Grade 6 Math Circles

Cryptography

Introduction

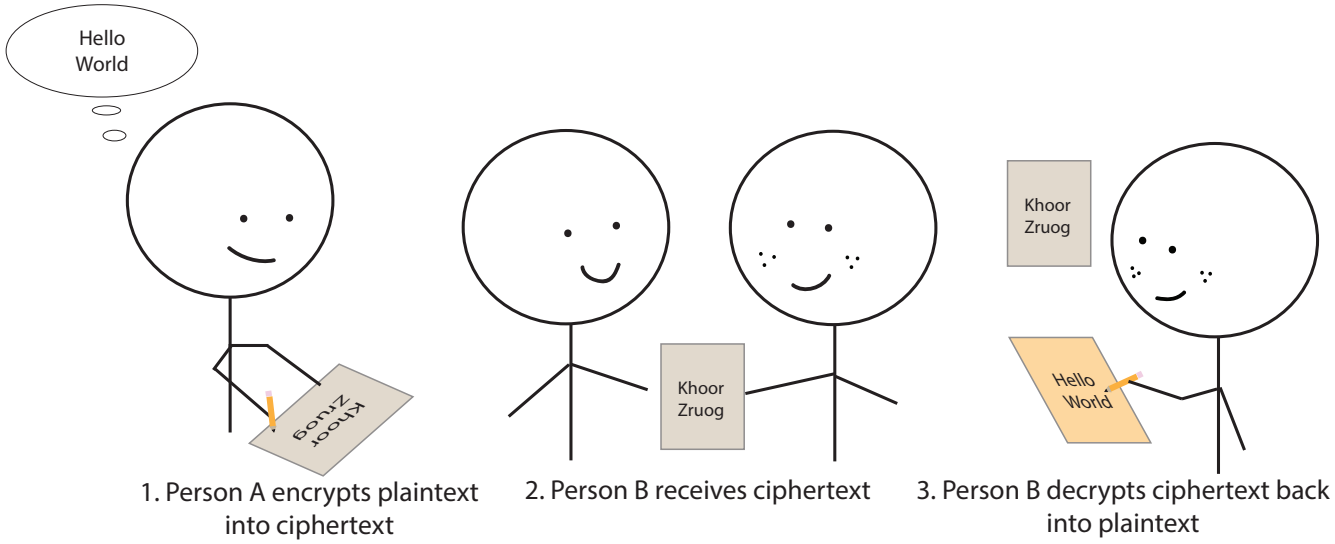
Cryptography is the study of hidden writing, or reading and writing secret messages or codes. The word cryptography comes from the Greek word *kryptos* ($\kappa\rho\upsilon\tau\varsigma$), meaning "hidden", and *graphein* ($\gamma\rho\alpha\phi\omega$), meaning "writing". There are some key words that will come up frequently in today's lesson.

Plaintext: The original message or information the sender wants to encode or hide

Encryption: The process of encrypting plaintext such that only authorized parties, such as the sender and intended receiver, can read it

Ciphertext: The plaintext that was encrypted using a cipher (the method of encryption)

Decryption: The process of decoding ciphertext back into its original plaintext



Stop and Think

In our everyday lives, what are some forms of written information that we would want to keep a secret from certain individuals? How often do you use some form of encryption?

The ciphers we will look at in this lesson are substitution ciphers. **Substitution ciphers** take letters in the alphabet and replace them with other letters in the alphabet, switching them with other letters such that the phrase "Hello!" becomes the gibberish "Svool!" to an eavesdropper.

Atbash Cipher

Originally created using the Hebrew alphabet, the **Atbash cipher** is created by reversing the alphabet. That is that A becomes Z, B becomes Y, C becomes X, and so on, until Z becomes A.

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

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

This is more easily represented below.

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

Example 1

The plaintext `ATBASH CIPHER` can be encrypted using the Atbash cipher, giving us the ciphertext `ZGYZHS XRKSVI`.

Exercise 1

Encrypt and decrypt the respective plaintext and ciphertext using the Atbash cipher.

