Ch 0. Classical Cryptosystems

이 훈 재 (李 焄 宰)
Hoon-Jae Lee
CNSL
Cryptography and Network Security Lab.
hjlee@dongseo.ac.kr

http://kowon.dongseo.ac.kr/~hjlee
http://crypto.dongseo.ac.kr

Ch 0. Classical Cryptosystems

- Agenda
  - Shift Ciphers
  - Affine Ciphers
  - Vigenere Ciphers
  - Substitution Ciphers
  - Sherlock Holmes
  - Playfair and ADFGX ciphers
  - Block ciphers
  - Binary Numbers and ASCII
  - One-Time Pads
  - Pseudo-random Bit Generation
  - Linear Feedback Shift Register Sequences
  - Enigma

Cryptography - Long History

1900 BC - Non-standard Egyption hierogliphs
1700 BC - Clay of Phaistos: still unrecovered
600 BC - Book of Jeremiah encoded
60 BC - Caesar used encryption
790 AD - First writing (by reference)
1200 AD - Roger Bacon describes methods
1518 AD - First printed book on cryptography
1861 AD - 1st U. S. patent issued
1927 AD - Used during prohibition by criminals
1942 AD – Wartime use (Germany/Japan/US)
1976 AD - Public key Cryptography invented
1977 AD – DES published (FIPS – 46)
RSA published (Rivest, Shamir, Adleman)
2001 AD – AES published (FIPS – 197)

Cryptography: past, present and future

Old days...

Classical Cryptography

- Essentially secret writing (encryption)

Today

Conventional Cryptography (symmetric key)

Public Key cryptography (asymmetric key)

- Besides encryption, there are many other cryptographic algorithms and schemes: e.g. proof of knowledge, digital signature, message authentication, secret sharing, etc.

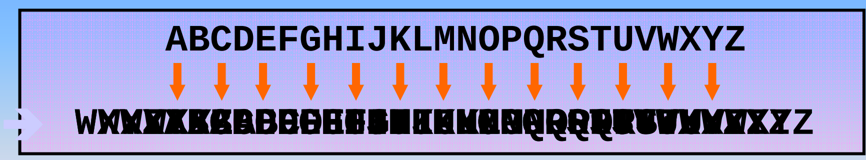
Future...

Quantum Cryptography

Classical Cryptography: Two Main Techniques

1. Shift
2. Substitution
   - Mono-alphabetic substitution
   - Poly-alphabetic substitution

Shift Cipher



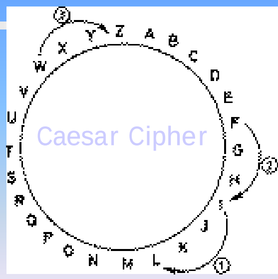
Ciphertext: AIPGSQIXSGMXCYRMZIVWMXCSJLSRKOSRK

Plaintext: [illegible]

Total number of possible shifts = 26

…can you write a computer program to automate this cracking machine?…

Shift Cipher

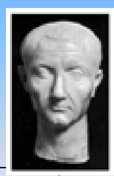


- A shift cipher can also be described as
  Encryption $E_K(x) = x + K \bmod 26$
  Decryption $D_K(x) = x - K \bmod 26$
  for English alphabet by setting up a correspondence between alphabetic characters and residues modulo 26.
  - K is the Key
- When K=3, the shift cipher is also known as Caesar Cipher
  - reputedly used by Julius Caesar (100 – 44 B.C.)

  - Plaintext: I CAME I SAW I CONQUERED
  - Ciphertext: L FDPH L VDZ L FRQTXHUHG

Shift Cipher

❑ Caesar's cipher (monoalphabetic)
(1st century B.C.) ** Flash Demo **



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

$$Y_i = (X_i + Z_i) \bmod 21$$
$$(A = 0, B = 1, \ldots) \quad (Z_i = 3)$$

Message	V	I	N	I	V	I	D	I	V	I	N	C	I
Key	D	D	D	D	D	D	D	D	D	D	D	D	D
Cryptogram	B	M	Q	M	B	M	G	M	B	M	Q	F	M

Security Strength of a Shift Cipher

- Only have 26 possible shifts
- Brute-force Attack : simply try each possible shift in turn
  - a.k.a. exhaustive key search
- Example:

Ciphertext –	**mjaiamwlxsvitpegipixxivw**	
Trial 1	lizhzlvkwruhsodfhohwwhuv	(shift backward by 1)
Trial 2	khygykujvotgrncegngvvgtu	(shift backward by 2)
Trial 3	jgxfxjtiupsfombdfmfuufst	(shift backward by 3)
Plaintext -	**ifwewishtoreplaceletters**	(shift backward by 4)

Hence K=4.

- The major problem of shift ciphers:

  "the key space is too small against brute-force attack"

- The complexity of brute-force attack is O(n).

Vigenere Cipher

- Invented in the 16th century
- A kind of poly-alphabetic substitution cipher
- Using the correspondence A ↔ 0, B ↔ 1, ..., Z ↔ 25.
- Keyword: **CIPHER** ↔ (2, 8, 15, 7, 4, 17)
- Plaintext: **thiscryptosystemisnotsecure**

19	7	8	18	2	17	24	15	19	14	18	24
2	8	15	7	4	17	2	8	15	7	4	17
21	15	23	25	6	8	0	23	8	21	22	15
18	19	4	12	8	18	13	14	19	18	4	2
2	8	15	7	4	17	2	8	15	7	4	17
20	1	19	19	12	9	15	22	8	25	8	19
20	17	4									
2	8	15									
22	25	19									

- Ciphertext: **VPXZGIAXIVWPUBTTMJPWIZITWZT**

Vigenere Cipher

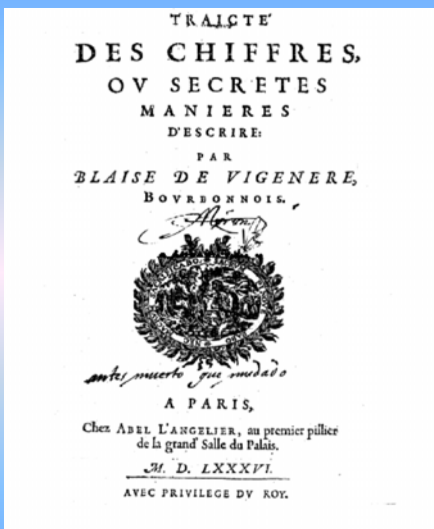
- Vigenère's cipher (polialphabetic) (1586)

  Key: Z_i = L, O, U, P

- Encipherment: $Y_i = (X_i + Z_i) \bmod 26$
- Decipherment: $X_i = (Y_i - Z_i) \bmod 26$

Message	P A R I S	V A U T	B I E N	U N E	M E S S E
Key	L O U P L	O U P L	O U P L	O U P	L O U P L
Cryptogram	A O L X D	J U J E	P C T Y	I H T	X S M H P

Vigenere Cipher


TRAICTE
DES CHIFFRES,
OV SECRETES
MANIERES
D'ESCRIRE:
PAR
BLAISE DE VIGENERE,
BOVRBONNOIS.

