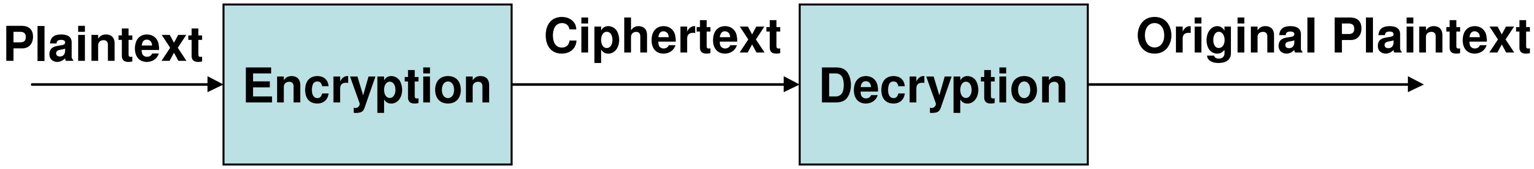
Module 1: Classical Symmetric Ciphers

Dr. Natarajan Meghanathan
Professor of Computer Science
Jackson State University
E-mail: natarajan.meghanathan@jsums.edu

Introduction to Cryptography

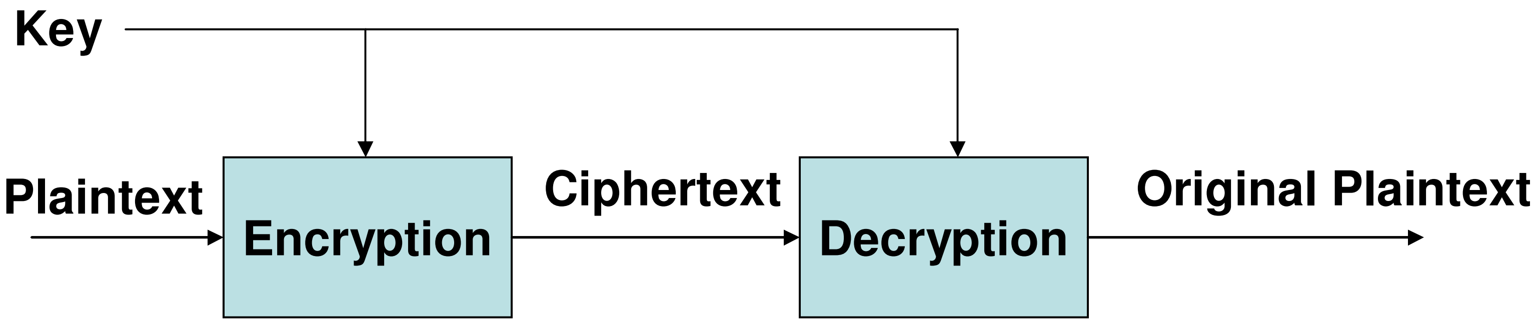
- Terms and Concepts
  - Cryptography: Performing encryption and decryption
  - Encryption: the process of transforming a message so that its meaning is not obvious
  - Decryption: the process of transforming an encrypted message back into its original, normal form
  - Cryptosystem: A system for encryption and decryption
  - Plaintext: Original, unencrypted, form of a message
  - Ciphertext: The encrypted form of a message

  - Formal notation: We seek a cryptosystem for which $P = D(E(P))$, where P is the Plaintext, E is the Encryption rule, $C = E(P)$ is the ciphertext, D is the Decryption rule.



Types of Encryption

- Symmetric encryption: The same key performs, both encryption and decryption.
  - P = D(K, E(K, P))



Key

Plaintext

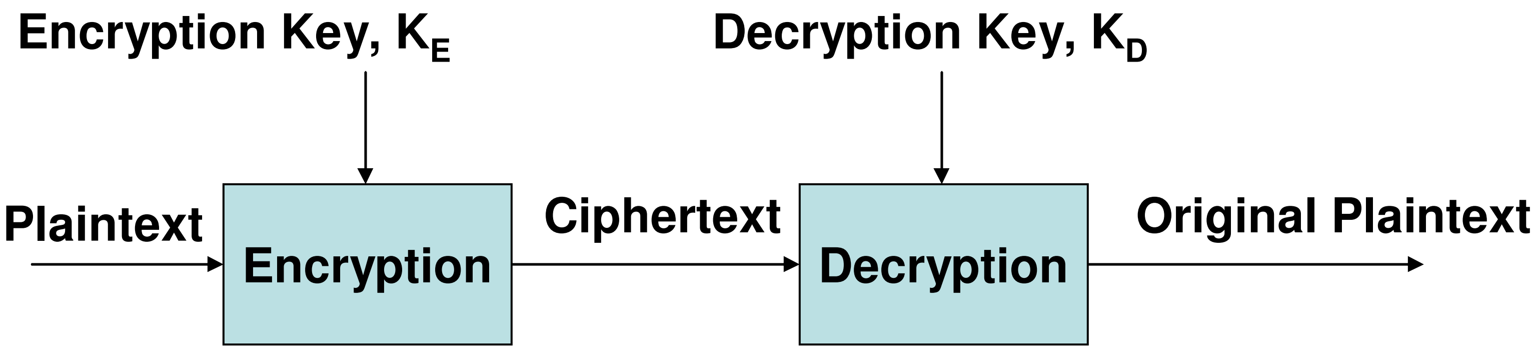
Encryption

Ciphertext

Decryption

Original Plaintext

- Asymmetric encryption: distinct, very different keys, one for encryption and the other for decryption only



