Factorization & Primality Testing

Çetin Kaya Koç

http://cs.ucsb.edu/~koc
koc@cs.ucsb.edu

Primes

- Natural (counting) numbers: $\mathcal{N} = \{1, 2, 3, \ldots\}$
- A number $p \in \mathcal{N}$ is called prime if it is divisible only by 1 and itself
- $p = 1$ is not considered prime; 2 is the only even prime
- Primes: 2, 3, 5, 7, 11, 13, ...
- There are infinitely many primes
- Every natural number n is factored into prime powers uniquely:

$$n = p_1^{k_1} \times p_2^{k_2} \times \cdots \times p_m^{k_m}$$

For example: $1960 = 2^3 \times 5^1 \times 7^2$

Primes

- The number of primes less than or equal to n is $\frac{n}{\log_e(n)}$

n	$n/\log_e(n)$	exact
10^2	21.7	25
10^3	144.8	168
10^6	72382.4	78498
10^9	4.8×10^7	50847534

- As we can see, primes are in abundance; we do not have scarcity
- The odds of selecting a prime is high for small numbers: if we select a 2-digit integer, the probability that it is prime is $25/100 = 25\%$
- The odds of selecting a prime less than 10^6 is $78498/10^6 \approx 7.8\%$
- If we make sure that this number is not divisible by 2 or 3, (which makes up 2/3 of integers), the odds increase to 23.5%

Primes

- As the numbers get larger, which would be the case for cryptographic applications, the ratio becomes less and less
- The ratio of 1024-bit (308-digit) primes to the 308-digit numbers is

$$\frac{1}{\log_e(2^{1024})} \approx \frac{1}{714}$$

- Therefore, if we randomly select a 308-digit integer, the probability that it is prime is 1/714
- If we remove the multiples of 2 and 3 from this selected integer, the odds of choosing a 308-digit prime at random is improved by a factor of 3 to 1/238

Checking for Primality vs Factoring

- Primality testing: Is $n \in \mathcal{N}$ prime?
  The answer is yes or no (we may not need the factors if n is composite)
- Factoring: What is the prime factorization of $n \in \mathcal{N}$?
  The answer is $n = p_1^{k_1} \times \cdots \times p_m^{k_m}$
- Is $2^{101} + 81 = 2535301200456458802993406410833$ prime?
  The answer: Yes
- Is $2^{101} + 71 = 2535301200456458802993406410823$ prime?
  The answer: No
- Factor $n = 2^{101} + 61 = 2535301200456458802993406410813$
  The answer: $n = 3 \times 19 \times 1201 \times 37034944570408560161757109$

Factorization by Trial Division

- Trial division (exhaustive search): Find a prime factor of $n \in \mathcal{N}$ by dividing n by numbers that are smaller than n
- Observation 1: We do not need to divide n by composite numbers; it is sufficient that we only try primes, for example, if n is divisible by 6, then we could have discovered earlier that it was divisible by 2
- Observation 2: One of the factors of n must be smaller than $\sqrt{n}$, otherwise if $n = pq$ and $p > \sqrt{n}$ and $q > \sqrt{n}$ implies $pq > n$
- Trial division finds a prime factor of $n \in \mathcal{N}$ by dividing n by k for $k = 2, 3, \ldots, \sqrt{n}$
- Trial division requires $O(\sqrt{n})$ divisions (in the worst case); if n is a k-bit number, then $n = O(2^k)$ and the number of divisions is $O(2^{k/2})$ which is exponential in k

Factorization by Trial Division

- For example, finding the factorization $2^{101} + 61$ requires about 2^{50} divisions; assuming one division requires 1 μs, this would take 35 years!
- However, this is the worst case analysis, in the sense that a prime divisor is as large as it can be $\approx \sqrt{n}$
- If n has a small divisors, they will be found more quickly
- For example, $2^{101} + 61$ has smaller factors such as 3, 19, and 1201, and thus, the trial division algorithm would quickly find them
- Therefore, we conclude that if $n = p \times q$ such that $p, q \approx \sqrt{n}$, then the trial division would take the longest time