A PARIS,
Chez ABEL L'ANGELIER, au premier pillier
de la grand' Salle du Palais.
M. D. LXXXVI.
AVEC PRIVILEGE DV ROY.

Blaise de Vigenère (1523-1596)

Vigenere Cipher

VIGENÈRE’S TABLE (1586)

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

$$P \rightarrow 15$$
$$L \rightarrow 11$$
$$15 + 11 = 0 \bmod 26$$
$$0 - 11 = -11$$
$$-11 + 26 = 15 \rightarrow P$$

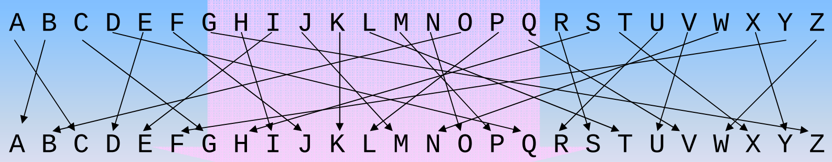
Note that the modulus of a negative value is computed by repeatedly adding the base until a positive value is obtained.

Vigenere Cipher

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

Simple Substitution

EIMBULJIWLNYANJMVLIURAHIWAI



DEPARTMENTOFCOMPUTERSCIENCE

- A key is a *random permutation* of the alphabetic characters.
- E.g.

a	b	c	d	e	f	g	h	i	j	k	l	m
X	N	Y	A	H	P	O	G	Z	Q	W	B	T

n	o	p	q	r	s	t	u	v	w	x	y	z
S	F	L	R	C	V	M	U	E	K	J	D	I

- What's the ciphertext of "**graduate of design and it**"?

Simple Substitution

- Total number of possible permutations

26!

- 26! = 403,291,461,126,605,635,584,000,000 (i.e. 27 digits)
- Maybe… also write a computer program to try them all exhaustively… (so-called **Brute-force Attack**)
- **Calculation**: suppose we have <u>one million</u> 3GHz PCs can try <u>3 billion permutations per second</u>, the machines will take **4263 years** to try all 26! permutations.
- Question: any better cracking algorithm?

Character Frequency Attack (Statistical Attack)

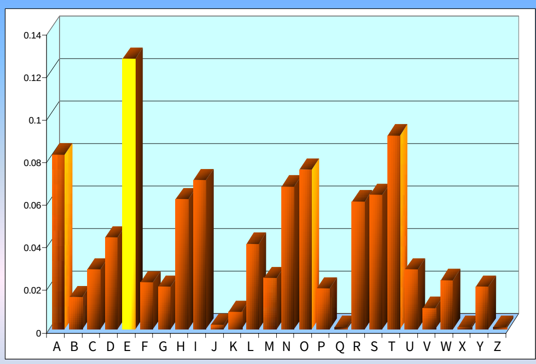
- Brute-force attack against simple substitution becomes less efficient for large alphabet size.
- However, simple substitution does not change relative letter frequencies.

Character Frequency Attack

- in most languages, letters are not equally common
- in English, **e** and **t** are by far the most common letters
- Probability of occurrences of the 26 English letters (obtained by Beker and Piper)

letter	probability	letter	probability
A	.082	N	.067
B	.015	O	.075
C	.028	P	.019
D	.043	Q	.001
E	.127	R	.060
F	.022	S	.063
G	.020	T	.091
H	.061	U	.028
I	.070	V	.010
J	.002	W	.023
K	.008	X	.001
L	.040	Y	.020
M	.024	Z	.001

Character Frequency Attack (Statistical Attack)



- May also be useful to consider sequences of two or three consecutive letters called **digrams** and **trigrams**, respectively.
- e.g. common diagrams (in decreasing order): TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, …
- e.g. common trigrams (in decreasing order): THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, …

Cryptanalysis of The Substitution Cipher

- May also be useful to consider sequences of two or three consecutive letters called **digrams** and **trigrams**, respectively.
- e.g. common diagrams (in decreasing order): TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, …
- e.g. common trigrams (in decreasing order): THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, …

