PART III - CRYPTOGRAPHY

You may know what this means · · ·

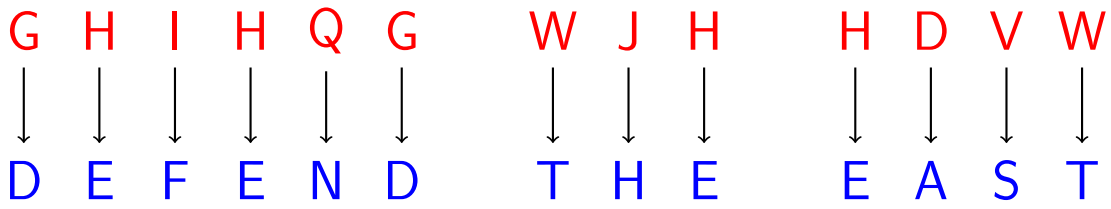
"Fascinating – I used to work on codes myself, during the war."

But do you know what this means?

GHIHQG WJH HDVW ZDOO RI WKH FDVWOH

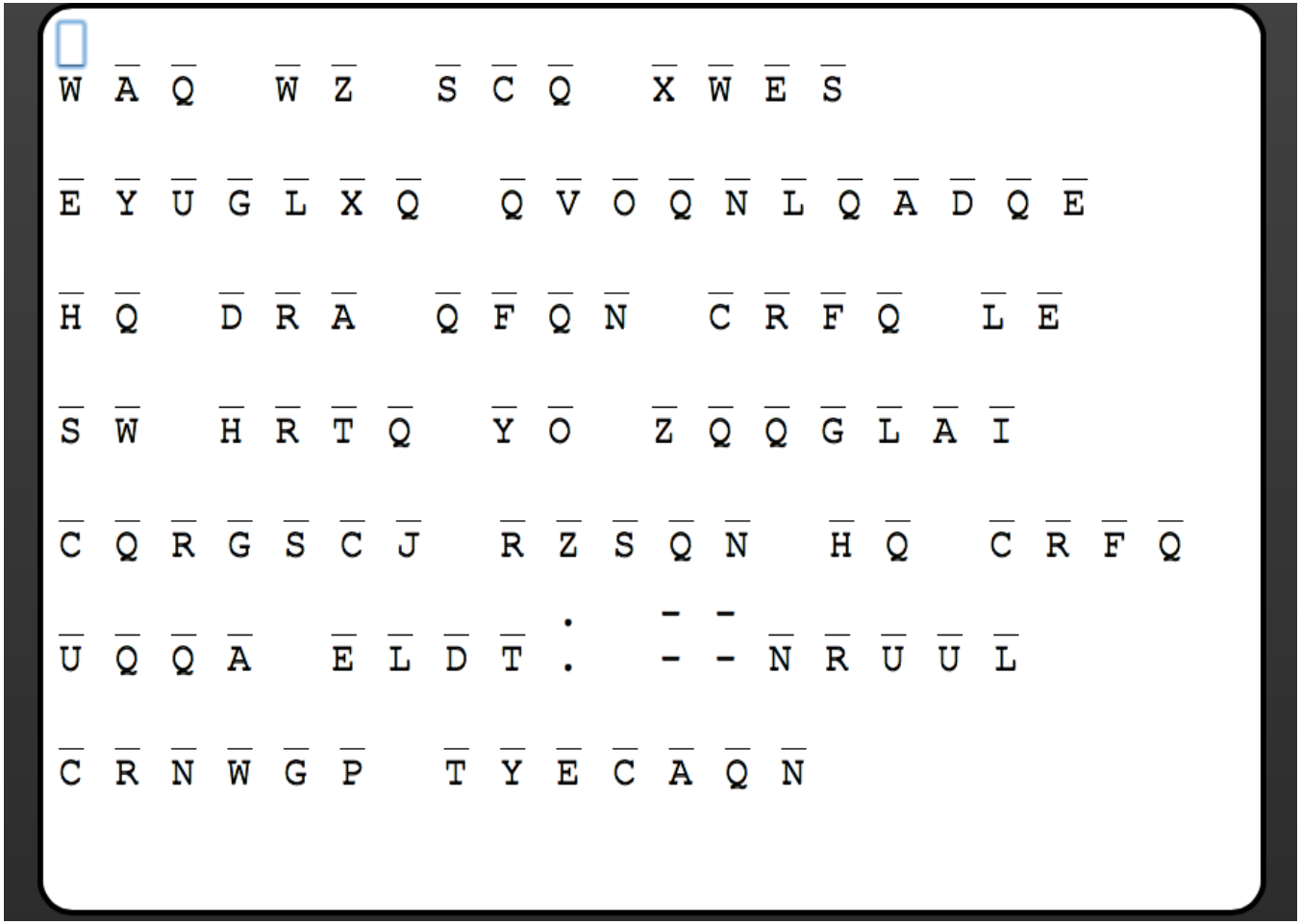
Julius Caesar (100 BC - 44BC) could have sent this message! Because he didn't trust his generals, when he sent messages he did something like replace every "A" with a "D", every "B" with an "E", every "C" with an "F", etc. See the pattern? We call this the key which allows us to decode the message. It is called the Caesar Cipher.

GHIHQG WJH HDVW ZDOO RI WKH FDVWOH



DEFEND THE EAST WALL OF THE CASTLE

Have you downloaded an app for Cryptograms or seen a puzzle like this in a magazine or newspaper?



Goals for Part III

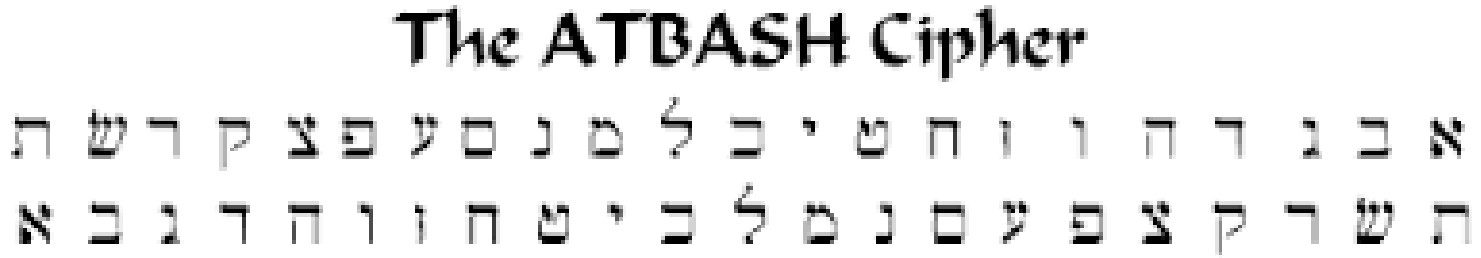
1. To learn basic terminology of crytography
2. To see how to code and decode simple ciphers if the key is given
3. To see how to decode simple substitution ciphers without a key using frequency of letters and words
4. To understand the difference between classical cryptography and modern cryptography.
5. To see how Public Key Encryption works.
6. To see how a Digital Signature works.

Goals for this lecture:

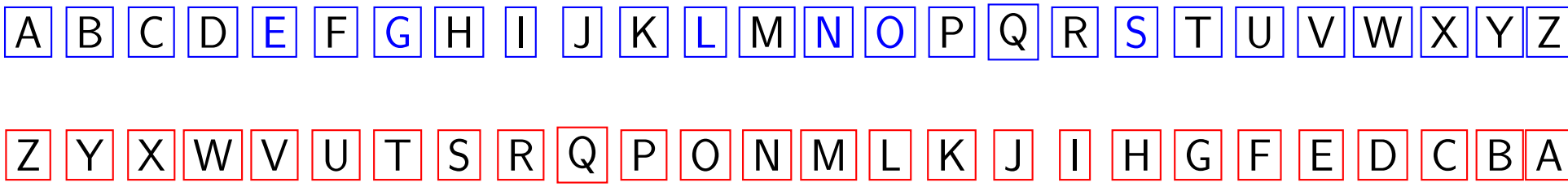
1. See how ciphers have been used in the past.
2. Learn basic terminology
3. Learn how to write a message using the Caesar cipher
4. Learn how to decode a message using the Caesar cipher
5. To understand why this type of cipher is not secure

Atbash Cipher

An even earlier code (circa 500 BC) than the Caesar cipher was used for the Hebrew alphabet. Instead of shifting a letter to the right it replaces the first letter of the Hebrew alphabet (aleph) with the last letter (tav); then the second letter (beth) for the second from last letter (shin), etc. So it uses a substitution where the letters are reflected.



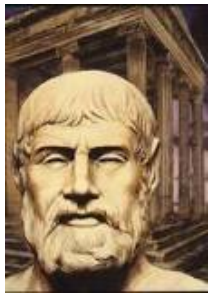
If we used an Atbash-like cipher with our alphabet we would make a substitution like the following.



Other Early Attempts at Secrecy

Early attempts were often messy. One strategy was to hide it on the messenger.

- The Chinese wrote messages on silk and encased them in a ball of wax which would then be hidden *in* the messenger.
- The Greeks (around 500 BC) used a different approach. Histaeus wanted to send a message to Aristogorus to urge revolt against the Persians but he needed secrecy. He shaved the head of his most trusted slave and tattooed the message on his head. When the hair grew back the slave was sent to Aristogorus who shaved his head and read the message. I guess they weren't in a hurry!



The Spartans used a tool called a scytale.

A piece of parchment is wound around a baton and a message is written on it. Then the parchment is removed and sent to the recipient. Of course the recipient must have the same size baton for it to work.

How could you break this code?

There were no significant improvements in creating more complex ciphers until the middle ages.

Leon Battista Alberti, a Venetian, created something called a Cipher disk and described how to use it in a 1467 treatise. One disk is fixed and you turn the other.

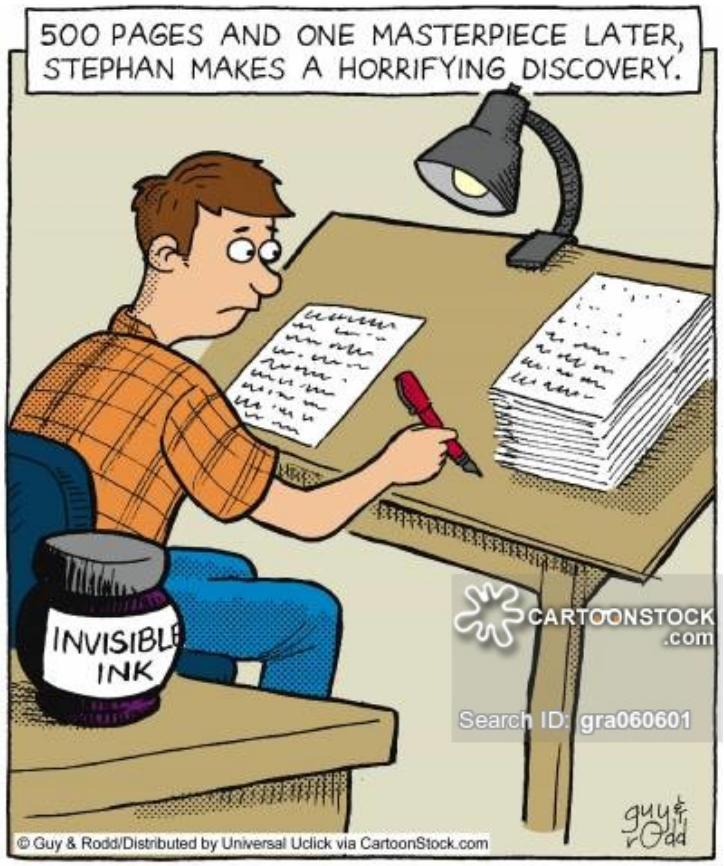
You turn the disk and the letters that match up from one disk to another give the substitution. Note that the letters are not in order in the center disk. In this picture the substitution V for A, Z for B, etc. is used.

Invisible Ink

Using invisible ink is a less intrusive method of guaranteeing secrecy. The simplest choices are organic compounds such as lemon juice, milk, or urine. During the American revolution both sides made extensive use of chemical inks that required special developers. It was used up through World War I.

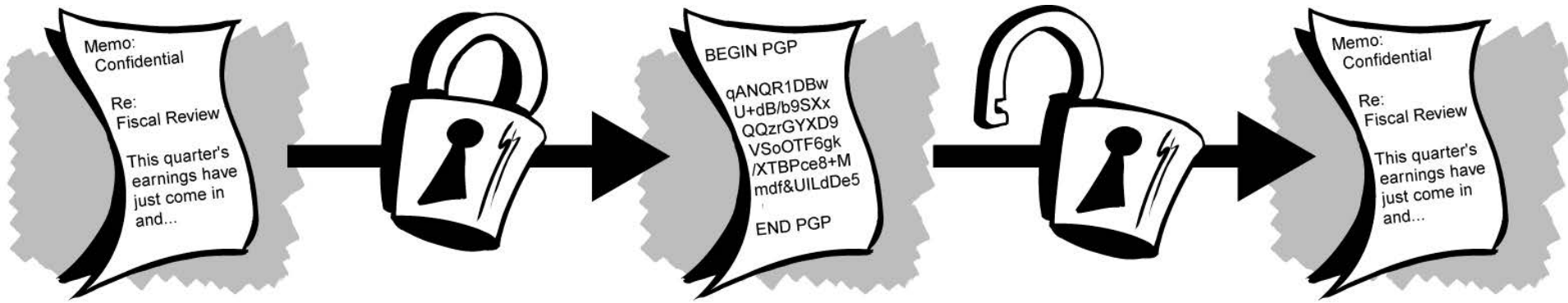
An invisible ink letter treated with a chemical reagent from British spy Benjamin Thompson, 1775 (Clements Library, University of Michigan).

There is a book entitled **Invisible Ink:Spycraft of the American Revolution** by John Nagy which you might find interesting.



Encryption & Decryption

- We call PLAINTEXT data that can be read and understood without any special measures. Examples: these notes, books, magazines, etc.
- ENCRYPTION is the method of disguising plaintext so that the information is hidden.
- CIPHERTEXT is encrypted plaintext. Example: the encrypted message on the second slide using the Caesar Cipher.
- DECRYPTION is the process of changing ciphertext to plaintext.



PLAINTEXT CIPHERTEXT PLAINTEXT

ENCRYPTION DECRYPTION

What is Cryptography?

- Cryptography is the study of using methods to encrypt information or decrypt ciphertext.
- In the past, cryptography has been used to protect secrets such as military information.
- Now we use methods to encrypt sensitive data such as personal information and to transmit information across insecure networks so it can only be read by the intended recipient.
- In the far past, people used pen & pencil to encrypt messages but now mathematical algorithms are used.
- Cryptography is everywhere now.
  - Web traffic
  - Wireless traffic like blue Tooth

- Encrypting files on your computer disk
- Content protection (DVDs, Blu-ray, etc.)
- Digital signatures
- etc.

- We will first look at simple encryptions methods which can be used to encrypt or decrypt messages by hand.
- Then we will investigate the ideas behind sophisticated encryption methods that are used in your daily life.

Here is a quote that is worthwhile to keep in mind.

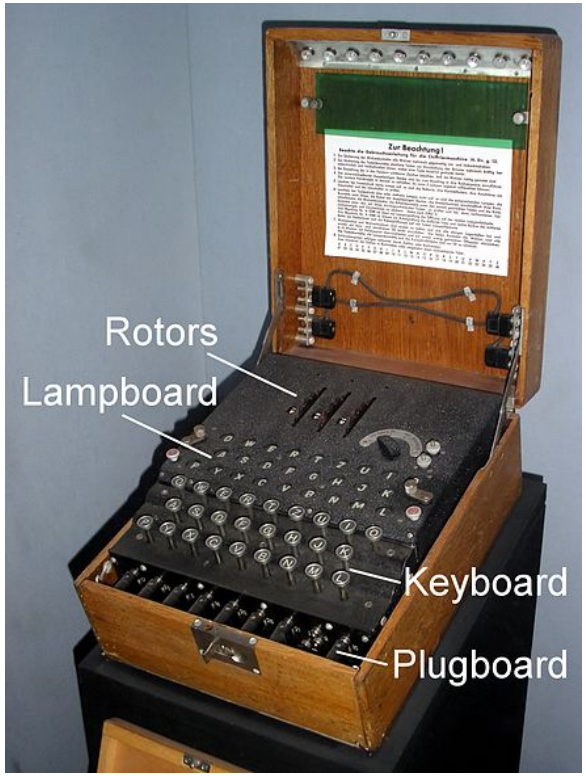
"There are two kinds of cryptography in this world: cryptography that will stop your kid sister from reading your files and cryptography that will stop major governments from reading your files."

– Bruce Schneier

Codebreaking in World War II

Did you see the movie The Imitation Game or read a book about Alan Turing such as the *Enigma Machine*?

- After WWI the Germans built a machine called the enigma machine to protect diplomatic and military communications.



- An operator typed the text of the communication on the keyboard and the machine encrypted the message.
- Three Polish cryptologists (working for Polish military) developed a mechanical

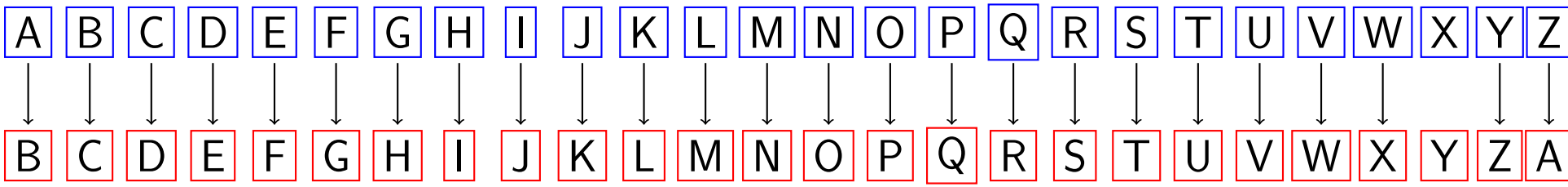
device for breaking the Enigma ciphers but the Germans then added additional complexity to the machines to make decryption much more difficult.

- Crytographers during WWII (led by Alan Turing) devised an electromechanical machine (a forerunner of a computer) which found the key to encrypt the messages and thus they could decrypt the messages.
- This played a pivotal role in reading encrypted German messages that enabled the Allies to defeat the Nazis in many crucial engagements.
- For more information on the Enigma machine go to `cryptomuseum.com` There is an app there you can download to simulate the Enigma machine.

Shifted Ciphers (Caesar Cipher)

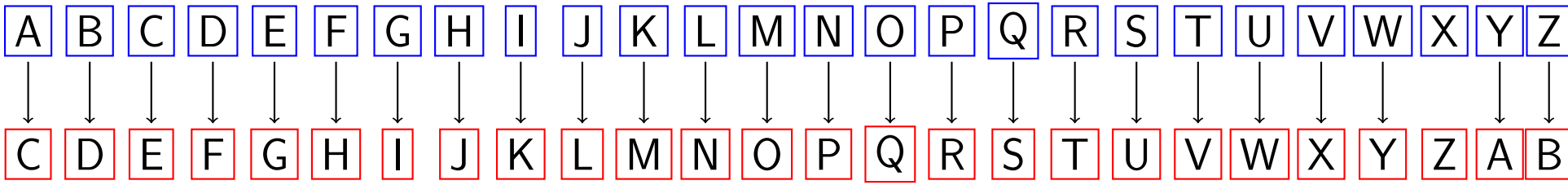
- The example we showed on the first slide used the Caesar cipher where each letter of the alphabet was replaced by the letter which was three letters after it (in a circular fashion).
- However, we can use any number of letters between 1 and 25 to shift the letter.
- This means that there are only 25 possible ways to decrypt any message using this simple shift method.

Shift 1 letter to right

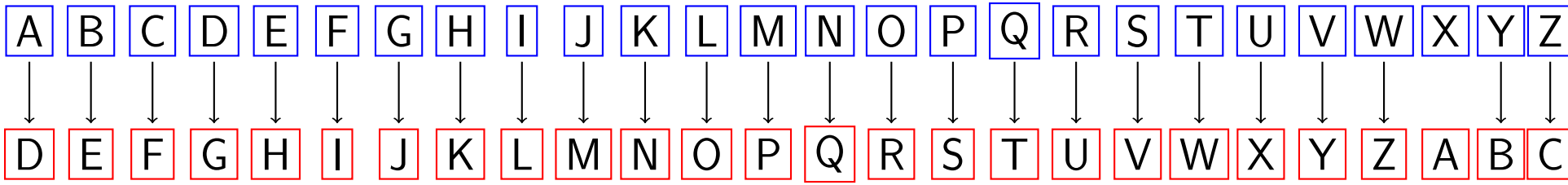


"Z" is replaced by "A" in this cipher because we shift the alphabet in a cyclic manner.

Shift 2 letters to right



Shift 3 letters to right

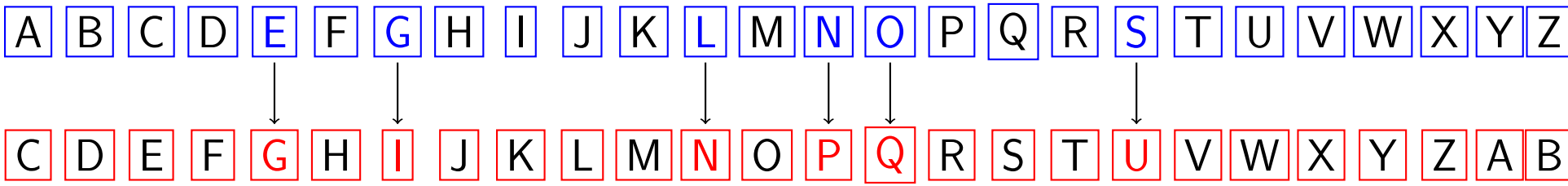


Example. Encrypt the message

GO NOLES

using a shifted cipher where you shift by 2 letters.

We first write the letters of the alphabet and then under it, the letters of the alphabet shifted by two letters



To do this we take a "G" from the top line and see that below it is an "I", below the "O" is a "Q", etc.

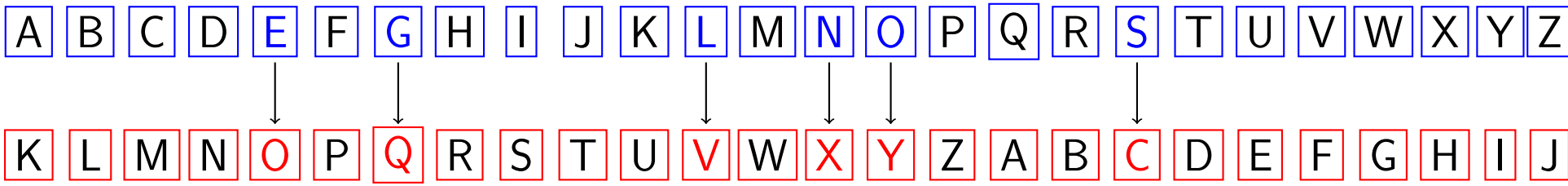
IQ PQNGU

Example. Encrypt the message

GO NOLES

using a shifted cipher where you shift by 10 letters.

We first write the letters of the alphabet and then under it, the letters of the alphabet shifted by ten letters.



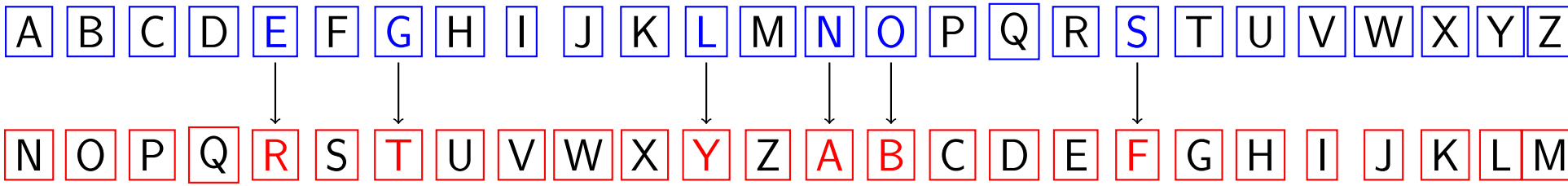
QY XYVOC

Example. Encrypt the message

GO NOLES

using a shifted cipher where you shift by 13 letters.

We first write the letters of the alphabet and then under it, the letters of the alphabet shifted by ten letters.



TB ABYRF

Note that because we have shifted by 13 characters and our alphabet has 26 characters (2×13) it is called a symmetric cipher and given the name ROT13. This is because to encrypt and decrypt you use the same key. For example, A $\rightarrow$ N and N $\rightarrow$ A. ROT13 is used in online forums as a means of hiding spoilers, punchlines, puzzle solutions,

and offensive materials from the casual glance. It is the basis for many puzzles and games and is thought to be the equivalent of a magazine printing the answers to a puzzle upside down.

Why did the chicken cross the road?

Gb trg gb gur bgure fvqr!

SOCRATIVE QUIZ - PartIII_Practice_Quiz1

IUZGAZ34E

1. Encrypt FSU using a 5 letter shift.
2. Encrypt FSU using a 6 letter shift.

HINT for # 1

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

F	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

Decrypting a Shifted Cipher

Decrypting a shift cipher is harder than encrypting because we have to go backwards! To encrypt all we need is to follow the key and do it once but with decrypting a shift cipher we have to consider possibly all choices. Lucky for us for a shifted cipher there are only 25 possibilities.

Consider the encrypted message QY XYVOC

If this had been encrypted with a 1-letter shift then we write the letter before "Q" (which is "P"), then the letter before "Y" which is "X", etc. to get PX WXUNB which clearly isn't decrypted. So the message wasn't encrypted using a 1-letter shift.

If the original message was formed using a 2-letter shift then we write the letter two before "Q", etc. or more simply, one letter before the 1-letter shift PX WXUNB to get OW VWUMA.

