RSA Encryption and Prime Factorization

People have needed to communicate information secretly for almost as long as we've been around. We can easily see how this can benefit finance or military, but it's even used in our day-to-day as computers communicate with each other. The earliest form of cryptography used what are known as **symmetric-key ciphers**, where two parties had access to a secret key that could both encrypt and decrypt messages. Of course, this requires the parties to have a way to communicate secretly in the first place. As technology advanced, the need for more sophisticated methods became necessary.

The RSA Cryptosystem—named after Ron Rivest, Adi Shamir, and Len Adleman, the first to publish[1] this method—is what is known as a **asymmetric-key cipher**, where everyone is allowed to encrypt with the public key, but only the holder of the private key can decrypt, making it great for one-way communications! While relatively new, it is built on notions, theorems, and work that has long existed in mathematics (we've covered most of it in class!).

RECALL: The **unit group** of n is the set $\mathbb{Z}_n^\times := \{a \in \mathbb{Z}_n \mid a \text{ is a unit in } \mathbb{Z}_n\}$.

RECALL: Euler's phi function satisfies the following properties:

(1) If p is prime and n is a positive integer, then $\phi(p^n) = p^{(n-1)}(p-1)$.
(2) If m, n are positive coprime integers, then $\phi(mn) = \phi(m)\phi(n)$.

(1) Generating an RSA Key:
   (a) Let $p = 47$ and $q = 59$. Calculate $n = pq$ and find $\phi(n)$.
   (b) Let $e = 17$. Explain why e has an inverse modulo $\phi(n)$.
   (c) Find $d = e^{-1} \pmod{\phi(n)}$.

(a) $n = 47 \cdot 59 = 2773$. By the proposition, $\phi(2773) = \phi(47) \cdot \phi(59) = (47-1)(59-1) = 2668$.
(b) 17 has an inverse modulo $\phi(n)$ if and only if $\gcd(17, \phi(n)) = 1$. 17 is prime, and 17 does not divide 2668, so 17 has an inverse.
(c) We apply the Euclidean Algorithm to find the inverse of 17:

$$2668 = 17 \cdot 156 + 16$$
$$17 = 16 \cdot 1 + 1$$

Thus we find after algebra that $1 = 17 \cdot 157 + 2668(-1)$, and so $d = 157$.

(2) Encoding and Encrypting:
   (a) Encode the message "HI" into an integer m by converting the letters into numbers according to the table below and concatenating them in order.[2]

00	A 01	B 02	C 03	D 04	E 05	F 06	G 07	H 08
I 09	J 10	K 11	L 12	M 13	N 14	O 15	P 16	Q 17
R 18	S 19	T 20	U 21	V 22	W 23	X 24	Y 25	Z 26

   (b) Find $y \equiv m^e \pmod{n}$.

[1]Clifford Cocks, an English mathematician, had actually developed a version of this four years prior, but he didn't think it was worth publishing!
[2]For example, "DOG" becomes $041507 = 41507$.

(a) $H = 08$ and $I = 09$, so our integer is 809.
(b) $809^{17} \equiv 522 \pmod{2773}$. If we wish to do this by hand, we could first calculate $809^2 \pmod{2773}$, then 809^4, 809^8, 809^{16}, and finally 809^{17}.

(3) Decoding and Decrypting:
  (a) Find $x \equiv y^d \pmod n$ using any techniques[3] from class.
  (b) Decode x into a message by reversing the encoding in (2a).
  (c) Explain why $m^{ed} \equiv m \pmod n$.
  (d) Encode the message "CAT" as an integer m, then find and compare $y \equiv m^{17} \pmod{2773}$ and $x \equiv y^{157} \pmod{2773}$. Explain why $x \neq m$.

