

MATH 13150: Freshman Seminar
Unit 18

1. THE RSA ALGORITHM

In this chapter, we'll learn how the RSA algorithm works.

1.1. Bob and Alice. Suppose that Alice wants to send a message to Bob over the internet and she wants to make sure that any hackers who intercept the message will not be able to read it. As is customary in cryptography, she encodes the message using a coding technique, called a key. A new feature is that Bob chooses the coding technique without consulting Alice, and everyone, including the hacker, can know her coding technique. However, Bob has special knowledge which enables him to decode the message. Bob is the only one with this special knowledge, so Bob is the only one who can decode Alice's secret message. As a consequence, Alice cannot decode her own message (not a big deal as long as she kept her original unencoded message). More importantly, a spy named Eve also cannot decode Alice's message.

This procedure is called "asymmetric" (or "not symmetric") because the sender and receiver have different information about the code. This is different from the coding techniques discussed in chapters one through four of the CodeBook, where the sender and receiver have the same information. For example, in the Enigma coding technique, both the sender and receiver have an Enigma machine and the keyword for the day.

RSA ALGORITHM: Let's say that Alice wants to send Bob a secret number called "a". If Alice can send a single number to Bob, then she can send any combination of numbers, and this means she can send any message she wants.

Bob to Alice "Message Coming". This is public information.

Bob picks two prime numbers p and q and computes their product, $N = p \cdot q$. He also computes $\phi(N) = \phi(p \cdot q) = (p - 1) \cdot (q - 1)$. He chooses a number k which is relatively prime to $\phi(N)$.

Bob to Alice "k, N". This is public information, but the primes p and q are private information, known only to Bob, and anyone who can factor N .

Alice computes $b \equiv a^k \pmod{N}$. This encodes the secret number "a" as the new number "b".

Alice to Bob: "b". This is public.

Bob knows that $b \equiv a^k \pmod{N}$, so he knows that $\sqrt[k]{b} \equiv a \pmod{N}$. So he computes $\sqrt[k]{b} \pmod{N}$, and he knows this is the secret number that Alice wanted to send him.

EXAMPLE: Let's see how this works in a small example. Note that the only mathematical operations going on are computing powers and kth roots in modular arithmetic, and we've learned how to do them.

Suppose Alice wants to send the secret number $a = 2$ to Bob.

Alice to Bob "Message Coming". This is public.

Bob chooses $p = 3$ and $q = 11$, and computes $N = p \cdot q = 3 \cdot 11 = 33$. He also computes $\phi(N) = \phi(3 \cdot 11) = 2 \cdot 10 = 20$. He chooses $k = 7$, which is OK because $\gcd(7, 20) = 1$.

Alice to Bob “k=7, N=33”. This is public information.

Alice computes $b \equiv 2^7 \pmod{33}$. An easy computation shows that $2^7 \equiv 29 \pmod{33}$, so $b = 29$. This is the encoded number.

Alice to Bob “b=29”. This is public information.

Now Bob knows that $b \equiv a^7 \pmod{33}$, so $\sqrt[7]{b} \equiv a \pmod{33}$. So he just has to compute $\sqrt[7]{29} \pmod{33}$. Let’s try that out.

We first write 1 as a combination of 7 and 20. A moment’s calculation gives

$$1 = 3 \cdot 7 - 20, \text{ so}$$

$$1 \equiv 3 \cdot 7 \pmod{20} \text{ and}$$

$29 \equiv 29^1 \equiv 29^{3 \cdot 7} \pmod{33}$, using the general rule on page 3 of Unit 15, which comes from Euler’s theorem. Now take the 7th root of each side to get:

$$\sqrt[7]{29} \equiv 29^3 \pmod{33}. \text{ We compute}$$

$29^3 \equiv 2 \pmod{33}$. So Bob knows that the secret number is $a = 2$, and he has successfully decoded Alice’s message.

Let’s think for a moment about the security of this encoding technique. The secret information that only Bob knows is that $p = 3$ and $q = 11$. If a spy is looking at the public information, they will see that $N = 33$, and they will be able to find p and q by taking the prime factorization of 33. So in this case, the secret message is not very well-protected since 33 is easy to factor. We will see later that if we take p and q to be bigger primes, then the secret message is much better protected.

