SEARCH FOR ALL 32-BIT PRIMES BY SIEVE OF ERATOSTHENES
Efficient Storage Of All 32-bit Primes List In A File And In RAM
Simple Random Generator Of 64-bit Prime Numbers
September, 2023

Abstract

Prime numbers are widely used in cryptographic systems. However, the strength of popular ciphers depends on the selection of the numbers considered prime. Systems with short primes are considered weak because the secret number(s) can be guessed by the attacker. The large primes are difficult to compute, thus a cryptographic system must either use the known prime number that has the mathematical proof or depend on the primality tests. This approach has an obvious flaw. The known prime numbers are available for prior study by the attacker. The primality tests generally do not guarantee that the number in question is the prime.

One possibility to overcome this issue is to devise a system that uses many relatively short primes in a way where the attacker must know all of them before he can decipher the message. The "short primes" must be abundant and easy to compute to support such a system.

We demonstrate that the regular modern computer can build and manipulate the complete list of 32-bit prime numbers. We also discuss the random selection of a prime within 64-bit range.

Let's use the sieve of Eratosthenes, the well-known and very simple method to find small prime numbers. The very simple yet mathematically strict foundation makes the practical computer implementation a good scholarly exercise. This article targets an audience without a mathematical background.

Notations

Let us remind readers with no mathematical background of the usual notations and facts from the theory. We also introduce a few non-standard notations here.

The set of natural numbers (positive integers) is represented by the letter N. The set of whole numbers (positive, negative integers, and zero) is represented by the letter Z. Let's also use the non-standard notation N_k for the whole numbers greater or equal to k. For example, $N = N_1$ by definition. The 0-based index used in many popular programming languages is N_0.

The set of rational numbers Q consists of all possible fractions a/b, where a is a whole number, and b is a natural number.

A natural number is *prime,* if it divides by two numbers, 1 and itself. (1 is not considered a prime.) The series of prime numbers is infinite and starts with

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, ...$$

In contrast, *a composite* number can be represented by a product of two or more primes. Let's use the letter P for the set of the prime numbers. Similar to the restricted whole numbers, we use the notation P_k for the prime numbers greater or equal to k. Our most interest is P_5.

We also will use so-called *prime candidates,* P^*, the natural numbers that are not divisible by 2 and 3. Enumerating of the prime candidates is still simple, compared to the indexing of all odd numbers, while it further accelerates the prime numbers search compared to considering just the odd numbers. For the task of computing primes, we can skip the beginning of the list. We are considering mostly P_5^*.

All prime numbers are prime candidates, but not all candidates are primes. All candidates are integers:

$$P_5 \subset P_5^* \subset N_5$$

Let's also remind the common mathematical symbols: $\forall$ — for each, any; $\exists$ — exists; $\exists!$ — exists and only one; $a \in A$ — the element a belongs to the set A; $A \subset B$ — all elements of A belong to B, A is subset of B; $\Rightarrow$ — therefore, ergo; $\equiv$ — identity, always equal.

When a is a fractional number, $\lfloor a \rfloor$ is its whole part of a, less of equal a.

The remainder of division $a \in Z$ by $b \in N$ is written as $a \bmod b$. Note that when $a < 0$, the division is according to Euclid's division lemma (see: Facts), so the remainder is always positive, counted from the multiply of b upwards. If a divides by b, it can be written as $a \mid b$. It happens if and only if $a \bmod b = 0$.

Facts from the theory of numbers

Euclid's division lemma

$$\forall A \in Z, \quad q \in N \quad \exists!\, k, r \in Z, \quad 0 \le r < q, \quad \text{such as} \quad A = kq + r \tag{1}$$

Corollary. Any rational number $a = A/q$ can be uniquely represented as the sum of the whole number $k = \lfloor a \rfloor$ and fractional part ε, where $0 \le \varepsilon < 1$. **Proof.** By Euclid's lemma, $a = (kq + r)/q = k + \varepsilon$, where $0 = 0/q \le \varepsilon = r/q < q/q = 1$. On top of that, $k \in Z$ and $r \in N$ are unique. *Q.e.d.*

The remainder of a division can be expressed through Euclid's lemma. If $a = kb + r$ as defined in (1), then $a \bmod b = r$, and $0 \le r < b$. **Example.** $(-1) \bmod 6 = 5$.