Continuing in this manner, we get the following table and see that the message was encrypted with a 10 letter shift.

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

Shift 0	QY XYVOC
Shift 1	PX WXUNB
Shift 2	OW VWUMA
Shift 3	NV UVTLZ
Shift 4	MU TUSKY
Shift 5	LT STRJX
Shift 6	KS RSQIW
Shift 7	JR QRPHV
Shift 8	IQ PQNGU
Shift 9	HP OPMFT
Shift 10	GO NOLES
Shift 11	FN MNKDR

We have to do at most 25 of these shifts to find the answer. Easy for us to do and even easier for a computer!

Because we have to do at most 25 shifts it doesn't really matter whether we shift backwards or forwards. So for the previous problem we could decrypt it by shifting to the right.

QY XYVOC $\Longrightarrow$ RZ YZWPD $\Longrightarrow$ SA ZAXQE

$\Longrightarrow$ TB ABYRF $\Longrightarrow$ UC BCZSG $\Longrightarrow$ VD CDATH

$\Longrightarrow$ WE DEBUI $\Longrightarrow$ XF EFCVJ $\Longrightarrow$ YG FGDWK

$\Longrightarrow$ ZH GHEXL $\Longrightarrow$ AI HIFYM $\Longrightarrow$ BJ IJGZN

$\Longrightarrow$ CK JKHAO $\Longrightarrow$ DL KLIBP $\Longrightarrow$ EM LMJCQ

$\Longrightarrow$ FN MNKDR $\Longrightarrow$ GO NOLES

In this example it was shorter to shift backwards but this is not the case in general as the following decryption illustrates.

Backward Shift	Decryption
0	LMTK PTKL
1	KLSJ OSJK
2	JKRI NRIJ
3	IJQH MQHI
4	HIPG LPGH
5	GHOF KOFG
6	FGNE JNEF
7	EFMD IMDE
8	DELC HLCD
9	CDKB GKBC
10	BCJA FJAB
11	ABIZ EIZA
12	ZAHY DHYZ
13	YZGX CGXY
14	XYFW BFWX
15	WXEV AEVW
16	VWDU ZDUV
17	UVCT YCTU
18	TUBS XBST
19	STAR WARS

Forward Shift	Decryption
0	LMTK PTKL
1	MNUL QULM
2	NOVM RVMN
3	OPWN SWNO
4	PQXO TXOP
5	QRYP UYPQ
6	RSZQ VZQR
7	STAR WARS

Typically in a shift cipher the key is given as a shift to the right. So to describe the key used in this cipher, we say that we shift the alphabet 19 letters to the right. Equivalently, we could describe it as 7 letters to the left.

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

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

It's a bit tedious to go through all of these shifts every time. In fact, this is the Brute Force approach for decrypting a shifted cipher. Is there anything we can do to speed up the process?

1. First we write down the alphabet across the top of a square. Then in a second row we write the alphabet shifted by one letter, in the third row the alphabet shifted by 2 letters, etc. It is called a Vigenere square.
2. Next we use some things we know about the English language. First, E is the most common letter followed by T. Also we can take a short word and see which shift makes sense. We will explore more about the frequencies of letters and words in the English language in the next lecture.
3. This Vigenere square makes encrypting messages with a shifted cipher easy too.

	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
S1	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	A
S2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
S3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
S4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
S5	F	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
S6	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
S7	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	G
S8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
S9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
S10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
S11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
S12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
S13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
S14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
S15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
S16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
S17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
S19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
S20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
S21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
S22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
S24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
S25	Z	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

Example. Use the fact that "E" is the most frequently appearing letter in the alphabet and "T" the second most to decode the message

XJINODOPODJI

- The letter "O" appears 3 times in this expression so we first see if this represents "E". So we look in the column with "E" at the top and go down the column until we see an "O". This occurs in Shift 10 the letter "O" represents "E" and so the decrypted message is NZ $\cdots$ which doesn't make sense.
- Next we try to see if "O" represents "T". We look at the column with "T" at the top and go down the column until we find an "O". If we look at the row beginning with "V" we see that this is the case so we apply Shift 21 to get CONSTITUTION

	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
S1	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	A
S2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
S3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
S4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
S5	F	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
S6	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
S7	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	G
S8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
S9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
S10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
S11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
S12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
S13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
S14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
S15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
S16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
S17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
S19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
S20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
S21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
S22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
S24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
S25	Z	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

Example. Use the fact that "E" is the most frequently appearing letter in the alphabet, "T" the second most and "A" the third to decode the message

IZUILQTTW

- The letter "I" appears 2 times in the encrypted message and the letter "T" appears two times (consecutive).
- We first see if "I" represents "E". We look at the top row and find the column with "E" and then go down the column until we find an "I" which occurs in Shift 4 but the decryption would begin EVQEH which doesn't make sense.
- Next we try to see if "I" represents "T". If we look at the column with "T" at the top and go down the column until we find an "I" we see that this occurs in Shift 15. In this case we get TKFTW which doesn't make sense.
- Next we try to see if "I" represents "A". If we look at the column beginning with "A" we see that using Shift 8 gives this representation and we get ARMADILLO

	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
S1	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	A
S2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
S3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
S4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
S5	F	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
S6	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
S7	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	G
S8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
S9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
S10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
S11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
S12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
S13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
S14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
S15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	CD	E	F	G	H	I	J	K	L	M	N	O	
S16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
S17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
S19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
S20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
S21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
S22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
S24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
S25	Z	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

Example. Decrypt the message

OZZ WG KSZZ

- The letter "Z" appears 4 times but it probably doesn't represent an "E" (we would have "EE" twice) or a "T" (we would have "?tt") or an "A" so we have to use a different strategy.
- The first word has a double letter at the end which is probably a consonant so the first letter is probably a vowel. Let's try this first.
- Looking at our square (see next slides with Vignere square highlighted), we now ask ourselves in which shifts does O represent a vowel. To do this we start with "O" at the top row and follow the diagonal consisting of "O" and see which shift gives a vowel. The first encountered is "I", then "E", etc. We omit "O" because this is the unshifted alphabet and we know shift 0 doesn't make sense.
  - Shift 6 O $\Longrightarrow$ I
  - Shift 10 O $\Longrightarrow$ E
  - Shift 14 O $\Longrightarrow$ A

- Shift 20 O $\Longrightarrow$ U

- Now we look at what "Z" represents in each of these four shifts.
  - Shift 6 OZZ $\Longrightarrow$ ITT
  - Shift 10 OZZ $\Longrightarrow$ EPP
  - Shift 14 OZZ $\Longrightarrow$ ALL
  - Shift 20 OZZ $\Longrightarrow$ UFF
- The only one that makes sense is Shift 14 so the message is decrypted as

ALL IS WELL

	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
S1	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	A
S2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
S3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
S4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
S5	F	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
S6	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
S7	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	G
S8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
S9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
S10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
S11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
S12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
S13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
S14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
S15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
S16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
S17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
S19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
S20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
S21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
S22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
S24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
S25	Z	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

We have seen two approaches for decrypting the Shift Cipher

1. Brute Force. In this approach we try to decrypt the message by shifting the alphabet by 1 letter and if this doesn't work we try a shift of 2 letters, etc. We know that we have to do at most 25 shifts to decrypt the message.

2. Use our knowledge of the English language. In this approach we use common knowledge about our language to make educated guesses for letters and see if a particular shift gives the decoded message. A Vignere square is helpful.

For a Shift Cipher it doesn't really matter which approach we choose because the Brute Force approach only takes a maximum of 25 tries. But for other ciphers we will see that the number of possibilities is so large that a brute force approach is not practical.

SOCRATIVE QUIZ - PartIII_Quiz1

IUZGAZ34E

1. _____ is the science and art of transforming messages to make them secure.
   a. Cryptography
   b. Encryption
   c. Decryption
   d. Caesar Cipher
2. _____ is the original message before encryption.
   a. Key
   b. Ciphertext
   c. Plaintext
   d. None of the above
3. _____ is the message after encryption.

a. Key
b. Ciphertext
c. Plaintext
d. None of the above

4. _____ is the procedure that transforms plaintext to ciphertext.

a. Key
b. Encryption
c. Decryption
d. Cryptography

5. _____ is the procedure that transforms ciphertext to plaintext.

a. Key
b. Encryption
c. Decryption
d. Cryptography

6. In a Shift Cipher when the Key is shift to right 4 letters, how is the letter "P" encrypted?

a. R
b. S
c. T
d. U

7. In a Shift Cipher when the Key is shift to LEFT 2 letters, how is the letter "P" decrypted?

a. N
b. R
c. M
d. S

8. In a Brute Force approach to decrypting a Shift Cipher where a 5 letter alphabet is used, the maximum number of tries is

a. 6
b. 5
c. 4
d. 3

9. If a Caesar shift where we shift 2 characters to the right, then "FSU" is encrypted as
    a. GTV
    b. HTV
    c. IVX
    d. HUW
10. If a Caesar shift where we shift 3 characters to the right, then "ZOO" is encrypted as
    a. CRR
    b. YNN
    c. CPP
    d. APP
    e. BQQ

Goals for this Lecture

1. To understand what a monoalphabetic substitution cipher is.
2. To learn how to encrypt and decrypt a message using a simple substitution cipher if the key is given.
3. To understand that a "Brute Force" approach to decrypting a simple substitution cipher is not feasible.
4. To learn how to interpret and make frequency plots for the occurrence of letters and words in a given text.
5. To see how to use frequency analysis to decrypt a simple substitution cipher.
6. To understand the weaknesses in a monoalphabetic substitution cipher.

Simple Substitution Ciphers

- In the examples in the previous lecture, the alphabet was shifted but the letters remained in the same order. This type of cipher is the simplest type of monoalphabetic substitution cipher which uses only one fixed alphabet.
- For example, when we used a Caesar Cipher with Shift 3 to right we used the single alphabet starting with "D" as a substitution alphabet

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

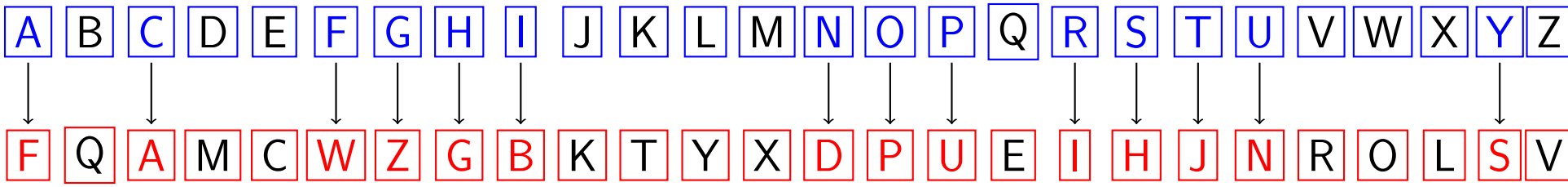
- Also the Atbash cipher is a monoalphabetic cipher but NOT a shifted cipher. Here we use the following alphabet as a substitution.

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

- We saw that it was easy to decrypt a message encrypted with a shifted (monoalphabetic) key because in English we have to do at most 25 shifted versions of the message to find the answer. We can also shorten our work by investigating the mostly commonly appearing letters.
- However, suppose we still just use the English alphabet but we mix up the order of the letters and then use this as a substitution, decrypting the message becomes a lot harder to do by hand. For example, suppose we use the key

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

Encrypting a message is always straightforward. For example, if we have the key



then the message

CRYTOGRAPHY IS FUN

is encrypted as

AISJPZIFUGS BH WND

Decrypting a message is easy if we have the key. We just take the letter from the second line and write the corresponding letter on the upper line.

For example, given the Atbash key decrypt the message HLXIZGREV.

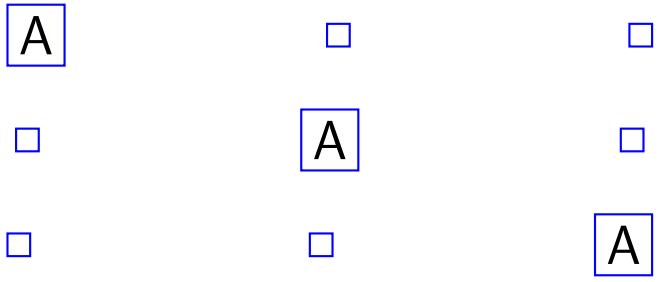
- We first find "H" on the second line (the key) and see that it represents "S"
- Next we find "L" on the second line and see that it represents "O".
- Continuing in this manner we get SOCRATIVE as the decrypted answer.

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

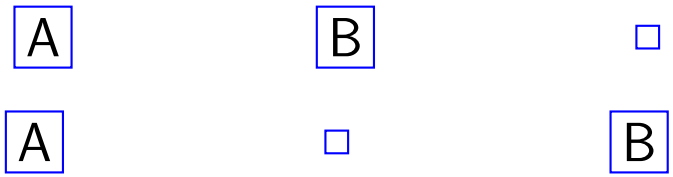
To decrypt a message (without a key) when a simple substitution cipher is used with the letters of the alphabet mixed up can we just use "Brute Force", i.e., try all possibilities?

- Clearly we can't try all possible combinations of the alphabet by hand but is it reasonable to write a computer algorithm which would use the Brute Force method of trying all possible combinations as we did for the shifted encryption?
- The answer is NO!
- To see this we first look at a simplified alphabet.

First assume we have an alphabet of 3 letters, A, B, C. we see that there are 3 places to put the letter "A"

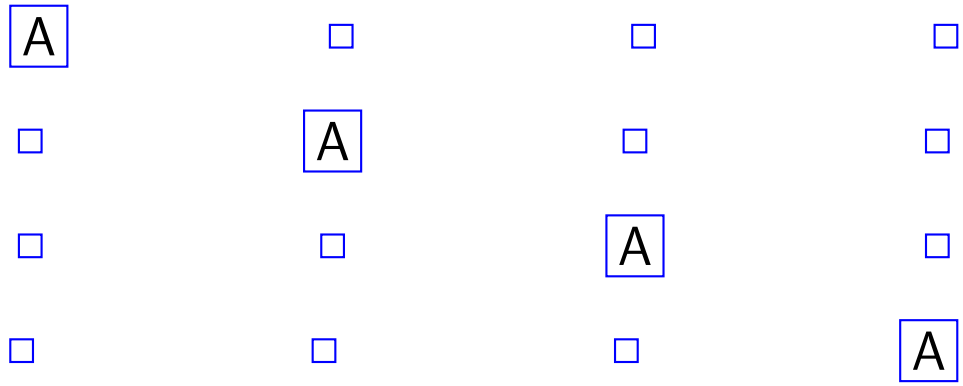


Now in each of the three cases there are two places to put the letter "B". For example,



And since there are 3 locations for "A" and for each location there are 2 places to put "B" we have 3×2 possibilities. Now in each of these six options there is only one place left for the letter "C" so we have $3 \times 2 \times 1 = 6$ possibilities.

Now assume we have an alphabet of 4 letters, A, B, C, D and we want to see how the number of possible combinations increase. We see that there are 4 places to put the letter "A"



Now in each of the four cases there are three places to put the letter "B". For example,

A	B	□	□
A	□	B	□
A	□	□	B

And since there are 4 locations for "A" and for each location there are 3 places to put "B" we have 4×3 possibilities. Now in each of these 12 options there are two places left for the letter "C". For example,

A	B	C	□
A	B	□	C

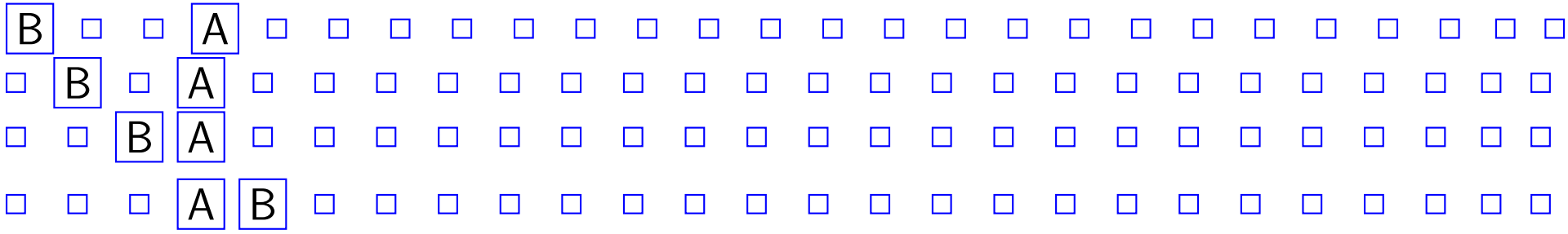
so we have $4 \times 3 \times 2 = 24$ possibilities. Finally in each of the 24 options there is only one place left to put the letter "D" so we get $4 \times 3 \times 2 \times 1 = 24$. The number of possibilities has increased by a factor of 4.

Now we want to do the same thing, except when we have 26 letters in the alphabet.

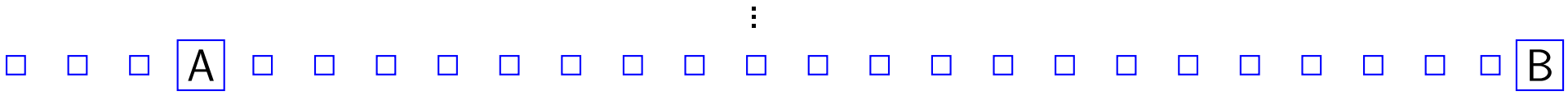
- First we see that there are 26 places to put the letter "A"

26 choices of where to put "A"

- After we have fixed the letter "A", we fix the letter "B". Now since "A" takes up 1 placeholder there are only 25 places where we can put the letter "B". But there are 26 different places where we can put "A" and for each of these choices there are 25 places to put "B". This means there are $26 \times 25 = 650$ possibilities for where "A" and "B" can go.



⋮



Now there are 26×25 possible combinations of "A" and "B".

- Once we fix "A" and "B" there are 24 places left to put "C". So for EACH of the 650 (26×25) possible combinations for "A" and "B" there are 24 possibilities of where to put "C". So far we have $26 \times 25 \times 24 = 15,600$ possible combinations.

- After we set the locations of "A", "B", and "C" there are 23 locations left to place "D". This means that for the first 4 letters of the alphabet there are $26 \times 25 \times 24 \times 23 = 358,800$ possibilities.
- After we have placed "A" – "Y" there is only one location left for "Z".
- So we have that the total number of possibilities is

$$26 \times 25 \times 24 \times 23 \times \cdots \times 2 \times 1 = 26!$$

- How big is 26!?

1!	1	
2!	2	
3!	6	
4!	24	
5!	120	
6!	720	
8!	40,320	
10!	3,628,800	
20!	2,432,902,008,176,640,000	
26!	403,291,461,126,605,635,584,000,000	$\approx 10^{26}$

Get the idea?

Example. The Russian alphabet consists of 33 letters.

1. If a substitution cipher is used for the Russian alphabet instead of our alphabet, will there be

   (a) more

   (b) less

   (c) the same number

   possible solutions to the encrypted message?

2. Which of the following indicates the number of possible keys for a substitution cipher using the Russian alphabet?

   (a) $26 \times 25 \times 24 \times \cdots \times 2 \times 1$

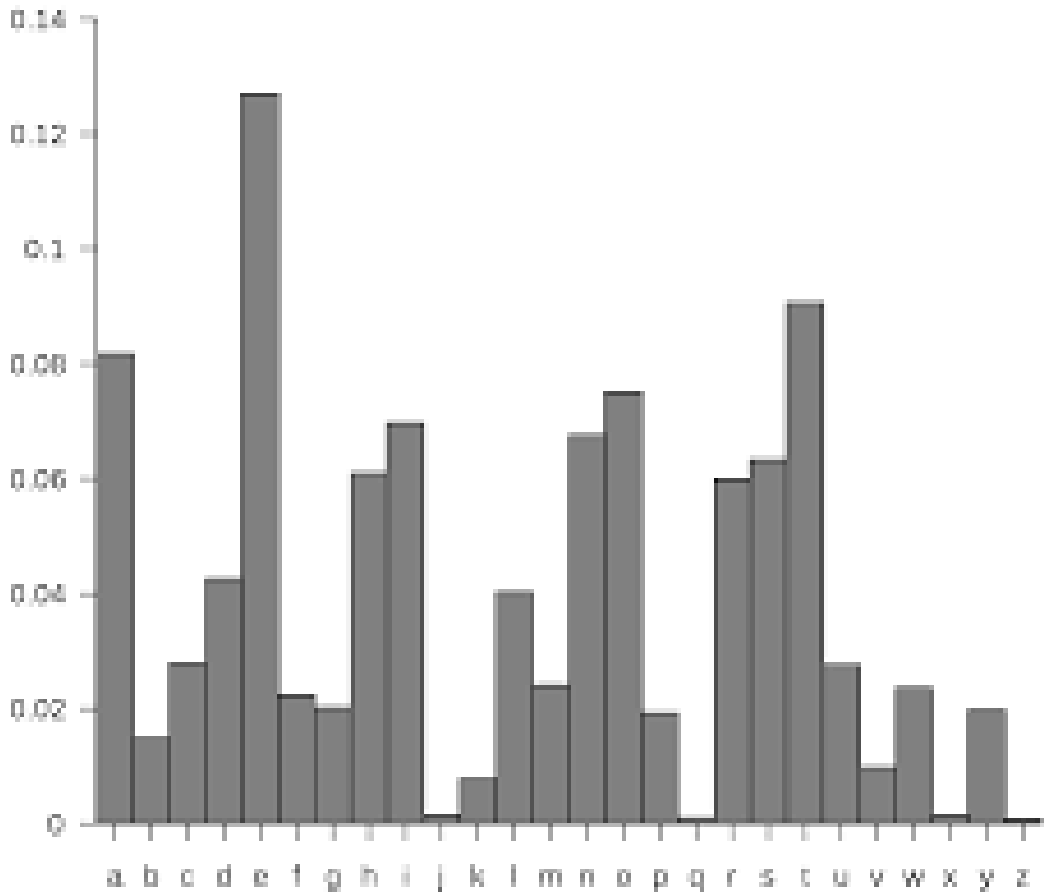
   (b) $33 \times 26 \times 25 \times 24 \times \cdots \times 2 \times 1$

   (c) $33 \times 32 \times 31 \times \cdots \times 2 \times 1$

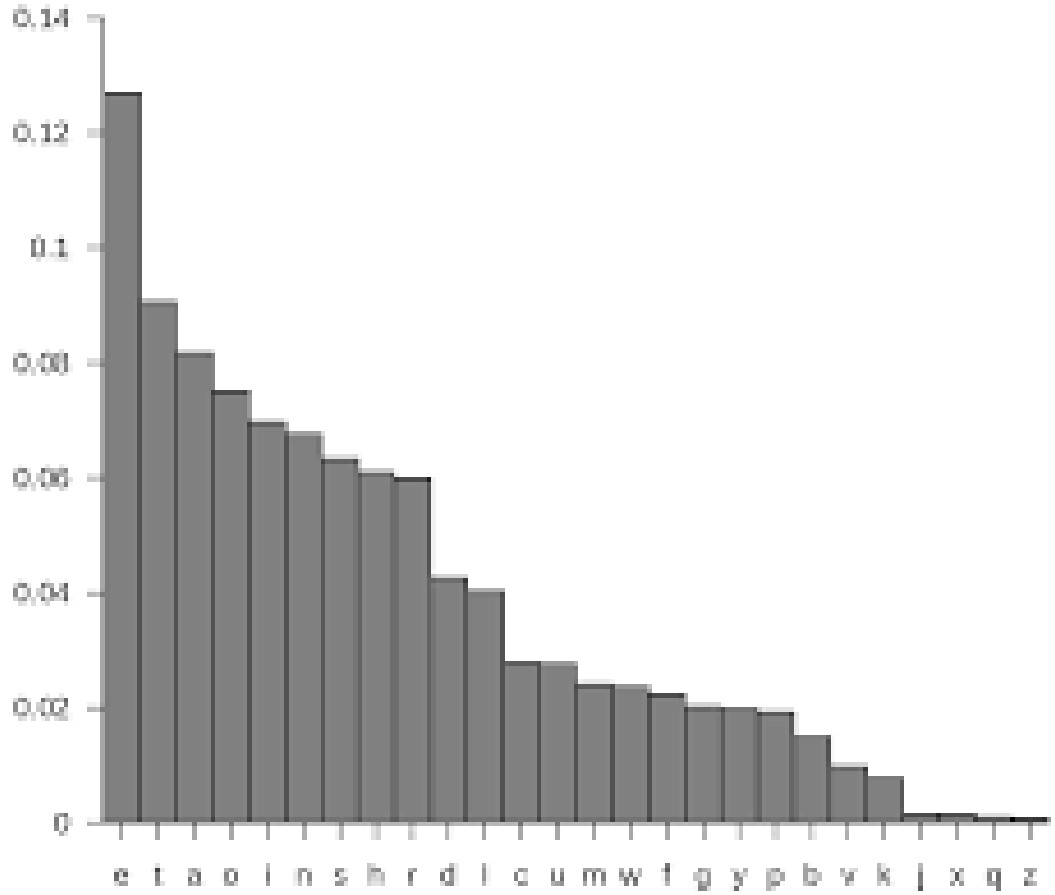
   (d) $(26 \times 25 \times 24 \times \cdots \times 2 \times 1)/33$

- Even with this simple substitution cipher the possibilities are too large to do an exhaustive search with a computer algorithm in a reasonable amount of time.
- What can we do instead?
- In any language certain letters (and words) occur more commonly than others.
- Before we said that in English the most commonly occurring letter is "E" and the second most commonly occurring letter is "T", the third most is "A". We can expand on this. Of course this type of frequency analysis works best for longer messages.
- If you were encrypting a code you probably wouldn't help someone trying to decrypt it by showing where one word ended and the next one started. If we assume there are no spaces indicated between words then one has to encode a letter or symbol for a space. This letter/symbol should occur most frequently.