Encryption Key, K_E

Decryption Key, K_D

Plaintext

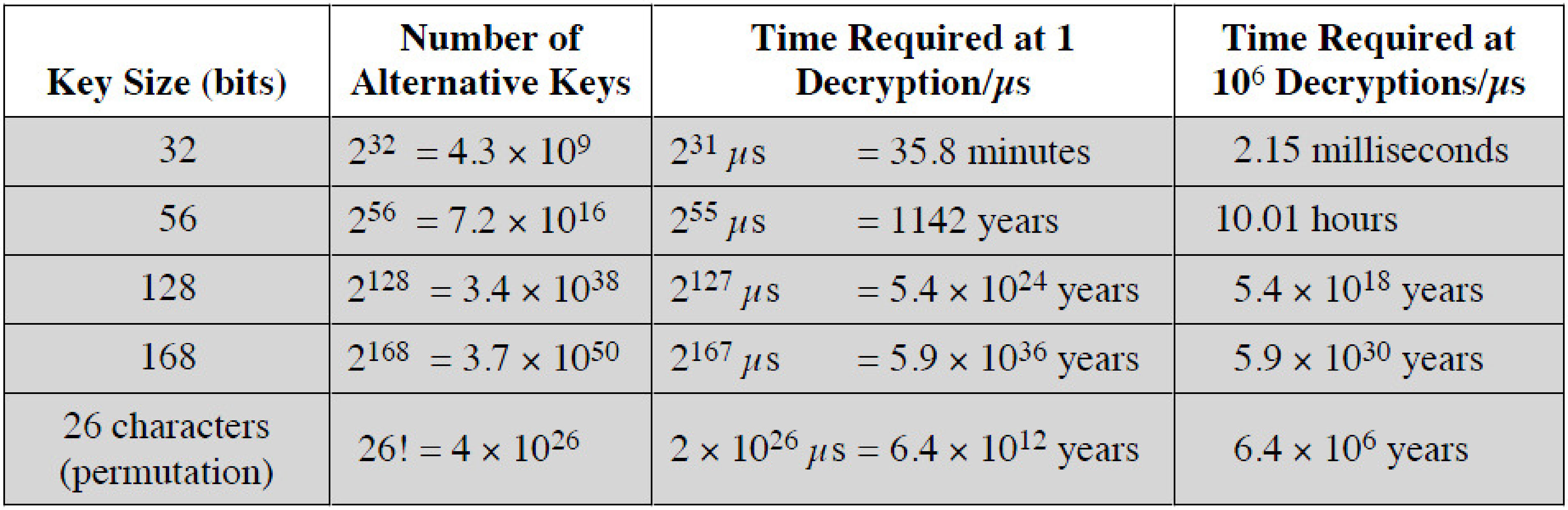
Encryption

Ciphertext

Decryption

Original Plaintext

Average Time required for Exhaustive Search



Key Size (bits)	Number of Alternative Keys	Time Required at 1 Decryption/μs	Time Required at 10^6 Decryptions/μs
32	$2^{32} = 4.3 \times 10^9$	$2^{31}\ \mu s = 35.8$ minutes	2.15 milliseconds
56	$2^{56} = 7.2 \times 10^{16}$	$2^{55}\ \mu s = 1142$ years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	$2^{127}\ \mu s = 5.4 \times 10^{24}$ years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	$2^{167}\ \mu s = 5.9 \times 10^{36}$ years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	$2 \times 10^{26}\ \mu s = 6.4 \times 10^{12}$ years	6.4×10^6 years

Source: Table 2.2 from William Stallings – Cryptography and Network Security, 5th Edition

Cryptanalysis

- Cryptanalysis is the process of breaking an encryption.
- A cryptanalyst can attempt to do any or all six of the following:
  - Break a single message
  - Recognize patterns in encrypted messages, to be able to break subsequent ones by applying a straightforward decryption algorithm
  - Infer some meaning regarding the communication (for example, the length of the communication, the frequency of the communication)
  - Deduce the key to break subsequent messages easily
  - Find weaknesses in the implementation or the environment of use of encryption
  - Find general weaknesses in an encryption algorithm, without necessarily having intercepted any messages

- A cryptanalyst would work with a variety of pieces of information:
  - Encrypted messages, known encryption algorithms, intercepted plaintext, data items known or suspected to be in a ciphertext message, mathematical and/or statistical tools and techniques, properties of languages, computers, etc.

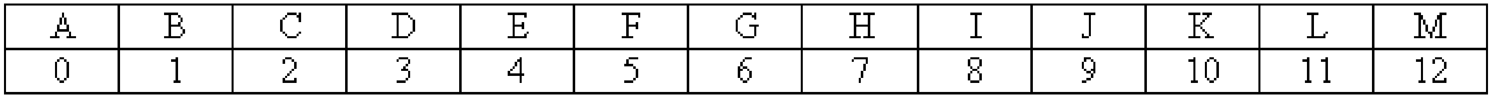
Cryptanalysis

Type of Attack	Known to Cryptanalyst
Ciphertext only	• Encryption algorithm, • Ciphertext
Known Plaintext	• Encryption algorithm, • Ciphertext • One or more plaintext-ciphertext pairs formed with the secret key
Chosen Plaintext	• Encryption algorithm, • Ciphertext • Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key
Chosen Ciphertext	• Encryption algorithm, • Ciphertext • Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

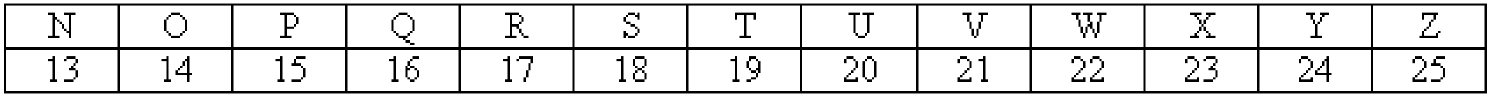
Source: Table 2.1 from William Stallings – Cryptography and Network Security, 5th Edition

Representing Characters

- Conventions/ Assumptions:
  - The plaintext is written in UPPERCASE letters and the ciphertext in lowercase letters
  - We use a numeric encoding for the letters as shown below as most encryption algorithms are based on mathematical transformations.



Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Code	0	1	2	3	4	5	6	7	8	9	10	11	12

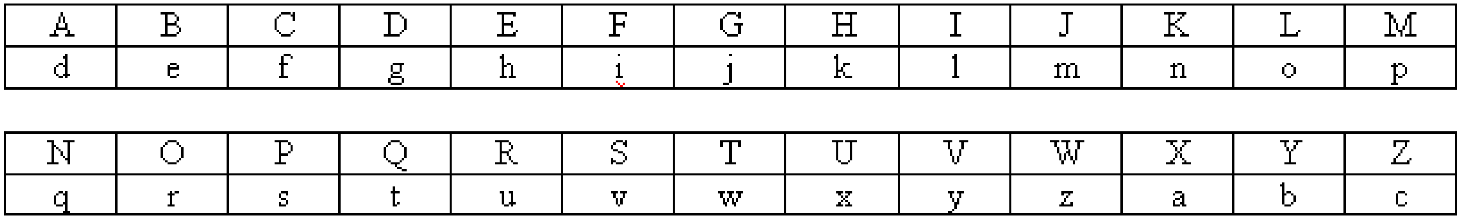


Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Code	13	14	15	16	17	18	19	20	21	22	23	24	25