Exercise: Ciphertext obtained from a substitution cipher

```
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

What's the message?

Poly-alphabetic Substitution Ciphers

- In both the Shift Cipher and the Substitution Cipher, once a key is chosen, each alphabetic character is mapped to a unique alphabetic character – monoalphabetic.
  - Small key space (O(n) for Shift Cipher)
  - Vulnerable to statistical attacks (Substitution Cipher).
- Alternative: **Polyalphabetic Substitution Ciphers**
  - We want large key space and less vulnerable to statistical attacks
- Approach: for different location or different time, same letter is to be substituted by different letter.

Sherlock Holmes

- Arthur Conan Doyle: "*The Adventure of the Dancing Men*" - A Sherlock Holmes Adventure.

"*Speaking roughly, T, A, O, I, N, S, H, R, D, and L are the numerical order in which letters occur; but T, A, O, and I are very nearly abreast of each other, and it would be an endless task to try each combination until a meaning was arrived at.*"

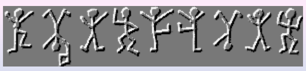
Read the story online and see the images and analysis of the decoding at http://camdenhouse.ignisart.com/canon/danc.htm

Sherlock Holmes

- Dancing Men

「AM HERE ABE SLANEY」
(나는 여기에 왔다. 에이브 슬레이니)

「NEVER」



「A T E L R I G E S」

「ELSIE PREPARE TO MEET THY GOD」
(엘시, 하느님 곁으로 갈 준비를 하라.)

「COME E L S I E 」

COME HERE AT ONCE
(곧 오라)

Monoalphabetic Ciphers - Playfair

Playfair - Charles Wheatstone 1854

Multiple letter encryption mapping two letters into a two cipher letters. Masks the symbol frequency better than simpler ciphers. Used by British in the Boer War, WWI, and to some extent in WWII.

Maps letters into a 5 x 5 matrix (Z is omitted) and follows three rules.

The matrix is populated and both ends know the mapping.

Playfair

Mapping is a spiral starting at lower-right corner.

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

Playfair Rules

Arrange plaintext into pairs. If a double letter (e.g., tt) Insert an X. If an odd number, insert an X pad at the end.

1. If pair is in same row, cipher pair is two letters to the right wrapped to left column (IG = HF; XB = QL).
2. If pair is in same column, cipher pair is below, wrap to top (FQ = SP; UN = VH; FS = SR).
3. If pair is at corners of a rectangle of letters, 1st encrypts to corner of same row, 2nd to corner in its row (EK = IC; UR = SV; AI = ME).

Playfair Example

Plain = ME Rx RI LY WE RO Lx LA LO NG

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

Cipher = AI YQ KF XK BH YP WQ BM XM OH

Product Ciphers – ADFGX cipher

Combination of substitution and transposition - German ADFGX cipher used in WW1 (2 step process).

1: Transpose one plaintext character into a limited set of 2-character symbols (the inner matrix can be changed):

	A	D	F	G	X
A	n	b	x	r	u
D	q	o	k	d	v
F	a	h	s	g	f
G	m	z	c	l	t
X	e	i	p	j	w

Product Ciphers – ADFGX cipher

M: forced to retreat ten km to abbeville few casualties

	A	D	F	G	X
A	n	b	x	r	u
D	q	o	k	d	v
F	a	h	s	g	f
G	m	z	c	l	t
X	e	i	p	j	w

forced becomes:
f = FX; o = DD; r = AG; c = GF; e = XA; d = DG

= FXDDAGGFXADG

Product Ciphers – ADFGX cipher

Step 2 = transposition using a sequence of numbers between 1 & 20 arranged in scrambled order (with order changed as often as needed). Example key (the numbers):

8	9	14	7	19	13	16	1	15	6	3	10	17	2	20	5	11	18	4	12
F	X	D	D	A	G	G	F	X	A	D	G	G	X	D	D	A	G	X	A
G	X	A	G	X	A	F	A	G	X	G	X	X	A	A	A	D	F	G	A
G	X	D	D	F	A	A	D	A	D	X	A	D	X	X	D	G	G	G	G
X	A	F	X	X	A	X	X	G	F	F	A	F	F	A	X	F	A	G	G
G	X	X	D	X	A	F	F												

For: “forced to retreat ten km to abbeville few casualties”

Product Ciphers – ADFGX cipher

Output is taken a column at a time from the transpose matrix in numeric order (i.e., 1,2,3, etc) and blocked in five character groups. For the message on the previous slide (forced to retreat ten km to abbeville few casualties):

FADXF XAXFD GFXFG GGDAD XAXDF DGDXD
FGGXG XXXAX GXAAA DGFAA GGGAA AADAD
FXXGA GGFAX FGXDF GFGAA XFXXD AXA

Not very strong. A Frenchman broke it in 3.5 months. Later the code was changed - took 24 hours to break.

The Hill Cipher

Let $m \geq 2$ be an integer. Let $P = C = (\mathbb{Z}_{26})^m$ and let

$$K = \{m \times m \text{ invertible matrices over } \mathbb{Z}_{26}\}$$

For a key k, we define

$$e_k(p) = pk$$

and

$$d_k(c) = ck^{-1},$$

where all operations are performed in $\mathbb{Z}_{26}$.

- As well as Vigenère, this cipher is polyalphabetic.
- It was invented by Lester S. Hill in 1929.

The Hill Cipher

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

- To encrypt a message using the Hill Cipher one should perform the following sequence of steps:
  - Using the table, plaintext message has to be expressed as sequence of integers in such a way that $p = (p_1, \ldots, p_m)$
  - The key is an $m \times m$ matrix
  - The resulting ciphertext $c = e_k(p) = p \times k$ will be a string $(c_1, \ldots, c_m)$ of length m
- To decrypt the ciphertext one should apply the inverse linear transformation. In other words
  $p = c \times k^{-1}$, where $kk^{-1} \bmod 26 = I$.

The Hill Cipher

Example:

- Encrypting a message using the Hill Cipher is very simple. D ecryption however is more challenging.
- We are given the following ciphertext "CKHUMD", with

$$k = \begin{pmatrix} 11 & 3 \\ 8 & 7 \end{pmatrix}$$

- To decrypt the ciphertext produced with the Hill Cipher, one should apply the inverse linear transformation k^{-1}.

$$k^{-1} = (11\times 7 - 3\times 8)^{-1}\begin{pmatrix} 7 & -3 \\ -8 & 11 \end{pmatrix} = \begin{pmatrix} 7 & 23 \\ 18 & 11 \end{pmatrix},$$

