Cryptography CS 555

Lecture 2: Basic Ciphers

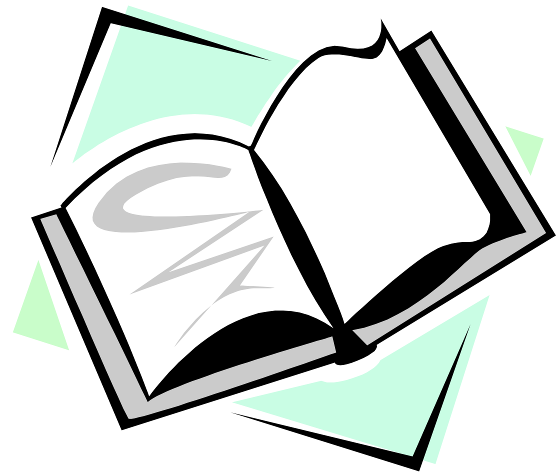
Department of Computer Sciences
Purdue University

Lecture Outline

- Shift and substitution ciphers.
- Attacks on shift and substitution ciphers.
- Vigenere cipher.
- Attacks on Vigenere: Kasisky Test and Index of Coincidence
- Cipher machines: Jefferson Wheel and Enigma machine.

Recommended Reading

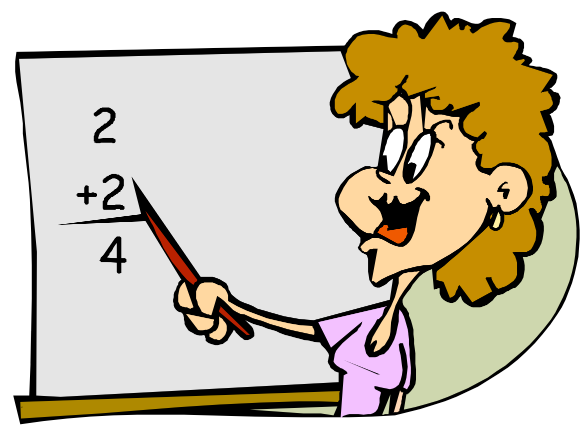
- Chapter 1 from Stinson



Ciphers

- **Substitution ciphers**: same alphabet used for encryption/decryption, a fix permutation of that alphabet defines the rule.
- **Transposition ciphers**: letters in the ciphertext are the same letters, with the same frequency as in the plaintext, but rearranged (using matrices).
- **Product ciphers**: composition of several ciphers, alternating between substitution and transposition.

Begin Math



Cartesian Product

Definition:

Given two sets *A* and *B*, the **Cartesian product** (or *direct product*) of the two sets, written as *A* x *B* is the set of all ordered pairs with the first element of each pair selected from *A* and the second element selected from *B*.

$$A \times B = \{ (a,b) \mid a \in A \text{ and } b \in B\}$$

Example:

A = {1, 2}, B = {a, b, c}.

A x B = {(1, a), (2, a), (1, b), (2, b), (1, c), (2, c)}

Binary Operation

Definition:

Given a set A, a **binary operation**, *, defined on A, is a function from the Cartesian product *A* x *A* to *B*. If B = A, i.e. * takes values in the same set *A*, it is said that the operation is closed on A .

Example:

For a, b Z, define $a * b = a + b$.

For a, b Z, define $a * b = ab$.

For a, b Z, define $a * b = \min\{a, b\}$.

For a, b Z, define $a * b = a/b$.

Modulo Operation

Definition:

$$a \bmod n = r \iff \exists q, \text{ s.t. } a = q \cdot n + r$$

$$\text{where } 0 \le r \le n - 1$$

Example:

7 mod 3 = 1

-7 mod 3 = 2

Definition (Congruence):

$$a \equiv b \bmod n \iff a \bmod n = b \bmod n$$

Groups

Definition:

A group $(G, *)$ is a set G on which a binary operation is defined which satisfies the following axioms:

- Closure: For all $a, b \in G$, $a * b \in G$.
- Associative: For all $a, b, c \in G$, $(a * b) * c = a * (b * c)$.
- Identity: $\exists\, e \in G$ s.t. for all $a \in G$, $a * e = a = e * a$.
- Inverse: For all $a \in G$, $\exists\, a^{-1} \in G$ s. t. $a * a^{-1} = a^{-1} * a = e$.

