The trial division or "naif" factoring algorithm.

Let $n \in \mathbf{N}$ be a composite number. Then n has at least a prime factor $p \leq \sqrt{n}$. The trial division factoring algorithm consists of dividing n by all the primes $s \leq \sqrt{n}$, until the smallest prime factor p is found (assuming that the list of primes up to $\sqrt{n}$ is already available).

By the Prime Number Theorem, $\pi(p)$, the number of primes $s \leq p$, can be approximated with $\frac{p}{\ln p}$. To obtain the smallest prime factor p of n, it takes $\frac{p}{\ln p}$ divisions with remainder $n : s$, for $s \leq p$. Then the complexity of this algorithm can be estimated as

$$\mathcal{O}(\frac{p}{\ln p} \log n \log p) = \mathcal{O}(p \log n),$$

whcih is exponential in the smallest factor p.

The Pollard ρ factoring algorithm.

The Pollard ρ factoring algorithm was introduced by John Pollard in 1975. It is based on the birthday paradox and it is very efficient in determining the small factors of a given number.

The birthday paradox. *Let X be a set with N elements. Pick a random element from X. If this action is repeated k times, then the probability $p(k)$ that some element is picked twice can be estimated as*

$$p(k) > 1 - e^{-k(k-1)/2N}.$$

Proof. The probability $p(k)$ that *some* element is picked twice is given by

$$p(k) = 1 - \frac{N(N-1)\dots(N-k+1)}{N^k} = 1 - (1 - \frac{1}{N}) \cdot \ldots \cdot (1 - \frac{k-1}{N}) = 1 - \prod_{i=1}^{k-1}(1 - \frac{i}{N}).$$

From the inequality $1 - x < e^{-x}$, for $x \neq 0$, one has

$$\prod_{i=1}^{k-1}(1 - \frac{i}{N}) < \prod_{i=1}^{k-1} e^{-\frac{i}{N}} = e^{-\sum_{i=1}^{k-1} \frac{i}{N}} = e^{-k(k-1)/2N},$$

from which the required estimate follows

$$p(k) > 1 - e^{-k(k-1)/2N}.$$

Remark. In particular,

$$p(k) > \frac{1}{2}, \quad \text{for } k > 1,177\sqrt{N}; \qquad p(k) > 1 - e^{-5} \sim 0.9932, \quad \text{for } k > 3,16\sqrt{N}.$$

In short, we may say that picking k times an element out of a set of N elements, there is a positive probability to pick the same element twice.

Conversely, if picking k times an element out of a set of a unknown cardinality N we got distinct elements, then N was likely larger than k^2.

How the birthday paradox is applied to the factorisation of an integer n?

In this case the set X is $\mathbf{Z}_n$, the group of residue classes modulo n. Let $\{x_j\}_{j=1}^k$ be a random sequence of length k in $\mathbf{Z}_n$. For every divisor p of n, the sequence descends to a random sequence in $\mathbf{Z}_p$.

• If $p < k^2$ then, by the birthday paradox, with high probability two elements of the sequence coincide in $\mathbf{Z}_p$. The smaller p is with respect to k^2, the more probable is a coincidence to occur in $\mathbf{Z}_p$.

• Even *without knowing the factorisation of* n, a coincidence in some $\mathbf{Z}_p$ can be detected by taking

$$\gcd(x_i - x_j \bmod n, n), \quad i, j = 1, \ldots, k.$$

Indeed, one has $\gcd(x_i - x_j \bmod n, n) > 1$ if and only if there is a divisor p of n for which $x_i = x_j \bmod p$. (it is unlikely that $\gcd(x_i - x_j \bmod n, n) = n$, unless $k \gg \sqrt{n}$). *In this way* $\gcd(x_i - x_j \bmod n, n)$ *will provide a non trivial factor of* n.

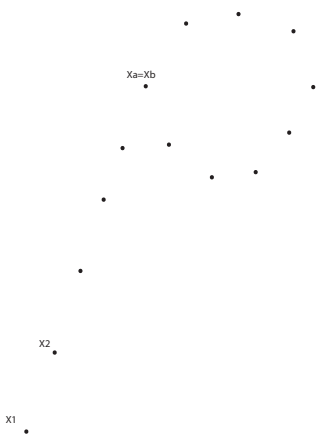
• A crucial idea in the Pollard ρ algorithm is to produce a random sequence in $\mathbf{Z}_n$ by evaluating the iterates of some function $f\colon \mathbf{Z}_n \to \mathbf{Z}_n$ on a random element $x_0 \in \mathbf{Z}_n$

$$x_1 = f(x_0), \quad x_2 = f^{(2)}(x_0) = f \circ f(x_0), \quad \ldots, \quad x_k = f^{(k)}(x_0) = f \circ f \ldots \circ f(x_0).$$

The function which is usually taken is $f(x) = x^2 + 2 \bmod n$. Once in the sequence a coincidence occurs

$$x_a = x_b, \quad \text{for some } b > a \geq 1, \tag{$*$}$$

then the sequence becomes periodic of period $b - a$, describing the letter ρ which gives the name to the algorithm.



The letter ρ *described by the sequence.*

• Another crucial idea of the algorithm is the following observation, which the reduces the number of gcd's to be performed from k^2 to k.

Lemma. *(Floyd's trick) If for some* $1 \leq a < b$ *one has*

$$x_a = x_b \mod p,$$

for some divisor p *of* n*, then there exists* $a \leq m \leq b$ *such that*

$$x_m = x_{2m} \mod p.$$

Dim. Suppose that there exist a and $b > a$ such that $x_b = x_a \bmod p$. From then on the sequence becomes periodic of period $b - a$. We claim that there exist $a \leq m \leq b$ and $s \in \mathbf{N}$ for which

$$m + s(b - a) = 2m \quad \Leftrightarrow \quad m = s(b - a).$$

Let's check that $s = [\frac{a}{(b-a)}] + 1$ works. Set $m = ([\frac{a}{(b-a)}] + 1)(b-a)$. One has indeed

$$m = ([\frac{a}{(b-a)}] + 1)(b-a) \geq \frac{a}{(b-a)}(b-a) = a$$

$$m = ([\frac{a}{(b-a)}] + 1)(b-a) \leq (\frac{a}{(b-a)} + 1)(b-a) = b.$$

As a result, one has

$$x_m \equiv x_{2m} \bmod p,$$

as claimed.

• Because of the above Lemma, one should form two parallel sequences

$$x_i = f^{(i)}(x_0), \quad \text{and} \quad y_i = f^{(2i)}(x_0),$$

and take

$$\gcd(x_i - y_i \bmod n, n), \qquad \text{for} \quad i = 1, \ldots k.$$

By the birthday paradox this procedure is likely to detect any non trivial factor p of n of size $\leq k^2$.

Remark. The Pollard ρ algorithm can be also be applied to solve the DLP in a cyclic group G. In order to apply the birthday paradox one must be able to produce random sequences in G, which at the same time can be controlled and allow us to deduce the information we want.

For example, if G is the group of points of an elliptic curve over $\mathbf{Z}_p$, with p prime, the construction of f is more complicated. See L. Washington.