Frequency plot for letters in the English language



Frequency plot in order of highest occurring to least occurring



Example Assume that you are on Wheel of Fortune and the following vowels and consonants have been guessed.

Vowels: E, A Consonants: N, S, T, R

1. Use the letter frequency plot to make an "educated" guess for the next vowel.

2. Use the letter frequency plot to make an "educated" guess for the next consonant.

Example - explanation Assume that you are on Wheel of Fortune and the following vowels and consonants have been guessed.

Vowels: E, A Consonants: N, S, T, R

1. Use the letter frequency plot to make an "educated" guess for the next vowel.

   From the plot, the order of the vowels from most frequent to least frequent is

   E, A, O, I, U, Y

   so if "E" and "A" have already been guessed, the next most frequent vowel is "O".

2. Use the letter frequency plot to make an "educated" guess for the next consonant.

   From the plot, the order of the consonants starting with most frequent is

   T, N, S, H, R, D, L, C, $\cdots$

   so if "N", "S", "T" and "R" have already been guessed, the next most frequent vowel is "H".

How can we use this frequency information to assist in decrypting a code?

- We first make a frequency count for the letters in the encrypted code. All we have to do is count how many times the letter appears in the encrypted message and tabulate or plot this.
- For simplicity we will only consider ciphers that use letters, i.e., no other symbols will be used. In addition, we assume that a space is encrypted with a letter of the alphabet. Of course this means that one letter of the alphabet doesn't appear in the plaintext.
- After we have the frequency of the encrypted letters we "guess" that the most frequent occuring encrypted letter is the space. The next most frequent encrypted letter might be "E".
- Of course the frequency plot for the encrypted letters might not directly correspond to those in the English language because these were determined by analyzing lots of texts, not just one passage. Later we will also use frequency of English words and letter combinations to decrypt the message.
- Of course a frequency count is more useful for longer messages than very short messages.

Example Make a letter frequency table for the following encrypted message. The message is given in groups of 5 letters but this does not indicate spaces in the message. The space is encrypted with a letter of the alphabet.

BMCAX TPAMH BMAPN BCANO
AHLLA BMHBA BMCDA TPBCP
RAUDA TPBCV FCBTN PAIMT
FMABM CDARV CHKAP NBANO

We now count how many times the encrypted letter "A" appears, how many times "B" appears, etc.

A	15	E	0	I	1	M	8	Q	0	U	1	Y	0
B	11	F	2	J	0	N	5	R	2	V	2	Z	0
C	8	G	0	K	1	O	2	S	0	W	0		
D	3	H	4	L	2	P	7	T	5	X	1		

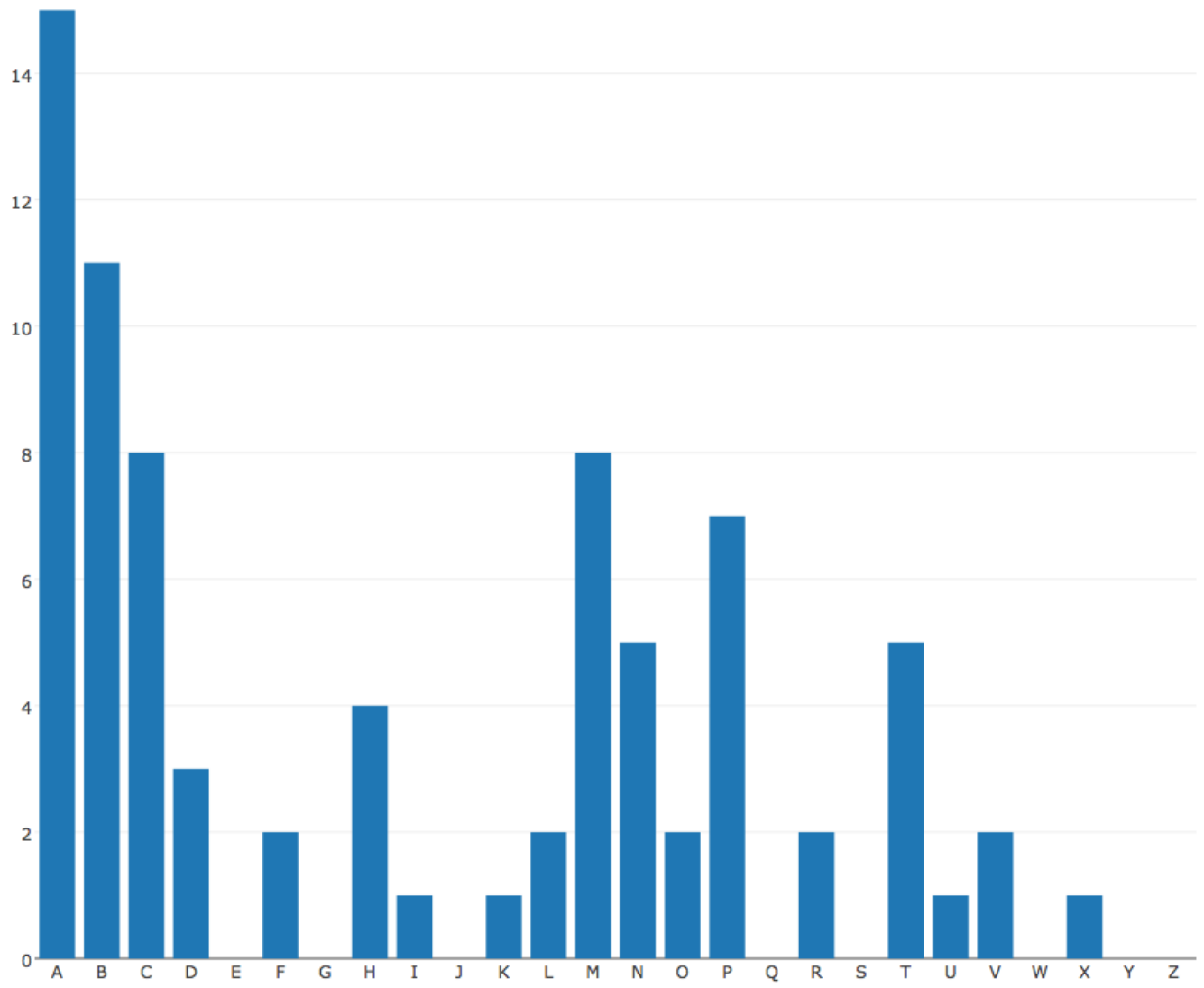
The letter "A" occurs the most often and so as a first guess we assume that it encrypts

a space so the word breaks are as follows.

BMC	XTP	MHBM	PNBC	NO
HLL	BMHB	BMCD	TPBCPR	
UD	TPBCVFCBTNP	IMTFM		
BMCD	RVCHK	PNB	NO	

Example Using the frequency table from the previous example, use PLOTLY to make a frequency plot.

A	15	E	0	I	1	M	8	Q	0	U	1	Y	0
B	11	F	2	J	0	N	5	R	2	V	2	Z	0
C	8	G	0	K	1	O	2	S	0	W	0		
D	3	H	4	L	2	P	7	T	5	X	1		


14
12
10
8
6
4
2
0
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

Example. Use the frequency table below for an encrypted message to answer the following questions.

A 1	E 2	I 5	M 3	Q 2	U 3	Y 7
B 7	F 2	J 9	N 1	R 1	V 5	Z 4
C 4	G 3	K 12	O 5	S 11	W 1	
D 6	H 1	L 15	P 9	T 3	X 20	

1. Which letter do you believe encrypts a space?
2. What letters would you guess first that encrypts the letter "E"?

- So far we know how to make a frequency table or plot for the letters occurring in the encrypted message.
- Another strategy that is useful in decryption is to look at the most commonly occurring words in the English language.
- When you do a Google search for "most common English words" you may find that different websites may have slightly different orderings of words. This is because different texts (and number of texts) have been used to count the number of occurrences of words.
- For the information in the next two slides we have used the website `www.world-english.org`
- On the next slide we list the English words starting with the most frequent.
- For our application it is more useful to list the word frequency per word length starting with the most frequent. In the following slide we do this. For example, for a 1-letter word there are only two choices – "a" and "I". The word "a" occurs more often than the word "I". For other words lengths there are more possibilities than we listed in the table but the most common ones are noted here.

Top 100 Most Commonly Used Words in English Language

Rank	Word	Rank	Word	Rank	Word	Rank	Word	Rank	Word
1	the	21	be	41	your	61	them	81	did
2	of	22	at	42	when	62	would	82	my
3	to	23	one	43	up	63	write	83	sound
4	and	24	have	44	use	64	like	84	no
5	a	25	this	45	word	65	so	85	most
6	in	26	from	46	how	66	these	86	number
7	is	27	or	47	said	67	her	87	who
8	it	28	had	48	an	68	long	88	over
9	you	29	by	49	each	69	make	89	know
10	that	30	not	50	she	70	thing	90	water
11	he	31	but	51	which	71	see	91	than
12	was	32	some	52	do	72	him	92	call
13	for	33	what	53	their	73	two	93	first
14	on	34	there	54	time	74	has	94	people
15	are	35	we	55	if	75	look	95	may
16	with	36	can	56	will	76	more	96	down
17	as	37	out	57	way	77	day	97	side
18	I	38	other	58	about	78	could	98	been
19	his	39	were	59	many	79	go	99	now
20	they	40	all	60	then	80	come	100	find

Most Frequent Words by Length (www.world-english.org)

1-letter words	a	I									
2-letter words	of	to	in	is	it	he	on	as	be	at	or
	by	we	up	an	do	if	so	go	no		
3-letter words	the	and	you	was	for	are	his	one	had	not	
	out	all	use	she	her	see	him	two	has	day	
	but	can	did	who	may	now					
4-letter words	that	with	they	have	this	from	some	what	were	your	
	when	word	said	each	time	many	then	them	like	long	
	make	look	more	come	most	over	know	than	call	down	side
	been	find									

Example. For a previous example we did a frequency analysis on an encrypted message and determined where the words begin and end by deciding which letter ("A") encrypted a space. We have the following encrypted message with the corresponding frequency table. With some given hints we want to see how to partially decrypt this message.

BMC XTP MHBM PNBC NO
HLL BMHB BMCD TPBCPR
UD TPBCVFCBTNP IMTFM
BMCD RVCHK PNB NO

A 15	E 0	I 1	M 8	Q 0	U 1	Y 0
B 11	F 2	J 0	N 5	R 2	V 2	Z 0
C 8	G 0	K 1	O 2	S 0	W 0	
D 3	H 4	L 2	P 7	T 5	X 1	

Now the most frequently occuring letters are "B" , "C" and "M". If we look at the first encrypted word we have BMC and two other places we have BMCD so a good

starting guess is

$$B \to T \text{ and } C \to E$$

What would you guess that "M" encrypts and "H" encrypts?

We have the 3-letter encrypted word BMC and so far we have part of it decrypted to give TME so we make the guess "M" encrypts "H".

We have a 4-letter word BMHB and so far we have THHT . Clearly H has to encrypt a vowel. We know that C encrypts "E" so it is either "A", "I", "O" or "U". The only one that makes sense is THAT so "H" encrypts "A".

What would you guess that "L" encrypts?

We have the 3-letter encrypted word HLL and we know that "H" is "A". If we look at the common 3-letter words beginning with "A" and having a double letter next we see that "ALL" is the only choice so "L" must encrypt itself.

Other Simple Substitution Ciphers

So far we have only mixed up the alphabet to obtain the key but we could also use a mixture of letters, numbers and symbols. For example, the key could be

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
B	1	&	A	8	4	H	*	T	!	5	X	N	+	L	=	7	V	W	U	3	G	6	S	Q	R

To encrypt a message we proceed as usual. For example, to encrypt GO NOLES we see that a "G" is represented by an "H", an "O" by an "L", an "N" by a "+", etc. Now if we want to encrypt the "space" then we choose a letter or symbol that hasn't been used in the key. For example the number "2" hasn't been used so if we use that the encrypted message is

HL2+LX8W

This is still a monoalphabetic substitution cipher and we can still use frequency analysis of English to decode the message.

Now let's see how to decrypt a message. Consider the encrypted message given in 5 symbol blocks.

*BG8C +LC48 BVCL4 C=8V4 8&UTL
+CQL3 C6TXX C+8G8 VCV8B &*CTU

We create the frequency table for the encrypted message

B	3	C	9	G	2	L	4	Q	1	U	2	V	4
X	2	3	1	4	3	*	2	+	3	=	1	&	2
8	7												

We see that "C" occurs the most often and so it probably encrypts a space. We have the message

*BG8 +L 48BV L4 =8V48&UTL+
QL3 6TXX +8G8V V8B&* TU

The number "8" is by far the next most frequent letter so we substitute an "E" and see if it seems to make sense.

*BGE +L 4EBV L4 =EV4E&UTL+
QL3 6TXX +EGEV VEB&* TU

If we concentrate on the 2-letter words we have three

+L L4 TU

The most common 2-letter words are

of, to, in, is, it, he, on, as, be , at , or, by, we, up, an, do, if, so, go , no

Notice that L is the second letter of +L and the first letter of L4 and of the listed 2-letter words only an "O" works. We have

*BGE +O 4EBV O4 =EV4E&UTL+
QO3 6TXX +EGEV VEB&* TU

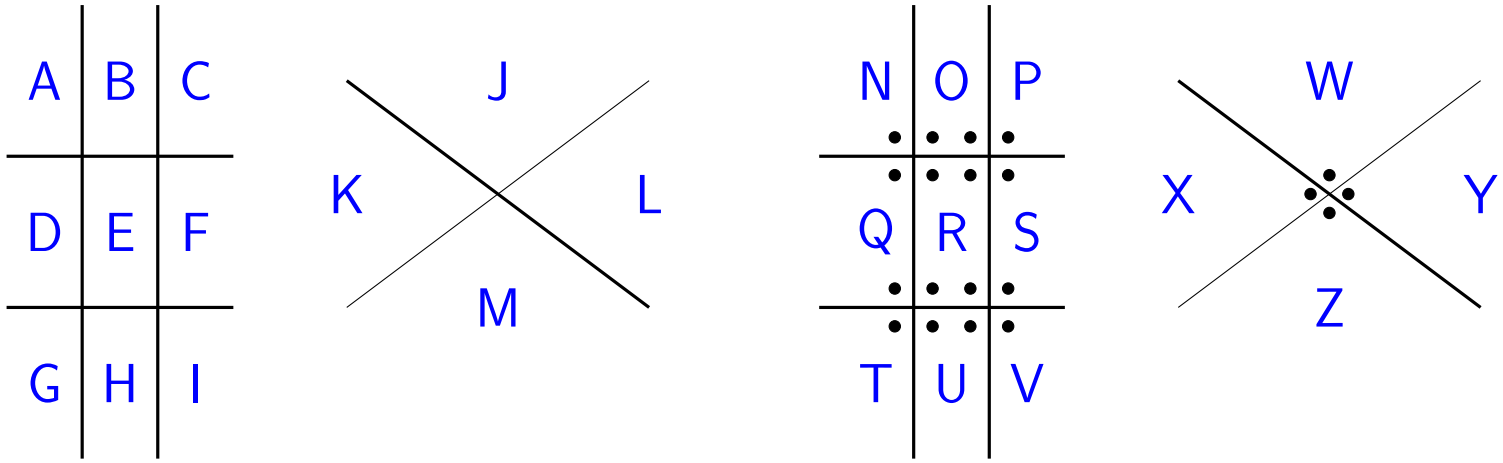
The most frequently occurring letter in English is "E" which we have decoded and the next is "T" and then "A". The letter V occurs 4 times, the letter B 3 times, the number 4 3 times, + three times. So these are all candidates for "T" and "A".

You can practice decoding by finishing this exercise. If you get stuck the key is just the one that is given above.

The Pigpen Cipher

Another simple substitution cipher is the Pigpen cipher also known as the Masonic cipher, Napoleon cipher, tic-tac-toe cipher. It is interesting because it is a geometric substitution cipher and doesn't just use letters or numbers it uses uncommon symbols.

To determine the key we write the letters of the alphabet on 4 grids. For example,



How do we encrypt letters using this key?

An A is encrypted as 

A B is encrypted as 

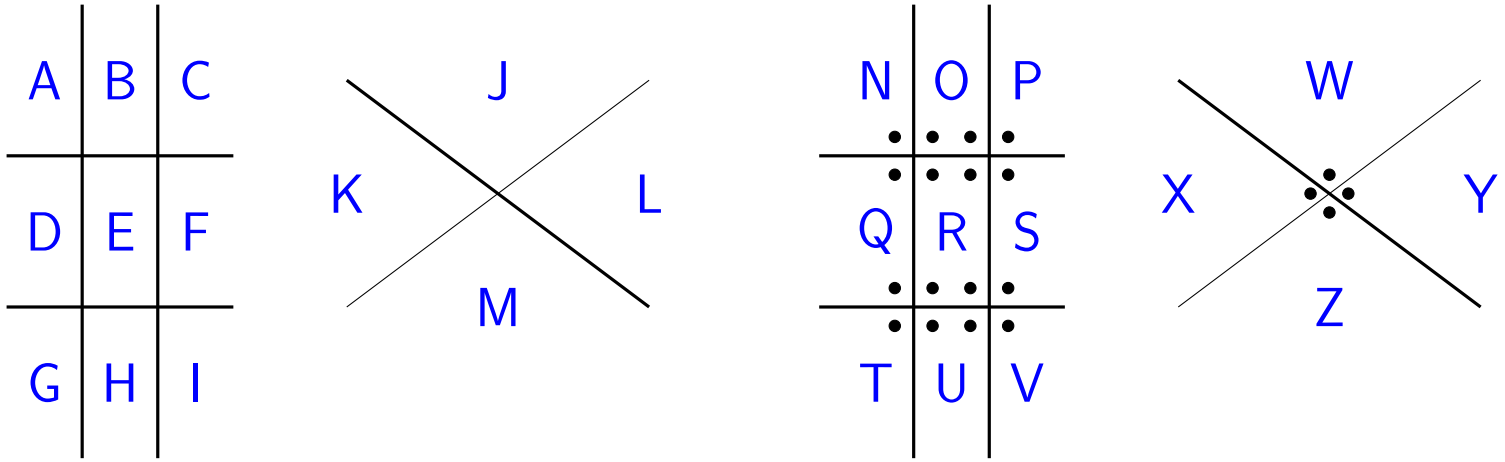
An M is encrypted as 

A P is encrypted as 

A Z is encrypted as 

Example. Use the key above to encrypt GO NOLES.

Explanation



G →

O →

N →

O →

L →

E →

S →

To decrypt a message using the Pigpen cipher we do exactly as before. We make a frequency table/plot of the symbols contained in the encrypted message and then use frequencies of letters and words in the English language.

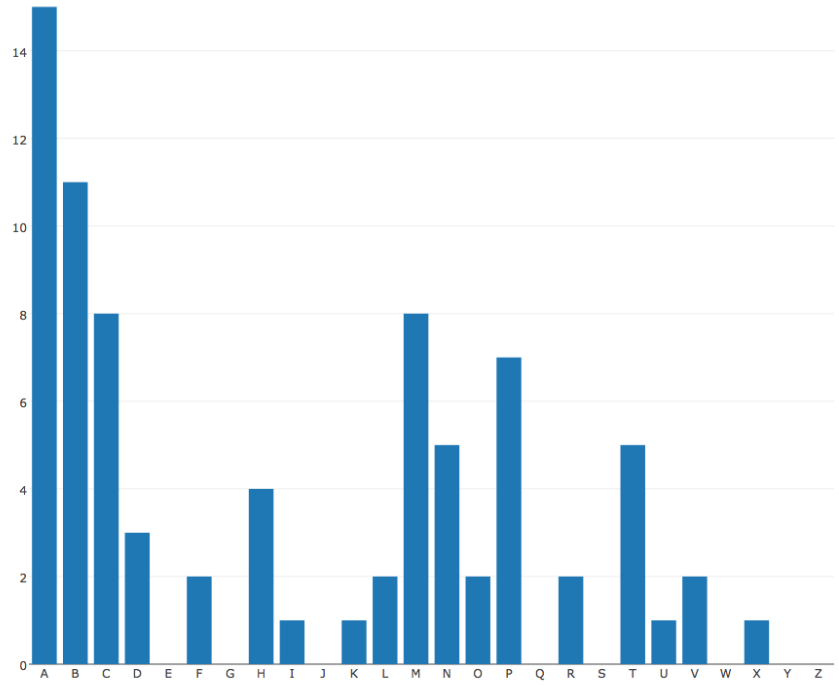
Even if we mixed up the order of the alphabet in the key above, we still approach the decryption in the same way and it still takes the same amount of work to decipher.

Weaknesses of Monoalphabetic Ciphers

- In the simple substitution ciphers we take the alphabet or set of symbols to be fixed throughout the encryption process. For example, "A" might always be encrypted as "P", "B" always encrypted as "3",etc. These are monoalphabetic ciphers.
- All monoalphabetic cipers are very susceptible to decryption.
- Why are they relatively easy to decrypt? Because although the symbols in the key themselves change, their frequency does not. So frequency analysis of the language of the original plaintext can be used. This frequency analysis approach was first documented by an Arabic mathematician in the 9th century.

SOCRATIVE QUIZ - PartIII_Quiz2

IUZGAZ34E



1. The frequency plot above represents the frequency of occurrence of letters in an encrypted message. Assuming the "space" is encrypted, what letter do you think represents it?
   a. A
   b. B
   c. C
   d. M
2. The frequency plot above represents the frequency of occurrence of letters in an encrypted message. How many letters appear 3 times?
   a. 1
   b. 2
   c. 3
   d. 4
3. The frequency plot above represents the frequency of occurrence of letters in an encrypted message. How many letters appear once?
   a. 1

b. 2

c. 3

d. 4

4. The frequency plot above represents the frequency of occurrence of letters in an encrypted message. How many letters of the alphabet don't appear in the encrypted message?

   a. 5

   b. 6

   c. 8

   d. 9

5. Which of the following ciphers is NOT a monoalphabetic cipher?

   a. Caesar Cipher with shift 2

   b. Atbash Cipher

   c. Pigpen Cipher

   d. None of the above

6. If we use a simple substitution cipher using 26 letters in random order, how many possibilities are there for the key?

a. 25
b. 25!
c. 26
d. 26!

7. If we do a letter frequency table or plot for an encrypted message where a space is NOT encrypted, what do typically the two most frequent letters represent?
   a. A, E
   b. E, N
   c. A, T
   d. E, T
8. The most common word in the English language is _____
   a. A
   b. I
   c. AN
   d. THE
9. If an encryption key contains both letters and numbers in a random fashion then it is more difficult to decrypt than one which just uses letters in random order.

10. A weakness of all monoalphabetic ciphers is that one can use frequency analysis for the language the plaintext is written in to decode the message.

Goals for lecture

1. In this lecture we will do an example where we completely decrypt a message which was encrypted using a monoalphabetic substitution cipher. We will use a strategy which uses frequency analysis and describe the approach in a step-by-step manner.

2. We will look at ways to improve the complexity of simple substitution ciphers.

Decryption of a Message Using a Monoalphabetic Substitution Cipher

Assume this encrypted message has 53 5-letter groups with no punctuation and assume that we know that a letter represents a space.

Step 1. Determine which letter encrypts a space by doing a frequency analysis and then write the message in terms of actual word length.

MJZYB	LGESE	CNCMQ	YGXY S	PYZD Z	PMYGI	IRL LC
PAYCK	YKGWZ	MCWZK	YFRCM	ZYVCX	XZLZP	MYXLG
WYTJS	MYGPZ	YWCAJ	M YCWS	ACPZY	XGLYZ	HSWBN
ZYXZT	YTGRN	VYMJC	POYMJ	SMYCX	YMJZL	ZYSLZ
YMTZP	MQYMJ	LZZYB	ZGBNZ	YCPYS	YLGGW	YMJZP
YMJZL	ZYCKY	SPYZD	ZPKYI	JSPIZ	YMJSM	YMJZL
ZYSLZ	YMTGY	GXYMJ	ZWYTC	MJYMJ	ZYKSW	ZYECL
MJVSQ	YERMY	MJCKY	CKYKG			

We now perform the tedious task of counting how many times "A" appears, how many times "B" appears, etc. to get the following table.

A	3	E	4	I	4	M	27	Q	3	U	0	Y	49
B	4	F	1	J	17	N	4	R	4	V	3	Z	33
C	18	G	14	K	9	O	1	S	14	W	9		
D	2	H	1	L	14	P	13	T	6	X	8		

- Since Y occurs most often by far we assume that it represents a space.
- We now replace Y with a space to see the word length.

MJZ
BLGESECNCMQ
GX
SP
ZDZPM
GIIRLLCPA

CK
KGWZMCWZK
FRCMZ
VCXXZLZPM
XLGW
TJSM

GPZ
WCAJM
CWSACPZ
XGL
ZHSWBNZ
XZT

TGRNV
MJCPO
MJSM
CX
MJZLZ
SLZ

MTZPMQ
MJLZZ
BZGBNZ
CP
S
LGGW

MJZP
MJZLZ
CK
SP
ZDZPK
IJSPIZ

MJSM
MJZLZ
SLZ
MTG
GX
MJZW

TCMJ
MJZ
KSWZ
ECLMJVSQ
ERM
MJCK

CK
KG

Step 2. We now list the encrypted words by length (for length 1-4) and their number of occurrences which is given in parentheses in the table below.

1-letter words	S (1)					
2-letter words	GX (2)	SP(2)	CK(3)	CX(1)	CP(1)	KG(1)
3-letter words	MJZ(2)	GPZ(1)	XGL(1)	XZT(1)	SLZ(1)	MTG (1)
	ERM(1)					
4-letter words	XLGW(1)	TJSM(1)	MJSM(2)	LGGW(1)	MJZP(1)	MJZW(1)
	TCMJ (1)	KSWZ(1)	MJCK(1)			

Step 3. We now look at the next most frequently appearing letters and use frequency analysis of letters and words to begin the decoding process. Start with short words.