Factorization by Trial Division

- The number of divisions for factoring n with large prime factors is exponential in terms of the number of bits in n
- Trial division starts from $k = 2$ and increases k until $\sqrt{n}$, and thus, it is very successful on numbers which have small prime factors: these factors would be found first, reducing the size of the number to be factored
- For example, given $n = 1227331068230022428624 11$, we would find the smaller factors 17, 31, and 101 first, and divide them out

$$\frac{n}{17 \times 31 \times 101} = m = 230584302746730 4993$$

and then continue to factor m which is smaller in size than n

Fermat's Trial Division

- Fermat's idea was that if n can be written as the difference of two perfect squares:

$$n = x^2 - y^2$$

then, we can write

$$n = (x - y)(x + y)$$

and therefore, we can find two factors of n

- As opposed to the standard trial division algorithm, Fermat's method starts $x \approx \lceil \sqrt{n} \rceil$ and $y = 1$, and increases y until we find a y value such that $x^2 - y^2 = n$
- Since $x \approx \lceil \sqrt{n} \rceil$, Fermat's methods finds a factor that is closer to the size of $\sqrt{n}$ before it finds a smaller factor

Fermat's Trial Division

- For example, consider $n = 302679949$, we have $\lceil \sqrt{n} \rceil = 17398$
- We start with $x = 17398$ and $y = 1$, increase y as long as $x^2 - y^2 \leq n$
- We either find a y such that $x^2 - y^2 = n$ or the selected value of x does not work, i.e., we cannot find y such that $x^2 - y^2 = n$, then we increase x as $x = x + 1$ and start with $y = 1$ again
- It turns out for $x = 19015$, we find $y = 7674$ such that

$$x^2 - y^2 = 19015^2 - 7674^2 = 302679949 = n$$

therefore, n is factored as $n = (x - y)(x + y)$ such that

$$n = (19015 - 7674)(19015 + 7674) = 11341 \times 26689$$

Kraitchik's Method

- Instead of looking for x and y satisfying $x^2 - y^2 = n$, we can also search for "random" x and y such that

$$x^2 = y^2 \pmod{n}$$

- For such a pair (x, y), factorization of n is not guaranteed; we only know the difference of the squares is a multiple of n:

$$x^2 - y^2 = (x - y)(x + y) = 0 \pmod{n}$$

- Since n divides $(x - y) \times (x + y)$, we have $1/2$ chance that prime divisors of n are distributed among the divisors of both of these factors
- The GCD$(n, x - y)$ will be a nontrivial factor, the GCD will be neither 1 nor n

Kraitchik's Method

- For $n = 221 = 13 \times 17$, we find $x = 4$ and $y = 30$, such that $4^2 = 16 \pmod{221}$ and $30^2 = 900 = 16 \pmod{221}$, and therefore,

$$\text{GCD}(221, 30 - 4) = \text{GCD}(221, 26) = 13$$

- In fact, there are many (x, y) such that $x^2 = y^2 \pmod{221}$, which gives us a higher chance of finding a pair (x, y):

$$(2, 15), (3, 88), (5, 73), \ldots, (11, 28), \ldots$$

- Note that we still perform an exhaustive search to find a pair (x, y)
- There is an algorithm due to Dixon to find factors slightly more efficiently by expressing them into small prime powers, and working with the exponents, i.e., $r = 2^8 \times 3^6 \times 5^2 \times 7^0 \times 11^8$ implies that r is a square