where $(11\times 7 - 3\times 8)^{-1}$ is the reciprocal of the residue of $(11\times 7 - 3\times 8)^{-1}(\mathrm{mod}\,26)$

- After k-1 is found, it is easy to find the corresponding plaint ext, which is "matrix" in our case.

Hill Cipher

- Multiletter cipher
- Takes m successive plaintext letters and substitutes for them m ciphertext letters
- 3x3 Hill cipher:

$$\mathbf{c_1 = (k_{11}p_1 + k_{12}p_2 + k_{13}p_3) \bmod 26}$$
$$\mathbf{c_2 = (k_{21}p_1 + k_{22}p_2 + k_{23}p_3) \bmod 26}$$
$$\mathbf{c_3 = (k_{31}p_1 + k_{32}p_2 + k_{33}p_3) \bmod 26}$$

- $C = E_K(P) = KP$; $P=D_K(C)=K^{-1}C = K^{-1}KP = P$
- m x m Hill cipher hides (m-1)-letter frequency info
- Strong against for the plaintext-only attack, but easily broken with known plaintext attack
  - with m plaintext-ciphertext pairs, each of length m; $K = CP^{-1}$

One-time Pad

- Polyalphabetic substitution with the keyword as long as the plaintext and the keyword has no statistical relationship to the plaintext.

Vernam Cipher (1918)

- Works on binary string rather than letters
- $C_i = P_i \oplus K_i$

  where K_i is chosen randomly

- The only cryptographic system that can be proved to be unconditionally secure
- Can be shown to be information theoretically secure.

Pseudorandom Number Generators (PRNGs)

- Algorithmic technique to create "rand om numbers"
  - Although not truly random, can pass ma ny tests of "randomness"

Real Random Test

- Coin-Tossing ~Good for Randomness Experiment

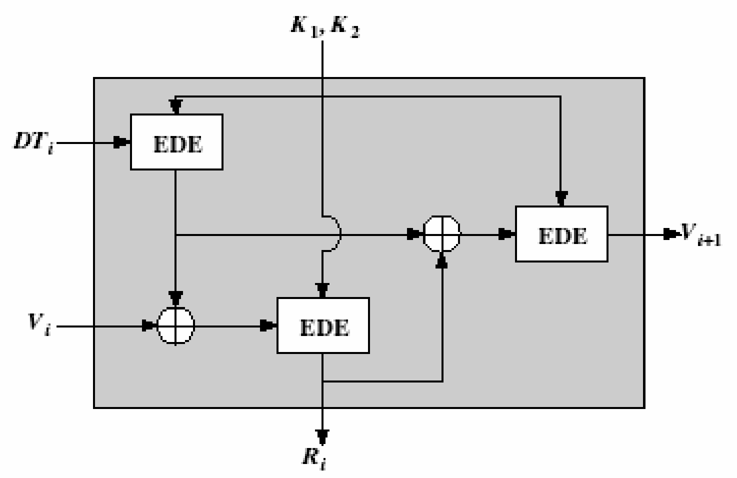
N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A																				
B																				

- Five Basic Test Criterion (Menezes)
  - Frequency Test : 0, 1 distribution (~ideal=N/2)
  - Serial Test : 0/0, 0/1, 1/0, 1/1 distribution (~ideal=N/4)
  - Poker Test : 00, 01, 10, 11 distribution (~ideal=N/8)
  - Run Test : 0'run, 1'run, 00'run, 11'run, 000'run, 111'run..
    (~ideal : short run > long run)
  - Correlation Test : A(0), A(1), A(2),....,A(20)(~ideal=peak)

Linear Congruential Generator

- Common iterative technique using:
  $X_{n+1} = (aX_n + c) \bmod m$
- Given suitable values of parameters can pr oduce a long random-like sequence
- Suitable criteria to have are [PARK88]:
  - T_1: function generates a full-period
  - T_2: generated sequence should appear random
  - T_3: efficient implementation with 32-bit arithmet ic
- Note that an attacker can reconstruct sequ ence given a small number of values

ANSI X9.17 PRNG



$R_i = \mathrm{EDE}_{k_1.k_2}[v_i \oplus \mathrm{EDE}_{k_1.k_2}[\mathrm{DT}_i]]$

$v_{i+1} = \mathrm{EDE}_{k_1.k_2}[R_i \oplus \mathrm{EDE}_{k_1.k_2}[\mathrm{DT}_i]]$

Blum Blum Shub (BBS) Generator

- Based on public key algorithms
- Use least significant bit from iterative equation:
  - $X_0 = s^2$ mod n
  - For i=1 to ∞
  - $X_i=(X_{i-1})^2$ mod n
  - $B_i=X_i$ mod 2
    - ✓ where n=p.q, and primes p,q=3 mod 4
- Unpredictable, passes **next-bit** test
- Security rests on difficulty of factoring N
- Is unpredictable given any run of bits
- Slow, since very large numbers must be used
- Too slow for cipher use, good for key generation

Example of BBS

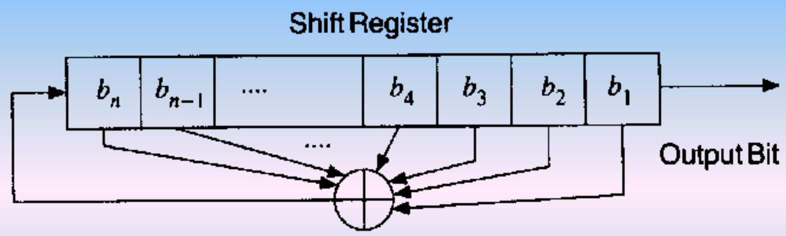
s	X_i	B_i
0	20749	
1	143135	1
2	177671	1
3	97048	0
4	89992	0
5	174051	1
6	80649	1
7	45663	1
8	69442	0
9	186894	0
10	177046	0

s	X_i	B_i
11	137922	0
12	123175	1
13	8630	0
14	114386	0
15	14863	1
16	133015	1
17	106065	1
18	45870	0
19	137171	1
20	48060	0

Linear feedback shift registers(LFSR) -1

- Characteristic
  - well-suited to hardware implementation.
  - can produce sequences of large period.
  - can produce sequences with good statistical properties
  - can be readily analyzed using algebraic techniques

Linear feedback shift registers(LFSR) -2



- feedback shift register is consisted of:
  - shift register
  - feedback function

Linear feedback shift registers(LFSR) -3

- LFSR is denoted *<L,C(D)>*
- *connection polynomial* :
  $C(D)=1+c_1D+c_2D^2+\ldots+c_LD^L \in Z_2[D]$
- *nonsingular* : degree of C(D) = L
- *singular* : degree of C(D) < L
- *initial state*
- If C(D) is primitive polynomial of degree L, then *<L,C(D)>* is called a *maximum-length LFSR*
- *m-sequence* : the output of a maximum-length LFSR with non-zero initial state

Rotor Machines

- In the beginning of twentieth century mechanical e ncryption devices started to be developed, in orde r to automate encryption/decryption process.
- Rotor machines were using a substitution cipher, which was rotated each cycle. Actually this idea w as already used during the American Civil War.
- The best well known rotor machine is Enigma.

Rotor Machines - Enigma

- Enigma used three rotors chosen from a set of fiv e. The three rotors were interconnected, so first ro tor would turn the second each full iteration, and s econd would turn the third.
- A number of additional mechanisms were used to make the cipher more secure.

- However incorrect usage of the device allowed Alli es to break the code.

Enigma

- Suppose we take a conversion for the first letter of the message and a different mapping for the next letter and a different mapping for the next letter …
- That is what we did with Vigenere
- Add additional encodings. Rotate from a fixed starting point through 26 positions of the first set of columns, then iterate a second set of columns. Now have 676 different mappings.
- To decode, must figure out the wiring inside each phase, and the order in which they are arranged in the machine.

Enigma