  - We can perform arithmetic on the letters of a message. For example, A + 3 = D, K – 1 = J
  - Arithmetic is performed as if the above alphabetic table were circular. In other words, all arithmetic is with respect to modulo 26. The result of every arithmetic operation is between 0 and 25.
    - For example, Y + 3 = B

Substitution Ciphers

- Idea: Use a correspondence table and substitute a character or symbol for each character of the original message
- Goal of substitution is Confusion: an attempt to make it difficult for a cryptanalyst or an intruder to determine how a message and key were transformed into ciphertext.
- Caesar Cipher
  - Each letter is translated to the letter a fixed number of times after it in the alphabet table
  - Caesar cipher uses shift by 3.
  - $C_i = E(p_i) = p_i + 3$



Plaintext	A	B	C	D	E	F	G	H	I	J	K	L	M
Ciphertext	d	e	f	g	h	i	j	k	l	m	n	o	p

Plaintext	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Ciphertext	q	r	s	t	u	v	w	x	y	z	a	b	c

  - Example:
    - Plaintext: TREATY IMPOSSIBLE
    - Ciphertext: wuhdwb lpsrvvleoh
  - Complexity of the encryption algorithm: Length of the message

Cryptanalysis of the Caesar Cipher

- On a closer look at the result of applying Caesar's encryption technique to "TREATY IMPOSSIBLE", we get the following clues from the ciphertext even if did not know the plaintext:
  - The break in between the two words is preserved in the ciphertext
  - Double letters are preserved (SS is translated to vv)
  - When a letter is repeated, it maps to the same ciphertext as before (look at letters T, I, E in the plaintext)
- Consider you are given the following ciphertext and you want to determine the plaintext: "`wklv phvvdjh lv qrw wrr kdug wr euhdn`"
  - As a start, assume that the coder was lazy, and has allowed the blank space to be translated to itself. Hence, the message has actually been enciphered with a 27-symbol alphabet: A through Z and a blank-space separating the words.
  - If this assumption is true, knowing where the spaces helps to find out what are the small words.
  - The English language has very few short words like *am*, *is*, *to*, *be*, *he*, *she*, *we*, *and*, *you*, *are*, and so on.
- There is a strong clue in the repeated `r` of the word `wrr`.
  - Two very common three-letter words having the pattern `xyy` are `see` and `too`; other less common possibilities are `add`, `odd` and `off`. Try the more common word first.

Cryptanalysis of the Caesar Cipher

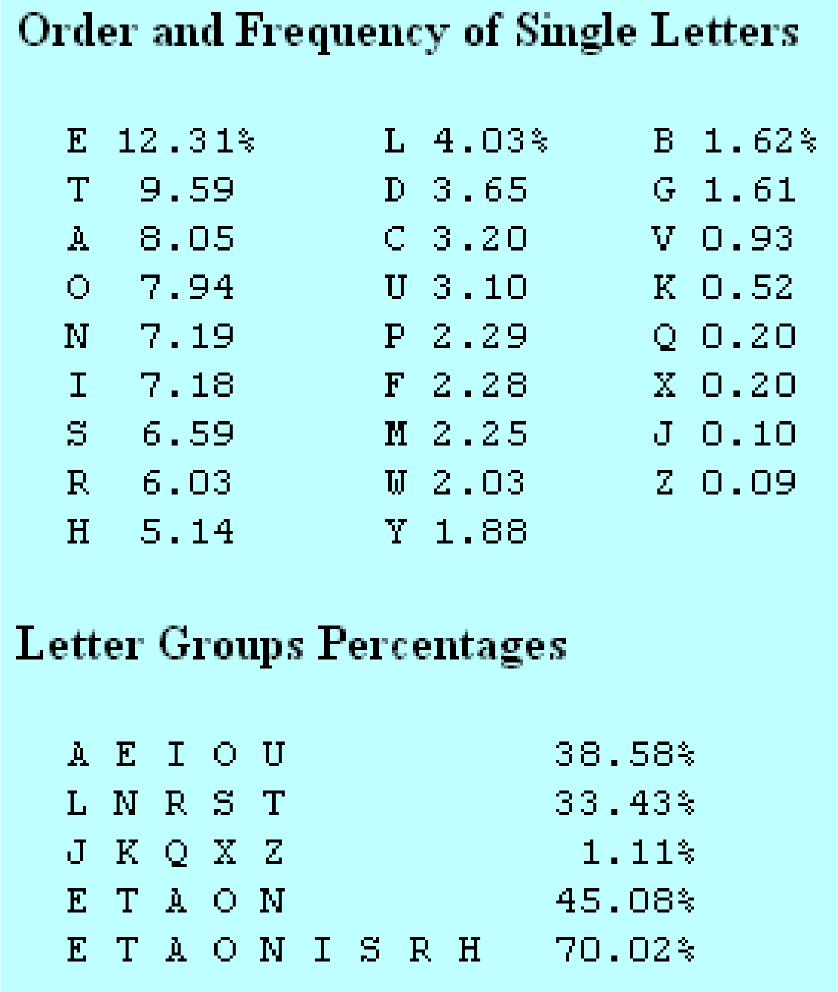
  - Also, the combination `wr` appears in the ciphertext too, so you can determine whether the first two letters of the three-letter word form a separate word by themselves.
  - `wklv phvvdjh lv qrw wrr kdug wr euhdn`
  - `T--- ------- -- -OT TOO ---- TO -----`
  - The `-OT` could be `cot`, `dot`, `got`, `hot`, `lot`, `not`, `pot`, `rot` or `tot`. A likely choice is not. So `q` = `N`
  - The word lv is also the end of the word wklv.
    - lv cannot be SO, because then wklv is T-SO. There is no such word
    - lv cannot be IN, because we have q = N
    - lv has to be IS, so wklv is THIS
  - `wklv phvvdjh lv qrw wrr kdug wr euhdn`
  - `THIS --SS--- IS NOT TOO H--- TO -----`
- By now, we should be able to figure out that the shift has been by three characters for each character in the plaintext. So, the plaintext for the given ciphertext is:
  - `wklv phvvdjh lv qrw wrr kdug wr euhdn`
  - `THIS MESSAGE IS NOT TOO HARD TO BREAK`

Cryptanalysis of Substitution Ciphers

- <u>Some clues to break the code more quickly</u>
  - The frequency with which certain letters are used can help us to break the code more quickly.
    - The letters E, T, O, A occur more often the letters J, Q, X, Z
  - The nature and context of the text being analyzed affects the distribution
    - In a medical article in which the term x-ray may be used often, the letter x would have an uncommonly high frequency
  - Letters appear to each other with predictable frequency
  - In usual English, EN, RE, ER,…, and ENT, ION, AND,… are most frequently-occurring coincident pairs (digrams) and triples (trigrams) of letters
  - Digrams and trigram frequencies are well-known for all written languages
  - Frequency distribution may not give complete decryption, due to peculiarities of plaintext, but considerably narrows down choices.

