Cryptanalysis
Concept and Practice

PURDUE
POLYTECHNIC

Group Activity:
What did the Japanese and Soviets do wrong?

Form 8 Groups:

1. Bring your individual answers to this question into your groups.
2. Discuss what you believe the Japanese and Soviets could have done differently to reduce the probability that their cryptosystems would have been broken.
3. Form one, combined written list based on your discussions

We will add your group's contribution to the whiteboard

Key Concept #1: Perfect Secrecy

Cryptosystems that exhibit *"Perfect Secrecy"*

- In fancy math terms: The conditional probability of the plaintext message after the ciphertext is known is the same as the conditional probability of the plaintext message before the ciphertext is known.

$$\Pr[C = c|M = m] = \Pr[C = c]$$

- More Simply: The ciphertext reveals nothing about the plaintext message other than its length.
- To prove this, we would show that, for every ciphertext and plaintext of the same length, there is a key which enciphers the given plaintext into the given ciphertext.

Key Concept #2: Kerckhoffs's Principle

Cryptosystems that follow "Kerchoffs's Principle"

- Have no secret components except for the key.
- This concept is also often attributed to Claude Shannon, who phrased it as, "The enemy knows the system."
- You are also familiar with its opposite axiom: "Security Through Obscurity".

Classical Ciphers

Classical Ciphers are of two forms: substitution and transposition.

- Substitution ciphers substitute letters of the plaintext message with other letters to produce the ciphertext message.
  - examples: Caesar, Vigenere
- Transposition ciphers move the plaintext letters around in a pattern to produce the ciphertext message.
  - examples: writing the message backward.

Modern ciphers use other methods to encrypt messages, but those methods sometimes combine multiple substitutions and transpositions as part of the algorithm.

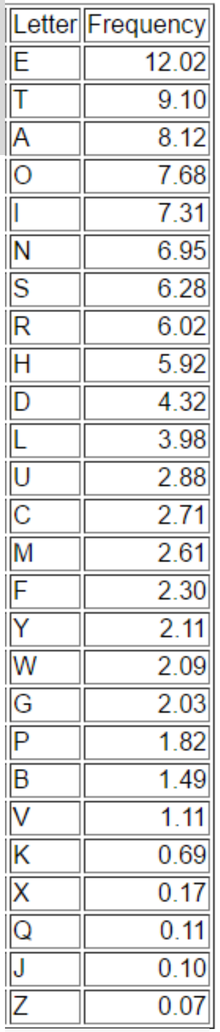
  - example: DES

Breaking and Distinguishing Classical Ciphers

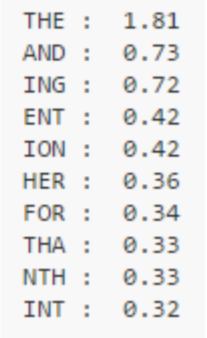
Simple Substitution Ciphers can be broken using frequency analysis and trial and error.

In frequency analysis, the cryptanalyst compares the frequency of occurrences of the letters in the ciphertext and compares them to the frequency of use in the native language of the text (in this case, English).

Frequency analysis is also used to distinguish between substitution ciphers from transposition.



Letter	Frequency
E	12.02
T	9.10
A	8.12
O	7.68
I	7.31
N	6.95
S	6.28
R	6.02
H	5.92
D	4.32
L	3.98
U	2.88
C	2.71
M	2.61
F	2.30
Y	2.11
W	2.09
G	2.03
P	1.82
B	1.49
V	1.11
K	0.69
X	0.17
Q	0.11
J	0.10
Z	0.07



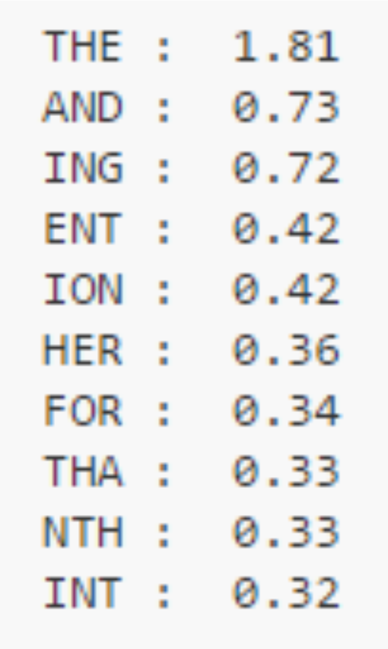
```
THE :  1.81
AND :  0.73
ING :  0.72
ENT :  0.42
ION :  0.42
HER :  0.36
FOR :  0.34
THA :  0.33
NTH :  0.33
INT :  0.32
```

