algorithm to solve the DLP $g^x = h$. You might find useful the identity

$$g^7 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

4 p

Solution: Since $g^7 = \mathrm{id}$ and 7 is a prime number, we find $N := \mathrm{ord}(g) = 7$. Hence $2 < \sqrt{N} < 3$, so $n = 1 + \lfloor \sqrt{N} \rfloor = 3$. We now create two lists:

$$\{e, g, g^2, g^3\} \qquad \text{and} \qquad \{h, hg^{-3}, hg^{-6}, hg^{-9}\}.$$

One computes

$$g^2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

$$g^3 = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}$$

$$hg^{-6} = hg = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Without computing more, we find a match: $hg^{-6} = e$. Hence

$$h = g^0 \cdot g^6 = g^6,$$

so $x = 6$.

c) What is the running time of Shank's algorithm for solving the DLP in $\mathbb{F}_p^*$? Explain. 4 p

Solution: Shanks' algorithm takes $\mathcal{O}(\sqrt{N}\log N)$ steps, where $N = \mathrm{ord}(g)$. Set $n = 1 + \lfloor\sqrt{N}\rfloor$. The creation of the lists $\{e, g, \ldots, g^n\}$ and $\{h, hg^{-n}, \ldots, hg^{-n^2}\}$ takes approximately $2n$ steps, since we can compute $u := g^{-n}$ one time and construct the second list as $\{h, hu, \ldots, hu^n\}$. Finding a match between the two lists of length $n+1$ takes $\mathcal{O}(n \log n)$ steps, so this determines the running time of the algorithm. Since $n \approx \sqrt{N}$, this comes down to a total running time of $\mathcal{O}(\sqrt{N}\log N)$.

4. a) Describe the Pohlig-Hellman algorithm. 4 p

Solution: The Pohlig-Hellman algorithm is a method to efficiently solve the discrete logarithm problem $g^x = h$ in a group G when $\mathrm{ord}(g) = N$ is composite. It consists of two parts.

Part 1: Suppose $N = p^e$, where p is a prime and e is a positive integer. We can solve $g^x = h$ as follows:

1. We look for x in the form
$$x = x_0 + x_1 p + x_2 p^2 + \ldots + x_{e-1}p^{e-1},$$
where $0 \leq x_i < p$ for $i = 0, \ldots, e-1$. This is possible, since if a solution x exists, we can assume that it satisfies $0 \leq x \leq N-1$.
2. Suppose we know $x_0, \ldots, x_{i-1}$ for some $i \geq 0$. We can solve x_i from the equation
$$(g^{p^{e-1}})^{x_i} = \left(h \cdot g^{-x_0 - x_1 p - \ldots - x_{i-1}p^{i-1}}\right)^{p^{e-i-1}}.$$

Part 2: For general N, write
$$N = q_1 \ldots q_n$$
where the q_i are pairwise coprime. Then we can solve $g^x = h$ as follows:

1. For each i, solve the discrete logarithm problem
$$(g^{N/q_i})^{x_i} = h^{N/q_i}.$$
2. Use the Chinese remainder theorem to find x such that $x \equiv x_i \pmod{q_i}$; this x solves the original DLP.

b) Using a cryptosystem based on the DLP in $\mathbb{F}_p^*$, how should you choose the modulus p in order to shield against the Pohlig-Hellman algorithm? 2 p

Solution: One should choose p such that $p-1$ can't be factorised into powers of small primes. Ideally, one should choose p such that $\frac{1}{2}(p-1)$ is prime.

c) What is the running time of the Pohlig-Hellman algorithm together with the naive algorithm to solve a DLP in a group with N elements? 3 p

Solution: The naive algorithm to solve a DLP $g^x = h$ in a group with N elements takes at most N steps: we have $\mathrm{ord}(g) \leq N$, so we have to try at most N values of x to find a solution.
Now suppose $N = p_1^{e_1} \ldots p_n^{e_n}$ as before. Then Part 2 of the Pohlig-Hellman algorithm involves solving n DLPs for elements of order $p_i^{e_i}$ for each i, and by Part 1, each of these amounts to solving e_i DLPs for elements of order p_i. Thus, in total this takes $\mathcal{O}(\sum_i e_i p_i)$ steps. The Chinese remainder theorem has a running time of $\mathcal{O}(\log N)$. Thus, the total running time is $\mathcal{O}(\sum e_i p_i + \log N)$.