- German engineer, Artur Scherbius (1878-1929) invented a machine of this type around 1918 and bought the patent rights to one invented in Holland also. He added a reflecting cylinder, which allowed the same machine to encode and decode. He called the machine enigma, from the Greek for riddle.
- The enigma used by the Germans in WWII had three rotors, and later four.

Enigma - 2

Cryptography and Network Security
-W. Stalling (4th Ed.)

Hoon-Jae Lee
CNSL
Cryptography and Network Security Lab.
hjlee@dongseo.ac.kr

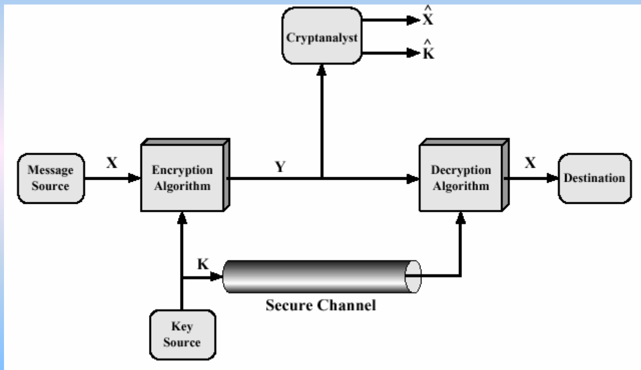
http://kowon.dongseo.ac.kr/~hjlee
http://crypto.dongseo.ac.kr

Chapter 2 – Classical Encryption Techniques

Many savages at the present day regard their names as vital parts of themselves, and therefore take great pains to conceal their real names, lest these should give to evil-disposed persons a handle by which to injure their owners.
—***The Golden Bough,*** Sir James George Frazer

Conventional Encryption Model

- Plaintext & Ciphertext
- Model of Conventional Cryptosystem



Basic Terminology

- **plaintext** - the original message
- **ciphertext** - the coded message
- **cipher** - algorithm for transforming plaintext to ciphertext
- **key** - info used in cipher known only to sender/receiver
- **encipher (encrypt)** - converting plaintext to ciphertext
- **decipher (decrypt)** - recovering ciphertext from plaintext
- **cryptography** - study of encryption principles/methods
- **cryptanalysis (codebreaking)** - the study of principles/ methods of deciphering ciphertext *without* knowing key
- **cryptology** - the field of both cryptography and cryptanalysis

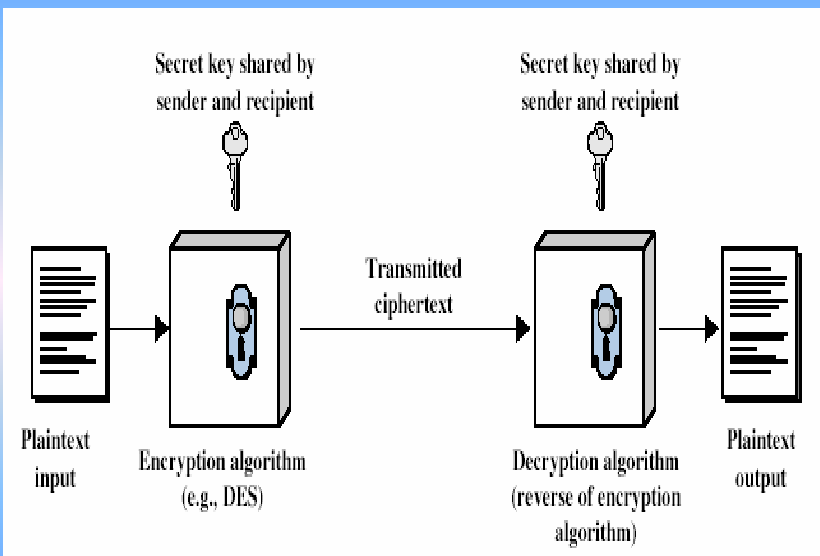
Cryptography

- Cryptographic systems
  - The type of operations used for transforming plaintext to ciphertext
    - Substitution
    - Transposition
    - Product
  - The number of keys
    - symmetric, single-key, secret-key, conventional encryption
    - asymmetric, two-key, public-key encryption
  - The way in which the plaintext is processed
    - Block cipher
    - Stream cipher

Symmetric Encryption

- or conventional / private-key / single-key
- sender and recipient share a common key
- all classical encryption algorithms are private-key
- was only type prior to invention of public-key in 1970's

Symmetric Cipher Model



Requirements

- two requirements for secure use of symmetric encryption:
  - a strong encryption algorithm
  - a secret key known only to sender / receiver

    $Y = E_K(X)$

    $X = D_K(Y)$
- assume encryption algorithm is known
- implies a secure channel to distribute key

Cryptanalysis

- Process of attempting to discover plaintext and key
- Various type of cryptanalytic attacks
  - Ciphertext only attack
  - Known plaintext attack
  - Chosen plaintext attack
  - Chosen ciphertext attack
  - Chosen text attack
- Types of security
  - unconditionally secure
  - computationally secure

Types of Cryptanalytic Attacks

- ❑ ciphertext only
  - ➢ only know algorithm / ciphertext, statistical, can identify plaintext
- ❑ known plaintext
  - ➢ know/suspect plaintext & ciphertext to attack cipher
- ❑ chosen plaintext
  - ➢ select plaintext and obtain ciphertext to attack cipher
- ❑ chosen ciphertext
  - ➢ select ciphertext and obtain plaintext to attack cipher
- ❑ chosen text
  - ➢ select either plaintext or ciphertext to en/decrypt to attack cipher

Brute Force Search

- ❑ always possible to simply try every key
- ❑ most basic attack, proportional to key size
- ❑ assume either know / recognise plaintext

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

More Definitions

- unconditional security
  - no matter how much computer power is available, the cipher cannot be broken since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext
- computational security
  - given limited computing resources (eg time needed for calculations is greater than age of universe), the cipher cannot be broken

Steganography

- an alternative to encryption
- hides existence of message
  - using only a subset of letters/words in a lon ger message marked in some way
  - using invisible ink
  - hiding in LSB in graphic image or sound file
- has drawbacks
  - high overhead to hide relatively few info bits

Steganography

- Cryptography vs. Steganography
- Example
  - Character marking, Invisible ink, Pin punctures, Typewriter correction ribbon
- Drawbacks
  - A lot of overhead,
  - worthless for being discovered
- First encryption and steganography

Classical Encryption Techniques

- Substitution Techniques
  - Caesar Cipher
  - Monoalphabetic Cipher
  - Playfair Cipher
  - Hill Cipher
  - Polyalphabetic Cipher
- Transposition Techniques
  - Rotor Machines

Classical Substitution Ciphers

- where letters of plaintext are replaced by other letters or by numbers or symbols
- or if plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns

Caesar Cipher

- earliest known substitution cipher
- by Julius Caesar
- first attested use in military affairs
- replaces each letter by 3rd letter on
- example:

  meet me after the toga party

  PHHW PH DIWHU WKH WRJD SDUWB

Substitution Ciphers(1) - Caesar cipher

- Simple-alphabetic substitution ciphers;
  - Letters of plaintext ⇒ Other letter or by numbers or symbols
  - Caesar cipher
    - ✓ Plain : a b c d … … x y z
    - ✓ Cipher : D E F G … … A B C
    - ✓ Encryption algorithm
      C=E(p)=(p+k) mod 26, k=1,2,…,25
    - ✓ Decryption algorithm
      p=D(C)=(C-k) mod 26
    - ✓ Plaintext: **meet me after the toga party**
    - ✓ Ciphertext:**PHHW PH DIWHU WKH WRJD SDUWB (k=3)**

Caesar Cipher

- can define transformation as:

```
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
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
```

- mathematically give each letter a number

```
a b c d e f g h i j k  l  m
0 1 2 3 4 5 6 7 8 9 10 11 12
n  o  p  q  r  s  t  u  v  w  x  y  Z
13 14 15 16 17 18 19 20 21 22 23 24 25
```

- then have Caesar cipher as:

$C = \mathrm{E}(p) = (p + k) \bmod (26)$

$p = \mathrm{D}(C) = (C - k) \bmod (26)$

Cryptanalysis of Caesar Cipher

- only have 26 possible ciphers
  - A maps to A,B,..Z
- could simply try each in turn
- a **brute force search**
- given ciphertext, just try all shifts of letters
- do need to recognize when have plaintext
- eg. break ciphertext "GCUA VQ DTGCM"

Monoalphabetic Cipher

- rather than just shifting the alphabet
- could shuffle (jumble) the letters arbitrarily
- each plaintext letter maps to a different random ciphertext letter
- hence key is 26 letters long

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

- Example

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

Monoalphabetic Cipher Security

- now have a total of 26! = 4 x 10^{26} keys
- with so many keys, might think is secure
- but would be **!!!WRONG!!!**
- problem is language characteristics

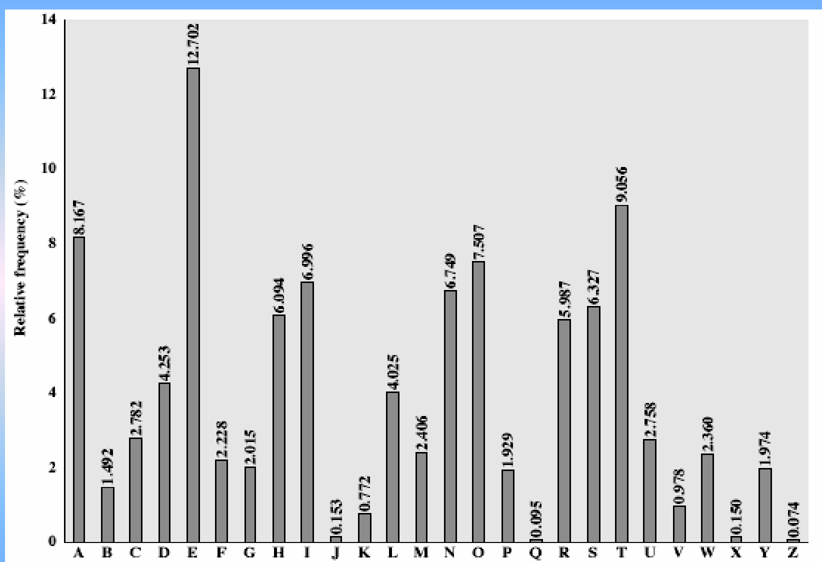
Substitution Ciphers(2)

- Monoalphabetic substitution ciphers;
  - Caesar Cipher is not secure for 25 keys
  - Any permutation assignment→ 26!=4*10^{26}
    - plaintext : a b c d e f g … z
    - ciphertext: k z g w j s m … n
  - Attack: frequency attack
    - "e" possibility-12.75%, t-9.25% r-8.5%,n/I-7.75% …
    - j/z-0.25

Language Redundancy and Cryptanalysis

- human languages are **redundant**
- eg "th lrd s m shphrd shll nt wnt"
- letters are not equally commonly used
- in English **e** is by far the most common letter
- then T,R,N,I,O,A,S
- other letters are fairly rare
- cf. Z,J,K,Q,X
- have tables of single, double & triple letter frequencies

English Letter Frequencies



Use in Cryptanalysis

- key concept - monoalphabetic substitution ciphers do not change relative letter frequencies
- discovered by Arabian scientists in 9th century
- calculate letter frequencies for ciphertext
- compare counts/plots against known values
- if Caesar cipher look for common peaks/troughs
  - peaks at: A-E-I triple, NO pair, RST triple
  - troughs at: JK, X-Z
- for monoalphabetic must identify each letter
  - tables of common double/triple letters help

Example Cryptanalysis

- given ciphertext:

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ
```