- The next two most frequently appearing letters are Z (33 appearances) and M (27 occurrences) so these are candidates for E and T. We could try each of these and see if we could guess some words. In fact, we know that "the" is the most common 3-letter word and both "T" and "E" appear in it so it's a safe guess that a 3-letter word with M and Z is "THE"; of course it could be other words too (like "TIE"). The encrypted word MJZ appears twice and all other 3-letter words occur once so a good beginning guess is that

$$\begin{aligned} M &\to T \\ J &\to J \\ Z &\to E \end{aligned}$$

- We also note that there is a single 1-letter word which should be either "A" or "I" but we can't make a choice from this. However, notice that we have a four-letter word THST so we suspect it is THAT . Thus S $\to$ A

Below is the partially decrypted message. As before, a red letter means encrypted and a blue letter is decrypted.

THE	BLGEAECNCTQ	GX	AP	EDEPT	GIIRLLCPA
CK	KGWETCWEK	FRCTE	VCXXELEPT	XLGW	THAT
GPE	WCAHT	CWAACPE	XGL	EHSWBNE	XET
TGRNV	THCPO	THAT	CX	THELE	ALE
TTZPMQ	THLEE	BEGBNE	CP	A	LGGW
THEP	THELE	CK	AP	EDEPK	IHAPIE
THAT	THELE	ALE	TTG	GX	THEW
TCTJ	THE	KAWE	ECLTHVSQ	ERT	THCK
CK	KG				

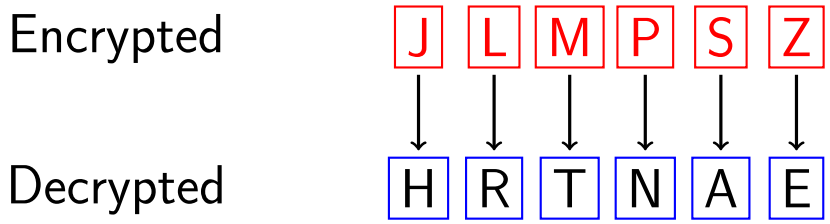
- We have the three-letter word ALE. We check the frequency of words by length and see that this is probably ARE making L an R.
- Also we have a word AP. Checking the frequency of two-letters words we have

"at", "an" and "as" but it can't be "at" because we know that M is T. Thus P must be S or an N. However, it can't be an S because look at the word THEP. If it is an S then this becomes THES whereas if it is an N then it becomes THEN. Thus P represents an N.

Making the replacements S $\rightarrow$A, L $\rightarrow$R, and P $\rightarrow$N. Our message becomes (where the blue letters are plaintext and the red are encrypted)

THE	BRGEAECNCTQ	GX	AN	EDENT	GIIRRRCNA
CK	KGWETCWEK	FRCTE	VCXXERENT	XRGW	THAT
GNE	WCAHT	CWAACNE	XGR	EHAWBNE	XET
TGRNV	THCNO	THAT	CX	THERE	ARE
TTZNMQ	THREE	BEGBNE	CN	A	RGGW
THEN	THERE	CK	AN	EDENK	IHANIE
THAT	THERE	ARE	TTG	GX	THEW
TCTJ	THE	KAWE	ECRTHVAQ	ERT	THCK
CK	KG				

So far we have decrypted the following letters



- Now we look at the remaining 2-letter words which we haven't decrypted; their number of occurences is in parentheses. .
  CK(3), CX(1), GX(2), KG(1)
- The most common 2-letter words (in order) are

  of to in is it he on as be at or by we up an do if so go no
- We already know that T is represented by M so we can rule out "to", "it', "at" so that leaves us with

of is in he on as be or by we up an do if so go no

- We also know that P decrypted is N and Z decrypted is E so we can rule out any of these 2-letters words which contain "N" or "E",i.e., "in", "on", "an", "no" and "he", "be", "we" which leaves us with the top ten possible 2-letter words

  of is as or by up do if so go

- Two of our code words start with a "C" (including the one that occurs the most) and 2 of our top English words begin with an "O" (including the most common and 4th most frequent) so let's guess that "C" is an "O" which gives

  $$\text{CK , CX , GX , KG} \implies \text{OK , OX, GX , KG}$$

  This suggests that K, X represent F ,R but we also have a 2-letter word KG so it doesn't look like K is either "F" or "R" because there are no common 2-letter word beginning with "F" or "R". So it looks like our guess that "C" represents an "O" is incorrect!

- So probably GX or KG represents OF. Now KG is the last word of the message so it doesn't make sense that KG would be OF so we guess that G represents O and X represents F. To reinforce this assumption look at the word VCXXERENT; if X represents F then this is probably DIFFERENT.

- The last encrypted word of the message is KG and since we guessed that G rep-

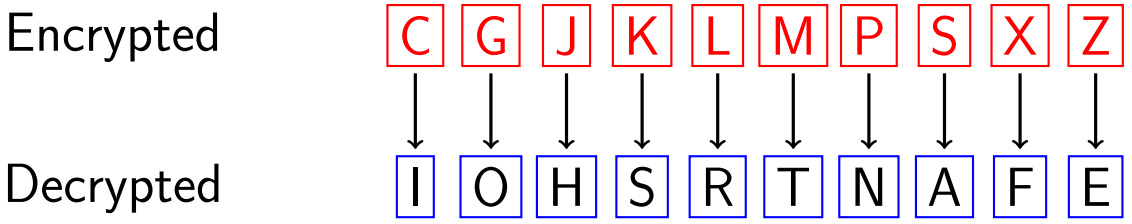
resents O we have KO. What 2-letter words end with "O" that could be the last word of the message? Continuing down the list of common 2-letter words we find the following words that end in "O" and don't contain the letters T, H, E, A, R, N, O, F which we have already decoded.

do, so, go

It's not clear how to proceed from here so let's move on to the other 2-letter words.

- The other 2-letter words are CK and CF. Remember that CK occurs the most frequently so our next most frequent 2-letter word is "IS" which would mean that C $\rightarrow$ I and K $\rightarrow$ S. This would mean that CF is IF and KO is SO; both make sense.

So far we have decrypted the following.



THE	BROEAEINITQ	OF	AN	EDENT	OCCRRRINA
IS	SOWETIWES	FRITE	VIFFERENT	FROW	THAT
ONE	WIAHT	IWAAINE	FOR	EHAWBNE	FET
TORNV	THINO	THAT	IF	THERE	ARE
TTENTQ	THREE	BEOBNE	IN	A	ROOW
THEN	THERE	IS	AN	EDENS	IHANIE
THAT	THERE	ARE	TTG	OF	THEW
TITJ	THE	SAWE	EIRTHVAQ	ERT	THIS
IS	SO				

SOCRATIVE QUIZ - PartIII_Practice_Quiz2

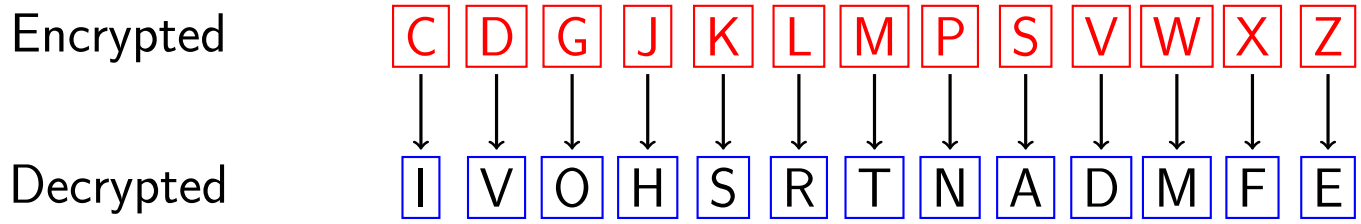
IUZGAZ34E

1. Looking at the partially decrypted words SOWETIWES and SAWE, what letter do you think that "W" encrypts?
2. Looking at the partially decrypted word VIFFERENT , what letter do you think that "V" encrypts?
3. Looking at the partially decrypted word EDENT , what letter do you think that "D" encrypts?

Using the facts (from the quiz) that

$$W \to M \qquad V \to D \qquad D \to V$$

we have so far decrypted the following letters which gives the partially decrypted message below.



THE	BROEAEINITQ	OF	AN	EVENT	OIIRRRINA
IS	SOMETIMES	FRITE	DIFFERENT	FROM	THAT
ONE	MIAHT	IMAAINE	FOR	EHAMBNE	FET
TORNV	THINO	THAT	IF	THERE	ARE
TTENTQ	THREE	BEOBNE	IN	A	ROOM
THEN	THERE	IF	AN	EDENS	IHANIE
THAT	THERE	ARE	TTG	OF	THEM
TITJ	THE	SAME	EIRTHDAQ	ERT	THIS
IS	SO				

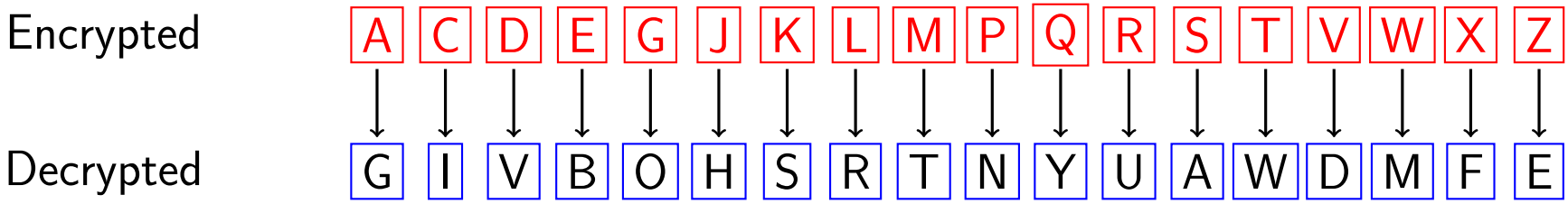
From the word MIAHT we see that A $\rightarrow$ G

From the words THAT and FET we see that T $\rightarrow$ W

This means that the word TTENTQ becomes TWENTQ so Q $\rightarrow$ Y

Now the word EIRTHDAQ becomes EIRTHDAY so E $\rightarrow$ B.

This makes sense because then ERT becomes BRT so R $\rightarrow$ U.



THE	BROBABINITY	OF	AN	EVENT	OIIURRING
IS	SOMETIMES	FUITE	DIFFERENT	FROM	WHAT
ONE	MIGHT	IMAGINE	FOR	EHAMBNE	FEW
WOUND	THINO	THAT	IF	THERE	ARE
TWENTY	THREE	BEOBNE	IN	A	ROOM
THEN	THERE	IF	AN	EVENS	IHANIE
THAT	THERE	ARE	TWO	OF	THEM
WITH	THE	SAME	BIRTHDAY	BUT	THIS
IS	SO				

We need to know the following 6 encrypted letters to finish decrypting the message.

F, H, B, N, O, I

From the word FUITE we see that F $\rightarrow$ Q.

From the words OIIURRING and IHANIE we see that I $\rightarrow$ C.

From the word WOUND we see that N $\rightarrow$ L.

Now we have the word BEOBLE so B $\rightarrow$ P.

Now the word EHAMBNE now becomes EHAMPLE so H $\rightarrow$ X.

Lastly we have the word THINO we see that O $\rightarrow$ K.

The final decrypted message is :

THE	PROBABILITY	OF	AN	EVENT	OCCURRING	
IS	SOMETIMES	QUITE	DIFFERENT	FROM	WHAT	
ONE	MIGHT	IMAGINE	FOR	EXAMPLE	FEW	
WOULD	THINK	THAT	IF	THERE	ARE	
TWENTY	THREE	PEOPLE	IN	A	ROOM	
THEN	THERE	IS	AN	EVENS	CHANCE	
THAT	THERE	ARE	TWO	OF	THEM	
WITH	THE	SAME	BIRTHDAY	BUT	THIS	IS
SO						

Making the Cipher more Complex

- In the previous simple substitution ciphers the cipher alphabet is fixed throughout the encryption process. For example, "A" might always be encrypted as "P", "B" always encrypted as "3", etc. These are called monoalphabetic ciphers.
- All monoalphabetic cipers are very susceptible to decryption.
- What if we allowed "A" to be encrypted as "P" in one place but as "T" in another place. These types of methods are called polyalphabetic ciphers. An example is the Alberti cipher which was developed in the 15th century.
- Polyalphabetic ciphers make decryption much more difficult.
- What would a key look like for such a cipher? First, we look at the Alberti cipher.

Alberti Cipher

Recall that the Alberti cipher used a physical apparatus which consisted of two disks, one fixed and one movable.

For simplicity, assume that the large outer disk consists of the letters

ABCDEFGHIJKLMNOPQRSTUVWXYZ

which will represent the plaintext. Assume the inner disk consists of

abcdefghijklmnopqrstuvwxyz

which will be the encrypted letters.

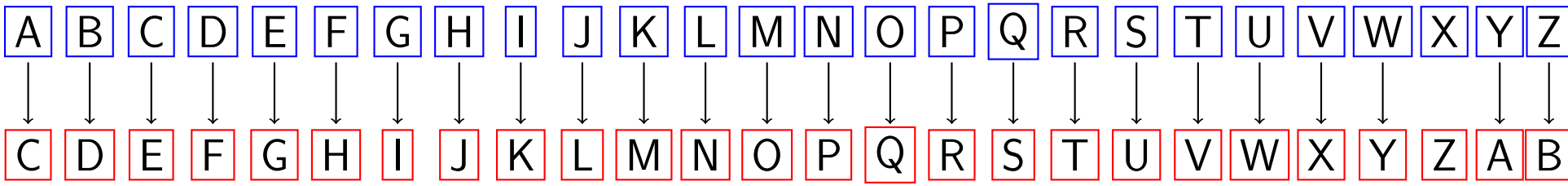
Assume you turn the disk so that "A" lines up with "a", "B" lines up with "b" etc. If you rotate the disk by 2 letters then "A" lines up with "c", "B" lines up with "d", i.e., you have done a Caesar shift to the right with shift two. Now choose a number, say 4, which we will call the period. You encrypt the first four letters of the plaintext and then you rotate the disk by one letter which gives a shift of three from the original. You encrypt the next four letters of the plaintext with this Shift 3 encryption and then rotate the disk again.

Example Encrypt the message

GO SEMINOLES

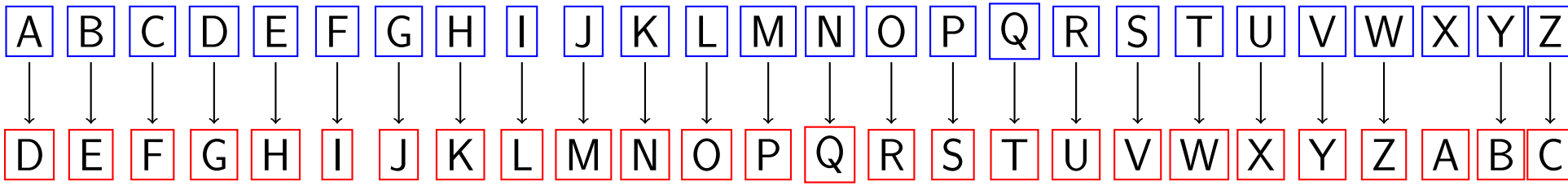
using the Alberti cipher with an original shift of 2, a period of 3 (i.e., every 3 letters rotate disk by 1) assuming the disks are set up as described above.

We begin by using a Caesar shift 2 to encrypt the first three letters of the message GO S



We have IQ UEMINOLES.

To encrypt the next 3 letters EMI we now rotate the disk by 1 to use a Caesar shift 3.



to encrypt EMI as HPL so that we have IQ UHPLNOLES.

Now we rotate the disk by 1 and use a Caesar 4 shift to encrypt NOL as RSP. Note that the "O" in GO and in SEMINOLES are encrypted differently.

Now we rotate the disk by 1 and use a Caesar 5 shift to encrypt ES as JX to get the encrypted message

IQ UHPLRSPJX

It is important to realize that if we do a frequency table for the letters occurring in the encrypted message, it won't help us because "O" is encrypted as "Q" in one place and as "S" in another. Likewise "S" is encrypted as "U" in the first occurrence and as "X" in the last occurrence. This is what makes polyalphabetic ciphers more difficult to break.

Vigenere Cipher

- In this cipher we use a text string such as a word as a key.
- Let's say the key is the word NOLES. We look at the numerical position in the alphabet of each letter in the word.

Letter	Position	Letter	Position	Letter	Position	Letter	Position	Letter	Position
A	1	B	2	C	3	D	4	E	5
F	6	G	7	H	8	I	9	J	10
K	11	L	12	M	13	N	14	O	15
P	16	Q	17	R	18	S	19	T	20
U	21	V	22	W	23	X	24	Y	25
Z	26								

We have

$$N \to 14, \quad O \to 15 \quad L \to 12 \quad E \to 5 \quad S \to 19$$

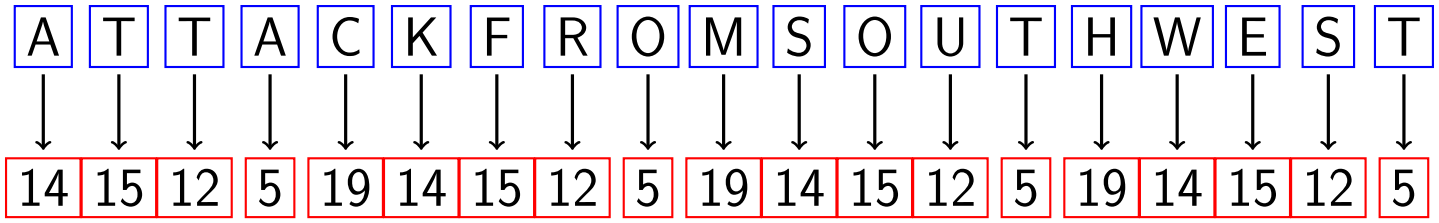
Our key is

$$14 \quad 15 \quad 12 \quad 5 \quad 19$$

- Now assume that the message we want to encrypt is

$$\text{ATTACK FROM SOUTHWEST}$$

- To encrypt the message we write below it the key 14,15,12,5,19 cyclically.



We then shift each letter of the plaintext by the number below it. We can use our position table to easily determine this. For example, the first "A" is replaced by the letter 14 positions to the right so since "A" is in the first position we add 14 to get the letter in the 15th position.

Letter	Position	Letter	Position	Letter	Position	Letter	Position	Letter	Position
A	1	B	2	C	3	D	4	E	5
F	6	G	7	H	8	I	9	J	10
K	11	L	12	M	13	N	14	O	15
P	16	Q	17	R	18	S	19	T	20
U	21	V	22	W	23	X	24	Y	25
Z	26								

- We shift "A" by 14 letters so it is replaced by the 15th letter in the alphabet which is "O"
- We shift the first "T" by 15 characters. Note that "T" is the 20th letter in the alphabet so if we shift it by 6 characters we get "Z" so we need to shift it by 9 more characters and the 9th letter of the alphabet is "I"
- The next "T" is shifted by 12 characters so we need the 6th letter of the alphabet so it is encrypted as "F".
- The next "A" is shifted by 5 characters and it is encrypted as "F"
- Continuing in this manner the encrypted message is then

OIFFVYUDTFGDGYAKTEY

- Again, frequency analysis doesn't help us to decode the message because, e.g., in one place "A" is replaced by "O" and in another place by "F".
- This code was easy to use because all you have to do is give the key word to the recepient and you can change the key word often. It was very hard to break before computers.

SOCRATIVE QUIZ - PartIII_Practice_Quiz3

IUZGAZ34E

Assume we have a Vigenere Cipher and the key is 5 2 7 15 21

1. If the first letter of the message is "N", what is this encrypted as?
2. If the second letter of the message is "O", what is this encrypted as?
3. If the third letter of the message is "T", what is this encrypted as?

Letter	Position	Letter	Position	Letter	Position	Letter	Position	Letter	Position
A	1	B	2	C	3	D	4	E	5
F	6	G	7	H	8	I	9	J	10
K	11	L	12	M	13	N	14	O	15
P	16	Q	17	R	18	S	19	T	20
U	21	V	22	W	23	X	24	Y	25
Z	26								

What if we are given the key and want to decrypt a message?

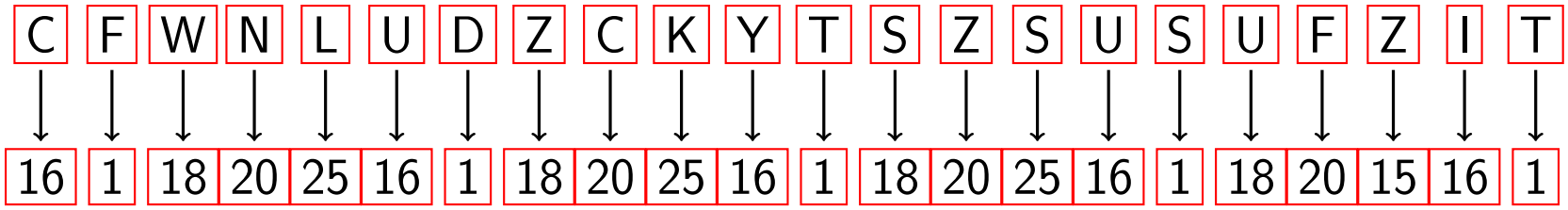
- Suppose we have the key word PARTY in our possession.
- We want to decrypt the message

  CFWNLUDZCKYTSZSUSUFZIT

- We first associate each letter in the key word with its numerical position in the alphabet.

  $$P \to 16, \quad A \to 1 \quad R \to 18 \quad T \to 20, \; Y \to 25$$

  Our key is now 16 1 18 20 25
- Now under each letter in the encrypted message write the corresponding number. To do this just write 16 1 18 20 25 in a cyclic manner.



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

- This says that the first encrypted "C" was encrypted by taking a letter of the plaintext alphabet and shifting it 16 letters to the right. So we have to shift to the left (or backwards) 16 letters to get the corresponding plaintext letter. C $\rightarrow$ M
- Now the second encrypted letter is "F" and it was encrypted by taking a plaintext letter and shifting one character to right so we shift to the left 1 character to get F $\rightarrow$ E
- Now "W" was formed by taking a plaintext character and shifting to the right 18 letters so we shift to the left 18 letters to get W $\rightarrow$ E

MEETMECHILISAFTERCLASS

MEET ME CHILIS AFTER CLASS

Note that the string SZSUS in the encrypted message forms a word (AFTER). Clearly there are three "S" letters in the encrypted message but each one represents a different letter!

Variants of Vigenere Cipher

- The key word (string of letters) is the same length as message.
- The key word is a randomly generated string of letters.
- The key word is only used once.
- If all three of the above criteria are used, then the cipher is called a one-time Pad
- One-time Pad is impossible to break by brute force approach (even with a computer). Even if you break it for one message, the key is different for the next message.
- Why? To break the ciphertext by brute force you need to try all possibilities (i.e., 26 letters) for each encrypted letter. So if the encrypted message is 10 letters long (which is short) , there are 26 plaintext choices for the first encrypted letter, 26 plaintext choices for the second, etc.

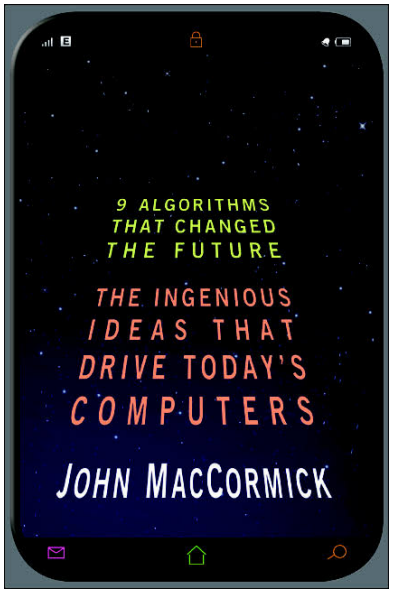
$$26 \times 26 \times 26 \times 26 \times 26 \times 26 \times 26 \times 26 \times 26 \times 26 = 26^{10} \approx 10^{14}$$

Reading Assignment for Next Time

Algorithm 4 in your text is

Public Key Cryptography: Sending Secrets on a Postcard

Read pages 38–43



SOCRATIVE QUIZ - PartIII_Quiz3

IUZGAZ34E

Letter	Position	Letter	Position	Letter	Position	Letter	Position	Letter	Position
A	1	B	2	C	3	D	4	E	5
F	6	G	7	H	8	I	9	J	10
K	11	L	12	M	13	N	14	O	15
P	16	Q	17	R	18	S	19	T	20
U	21	V	22	W	23	X	24	Y	25
Z	26								

1. Which of the following is NOT a monoalphabetic cipher?
   a. Caesar cipher
   b. Substitution cipher using symbols instead of letters
   c. Atbash cipher
   d. Vigenere cipher
2. Which of the following is a polyalphabetic cipher?

a. Caesar cipher

b. Alberti cipher

c. Vigenere cipher

d. (b) and (c)

3. The main difference between a monoalphabetic cipher and a polyalphabetic cipher is that a letter such as "A" is always encrypted the same way in a polyalphabetic cipher but in a monoalphabetic cipher it can be encrypted by different letters in different places.

4. The reason that a polyalphabetic cipher is more difficult to decrypt than a monoalphabetic cipher is that frequency analysis can't be used.

5. In a Vigenere cipher if the key word is FLORIDA then to encrypt the phrase

MEETMETONIGHTUSUALPLACE

the first "E" is encrypted by shifting how many letters?

(a) 6

(b) 7

(c) 12

(d) 15

6. In a Vigenere cipher if the key word is FLORIDA then to encrypt the phrase

MEETMETONIGHTUSUALPLACE

the second "T" is encrypted by shifting how many letters?