- Short messages give a cryptanalyst little to work with as the latter works by finding patterns (possible to obtain more with long messages). So, shorter messages are fairly more secure with simple encryption algorithms.

Useful English Language Statistics

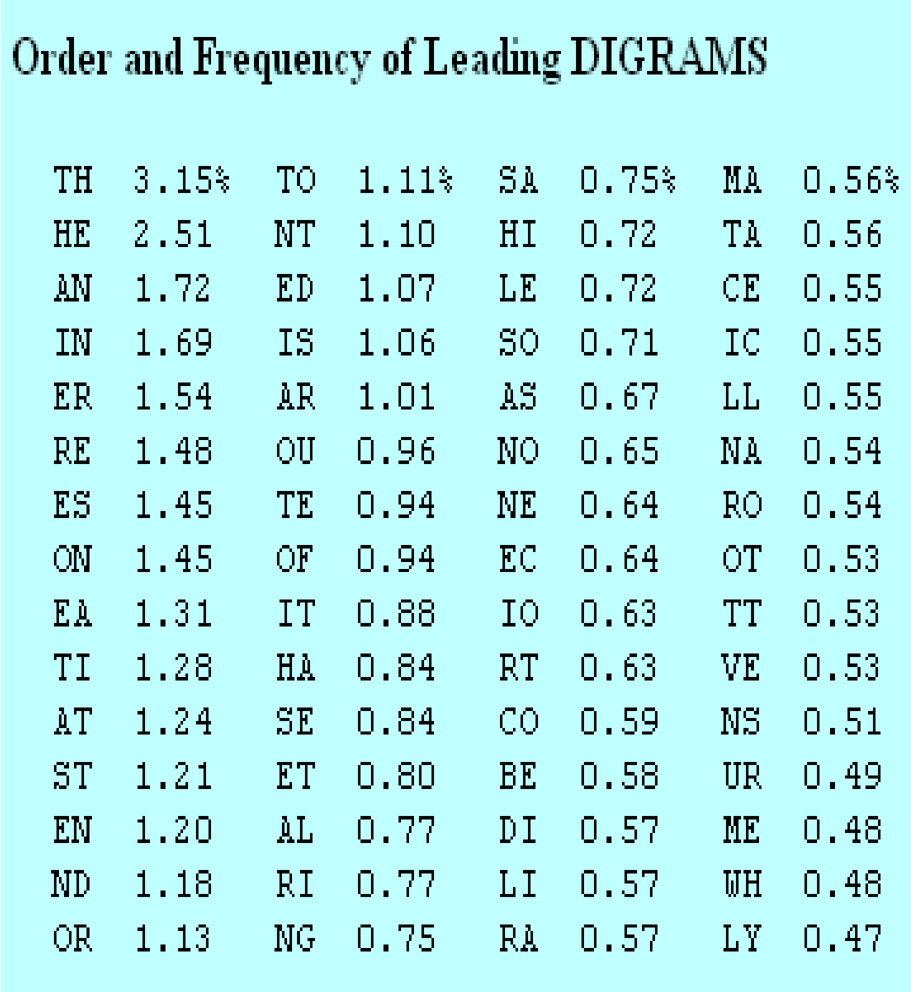


Order and Frequency of Single Letters

E	12.31%	L	4.03%	B	1.62%
T	9.59	D	3.65	G	1.61
A	8.05	C	3.20	V	0.93
O	7.94	U	3.10	K	0.52
N	7.19	P	2.29	Q	0.20
I	7.18	F	2.28	X	0.20
S	6.59	M	2.25	J	0.10
R	6.03	W	2.03	Z	0.09
H	5.14	Y	1.88		

Letter Groups Percentages

A E I O U	38.58%
L N R S T	33.43%
J K Q X Z	1.11%
E T A O N	45.08%
E T A O N I S R H	70.02%



Order and Frequency of Leading DIGRAMS

TH	3.15%	TO	1.11%	SA	0.75%	MA	0.56%
HE	2.51	NT	1.10	HI	0.72	TA	0.56
AN	1.72	ED	1.07	LE	0.72	CE	0.55
IN	1.69	IS	1.06	SO	0.71	IC	0.55
ER	1.54	AR	1.01	AS	0.67	LL	0.55
RE	1.48	OU	0.96	NO	0.65	NA	0.54
ES	1.45	TE	0.94	NE	0.64	RO	0.54
ON	1.45	OF	0.94	EC	0.64	OT	0.53
EA	1.31	IT	0.88	IO	0.63	TT	0.53
TI	1.28	HA	0.84	RT	0.63	VE	0.53
AT	1.24	SE	0.84	CO	0.59	NS	0.51
ST	1.21	ET	0.80	BE	0.58	UR	0.49
EN	1.20	AL	0.77	DI	0.57	ME	0.48
ND	1.18	RI	0.77	LI	0.57	WH	0.48
OR	1.13	NG	0.75	RA	0.57	LY	0.47

Monoalphabetic Cipher

- The key is 26 characters long – each plaintext letter maps to a different randomly chosen ciphertext letter.
- Even though, we have 26! Keys (there can be 26! Permutations of 26 characters), the cipher does not sufficiently obscure the underlying language characteristics and could be prone to successful cryptanalytic attacks using the statistics of the English language.

```
Plain:   abcdefghijklmnopqrstuvwxyz
Cipher:  DKVQFIBJWPESCXHTMYAUOLRGZN
```

