Final Project

Cryptography

Runyu Wang

1.2

The most frequently used letters in English alphabet are: E, T, A, O, and I. This information is crucial to cracking Caesar Cryptosystem, especially when all punctuation and spaces are removed. By finding the most frequently used letters in the ciphertext and link it to one of the letters with the highest frequency, we might just crack the code. Or we can find how far these frequently used cipher letters are apart on alphabet, and compare them with the distance of letters among "E, T, A, O, and I."

The distance from A to E is 4. The distance from E to I is also 4. Etc.

1.3

We found that the most frequently used letter in the ciphertext is "K" with 14 appearances and the second most frequently used letter is "O" with 10 appearances "K" and "O" happens to have a distance of 4 on the alphabet. Let's try to link "K" with the known most frequently used letter in English: "E"

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F

We can see that "O" happens to land on "I" as well. Let's try to decrypt with this table:

"amathematicalproblemshouldbedifficultinordertoentice

usyetnotcompletelyinaccessiblelestitmockatoureffortsdavidhilbert"

We can add in spaces and punctuation to make it more readable:

"A mathematical problem should be difficult in order to entice

us, yet not completely inaccessible lest it mock at our efforts. -David Hilbert"

1.4

Now let's associate the alphabet with number from 0-25. We can think of the alphabet as a Complete Residue System (CRS) modulo 26. We have proven in Lesson 5 question 2.13 that if S is a CRS mod m, then S + b is also a CRS modulo m. Here, encryption key is the added number b. It will have a value of 0-25 mod 26, and users can add key value to each of the letter in the plaintext to get the ciphertext. (Of course, the addition I referred to is within modular arithmetic.) And no matter the value, the encrypted alphabet would still be a complete alphabet system

1.5

I agree with my imaginary partner, Michael Jordan, that a=101, c=9901, b=100, and m=1000000. Using this affine cryptosystem, we can send encrypted integers between 0 and 999999 to each other. My plaintext message is N=950115. I would compute $101\times950115+100=95961715\equiv961715(\text{mod } 1000000)$ and send the ciphertext $\widetilde{N}=96171$. Now the GOAT will compute:

$$c\cdot\left(\widetilde{N}-b\right)\equiv 9520950115\equiv 950115(\text{mod } 1000000)$$

which is my original text.

1.7

Consider the encryption function: $\widetilde{N}\equiv a\times N+b(\text{mod } m)$

The plaintext is encrypted by multiplying a then add b mod m. In order to get the original plain text from ciphertext, the first reverse step is subtracting b from $\widetilde{N}$. This is quite straightforward. Now we have $\widetilde{N}-b\equiv a\times N(\text{mod } m)$. We picked c so that $a\cdot c\equiv 1(\text{mod } m)$, meaning that a and c are inverse of each other. We can multiply the both side of the congruence equation by c, and we get

$c\cdot(\widetilde{N}-b)\equiv c\times a\times N(\text{mod } m)$, where $c\cdot a\equiv 1(\text{mod } m)$. Hence, we get $c\cdot\left(\widetilde{N}-b\right)\equiv N\ (\text{mod } m)$. This is how the decryption of this cryptosystem works.

1.8

We know that the Evil Realtor's association is using an affine cryptosystem with m= 100000. We also found out that two most frequently used ciphertexts are "84576" and "33154". We can try to link them to the two known hottest zip codes "49508" and "80922."

Let's still consider this encryption function: $\widetilde{N}\equiv a\times N+b(\text{mod } m)$

<u>Case 1</u> plaintext: "49508"→ ciphertext: "84576"

plaintext: "80922"→ ciphertext: "33154"

$$84576\equiv 49508\cdot a+b\ (\text{mod } 100000)$$

$$33154\equiv 80922\cdot a+b\ (\text{mod } 100000)$$

Subtracting the two congruence equations we get:

$$51422\equiv -31414\cdot a\ (\text{mod } 100000)$$

Equivalent to:

$$51422\equiv 68586\cdot a\ (\text{mod } 100000)$$

We can solve this with Bachet's theorem:

With the help of technology, we can find that GCD (68586,100000) =2

68586÷2=34293

51422÷2=25711

100000÷2=50000

Again, with the help of technology, we can find that the inverse of 34293 mod 50000 is 12157.

a = 12157×25711≡68627 (mod 100000). We can substitute this value back to obtain the value of b=-940≡ 99060.

