1 Modern Cryptography

(Continued from previous notes.)

1.1 RSA

RSA is a cryptosystem named after Ron Rivest, Adi Shamir, and Leonard Adleman, who publicly described the algorithm in 1977. The GCHQ mathematician Clifford Cocks developed an equivalent system back in 1973, but his work was not declassified until 1997.

1.2 Converting Text Messages to Numbers

We first need to talk about how to convert text messages into integers, since RSA only allows us to transfer integers. A variety of methods could be employed to make this happen, but one simple idea is for someone to encode a message in the usual way (remove all non-alphabet characters and capitalize everything) and regard the resulting string as a number in base 26.

(Example.) Suppose Alice has a message `HIBOB`. Using the usual letter-to-number correspondence (where `A` is `0`, `B` is `1`, and so on), these numbers correspond to `7`, `8`, `1`, `14`, `1`, in that order. Then, we can construct the base 26 integer,

$$1 \cdot 26^0 + 14 \cdot 26^1 + 1 \cdot 26^2 + 8 \cdot 26^3 + 7 \cdot 26^4 = 3340481_{26}.$$

This is an **integer representation** of the message `HIBOB` in the sense that there is a straightforward algorithm to recover the plaintext; we can use the same algorithm we used to write a number, but dividing repeatedly by 26 instead, to recover the letter-to-number correspondence:

$$\begin{aligned} 3340481 &= 128480 \cdot 26 + 1 \\ 128480 &= 4941 \cdot 26 + 14 \\ 4941 &= 190 \cdot 26 + 1 \\ 190 &= 7 \cdot 26 + 8 \\ 7 &= 0 \cdot 26 + 7. \end{aligned}$$

Looking at the remainders, from bottom to top, gives us `7 8 1 14 1`, which is exactly the correspondence.

(Exercise.)

- Find the integer representation of `GAIA`.

  Each letter in `GAIA` corresponds to the numbers `6 0 8 0`, respectively. So,

  $$0 \cdot 26^0 + 8 \cdot 26^1 + 0 \cdot 26^2 + 6 \cdot 26^3 = 105664_{26}.$$

- Find the text corresponding to the integer `245405438`.

We have

$$\begin{aligned} 245405438 &= 26 \cdot 9438670 + 18 \\ 9438670 &= 26 \cdot 363025 + 20 \\ 363025 &= 26 \cdot 13962 + 13 \\ 13962 &= 26 \cdot 537 + 0 \\ 537 &= 26 \cdot 20 + 17 \\ 20 &= 26 \cdot 0 + 20. \end{aligned}$$

Looking at the remainders from bottom to top gives us `20 17 0 13 20 18`, which is `URANUS`.

(Exercise.) Let's say you wanted to preserve spaces in your message. How would you modify the above method of associating an integer to text to make this happen?

Instead of base 26 (which only allows for all 26 capital letters), we can use base 27, where numbers 0-25 corresponds to a letter and number 26 corresponds to a space. Then, the usual process of converting text to integer and integer to text is exactly the same.

1.3 How RSA Works

Bob starts by picking two distinct large integers p and q. He computes $n = pq$ and $\phi(n) = (p-1)(q-1)$. He picks a random integer d between 0 and $\phi(n)$ such that $\gcd(d, \phi(n)) = 1$. He then computes $e \equiv d^{-1} \pmod{\phi(n)}$ (recall that this is the *inverse* of d mod $\phi(n)$.), for example by using the extended Euclidean algorithm. He then publishes the pair (n, e) for the world to see; this is his **public encryption key**. He keeps the remaining numbers private.

Suppose Alice has a secret integer m between 0 and $n-1$ and she wants to send that integer to Bob. She encrypts m by computing $c = m^e \pmod{n}$, and this is the ciphertext that she sends to Bob.

When Bob receives c, he computes $c^d \pmod{n}$. As it turns out, this must be m, so he has received Alices' message.

(Exercise.) Suppose Bob picks the primes $p = 3$ and $q = 5$. We have $n = pq = 15$ and $\phi(n) = (p-1)(q-1) = 8$. Suppose Bob further chooses $d = 3$ (which is relatively prime to 8).

1. What is Bob's public encryption key?

We have n, so we need to find e. To do so, we make use of the Modular Inversion Theorem to find the inverse of 3 mod 8, which in particular means we need to find the Bezout coefficients. We begin by computing $\gcd(3, 8)$ (trivially, we know this is 1, but we'll still do it);

a	**b**	$\mathbf{b = aq + r}$	**q**	**r**
3	8	$8 = 3q + r$	2	2
2	3	$3 = 2q + r$	1	1
1	2	$2 = 1q + r$	2	0

Here, we find that $\gcd(3, 8) = 1$. Notice that the operations we performed are

- (Eq. 1) $8 = 3 \cdot 2 + 2 \implies 2 = 8 + 3(-2)$
- (Eq. 2) $3 = 2 \cdot 1 + 1 \implies 1 = 3 + 2(-1)$
- (Eq. 3) $2 = 1 \cdot 2 + 0$

So, working backwards from the last equation with a remainder, we have

$$\begin{aligned} 1 &= 3 + 2(-1) \\ &= 3 + \underbrace{(8 + 3(-2))}_{\text{Eq. 1}}(-1) \\ &= 3 + 8(-1) + 3(-2)(-1) \\ &= 3 + 8(-1) + 3(2) \\ &= 3(3) + 8(-1). \end{aligned}$$

From this, it follows that the Bezout coefficients are 3 and -1. In particular, we find that $x = 3$ and so 3 is the inverse of 3 mod 8. Therefore, Bob's public encryption key is $(n, e) = \boxed{(15, 3)}$.