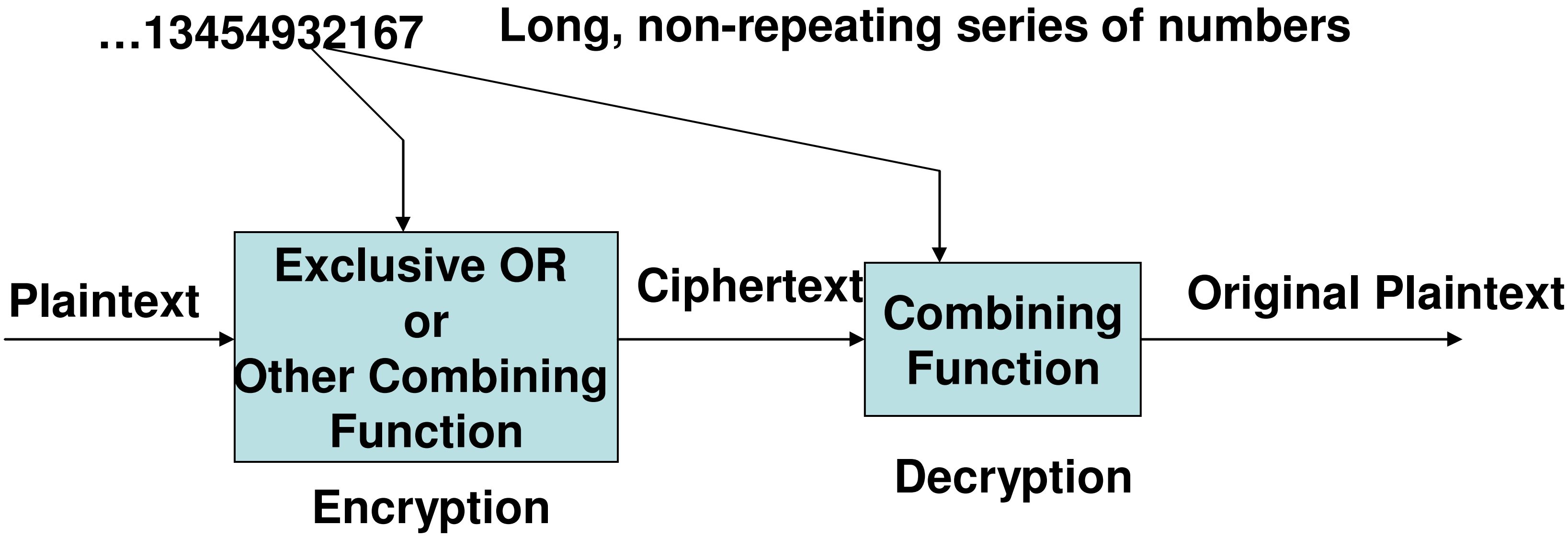
```
Plaintext:    ifwewishtoreplaceletters
Ciphertext:   WIRFRWAJUHYFTSDVFSFUUFYA
```

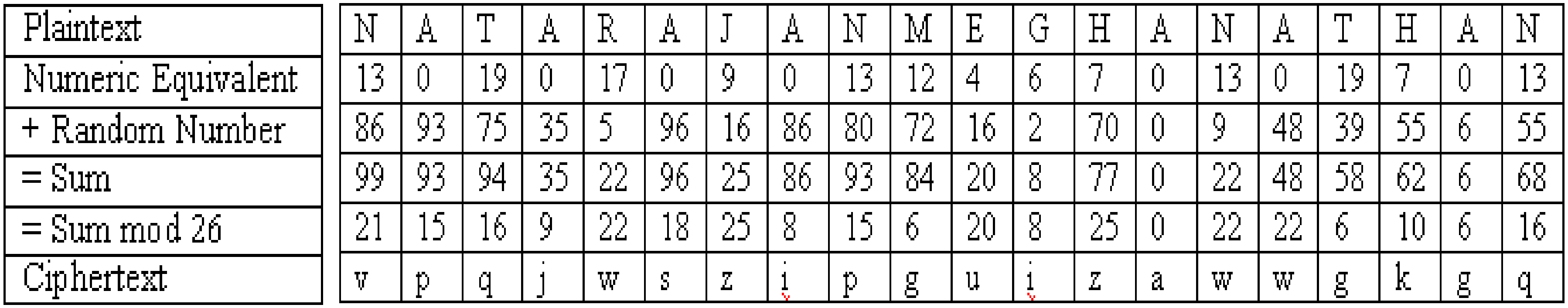
One-Time Pads

- The key size is as long as the message to be encrypted.
- The key is a non-repeating, arbitrarily long series of unpredictable (random) numbers, to which both sender and receiver, but nobody else has access.
- One-time pad is the only theoretically unbreakable cipher, because:
  - If the random numbers are truly unpredictable and non-repeating, there is no way to tell how a particular plaintext character was enciphered. Even if it is known that A is enciphered to W this time, that gives no knowledge about the encipherment of the next A or the plaintext character that follows A.

- The Vernam Cipher
  - A type of one-time pad
  - Combines a long non-repeating sequence of randomly generated numbers with the plaintext
  - As long as the random numbers are non-repeating, the Cipher code is immune to cryptanalytic attacks.

Vernam Cipher





Plaintext	N	A	T	A	R	A	J	A	N	M	E	G	H	A	N	A	T	H	A	N
Numeric Equivalent	13	0	19	0	17	0	9	0	13	12	4	6	7	0	13	0	19	7	0	13
+ Random Number	86	93	75	35	5	96	16	86	80	72	16	2	70	0	9	48	39	55	6	55
= Sum	99	93	94	35	22	96	25	86	93	84	20	8	77	0	22	48	58	62	6	68
= Sum mod 26	21	15	16	9	22	18	25	8	15	6	20	8	25	0	22	22	6	10	6	16
Ciphertext	v	p	q	j	w	s	z	i	p	g	u	i	z	a	w	w	g	k	g	q

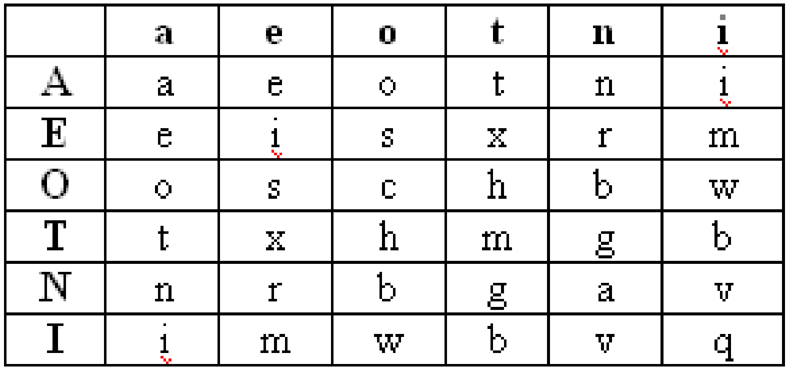
Book Ciphers

- Book cipher is a variation of the well-known Vignere cipher
- The key comes from a text portion starting from a certain page of a book. Both the sender and receiver should have the same edition of the book.
- Consider encrypting the message `MACHINES CANNOT THINK`
- Using the Key: `i am i exist that is certain`
- The ciphertext is the character corresponding to the cell at the intersection of the row of the plaintext character and the column of the character in the key
- Cryptanalysis becomes difficult with more flatter frequency distribution.
- Encryption: $C_i \equiv P_i + K_i \pmod{26}$
- Decryption: $P_i \equiv C_i - K_i \pmod{26}$