5. a) Describe the RSA public key cryptosystem and explain what role Euler's theorem plays in it. 3 p

In the RSA public key cryptosystem, Alice picks two large primes p and q, and publishes the modulus $N := pq$ and a public key e satisfying $\gcd(e, (p-1)(q-1)) = 1$. This allows her to compute $d := e^{-1} \pmod{(p-1)(q-1)}$.
When Bob wants to send Alice a message m, he can send her the value $m^e \pmod N$, which Alice can decrypt because $(m^e)^d \equiv m \pmod N$. This last statement follows from Euler's theorem, which says that if $N = pq$ is a product of two primes, $g := \gcd(p-1, q-1)$, and $\gcd(a, N) = 1$, then $a^{(p-1)(q-1)/g} \equiv 1 \pmod N$. Indeed, since $d = e^{-1} \pmod{(p-1)(q-1)}$, we can write

$$de = 1 + k\frac{(p-1)(q-1)}{g}$$

for some $k \in \mathbb{Z}$, and consequently

$$(m^e)^d = m^{de} = m^{1+k(p-1)(q-1)/g} = m \cdot (m^{(p-1)(q-1)/g})^k \equiv m \pmod{(p-1)(q-1)}$$

b) Solve the congruence:

$$x^{27} \equiv 52 \mod 55$$

4 p

Solution 1: We use Euler's theorem. Note that $55 = 5 \cdot 11$. We find that 27 is coprime to $(p-1)(q-1) = 4 \cdot 10 = 40$, so by Euler's theorem, the unique solution is

$$x = 52^d \pmod{55}, \quad \text{where} \quad d = 27^{-1} \pmod{20},$$

since $20 = 4 \cdot 10/\gcd(4, 10)$. By inspection, we notice that $3 \cdot 27 \equiv 3 \cdot 7 = 21 \equiv 1 \pmod{20}$. Thus,

$$x \equiv 52^3 \equiv (-3)^3 = -27 \equiv 28 \pmod{55}.$$

Solution 2: Note that $55 = 5 \cdot 11$. By the Chinese remainder theorem, it is enough to solve the congruences

$$x_1^{27} \equiv 2 \pmod 5$$

and

$$x_2^{27} \equiv 8 \pmod{11}$$

and lift these to a solution modulo 55. Clearly $x = 0$ is not a solution to either equation, so for $p \in \{5, 11\}$ we may assume that $x^{p-1} \equiv 1 \pmod p$ (by Fermat's little theorem). Hence we have to solve

$$x_1^3 \equiv 2 \pmod 5$$

and

$$x_2^7 \equiv 8 \pmod{11}.$$

Some trial and error gives $x_1 \equiv 3 \pmod 5$ and $x_2 \equiv 6 \pmod{11}$. Finally, finding $0 \le x < 55$ which reduces to x_1 modulo 5 and to x_2 modulo 11 is easy by checking which of the numbers $6 + 11n$ work for $0 \le n < 5$. This gives the solution $x = 28$.

c) Alice and Bob both create keys for the RSA cryptosystem. They both choose the modulus $N = 8549$, but Alice's encryption key is $e_A = 5$ while Bob's is $e_B = 4187$. Eve encrypts the message $m = 44$ using both keys and finds that the ciphertexts coincide. Using this information help Eve factor the modulus N. (*Hint:* $93^2 = 8649$.) 5 p

Solution: We are given that $44^5 \equiv 44^{4187} \pmod{8549}$, so $44^{4182} \equiv 1 \pmod{8549}$. Hence also $44^{2\cdot 4182} = 44^{8364} \equiv 1 \pmod{8549}$. It seems reasonable to believe that $(p-1)(q-1) = 8364$ given this information. This gives $p + q = pq - 8364 + 1 = 186$. Then we have

$$(X-p)(X-q) = X^2 - (p+q)X + pq = X^2 - 186X + 8549,$$

which we can solve with the quadratic formula. Noting that $186 = 2 \cdot 93$ and using the hint $93^2 = 8649$, we obtain the solutions

$$\frac{186 \pm \sqrt{186^2 - 4 \cdot 8549}}{2} = \frac{2 \cdot 93 \pm 2\sqrt{93^2 - 8549}}{2} = 93 \pm \sqrt{100}.$$

Thus $N = 83 \cdot 103$.