The following trivial observation will be frequently used. $\forall k \in Z$

$$\begin{aligned} 0 &\le k \bmod 2 \le 1 \\ -1 &\le -(k \bmod 2) \le 0 \\ 0 &\le 1 - (k \bmod 2) \le 1 \end{aligned}$$

Note that for the prime candidate it becomes the equilibrium:

$$\forall q \in P_5^* \quad q \bmod 2 = 1$$

The main theorem. Let us rephrase it in the form:

$$\forall A \in N_2 \quad \exists p_1, p_2, \dots p_k \in P, \quad \text{and} \quad \alpha_1, \alpha_2, \dots \alpha_k \in N, \quad \text{such as} \quad A = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k} \tag{2}$$

(Remark. The actual statement is more strict. With some lengthy clarification, the decomposition (2) is unique. But later on, we only use the fact that it is possible!)

Let us introduce a few simple statements that we will also use frequently.

Lemma A. Let $a, b \in Z$. If $a > b$ then $a \ge b + 1$.

Proof. Let $t = a - b > 0$, and $t \in Z$. In other words, t can be 1, 2, 3, etc. That is, $t \ge 1$. *Q.e.d.*

Lemma B. Let $a, b \in Q$. If $a \le b$ then $\lfloor a \rfloor \le \lfloor b \rfloor$.

Proof by contadiction. Let's write (using the Corollary of Euclid's division lemma):

$$\begin{cases} a = a_0 + \varepsilon, \quad a_0 \in Z, \quad 0 \le \varepsilon < 1 \\ b = b_0 + \delta, \quad b_0 \in Z, \quad 0 \le \delta < 1 \end{cases}$$

and let the lemma statement be false, $a_0 > b_0$. By Lemma A, $a_0 \ge b_0 + 1$, or $a_0 - b_0 \ge 1$. Write:

$$a - b = a_0 + \varepsilon - b_0 - \delta = (a_0 - b_0 - 1) + \varepsilon + (1 - \delta) > 0,$$

because the first two members are non-negative, and the last one is positive. Therefore, $a > b$. Contradiction proves the lemma. *Q.e.d.*

Indexes of prime candidates and their properties

1) Elements of P_5^* can be represented in the form $6m \pm 1$, where $m \in N$. If $p \in P_5$, then $p \in P_5^*$.

2) Consider the enumeration function

$$Q(k) = 3k + 5 - (k \bmod 2), \quad k \in N_0. \tag{3}$$

This function represents all possible prime candidates. The set of all $Q(k)$-s is the same as P_5^*. For all $k \in N_0$, $Q(k) \geq 5$.

3) The reversal of the enumeration function Q is

$$K(q) = \left\lfloor \frac{q - 2 - (q \bmod 2)}{3} \right\rfloor, \quad \forall q \in N_5,\ K(q) \geq 0 \tag{4}$$

If $q \in P_5^*$, then $Q(K(q)) = q$. If $k \in N_0$, then $K(Q(k)) = k$.
Corollary. If $Q(K(q)) = q$ for $q \in N_5$, then $q \in P_5^*$.

4) Let $k_1, k_2 \in N_5$. If $k_1 < k_2$, then $Q(k_1) < Q(k_2)$.
Corollary. If $k_1 \neq k_2$, then $Q(k_1) \neq Q(k_2)$.

5) Let $q_1 \in N_5$, $q_2 \in P_5^*$. If $q_1 > q_2$, then $K(q_1) > K(q_2)$.
Corollary. If $q_1, q_2 \in P_5^*$ and $q_1 \neq q_2$, then $K(q_1) \neq K(q_2)$.

6) Let $q_1, q_2 \in N_5$. If $q_1 \leq q_2$, then $K(q_1) \leq K(q_2)$.
Note that unlike in Propety 5, q_2 doesn't have to be the prime candidate.
Corollary. $\forall q_0 \in N_5$, the $q^* = Q(K(q_0)) \in P_5^*$ is the nearest prime candidate such that $q^* \geq q_0$.

7) Let $m \in N_2$. Then $2^{2m} - 1 \notin P_5^*$. It divides by 3.

8) Let $m \in N$. Then $2^{2m} + 1 \in P_5^*$.
Corollary. For $m \in N_2$, $K(2^{2m} - 1) = K(2^{2m} + 1)$. If $p \in P$ and $p < 2^{2m}$, then $K(p) < K(2^{2m} - 1)$.
In other words, search for the primes in the 2^{2m} range shall stop at index $K(2^{2m} - 1) - 1$. Note that no portion of the last expression exceeds $2^{2m} - 1$.