Example:

$(Z, +)$

$(Z_n$, addition modulo) where $Z_n = \{0, 1, \ldots, n - 1\}$

Groups

Definition:

A group $(G, *)$ is called an abelian group if $*$ is a commutative operation:

Commutative: For all $a, b \in G$, $a * b = b * a$.

Example:

$(R, +)$

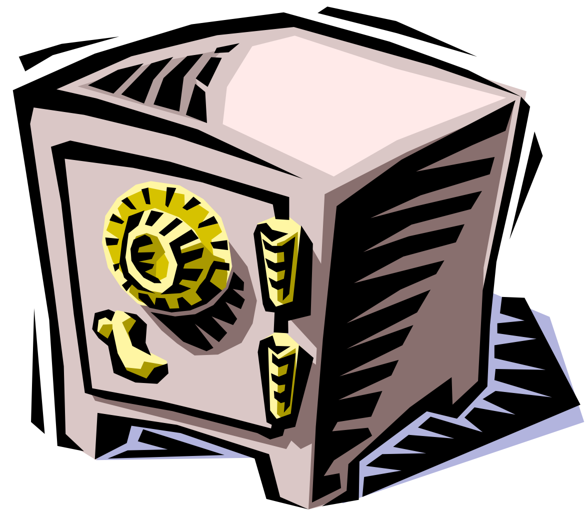
End Math

Shift Cipher

- defined over Z_{26} as follows:

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

- Convert each letter in the plaintext P to it's corresponding number.
- Key K, $0 \le K \le 25$.
- Let $P = C = Z_{26}$
- $e_k(P) = (P + K) \bmod 26$
- $d_k(C) = (C – K) \bmod 26$



Shift Cipher: An Example

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 = CRYPTOGRAPHYISFUN

K = 11

C = NCJAVZRCLASJTDQFY

C □ 2; 2+11 mod 26 = 13 □ N

R □ 17; 17+11 mod 26 = 2 □ C

…

N □ 13; 13+11 mod 26 = 24 □ Y

Shift Cipher: Cryptanalysis

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

- $e_k(P) = (P + K) \bmod 26$
- Can an attacker find K? YES: exhaustive search, key space is small (26 possible keys).
- Once K is found, very easy to decrypt

$$d_k(C) = (C - K) \bmod 26$$

- History: K = 3, Caesar's cipher

Substitution 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

$\pi(A)$ ... $\pi(Z)$

- Ciphertext, Plaintext, $\in Z_{26}$
- e_{π}(Plaintext) = π (Plaintext)
- d_{π}(Ciphertext) = π^{-1} (Ciphertext)

Example:

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

π = **BADC**EFGHIJKLMNOPQRSTUVWXYZ

BECAUSE $\rightarrow$ AEDBUSE

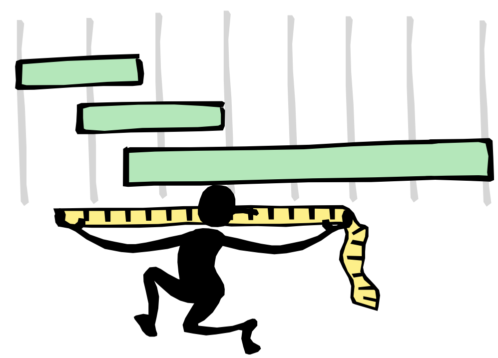
Substitution Ciphers: Cryptanalysis

- Each language has certain features: frequency of letters, or of groups of two or more letters.
- Substitution ciphers preserve the language features.
- Substitution ciphers are vulnerable to frequency analysis attacks.