(a) We can use the Chinese Remainder Theorem to solve the system of congruences:

$$x \equiv 522^{157} \pmod{47}$$
$$x \equiv 522^{157} \pmod{59}$$

(since this forms a unique congruence class modulo n.) We can reduce 522 modulo 47 and 59 to 5 and 50 respectively; we also know that $\phi(47) = 46$ and $\phi(59) = 58$, and that $5^{46} \equiv 1 \pmod{46}$ and $50^{58} \equiv 1 \pmod{59}$. Thus we can instead solve:

$$x \equiv 5^{19} \pmod{47}$$
$$x \equiv 50^{41} \pmod{59}$$

We can solve this system of congruences using techniques from class and find $x = 809$.
(b) This decodes into the original "HI".
(c) Since $ed \equiv 1 \pmod{\phi(n)}$, $ed = \phi(n) \cdot k + 1$ for some integer k. Recall by Euler's Theorem that $a^{\phi(n)} \equiv 1 \pmod n$, so we have $m^{ed} \equiv m^{\phi(n)\cdot k+1} \equiv m^{\phi(n)\cdot k} \cdot m \equiv 1^k \cdot m \equiv m \pmod n$.
(d) The result is 88. Actually, $x \equiv m \pmod n$, but since $m \geq n$, $m \neq x$.

(4) Creating your own key-pair:
  (a) Choose two large primes and compute $n = p \cdot q$ and $\phi(n)$.
  (b) Choose any $0 < e < \phi(n)$ in $\mathbb{Z}_n^\times$.
  (c) Write your n and e on the board; these make up your public key.
  (d) Find $d = e^{-1} \pmod{\phi(n)}$.

Results depend on choice of primes p and q and public key e.

(5) Sending messages[4]:
  (a) Find another group to exchange messages with. Come up with a message m and encrypt it using that group's n and e. Write your encrypted message on the board.
  (b) Once the other group has written their encrypted message for you on the board, decrypt it and see what they sent.
  (c) Pick any group's message on the board and see if you can decrypt it, using any techniques. What do you need to know before you can decrypt the message?

Results depend on choice of primes p and q, public key e, and message m.

For (5c), we need to find p and q in order to determine the private key d; the specific result will vary.

[3] HINT: Try using the Chinese Remainder Theorem to work with smaller numbers.
[4] If at any point you're waiting, work ahead on future problems!

(6) Factoring by **Trial Division**:

(a) Let $n = 1643$ be the product of two primes. Factor n by brute force, i.e., attempt to divide by each[5] prime up to n.

(b) There is a $200,000 cash reward for factoring a 617-digit product of two primes. Explain why this is unreasonable to do by Trial Division.

> (a) The factors are 31 and 53.
>
> (b) Based on prime approximations, we would expect to test roughly 10^{306} primes. If we could test 1,000,000 primes per second, it would still take 10^{293} years!

> THEOREM: If $a^2 \equiv b^2 \pmod{n}$, then $\gcd(a+b,n)\cdot\gcd(a-b,n) = n$. Furthermore, if $a \not\equiv \pm b \pmod{n}$, then $\gcd(a+b,n)$ and $\gcd(a-b,n)$ are non-trivial factors of n.

(7) Factoring by the **Continued Fraction Algorithm**:

(a) Let $n = 3053$ be the product of two primes. Find the **factor base** of n: the set of positive primes[6] $q_i \leq 7$ where $\left(\frac{n}{q_i}\right) = 1$.

(b) Check[7] that each element in the factor base is not a prime factor of n.

(c) Find the first[8] 5 convergents $C_k = \frac{p_k}{q_k}$ of $\sqrt{n}$. For each of these, compute $a_k \equiv p_k \pmod{n}$ and $b_k \equiv p_k^2 \pmod{n}$.

(d) Write each b_k as a product of primes in the factor base, if possible[9]. Find a nonempty set of pairs $(a_i, b_i), \ldots, (a_j, b_j)$ such that $b_i \cdots b_j$ is trivially a square modulo n and

$$a_i \cdots a_j \not\equiv \pm\sqrt{b_i \cdots b_j} \pmod{n}$$

.

(e) Let $A \equiv a_i \cdots a_j \pmod{n}$ and $B \equiv \sqrt{b_i \cdots b_j} \pmod{n}$. Calculate and compare $A^2 \pmod{n}$ and $B^2 \pmod{n}$.