Proof

1) According to Euclid's lemma (1), any whole number can be uniquely represented as $6t + \gamma$, where $t \in Z$ — some index number, and $\gamma = 0$, 1, 2, 3, 4, or 5. Let's note that $6t + 5 = 6(t + 1) - 1$. Since we are looking for numbers from N_5, let's rewrite the statement above. Any number $n \in N_5$ can be represented as one of the following, with $m \in N$:

a) $6m - 1$
b) $6m + 0$ | 2
c) $6m + 1$
d) $6m + 2$ | 2
e) $6m + 3$ | 3
f) $6m + 4$ | 2

Except (a) and (c), all other numbers divide by 2, 3, or both. Neither (a) nor (c) divides by 2 or 3. $P_5 \subset P_5^*$ because otherwise such a prime number $p \geq 5$ exists that divides by 2 or 3. *Q.e.d.*
That is,

$$\forall q \in P_5^* \quad \exists!\, m \in N,\ \gamma \in \{-1, 1\}, \quad \text{such as } q = 6m + \gamma. \tag{5}$$

2) Let's establish the relationship between the index k in formula (3) and the pair (m, γ) in (5). Let $k = 2s + t$, where $s \geq 0$ and $t = 0, 1$. We have $k \bmod 2 = t$. Write:

$$3(2s + t) + 5 - t = 6s + 2t + 5 = 6(s + 1) + (2t - 1),$$

which gives us the transformation formulas

$$\begin{cases} s = \lfloor k/2 \rfloor \\ t = k \bmod 2 \\ m = s + 1 \\ \gamma = 2t - 1 \end{cases} \qquad \begin{cases} s = m - 1 \\ t = (\gamma + 1)/2 \in \{0, 1\} \quad \textit{(verify!)} \\ k = 2s + t \end{cases}$$

$\forall k$ satisfying (3) $\exists (m, \gamma)$ satisfying (5), and vice versa, $\forall (m, \gamma)$ satisfying (5) $\exists k$ satisfying (3). *Q.e.d.*

P.S. Note that we only proved the existence of the alternative representation, not the uniqueness of the transformation. However, we will not need it in our study.

3) Verify that $K(q) \geq 0$ for any $q \geq 5$. We have

$$q - 2 - (q \bmod 2) \geq 5 - 2 - 1 = 2 \geq 0.$$

By Lemma B and expression (4), $K(q) \geq 0$. *Q.e.d.*

Verify the prime candidate transformation. By formula (5), $q = 6m + \gamma$, $m \geq 1$, and $\gamma \in \{-1, 1\}$. Evaluate the expression (4) using q:

$$K(q) = \left\lfloor \frac{q - 2 - (q \bmod 2)}{3} \right\rfloor = \left\lfloor \frac{(6m - 6) + (3 + \gamma)}{3} \right\rfloor = 2(m - 1) + \begin{cases} 1, & \gamma = 1 \\ 0, & \gamma = -1 \end{cases}$$

Evaluate expression (3). We have

$$K(q) \bmod 2 = \begin{cases} 1, & \gamma = 1 \\ 0, & \gamma = -1 \end{cases}$$

$$Q(K(q)) = 6(m - 1) + \left\{ \begin{matrix} 3, & \gamma = 1 \\ 0, & \gamma = -1 \end{matrix} \right\} + 5 - \left\{ \begin{matrix} 1, & \gamma = 1 \\ 0, & \gamma = -1 \end{matrix} \right\}$$

$$= 6m + \left\{ \begin{matrix} 1, & \gamma = 1 \\ -1, & \gamma = -1 \end{matrix} \right\} \equiv 6m + \gamma. \quad \textit{Q.e.d.}$$

Verify the index transformation. We have $k \bmod 2 \in \{0, 1\}$ and $Q(k) \bmod 2 \equiv 1$.

Using (3) and (4), write:

$$K(Q(k)) = \left\lfloor \frac{3k + 5 - (k \bmod 2) - 2 - 1}{3} \right\rfloor = \left\lfloor \frac{3k + 2 - (0 \text{ or } 1)}{3} \right\rfloor \equiv k.$$

Finally, to prove the Corollary, notice that $K(q) \in N_0$ $\forall q \in N_5$. By Property 2, $Q(k) \in P_5^*$ if $k \in N_0$. Therefore, $q \in P_5^*$. *Q.e.d.*

4) By Lemma A, $k_1 \leq k_2 - 1$. We have

$$3k_1 + 5 - (k_1 \bmod 2) \leq 3k_1 + 5 \leq 3(k_2 - 1) + 5 < 3k_2 + 5 - 1 \leq 3k_2 + 5 - (k_2 \bmod 2). \quad \textit{Q.e.d.}$$