Example: English Features

- The nine high-frequency letters E, T, A, O, N, I, R, S, and H constitute 70% of plaintext.
- EN is the most common two-letter combination, followed by RE, ER, and NT.
- Vowels, which constitute 40 % of plaintext, are often separated by consonants.
- The letter A is often found in the beginning of a word or second from last. The letter I is often third from the end of a word.
- And more …

Substitution Ciphers: Cryptanalysis



- The number of different ciphertext characters or combinations are counted to determine the frequency of usage.
- The cipher text is examined for patterns, repeated series, and common combinations.
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics.

Vigenere Cipher

Definition:

Given m, a positive integer, $P = C = (Z_{26})^m$, and $K = (k_1, k_2, \dots, k_m)$ a key, we define:

Encryption:

$e_k(p_1, p_2 \dots p_m) = (p_1+k_1, p_2+k_2 \dots p_m+k_m) \pmod{26}$

Decryption:

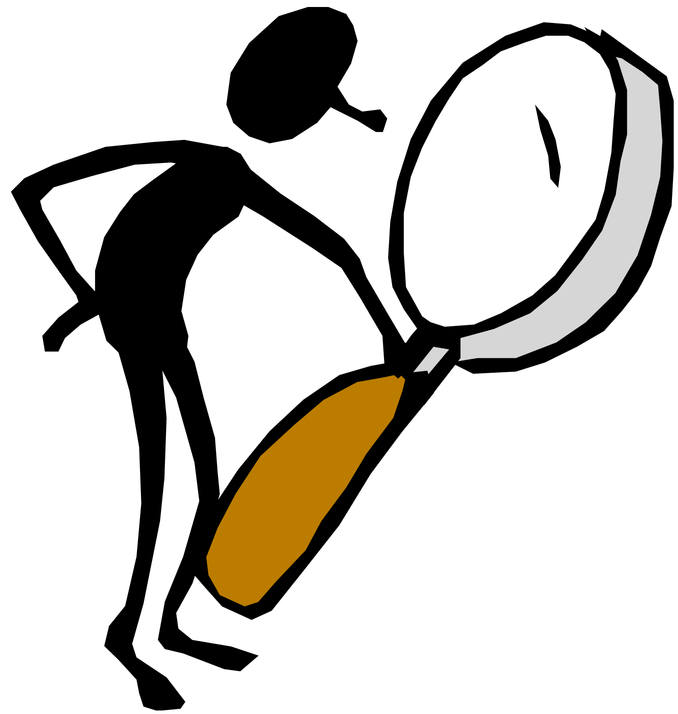
$d_k(c_1, c_2 \dots c_m) = (c_1-k_1, c_2-k_2 \dots c_m- k_m) \pmod{26}$

Example:

Plaintext: C R Y P T O G R A P H Y
Key: L U C K L U C K L U C K
Ciphertext: N L A Z E I I B L J J I

Vigenere Cipher: Cryptanalysis

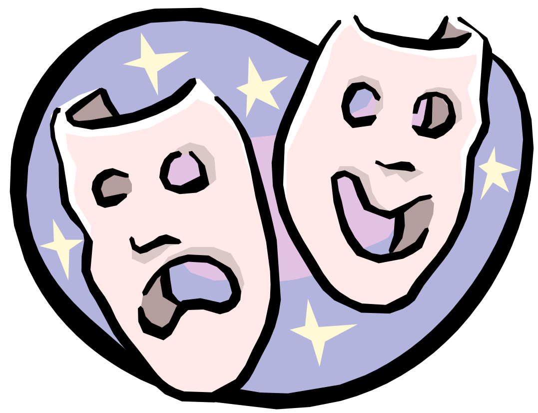
- Frequency analysis: for a given language, map frequency of letters to the known frequencies in that language.
- Substitution ciphers are vulnerable to frequency analysis because they preserve the features of the language.
- What about Vigenere?