(a) 1
(b) 2
(c) 4
(d) 9

7. In a Vigenere cipher if the key word is FLORIDA then to encrypt the phrase

MEETMETONIGHTUSUALPLACE

the last 'E" is encrypted by shifting how many letters?

(a) 1
(b) 2
(c) 4
(d) 9

8. In a Vigenere cipher if the key word is FLORIDA then to encrypt the phrase

MEETMETONIGHTUSUALPLACE

the "N" is encrypted by shifting how many letters?

(a) 6
(b) 12
(c) 15
(d) 18

9. In a Vigenere cipher if the key word is FLORIDA then to encrypt the phrase

MEETMETONIGHTUSUALPLACE

the last "E" is encrypted by shifting how many letters?

a. 6
b. 12
c. 15
d. 18
e. 9

10. In a Vigenere cipher if the key word is FLORIDA then to encrypt the phrase

MEETMETONIGHTUSUALPLACE

the first "O" is encrypted as

a. P
b. S
c. T
d. U

Goals for this Lecture

1. To discuss some of the differences between Classical Cryptography and Modern Cryptography
2. To introduce the concept of binary digits.
3. To introduce the two types of Keys in modern cryptography - Public and Private.
4. To introduce the two types of Encryption Systems.
5. To explain Symmetric Key Encryption and give examples.

Modern Cryptography

What are the major differences between Modern Cryptography and Classical Cryptography?

• While classical cryptography manipulates traditional characters (letters, digits) directly modern cryptography operates on binary bit sequences. What does this mean?

A binary digit can only be 0 or 1. In computer talk, binary digit is oftened shortened to bit. The reason that a binary digit is used is because it can be represented as on/off,

positive/negative, north/south, etc. Computers store their information in bits. Bits are the building blocks of computer data storage. A group of 8 bits is called a byte.

To begin to see how binary numbers correspond to our decimal numbers let's count from 1 to 20 in each system.

So 0 and 1 are easy in binary because they are the same as in our decimal system but what do we do when we want a binary number for 2? The same thing that we do when we increase 9 by 1 to get 10; so 2 in binary is just 10. Now for 3 in binary we do the same that we did when we increase 10 by 1 to get 11 so 3 is just 11 in binary. Thus 3 is the largest number we can make with 2 binary digits so for our system the largest two digit number is 99 so to increase it we go to 100. Thus 4 in binary is just 100.

Decimal	Binary	Decimal	Binary	Decimal	Binary	Decimal	Binary
1	1	6	110	11	1011	16	10000
2	10	7	111	12	1100	17	10001
3	11	8	1000	13	1101	18	10010
4	100	9	1001	14	1110	19	10011
5	101	10	1010	15	1111	20	10100

Notice how all odd numbers end in 1 in binary and even end in 0. When we say 10 in "binary talk" we say "one zero" not ten.

Binary Examples

1. Which is larger 101001 or 101010?
2. If 20 in binary is 10100, what is 21?
3. If 20 in binary is 10100, what is 22?
4. Write the largest 2 bit binary number. Write the largest 3 bit binary number. Write the largest 4 bit binary number.
5. There are 2 numbers which can be represented as a 2-digit binary number (2 and 3); there are 4 numbers which can be represented as a 3-digit binary number (100,101,110,111); there are 8 numbers which can be represented as a 4-digit binary number. How many different numbers can be represented by a 5-digit binary number? (In this problem we require that the first to the left is a 1 so that we don't consider numbers like 001 as a 3-digit binary number.)
6. How many different numbers can a 6-digit binary number represent?

The storage on your computer is probably measured in gigabytes (GB). Now remember that a byte is 8 binary digits or bits and giga is a prefix meaning 10^9 so one GB is 1,000,000,000 bytes or 8,000,000,000 bits. But how much is that in our decimal system?

Let's see how to convert from binary to decimal but we will use smaller numbers.

We know that our decimal number system is based on 10; that is, we use digits from 0 to 9. For example,

$$58,913 = 3 \times 1 + 1 \times 10 + 9 \times 100 + 8 \times 1,000 + 5 \times 10,000$$

Mathematically we write this using exponents as

$$58,913 = 3 \times 10^0 + 1 \times 10^1 + 9 \times 10^2 + 8 \times 10^3 + 5 \times 10^4$$

So then each digit to the left of the decimal point increases by a power of 10. Remember that $10^2 = 10 \times 10 = 100$, $10^3 = 10 \times 10 \times 10 = 1000$, etc.

The binary digit system is based on 2 so it uses digits from 0 to 1. So if we have a binary number it can be expressed in powers of $2^0, 2^1, 2^2, \cdots$. Recall that this power notation just means multiply the base of 2 times itself as many times as the exponent says. For example

$$2^2 = 2 \times 2 = 4 \qquad 2^3 = 2 \times 2 \times 2 = 8 \qquad 2^4 = 2 \times 2 \times 2 \times 2 = 16$$

Example. How many digits does the binary number represented by

$$1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 0 \times 2^3 + 1 \times 2^4 + 1 \times 2^5$$

have?

We have 6 different powers of 2 so it will be a 6 digit binary number.

Example. What number does

$$1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 0 \times 2^3 + 1 \times 2^4 + 1 \times 2^5$$

represent in binary?

The right most digit is 1 because it multiplies 2^0, the second digit to right is 1, the

third is 0, the fourth is 0, the next is 1 and finally the digit to the left is 1 to give 110011.

To write a binary number in our decimal system we take the right most digit and multiply by $2^0 = 1$, the second to the right and multiply by $2^1 = 2$, the third to the right and multiply by $2^2 = 4$, etc. For example, we have

$$100 = 0 \times 2^0 + 0 \times 2^1 + 1 \times 2^2 = 4$$

$$111 = 1 \times 2^0 + 1 \times 2^1 + 1 \times 2^2 = 7$$

and

$$10011 = 1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 0 \times 2^3 + 1 \times 2^4 = 1 + 2 + 0 + 0 + 16 = 19$$

in base 10. Recall that we showed previously that the binary 10011 represented 19 by counting starting from 1 in binary.

So representing a binary number as a decimal number is easy.

Example. Write the 7 digit binary number 1000000 in our decimal system.

$$\begin{aligned} 1000000 &= 0 \times 2^0 + 0 \times 2^1 + 0 \times 2^2 + 0 \times 2^3 + 0 \times 2^4 + 0 \times 2^5 + 1 \times 2^6 \\ &= 0 + 0 + 0 + 0 + 0 + 1(2^6) = 2^6 = 64 \end{aligned}$$

Example. Write the 6 digit binary number 110011 in our decimal system.

$$\begin{aligned} 110011 &= 1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 0 \times 2^3 + 1 \times 2^4 + 1 \times 2^5 \\ &= 1 + 2 + 0(4) + 0(8) + 16 + 32 = 51 \end{aligned}$$

Now we do the opposite, that is we write a number in our decimal system as a binary number (without counting from 1).

Example. Write 7 as a binary number.

First, let's determine how many digits the binary number will have. To do this we see what is the largest power of 2 that is ≤ 7. In this case we know that $2^2 = 4, 2^3 = 8$ so the largest power is 2. This means we will have 3 digits because remember that the right most digit in the binary number is multiplied by 2^0. So we have

$$7 = 1 \times 2^2 + ? \times 2^1 + ? \times 2^0$$

Now we have 7-4=3 left to write in binary. Clearly $2^1 = 2 < 3$ so the coefficient of 2^1 is 1. Thus

$$7 = 1 \times 2^2 + 1 \times 2^1 + ? \times 2^0$$

So far we have $4 + 2 = 6$ so all we have to write is 1 so we have

$$7 = 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

which gives the binary number 111.

Example. Write 51 as a binary number.

First, let's determine how many digits the binary number will have. To do this we see what is the largest power of 2 that is ≤ 51. In this case we know that $2^5 = 32, 2^6 = 64$

so the largest power is 5. This means we will have 6 digits because remember that the right most digit in the binary number is multiplied by 2^0. So we have

$$51 = 1 \times 2^5 + ? \times 2^4 + ? \times 2^3 + ? \times 2^2 + ? \times 2^1 + ? \times 2^0$$

We have to represent 51-32=19. Now $2^4 = 16$ which is smaller than 19 so we have a 1 multiplying 2^4. We have

$$51 = 1 \times 2^5 + 1 \times 2^4 + ? \times 2^3 + ? \times 2^2 + ? \times 2^1 + ? \times 2^0$$

Now all we have to write is 19-16=3. $2^3 = 8$ which is not smaller than 3 so the coefficient of 2^3 is 0. Likewise $2^2 = 4$ so its coefficient is 0. But $2^1 = 2$ which is smaller than 3 so its coefficient is 1. We have

$$51 = 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + ? \times 2^0$$

So far we have $2^5 + 2^4 + 2^1 = 32 + 16 + 2 = 50$ so all we have to do is add a 1 to get

$$51 = 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

so the binary number is 110011.

We will need to convert binary numbers to decimal because typically the size of the storage allocated for a number is given in terms of binary numbers, not our base 10 numbers.

SOCRATIVE QUIZ - PartIII_Practice_Quiz4

IUZGAZ34E

Note:

$$2^0 = 1, 2^1 = 2, 2^2 = 4, 2^3 = 8, 2^4 = 16, 2^5 = 32$$

1. Which of the following is NOT a valid binary number?
   (a) 1010111
   (b) 1010211
   (c) 1111111
   (d) 1000000
2. The next binary number after 1111 is
   (a) 1112
   (b) 10000

(c) 10001

(d) 10010

3. What is the largest 5 digit binary number?

(a) 10000

(b) 10010

(c) 10101

(d) 11001

(e) 11111

4. What is the binary number 101 in our base 10 number system?

(a) 101

(b) 4

(c) 5

(d) 6

5. What is the binary number 1010 in our base 10 number system?

(a) 8

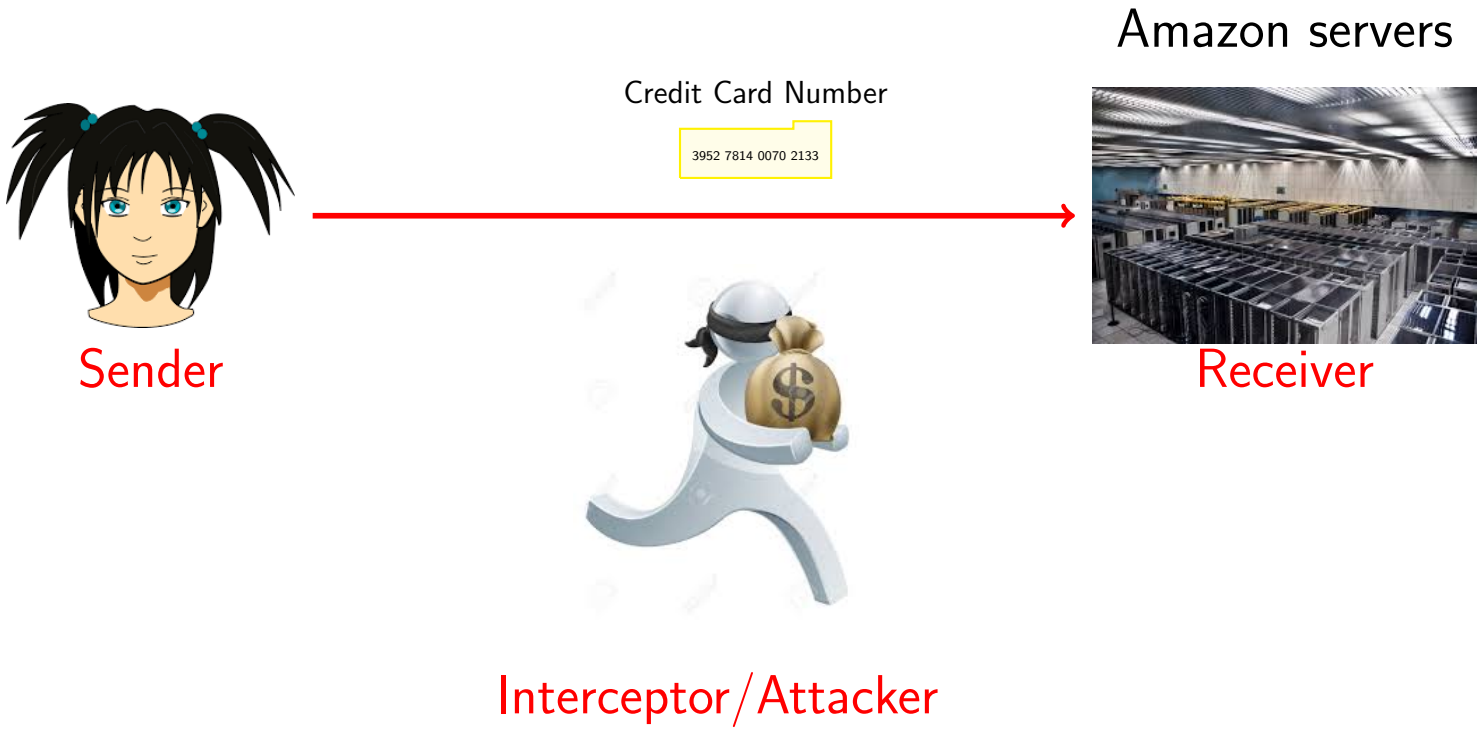
(b) 9

(c) 10
(d) 11

Differences between Classical and Modern Cryptography

- Classical cryptography is mainly based on "security through obscurity" while modern cryptography relies on mathematical algorithms for coding the information.
- In classical cryptography Keys were kept secret and only the parties encrypting and decrypting the messages know about them (in theory).
- In modern cryptography secrecy is obtained through a Key which is used to initialize an encrypting or decrypting algorithm. Even if the algorithm is known, it is impossible for an attacker to decrypt the message without a Key.
- Just like in our simple ciphers the encryption algorithm produces ciphertext given the original plaintext and the encryption Key. It's just that now it's a much more complicated procedure than simply shifting or substituting characters.
- Also we have a decryption algorithm that produces unique plaintext from the ciphertext and the decryption Key. For example, when a Shift Cipher was used our decryption algorithm just told us how far to shift our letters.

Modern Cryptography Problem



The sender wants to transfer sensitive data such as a credit card number to a receiver such as Amazon in such a way that an unauthorized person (the interceptor or attacker) can't get the information. The objective is that only the sender and receiver know the sensitive data.

Here's a simple example similar to the one in your text.

Suppose you want to tell your classmate a secret right now but your younger brother is in the room and you don't want him to know the secret.

Rule: You are only allowed to communicate by speaking out loud; i.e., you are NOT allowed to whisper in your friend's ear.

Simplifying Assumption: The "secret" you want to tell your friend is the number 15.

How can you do this? You think of something that both you and your classmate know but your brother won't know.

For example, suppose you have 7 math problems for homework tonight. Then you say double the number of math problems that we have for homework tonight and add 1. If your classmate remembers that you both have 7 problems then she computes

$$2 \times 7 + 1 = 15$$

and has the secret!

This is called a shared secret and the same basic idea is used in the simplest encryption algorithms.

Public Keys and Private Keys

In modern cryptography a Key is basically a long string of random numbers and letters. For example, the following string has 32 digits (in our base 10 decimal system). Remember though that the computer stores information in bits.

$$3048\ 0241\ 0019\ 1826\ 9786\ 2545\ 8912\ 9037$$

A Public Key is just what its name suggests – public. It is made available to everyone via a publicly accessible directory or repository.

A Private Key is just what its name suggests – private. Only the owner has the key. A Private Key can be shared as in our example.

The length of the Key determines the number of possiblities which in turn indicates how many steps it would take if one tried a Brute Force approach. We have already seen how to do this before. For example in a monoalphabetic cipher using our 26 letters there are $26 \times 25 \times \cdots \times 2 \times 1 = 26!$ combinations but with a polyalphabetic cipher using our 26 letters there are $26 \times 26 \times \cdots = 26^{26}$ possibilities.

However, just like memory on your computer, the Key length is described in terms of bits so we would like to know what actual length (in our decimal system) this corresponds to.

For example, if the Key length is 2 bits then we know that only 1,2,3 can be represented (as 1, 10, 11). Think of a bicycle lo so we have $3 \times 2 \times 1 = 3! = 6$ possible keys.

If the Key length is 3 bits then looking at our table where we counted to 20 we see that only 1-7 can be represented so we have $7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 7! = 5040$ possible keys.

If the Key length is 4 bits then looking at our table we see that the largest decimal number is 15 that can be represented so we have $15!$ possible keys.

Decimal	Binary	Decimal	Binary	Decimal	Binary	Decimal	Binary
1	1	6	110	11	1011	16	10000
2	10	7	111	12	1100	17	10001
3	11	8	1000	13	1101	18	10010
4	100	9	1001	14	1110	19	10011
5	101	10	1010	15	1111	20	10100

What is the largest number we can represent with 5 digits?

We know that it takes 5 bits to represent 20 but we can still represent a larger number with 5 bits. One way to determine this number is to keep counting in binary but an easier way is to realize that the largest 5 bit number is 11111 and ask what decimal number this represents. Recall that to do this we write

$$11111 = 1 \times 2^0 + 1 \times 2^1 + 1 \times 2^2 + 1 \times 2^3 + 1 \times 2^4 = 1 + 2 + 4 + 8 + 16 = 31$$

So using a binary key of length 11111 means we have numbers 1-31 in our decimal system. But this means we can only represent a 2 digit number up to 31 which would not be a very secure Key.

However, we would like to know a "rule of thumb" to convert a large binary number to decimal. We have the following formula for approximating this relationship.

If m is the number of binary digits in a number, then the number of decimal digits n is approximated by

$$n \approx 0.3m = 30\%(m)$$

For example, a 56-bit length Key is really a Key length of approximately $.3 \times 56$ decimal digits or about 17 decimal digits.

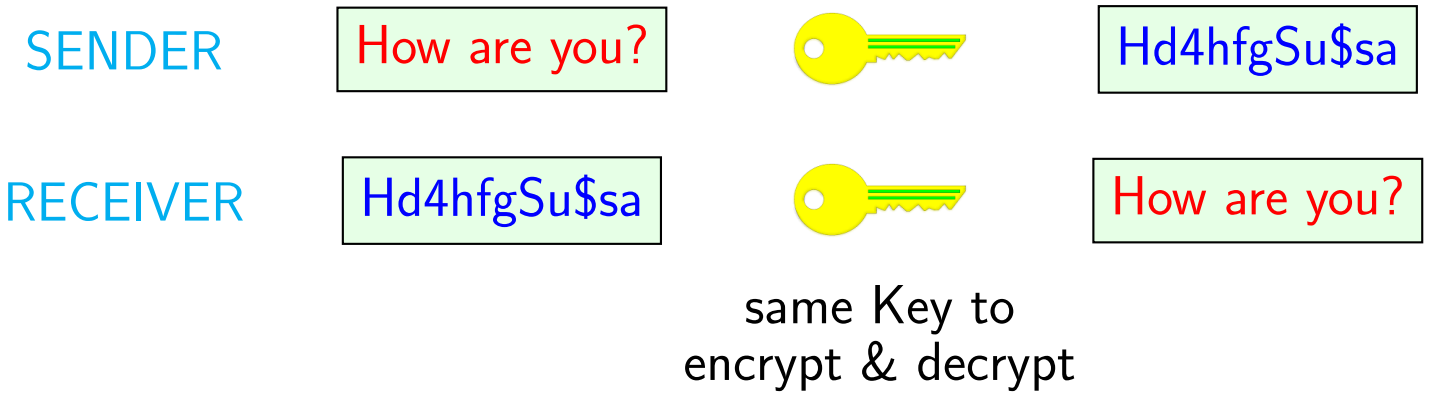
How did we get this "30%" rule?

It turns out that you actually need to use logarithms to show this. However, for our needs we will just make a table comparing the number of binary digits to the number of base 10 digits in a number. As you can see from the table, 30% is a rough estimate (this actually comes from the fact that $\log_{10} 2 = 0.301$

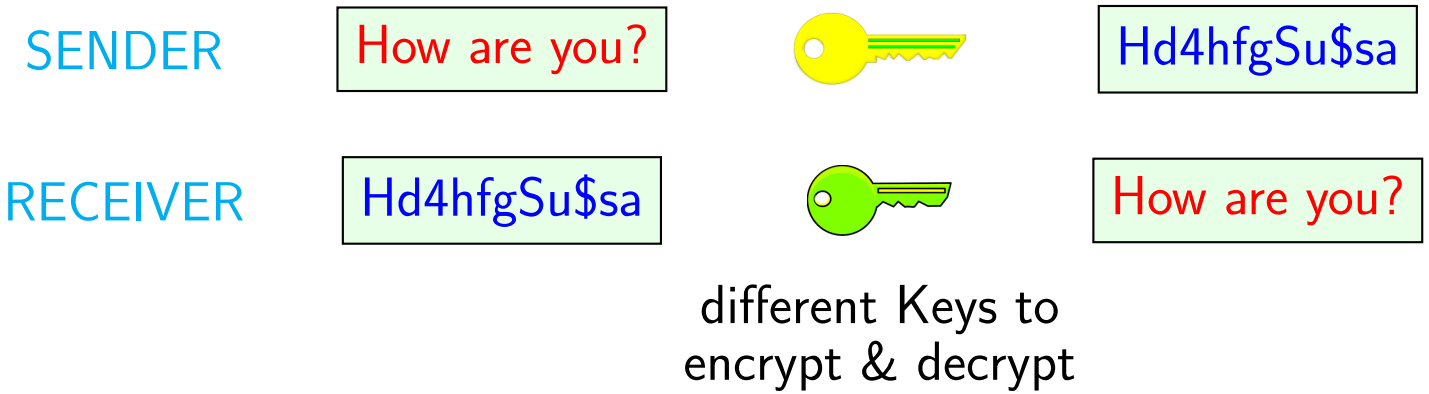
	# binary digits	# base 10 digits	Percent
$2^4 = 16$	5	2	40%
$2^5 = 32$	6	2	33%
$2^6 = 64$	7	2	28.5%
$2^7 = 128$	8	3	37.5%
$2^8 = 256$	9	3	33%
$2^9 = 512$	10	3	30%
$2^{10} = 1024$	11	4	36.4%
$2^{11} = 2048$	12	4	33%
$2^{12} = 4096$	13	4	30.8%
$2^{13} = 8192$	14	4	28.5%
$2^{14} = 16384$	15	5	33%

Types of Modern Encryption

- There are two basic types of encryption
  - Symmetric Encryption
  - Asymmetric or Public Key Encryption
- Symmetric encryption uses the same Keys for encrypting and decrypting.
- The ciphers we have considered so far (shift, simple substitution, Vigenere) all use symmetric encryption.

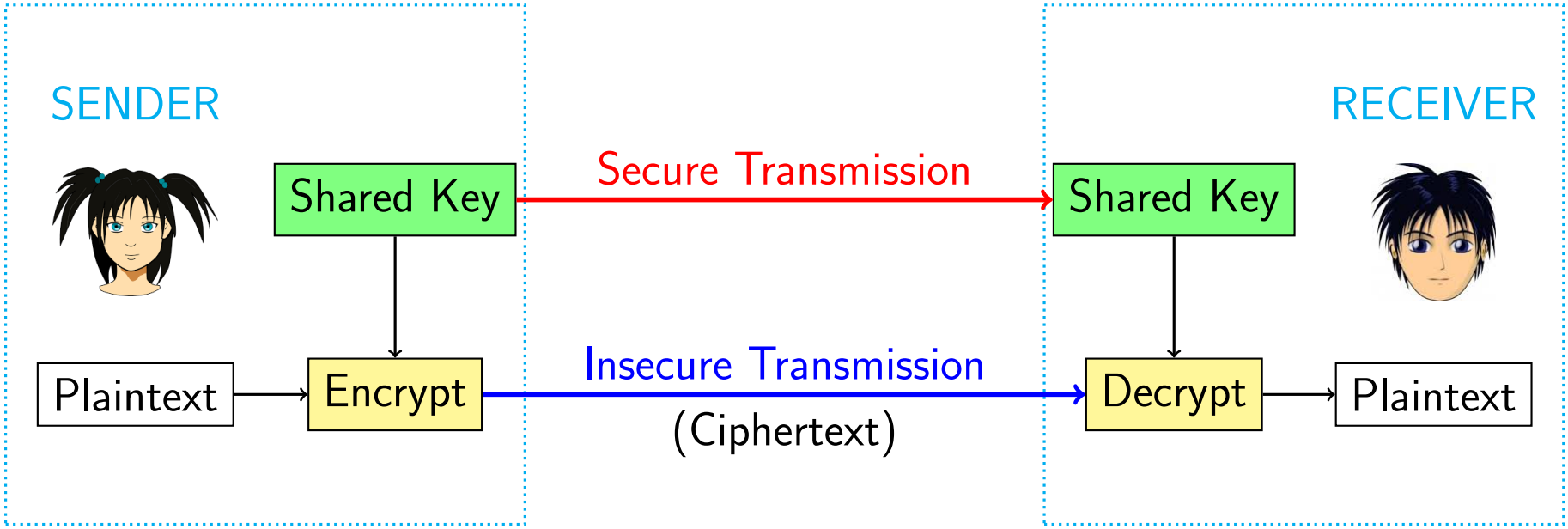


- Public Key (Asymmetric) encryption uses different Keys for encrypting and decrypting.



- Only the Green Key can decrypt what the Yellow Key encrypts.
- This takes sophisticated mathematical algorithms.

Symmetric Encryption System



Both Sender and Receiver agree upon a Private Key which is shared in a secure manner. The Sender encrypts the message with the Private Key and because it is encrypted she can send it using an insecure method. The Receiver also has the Private Key and uses it to decrypt the message.

The simple example of telling your classmate a secret without your brother knowing

is a Symmetric Encryption which used the Shared Private Key of knowing how many math homework problems were assigned.