6. a) Let $N = 44377$, $F(T) = T^2 - N$, and $a = \lfloor \sqrt{N} \rfloor + 1 = 210$. Characterize which of the numbers

$$F(a), F(a+1), F(a+2), \ldots, F(a+100)$$

are divisible by 5 and which are divisible by 11. 3 p

Solution: Let's start with the values modulo 5. We have $N \equiv 2 \pmod 5$, so 5 divides $F(T)$ if and only if $T^2 \equiv 2 \pmod 5$. But the squares modulo 5 are $0, 1, 4$, so we see that $F(T)$ is never divisible by 5.

We do the same for the values modulo 11. Now $N \equiv 3 \pmod{11}$, so 11 divides $F(T)$ if and only if $T^2 \equiv 3 \pmod{11}$. This equation has the two solutions $T \equiv 5, 6 \pmod{11}$. Since $a = 210 \equiv 1 \pmod{11}$, we see that $F(a+x)$ is divisible by 11 if and only if $x \equiv 4, 5 \pmod{11}$. Thus, the values divisible by 11 are

$$F(a+4),\ F(a+5),\ F(a+15),\ F(a+16),\ \ldots,\ F(a+92),\ F(a+93).$$

b) Now set $N = 3219577$, $F(T) = T^2 - N$, and $a = \lfloor \sqrt{N} \rfloor + 1 = 1794$. After computing $F(a+i)$ for $i = 0, \ldots, 350$, we found the following 13-smooth numbers:

$$\begin{aligned}
(a+7)^2 - N &= 2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 13 \\
(a+19)^2 - N &= 2^6 \cdot 3^4 \cdot 13 \\
(a+59)^2 - N &= 2^4 \cdot 3 \cdot 7^3 \cdot 13 \\
(a+73)^2 - N &= 2^7 \cdot 3^3 \cdot 7 \cdot 11 \\
(a+227)^2 - N &= 2^5 \cdot 3^3 \cdot 7 \cdot 11 \cdot 13 \\
(a+343)^2 - N &= 2^3 \cdot 3^7 \cdot 7 \cdot 11
\end{aligned}$$

Find at least four perfect squares one can form out of these numbers. 4 p

Solution: We have 6 primes ≤ 13, so we represent each of the found 13-smooth numbers by vectors in $\mathbb{F}_2^6$ whose i-th entry is the parity of the power of the i-th prime in its factorisation. Putting these into a matrix as column vectors, we obtain

$$\begin{pmatrix}
1 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 1 & 0
\end{pmatrix},$$

which we reduce with Gaussian elimination to

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

A general element in the kernel is $(a, b, 0, c, a + b, b + c)$. Taking for example $(a, b, c) \in \{(1,0,0), (0,1,0), (0,0,1), (1,1,0)\}$ gives four perfect squares corresponding to the vectors

$$\begin{aligned}
(1,0,0,0,1,0) &\to (2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 13) \cdot (2^5 \cdot 3^3 \cdot 7 \cdot 11 \cdot 13); \\
(0,1,0,0,1,1) &\to (2^6 \cdot 3^4 \cdot 13) \cdot (2^5 \cdot 3^3 \cdot 7 \cdot 11 \cdot 13) \cdot (2^3 \cdot 3^7 \cdot 7 \cdot 11); \\
(0,0,0,1,0,1) &\to (2^7 \cdot 3^3 \cdot 7 \cdot 11) \cdot (2^3 \cdot 3^7 \cdot 7 \cdot 11); \\
(1,1,0,0,0,1) &\to (2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 13) \cdot (2^6 \cdot 3^4 \cdot 13) \cdot (2^3 \cdot 3^7 \cdot 7 \cdot 11).
\end{aligned}$$

c) Write down all the checks for factors of N coming from the perfect squares you found in (b). You do not need to carry out the computations. 4 p