(a) `SLICE OF PI`

(b) `HRTNZ HFNNZGRLM`

Caesar Cipher

This is one of the most well-known substitution ciphers. It is named after Julius Caesar who would encrypt communication intended for his army. Caesar encrypted his messages by shifting over every letter of the alphabet by 3 units. The cipher is as follows:

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

Example 2

Suppose Caesar wanted to send the following message:

`BEWARE THE IDES OF MARCH`

Using the cipher above, we take the plaintext letters and rewrite their corresponding ciphertext letters to get:

`EHZDUH WKH LGHV RI PDUFK`

When Caesar's best friend Brutus receives this message, he can refer to the same cipher, but decrypt by comparing the ciphertext letters with the corresponding plaintext letters.

The idea of the Caesar shift can be improved by using shift numbers other than 3.

Exercise 2

Encrypt or decrypt the following messages using the shift number given in parentheses.

(a) `MY SALAD NEEDS DRESSING` (8)

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

(b) `GUR YRGGHPR UNF R PBYV` (13)

plaintext	A	B	C	D	E	F	G	H	I	J	K	L	M
ciphertext	N												

plaintext	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ciphertext													

The last Exercise shouldn't be too difficult since the shift number was provided. This is why the shift number, or **key**, is usually not sent along with the ciphertext itself, but agreed on before hand by both sender and receiver.

When talking about the effectiveness of ciphers, we also ask about the **security** of the cipher. A cipher is secure when it is very difficult to decrypt a given ciphertext without the shift number, key, or context of the message.

Breaking Ciphers

So far we have looked at certain ciphers and how to encrypt or decrypt their messages as intended. But a major component of cryptography is how ciphers are to be **broken**. To break a cipher means to decrypt the ciphertext using brute force or clever guesses.

Cryptography has its historical roots in warfare as we saw in the case of the Caesar cipher. If our communications to our allies are not encrypted when sent, enemies who intercept our messages gain unwanted access to our secret plans. By encrypting communications however, enemies may never understand our plans, or at the very least, we slow them down as they need to take some time to decipher them.

Notably World War II took place during a time where ciphers were becoming more and more complex. Emerging technologies were invented to encode, decode, and break ciphers much quicker. The Enigma Machine is a notable and well-known example of a cipher machine used by the Germans. **Alan Turing**, the "Father of Computer Science" was on the team of cryptographers who broke Enigma. This breakthrough not only contributed to the war ending sooner, but it was a major feat in computing and mathematics at the time.

Example 3

Suppose we intercepted the following Caesar ciphertext but do not know the shift number.

WZXLIV FLK NYRK ZK JRPJ

What are some things we could do to decode this message and figure out what it says?

Solution: There are a couple of different approaches. One way that will always work is to test all 26 unique shift numbers and apply them to the cipher text.

1: XAYMJW GML OZSL AL KSQK
2: YBZNKX HNM PATM BM LTRL
3: ZCAOLY ION QBUN CN MUSM
4: ADBPMZ JPO RCVO DO NVTN
5: BECQNA KQP SDWP EP OWUO
6: CFDROB LRQ TEXQ FQ PXVP
7: DGESPC MSR UFYR GR QYWQ
8: EHFTQD NTS VGZS HS RZXR
9: FIGURE OUT WHAT IT SAYS
10: GJHVSF PVU XIBU JU TBZT
11: HKIWTG QWV YJCV KV UCAU
12: ILJXUH RXW ZKDW LW VDBV
13: JMKYVI SYX ALEX MX WECW
14: KNLZWJ TZY BMFY NY XFDX
15: LOMAXK UAZ CNGZ OZ YGEY
16: MPNBYL VBA DOHA PA ZHFZ
17: NQOCZM WCB EPIB QB AIGA
18: ORPDAN XDC FQJC RC BJHB
19: PSQEBO YED GRKD SD CKIC
20: QTRFCP ZFE HSLE TE DLJD
21: RUSGDQ AGF ITMF UF EMKE
22: SVTHER BHG JUNG VG FNLF
23: TWUIFS CIH KVOH WH GOMG
24: UXVJGT DJI LWPI XI HPNH
25: VYWKHU EKJ MXQJ YJ IQOI
26: WZXLIV FLK NYRK ZK JRPJ

What was the shift used?

Applying the remaining 25 shifts to a Caesar ciphertext might have been a long and tedious process in the times of Caesar, it is very accessible to break Caesar ciphers with the use of modern technology. For example, I ran the ciphertext through a website that brute-forced all possible shifts in a couple of seconds.