…by percentage

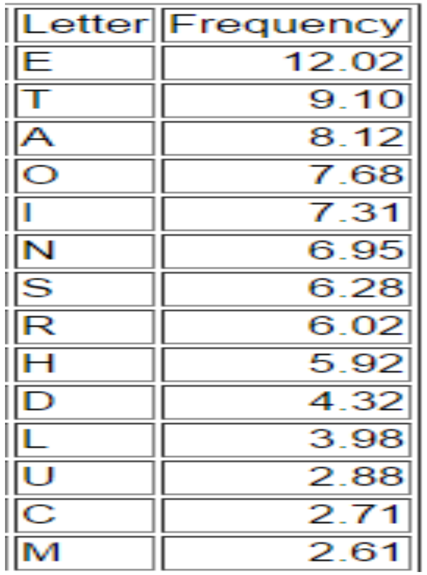
Activity

Attempt to decipher the following ciphertext created using a simple cipher. You have 5 minutes.

```
AOLWYLZPKLUAVMAOLBUPALKZAHALZ
THE  EA  E T  THE   TE     E
```



THE	:	1.81
AND	:	0.73
ING	:	0.72
ENT	:	0.42
ION	:	0.42
HER	:	0.36
FOR	:	0.34
THA	:	0.33
NTH	:	0.33
INT	:	0.32



Letter	Frequency
E	12.02
T	9.10
A	8.12
O	7.68
I	7.31
N	6.95
S	6.28
R	6.02
H	5.92
D	4.32
L	3.98
U	2.88
C	2.71
M	2.61

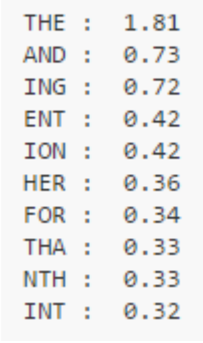
…by percentage

Breaking and Distinguishing Classical Ciphers

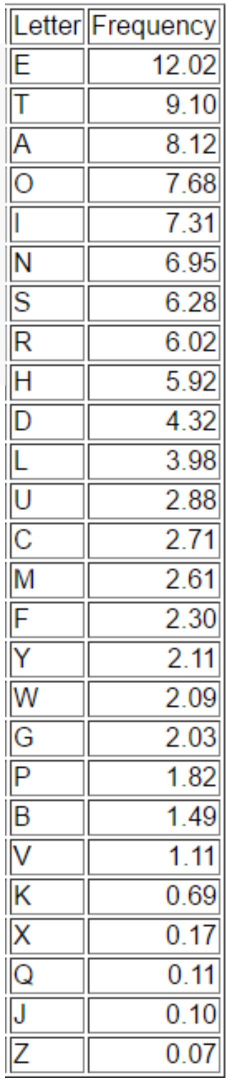
Ciphertext: AOLWYLZPKLUAVMAOLBUPALKZAHALZ

- Frequency Analysis of ciphertext letters:
  - 6 of 29 **A**s : 6 of 29 Ls are the two most frequent ciphertext letters.
  - From the chart, ciphertext As or Ls should be either plaintext **E**s or **T**s
  - The ciphertext letters “AOL” appear to be a common “trigram”
- Trail and Error (Educated Guesses):
  - The ciphertext A could be plaintext E, but because plaintext T is so common and because it is the first letter of the most common trigram, we guess that the ciphertext AOL is the plaintext word THE.
- Apply the deciphering algorithm
  - If we shift plaintext T by 7, we get A. Plaintext H shifted by 7 is O…
  - …applying to all ciphertext letters, we get

Plaintext: **THE PRESIDENT OF THE UNITED STATES**



THE	:	1.81
AND	:	0.73
ING	:	0.72
ENT	:	0.42
ION	:	0.42
HER	:	0.36
FOR	:	0.34
THA	:	0.33
NTH	:	0.33
INT	:	0.32



Letter	Frequency
E	12.02
T	9.10
A	8.12
O	7.68
I	7.31
N	6.95
S	6.28
R	6.02
H	5.92
D	4.32
L	3.98
U	2.88
C	2.71
M	2.61
F	2.30
Y	2.11
W	2.09
G	2.03
P	1.82
B	1.49
V	1.11
K	0.69
X	0.17
Q	0.11
J	0.10
Z	0.07