Solution: Above, we have found four numbers which are 13-smooth and perfect squares; say $b_1^2, \ldots, b_4^2$. We also know that these are the reductions of some $a_1^2, \ldots, a_4^2$ modulo N, and for each i, we want to compute $\gcd(N, a_i - b_i)$ in order to hopefully factor N. So we need to check the following quantities:

$$\begin{aligned}
&\gcd(N, (a+7)(a+227) - 2^4 \cdot 3^2 \cdot 7 \cdot 11 \cdot 13); \\
&\gcd(N, (a+19)(a+227)(a+343) - 2^7 \cdot 3^7 \cdot 7 \cdot 11 \cdot 13); \\
&\gcd(N, (a+73)(a+343) - 2^5 \cdot 3^5 \cdot 7 \cdot 11); \\
&\gcd(N, (a+7)(a+19)(a+343) - 2^6 \cdot 3^6 \cdot 7 \cdot 11 \cdot 13).
\end{aligned}$$

7. a) Consider the elliptic curve $E : y^2 = x^3 + x + 1$ over $\mathbb{F}_5$. Check that E indeed is an elliptic curve and that the points $P = (2, 4)$ and $Q = (3, 1)$ are on E, and calculate $P + Q$. 3 p

Solution: To check that an equation $y^2 = x^3 + ax + b$ is an elliptic curve over $\mathbb{F}_p$, we need that

$$\Delta = 4a^3 + 27b^2 \not\equiv 0 \pmod{p}.$$

In this case, $\Delta = 4 + 27 \equiv 1 \pmod{5}$, so E is an elliptic curve.

The point $(2, 4)$ is on the curve iff $4^2 \equiv 2^3 + 2 + 1 \equiv 11 \equiv 1 \pmod{5}$, which is true. Similarly, $(3, 1)$ is on the curve iff $1^3 \equiv 3^3 + 3 + 1 \equiv 31 \equiv 1 \pmod{5}$, which is also true.

We calculate $P + Q$ as follows. The line through P and Q is $y = 2(x - 3) + 1$. We obtain

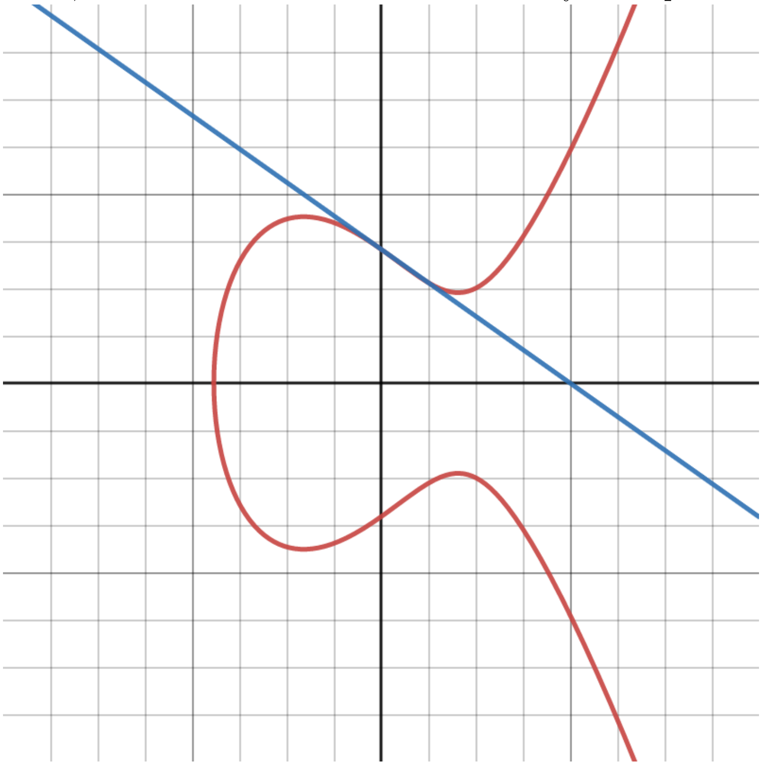
$$y^2 = x^3 + x + 1 = 4(x-3)^2 + 4(x-3) + 1,$$

so $x^3 + x^2 + x + 1 \equiv 0 \pmod{5}$. Since P and Q are points of intersection, we can factor out $(x - 2)(x - 3)$, which makes the last factor $(x - 4)$. Plugging in $x = 4$ in $y = 2(x - 3) + 1$ gives $y = 3$. Finally, to obtain $P + Q$, we need to replace y by $-y$, which gives

$$P + Q = (4, 2).$$

b) An *inflection point* of an elliptic curve E is a point P where the tangent line meets E with multiplicity 3. What is the order of such at point P? Draw a picture. 3 p