Advantages of Symmetric Encryption Systems

- Same Key is used to encrypt the data and to decrypt the data so only one Key is needed
- Generally faster than Public Key encryption systems

Disadvantages of Symmetric Encryption Systems

- Since both sender and receiver need the same Key it must be stored in a secure location that is accessible to both parties
- There needs to be a secure channel to transfer the Key
- Having to store the Key in a location, even if it is "secure", lends itself to attacks.
- Keys should be changed regularly to prevent attacks

Cast of Archetypal Characters in Cryptography

- Bob and Alice are used for standard characters in cryptography. Saying "Alice wants to send Bob a message" is more natural to say than "Party A wants to send Party B a message".
- Eve is an eavesdropper and is usually a passive attacker. She listens in on messages between Bob and Alice but can't modify the message.
- Mallory is a malicious attacker who can modify messages, substitute their own messages, etc. Securing a system against Mallory requires much stronger security measures than against Eve.
- Faythe is a trusted advisor or intermediary such as a repository of key service. May be a machine or a human.
- Craig is a password cracker.
- There are many other archetypal characters; see Wikipedia, "Alice and Bob" entry or the article "Alice and Bob: IT's inseparable couple". Check it out to see if your name is a standard cryptography character!

A Simple Example of a Symmetric Key Encryption

Alice wants to send a private message to Bob via postal mail. Unfortunately, she is convinced that the postman is reading the mail that she sends.

What can she do? She decides to put the message inside a lockbox, then mail the box to Bob. She buys a lockbox with two identical keys.

She doesn't trust her postman so she is afraid if she mails a key to Bob then he might make a copy. So she has to set up a meeting with Bob to give him a key but she only has to do this once, no matter how many messages she sends.

Alice then writes her message and locks it in the lockbox and mails it to Bob.

The mailman can see the lockbox but he can't open it. There is no way the postman can read the message so it is secure!

Bob can read the message because Alice and Bob each have a Key.

- This is a symmetric encryption system because the Key (in this case a physical key) is the same to encrypt (lock the lockbox) and decrypt (unlock the lockbox).
- The Key (here the physical object) only has to be securely sent once.
- This type of security was used in armored cars. At one location, items are loaded into the back of the van and an overseer locks the back with a key that he/she

keeps. An identical key has been given to the receiver of the goods. For security reasons the driver can't open the back so it can't be opened until it gets to its final destination.

But what if Bob lives in New Zealand?

TimeLine for First Symmetric Encryption Key Algorithms

May 1973	U.S. Government publishes a first request for a standard encryption algorithm
August 1974	U.S. Government publishes a second request for a standard encryption algorithm
March 1975	Data Encryption Standard (DES) published for review
November 1976	DES is approved as standard
1986	HBO begins using TV satellite scrambling system based on DES
June, 1997	Message encrypted with DES is publically broken
November 2001	Advanced Encryption Standard (AES) published
May 2002	AES becomes standard

DES (Data Encryption Standard)

1976, 56-bit Key length

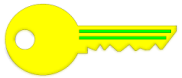
broken, 1997 (56 hours)

AES (Advanced Encryption Standard)

2001, 128-256 bit Key length

broken, 2009

Symmetric Key



Algorithms

Blowfish

1993, 448 bit Key length

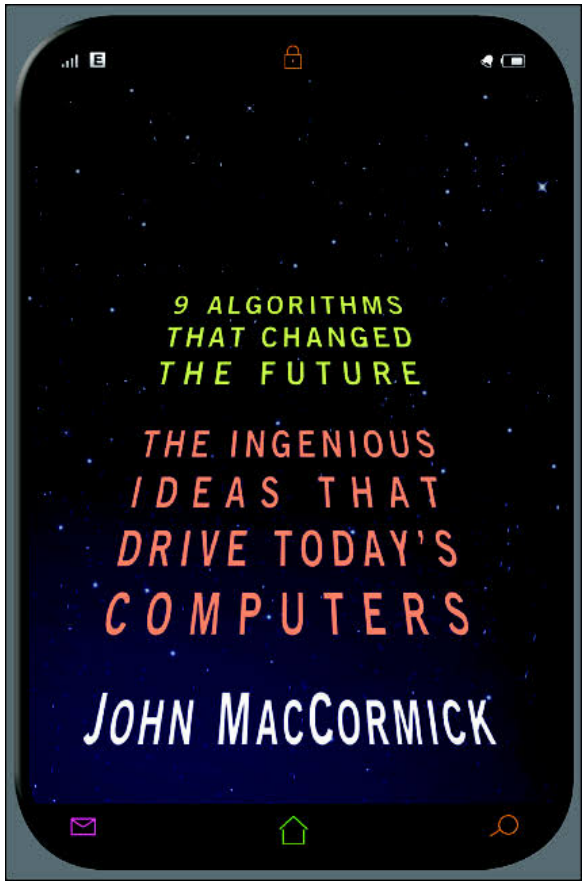
3 DES (Triple DES)

1998

International Data Encryption (IDEA)

1991, 128-bit Key length

Assignment: Read pages 44-59 of Algorithm 4



SOCRATIVE QUIZ - PartIII_Quiz4

IUZGAZ34E

1. Computers store information as _____.
   a. binary digits
   b. decimal numbers
   c. bits
   d. both (a) and (c)
   e. both (b) and (c)
2. The binary number 10011 is larger than the binary number 10100.
3. All binary numbers which end in a 0 represent even numbers in our decimal system.
4. If you are counting in binary numbers, what number comes after 111?

   (a) 1000

(b) 110
(c) 112
(d) 1001

5. If you are counting in binary numbers, what number comes after 1101?

(a) 1111
(b) 1110
(c) 1010
(d) 1001

6. The binary number 10 represents what decimal number?

(a) 9
(b) 5
(c) 11
(d) 3

7. Approximately how many digits in base ten (our system) do you think it would take to represent an 128-bit Key?

a. 128

b. 64

c. 38

d. 30

e. 19

8. In a(n) _____ encryption system, the same Key is used by both the sender and receiver.

a. symmetric

b. asymmetric

c. Public-Key

d. both (a) and (c)

e. both (b) and (c)

9. In a(n) _____ encryption system two different Keys are used.

a. symmetric

b. asymmetric

c. Public-Key

d. both (a) and (c)

e. both (b) and (c)

Goals for Lecture

1. To understand how Public Key Encryption differs from Symmetric Key Encryption
2. To understand the main idea behind Public Key Encryption by an analogy with mixing paint.
3. To introduce the concept of "clock arithmetic" and see how it is connected to modular arithmetic.
4. To get a glimpse of the mathematics behind Public Key Encryption by using modular arithmetic.

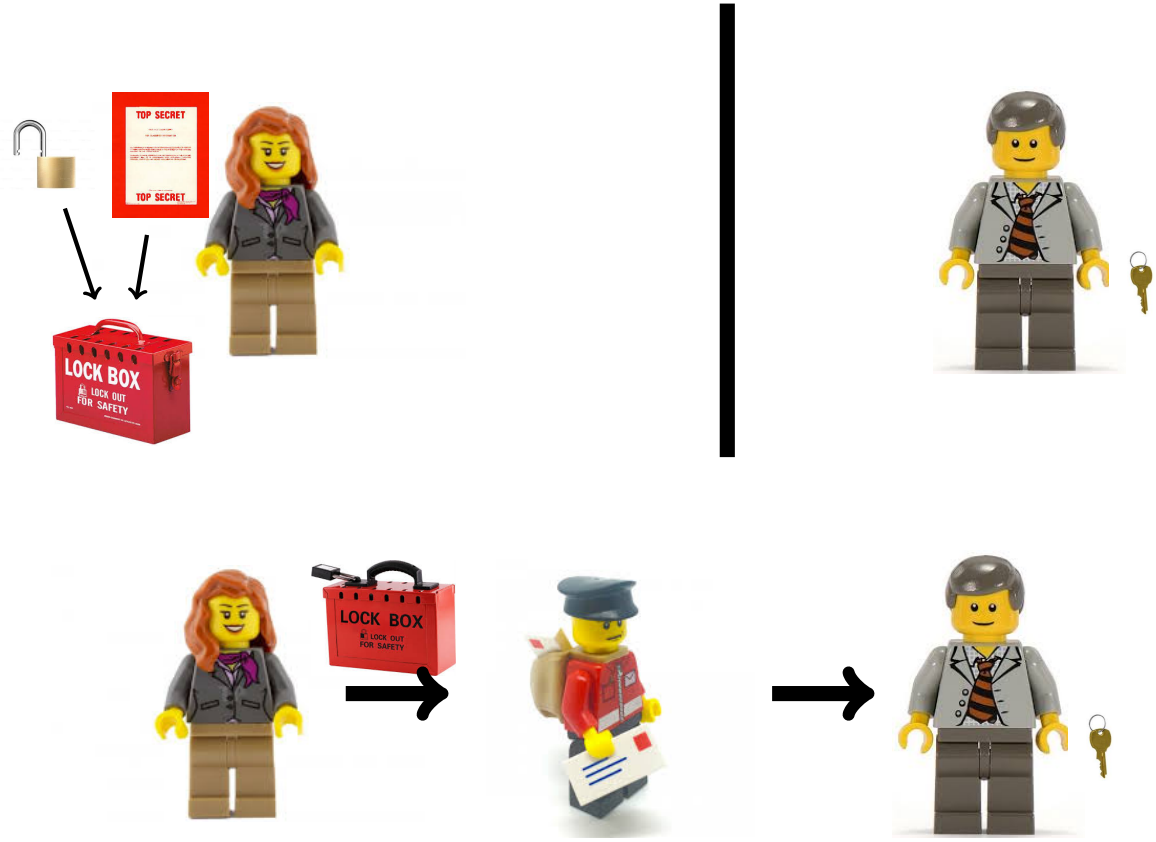
A Simple Example of Public Key Encryption

In Public Key Encryption one Key is used to encrypt the message and another Key is used to decrypt it.

Alice wants to send a private message to Bob via postal mail. Unfortunately, she is convinced that the postman is reading the mail that she sends. The problem is that Bob lives in New Zealand and Alice lives in Florida so they can't easily meet.

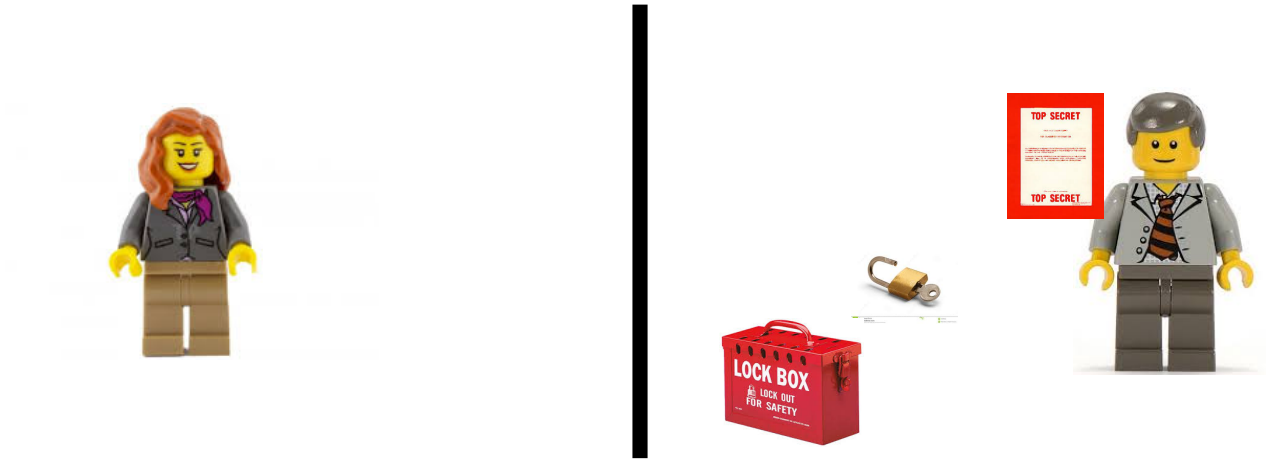
Bob buys a padlock with a key. He keeps the key and mails the unlocked padlock to Alice.

Alice buys a simple lockbox that closes with a padlock. She writes her message, puts it in the lockbox and locks it with the padlock Bob sent her and mails it back to Bob.



The postman can't read the message because there is only one key to the padlock which Bob has.

When Bob gets the package he uses his key to open the lockbox and read the message.



In order for Alice to send another message to Bob, he has to mail the unlocked lockbox and the opened padlock back to Alice.

It is Public Key system (asymmetrical) because Bob's Private Key is the physical key to decrypt the message (i.e., unlock the box) and the encryption is done by the padlock which is the Public Key because it was sent by regular mail so anyone could duplicate it.

Of course this only works for sending messages one-way. For two-way messages they each would have to have their own padlock and keep the key for it.

Shared Secret Key

- We know that in Public Key Encryption a different key is used to encrypt the message and another one to decrypt the message.
- In practice, Public Key Encryption typically uses a Shared Secret Key. Remember in the example when you wanted to tell your classmate a secret without your younger sibling hearing it you had to find a way for your classmate to figure our the secret without you saying it.
- The problem is that both the sender and receiver must figure out the shared secret Key from some public information without actually transmitting it back and forth.
- For example, the first time you sign on to a new secure website to make a purchase, then your computer and the receiver must agree on a secret code before your sensitive information is sent.
- To understand how this is accomplished we first look at an analogy of paint mixing and then we see how it could be done in practice.

A Simple Analogy for Forming a Shared Secret - Paint Mixing

Goal: Bob and Alice want to mix a secret paint color but Eve is eavesdropping and they don't want her to know their special color.

Strategy: Bob and Alice will mix the same color without Eve knowing and without directly sharing the formula for the new color. This formula will play the role of the shared secret key.

Assumptions:

1. Bob, Alice and Eve are in a room which has a curtained off area for Bob and another for Alice so that they can mix paint in private.
2. In the center of the room are paint pots of say 1,000,000 different colors.
3. There are several pots of each color of paint.

Rules:

1. Neither Bob nor Alice can give a paint pot directly to the other person.
2. If information (i.e., a paint pot) is to be shared, then it must be placed in the center of the room. Multiple copies of that particular color must be shared.

EVE

Bob and Alice want to mix a secret paint color without Eve knowing

Step 1. Before beginning, several pots of a single color have been placed in Bob's private corner and several pots of another color have been placed in Alice's private corner. We refer to this as Bob's or Alice's private color of paint. We assume that this was done without the knowledge of anyone else.

Let's assume that Bob has blue and Alice has crimson. Since there are a million different colors, the liklihood of Bob and Alice having the same color is small.

Only Bob knows that his private color is blue.

Only Alice knows that her private color is crimson.

Eve doesn't know either Bob's or Alice's private color.

Bob and Alice each have a secret paint color

Step 2. Alice chooses a public color of paint.

Alice announces to the room that the public color is pale yellow.

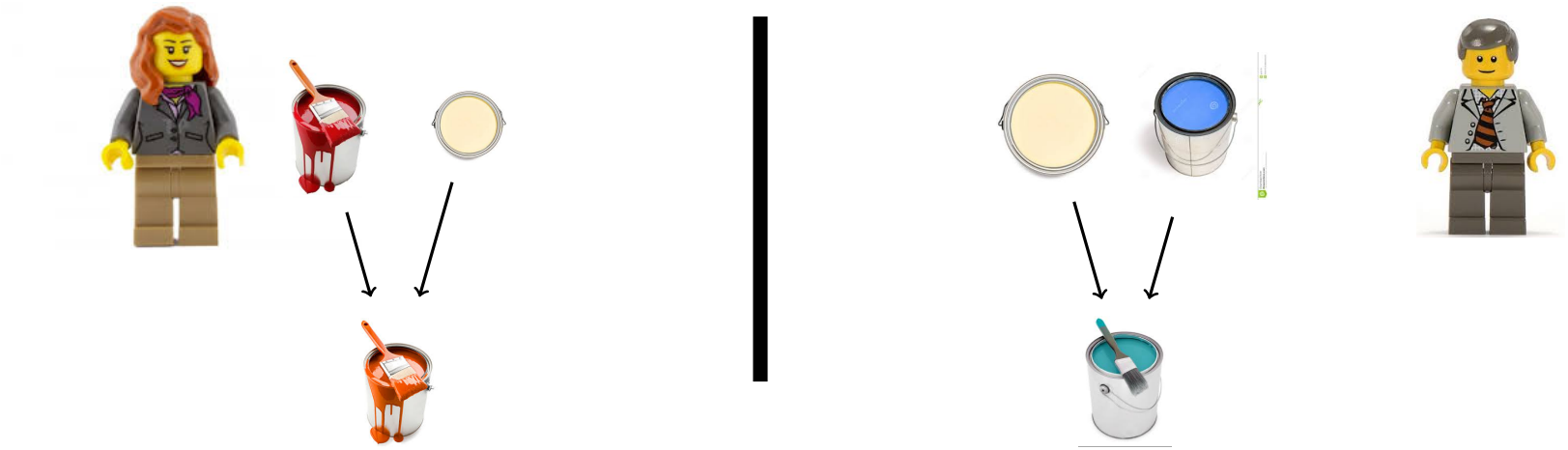
So Alice, Bob and Eve know the public color, pale yellow.

Everyone knows the public color is pale yellow

Step 3. Alice and Bob each get several paint pots of pale yellow paint. Behind their respective curtains they mix 1 pot of pale yellow with one pot of their private color. They repeat the process to get several public-private pots.

Alice now has paint pots that are half pale yellow and half crimson.

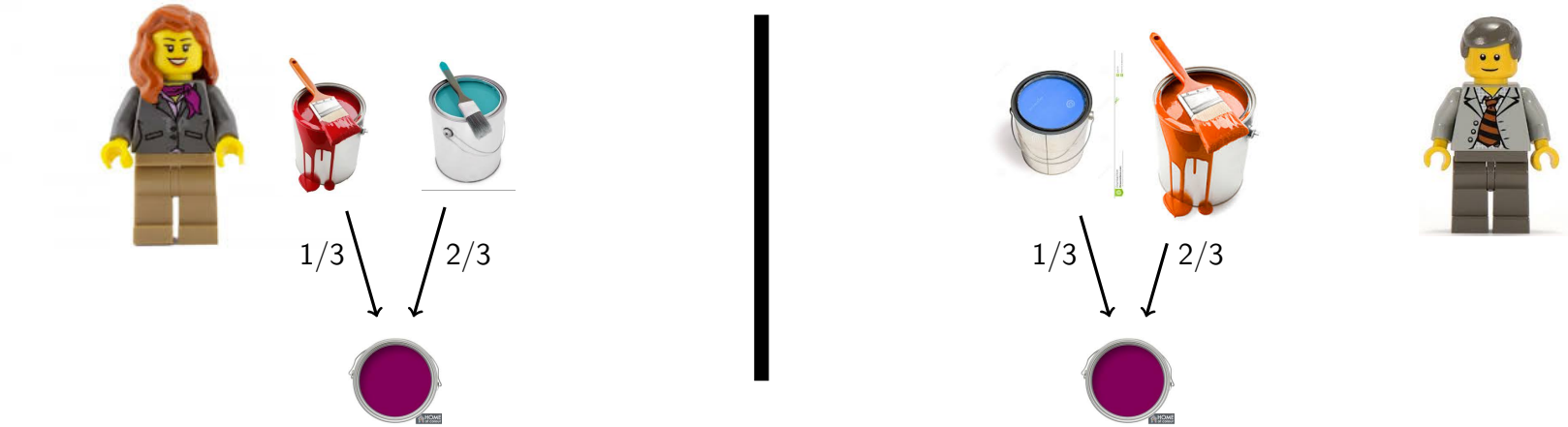
Bob now has paint pots that are half pale yellow and half blue.



Bob and Alice each mix a public-private color

Step 4. Alice and Bob place their public-private pots in the center of the room and take a pot of each other's back to their corners.

Step 5 Alice mixes two parts of Bob's public-private pot with one part of her private color. Bob mixes two parts of Alice's public-private pot with one part of his private color.



Bob and Alice each have a paint that is 1/3 pale yellow, 1/3 crimson and 1/3 blue

- Alice has a pot of paint that is 2/3 Bob's public-private color (equal parts blue and pale yellow $\Longrightarrow$ bluish green) and 1/3 of her private color (crimson) so her paint color is 1/3 pale yellow (public color) , 1/3 blue (Bob's private color) and 1/3 crimson (her private color).
- Bob has a pot of paint that is 2/3 Alice's public-private color (equal parts crimson and pale yellow $\Longrightarrow$ orangish) and 1/3 of his private color (blue) so his paint color is 1/3 pale yellow (public color) , 1/3 crimson (Alice's private color) and 1/3 blue (his private color).
- They both have the formula for a new paint color without broadcasting it.

Now we want to use the idea behind mixing paint and apply it to real-world applications.

In the real-world we use numbers instead of paint colors. Instead of mixing paints we perform operations on numbers.

Remember, when we mixed the paints together this process was irreversible so we need a mathematical process like this.

Clearly addition is reversible by using subtraction. That is, if we add 9 to a number, say 22 to get 31 then we can subtract 9 from 31 to get 22. This is because addition and subtraction are inverse operations.

Multiplication is reversible by using division. For example, if we multiply 12 by $3/4$ to get 9 then we divide 9 by $3/4$ (i.e., multiply by $4/3$) to get 12. This is because multiplication and division are inverse operations.

We now want to look at a type of arithmetic that does not have an inverse. Mathematicians call this modular arithmetic but to explain it we will call it clock arithmetic. You will see that you have been doing modular arithmetic most of your life without realizing it!

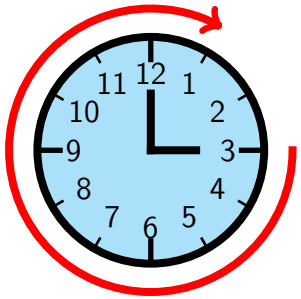
Clock Arithmetic (a.k.a. Modular Arithmetic)

Let's start with the typical clock with 12 numbers on it.

Suppose it is 3 o'clock and we want to know what time it will be in 10 hours. Clearly we don't say 13 o'clock. As we move around the clock (clockwise) adding 10 hours when we pass 12 we set the counter to 0. So instead of saying 13 o'clock we say 1 o'clock.

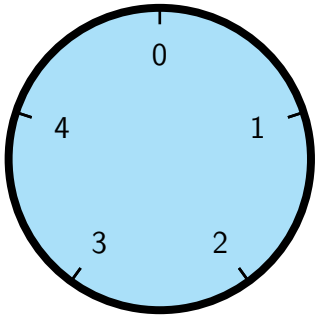
So using 12-hour clock arithmetic we say that $3 + 10 = 1$!

If there is a meeting that you know lasts 2 hours and it ends at 1 o'clock we know that it started at 11 o'clock so in 12-hour clock arithmetic we are saying $1 - 2 = 11$.



We use a 12 hour analog clock but we can do clock arithmetic in the same way even if we have a 5 hour or a 22 hour clock.

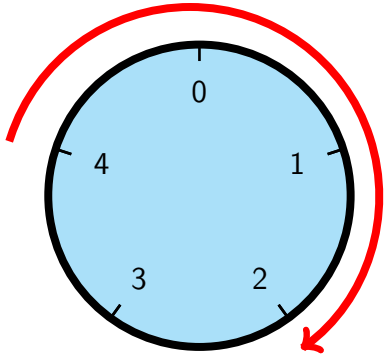
To make clock time correspond to modular arithmetic we adopt the convention that we use 0 instead of the last hour. For example, if we have a 5 hour clock the numbers on the dial are 0,1,2,3,4.



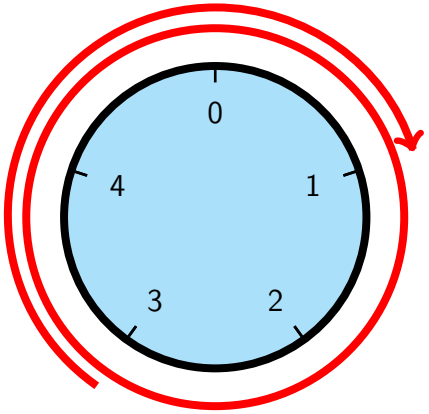
To do 5-hour clocktime arithmetic we simply mimic what we do on a 12 hour clock.

For example, if we want to add $4 + 3$ we get 7 normally but in 5-hour clocktime we get 2 by adding 3 hours to 4 and since when we add 1 hour we get 0, then two more hours gets us to 2.

$4 + 3 = 2$ in
5 hour clock arithmetic

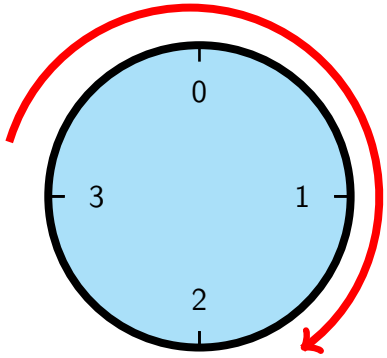


$3 + 8 = 1$ in
5 hour clock arithmetic

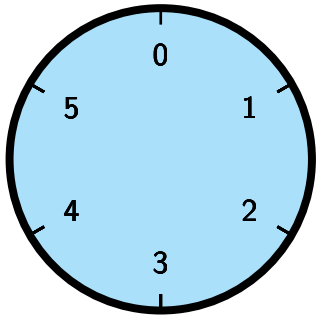


This approach works no matter how many hours we have on the clock

$3 + 3 = ?$ in
4 hour clock arithmetic



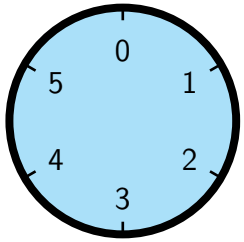
$3 + 8 = ?$ in
6 hour clock arithmetic



SOCRATIVE QUIZ - PartIII_Practice_Quiz5

IUZGAZ34E

Use this 6-hour clock to do the following clock arithmetic.