- Example for Book Cipher: Use a character grouping of size 5
  - Plaintext: `MACHI NESCA NNOTT HINK`
  - Key: `iamie xistt hatis cert`
  - Ciphertext: `uaopm kmkvt unhbl jmed`

Cryptanalysis of Book Ciphers

- The probability that a given character in the plaintext is any one of E, A, O, T, N or I is close to 50%.
- Similarly, the probability that a given character in the key (taken from a book) is any one of E, A, O, T, N or I is close to 50%.
- To break the cipher, assume that each letter of the ciphertext comes from a situation in which the plaintext letter (row selector) and the key letter (column selector) are both one of the six most frequent letters.
- A sub-table of the Vigenere tableau table that lists the intersections between these six characters is given below:



	a	e	o	t	n	i
A	a	e	o	t	n	i
E	e	i	s	x	r	m
O	o	s	c	h	b	w
T	t	x	h	m	g	b
N	n	r	b	g	a	v
I	i	m	w	b	v	q

Cryptanalysis of Book Ciphers

- Searching through the sub-table for possibilities, we have:

```
– Ciphertext:         uaopm kmkvt unhbl jmed
– Possible            ?AA?E ?E?NA ?AOO? ?EA?
  Plaintexts:          NO T  T IT  NTT   TE
                          I  I           I
- Actual Plaintext:   MACHI NESCA NNOTT HINK
```

- Out of the 25 predictions, 8 were correct. Hence, the percentage correctness in the predictions is 8/25 = 32%.

Transpositions (Permutations)

- A transposition is an encryption in which the letters of the message are rearranged.
- Goal of Transposition: To aim for diffusion, widely spreading the information from the message or key across the ciphertext.

- Columnar Transpositions
  - Agree on the number of columns to be used for the transposition.
  - The plaintext characters are rearranged into these many columns and sent one column after another.
  - If the message length is not a multiple of the number of columns, the last few columns would be short of a letter compared to the other columns. In such a case, fill these short columns with an infrequently appearing letter, say ‘x’
  - Difference with substitution ciphers:
    - Space required is directly related to the length of the message. Substitution ciphers required constant storage space (for the correspondence table)
    - The first column of the message cannot be output until the whole message is read. Delay associated with decrypting the ciphertext depends on the length of the message.

Example for Columnar Transposition

- Plain text: THIS IS A MESSAGE TO SHOW HOW COLUMNAR TRANSPOSITION WORKS
- Ciphertext: TSSOHLRSTOHAASOUTPIRIMGHWMROOKSEEOC NASNSISTWOANIWX
- At the receiver:
  - To figure out each column and the number of rows, the receiver divides the message length by the number of columns agreed upon.

T	H	I	S	I
S	A	M	E	S
S	A	G	E	T
O	S	H	O	W
H	O	W	C	O
L	U	M	N	A
R	T	R	A	N
S	P	O	S	I
T	I	O	N	W
O	R	K	S	X
1	2	3	4	5

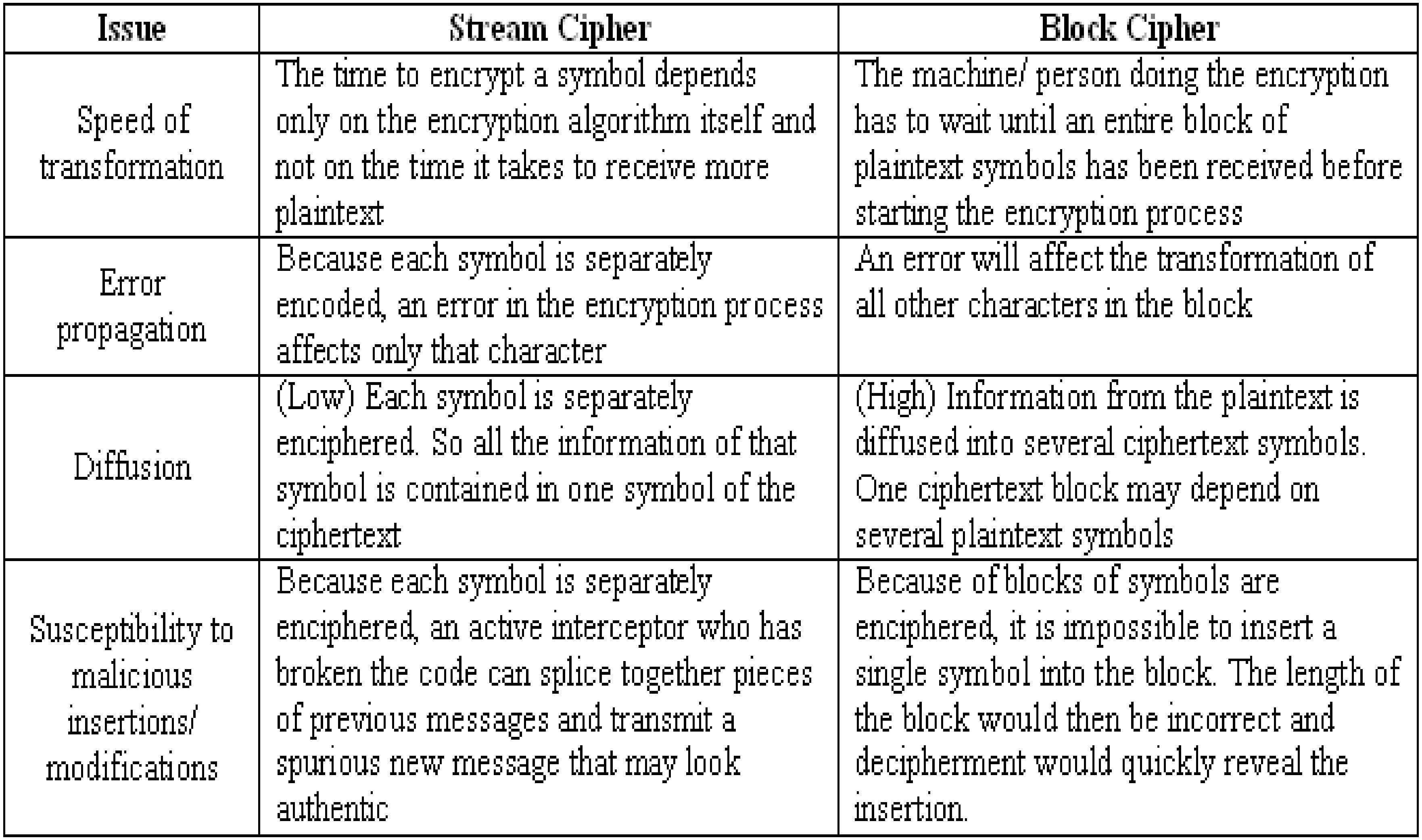
- To make it more secure, the sender and receiver could agree on a code word of length equal to the number of columns and then send the columns in the alphabetical order of the characters in the key word.
- Let the code word be ZEBRA. Then, the fifth column would be sent first, followed by the third column, followed by the second column and so on.
- Ciphertext: ISTWOANIWXIMGHWMROOKHAASOUTPIRSEEOC NASNSTSSOHLRSTO