(f) Apply the Theorem, and use the Euclidean Algorithm to find the prime factors of n.

> (a) The primes in range are 2, 3, 5, and 7. Of these, 3053 is a square modulo 2 and 7, thus the factor base is $\{2, 7\}$.
>
> (b) We can see that 2 does not divide 3053, and by the Euclidean Algorithm $3057 = 7 \cdot 436 + 1$ and is not a divisor.
>
> (c) Apply the Continued Fraction Algorithm:

k	β_k	α_k
0	$\sqrt{3053}$	55
1	≈ 3.93	3
2	≈ 1.06	1
3	≈ 15.03	15

k	β_k	α_k
4	$\cdots$	27
5	$\cdots$	1
6	$\cdots$	1

[5] HINT: Start by determining a reasonable upper bound for the smallest prime factor of n, and then divide and conquer.

[6] The upper bound of 7 was not arbitrary; $7 = \lfloor e^{\frac{1}{2}\sqrt{\ln(n)\ln(\ln(n))}} \rfloor$.

[7] If an element were to be a factor of n, then we can reduce n by that factor and try again.

[8] This choice was arbitrary. If we wish to do this in general, we'll take one convergent at a time until we find a solution.

[9] If b_k isn't possible, try $-b_k = (-1)p_1^{e_1}p_2^{e_2}\cdots p_k^{e_k}$.

We then evaluate the fractions:

$$C_0 = \frac{55}{1} \qquad C_1 = \frac{166}{3} \qquad C_2 = \frac{221}{4}$$

$$C_3 = \frac{3481}{63} \qquad C_4 = \frac{94208}{1705} \qquad C_5 = \frac{97889}{1768}$$

We then have:

k	p_k	a_k	b_k
0	55	55	-28
1	166	166	79
2	221	221	-7
3	3481	428	4
4	94208	2618	-61
5	97889	3046	49

(d) b_1 and b_4 cannot be written as a product of primes in the factor base, so we will not consider them. Of the remainder, we have:

$$b_0 = (-1) \cdot 2^2 \cdot 7 \qquad b_2 = (-1) \cdot 7^1 \qquad b_3 = 2^2 \qquad b_5 = 7^2$$

Any of the following can form trivial squares work:

i. $\{\ \big(55,\ (-1 \cdot 2^2 \cdot 7)\big),\ \big(221,\ (-1 \cdot 7)\big)\ \}$

ii. $\{\ (428,\ 2^2)\ \}$

- $\{\ (3046,\ 7^2)\ \}$: Since $3046 \equiv \pm 7 \pmod{3053}$, we discard this one.

iii. $\{\ (428,\ 2^2)\ (3046,\ 7^2)\ \}$

- $\{\ \big(55,\ (-1 \cdot 2^2 \cdot 7)\big),\ \big(221,\ (-1 \cdot 7)\big),\ (428,\ 2^2)\ \}$: $28 \equiv \pm 28$, so we discard.
- $\{\ \big(55,\ (-1 \cdot 2^2 \cdot 7)\big),\ \big(221,\ (-1 \cdot 7)\big),\ (428,\ 2^2),\ (3046,\ 7^2)\ \}$: $2857 \equiv \pm 196$, discard.

Further solutions will consider (i), but all of them will work.

(e) With (i), we find $A \equiv 2996$, $B \equiv 14$. We confirm that $A^2 \equiv 196 \equiv B^2 \pmod{3053}$.

(f) The Theorem tells us that we will get nontrivial factors of 3053 by calculating $\gcd(A + B, 3053) = \gcd(3010, 3053)$ and $\gcd(A - B, 3053) = \gcd(2982, 3053)$. Applying the Euclidean Algorithm:

$$3053 = 3010 \cdot 1 + 43 \qquad 3053 = 2982 \cdot 1 + 71$$

$$3010 = 43 \cdot 70 \qquad 2982 = 71 \cdot 42$$

A quick check reveals that $43 \cdot 71 = 3053$!