2. Suppose Alice wants to send Bob the message $m = 7$. What is the ciphertext c that Alice sends Bob?

We have
$$c = m^e \pmod{n} = 7^3 \pmod{15} = 13,$$
where $e = 3$ and $n = 15$ from Bob's encryption key.

3. Check that, if Alice sends the ciphertext c corresponding to $m = 7$ to Bob, that Bob actually recovers the original plaintext.

Bob computes
$$c^d \pmod{n} = 13^3 \pmod{15} = 7,$$
which is exactly the message that Alice sent.

4. Suppose Alice sends the ciphertext $c = 2$ to Bob. What is the corresponding plaintext?

Bob again computes
$$2^3 \pmod{15} = 8.$$

(Exercise.) Let's now take Eve's perspective to see why choosing large primes is crucial. Suppose Bob's RSA public key is $(35, 7)$ and Alice has just sent Bob the ciphertext $c = 17$. What is Bob's decryption key? What is Alice's plaintext message?

We know that Bob's public key is $(n, e) = (35, 7)$. So,

$$n = pq \implies 35 = 5 \cdot 7.$$

Therefore, $p = 5$ and $q = 7$. With this in mind, we know that

$$\phi(n) = \phi(35) = (5-1)(7-1) = 24.$$

With this in mind, we want to find the inverse of e mod $\phi(n)$; that is,

$$d \equiv e^{-1} \pmod{\phi(n)}.$$

Once again, we need to use the Modular Inversion Theorem to find the inverse of 7 mod 24. Let's begin by finding $\gcd(7, 24)$;

a	**b**	**b = aq + r**	**q**	**r**
7	24	$24 = 7q + r$	3	3
3	7	$7 = 3q + r$	2	1
1	3	$3 = 1q + r$	3	0

We find that $\gcd(7, 24) = 1$ and, more importantly, the equations used to get us to this value are

- (Eq. 1) $24 = 7(3) + 3 \implies 3 = 24 + 7(-3)$
- (Eq. 2) $7 = 3(2) + 1 \implies 1 = 7 + 3(-2)$
- (Eq. 3) $3 = 1(3) + 0$.

Starting from the last equation with a remainder and working backwards, we have

$$\begin{aligned} 1 &= 7 + 3(-2) \\ &= 7 + (\underbrace{24 + 7(-3)}_{\text{Eq. 1}})(-2) \\ &= 7 + 24(-2) + 7(-3)(-2) \\ &= 7 + 24(-2) + 7(6) \\ &= 7(7) + 24(-2). \end{aligned}$$

In particular, we find that the inverse of 7 mod 24 is 7, so $d = 7$ and this is Bob's decryption key.

With this decryption key in hand, we can decrypt Alice's plaintext message:

$$17^7 \pmod{35} = 3.$$

1.4 Why RSA Works

Theorem 1.1: RSA

Suppose p, q are distinct primes and $n = pq$, that d is an integer with $1 \leq d \leq \phi(n)$ and $\gcd(d, \phi(n)) = 1$, and that $e \equiv d^{-1} \pmod{\phi(n)}$. If $0 \leq m \leq n - 1$ and $c = m^e \pmod{n}$, then

$$c^d \pmod{n} = m.$$

1.5 Why RSA is Probably Secure

If Eve is eavesdropping on Alice and Bob's communication, she knows Bob's public key (n, e) and she sees Alice's ciphertext c. She knows that c is the eth power mod n of Alice's original message m, so the security of RSA relies on the presumed difficulty of the following problem:

(Discrete Root Problem.) Suppose you are given positive integers n, e, and c, and you know further that

- n is a product of two distinct primes (but you don't know which ones),
- e is invertible mod $\phi(n)$ (but you don't know which $\phi(n)$), and
- c is an eth power mod n.

Find the unique eth root of c mod n, i.e., the unique integer m such that $0 \leq m \leq n-1$ such that $m^e \equiv c \pmod{n}$.

There is most likely no fast way of doing this, except for the way that Bob uses, which requires some secret knowledge. Bob needs to know the decryption exponent d, which is an inverse of e mod $\phi(n)$, and knowing that would seem to require knowing $\phi(n)$. The following lemma shows that knowledge of $\phi(n)$ is actually equivalent to a knowledge of a factorization of n, which is believed to be hard to find quickly.

Lemma 1.1

Suppose p and q are distinct primes and $n = pq$. If Eve knows n and can quickly calculate $\phi(n)$¡ then she can also quickly find p and q.

In other words, the lemma tells us that, since it is believed that there is no fast factoring algorithm for classical (ie, non-quantum) computers, this tells us that Eve probably does not have a quick way of finding the decryption exponent d in the same way that Bob does.