Vigenere Cipher

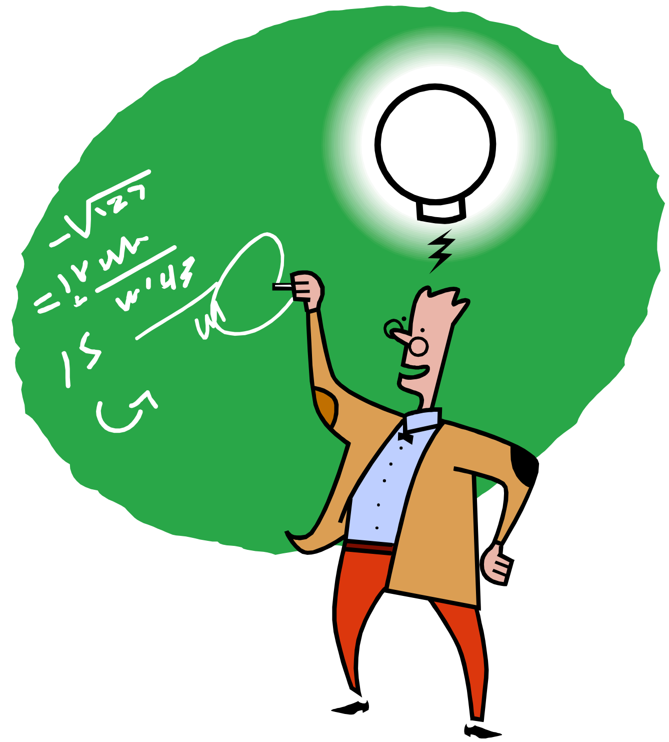
- Vigenere masks the frequency with which a character appears in a language: one particular letter is mapped to more than one letter. Makes the use of frequency analysis more difficult.

Plaintext:	C	R	Y	P	T	O	G	R	A	P	H	Y
Key:	L	U	C	K	L	U	C	K	L	U	C	K
Ciphertext:	N	L	A	Z	E	I	I	B	L	J	J	I



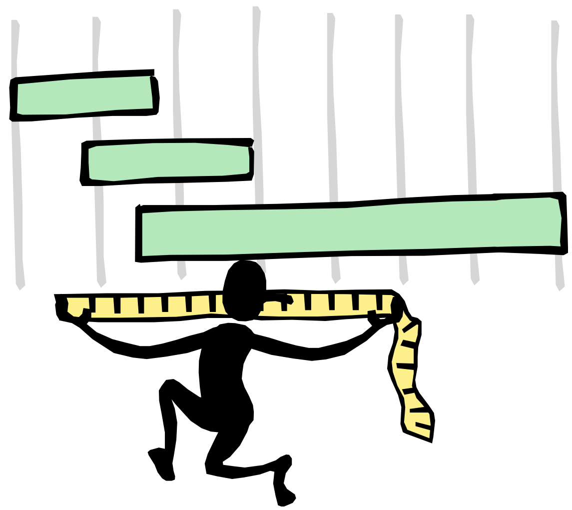
Vigenere Cipher: Cryptanalysis

- Any message encrypted by a Vigenere cipher is a collection of as many simple substitution ciphers as there are letters in the key. So…
  - Find the length of the key.
  - Divide the message into that many simple substitution encryptions.
  - Use frequency analysis to solve the resulting simple substitutions.



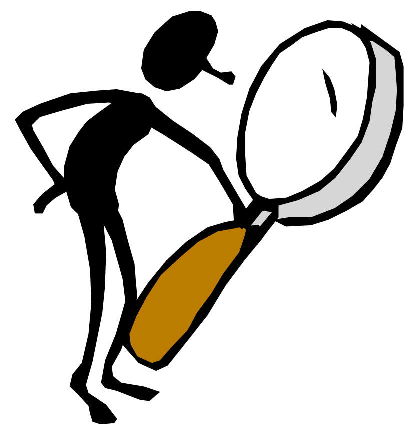
How to Find the Key Length?

- For Vigenere, as the length of the keyword increases, the letter frequency shows less English-like characteristics and becomes more random.
- Two methods to find the key length:
  - Kasisky test
  - Index of coincidence (Friedman)



Kasisky Test

- Note: two identical segments of plaintext, will be encrypted to the same ciphertext, if they occur in the text at the distance □, (□≡0 (mod m), m is the key length).
- Algorithm:
  - Search for pairs of identical segments of length at least 3
  - Record distances between the two segments: □1, □2, …
  - m divides gcd(□1, □2, …)



Index of Coincidence (Friedman)

Informally: Measures the probability that two random elements of an n-letters string x are identical.