Modern Factorization Methods

- Factorization in general seems to require exhaustive search: modern factorization algorithms differ from one another slightly in the way this search is constructed
- There is no known deterministic or randomized polynomial time algorithm for finding the factors of a given composite integer n, particularly, when $n = p \times q$ with size of p and q about half of the size of n
- The best integer factorization algorithm called GNFS (generalized number field sieve) algorithm requires a time complexity of

$$O\left(\exp\left(\left(\frac{64}{9}\,b\right)^{\frac{1}{3}}(\log b)^{\frac{2}{3}}\right)\right)$$

where b is the number of bits in n

Complexity of Factorization

- It is not known exactly which complexity classes contain the decision version of the integer factorization problem
- It is known to be in $\mathcal{NP}$ since a YES answer can be verified in polynomial time by multiplication: Are p and q factors of n?
- However, it is not known to be in $\mathcal{NP}$-complete; no such reduction proof is discovered
- Many people have looked for a polynomial time algorithm for integer factorization, and failed
- On the other hand, factorization problem can be solved in polynomial time on a quantum computer, using Shor's algorithm

Primality Testing

- The decision problem "Is n prime?" is called the primality testing
- Primality testing is easier than factorization, as might be expected, since we are not asking for the factors of n
- There are two very efficient randomized polynomial-time algorithms: **Fermat's method** and **Miller-Rabin method**
- There is also a deterministic polynomial-time algorithm invented in 2002: **The AKS algorithm**, due to three Indian computer scientists: Manindra Agrawal, Neeraj Kayal, and Nitin Saxena at the IIT Kanpur
- In the first version of their paper, time complexity was $O(b^{12})$, which was later improved to $O(b^{10.5})$ and then to $O(b^{7.5})$, where $b = \log(n)$

Fermat's Method

- Fermat's Little Theorem: If p is prime and $1 \leq a < p$, then

$$a^{p-1} = 1 \pmod{p}$$

- The contrapositive of Fermat's Little Theorem: If a and n satisfy $1 \leq a < n$ and $a^{n-1} \neq 1 \pmod{n}$, then n is composite
- Consider the list of $3^{n-1} \pmod{n}$ for $n = 4, 5, \ldots, 19$

n	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
3^{n-1}	3	1	3	1	3	0	3	1	3	1	3	9	11	1	9	1

- This shows that for all composite numbers in this range, $3^{n-1} \pmod{n}$ is distinct from 1, whereas all prime numbers satisfy $3^{n-1} = 1 \pmod{n}$

Fermat's Witness and Fermat's Liar

- Fermat's Little Theorem (and its contrapositive) provide good criteria for checking primality
- A number a in the range $a \in [1, n)$ and $\gcd(a, n) = 1$ is called a **Fermat's witness** for any $n \geq 2$, if $a^{n-1} \neq 1 \pmod{n}$
- Existence of a witness for n means n is a composite number
- A number a in the range $a \in [1, n)$ and $\gcd(a, n) = 1$ is called a **Fermat's liar** for an odd composite number $n \geq 3$, if $a^{n-1} = 1 \pmod{n}$
- Fermat's liar a is lying to us that n is prime, even though n is an odd composite number

Fermat's Witness and Fermat's Liar

- 2 is a witness for all composite n in the range $[2, 340]$ since if n is composite then $2^{n-1} \neq 1 \pmod{n}$, for $n = 2, 3, \ldots, 340$
- 2 is a liar for $n = 341$, since $2^{340} = 1 \pmod{341}$ even though it is not a prime number: $341 = 11 \cdot 31$
- 3 is a witness for 341 since $3^{340} = 56 \pmod{341}$
- Because of the existence of Fermat liars, the converse of Fermat's Little Theorem is not true: The condition that $a^{n-1} = 1 \pmod{n}$ does not imply that n is prime
- However, if n is a composite number, then there exists some Fermat's witness a for n

The Fermat Test

FERMAT(n)
Input: $n \geq 3$ is an odd integer
Step 1: Randomly choose a in the range $a \in [2, n-2]$
Step 2: $x := a^{n-1} \pmod{n}$
Step 2: `if` $x \neq 1 \pmod{n}$ `return` "n is composite"
`else return` "n is prime"

- Fermat's test is a randomized algorithm
- If the Fermat test gives the answer "n is composite", the number n is composite indeed
- However, if the Fermat test gives the answer "n is prime", the number n may or may not be prime, as there are Fermat's liars

The Fermat Test

- Consider $n = 143$ which is a composite number $143 = 11 \cdot 13$
- The table below shows Fermat's witnesses and liars for 143