EXAMPLE: Suppose that Alice wants to send to Bob the secret number $a = 4$ and Bob picks the primes $p = 7$ and $q = 11$, and Bob takes $k = 17$. Then after Alice tells Bob that a message is coming, Bob computes

$N = 7 \cdot 11 = 77$, and $\phi(N) = 6 \cdot 10 = 60$. It is OK for Bob to pick $k = 17$ since $\gcd(17, 60) = 1$. Then Bob sends to Alice

“k=17, N=77”, which is public information.

Alice then computes $b \equiv 4^{17} \pmod{77}$, and the modular arithmetic calculator shows that $4^{17} \equiv 16 \pmod{77}$, so $b = 16$. Then Alice sends to Bob the encoded message:

“b=16”, which is public information.

To decode the message, Bob has to compute:

$\sqrt[k]{b} \pmod{N} \equiv \sqrt[17]{16} \pmod{77}$. Since Bob knows that $\phi(77) = 60$, his first step is to write 1 as a combination of 17 and 60, which he does using the Euclidean algorithm as follows:

$$60 = 3 \cdot 17 + 9$$

$$17 = 9 + 8$$

$9 = 8 + 1$, and reversing the steps, we find:

$$1 = 9 - 8 = 9 - (17 - 9) = 2 \cdot 9 - 17, \text{ so}$$

$$1 = 2 \cdot (60 - 3 \cdot 17) - 17 = 2 \cdot 60 - 7 \cdot 17, \text{ so}$$

$$1 \equiv -7 \cdot 17 \pmod{60}, \text{ so}$$

$$1 \equiv 53 \cdot 17 \pmod{60}$$

This implies that:

$16 \equiv 16^1 \equiv 16^{53 \cdot 17} \pmod{77}$ using Euler's theorem, so

$\sqrt[17]{16} \equiv 16^{53} \pmod{77}$. The modular arithmetic calculator shows that:

$16^{53} \equiv 4 \pmod{77}$, so Bob knows the secret number was "a=4", and Bob has successfully decoded Alice's message.

1.2. The Role of Eve. Now we suppose there is a spy whose name is traditionally Eve. Eve can see all of Alice and Bob's public messages. In this section, we'll see how Eve's job works, and how it is more difficult than Bob's job. We'll do this by looking at a series of problems that illustrate the jobs of each of the three characters. These problems are the same as the format as the problems in the homework, and you will have to do problems of this sort on the final exam.

PROBLEM: Alice wants to send a secret number to Bob. If her secret number is $a = 7$ and Bob sends her $k = 5$ and $N = 51$, what should she send Bob?

SOLUTION: Alice's task is to compute $b \equiv a^k \pmod{N}$, so she computes

$b \equiv 7^5 \pmod{51}$. Since $7^2 \equiv 49 \equiv -2 \pmod{51}$,

$7^4 \equiv (-2)^2 \equiv 4 \pmod{51}$, so

$7^5 \equiv 7 \cdot 4 \equiv 28 \pmod{51}$, so Alice should send

"b=28". This solves the problem. You could also do this using a hand-held calculator for the purposes of an exam. It is even easier to do using the modular arithmetic calculator, but you can't use that on the final exam.

PROBLEM: Alice wants to send a secret number to Bob. If her secret number is $a = 8$ and Bob sends her $k = 31$ and $N = 2701$, what should she send Bob?

SOLUTION: She should send Bob $b \equiv 8^{31} \equiv 1176 \pmod{2701}$. So the answer is "b=1176".

I did the last problem using the modular arithmetic calculator, and while you could compute $8^{31} \pmod{2701}$ using a hand-held calculator, you will not have to do such a long computation on the final exam.

NOTE: Alice's job is really easy.

PROBLEM: Suppose Bob is receiving a secret message from Alice and he has chosen $p = 7$, and $q = 13$, and $k = 11$. If he receives "b=2" from Alice, then what is the secret number she is sending him?