Therefore, (a, b) = (68627, 99060) is a possible encryption key.

Case 2 plaintext: "49508"→ ciphertext: "33154"

plaintext: "80922"→ ciphertext: "84576"

$$33154 \equiv 49508 \cdot a + b \pmod{100000}$$

$$84576 \equiv 80922 \cdot a + b \pmod{100000}$$

With the same method we get that (a, b) = (81373, 18670) is another possible encryption key.

There are two possible encryption keys because there are two combinations of the most frequently used codes.

2.1

I sent my favorite number N=23 to my pal Michael with Massey-Omura cryptosystem. We agreed on the prime 31. I chose a=7 and c=13, where a× c ≡ 1 (mod 30); Michael chose b=11 and d=11, where b×d ≡ 1 (mod 30). We did not share this information with each other. I computed

$$N_1 = N^a = 23^7 \equiv 29 \pmod{31}$$

And I sent 17 to him. Michael Jordan received 17, and computed:

$$N_2 = N_1^b = 29^{11} \equiv 29 \pmod{31}$$

I received 23, and computed:

$$N_3 = N_2^c = 29^{13} \equiv 23 \pmod{31}$$

Michael receives 23, and computed:

$$N = N_3^d = 23^{11} \equiv 23 \pmod{31}$$

Hence Michael received my favorite number 23, and chose it as his jersey number.

2.2

We can look at the process of the number between Michael and me. I take my number N, and computed N^a; he then took the new number and compute $(N^a)^b$; I computed $((N^a)^b)^c$; Finally Michael computed $(((N^a)^b)^c)^d$. Now let's consider $(((N^a)^b)^c)^d$ modulo p, Michael and I pick (b, d) and (a, c) respectively so that

$$b \times d \equiv 1(\text{mod } (p-1))$$

$$a \times c \equiv 1(\text{mod } (p-1))$$

Given these two equations, there must exist integers m and k, such that

$$b \times d = (p-1) \times m + 1$$

$$a \times c = (p-1) \times k + 1$$

We can re-write $(((N^a)^b)^c)^d$ as $(N^{ac})^{bd}$, which is equivalent to

$$(N^{ac})^{(p-1)m+1} = ((N^{ac})^m)^{(p-1)} \times N^{ac}$$

We have proven in lesson 5 that $\varphi(p) = p - 1$ when p is a prime number. Whatever number we choose, since it is less than p, it is always relatively prime to p. (We have proven that as well.) Then we can invoke The Euler-Fermat theorem here:

$$((N^{ac})^m)^{(p-1)} \times N^{ac} \equiv 1 \times N^{ac} = N^{ac}(\text{mod } m)$$

Applying the same method, we can see that:

$$N^{ac} = N^{(p-1)\times k+1} = (N^k)^{p-1} \times N \equiv N(\text{mod } m)$$

Hence, we have proved that $(((N^a)^b)^c)^d \equiv N(\text{mod } m)$, when $a \times c \equiv 1\big(\text{mod } (p-1)\big)$ and $b \times d \equiv 1(\text{mod } (p-1))$. This communication will (more) secretly convey the plain text because the encryption keys are selected privately by both the sender and the receiver of the message.