Z	E	B	R	A
T	H	I	S	I
S	A	M	E	S
S	A	G	E	T
O	S	H	O	W
H	O	W	C	O
L	U	M	N	A
R	T	R	A	N
S	P	O	S	I
T	I	O	N	W
O	R	K	S	X
5	3	2	4	1

Stream and Block Ciphers

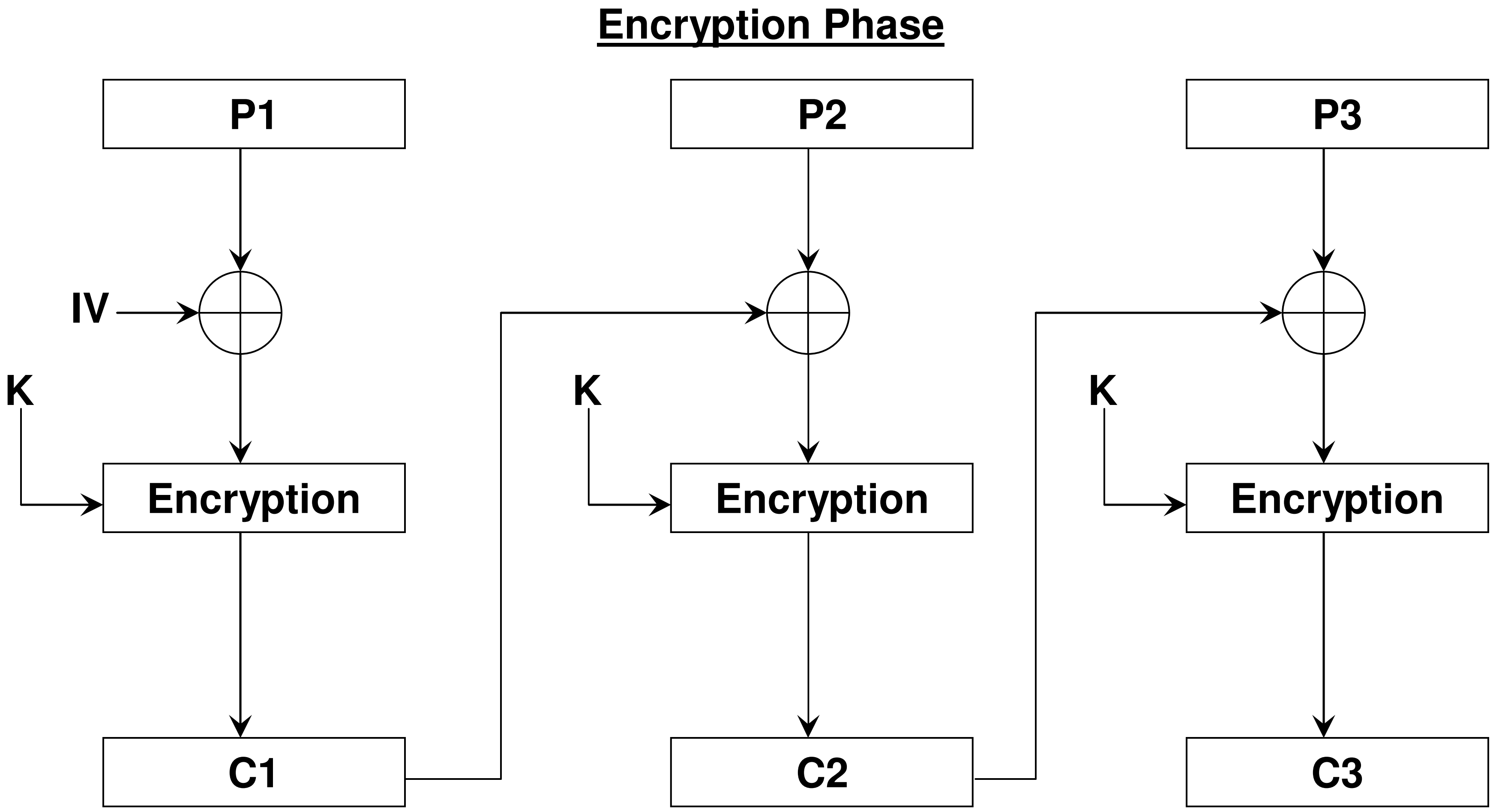
- Stream Cipher:
  - Converting one symbol of plaintext immediately into a symbol of ciphertext
  - The transformation depends only on the symbol, the key and the control information of the encipherment algorithm
  - Example: All substitution cipher algorithms
- Block Cipher:
  - Encrypts a `group` of plaintext symbols as `one block`
  - In columnar transposition, the entire message is translated as one block