- count relative letter frequencies (see text)
- guess P & Z are e and t
- guess ZW is th and hence ZWP is the
- proceeding with trial and error fially get:

```
it was disclosed yesterday that several informal but
direct contacts have been made with political
representatives of the viet cong in moscow
```

Playfair Cipher

- not even the large number of keys in a monoalphabetic cipher provides security
- one approach to improving security was to encrypt multiple letters
- the **Playfair Cipher** is an example
- invented by Charles Wheatstone in 1854, but named after his friend Baron Playfair

Playfair Key Matrix

- a 5X5 matrix of letters based on a keyword
- fill in letters of keyword (sans duplicates)
- fill rest of matrix with other letters
- eg. using the keyword MONARCHY

```
MONAR
CHYBD
EFGIK
LPQST
UVWXZ
```

Substitution Techniques – Playfair Cipher

Column

row

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

(keyword : monarchy)

1. balloon => ba lx lo on
2. AR => RM
3. MU => CM
4. HS => BP , EA => IM/JM

- Frequency analysis much more difficult
- it still leaves much of the structure of the plaintext language.

Encrypting and Decrypting

- plaintext encrypted two letters at a time:
  1. if a pair is a repeated letter, insert a filler like 'X', eg. "balloon" encrypts as "ba lx lo on"
  2. if both letters fall in the same row, replace each with letter to right (wrapping back to start from end), eg. “ar" encrypts as "RM"
  3. if both letters fall in the same column, replace each with the letter below it (again wrapping to top from bottom), eg. “mu" encrypts to "CM"
  4. otherwise each letter is replaced by the one in its row in the column of the other letter of the pair, eg. “hs" encrypts to "BP", and “ea" to "IM" or "JM" (as desired)

Security of the Playfair Cipher

- security much improved over monoalphabetic
- since have 26 x 26 = 676 digrams
- would need a 676 entry frequency table to analyse (verses 26 for a monoalphabetic)
- and correspondingly more ciphertext
- was widely used for many years (eg. US & British military in WW1)
- it can be broken, given a few hundred letters
- since still has much of plaintext structure

Substitution Techniques – Hill Cipher(1)

- Hill Cipher
  - Hill cipher takes m successive plaintext letters and substitutes for them m ciphertext letters.
  - For m = 3, Hill cipher can be described as follows

    $C_1 = (K_{11}P_1 + K_{12}P_2 + K_{13}P_3) \bmod 26$

    $C_2 = (K_{21}P_1 + K_{22}P_2 + K_{23}P_3) \bmod 26$

    $C_3 = (K_{31}P_1 + K_{32}P_2 + K_{33}P_3) \bmod 26$
  - This can be expressed in term of vectors and matrices :

$$\begin{pmatrix} C1 \\ C2 \\ C3 \end{pmatrix} = \begin{pmatrix} K11 & k12 & k13 \\ K21 & k22 & k23 \\ K31 & k32 & k33 \end{pmatrix} \begin{pmatrix} P1 \\ P2 \\ P3 \end{pmatrix}$$

Substitution Techniques – Hill Cipher(2)

- Decryption: requires the inverse of the matrix K

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix}$$

  - $C = E_k(P) = KP$
  - $P = Dk(C) = K^{-1}C = K^{-1}KP = P$

- Ciphertext-only attack → strong,
- known plaintext attack → weak

Polyalphabetic Ciphers

- another approach to improving security is to use multiple cipher alphabets
- called **polyalphabetic substitution ciphers**
- makes cryptanalysis harder with more alphabets to guess and flatter frequency distribution
- use a key to select which alphabet is used for each letter of the message
- use each alphabet in turn
- repeat from start after end of key is reached

Vigenère Cipher

- simplest polyalphabetic substitution cipher is the **Vigenère Cipher**
- effectively multiple caesar ciphers
- key is multiple letters long K = k1 k2 ... kd
- i^{th} letter specifies i^{th} alphabet to use
- use each alphabet in turn
- repeat from start after d letters in message
- decryption simply works in reverse

Example

- write the plaintext out
- write the keyword repeated above it
- use each key letter as a caesar cipher key
- encrypt the corresponding plaintext letter
- eg using keyword *deceptive*

```
key:       deceptivedeceptivedeceptive
plaintext: wearediscoveredsaveyourself
ciphertext:ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Aids

- simple aids can assist with en/decryption
- a **Saint-Cyr Slide** is a simple manual aid
  - a slide with repeated alphabet
  - line up plaintext 'A' with key letter, eg 'C'
  - then read off any mapping for key letter
- can bend round into a **cipher disk**
- or expand into a **Vigenère Tableau** (see text Table 2.3)

Security of Vigenère Ciphers

- have multiple ciphertext letters for ea ch plaintext letter
- hence letter frequencies are obscure d
- but not totally lost
- start with letter frequencies
  - see if look monoalphabetic or not
- if not, then need to determine numbe r of alphabets, since then can attach each

Substitution Techniques – Polyalphabetic Cipher(1)

- To improve on the simple monoalphabetic tech.
  - Different monoalphabetic substitution as one proceeds through the plaintext message
- Vigenere cipher
  - The 26 Caesar ciphers, with shifts of 0 through 25
  - Key letter

Substitution Techniques – Polyalphabetic Cipher(2)

- The Modern Vigenere Tableau

	ABCDEFGHIJKLMNOPQRSTUVWXYZ
A	ABCDEFGHIJKLMNOPQRSTUVWXYZ
B	BCDEFGHIJKLMNOPQRSTUVWXYZA
C	CDEFGHIJKLMNOPQRSTUVWXYZAB
D	DEFGHIJKLMNOPQRSTUVWXYZABC
E	EFGHIJKLMNOPQRSTUVWXYZABCD
F	FGHIJKLMNOPQRSTUVWXYZABCDE
G	GHIJKLMNOPQRSTUVWXYZABCDEF
H	HIJKLMNOPQRSTUVWXYZABCDEFG
I	IJKLMNOPQRSTUVWXYZABCDEFGH
J	JKLMNOPQRSTUVWXYZABCDEFGHI
K	KLMNOPQRSTUVWXYZABCDEFGHIJ
L	LMNOPQRSTUVWXYZABCDEFGHIJK
M	MNOPQRSTUVWXYZABCDEFGHIJKL
N	NOPQRSTUVWXYZABCDEFGHIJKLM
O	OPQRSTUVWXYZABCDEFGHIJKLMN
P	PQRSTUVWXYZABCDEFGHIJKLMNO
Q	QRSTUVWXYZABCDEFGHIJKLMNOP
R	RSTUVWXYZABCDEFGHIJKLMNOPQ
S	STUVWXYZABCDEFGHIJKLMNOPQR
T	TUVWXYZABCDEFGHIJKLMNOPQRS
U	UVWXYZABCDEFGHIJKLMNOPQRST
V	VWXYZABCDEFGHIJKLMNOPQRSTU
W	WXYZABCDEFGHIJKLMNOPQRSTUV
X	XYZABCDEFGHIJKLMNOPQRSTUVW
Y	YZABCDEFGHIJKLMNOPQRSTUVWX
Z	ZABCDEFGHIJKLMNOPQRSTUVWXY

Substitution Techniques – Polyalphabetic Cipher(3)

- Example

Key:	**deceptive**deceptivedeceptive
Plaintext:	wearediscoveredsaveyourself
Ciphertext:	zicvtwqngrzgvtwavzhcqyglmgj

- Strength
  - Multiple ciphertext letters for each plaintext letter
- Breaking this cipher
  - Monoalphabetic substitution or Vigenere cipher
    - ✓ Monoalphabetic : statistical properties of the ciphertext
    - ✓ Vigenere : determining the length of the keyword

Substitution Techniques – Polyalphabetic Cipher(4)

- Autokey system
  - Vigenere proposed

Key:	deceptivewearediscoveredsav
Plaintext:	wearediscoveredsaveyourself
Ciphertext:	zicvtwqngkzeiigasxstslvvwla

- Vernam cipher → **Perfect secrecy**
  - Gilbert Vernam(1918)
  - Length (Plaintext = Keyword)
  - Rejection of the statistical relationship
  - $C_i = p_i \oplus k_i$
  - $p_i = C_i \oplus k_i$
  - Running loop of tape
  - One-time pad – Joseph Mauborgne

Kasiski Method

- method developed by Babbage / Kasiski
- repetitions in ciphertext give clues to period
- so find same plaintext an exact period apart
- which results in the same ciphertext
- of course, could also be random fluke
- eg repeated "VTW" in previous example
- suggests size of 3 or 9
- then attack each monoalphabetic cipher individually using same techniques as before

Autokey Cipher

- ideally want a key as long as the message
- Vigenère proposed the **autokey** cipher
- with keyword is prefixed to message as key
- knowing keyword can recover the first few letters
- use these in turn on the rest of the message
- but still have frequency characteristics to attack
- eg. given key *deceptive*