2.3

Suppose the eavesdropper is able to compute the discrete logarithm problems, and he/she has the knowledge of the residues of N^a, N^b, and N^{ab} (mod p). N^{ab} is equivalent to $(N^a)^b$.Then he/she can compute b by:

$$b = \log_{N^a} N^{ab}$$

Note that $N^b \equiv N^{abc}$ according to the proof above. Then the last missing piece for the eavesdropper to crack the code is d since

$$N \equiv N^{abcd} = (N^b)^d$$

d can be found by finding the inverse of b mod(p-1), which can be efficiently computed. Therefore, the plaintext is up for grab if someone can compute the discrete logarithm problems while knowing residues of N^a, N^b, and N^{ab} (mod p).

$$c = \log_{N^{ab}} N^b \text{, since } N^b = N^{abc}$$

$$N = (N^a)^c$$

3.1

Michael Jordan and I agree publicly on p=31, and r=3, which is a primitive root modulo 31. I want to send him the number (minus 1) of the day of the month of my birthday, which is N = 14. Michael privately chooses b=5 and sends me the encryption key:

$$e = r^b = 3^5 \equiv 26 \pmod{31}$$

I privately choose a = 20, and send Michael the ciphertext:

$$\tilde{r} = r^a = 3^{20} \equiv 5 \pmod{31}$$

$$\widetilde{N} = Ne^a = 14 \times 26^{20} \equiv 9 \pmod{31}$$

Now the six-time NBA champion computes:

$$\widetilde{N}\tilde{r}^{p-1-b} = 9 \times 5^{31-1-5} \equiv 14 \pmod{31}$$

Now Michael receives the two-digit number that I intended to send him.

3.2

Now let's look at how this cryptosystem worked.

Michael first send me the encryption key $e = r^b$;

I send back the ciphertext, $\widetilde{N} = Ne^a = N(r^b)^a = Nr^{ab}$ along with a decryption key $\tilde{r} = r^a$

By computing $\widetilde{N}\tilde{r}^{p-1-b}$, Michael obtains

$$\widetilde{N}\tilde{r}^{p-1-b} = Nr^{ab} \times (r^a)^{p-1-b} = Nr^{ab} \times (r^a)^{p-1} \times (r^a)^{-b} = Nr^{ab-ab} \times (r^a)^{p-1}$$

Now we can invoke Euler-Fermat theorem once again, since $\varphi(p) = p - 1$ when p is a prime number:

$$(r^a)^{p-1} \equiv 1 \pmod{p}$$

$$Nr^{ab-ab} \times (r^a)^{p-1} \equiv N \times 1 \times 1 \equiv N \pmod{p}$$

Hence, we have proven that $\widetilde{N}\tilde{r}^{p-1-b} \equiv N \pmod{p}$, and that's why this cryptosystem works.

3.3

If an eavesdropper on ElGamal cryptosystem can invalidate Diffie-Hellman assumption, then he can obtain r^{ab} with the knowledge of r, $r^a$and r^b modulo p. Now this eavesdropper also has knowledge of Nr^{ab}, which is $\widetilde{N}$, therefore he/she can formulate this congruence equation mod p:

$$Nr^{ab} = \widetilde{N} \equiv N \cdot r^{ab} (\text{mod } p)$$

Now that r was chosen as a primitive root to the large prime p, then p must be relatively prime to r. If r and p are relatively prime, r contains no p in its prime factorization's factors. Then r^{ab} must be relatively prime to p as well. Therefore, $\text{GCD}(r^{ab}, p) = 1$. Hence $\text{GCD}(r^{ab}, p)$ always divides $\widetilde{N}$. According to Bachet's Theorem, the congruence equation:

$$\widetilde{N} \equiv N \cdot r^{ab} (\text{mod } p)$$

is solvable. The eavesdropper can find the inverse to r^{ab} (modp), i. Then $i \times \widetilde{N}$ is a solution to N.

Now the eavesdropper has acquired the plaintext.