The security of the Caesar cipher, or the Atbash cipher for that matter, can be easily challenged by a computer program or very determined codebreaker. This doesn't mean that all substitution ciphers can be broken as easily.

Consider this random substitution cipher I created just now with no pattern and no shift number in mind.

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

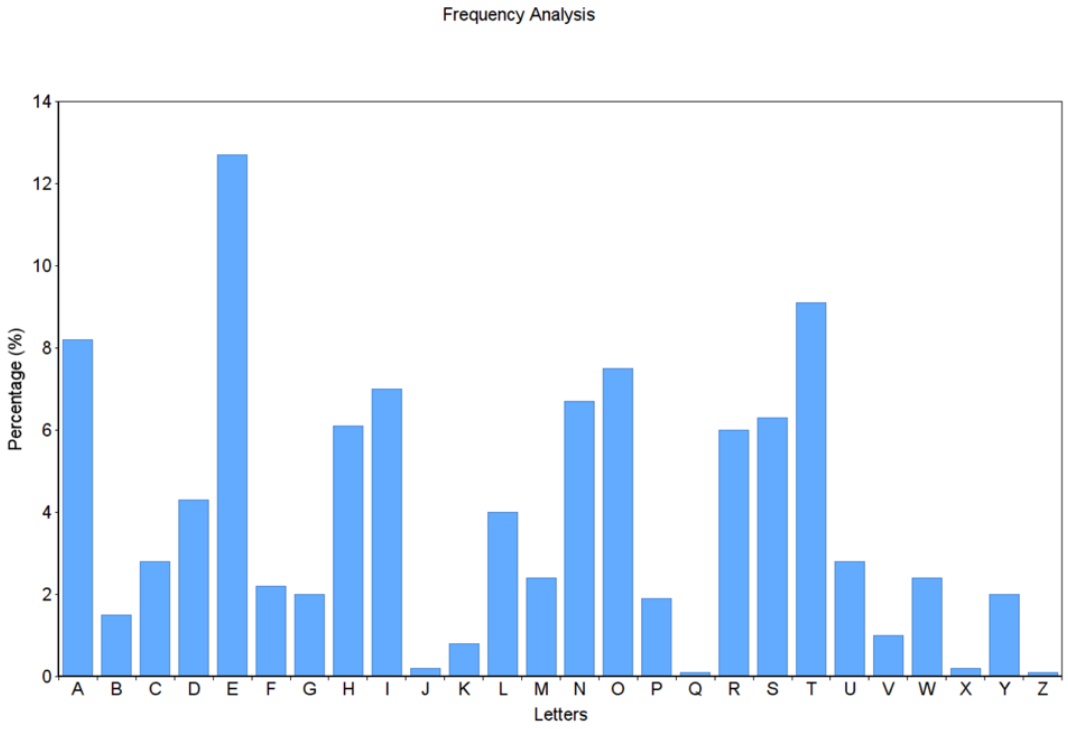
If I send out the message `HWUM MRTP RK NALZW?` without the cipher attached, anyone who intercepts the cipher may try to break it using the Caesar or Atbash ciphers without success. They will have to resort to other means of decoding it. This means that the security of substitution ciphers is not weak at all. You can easily upgrade my shift by not having the same letters map to each other (e.g. `A` encodes to `U`, but `U` encodes to `G`, `G` encodes to `W`, etc.).

Even if you are a computer programmer, there are $26! = 403291461126605635584000000$ possible arrangements of the English alphabet, meaning there are far too many possible substitution ciphers to brute force. This implies that we will need to be more clever with the ways that we narrow down the best approaches to breaking a cipher.

Frequency Analysis

One method of breaking an encryption is known as **frequency analysis**. Since we are dealing with letters, **frequency** is the number of times a letter occurs. This is particularly useful for substitution ciphers. By noting the most frequent letters in a ciphertext, we can begin to deduce what letters they might correspond to in the plaintext.

Below is the frequency graph that shows the average frequency of each letter in the English alphabet.



The most commonly used letter of the English alphabet is the letter E. Following E, we have the letters T, A, O, I, N, etc.