SOLUTION: Bob should compute $N = 7 \cdot 13 = 91$ and $\phi(N) = 6 \cdot 12 = 72$. He knows that the secret number $a \equiv \sqrt[k]{b} \pmod{N} \equiv \sqrt[11]{2} \pmod{91}$. To compute a , he first finds 1 as a combination of 72 and 11 using the Euclidean algorithm. The steps are:

$$72 = 6 \cdot 11 + 6$$

$$11 = 6 + 5$$

$6 = 5 + 1$, and when we reverse these, we find:

$$1 = 6 - 5 = 6 - (11 - 6) = 2 \cdot 6 - 11, \text{ so}$$

$$1 = 2 \cdot (72 - 6 \cdot 11) - 11 = 2 \cdot 72 - 13 \cdot 11, \text{ so}$$

$$1 \equiv -13 \cdot 11 \pmod{72}, \text{ so since } -13 \equiv 59 \pmod{72},$$

$1 \equiv 59 \cdot 11 \pmod{72}$. Using the general rule on page 3 of Unit 15 or Euler's theorem, we find:

$$2 \equiv 2^{59 \cdot 11} \pmod{91}, \text{ so}$$

$$\sqrt[11]{2} \equiv 2^{59} \pmod{91}. \text{ We compute}$$

$2^{59} \equiv 46 \pmod{91}$ using the modular arithmetic calculator, so we conclude:

$$\sqrt[11]{2} \equiv 46 \pmod{91}, \text{ so the secret number is}$$

"a=46". You can check this by computing $46^{11} \equiv 2 \pmod{91}$, so "a=46" is the secret number.

PROBLEM: Suppose Eve sees that Bob told Alice that "k=7, N=143" and Alice told Bob that "b=6". What is the secret number "a" that Alice is sending to Bob?

SOLUTION: To find a, Eve uses the knowledge that $b \equiv a^7 \pmod{143}$, to conclude that $a \equiv \sqrt[7]{6} \pmod{143}$. To compute $\sqrt[7]{6} \pmod{143}$, Eve must compute $\phi(143)$. To do this, Eve factors 143. Since $\sqrt{143}$ is about 11.9, Eve tries all primes up to 11. She sees that 2, 3, 5, 7 do not divide 143, but $143 = 11 \cdot 13$. From this, she knows that $\phi(143) = 10 \cdot 12 = 120$. She then writes 1 as a combination of 7 and 120. After a calculation using the Euclidean algorithm, Eve finds that:

$$1 = 120 - 17 \cdot 7, \text{ so}$$

$$1 \equiv -17 \cdot 7 \pmod{120}, \text{ so since } -17 \equiv 103 \pmod{120},$$

$$1 \equiv 103 \cdot 7 \pmod{120}. \text{ From this, she concludes that}$$

$$6 \equiv 6^{103 \cdot 7} \pmod{143}, \text{ so}$$

$$\sqrt[7]{6} \equiv 6^{103} \equiv 7 \pmod{143}. \text{ From this, Eve knows that}$$

The secret message is "a=7". To check this, she computes $7^7 \equiv 6 \pmod{143}$, so she knows her math is correct, and "a=7".

Eve also decides that Bob is lousy at security, since he picked such an easily factored number as 143.

PROBLEM: Suppose Eve sees that Bob told Alice "k=17, N=5671" and Alice told Bob that "b=4". What is the secret number "a" that Alice is sending to Bob?

SOLUTION: Alice knows that $b \equiv a^{17} \pmod{5671}$, so $a \equiv \sqrt[17]{4} \pmod{5671}$, and Alice has to compute "a". To do this, she needs to compute $\phi(5671)$, and to do this, she needs to factor 5671. She computes $\sqrt{5671} = 75. \dots$, so she knows a prime less than 5671 should divide 5671. She tries each of the primes 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, and 47 and finds that none of them divide 5671. She then tries 53 and sees that $5671 = 53 \cdot 107$. From this, she knows that $\phi(5671) = 52 \cdot 106 = 5512$. Her next job is to write 1 as a combination of 17 and 5512, which she does using the Euclidean algorithm. She finds:

$$5512 = 324 \cdot 17 + 4$$

$$17 = 4 \cdot 4 + 1, \text{ so}$$

$$1 = 17 - 4 \cdot 4 = 17 - 4 \cdot (5512 - 324 \cdot 17) = 1297 \cdot 17 - 4 \cdot 5512 \text{ and}$$

$$1 \equiv 1297 \cdot 17 \pmod{5512}, \text{ so}$$

$4 \equiv 4^{1297 \cdot 17} \pmod{5671}$ using the general rule on page 3 of Unit 15 or Euler's theorem, and