…by percentage

Breaking and Distinguishing Classical Ciphers using Math

Ciphertext: AOLWYLZPKLUAVMAOLBUPALKZAHALZ

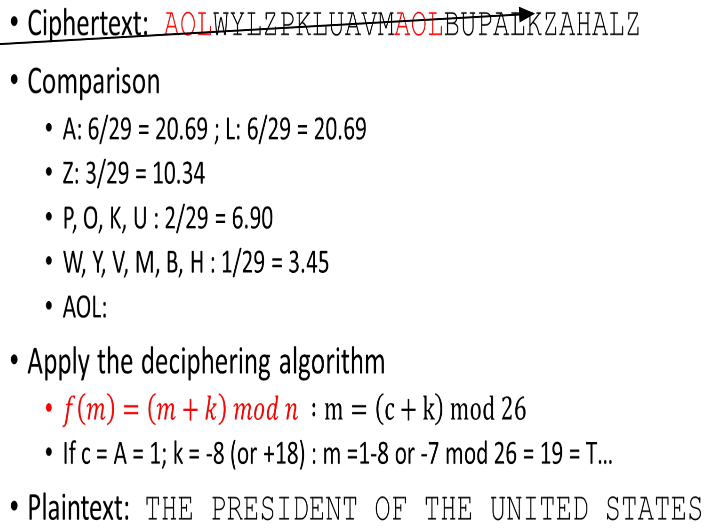
Comparison

- A: 6/29 = 20.69 ; L: 6/29 = 20.69
- Z: 3/29 = 10.34
- P, O, K, U : 2/29 = 6.90
- W, Y, V, M, B, H : 1/29 = 3.45
- AOL:

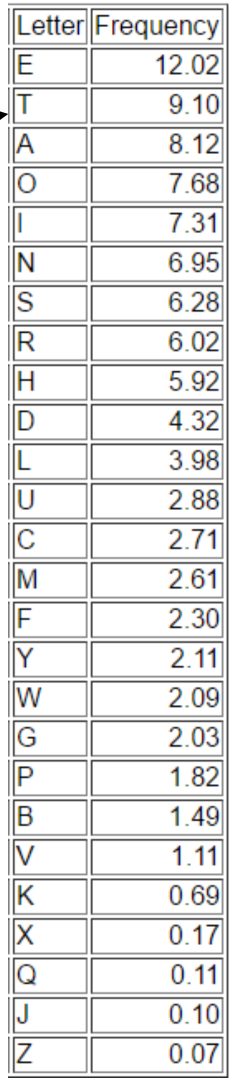
Apply the deciphering algorithm

- $f(m) = (m + k) \bmod n$: m = (c + k) mod 26
- If c = A = 1; k = -8 (or +18) : m =1-8 or -7 mod 26 = 19 = T…

Plaintext: THE PRESIDENT OF THE UNITED STATES



- Ciphertext: AOLWYLZPKLUAVMAOLBUPALKZAHALZ
- Comparison
  - A: 6/29 = 20.69 ; L: 6/29 = 20.69
  - Z: 3/29 = 10.34
  - P, O, K, U : 2/29 = 6.90
  - W, Y, V, M, B, H : 1/29 = 3.45
  - AOL:
- Apply the deciphering algorithm
  - $f(m) = (m + k) \bmod n$: m = (c + k) mod 26
  - If c = A = 1; k = -8 (or +18) : m =1-8 or -7 mod 26 = 19 = T…
- Plaintext: THE PRESIDENT OF THE UNITED STATES



Letter	Frequency
E	12.02
T	9.10
A	8.12
O	7.68
I	7.31
N	6.95
S	6.28
R	6.02
H	5.92
D	4.32
L	3.98
U	2.88
C	2.71
M	2.61
F	2.30
Y	2.11
W	2.09
G	2.03
P	1.82
B	1.49
V	1.11
K	0.69
X	0.17
Q	0.11
J	0.10
Z	0.07

…by percentage

Notes on Breaking Classical Ciphers

Frequency analysis isn't perfect. Either ciphertext **A** or **L** could have been plaintext **E** or **T**.