1. $5 + 2 = ?$
   (a) 1
   (b) 2
   (c) 3
   (d) 4
   (e) 5

2. $3 + 4 = ?$

   (a) 1
   (b) 2
   (c) 3
   (d) 4
   (e) 5

3. $5 + 5 = ?$

   (a) 1
   (b) 2
   (c) 3
   (d) 4
   (e) 5

4. $5 + 11 = ?$

   (a) 1
   (b) 2
   (c) 3

(d) 4

(e) 5

Relating Clock Arithmetic to Modular Arithmetic

Now let's see how this is related to modular arithmetic. When you divide two integers you get an integer plus a remainder. For example

$$\frac{13}{5} = 2 \quad \text{remainder 3}$$

$$\frac{26}{11} = 2 \quad \text{remainder 4}$$

In modular arithmetic we are only interested in the remainder. The notation we use is

$$13 \mod 5 = 2 \qquad 26 \mod 11 = 4$$

So the number before the modulus ($\mod$) is the numerator of the fraction and the number after it is the denominator and the answer is the remainder.

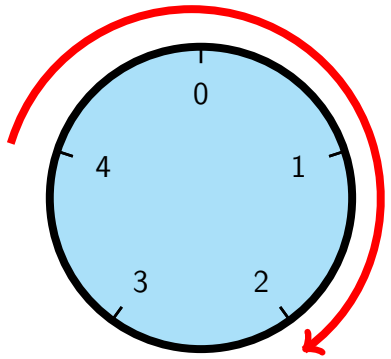
In the table below we give modular arithmetic using $\mod 5$. Note how the remainder goes from 1 to 4, then 0 and repeats.

$1 \mod 5 = 1$	$6 \mod 5 = 1$	$11 \mod 5 = 1$	$16 \mod 5 = 1$
$2 \mod 5 = 2$	$7 \mod 5 = 2$	$12 \mod 5 = 2$	$17 \mod 5 = 2$
$3 \mod 5 = 3$	$8 \mod 5 = 3$	$13 \mod 5 = 3$	$18 \mod 5 = 3$
$4 \mod 5 = 4$	$9 \mod 5 = 4$	$14 \mod 5 = 4$	$19 \mod 5 = 4$
$5 \mod 5 = 0$	$10 \mod 5 = 0$	$15 \mod 5 = 0$	$20 \mod 5 = 0$

Now if we return to our example of doing 5 hour clock arithmetic we saw that $4+3 = 2$ and notice that $7 \mod 5 = 2$. How do these relate?

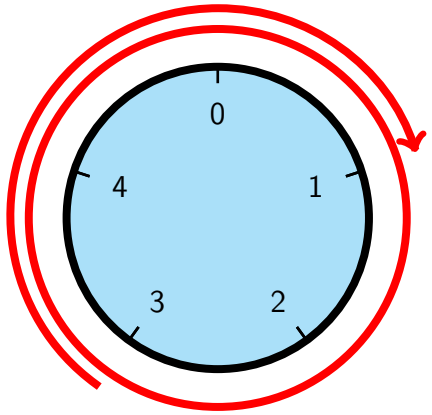
If we add 4 and 3 in the usual way we get 7 (the numerator in our fraction) and so since we are using a 5 hour clock this is our denominator and we write $7 \mod 5$.

$4 + 3 = 2$ in
5 hour clock arithmetic
$4 + 3 = 7, \frac{7}{5} = 1$ remainder 2
$7 \mod 5 = 2$



In our second example using a 5 hour clock we had $3+8=1$. If we add 3 and 8 we get 11 and $11 \mod 5 = 1$ because 5 goes into 11 twice with a remainder of 1.

$3+8=1$ in
5 hour clock arithmetic
$3+8=11, \frac{11}{5}=2$ remainder 1
$11 \mod 5 = 1$



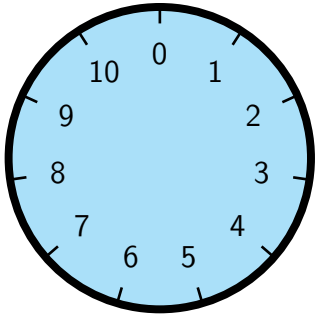
With modular arithmetic we can also do multiplication. For example

$$5 \times 4 = 20 \quad (5 \times 4) \mod 5 = 0 \qquad (5 \times 4) \mod 6 = 3,$$

$$2 \times 2 \times 2 \times 2 = 2^4 = 16 \qquad 2^4 \mod 3 = 1 \qquad 2^4 \mod 11 = 5$$

In the second one we have used power notation, i.e., exponentiation.

For our example using Public Key encryption we will use a clock with 11 hours because it's easy to multiply 11 by integers.



For the calculations below we need to know multiples of 11 so we make a table for reference. The first column is n the number we are multiplying 11 by and the second column is the result.

Multiplication of $n \times 11$

n	$11 \times n$	n	$11 \times n$	n	$11 \times n$	n	$11 \times n$	n	$11 \times n$
1	11	2	22	3	33	4	44	5	55
6	66	7	77	8	88	9	99	10	110
11	121	12	132	13	143	14	154	15	165
16	176	17	187	18	198	19	209	20	220
21	231	22	242	23	253	24	265	25	275

Now we can look at the table to immediately perform modular arithmetic where we use powers of integers. We have

$$213 \mod 11 = 4$$

To do this we look at the table and find that 209 and 220 sandwich 213. Since $11 \times 19 = 209$ and $11 \times 20 = 220$ we know that 19 is the largest integer which divides 213 and the remainder is 213 - 209 = 4.

Likewise

$$243 \mod 11 = 1$$

because $11 \times 22 = 242$ and $11 \times 23 = 253$ so the largest integer dividing 243 is 22 and the remainder is 243-242=1.

Example Use the table to determine each of the following.

1. 125 mod 11
2. 190 mod 11
3. 259 mod 11

SOCRATIVE QUIZ - PartIII_Practice_Quiz6

IUZGAZ34E

Use the table below to answer the following questions.

n	$11 \times n$	n	$11 \times n$	n	$11 \times n$	n	$11 \times n$	n	$11 \times n$
1	11	2	22	3	33	4	44	5	55
6	66	7	77	8	88	9	99	10	110
11	121	12	132	13	143	14	154	15	165
16	176	17	187	18	198	19	209	20	220
21	231	22	242	23	253	24	265	25	275

1. If $9^2 = 81$, what is $9^2 \mod 11$?

2. What is $256 \mod 11$?

3. Let n be any positive integer. What is the largest value that $2^n \mod 11$ can be?

(a) 275

(b) 15

(c) 11

(d) 10

We're almost ready to do the actual example. However, we will need results such as $5^4 \mod 11$, $9^7 \mod 11$ so instead of computing each of these separately, I'll precalculate some of them and put in a table. I have only included some terms that we will need in the examples. For homework you will be adding more; these are determined in exactly the same manner as the above examples.

n	1	2	3	4	5	6	7	8	9	10
2^n	2	4	8	5	10	9	7	3	6	1
3^n	3	9	5	4	1	3	9	5	4	1
4^n	4	5	9	3	1	4	5	9	3	1
5^n	5	3	4	9	1	5	3	4	9	1
6^n	6	3	7	9	10	5	8	4	2	1
9^n	9	4	3	5	1	9	4	3	5	1

Before we start the example, it's important to realize that even if we know that $3^n \mod 11 = 1$ we still can't figure out what n is. If we look at the second line of the table we see that n could be 5 or 10; continuing this pattern it could be 15, 20, 25, etc.

Public Key Encryption Example Using Modular Arithmetic

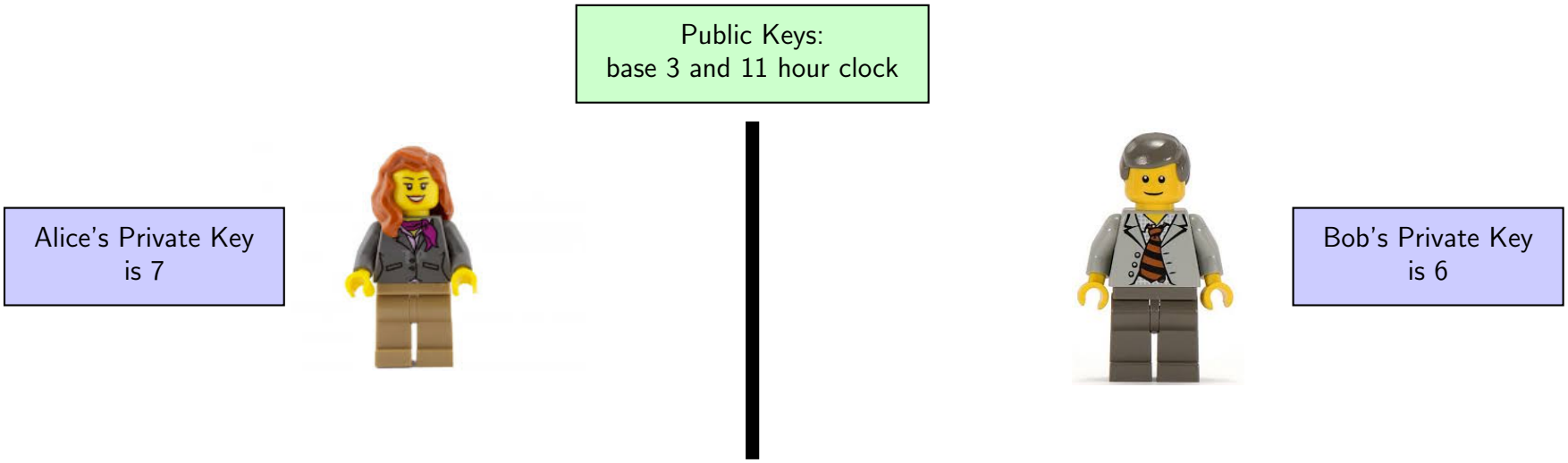
Goal: To create a shared secret in such a way that the secret can't be obtained by someone observing the communication.

Step 1. Bob and Alice each choose a Private Key.

Alice's Private Key is 7

Bob's Private Key is 6

Step 2. Bob and Alice agree on Public Keys. In particular they choose a base (so we can use our table it should be 2, 3 or 6) and the number of hours in the clock; again so we can use the table we choose 11.



Step 3 Alice and Bob each generate a number using their Private number, the public base and the 11 hour clock arithmetic. They use the following formula to get their Public-Private number (PPN)

$$\text{Alice:} \quad (\text{public base})^{\text{Alice's Private No.}} \mod 11$$

$$\text{Bob:} \quad (\text{public base})^{\text{Bob's Private No.}} \mod 11$$

Public Keys:
base 3 and 11 hour clock

Alice's Private Key
is 7

Alice's Public-Private number (PPN)
$3^7 \mod 11 = 9$

Bob's Private Key
is 6

Bob's Public-Private number (PPN)
$3^6 \mod 11 = 3$

Step 4 Bob and Alice exchange their Public-Private numbers. They then perform modular arithmetic to get the SAME number which is the shared secret. Alice and Bob use the following formulas

$$\text{Alice:} \quad (\text{Bob's PPN})^{\text{Alice's private number}} \mod 11$$

$$\text{Bob:} \quad (\text{Alice's PPN})^{\text{Bob's private number}} \mod 11$$

Public : base 3 and 11 hour clock
Alice's PPN =9
Bob's PPN=3

Alice's Private Key is 7
$3^7 \mod 11 = 9$

Bob's Private Key is 6
$9^6 \mod 11 = 9$

This works for any choice of numbers. Let's try a different set of Private numbers and base (but we will still use the 11 hour clock so we can use our table)

Step	Alice	Bob	Alice	Bob
Private #	8	2	4	9
Public base	5	5	2	2
PPN	$5^8 \mod 11 = 4$	$5^2 \mod 11 = 3$	$2^4 \mod 11 = 5$	$2^9 \mod 11 = 6$
Shared secret	$3^8 \mod 11 = 5$	$4^2 \mod 11 = 5$	$6^4 \mod 11 = 9$	$5^9 \mod 11 = 9$

Comments:

- Eve, an eavesdropper, knows the base, the number of hours in the clock and the two PPNs but she can't determine the secret number from these because she doesn't know Alice's or Bob's private number.
- In these examples we have chosen a small clock number. This is a problem because, e.g., when you are doing $\mod 11$ arithmetic then the possibilities for the shared secret are $0, 1, 2, \ldots, 10$ and so this is not very secure. In practice much larger numbers are used.

Public Key Encryption in Practice - Diffie-Hellman Algorithm

- The goal of the algorithm is the same as in the example – a shared secret is generated between two parties in such a way that it can't be seen by an observer. Even someone analyzing the communication at a later date can't figure out the secret.
- Suppose you sign on to a secure website (https) such as your electric company to pay your monthly bill.
- Your computer and the internet server create a shared secret in much the same way as in our example.

How is the algorithm different from our example?

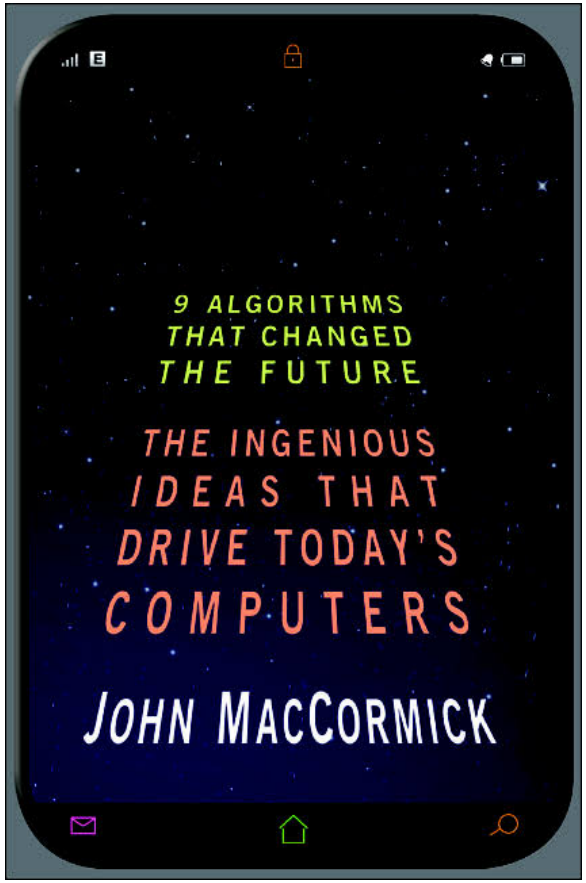
1. The clock size now is hundreds of digits in length so that the number of possibilities for private numbers is enormous.
2. It also turns out that the clock size should be a prime number. What is this? A prime number is only divisible by itself and 1. For example,

   $$1, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43$$

   are the first 15 prime numbers. No even number is prime except 2.
3. Another property of this algorithm is that the public base that is chosen must be related to the clock in a particular way. Let's go back to our table where we computed $b^n \mod 11$ for different bases b. If we compare the line in the table for 2^n and 3^n we notice that there is a difference. The line for 2^n contains ALL the numbers from 1 to 10 whereas the line for 3^n is missing 2, 6, 8, and 10. The second requirement of the algorithm is that the base is chosen so that all possible numbers on the clock (except 0) occur when we perform modular arithmetic with this base.

Reading Assignment for next time

Algorithm 9 - Digital Signatures: Who *Really* Wrote This Software



SOCRATIVE QUIZ - PartIII_Quiz5

IUZGAZ34E

1. $5 + 4 = ?$ in 7 hour clock arithmetic
   a. 1
   b. 2
   c. 3
   d. 4
2. $5^2 \mod 4 = ?$
   a. 1
   b. 2
   c. 3
   d. 4
3. One commonly used Public Key Encryption using a shared secret Key is

a. AES

b. DES

c. Blowfish

d. Diffie-Hellman

4. When using a Public Key Encryption algorithm with a Shared Secret Key then

a. the shared secret Key is transmitted to sender and receiver via a secure transmission

b. the sender and receiver both use public information and a mathematical computation to compute the Shared Secret Key

c. the sender uses public information and a mathematical computation to compute the Shared Secret Key and transmits to the receiver via a secure transmission

d. the receiver uses public information and a mathematical computation to compute the Shared Secret Key and transmits to the sender via a secure transmission

DIGITAL SIGNATURES

Goals for this lecture:

- To understand what a digital signature is and how the goal of it differs from Public Key Encryption.
- To see how digital signatures are used in practice.
- To see how modular arithmetic can be used in a manner similar to Public Key encryption to solve this problem

What are Digital Signatures used for?

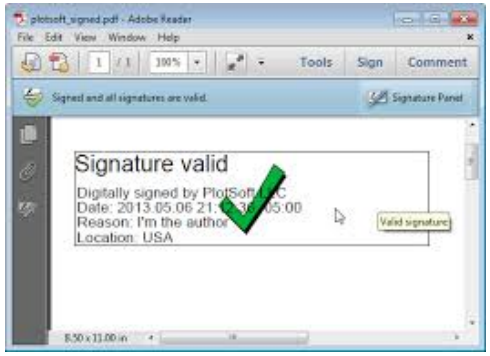
The phrase "digital signature" is a bit of an oxymoron. Clearly we use the term "signature" as something we can read but which can't be forged. However "digital" literally means "a string of digits" and we know that if we can read the string of digits then we can copy them.

The difference in digital signatures and other cryptography we have looked at is that the goal is NOT secrecy but authenticity. That is, we want to know that the digital signature is not forged.

The obvious use of digital signatures is for things that paper signatures are used for such as signing an apartment lease from long distance.

We will see that they have another important use too.

Sending hard copies (i.e., the physical item) of documents (such as contracts, etc.) back and forth between organizations takes a lot of time. A time saving device is to use digital signatures.



The receiver of the document needs to be sure that the signature was done by the sender and no one else, i.e., it needs to be verifiable. The receiver of the document needs to be sure that the signature was not forged.

Using a mobile ID you can sign documents or transfer money with the touch of your finger.

In the European Union a digital signature carries the same weight as a physical signature.

Studies have shown that using digital signatures can save up to a week's worth of time for each working age adult!

A study in Estonia (population 1.3 million) has shown that using digital signatures saves a stack of paper the height of the Eiffel Tower (984 ft) every month!

How are Paper Signatures Verified?


TITLE OF ACCOUNT

Southeast Bank Ltd.

_____ BRANCH _____ 200

I/We the undersigned, request you to open a Foreign Currency Account in US Dollar/Stg. Pound/Euro in my/our name (s). I/We agree to comply with and be bound by the rules and regulations of the Bank relative to this Account and any amendments thereto.
Declarations are on the reverse side.

FULL NAME	SIGNATURE
A/C HOLDER :	
NOMINEE :	
SPECIAL INSTRUCTION :	Introduced by
	Verified By / Approved by

Deb-07

Suppose you write a check for $1000 and your bank realizes that this is a large amount and wants to verify that you actually wrote it and that it is not forged. When you opened your account, the bank made you sign a signature card which they keep on file. The bank employee can physically go to the card and check your signature on the check against the one on file.

If someone has a document with your name signed on it and you claim that you didn't sign it, then prior copies of your signature must be compared with the one on the document to determine if it is a valid signature.

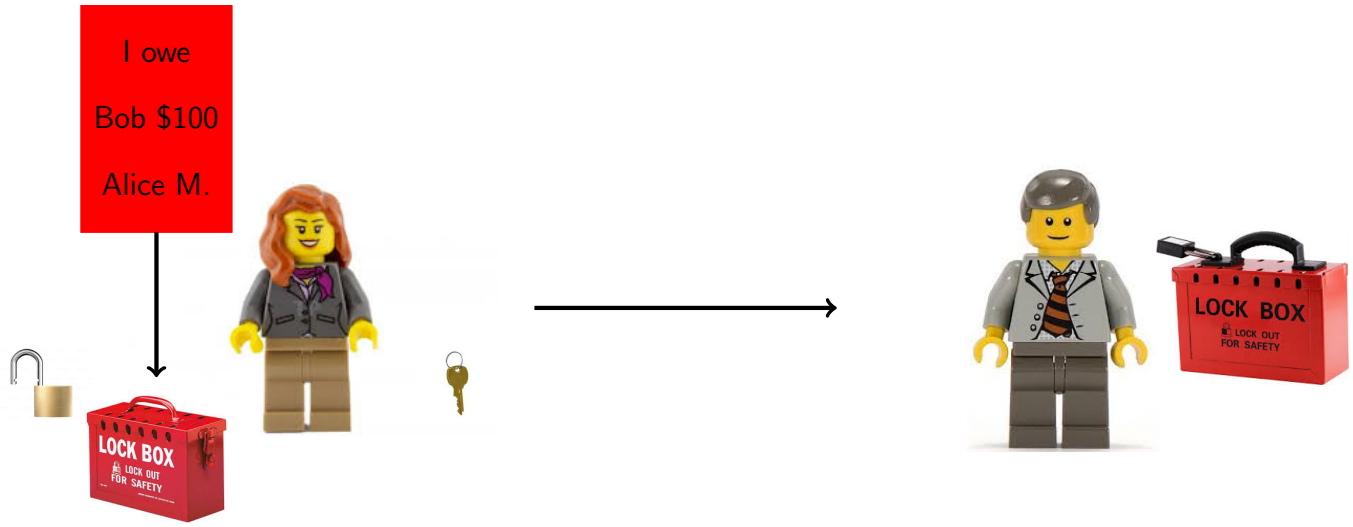
How can we mimic this procedure with a digital signature?

Lockbox Example

To understand how to verify a digital signature we first look at an example which uses physical objects instead of numbers and mathematics just like we did for Public Key Encryption.

Suppose Alice borrows $100 from Bob. Being fiscally responsible, Bob asks Alice for an IOU. However, Bob is always losing things so he doesn't want a slip of paper.

Alice buys a lockbox and only she has a key. She write the IOU and puts it inside the lockbox and keeps the key. She gives the lockbox to Bob which is big enough that he won't lose it.



Several months pass and Bob asks Alice if she can repay the $100. If Alice says "I don't remember borrowing money from you," Bob can produce the lockbox and ask Alice to give him the key. He then opens it and shows her the IOU.

The "signature" is verified since Alice is the only one with the key. Neither Bob nor anyone else could have opened the lockbox and forged the IOU.

What can go wrong?

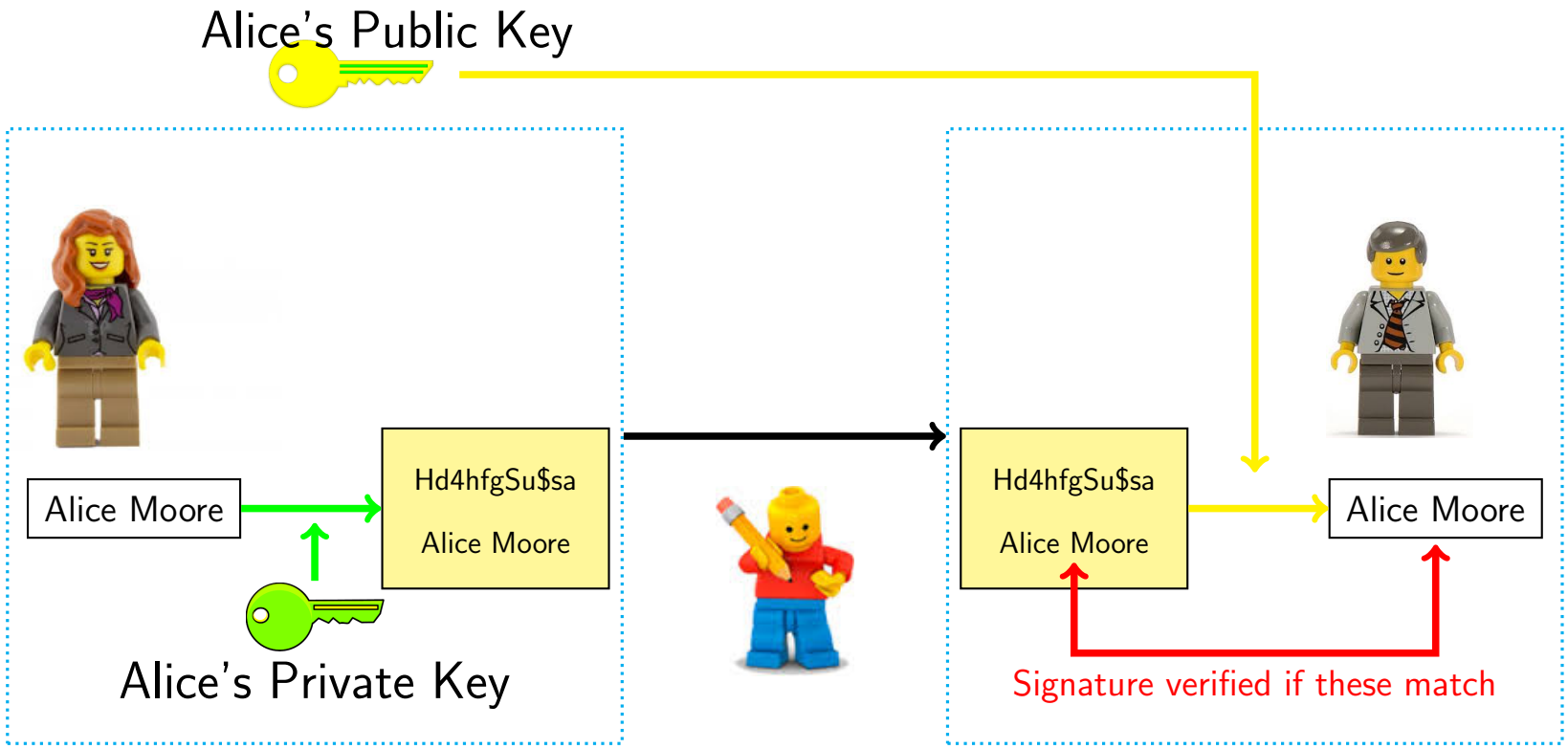
1. Alice could put any piece of paper (or nothing) into the lockbox instead of the IOU.
2. Alice could say that she lost the key.
3. Alice could give Bob the wrong key and when the lockbox wouldn't open she could respond "it isn't my lockbox."