$\sqrt[17]{4} \equiv 4^{1297} \equiv 4534 \pmod{5671}$, using her modular arithmetic calculator. She concludes that the secret number is:

“a=4534”. To check this, she compute $4534^{17} \equiv 4 \pmod{5671}$, so she’s got the secret message.

This time, Eve had to work quite a bit harder to find the factorization of 5671, so now Eve is a bit more impressed with Bob.

REMARK: We can also see how Bob’s job is different from Eve’s job. The difference is that Bob knows that $5671 = 53 \cdot 107$ when the problem begins, because Bob has picked the primes $p = 53$ and $q = 107$. The extra problem that Eve must solve that Bob does not need to solve is to factor 5671. As you can see in this last problem, that is not such a big deal, because you can test whether a prime less than 75 divides 5671 in a few minutes. However, if we were to pick $n = 41876041$, it would take you quite a long time to discover that $p = 5323$ and $q = 7867$. In fact, you would have to check more than 500 primes to find this factorization. This would take a person more than an hour, but a computer program could do it in a few seconds. The real security of the RSA algorithm depends on using much bigger primes, and these are chosen and manipulated by computer. In principle, it takes longer than the age of the universe to factor a number that is a product of two primes with 200 digits each.

1.3. Sending and receiving actual messages using RSA. To send a message using RSA, we can assign a number to each letter or number. For example, we can do this using the following table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
T	U	V	W	X	Y	Z	(space)											
21	22	23	24	25	26	27	28											
0	1	2	3	4	5	6	7	8	9									
29	30	31	32	33	34	35	36	37	38									

For example, the letter “A” is translated to “2”. The letter “R” is translated to “19”, and the number “3” is translated to “31”.

This is a relatively simple substitution encryption technique, and it can easily be broken using frequency analysis. Its only function is to translate letters into numbers. We can get a secure code by using RSA to encode each of the numbers from 2 to 38. If we also wanted our text to include punctuation marks, we could assign numbers to them, but let’s avoid this so that things are simpler.

EXAMPLE: Use RSA with $p = 43$, $q = 47$ and $k = 11$ to encode the message “The exam is at 145”.

To do this, first rewrite the message so there are no spaces to get “Theexamisat145”.

Next, change each letter to a number according to the table above. So since “T” is translated as “21”, the first number is “21”. Since “H” is translated as “9” and “h” is

the second letter in the message, the second number is “9”. We put dashes between numbers, and translate the entire message as:

21-9-6-6-25-2-14-10-20-2-21-30-33-34

Now we can encode each of these numbers using RSA. Since $p = 43$ and $q = 47$, $N = p \cdot q = 43 \cdot 47 = 2021$. To encode the first number 21, we just do Alice’s job with “k=11” and “N=2021” and compute

$$21^{11} \equiv 379 \pmod{2021}$$

The second number is encoded as $9^{11} \equiv 298 \pmod{2021}$, and so on. The message encoded via RSA is:

379-298-1283-1283-1156-27-784-445-1910-27-379-1310-673-820

This is the message that the sender would send to the receiver. To decode this message, the receiver would compute $\sqrt[11]{b} \pmod{2021}$ for each number in the message, and then translate the message back into letters and numbers by reversing the table.

