MA294 Lecture

Timothy Kohl

Boston University

January 25, 2024

RSA Encryption

For a modulus $m > 1$ we showed for $a \in U(m)$ that $a^{\phi(m)} \equiv 1 \ (mod \ m)$. A simple consequence of this is obtained if we multiply both sides by a, namely that

$$a^{\phi(m)+1} \equiv a \ (mod \ m)$$

and, in fact, this theorem is true even if $a \in \mathbb{Z}_m$ and not just in $U(m)$, which is particularly useful in its application to cryptography.

One of the other preparatory points to note/recall is that $\phi(rs) = \phi(r)\phi(s)$ if $gcd(r, s) = 1$, and in particular, if p and q are distinct primes, then certainly $gcd(p, q) = 1$ and so $\phi(pq) = \phi(p)\phi(q) = (p - 1)(q - 1)$.

The other operational fact, which is also at the heart of the security of the RSA algorithm we'll present, is that while it's easy to multiply two large numbers, it is computationally quite difficult to take a large number and find its constituent factors.

So, in particular, if p and q are 'large' primes then $m = pq$ is an even larger number, and if one were given such a large number without knowing 'p' and 'q' then it would be quite difficult to factor m to find p and q.

Indeed, it is the relatively difficulty of factorization, as we'll see, which is what makes the RSA algorithm secure.

And this also touches on the subject of 'quantum computers' (outside the scope of this course) which are basically computer systems for which factorization is feasible in a 'reasonable' amount of time, as compared with the 'ordinary' computers of today.

The one saving factor though, is that these quantum computers are extremely difficult to build and work with, and it is generally accepted that they haven't been used for this purpose...yet!

RSA Encryption

- Let $m = pq$ where p,q are 'big' primes. (Keep p,q secret.)
- Choose k such that $gcd(k, \phi(m)) = 1$.
- Find 'j' such that $kj \equiv 1 \ (mod \ \phi(m))$, i.e. $kj = t\phi(m) + 1$ (keep j secret!)
- Publish k and m. (This is the 'public' part of public-key encryption.)
- If a sender has a message, representable as a number a, where presumably $a < m$ then they compute a^k mod m which is the encrypted message they transmit.
- The recepient then takes a^k and computes $(a^k)^j$.

The end result is that

$$\begin{aligned} a^{kj} = a^{t\phi(m)+1} &= (a^{\phi(m)})^t a \\ &\equiv 1 \cdot a \ (mod \ \ m) \\ &\equiv a \ (mod \ \ m) \end{aligned}$$

So the recepient now has the message the sender wanted to transmit, the number a.

For a bit of history, the earliest versions used numbers $m = pq$ of around 100 digits which is about 330 binary digits.

[Indeed sometimes these numbers are given in units of how long they are in binary digits.]

And there was a challenge early on posed by the developers (R.S.A.) to see who could factor some large numbers, and as time passed, and techniques and computing power increased, the bit length had to be increased to continue the challenge.

There are many resources online describing this, which one may readily find.

Latin Squares

Consider $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ and all the possible $i + j$ for $i, j \in \mathbb{Z}_4$.

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

where we emphasize how the table entries in the 'inner' 4×4 square are filled in, based on adding the entries in the row and side column, e.g. $1 + 2 = 3$.

If we remove the top row, and left column we get the table:

0	1	2	3
1	2	3	0
2	3	0	1
3	0	1	2

and we observe the most important feature, namely that each row and each column contains exactly one element from $\{0, 1, 2, 3\}$ with no repeats.

What does this remind you of?

Definition

A latin square of order n is an $n \times n$ array in which each one of n symbols occurs once in each row and once in each column.

A more interesting example, for $n = 6$.

A	B	C	D	E	F
B	A	E	F	C	D
C	F	A	E	D	B
D	E	F	A	B	C
E	D	B	C	F	A
F	C	D	B	A	E

For a latin square L one may define $L(i,j)$ to mean the entry at the i-th row and j-th column.

So for above we have $L(2,3) = E$, $L(4,6) = C$ and we can fram the latin square condition in terms of the $L(i,j)$.

Namely L is a latin square iff $L(i,j) = L(i,k)$ only if $j = k$ and $L(i,j) = L(k,j)$ only if $i = k$.

Going back to the example with $\mathbb{Z}_4$ earlier, we have the following general statement.

Theorem

For each $m > 1$ the $m \times m$ array defined by $L(i,j) = i + j$ for $i, j \in \mathbb{Z}_m$ is a latin square.

Proof.

If $L(i,j) = L(i,k)$ then $i + j = i + k$, so can we conclude $j = k$ in $\mathbb{Z}_m$? Well, in $\mathbb{Z}_m$, every element as an additive inverse, so in particular

$$\begin{aligned} i + j &= i + k \\ (-i) + i + j &= (-i) + i + k \\ 0 + j &= 0 + k \\ j &= k \end{aligned}$$

and similarly $L(i,j) = L(k,j)$ implies $i = k$. □

So what this shows is that for *every* m there is an $m \times m$ latin square.

However, there are other examples as our 6×6 example earlier demonstrates, but even for 4×4 we can find another (actually natural) example.

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

as compared with the $\mathbb{Z}_4$ example (expressed with letters instead of numbers)

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Before going further we should point out that latin squares have a number of different applications.

One of these is in experimental design.

Suppose you have four test subjects, each of which is questioned separately by four interviewers over four rounds of questioning, and you want to make sure every interviewer interviews one and only one test subject, but that every interviewer interviews every subject.

If the four interviewers are $\{1, 2, 3, 4\}$ and the subjects are A, B, C, D and the 'rounds' $R1, R2, R3, R4$ then a latin square can be used to distribute the subjects between the interviewers in each round.

	1	2	3	4
$R1$	A	B	C	D
$R2$	B	C	D	A
$R3$	C	D	A	B
$R4$	D	A	B	C

This is a fairly obvious plan since the 'cyclic' rotation of subjects A,B,C,D, amongst the interviewers obviously guarntees that the conditions of the experiment are satisfied.

An alternate choice is the other one we saw above.

	1	2	3	4
$R1$	A	B	C	D
$R2$	B	A	D	C
$R3$	C	D	A	B
$R4$	D	C	B	A

A natural question is how many latin squares are there of a given size $n \times n$?

The first observation one can make which helps streamline the discussion is that for any latin square in the set $S = \{x_1, \dots, x_n\}$, one may assume that the first row and the first column are in the same order. i.e.

x_1	x_2	$\dots$	x_n
x_2	$\ddots$		
$\vdots$			
x_n			

and such latin squares are called reduced.

The number of (reduced) latin squares for small n can be determined:
$n = 1 \to 1$

$n = 2 \to 1$ (If the first row and column are given, there is only one way to fill in the remaing cell!)

$n = 3 \to 1$

A	B	C
B	?	?
C	?	?

$n = 4 \to 2$ (the ones we've seen already, and no others)
$n = 5 \to 56$ (This is not a typo!)
$n = 6 \to 9408$ (This is also not a typo!)
$n = 7 \to 16,942,080$ (Also not a typo.)

You get the idea.