Comparison: Stream and Block Ciphers



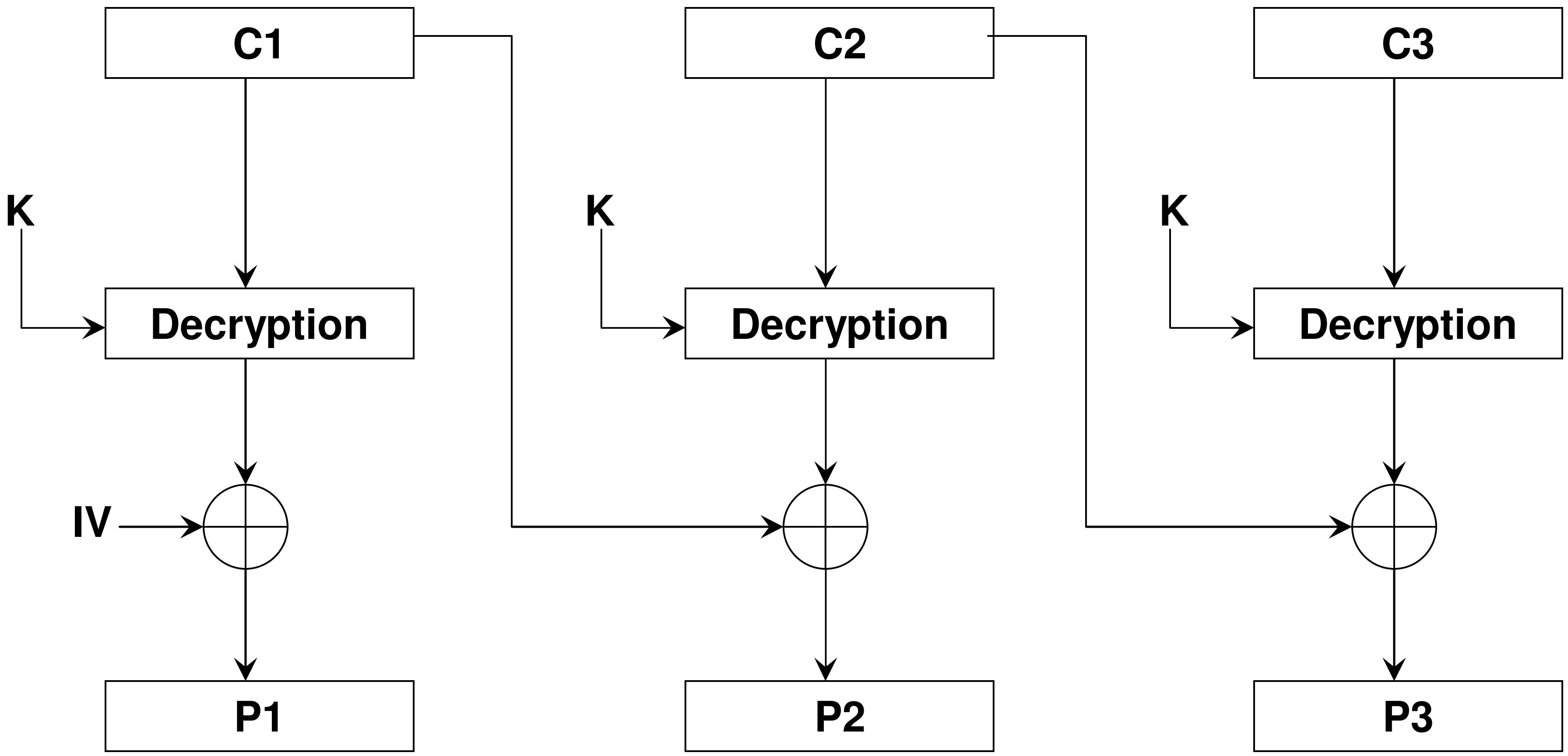
Issue	Stream Cipher	Block Cipher
Speed of transformation	The time to encrypt a symbol depends only on the encryption algorithm itself and not on the time it takes to receive more plaintext	The machine/ person doing the encryption has to wait until an entire block of plaintext symbols has been received before starting the encryption process
Error propagation	Because each symbol is separately encoded, an error in the encryption process affects only that character	An error will affect the transformation of all other characters in the block
Diffusion	(Low) Each symbol is separately enciphered. So all the information of that symbol is contained in one symbol of the ciphertext	(High) Information from the plaintext is diffused into several ciphertext symbols. One ciphertext block may depend on several plaintext symbols
Susceptibility to malicious insertions/ modifications	Because each symbol is separately enciphered, an active interceptor who has broken the code can splice together pieces of previous messages and transmit a spurious new message that may look authentic	Because of blocks of symbols are enciphered, it is impossible to insert a single symbol into the block. The length of the block would then be incorrect and decipherment would quickly reveal the insertion.

Cipher Block Chaining (CBC) to Create Avalanche Effect



Source (adapted from): Figure 6.4 from William Stallings – Cryptography and Network Security, 5th Edition

Cipher Block Chaining (CBC) to Create Avalanche Effect



Source: Figure 6.4 from William Stallings – Cryptography and Network Security, 5th Edition

Advantages and Limitations of CBC

- Advantages: A ciphertext block depends on all paintext blocks before it
- Any change to a plaintext block affects all of the succeeding ciphertext blocks – creates an Avalanche effect. This property can be used to compute a "Message Authentication Code" (MAC) for the entire plaintext and sent as part of the message.
- If the "integrity" of the message is the only required criterion, then we can send IV, P1, P2, …, P_{last_block}, MAC.
  - If any intruder changes any of the plaintext, the Avalanche Effect property of CBC requires that the MAC value computed by the destination to be different than what is sent by the sender as part of the message.

- Limitations: Needs an Initialization Vector (IV), which must be known to sender & receiver
- Since a recovered plaintext block (from CBC decryption) is not used for further decryptions and it is the only ciphertext of a previous block that is being used to decrypt the ciphertext of the current block, with CBC decryption, a change in a ciphertext block only affects two of the recovered plaintext blocks.
- The IV cannot be sent in clear text. Hence, the IV must be either a fixed value or must be sent encrypted offline (using schemes such as public-key encryption) before the rest of the message is sent.

Confusion and Diffusion

- Confusion
  - The interceptor should not be able to predict what will happen to the ciphertext by changing one character in the plaintext
  - One-time pad provides good confusion because one plaintext letter can be transformed to any ciphertext letter at different places in the output
  - A Caesar cipher is not a good example for confusion because an analyst who deduces the transformation of a few letters can also predict the transformation of the remaining letters
  - Substitutions are sometimes done using S-Boxes, which are nothing but table-driven substitutions.
  - Substitution produces confusion
- Diffusion
  - The cipher should spread the information from the plaintext to the entire ciphertext so that changes in the plaintext affect many parts of the ciphertext;
  - The plaintext statistics are disbursed throughout the ciphertext making it difficult to use the frequency analysis of digrams, trigrams, etc and the occurrence of common words in the plaintext.
  - Good diffusion means the interceptor needs much of the ciphertext to infer about the algorithm
  - Transpositions (permutations) produce diffusion.
  - Transpositions are sometimes done through P-boxes