Multiples of 11	11	22	33	44	55	66	77	88	99	110	121	132
Multiples of 13	13	26	39	52	65	78	91	104	117	130		
Fermat witnesses in $\mathbb{Z}_{143}^*$	2	3	4	5	6	7	8	9	10	14	15	16
	17	18	19	20	21	23	24	25	27	28	29	30
	31	32	34	35	36	37	38	40	41	42	43	45
	46	47	48	49	50	51	53	54	56	57	58	59
	60	61	62	63	64	67	68	69	70	71	72	73
	74	75	76	79	80	81	82	83	84	85	86	87
	89	90	92	93	94	95	96	97	98	100	101	102
	103	105	106	107	108	109	111	112	113	114	115	116
	118	119	120	122	123	124	125	126	127	128	129	133
	134	135	136	137	138	139	140	141				
Fermat liars	1	12	131	142								

The Fermat Test

- If we run the Fermat test on 143, the probability that it answers "n is composite" is $138/140 \approx 0.9857$, since there are only two (non trivial) Fermat liars
- In other words, the Fermat witnesses outnumber the Fermat liars clearly in this example
- If this were true for all odd composite numbers, we would have a no-biased Monte Carlo algorithm for the primality problem
- A no-biased Monte Carlo algorithm **always** gives correct "no" answers, but perhaps incorrect "yes" answers
- Unfortunately, if n is composite, the Fermat test does not say so with probability at least $1/2$ for **each given** n

Carmichael Numbers

- There exist composite numbers n for which all elements of $\mathcal{Z}_n^*$ are Fermat liars
- Such numbers are called Carmichael numbers
- The smallest Carmichael number: $561 = 3 \cdot 11 \cdot 17$
- The next 6 Carmichael numbers are 1105, 1729, 2465, 2821, 6601, 8911
- Note that Carmichael numbers have Fermat witnesses in $\mathcal{Z}_n - \mathcal{Z}_n^*$
- It was proven in 1994 by Alford, Granville, and Pomerance that there are infinitely many Carmichael numbers: Specifically they proved that there are at least $\sqrt[7]{n^2}$ Carmichael numbers between 1 and n
- Carmichael numbers have at least 3 prime factors

The Fermat Test

- Theorem: If $n \geq 3$ is an odd composite number that has at least one Fermat witness in $\mathcal{Z}_n^*$, then the Fermat test on input n gives the correct answer "n is composite" with probability at least $1/2$
- This theorem says that for many composite numbers (except Carmichael numbers) the Fermat test has a good probability bound
- The reason why the Fermat test is not a Monte Carlo algorithm for "is n prime?" problem is that $\mathcal{Z}_n^*$ contains too many Fermat liars for infinitely many numbers n, namely Carmichael numbers
- Given a Carmichael number n as input, the Fermat test gives the wrong answer "n is prime" with probability

$$\frac{\phi(n)}{n} \approx \prod(1 - \frac{1}{p}) \lesssim 1$$

The Miller-Rabin Test

MILLER-RABIN(n)

Input: $n \geq 3$ is odd, such that $n - 1 = 2^k \cdot m$, for odd m

Step 1: Randomly a in the range $a \in [1, n-1]$

Step 2: $x := a^m \pmod{n}$

Step 3: `if` $x = 1 \pmod{n}$ `return` "n is prime" and `halt`

Step 4: `for` $j = 0, 1, \ldots, k-1$

Step 5: `if` $x = -1 \pmod{n}$, `return` "n is prime" and `halt`
`else` $x := x^2 \pmod{n}$