To alleviate these problems, Bob could designate a trusted third party such as a bank. Alice would sign the IOU in the presence of the third party, give them the key for safe keeping and give Bob the lockbox. To verify check signatures the bank kept a signature card but now it will keep the physical key.

Now when Bob wants to see the IOU, he goes to the bank, gets the key and opens the lockbox in front of witnesses. The fact that the lockbox opens with Alice's key (which the bank is holding) authenticates the document inside.

To make this approach practical using computers we use numbers (hence the adjective "digital") in place of lockboxes and mathematical operations instead of physical keys just like we did in Public Key Encryption.

Suppose now Alice wants to sign a document and send it to Bob. It does not need to be sent securely but Bob needs to know that it was Alice that signed the document. How can we do this?



The procedure is almost the reverse of Public Key Encryption. Recall that in Public Key Encryption the goal was for Alice to send a message securely to Bob. Now Alice wants to send a document with a digital signature to Bob but she doesn't want to see it. The goal is that the receiver (Bob) can verify it, i.e., know that it was sent by Alice and not by someone else.

In practice, the lockboxes will be represented by numbers and the locking/unlocking procedures will use modular arithmetic (i.e., clock arithmetic) as we did in Public Key Encryption. Also we will use a clock with 11 hours because it is so easy to multiply numbers by 11.

In Public Key encryption we used addition when using modular arithmetic but here we use multiplication; it works the same way though.

Let's review how to do this before we proceed. Remember that when using modular arithmetic all we are concerned about is the remainder.

Since we are using an 11 hour clock all of our expression will be of the form $m \times n \mod 11$ where m, n are integers. For example,

$$4 \times 3 \mod 11 \quad \rightarrow \quad 12 \mod 11 = 1$$

because when we divide 12 by 11 we get 1 with remainder $= 1$. Also

$$9 \times 3 \mod 11 \quad \rightarrow \quad 27 \mod 11 = 5$$

since $27/11 = 2$ remainder 5 and

$$5 \times 6 \mod 11 \quad \rightarrow \quad 30 \mod 11 = 8$$

since $30/11 = 2$ remainder 8.

Socrative PartIII Practice Quiz7

IUZGAZ34E

1. What is (4 $\times$ 5) mod 11 ?
2. What is (3$\times$ 6) mod 7 ?
3. What is (2^4) mod 11 ?

When you use software like ADOBE to digitally sign a document you simply type your name. Since computers prefer numbers, the first thing that happens is that your name is changed into a string of numbers like 3716590087214457 8. Now remember that computers use binary numbers but for our purposes we won't complicate the discussion by using binary numbers.

To make things easy we start by assuming Alice's name can be transformed using only a single digit. Once we understand the approach we can see how to do the case when it is represented with multiple digits.

Assume Alice's signature is represented by the digit 3.

Alice must have a Private Key to encrypt her digital signature of 3. Assume Alice chooses 6.

Alice selects a Public Key which in our case is the clock size of 11 and a number less than 11; here we choose 2.

Alice uses multiplicative modular arithmetic to encrypt her digital signature of 6

$$3 \times 6 = 18 \quad \Longrightarrow \quad 18 \mod 11 = 7$$

So Alice's digital signature is encrypted as 7. This is sent to Bob but anyone can see it. In addition Alice sends her digital signature of 3.

How does Bob unlock the signature and know that only Alice has sent it?

Because he unlocks the encrypted signature with her Public Key. How does he do this? He takes the encrypted signature (7) and uses the Public Key of 2 to perform multiplicative modular arithmetic

$$7 \times 2 = 14 \quad \Longrightarrow \; 14 \mod 11 = 3$$

Now he compares this with the original signature 3 which Alice has also sent in its original form (3). If they match, then he knows that the encrypted document that Alice sent is not forged.

This verification will work no matter what Alice's original digital signature is, as illustrated in the table below.

Digital Signature	Private Key	Encryption	Public Key	Decryption
1	6	$1 \times 6 \mod 11 = 6$	2	$6 \times 2 \mod 11 = 1$
3	6	$3 \times 6 \mod 18 = 7$	2	$7 \times 2 \mod 14 = 3$
4	6	$4 \times 6 \mod 11 = 2$	2	$2 \times 2 \mod 11 = 4$
5	6	$5 \times 6 \mod 11 = 8$	2	$8 \times 2 \mod 11 = 5$
9	6	$9 \times 6 \mod 11 = 10$	2	$10 \times 2 \mod 11 = 9$

Of course the Private Key and the Public Key have to be related. For example, Alice's digital signature was represented by 3 (which everyone knows) but if someone intercepts this and encrypts it with 5 instead of 6 as Alice did then Bob will find that the signature is a forgery because

$$3 \times 5 = 15 \implies 15 \mod 11 = 4$$

is the encrypted signature Bob receives and decrypting it with Alice's Public Key (2) gives

$$4 \times 2 = 8 \implies 8 \mod 11 = 8$$

and $3 \neq 8$.

How did we know that 2 was the Key which decrypted Alice's encryption Key of 6?

It turns out that the basis is an algorithm that was written over 2000 years ago!

Euclid's Algorithm

Euclid's Algorithm (Greek mathematician around 300 BC) is a clever way of computing the greatest common divisor (GCD or gcd) of two integers; i.e., finding the largest integer which divides both numbers evenly (without a remainder). This is probably the oldest algorithm still in use.

The original importance of this idea was probably for measurements in geometry. For example, suppose you have two rulers of length 18" and 24" and you wanted to find the length of a third ruler which is as long as possible and that you can use it as a scale for the other two rulers. For example, you could have rulers of length 1", 2", 3", and 6" so the longest is 6" for rulers of length 18 and 24. This is equivalent to finding the gcd(18,24).

As another example, suppose we have the integers 16 and 24 and we want to find the largest integer which divides both evenly. Clearly the gcd has to be ≤ 16, the smaller of the two integers. Determining the gcd in this case is easy because we know that $8 \times 2 = 16$, $8 \times 3 = 24$ and no integer larger than 8 divides them both so the

gcd(16,24) is 8.

Before we look at Euclid's algorithm we see what a Brute Force algorithm for finding the gcd of two numbers would look like.

To determine gcd(16,24), we first note that the largest the gcd can be is 16 so the first guess is 16 and if that doesn't work we try 15, then 14, etc. When we find the first integer that divides both, then this is the answer because we want the largest such integer. To check to see if an integer divides 16 or 24 evenly we can perform modular arithmetic because all we have to do is check to see if the remainder is zero.

Why don't we start at 1 and increase the guess by 1 each time?

Brute Force Algorithm for finding gcd(16,24)

Step 1. Set gcd $= 16$

Step 2. Determine remainder $R_1 = 16 \mod$ gcd.

Step 3. If remainder $R_1 \neq 0$ then set gcd $=$ gcd - 1 and go to Step 2

If remainder $R_1 = 0$ then determine new remainder $R_2 = 24 \mod$ gcd and continue.

Step 4. If $R_2 = 0$, then stop.

Else if $R_2 \neq 0$ then gcd $=$ gcd -1 and go to Step 2.

Loop	Current guess for gcd	R_1	Action	R_2	Action
1	16	$16 \mod 16 = 0$	calculate R_2	$24 \mod 16 = 8$	go to Step 2
2	15	$16 \mod 15 = 1$	reduce gcd & go to Step 2		
3	14	$16 \mod 14 = 2$	reduce gcd & go to Step 2		
4	13	$16 \mod 13 = 3$	reduce gcd & go to Step 2		
5	12	$16 \mod 12 = 4$	reduce gcd & go to Step 2		
6	11	$16 \mod 11 = 5$	reduce gcd & go to Step 2		
7	10	$16 \mod 10 = 6$	reduce gcd & go to Step 2		
8	9	$16 \mod 9 = 7$	reduce gcd & go to Step 2		
9	8	$16 \mod 8 = 0$	go to Step 3	$24 \mod 8 = 0$	Found GCD

To generalize this algorithm for any two integers we have the following.

Brute Force Algorithm for finding $\gcd(a, b)$ where $a < b$

Step 1. Set gcd $= a$

Step 2. Determine remainder $R_1 = a \mod$ gcd.

Step 3. If remainder $R_1 \neq 0$ then set gcd $=$ gcd -1 and go to Step 2.

If remainder $R_1 = 0$ then determine new remainder $R_2 = 24 \mod$ gcd and continue.

Step 4. If $R_2 = 0$, then stop.

Else if $R_2 \neq 0$ then gcd $=$ gcd -1 and go to Step 2.

Now it only took 9 steps using the Brute Force approach to find gcd(16,24) but what if our numbers were much larger such as finding the gcd(105,252)? It turns out that the gcd is 21 so we have to do 105-21=84 recursive loops to get the result. Note that 105=21×5 and 252=21×12.

How can we characterize the number of loops it takes to do the Brute Force Algorithm for finding gcd(a,b)?

It turns out that 17= gcd(6409,42823) so we have to do 6409-17=6,392 recursive loops. But if we wanted to find gcd(6409,12818) we see by observation that we get the answer on the first step of the Brute Force algorithm because gcd(6409,12818)=6409. So all we can say about the number of recursive loops that it takes for the Brute Force approach is that it is $\leq$ the smaller of the two numbers.

Now let's see how Euclid's algorithm works. First we look at a simplified version and then see how we can improve on it.

The basis for the algorithm is that if we have an integer n that divides both a and b then it also divides $a-b$. For example, 5 = gcd(15,100) but 5 also divides 100-15=85 and 5 = gcd(15,85) so instead of finding gcd(15,100) we could find gcd(15,85). First we do two examples using this simplified version of Euclid's algorithm based upon this observation and then write the pseudo-code.

Example Find gcd(16,24) using the Simplified Euclid's algorithm. Compare with the Brute Force approach.

We first note that gcd(16,24) = gcd(16,24-16)=gcd(16,8). We note that 8 divides 16 and itself so it must be the gcd(16,24). What we are really doing here is dividing 24 by 16 and getting the remainder; in this case 8. We can use modular arithmetic for this and we have

$$24 \mod 16 = 8$$

so now gcd(16,24) = gcd(16,8) where 8 is the remainder. Doing this again we have

$$16 \mod 8 = 0$$

and we are done. Recall that the Brute Force approach took 9 loops to find gcd(16,24) and we found it here in two steps.

Example Find gcd(11,24) using the Simplified Euclid's algorithm. How many steps does it take?

On the first step we note that

$$24 \mod 11 = 13$$

so

$$\gcd(11, 24) = \gcd(11, 13).$$

Now on the next step we have

$$13 \mod 11 = 2$$

so that

$$\gcd(11, 13) = \gcd(11, 2).$$

Continuing we have $11 \mod 2 = 1$

$$\gcd(11, 2) = \gcd(2, 1)$$

and then

$$2 \mod 1 = 0$$

so we have

$$\gcd(11, 24) = \gcd(1, 1)$$

and the gcd(11,24) =1. We knew this had to be the answer because 11 is prime, which means that the only divisors it has is 1 and itself.

Simplified Euclid's Algorithm for finding $\gcd(a, b)$

Step 1. If $a > b$, exchange a and b

Step 2. Calculate remainder R = b mod a.

Step 3. If R=0, then a is the gcd(a,b) and stop.

If R $\neq 0$ set b=a and a = R; go to Step 2.

Note that by setting the remainder R to be the number we are dividing by in modular arithmetic we are guaranteed that the remainder is smaller than the number we divided by in the previous step.

We can improve on this algorithm. Recall that we said that if 5 divides 15 and 100 then it also divides 100-15=85 so that gcd(15,100) = gcd(15,85). However it is also true that 5 divides 100- (2×15)=70 so we could find gcd(15,70). Continuing in this manner we see that 100- (3×15)=55 so we could find gcd(15,55) and 100- (4×15)=40 so we could find gcd(15,40); 100- (5×15)=25 so we could find gcd(15,25) and finally 100- (6×15)=10 so we could find gcd(10,15). It will be quicker to find the gcd of small integers so the first step of Euclid's algorithm is to reduce the problem of finding

gcd(15,100) to finding gcd (10,15). The way we do this

$$\frac{100}{15} = 6 \quad \text{remainder=10} \implies 100 = 6(15) + \frac{10}{15}$$

and we know that the remainder can be found by using modular arithmetic

$$10 = 100 \mod 15$$

.

Now what do we do? We want to find the gcd(10,15) so we write

$$\frac{15}{10} = 1 \quad \text{remainder=5} \implies 15 = 10 + \frac{5}{15} \implies 5 = 15 \mod 10$$

so we have reduced our problem to determining gcd(5,10) and so we write

$$\frac{10}{5} = 2 \quad \text{remainder 0} \implies 10 = 2 \times 5 \implies 0 = 10 \mod 5$$

and the answer is 5, the last number we divided by. We know we have the answer because our remainder is 0.

Let's do one more example before we write out the algorithm.

Example Find the gcd(105,252) using Euclid's algorithm.

We divide 252 by 105 and find the remainder

$$\frac{252}{105} = 2 \quad \text{remainder} = 42 \implies 42 = 252 \mod 105$$

which gives us that gcd(105,252)=gcd(105,42). Now we have

$$\frac{105}{42} = 2 \quad \text{remainder} = 21 \implies 21 = 105 \mod 42$$

and we have reduced the problem to finding gcd(21,42). We have

$$\frac{42}{21} = 2 \quad \text{remainder} = 0 \implies 0 = 42 \mod 21$$

and we are done because the remainder is zero. Note that in each case the remainder has to be less than the number you are dividing by. Thus gcd (105,252)=21 and this was found by executing the following 3 steps.

1. calculate 252 mod 105 (which is 42)
2. calculate 105 mod 42 (which is 21)

3. calculate 42 mod 21 (which is 0)

So what would the steps be if we wanted to find gcd(a,b) where a$<$ b

1. calculate c$=$ b mod a
2. if c $\neq$ 0 calculate d $=$ a mod c
3. if d $\neq$ 0 calculate e $=$ c mod d
4. if e $\neq$ 0 calculate f $=$ d mod e
5. etc.

Here is the pseudo-code for Euclid's algorithm.

Euclid's Algorithm for finding gcd(a,b) where a$<$b

Step 1. Calculate c $=$ b mod a

Step 2. If c $=$ 0 , set gcd$=$a and stop.

If c $\neq 0$ set b$=$a and a $=$ c; go to Step 1.

Let' s make a table like we did for the Brute Force algorithm for determining gcd(105,252).

Loop	c	Action at Step 2
1	42=252 mod 105	set a=42, b=105 & go to Step 1
2	21 = 105 mod 42	set a=21, b=42 & go to Step 1
3	0 = 42 mod 21	done; set gcd = 21

Now let's look at Euclid's algorithm for determining gcd(6409,42823) which took over 6000 loops for the Brute Force approach.

Loop	c	Justification for c	Action at Step 2
1	4369=42823 mod 6409	42823/6409 = 6 + 4369/6409	set a=4369, b=6409 & go to Step 1
2	2040 = 6409 mod 4369	6409/4369=1 + 2040/4369	set a=2040, b=4369 & go to Step 1
3	289 = 4369 mod 2040	4369/2040=2 + 289/2040	set a=289, b=2040 & go to Step 1
4	17 = 2040 mod 289	2040/289= 7+17/289	set a=17, b=289 & go to Step 1
5	0 = 289 mod 17	289/17 = 17	stop, set gcd = 17

This took only 5 steps compared with more than 6000 for the Brute Force approach!

Socrative PartIII_Practice_Quiz8

IUZGAZ34E

1. The gcd(12,40) is
   (a) 12
   (b) 6
   (c) 4
   (d) 3
2. To determine the gcd(21,49) by the Brute Force method you start at 2 and increase the number until you find the first integer that divides both 21 and 49 evenly and that is your greatest common divisor.
3. In the first recursive loop of the SIMPLIFIED Euclid's algorithm for determining the gcd (21,49) one rewrites this problem in terms of
   (a) gcd(21,70)

(b) gcd(21,28)

(c) gcd(21,21)

(d) gcd(21,7)

4. In the first recursive loop of Euclid's algorithm for determining the gcd (21,49) one rewrites this problem in terms of

   (a) gcd(21,70)

   (b) gcd(21,28)

   (c) gcd(21,21)

   (d) gcd(21,7)

5. Euclid's algorithm for determining gcd(a,b) terminates when b mod a $= 0$ when a $<$ b.

Now that we have seen Euclid's algorithm for determining the greatest common divisor of two integers, we want to see how this helps us to find the relationship that the Public and Private Keys must have.

Recall that in our example, Alice chose a Private Key of 6 and we said that this means the Public Key (for an 11 hour clock) had to be 2. You can actually swap these around and have the Public Key of 6 and the Private Key of 2 and it still works because they are related in the same way.

Now what we want is to find the Public Key for an 11 hour clock that reverses the result of her Private Key of 6. What we want to find is the inverse operation; e.g., if we add 6 to a number x then to get x back if we subtract 6 from the result.

Specifically we want to know the following. If her digital signature is represented by 3 and it is encrypted using the Private Key of 6 then we get the encrypted signature 7 from the expression

$$(3 \times 6) \mod 11 = 7 \implies 7 = 18 - (11 \times 1)\,.$$

We want to know n such that

$$(\mathrm{n} \times 7) \bmod 11 = 3$$

so we are really asking what is the inverse of 18 mod 11, i.e., what operation reverses this process so that you get 3 back. This is equivalent to finding an n such that

$$(n \times 7) - 11\,k = 3$$

for some integer value of $k > 0$ where

$$(n+ 7)/11 = k + 3/11\,.$$

So if we can write 3 as a combination of a multiple of 7 minus a multiple of 11 we are done. By inspection n=1 doesn't work because we have $7 - 11k \neq 3$ for any k. We can try n=2 to get $14 - 11k = 3$ and this holds for k=1. So for this easy case we were able to find the result by inspect.

Now let's see how Euclid's algorithm can be used to find the answer of n=2, k=2. We use Euclid's algorithm to find gcd(11,18).

$$\begin{aligned} 7 = 18 \mod 11 &\implies 18 - (11 \times 1) = 7 \\ 4 = 11 \mod 7 &\implies 11 - (7 \times 1) = 4 \\ 3 = 7 \mod 4 &\implies 7 - (4 \times 1) = 3 \\ 1 = 4 \mod 3 & \end{aligned}$$

Now we write 3 as a combination of 11 and 7 to get

$$3 = 7 - 4 = 7 - (11 - 7) = 2 \times 7 - 11$$

so we have found n=2, k=1.

Just like Public Key Encryption, in practice the clock size is very large and it is a prime number. Any positive number less than the clock size can be used to encrypt the digital signature. Then a computer program which implements Euclid's algorithm is used to determine the Public Key which decodes the encrypted message to get the original digital signature.

Unfortunately, the multiplicative modular arithmetic approach is flawed. Remember that when we chose Alice's Private Key of 6 and the Public Key of 2 we said that these could be reversed; i.e., choose a Private Key of 2 and a Public Key of 6. This means that in the first case we can use Euclid's algorithm to get the Public Key of 2 but we can also use it to get the Private Key because it works in reverse.

If we can't use the multiplicative modular arithmetic approach in practice, then what can we use?

We use exponentiation in place of multiplication. Remember we have used exponentiation such as

$$2^4 = 2 \times 2 \times 2 \times 2 = 16$$

or

$$5^3 = 5 \times 5 \times 5 = 125$$

and we call this "raising a number to a power." Of course it is just multiplication but in a much more complicated way.

Terminology:

$$2^4 \quad \Longrightarrow \quad 2 \text{ is called the base and 4 is called the exponent}$$

For reasons you will see in a minute we will use a clock size of 22 and we will be raising numbers to either the third or seventh powers only.

Suppose we want to calculate $2^5 \mod 22$. We proceed as before

$$32 = 2^5 \implies 2^5 \mod 22 = 10$$

since $1 \times 22 + 10 = 32$.

As before the clock size (here 22) is made public. Assume that Alice's digital signature is represented as 4. This is the base in the exponentiation operation. For the Private Key Alice chooses the exponent, for example 3. Now the digital signature is encrypted as 20 because

$$4^3 \mod 22 = 64 \mod 22 = 20$$

To decrypt the message we use the power of 7 in our exponentiation. We have

$$20^7 \mod 22 = 1,280,000,000 \mod 22 = 4$$

which is the original message. The reason we used 3 and 7 as our powers is that 7 is known to reverse the process of exponentiation with a power of 3 using mod 22.

Example. Encrypt and decrypt a digital signature of 5 using a 22 hour clock and encrypting with a power of 3 and decrypting with a power of 7.

To encrypt we have

$$5^3 \mod 22 = 125 \mod 22 = 15$$

since $22 \times 5 = 110$ so the encrypted value is 15.

To decrypt we use as the base the encrypted value of 15 and 7 as the exponent. We have

$$15^7 \mod 22 = 170,859,375 \mod 22 = 5$$

which is the same as the original signature. Here $170,859,375 = 7,766,335 \times 22 + 5$.

Why is exponentiation better than multiplication?

As before, the clock size is public and the Public Key which decrypts the signature is known to everyone. However, unlike modular arithmetic using multplication, knowing these two things are not enough to obtain the Private Key which was chosen to encrypt the digital signature. The user can use an algorithm to determine the Private Key once the Public Key is chosen but reversing the operation without knowledge of the Private Key is very secure at present.

This approach of using exponentiation instead of multiplication is a well known algorithm called the RSA digitial signature algorithm, named after its inventors.

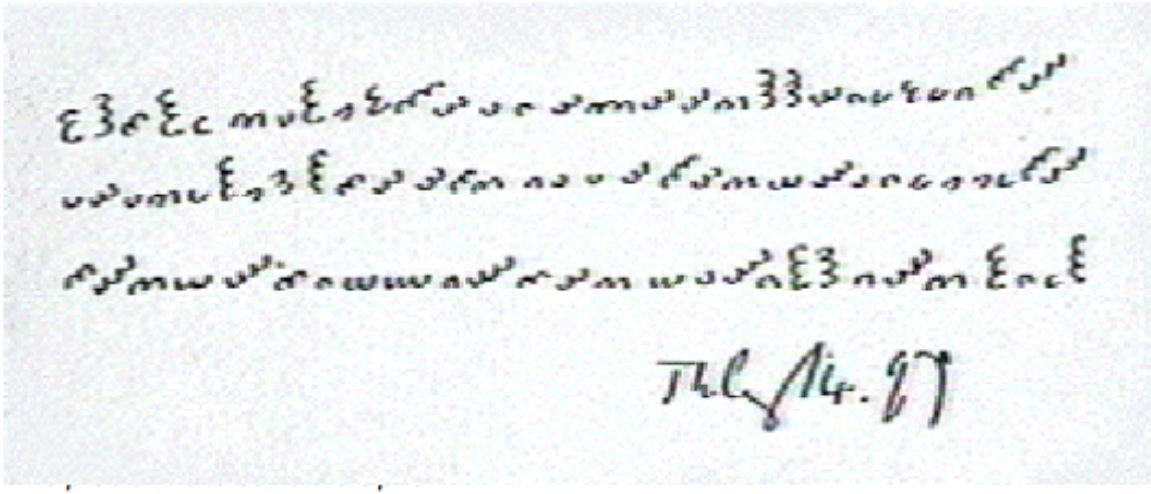
Most of us have very few occasions to use digital signatures in place of printed signatures. However, we use digital signatures many times without realizing it.

For example, suppose you go to the Web and download a new piece of software such as a new version of Adobe FlashPlayer. Most of the time the software is "signed" so your computer has to "unlock" the signature using the signer's Public Key and compares it with the signature that appears on the software. If it doesn't match, you get a message such as the one below. So your computer is using digital signatures a lot!

Unsolved Ciphers

Dorabella Cipher

The Dorabella cipher is contained in a note written by the famous composer Sir Edward Elgar to a young lady companion, Dora Penny, in 1897. The cipher remains unsolved.



Zodiac 340 Cipher

The Zodiac killer was responsible for several murders in the San Francisco area in the late 1960s. He sent several letters and ciphers to newspapers during this time. His first cipher was solved but his second, known as the 340 cipher, remains unsolved.

It is important to realize that encryption methods have to change as the hardware and software capabilities improve. Thus there are always new algorithms for encrypting and decrypting.

We have to stay ahead of the hackers!

Socrative PartIII_Quiz6

IUZGAZ34E

1. Unlike Public Key Encryption, the goal of a Digital Signature algorithm is authenticity not secrecy.
2. The idea of digital signatures is often used to verify the author of software that we download.
3. What is 2^4 mod 11?
   (a) 0
   (b) 2
   (c) 5
   (d) 9
4. Encrypting a digital signature using multiplicative modular arithmetic is a very secure strategy.

5. Euclid's algorithm terminates when the remainder is zero.
6. In the first recursive loop of the SIMPLIFIED Euclid's algorithm for determining the gcd(14,49) one reduces the problem in terms of
   (a) gcd(7,14)
   (b) gcd(14,35)
   (c) gcd(14,42)
   (d) gcd(42,49)
7. In the first recursive loop of Euclid's algorithm for determining the gcd(14,49) one reduces the problem in terms of
   (a) gcd(7,14)
   (b) gcd(14,35)
   (c) gcd(14,42)
   (d) gcd(42,49)
8. Euclid's algorithm can be used to find the Public Key once a Private Key is chosen when exponential modular arithmetic is used.
9. The gcd(120,45) is 5. About how many steps would you have to do to find this with a Brute Force approach?

(a) 10

(b) 20

(c) 30

(d) 40

10. The gcd(120,45) is 5. About how many steps would you have to do to find this with Euclid's Algorithm?

(a) < 5

(b) more than 5 but < 10

(c) more than 10 but < 20

(d) more than 20