For example, to decode the first number 379, the receiver should compute

$$\sqrt[11]{379} \pmod{2021}.$$

Since $2021 = 43 \cdot 47$, $\phi(2021) = 42 \cdot 46 = 1932$. First write 1 as a combination of 1932 and 11 as follows:

$$1932 = 175 \cdot 11 + 7$$

$$11 = 7 + 4$$

$$7 = 4 + 3$$

$$4 = 3 + 1, \text{ which gives}$$

$$1 = 4 - 3 = 4 - (7 - 4) = 2 \cdot 4 - 7 = 2 \cdot (11 - 7) - 7 = 2 \cdot 11 - 3 \cdot 7, \text{ so}$$

$$1 = 2 \cdot 11 - 3 \cdot (1932 - 175 \cdot 11) = 527 \cdot 11 - 3 \cdot 1932, \text{ so}$$

$$1 \equiv 527 \cdot 11 \pmod{1932}, \text{ and}$$

$$379 \equiv 379^1 \equiv 379^{527 \cdot 11} \pmod{2021}, \text{ and}$$

$$\sqrt[11]{379} \equiv 379^{527} \equiv 21.$$

Then we translate “21” as “t” by reversing the table at the beginning of this section.

It is now easy to decode the rest of the message, and requires little extra work. For example, to decode the second number “298” in the encoded message, we compute

$$\sqrt[11]{298} \equiv 298^{527} \equiv 9 \pmod{2021}, \text{ which translates back to “h”}.$$

By following this process, we ultimately get the original message “theexamisat145”.

EXAMPLE: Suppose the receiver has chosen $p = 43$, $q = 47$, and $k = 31$, and has received the encoded message

1571-574-1571-1209-823-1569-1384-1435-1689

What is the original message?

We already computed $N = 2021$ and $\phi(N) = 1932$ in the last example. To decode the message, compute $\sqrt[31]{b} \pmod{2021}$ for each number b in the sequence. We start with $b = 1571$. The first step is to write 1 as a combination of 31 and 1932, which gives:

$$1 = 187 \cdot 31 - 3 \cdot 1932.$$

$$1 \equiv 187 \cdot 31 \pmod{1932}, \text{ so}$$

$$1571 \equiv 1571^1 \equiv 1571^{187 \cdot 31} \pmod{2021}, \text{ so}$$

$$\sqrt[31]{1571} \equiv 1571^{187} \equiv 10 \pmod{2021}. \text{ Since 10 translates as “i”, the first letter is “i”}.$$

Similarly, $\sqrt[31]{574} \equiv 574^{187} \equiv 13 \pmod{2021}$, and 13 translates as “l”, so the second letter is “l”.

Continuing in this way, you see that the message is “likecats”, which you can read as “I like cats”.

I’ll hand out an RSA project, where you’ll get a chance to try to decode a message which gives the answer to the first question on the final exam. In most math classes, the problems on the final exam are a closely guarded secret, but I figure that in a class about the RSA algorithm and coding, it would be a good idea to give you an incentive to decode an actual message. In the project, I’ll pick bigger primes, so it will not be so easy for you to decode the message. For this, it will be very useful for you to have a list of many primes, which you can find at the website:

<http://primes.utm.edu/lists/small/1000.txt>

1.4. Some comments: In this section, I’ll make some comments about the RSA algorithm. You don’t need to know anything from this section for the course.

(1) Even though you weren’t aware of the RSA algorithm before you took this course, you’ve probably used it. Almost any web browser has some version of the RSA algorithm programmed into it. The computer does the calculations for you, so you don’t need to know how to compute k th roots in modular arithmetic. Even now, you’re not going to start performing the RSA algorithm yourself when you’re sending a secure message, because your computer does it much better than you could do it.

(2) Although we have learned some abstract mathematics and seen how it can be applied to a real world problem, I don’t expect you to start using the RSA algorithm in practice. Instead, the purpose of this course is more to give you some idea of how mathematical ideas which look completely abstract and unrelated to real life can be used to solve practical problems. Hopefully, this will give you some appreciation of the power and reach of mathematics. Further, hopefully it will give you the idea that even if you don’t necessarily want to do math, some very interesting and sophisticated mathematical ideas are in reality quite simple. After all, we’ve learned how the RSA algorithm works in a one semester course, and we only needed high school algebra to do this. Many very important mathematical ideas have fairly simple conceptual explanations.