In knowing the most frequent letters of a message, we can begin to take guesses at possible substitutions of its letters. If the message is long enough, we might even observe common groups of letters being used multiple times, indicating frequently used words of the language.

Exercise 3

Consider the ciphertext

`RXHX GX WF UWUYK`

What is the most common letter in the ciphertext? Assuming that the plaintext is in English, what letter in the ciphertext likely corresponds to the plaintext letter `E`?

Note that you do not have to decode this cipher.

Vigenère Cipher

Let us look at one more substitution cipher that can be considered more secure than ciphers we have seen before. The **Vigenère cipher** was invented in by Giovan Battista Bellaso in 1553 despite being named after Blaise de Vigenère.

Vigenère encryption uses a **keyword**. Using this keyword, we will apply several Caesar ciphers at a time. We will also need to translate the alphabet into numbers:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Example 4

Suppose we use the keyword `CODE` and we want to encrypt the following plaintext:

`TIME FOR COOKIE`

To begin, write out the plaintext and keyword on a table, repeating the keyword until the end of the plaintext.

keyword	C	O	D	E	C	O	D	E	C	O	D	E	C
shift number													
plaintext	T	I	M	E	F	O	R	C	O	O	K	I	E
ciphertext													

We then translate each letter in our keywords into numbers and write them underneath in the shift number row. These correspond to the shift numbers you will use to **apply a Caesar shift** to each plaintext letter.

keyword	C	O	D	E	C	O	D	E	C	O	D	E	C
shift number	2	14	3	4	2	14	3	4	2	14	3	4	2
plaintext	T	I	M	E	F	O	R	C	O	O	K	I	E
ciphertext													

Starting on the left column, our plaintext letter is `T` and our shift number is 2 (from our keyword letter C). Thus to get our ciphertext we apply a Caesar shift of 2 on `T`.
With a Caesar Shift of 2:

plaintext	...	O	P	Q	R	S	**T**	U	V	W	X	Y	Z
ciphertext	...	Q	R	S	T	U	**V**	W	X	Y	Z	A	B

We get `V` for our ciphertext letter as shown above. (Note we could also just count 2 letters to the right from `T` in the alphabet to find our ciphertext letter `V`.) Our original chart now looks like this

keyword	C	O	D	E	C	O	D	E	C	O	D	E	C
shift number	2	14	3	4	2	14	3	4	2	14	3	4	2
plaintext	T	I	M	E	F	O	R	C	O	O	K	I	E
ciphertext	**V**												

The next column of the chart we have the letter `I` and a shift number of 14. So we apply a Caesar shift of 14 to the letter `I`.
With a Caesar Shift of 14:

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

We get `W` for our ciphertext letter. We update the chart as follows.

keyword	C	O	D	E	C	O	D	E	C	O	D	E	C
shift number	2	14	3	4	2	14	3	4	2	14	3	4	2
plaintext	T	I	M	E	F	O	R	C	O	O	K	I	E
ciphertext	V	**W**											

As you can see, each letter in the ciphertext is determined by the shift number and the plaintext letter. For each plaintext letter, we apply a Caesar cipher using the corresponding shift number.

The Vigenère Cipher is much harder to break compared to the ciphers from before. In the last example, the two `O`'s in `COOKIE` will be shifted by 2 and 14 respectively, leading to them not sharing the same ciphertext letter despite being the same plaintext letter. In the Problem Set, we will work on decrypting Vigenère ciphertexts and explore how frequency analysis can be used to break it.

Exercise 4

Complete the Vigenère encryption from Example 5. Note, the final ciphertext will have spaces at the same locations as the plaintext:

TIME FOR COOKIE

keyword	C	O	D	E	C	O	D	E	C	O	D	E	C
shift number	2	14	3	4	2	14	3	4	2	14	3	4	2
plaintext	T	I	M	E	F	O	R	C	O	O	K	I	E
ciphertext	V	W											

The following Caesar shift tables will be helpful.

Caesar shift of 2:

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

Caesar shift of 3:

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

Caesar shift of 4:

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

Caesar shift of 14:

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

Conclusion

While we might not have seen as much algebra or numbers in this week's lesson, cryptography is nevertheless a topic of great interest and ongoing research in the intersection of pure and applied mathematics. It has never been harder to find a secure cipher when it has never been easier to use technology to assist in breaking them.