Corollary: if $k_1 \neq k_2$, then either $k_1 < k_2$ or $k_1 > k_2$. By Property 4, it makes either $Q(k_1) < Q(k_2)$ or $Q(k_1) > Q(k_2)$, but not equal. *Q.e.d.*

5) Consider $\delta = q_1 - q_2 > 0$, and express it by Euclid's lemma in form $\delta = 2t + \varepsilon$. Statement. $\delta + \varepsilon \geq 2$. Proof. Consider separately the cases $\varepsilon = 0$ and $\varepsilon = 1$. Because of Lemma A,

$$\delta \geq 1 \quad \Rightarrow \quad \begin{bmatrix} \varepsilon = 0, & t > 0 \\ \varepsilon = 1, & t \geq 0 \end{bmatrix} \quad \Rightarrow \quad t \geq 1 \quad \Rightarrow \quad \begin{bmatrix} \varepsilon = 0, & \delta \geq 2 \\ \varepsilon = 1, & \delta \geq 1 \end{bmatrix} \quad \Rightarrow \quad \delta + \varepsilon \geq 2.$$

Because of Property 3, $\exists k_2 = K(q_2)$ such that

$$q_2 = 3k_2 + 5 - (k_2 \bmod 2)$$

Express $q_1 \bmod 2$:

$$q_1 = q_2 + \delta = (q_2 - 1) + 2t + 2\varepsilon + (1 - \varepsilon)$$
$$q_1 \bmod 2 = 0 + 0 + 0 + (1 - \varepsilon)$$

Write the expression for $K(q_1)$ and establish its connection with $K(q_2)$:

$$K(q_1) = \left\lfloor \frac{q_1 - 2 - (q_1 \bmod 2)}{3} \right\rfloor = \left\lfloor \frac{3k_2 + 5 - (k_2 \bmod 2) + \delta - 2 - (1 - \varepsilon)}{3} \right\rfloor$$
$$= \left\lfloor \frac{(3k_2 + 3) + (1 - (k_2 \bmod 2)) + (\delta + \varepsilon - 2)}{3} \right\rfloor \geq k_2 + 1 > k_2 = K(q_2), \quad Q.e.d.$$

Corollary: similar to corollary in Property 4, if $q_1 \neq q_2$, then either $q_1 < q_2$ or $q_1 > q_2$. If both $q_1, q_2 \in P_5^*$, then, by Property 5, either $K(q_1) < K(q_2)$ or $K(q_1) > K(q_2)$. *Q.e.d.*

6) Split the condition. If $q_1 = q_2$, then obviously $K(q_1) = K(q_2)$. We need to prove the statement in case $q_1 < q_2$. Let's compare

$$A_2 = \frac{q_2 - 2 - (q_2 \bmod 2)}{3} \quad \vee \quad A_1 = \frac{q_1 - 2 - (q_1 \bmod 2)}{3}.$$

By Lemma A, $q_2 - q_1 - 1 \geq 0$. Write

$$3(A_2 - A_1) = (q_2 - q_1 - 1) + (q_1 \bmod 2) + (1 - (q_2 \bmod 2)) \geq 0,$$

because all three members are not less than 0. Then, taking the whole parts of A_1 and A_2, by Lemma B, we conclude that $K(q_2) \geq K(q_1)$. *Q.e.d.*

Prove the Corollary by contradiction.

a) q^* cannot be smaller. It $q^* < q_0$, then $K(q^*) < K(q_0)$ (Property 5), and $q^* < Q(K(q_0))$ (Property 4). Since $Q(K(q_0)) = q^*$, we have $q^* \neq q^*$. Contradiction.

b) q^* is the next. Otherwise, $\exists\, q_+ \in P_5^*$, such that $q_0 \leq q_+ < q^*$.

$$\begin{cases} K(q_+) \geq K(q_0) & \text{(Property 6)} \\ K(q_+) < K(q^*) & \text{(Property 5)} \end{cases} \Rightarrow \begin{cases} Q(K(q_+)) \geq Q(K(q_0)) = q^* & \text{(Property 4)} \\ Q(K(q_+)) < Q(K(q^*)) = q^* \end{cases}$$

That it, a number $Q(K(q_0))$ exists that is less than q^* and is not less than q^* at the same time. This cannot happen. *Q.e.d.*

7) Prove by induction. $m \in N_2$. Verify at starting condition $m = 2$. We have $2^4 - 1 = 15 \mid 3$.