(3) You might wonder how the primes p and q are chosen and how k is chosen. In practice, p and q have to be bigger than any number a that will be sent in the message, and this will ensure that a is relatively prime to $N = p \cdot q$. Since all characters can be expressed with fewer than 100 numbers, any primes bigger than 100 will do. To choose k , we just have to pick a number that is relatively prime to $p - 1$ and $q - 1$. I usually pick k to be prime, since then it is easy to see whether k divides p or q . For example, if $p = 67$ and $q = 103$, then $p - 1 = 66$ and $q - 1 = 102$. The primes dividing either of these numbers are 2, 3, 11 and 17, so if I pick $k = 13$, that’ll work. If p and q are 100 digits long, it may not be so easy to factor $p - 1$ and $q - 1$, but we can just pick a prime k that is 20 digits long, and see if k divides $p - 1$ or $q - 1$. If it does, then we just try another prime in place of k and after a short time, we’ll find one that does not divide $p - 1$ or $q - 1$.

(4) When you do a single decoding problem, like when you're playing the role of Bob or Eve, it looks like as much of the work is involved in writing 1 as a combination of k and $\phi(N)$ as in factoring N . This is misleading. Although the process of writing 1 as a combination of two numbers is somewhat difficult to learn, once you've learned how to do it, it is routine, and it is very easy to write a computer program telling the computer to do it. Further, even if the numbers involved are very big, a computer can do it really quickly, and the computer is not going to make mistakes using the distributive law. On the other hand, there are no really good tricks known for factoring a number N as a product of two primes. The way to do this is to go through a list of primes until you find one that divides N . If N has 300 digits, then there are something on the order of 10^{297} primes to check. Even for a computer, that is going to take a very, very long time to check.

(5) It is interesting to note that although RSA depends on the fact that there is no known procedure for quickly factoring a number as a product of two primes, it is entirely possible that someone will think of one. If someone thought of one, they might very well not tell anyone. They could then intercept many secret messages, and for example, transfer all the money in your bank account to a numbered account in the Cayman islands. A more realistic scenario involves the NSA, or National Security Agency, which is a government agency. The NSA employs many mathematicians and computer scientists who work on coding and decoding problems. If the NSA were to find a way to break the RSA algorithm, they probably wouldn't tell anyone. Instead, they would use their method to intercept all kinds of secret messages sent by other governments and other groups.

(6) In fact, there is in theory a kind of computer called a quantum computer, and a fully operational quantum computer can factor any number quickly, and hence could break RSA. Some primitive quantum computers have been built, but they can only multiply small numbers together at this point. It is not clear whether it is feasible to build a quantum computer. On the other hand, an operational quantum computer could also use principles of quantum mechanics to produce an unbreakable code.

(7) An algorithm like RSA does not need to be perfect to be effective. There are old-fashioned espionage techniques, and you read about them in the CodeBook. If a country, like Poland before World War 2, wants to know some secret information, they can always try to capture or otherwise find someone on the other side, and try to induce them to give up the information. They can also try infiltrating a mole into the other side. As long as these methods are much easier than breaking RSA, RSA will have value.

EXERCISES:

- (1) Suppose Alice wants to send to Bob the secret number "a=5" and Bob has sent Alice "k=13" and "N=201". What is the encoded message that Alice should send Bob?
- (2) Suppose Alice wants to send to Bob the secret number "a=2" and Bob has sent Alice "k=17" and "N=403". What is the encoded message that Alice sends Bob?

- (3) Suppose that Alice has sent Bob the encoded message “ $b=10$ ” and Bob has chosen $p = 7$ and $q = 11$. If $k = 43$, what is Alice’s original message?
- (4) Suppose that Alice has sent Bob the encoded message “ $b=2$ ” and Bob has chosen $p = 7$, $q = 19$ and $k = 5$. What is Alice’s secret message?
- (5) Suppose that Eve sees that Bob has told Alice to use “ $k=13$ ” and “ $N=85$ ”, and Alice has sent Bob the encoded message “ $b=4$ ”. What is the secret message?
- (6) Suppose that Eve sees that Bob has told Alice to use “ $k=13$ ” and “ $N=1121$ ”, and Alice has sent Bob the encoded message “ $b=7$ ”. What is the secret message?
- (7) Suppose that Eve sees that Bob has told Alice to use “ $k=17$ ” and “ $N=123241$ ”, and Alice has sent Bob the encoded message “ $b=21295$ ”. If Eve knows that Alice is sending Bob the time when they will rob a bank, when should she alert the bank to expect the robbers?