Definition:

Suppose $x = x_1 x_2 \ldots x_n$ is a string of n alphabetic characters. Then $I_c(x)$, the index of coincidence is:

$$I_c(x) = P(x_i = x_j)$$

Index of Coincidence (cont.)

- Reminder: binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- Consider the plaintext x, and $f_0, f_1, \ldots f_{25}$ the frequencies with which A, B, … Z appear in x and $p_0, p_1, \ldots p_{25}$ the probabilities with which A, B, … Z appear in x.
- We want to compute $I_c(x)$.

Index of Coincidence (cont.)

- We can choose two elements out of the string of size n in $\binom{n}{2}$ ways
- For each i, there are $\binom{f_i}{2}$ ways of choosing the elements to be i

$$I_C(x) = \frac{\sum_{i=0}^{S} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{S} f_i(f_i - 1)}{n(n-1)} \approx \frac{\sum_{i=0}^{S} f_i^2}{n^2} = \sum_{i=0}^{S} p_i^2$$

Index of Coincidence of English

- For English, S = 25 and p_i can be estimated

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

$$I_c(x) = \sum_{i=0}^{25} p_i^{\,2} = 0.065$$

Finding the Key Length

$y = y_1 y_2 \ldots y_n$, m is the key length

$$\begin{array}{cc} \begin{pmatrix} y_1 & y_{m+1} & \cdots & y_{n-m+1} \\ y_2 & y_{m+2} & \cdots & y_{n-m+2} \\ \cdots & \cdots & \cdots & \cdots \\ y_m & y_{2m} & \cdots & y_n \end{pmatrix} & \begin{matrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ \ldots \\ \mathrm{y}_\mathrm{m} \end{matrix} \end{array}$$

Guessing the Key Length

- If m is the key length, then the text ``looks like'' English text

$$I_c(y_i) \approx \sum_{i=0}^{i=25} p_i^{\,2} = 0.065 \quad \forall 1 \le \mathrm{i} \le \mathrm{m}$$