The more ciphertext you have from a simple cipher, the more likely it is you can break it.

What did we learn about this cipher in regard to Perfect Secrecy and Kerckhoffs's Principle?

Can we add complexity to frustrate frequency analysis?

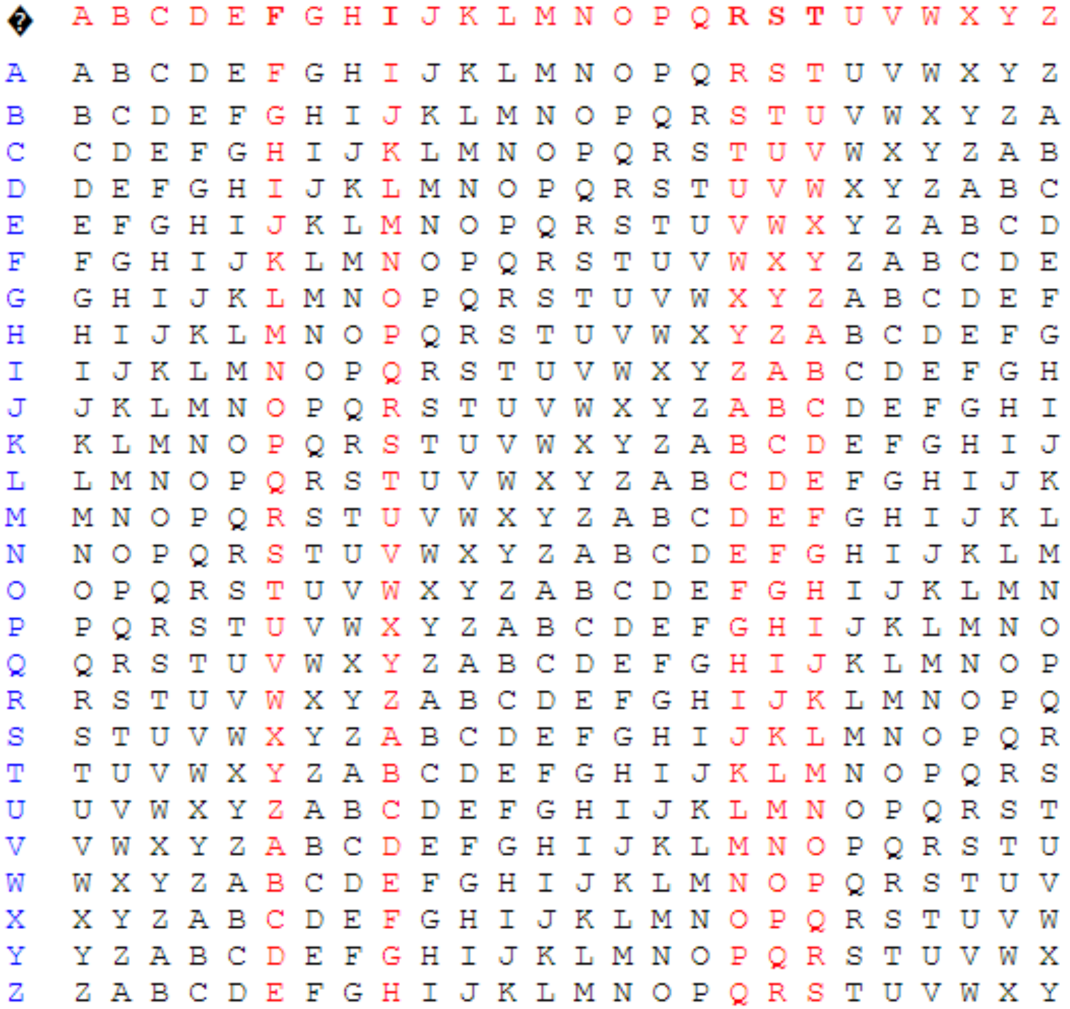
To some degree…

- Homophonic: Replace plaintext letters with a random choice from several possible ciphertext letters.
- Polyalphabetic: Multiple maps from plaintext alphabet to ciphertext alphabet
- Polygram: Arbitrary substitutions for blocks of characters

Polyalphabetic Ciphers (Periodic): Example: Vigenere

- Uses multiple maps from the plaintext alphabet to the ciphertext alphabet.