Solution: If the tangent line meets E with multiplicity 3 in a point P, it follows that the third point of intersection of the tangent line at P with E is again P. By definition, this point is $-2P$, so we have $-2P = P$ and thus $3P = 0$. Hence P has order 1 or order 3.
Your picture should contain a non-vertical tangent line to an elliptic curve which “crosses” the curve, such that it will not intersect it in any other point. Example:



c) Let E be an elliptic curve over $\mathbb{F}_{53}$. Explain why the number of points on E is between 39 and 69. 3 p

Solution: By Hasse's theorem, the number of points $|E(\mathbb{F}_p)|$ on an elliptic curve over $\mathbb{F}_p$ satisfy $|E(\mathbb{F}_p)| = p + 1 - t_p$, where $|t_p| \leq 2\sqrt{p}$. In this case, $2\sqrt{p} = 2\sqrt{53} < 2\sqrt{64} = 16$, so

$$39 = 54 - 15 \leq |E(\mathbb{F}_{53})| \leq 54 + 15 = 69.$$

d) Why is the fast powering algorithm particularly fast on an elliptic curve compared to an arbitrary group? 1 p

Solution: Subtracting points on elliptic curves is equally time-efficient as adding points. Therefore, when computing nP, one need not restrict themselves to expanding n as a binary number, but can also allow for minus signs between powers of 2. This will lead to a more efficient fast powering algorithm because the number of steps to compute nP may be reduced.

8. a) Describe the elliptic curve Diffie-Hellman key exchange. How should the public parameters be chosen? 4 p

Solution: A trusted party publishes a prime number p, an elliptic curve $E/\mathbb{F}_p$, and a point P on E. If Alice and Bob want to perform a Diffie-Hellman key exchange, they should work as follows:

- Alice picks an integer n_A and sends Bob the point $n_A P$.
- Bob picks an integer n_B and sends Alice the point $n_B P$.
- Alice computes $n_A(n_B P)$ and Bob computes $n_B(n_A P)$. This is their shared secret key.

For safety reasons, the public parameters should be chosen in such a way that the ECDLP has no easy solutions. For instance, one should avoid points P such that $\mathrm{ord}(P)$ is a product of powers of small primes, as this would make the Pohlig-Hellman algorithm a feasible attack. Similarly, one should avoid pairs (p, E) with $|E(\mathbb{F}_p)| = p + 1$ (i.e. supersingular elliptic curves), since in this case the MOV algorithm yields a feasible attack.

b) What is the main benefit of cryptosystems based on elliptic curves compared to those based on $\mathbb{F}_p^*$? 3 p

Solution: The main benefit is that the DLP in $\mathbb{F}_p^*$ can be solved in subexponential time using the index calculus (meaning $\mathcal{O}(p^\epsilon)$ for every $\epsilon > 0$), whereas the fastest known algorithm to solve the ECDLP in $E(\mathbb{F}_p)$ takes $\mathcal{O}(\sqrt{p})$ steps.

c) Describe Lenstra's factorization algorithm. What kinds of numbers does it factor particularly efficiently? 5 p

Solution: In Lenstra's factorization algorithm, one starts with a number $N = pq$ which one wants to factorize. One then picks an elliptic curve E over $\mathbb{Z}/N\mathbb{Z}$ and a point P on it. One then calculates $2!P, 3!P, 4!P, \ldots$. At any step, it may be that the computation of $n!P$ fails because one needs to compute the inverse of a denominator, which does not exist in $\mathbb{Z}/N\mathbb{Z}$; it follows that

$$\gcd(\text{denominator}, N) > 1,$$

and the hope is that this gcd is a proper factor of N. If it is, we are done; if not, we try again with a different elliptic curve and a different point.

Lenstra's factorisation algorithm has a running time which only depends on the smallest prime factor of N. Thus, it factors N particularly efficiently if it has a relatively small prime factor.