If the premise stands at m, verify that it also stands at $m + 1$. Let $2^{2m} - 1 = 3s$.

$$2^{2(m+1)} - 1 = 4 \cdot 2^{2m} - 1 = 4 \cdot (2^{2m} - 1) + 3 = 12s + 3 \mid 3, \quad Q.e.d.$$

8) Let $q_- = 2^{2m} - 1 = 3s$ again (Property 7). Then $q_+ = 2^{2m} + 1 = 3s + 2$. Let's also note that both q_- and q_+ are odd. Therefore, s is also odd, otherwise q_- is even.

$$K(q_+) = \left\lfloor \frac{q_+ - 2 - 1}{3} \right\rfloor = \left\lfloor \frac{3s - 3 + 2}{3} \right\rfloor = s - 1.$$

$K(q_+)$ is even, therefore

$$Q(K(q_+)) = 3K(q_+) + 5 = 3(s - 1) + 5 = q_+.$$

By Corollary of Property 3, $q_+ \in P_5^*$, *Q.e.d.*
Calculate $K(q_-)$ at $m \geq 2$.

$$K(q_-) = \left\lfloor \frac{q_- - 2 - 1}{3} \right\rfloor = \left\lfloor \frac{3s-3}{3} \right\rfloor = s - 1 = K(q_+).$$

By Property 5, if $p < 2^{2m} < q_+$, then $K(p) < K(q_+) = K(q_-)$. *Q.e.d.*

The sieve of Eratosthenes for small primes

Let us verify whether a prime candidate is the prime. While we need to make sure that it doesn't divide by any possible number except 1 and itself, we only *have to* test the divisors that are *less* than the prime candidate.

Statement. If $T \in N$ does not divie by any prime number $p_k < T$, then T is prime.

Proof. a) T does not divide be any $q > T$, otherwise $\exists\, k \geq 1$ such that $T = kq$. This is impossible: $kq \geq q > T$.

b) Let $q \notin P$ be divisor of T, and $q < T$. We can write: $T = kq$, and $k > 1$. Because of the main theorem, q can be represented as a product of several prime numbers. (Or q can be a prime.) q can be represented in form $q = k_q q_p$, where $k_q \geq 1$, and $q_p \in P$. But in this case, p can be represented in the form $T = kq = (kk_q) \cdot q_p$. Since $kk_q > 1$, we found the prime divisor $q_p < T$.

Summarising, if there is a composite divisor of T less than T, then there is also a prime divisor of T less than T. If T doesn't divide by any of the prime numbers less than itself, then it doesn't divide by any (maybe composite) numbers less than itself. *Q.e.d.*

This idea makes the sieve of Eratosthenes. We have a growing list of prime numbers starting with $2, 3, \ldots$, and exclude any number that divides by anything in the prior list. When we come across a number that was not excluded, it is the prime.

Lemma 1. To verify whether T is a prime, it is enough to try primes up to $\sqrt{T}$. Specifically, *the stop condition* is $p_k^2 > T$. If T doesn't divide by any of the smaller prime numbers p_1, p_2, ... p_{k-1}, then T is prime.

Proof by contradiction. Let's assume that T doesn't divide by any prime number less than p_k. If T divides by any number $p_k \leq q < T$, we can write $T = nq$, where $n > 1$. We have $n < p_k$, otherwise $T = nq \geq p_k^2 > T$. By the main theorem, n has a prime divisor $n_p \leq n < p_k$, and n_p is also the divider of T. Contradiction. *Q.e.d.*

Theorem 1. The stop condition is coming from the test divisions. Let neither of the prime numbers p_1, p_2, ... p_{k-1} be a divisor of T. Considering the next prime number p_k, if $\lfloor T/p_k \rfloor < p_k$, then T is prime.

Proof. Let $q = \lfloor T/p_k \rfloor$. By Euclid's lemma and Lemma A,

$$T = qp_k + r < qp_k + p_k = (q+1)p_k \leq p_k^2.$$

That is, $T < p_k^2$. Conditions of Lemma 1 are satisfied, which completes the proof. *Q.e.d.*

Theorem 2. To test a $T < 2^{2N}$ for primality, we only need to consider prime divisors $p_k < 2^N$. No need to look for primes above 2^N.

Proof. If $k-1$ is the largest index that provides $p_{k-1} < 2^N$, then $p_k > 2^N$. By the conditions of the Theorem and by Lemma 1,

$$p_k^2 > 2^{2N} > T,$$

and we don't need to consider p_k or any larger test numbers. *Q.e.d.*

Remark. As we study the computer implementation, N refers to the bit length of an integer supported by the platform. Typical integer sizes in modern computers are 8, 16, 32, or 64 bits. As the natural numbers are concerned, the signed integer types have sizes 7, 15, 31, or 63 bits.

Prime gap, index prime gap