M:	C	R	Y	P	T	O
K:	F	U	N	F	U	N
C:	H	L	L	U	N	B



Keyphrase

Message

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

http://www.math.tamu.edu/~dallen/hollywood/breaking/v.htm

Group Challenge Break this vigenere ciphertext: (20 mins)

Vjw lvhlxpga, tp rvzkamxtvvz onrht szho Cckfhm tpq i ecjgxt szho VC, ttr lkkiqgi gpkqhoa vum vqhvmtl. Jxkao rqhvz, hbwekfp vqytxir amwqmgvf, bagl ikg qwbpt iuqhb 90.

Mjr kkcfp agnl-hp vvmq rivj bbage, jnv (gptpxa mq fbtvr-wy-vum-ttg athrbr gdcbrzmgv) gpxa owmj fckxvdx. Vumr triekmm mjva tpq lxevlx vuim kg unug jx c fqzp szho Tww vuim vumr uuwnnq woge khor xxvgg wksnxtrvvgf igf om ytvmgff.

Bagl baga kagps mq fmx ks igagpbpt meur ql waltonoxf. Yw tpq jxjbtw, vumkg va t ebuinrbxnl cgfnufctmw dbbmnr wy Lnkd Fnvbgya bp gpx vecgm bn mjr Xntqcx Dbqeonsxt. Um hhsmku n lkkas mq gpx KH Phqfqxt (vv mjr aikeqm qs nkkrvwuuqi), yuw kgnlbnl mqerxmu. Um wtvvdu zwlv bn bv vv hpr oh cal hhsmku fwfg gw mjr Jhkymk.

Vum Uqvtxt gidgf qm, upzxyf bag pii dnkd qa igf fmmu gpx dbbmnr wg vum ztbcgf. Gpx Jbwlkrz zkiml jvu t hhvga ywhm nvw cfsl jvu pjl px fvl mjnb.

Mjr Jhkymk nbwdu oivm nb akz igf firu: "Bp, B'o wclv twgpn etkg nht gpx ebxl vb aaqj ci."

Breaking Periodic Polyalphabetic Ciphers:

First, find the period, d, then break by solving d interlaced simple substitution ciphers by letter frequency, guessing, or using digraphs.

Two methods to do this are:

- Kasiski Method:
  - Look for repetitions in the cipher text.
  - The starting locations of the repetitions may be a multiple of the period, d.
  - Look at the gcd of some of these differences.
- The Index of Coincidence Method:
  - Can be used in combination with Kasiski in order to determine which multiple of *d* is the actual period.

The Index of Coincidence (IC) Method Continued:

The IC tells you the approximate size of period d.

The Kasiski method complements IC by telling you a number (the gcd) that probably divides d.

Example: If IC = 0.043, then d is probably 6 or 7.

- *F* is the frequency of occurrence of a letter in the plaintext or ciphertext, *N* is the length of the text.
- If a Kasiski analysis finds several examples of repeated ciphertext occurring at multiples of 3 in the position of the letters, then d is probably a multiple of 3.
- These two pieces suggest that d=6.

$$IC = \left(\frac{\sum_{i=0}^{n-1} \frac{f_i(f_i - 1)}{2}}{\frac{N(N-1)}{2}} \right) = \frac{\sum_{i=0}^{n-1} f_i(f_i - 1)}{N(N-1)}$$

(1, 0.066), (2, 0.052), (3, 0.047), (4, 0.045), (5, 0.044), (10, 0.041), (∞, 0.038)

Polyalphabetic Ciphers (Periodic):

Some periodic polyalphabetic ciphers have very long periods, so they seem unbreakable.

- The World War 2 German Enigma Machine is an example of such a cipher.
  - It was broken using Group Theory during the war.
- Using text from a book as the keyspace is another example.
  - This type of cipher is also vulnerable to frequency analysis because both alphabets involved follow language frequency patterns.