3.4

Diffie-Hellman assumption states that with knowledge only of p, r, $r^a$and r^b modulo p, it is computationally impossible to determine r^{ab}. If one knows how to crack ElGamal cryptosystem, he/she can retrieve plaintext N given p, r, r^a, r^b, and Nr^{ab} modulo p. Since p is a prime, and N is less than p, N and p must be relatively prime. There must exist N's inverse modulo p, N^{-1}, and it is very efficient to find modular inverses.

He/she now just has to multiply Nr^{ab} with N^{-1} to obtain r^{ab}. This process doesn't seem infeasible to compute. Hence if someone knows how to crack ElGamal cryptosystem, Diffie-Hellman assumption will not hold.

3.5

Suppose someone knows how to compute discrete logarithms efficiently. Then with the knowledge only of p, r, $r^a$and r^b modulo p, he/she can compute the value of a, by computing $a = \log_r r^a \ (\text{mod } p)$, as well as the value of b, by computing $b = \log_r r^b \ (\text{mod } p)$.

Now the values of a and b are obtained, computing r^{ab} is easy. Therefore, if one can compute the discrete logarithms efficiently, Diffie-Hellman assumption will not hold.

4.1

I want to send N= 617, the area code of my phone number to Michael Jordan using RSA encryption, so that maybe we can finally start communicating using text. Michael privately chooses two primes $p = 23$ and $q = 31$, as well as two integers $b = 101$ and $d = 281$ such that: $bd \equiv 1 \ (\text{mod}(p-1) \cdot (q-1))$.

He sends over publicly the modulus, $m = pq = 23 \times 31 = 713$, as well as the encryption key, b=101.

I then compute the ciphertext $\widetilde{N} = N^b = 617^{101} \equiv 421 (\text{mod } 713)$, and send over $\widetilde{N} = 421$ to Michael Jordan.

The Hall-of-Famer computes $\widetilde{N}^d = 421^{281} \equiv 617 \ (\text{mod} 713)$, and receives my phone number's area code.

4.2

$m = p \times q$, where p and q are distinct prime numbers. This is also m's prime factorization form. Then m must be a squarefree since the two of its prime factors both only have first power. Since p and q are distinct prime numbers, p and q must be relatively prime. Then $\varphi(m) = \varphi(pq) = \varphi(p) \times \varphi(q)$. We know that $\varphi(x)$ is always equal to x-1 when x is a prime number. Then $\varphi(m) = (p-1) \cdot (q-1)$.

Now that Michael privately chose b and d, such that $bd \equiv 1 \ (\text{mod}(p-1) \cdot (q-1))$, meaning $bd \equiv 1 \ (\text{mod } \varphi(m))$.

According to Fermat's little theorem, given m is squarefree, and $bd \equiv 1 \ (\text{mod } \varphi(m))$:

$$N^{bd} \equiv N (\text{mod } m)$$

With the encryption key public, I can send any ciphertext in the form $\widetilde{N} = N^b$, within range of m. For Michael to decrypt the message, he only has to compute $\widetilde{N}^d$ to obtain N^{bd}, which is congruent to the original plaintext modulo m, according to our proof above.

4.3

Suppose an eavesdropper learnt of b and the modulo m residue of N^b. He also has a solution to the factorization problem. Then he/she can find out the values of p and q, the two prime factors with just the knowledge of m.

Now that p and q is acquired, it is easy to compute the inverse of b modulo $(p-1)(q-1)$, which is the value of d. The eavesdropper only has to compute the last step, $N^{bd} \equiv N (\text{mod } m)$, to get the plaintext.

Therefore, if someone, who knows a solution to the factorization problem, intercepts the public key and ciphertext, the encryption is not secure.