```
key:       deceptivewearediscoveredsav
plaintext: wearediscoveredsaveyourself
ciphertext:ZICVTWQNGKZEIIGASXSTSLVVWLA
```

One-Time Pad

- if a truly random key as long as the message is used, the cipher will be secure
- called a One-Time pad
- is unbreakable since ciphertext bears no statistical relationship to the plaintext
- since for **any plaintext** & **any ciphertext** there exists a key mapping one to other
- can only use the key **once** though
- have problem of safe distribution of key

Transposition Ciphers

- now consider classical **transposition** or **permutation** ciphers
- these hide the message by rearranging the letter order
- without altering the actual letters used
- can recognise these since have the same frequency distribution as the original text

Row Transposition Ciphers

- a more complex scheme
- write letters of message out in rows over a specified number of columns
- then reorder the columns according to some key before reading off the rows

```
Key:        4 3 1 2 5 6 7
Plaintext:  a t t a c k p
            o s t p o n e
            d u n t i l t
            w o a m x y z
Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ
```


Transposition Techniques(1)

- Permutation on the plaintext letters
- Rail fence technique
  - Written down as a sequence of diagonals and the read off as a sequence of rows
  - Ex) meet me after the toga party

m		e		m		a		t		r		h		t		g		p		r		y					
	e		t		e		f		e		t		e		o		a		a		t						

  - MEMATRHTGPRYETEFETEOAAT

Transposition Techniques(2)

- more complex scheme
  - Message: row by row ➔ column by column
  - Key = column ordering
  - Example

Key :	4	3	1	2	5	6	7
Plaintext:	a	t	t	a	c	k	p
	o	s	t	p	o	n	e
	d	u	n	t	i	l	t
	w	o	a	m	x	y	z

Ciphertext :
TTNAAPTMTSUOAODWCOIXKNLYPETZ

Rail Fence cipher

- write message letters out diagonally over a number of rows
- then read off cipher row by row
- eg. write message out as:

```
m e m a t r h t g p r y
 e t e f e t e o a a t
```

- giving ciphertext
  MEMATRHTGPRYETEFETEOAAT

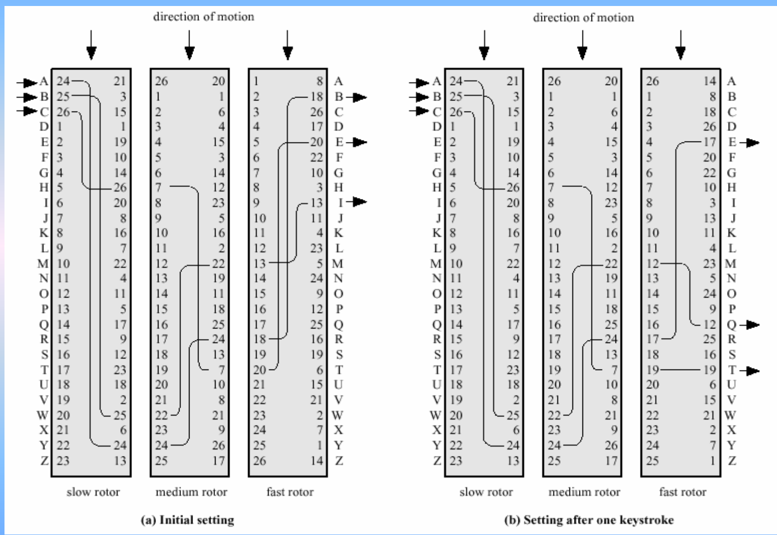
Product Ciphers

- ciphers using substitutions or transpositions are not secure because of language characteristics
- hence consider using several ciphers in succession to make harder, but:
  - two substitutions make a more complex substitution
  - two transpositions make more complex transposition
  - but a substitution followed by a transposition makes a new much harder cipher
- this is bridge from classical to modern ciphers

Rotor Machines

- before modern ciphers, rotor machines wer e most common product cipher
- were widely used in WW2
  - German Enigma, Allied Hagelin, Japanese Purpl e
- implemented a very complex, varying subst itution cipher
- used a series of cylinders, each giving one substitution, which rotated and changed af ter each letter was encrypted
- with 3 cylinders have 26^3=17576 alphabets

Transposition Techniques - Rotor Machine(1)



Transposition Techniques - Rotor Machine(2)

- A set of independently rotating cylinders
  - Three Cylinders : 26 X 26 X 26 = 17,576
  - Four Cylinders : 17,576 X 26 = 456, 976
  - Five Cylinders : 456,976 X 26 = 11,881,376

Rotor Machine(3)

My pictures at Cryptography Co., USA, 2002

2007-10-02 CNSL-Internet-DongseoUniv. 105

Rotor Machine(4)

My pictures at Cryptography Co., USA, 2002

2007-10-02 CNSL-Internet-DongseoUniv. 106

Summary

- have considered:
  - classical cipher techniques and terminology
  - monoalphabetic substitution ciphers
  - cryptanalysis using letter frequencies
  - Playfair ciphers
  - polyalphabetic ciphers
  - transposition ciphers
  - product ciphers and rotor machines
  - stenography