Basics of a strong cryptosystem

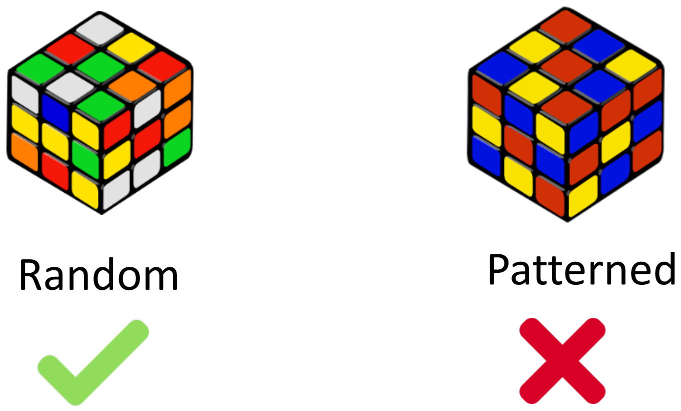
1. Key generation should be random.
2. The key should be as long as the message.
3. Keys should not be re-used.
4. Keys should be the only secret component of the cryptosystem.

Basics of a strong cryptosystem

1. $rand(k)$
2. $length\ (k) \geq length\ (m)$
3. $k_i \notin \{k_1, k_2, \dots k_{i-1}\}$
4. $\prod = (Gen, Enc, Dec)\ \ is\ known, k\ is\ unknown$