4.5

m is squarefree; $k \equiv 1 (\text{mod} \varphi(m))$; $m' = \frac{m}{GCD(a,m)}$. Let's show that $a^k \equiv a (\text{mod } m')$

$m' = \frac{m}{GCD(a,m)}$, so m' is a divisor of m. According to question 3.13 in lesson 5, if m' is a divisor of m, then $\varphi(m')$ is also a divisor of $\varphi(m)$. Then we have: $\varphi(m) = j \times \varphi(m'), j \in \mathbb{Z}$

Given $k \equiv 1 (\mathrm{mod} \varphi(m))$, we can re-write k in the form: $k = i \times \varphi(m) + 1, i \in \mathbb{Z}$

Combining the equations above, and we get:

$$k = i \times j \times \varphi(m') + 1$$

Hence, we can re-write a^k in the form:

$$a^k = a^{i \cdot j \cdot \varphi(m')+1} = (a^{\varphi(m')})^{i \cdot j} \times a$$

According to question 3.24 in lesson 3, if m is squarefree, which is true is this case, then $\mathrm{GCD}\left(a, \frac{m}{\mathrm{GCD}(a,m)}\right) = 1$. In other words, a and m′ are relatively prime.

According to Euler-Fermat Theorem, if a and m′ are relatively prime, then $a^{\varphi(m')} \equiv 1 \pmod{m'}$. Let's substitute this back to the last equation modulo m′, and we have:

$$a^k = (a^{\varphi(m')})^{ij} \times a \equiv 1^{ij} \times a \equiv a (\mathrm{mod}\ m')$$

■

4.6

Since a is a multiple of GCD(a, m), a and any of a′s multiples will have no residue on division by GCD(a, m). In other words, $a \equiv 0 \left(\mathrm{mod}\ \mathrm{GCD}(a, m)\right)$, and $a^k \equiv 0 \left(\mathrm{mod}\ \mathrm{GCD}(a, m)\right)$. Therefore:

$$a^k \equiv a \equiv 0 \left(\mathrm{mod}\ \mathrm{GCD}(a, m)\right)$$

■

4.7

Let's deduce the extended version of Fermat's little theorem.

We now know that:

$$a^k \equiv a (\mathrm{mod}\ m'), \text{where } m' = \frac{m}{\mathrm{GCD}(a, m)}$$

$$a^k \equiv a (\mathrm{mod}\ \mathrm{GCD}(a, m))$$

From these two congruence equations, we can write the following equations. For readability, let's write GCD(a, m) as g:

$$a^k = a + m' \times c, \qquad c \in \mathbb{Z}$$

$$a^k = a + GCD(a, m) \times d = a + g \times d, \qquad d \in \mathbb{Z}$$

Then we have:

$$a + m' \times c = a + g \times d$$

$$m' \cdot c = g \cdot d$$

Let's take a look at m in its prime factorization form:

$$m = p_1 \cdot p_2 \cdots \cdot p_k$$

None of its prime factors have exponentiation higher than 1, since m is squarefree. Therefore, all of its prime factor, p_i, is either a divisor of a or not. The product of all of its prime factors that divides a is $GCD(a, m)$. The product of all of its prime factors that does not divide a is $\frac{m}{GCD(a,m)}$.There exist no p_i such that it is a factor of both $GCD(a, m)$ and $\frac{m}{GCD(a,m)}$, since m is squarefree. Then we can conclude that $GCD(a, m)$, g, and $\frac{m}{GCD(a,m)}$, m', are relatively prime.

Let's look at the equation above:

$$m' \cdot c = g \cdot d$$

This means m' divides $g \cdot d$. Since m' and g are relatively prime, m' does not divide g. By Euclid's Lemma, m' must divides d. Then there must exist an integer x, such that

$$d = x \cdot m'$$

Let's substitute this value into the equation

$$a^k = a + g \cdot d = a + g \cdot x \cdot m' = a + x \cdot GCD(a, m) \cdot \frac{m}{GCD(a, m)} = a + x \cdot m$$

$x \cdot m \equiv 0 \pmod{m}$, since $x \cdot m$ is a multiple of m.

Therefore, we have:

$$a^k = a + x \cdot m \equiv a + 0 \equiv a \pmod{m}$$

■