Step 6: `return` "n is composite" and `halt`

The Miller-Rabin Example

- $n = 561$ implies $n - 1 = 560 = 2^4 \cdot 35$, thus $k = 4$ and $m = 35$
- Pick $a = 2$ and compute $x := 2^{35} = 263 \pmod{561}$; $x \neq 1$
- $j = 0 \rightarrow x \neq -1 \pmod{561}$; $x := 263^2 = 166 \pmod{561}$
- $j = 1 \rightarrow x \neq -1 \pmod{561}$; $x := 166^2 = 67 \pmod{561}$
- $j = 2 \rightarrow x \neq -1 \pmod{561}$; $x := 67^2 = 1 \pmod{561}$
- $j = 3 \rightarrow x \neq -1 \pmod{561}$; $x := 1^2 = 1 \pmod{561}$
- Therefore, n is composite

Square Roots of 1 Mod n

- An element $x \in \mathcal{Z}_n$ is a quadratic residue mod n if and only if there is some $a \in [1, n)$ such that $x = a^2 \pmod{n}$
- For example, 3 is quadratic residue mod 11 since $3 = 5^2 \pmod{11}$
- If $x = 1$, then a is said to be square root of 1 mod n
- Trivially, 1 and -1 are always square roots of 1 mod m since $1^2 = 1 \pmod{n}$ and $(n-1)^2 = (-1)^2 = 1 \pmod{n}$
- The prime number 23 has 2 square roots of 1, namely 1 and 22
- The composite number $143 = 11 \cdot 13$ has 4 square roots of 1, namely 1, 12, 131, and 142

Square Roots of 1 Mod n

- Theorem: Every prime number n has only two trivial square roots of 1 mod n, namely $\pm 1 \pmod{n}$
- Hence, if n has a nontrivial (other than ± 1) square root of 1, then n must be composite
- If $n = p_1 p_2 \cdots p_k$ is composite, where $p_i > 2$ are prime numbers then the Chinese Remainder Theorem can be used to show that n has exactly 2^k square roots of 1 mod n
- The square roots of 1 mod n are all numbers $a \in [1, n)$ such that $a = \pm 1 \pmod{p_i}$ for $i = 1, 2, \ldots, k$
- Unless n has extraordinarily many prime factors, we cannot find nontrivial square roots of 1 mod n by picking random numbers a

Miller-Rabin Witnesses and Miller-Rabin Liars

- Let $n \geq 3$ be any odd number and $a \in \mathcal{Z}_n^*$, and express $n - 1 = 2^k \cdot m$ with m is odd
- We say a is a **Miller-Rabin witness** for n if and only if none of the following are true

$$\begin{aligned} a^m &= 1 \pmod{n} \\ a^{2^j m} &= -1 \pmod{n} \quad \text{for} \quad \exists j \in \{0, 1, \ldots, k-1\} \end{aligned}$$

- We say a is a **Miller-Rabin liar** for n if and only if n is a composite number and a is not a Miller-Rabin witness

Miller-Rabin Witnesses and Miller-Rabin Liars

- Consider the Carmichael number $n = 561 = 3 \cdot 11 \cdot 17$
- We have $n - 1 = 560 = 2^4 \cdot 35$, and thus $k = 4$ and $m = 35$
- By enumeration, we can show that 561 has 4 Miller-Rabin liars: 1, 50, 101, and 460, i.e., for each one of these numbers a either the first condition or the second condition is satisfied

a	a^{35}	a^{70}	a^{140}	a^{280}	a^{560}
1	1	1	1	1	1
50	-1	1	1	1	1
101	-1	1	1	1	1
460	1	1	1	1	1

The rest of numbers in $\mathcal{Z}^*_{561}$ are all Miller-Rabin witnesses

The Miller-Rabin Test

- Theorem: If there exists a Miller-Rabin witness for n, then n is composite
- Theorem: If $n \geq 3$ is an odd composite number, then there are at most $\frac{n-1}{4}$ Miller-Rabin liars
- Theorem: The Miller-Rabin Test has an error probability of at most $1/4$
- The Miller-Rabin test is very efficient and has a very good probability bound — it is the preferred algorithm for generating large primes used in the RSA algorithm, the Diffie-Hellman key exchange algorithm, or any of the public-key cryptographic protocols where large primes are needed
- There is another probabilistic algorithm for primality testing, called Solovay-Strassen test, however, it is less efficient and less accurate, and therefore, less popular