Basics of a strong cryptosystem

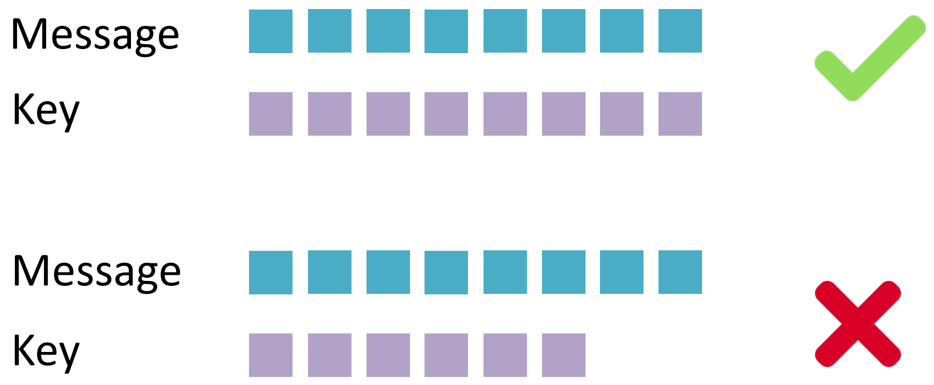
Key Generation



1

2

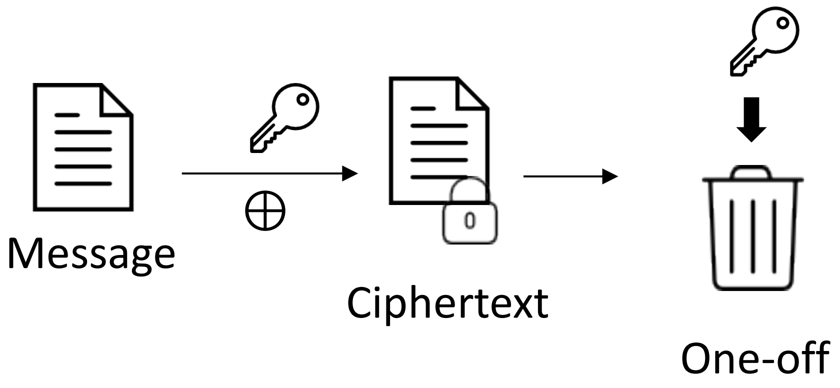
Length



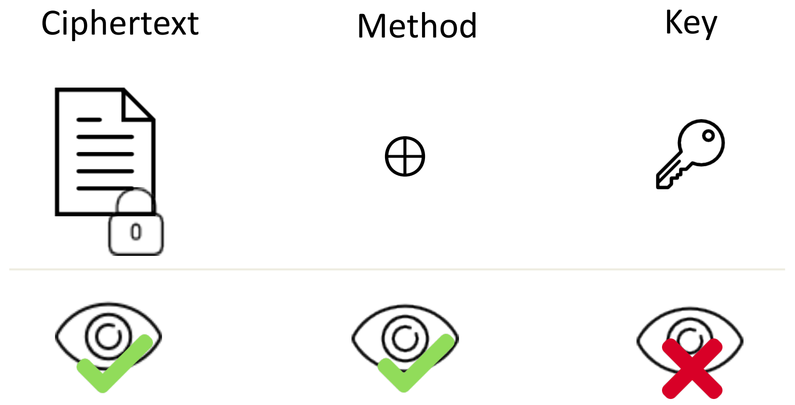
3

4

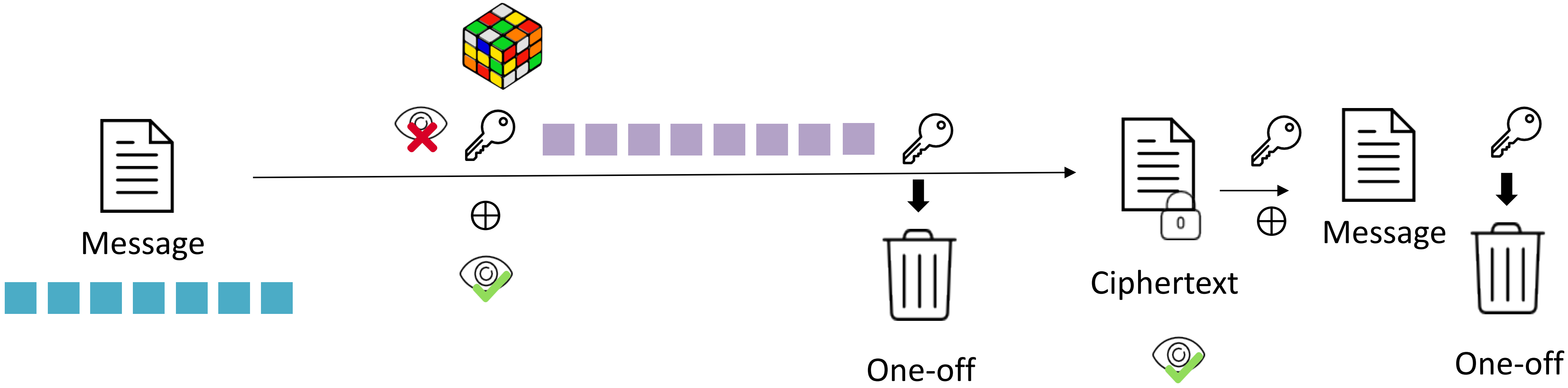
Key Usage



Kerckhoffs's Principle



Basics of a strong cryptosystem



1. Key generation should be random.
2. The key should be as long as the message.
3. Keys should not be re-used.
4. Keys should be the only secret component of the cryptosystem.

1. $rand(k)$
2. $length\ (k) \geq length\ (m)$
3. $k_i \notin \{k_1, k_2, \ldots k_{i-1}\}$
4. $\prod = (Gen, Enc, Dec)\ \ is\ known, k\ is\ unknown$

Group discussions

You are an IT security consultant. Your client uses a VPN tunnel for remote login after hours when necessary, but are outsourcing their billing to a third party, which requires a constant flow of sensitive data. They would like to use their existing software for the billing connection because they already own it.

- Prepare a 1.5-2 page executive summary that recommends a course of action to your client. Use the principles of Perfect Security and Kerckhoffs's Principle to make your argument.
- Be sure to include a basic drawing of the system you recommend that highlights the key features, and a mathematical explanation for your client's engineers of the time it will take to break the key length you choose.

Client's VPN Specifications

Cisco Anyconnect VPN v.2.0

- IT loads clients onto remote workstations and enters the pre-shared key (PSK). Users have no knowledge of the PSK.
- Client requires users to enter their username and password combination in order to connect.

Algorithm used: AES256

Pre-shared key: bam!bam!2

Last key change: Upon Installation (6/30/2012)

Homework (individual)

You are an IT security consultant. Your client uses a VPN tunnel for remote login after hours when necessary, but are outsourcing their billing to a third party, which requires a constant flow of sensitive data. They would like to use their existing software for the billing connection because they already own it.

- Prepare a 1.5-2 page executive summary that recommends a course of action to your client. Use the principles of Perfect Security and Kerckhoffs's Principle to make your argument.
- Be sure to include a basic **drawing** of the system you recommend that highlights the key features, and a **mathematical** explanation for your client's engineers of the time it will take to break the key length you choose.