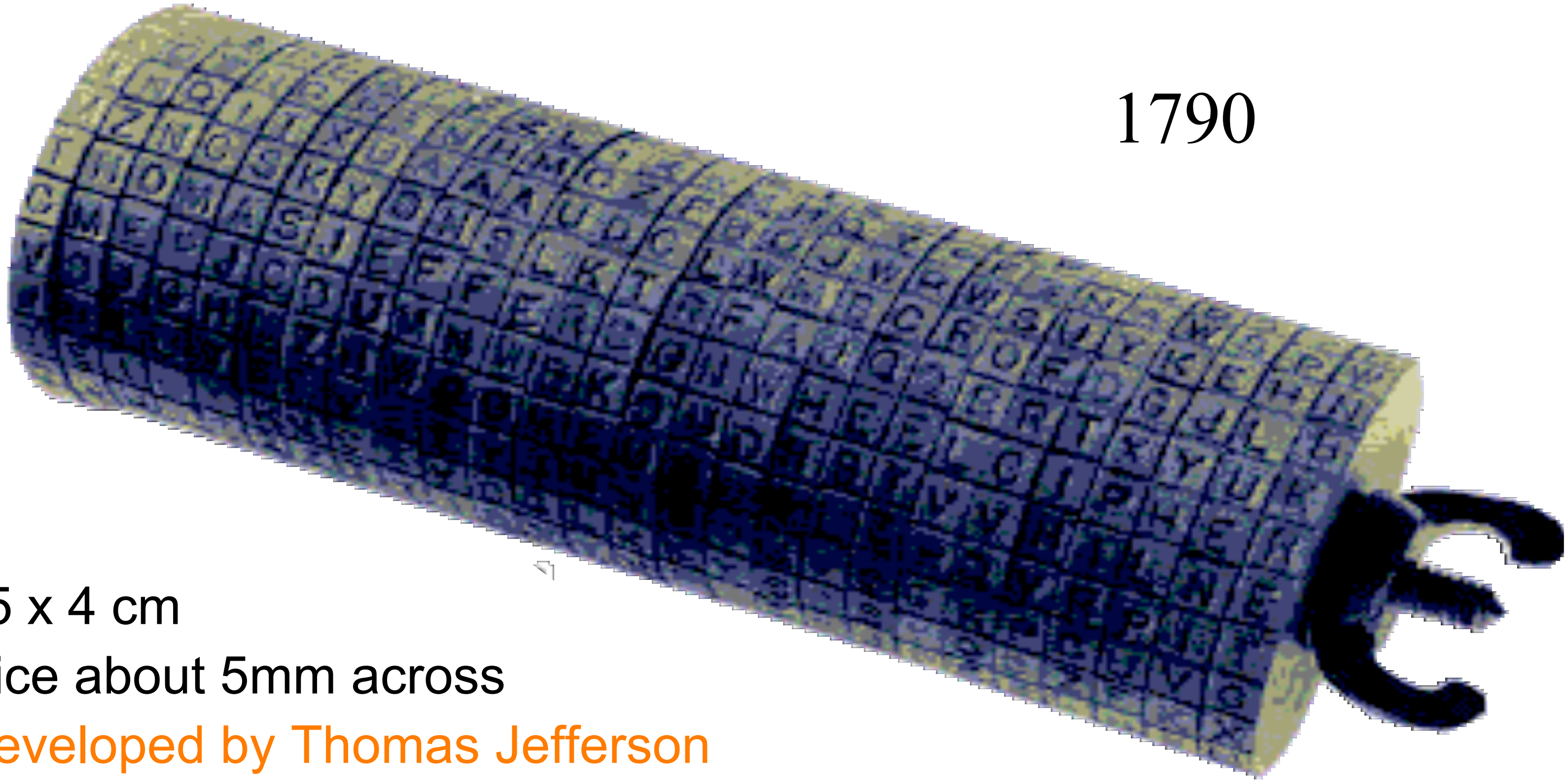
- If m is not the key length, the text ``looks like'' random text and:

$$I_c \approx \sum_{i=0}^{i=25} \left(\frac{1}{26}\right)^2 = 26 \cdot \frac{1}{26^2} = \frac{1}{26} = 0.038$$

Cipher Machines

- Used to encrypt/decrypt data
- Examples: Jefferson Cipher, Enigma Machine, Purple Machine

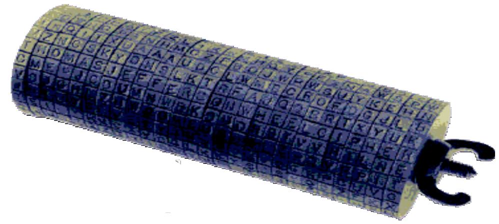
Jefferson Wheel Cipher



1790

- 15 x 4 cm
- slice about 5mm across
- Developed by Thomas Jefferson
- 25 wheels, each wheel had the letters of the alphabet on it in a different random order, wheels were set on a common axle, in the specified order

Jefferson Cipher



- Encode: rotate each wheel such the message appears along one side of the cylinder, the cylinder is then turned and another line is copied out at random.
- Decode: use the cylinder to enter the ciphertext, and then turn the cylinder examining each row until the plaintext is seen.
- Same cylinder must be used for both encryption and decryption.

Enigma Machine

- Encryption machine used by germans in the WWII, relies on elctricity
- Plug board: allowed for pairs of letters to be remapped before the encryption process started and after it ended.
- Light board
- Keyboard
- Set of rotors: user must select three rotors from a set of rotors to be used in the machine. A rotor contains one-to-one mappings of all the letters.
- Reflector (half rotor).

How Does it Work?

- Current passes through:
  - the plug board,
  - the three rotors,
  - the reflector which reverses the current,
  - back through the three rotors,
  - back through the plug board
  - then the encrypted letter is lit on the display.
- For each letter, the rotors rotate. The rotors rotate such as the right most rotor must complete one revolution before the middle rotor rotated one position and so on.

Letters Remapped

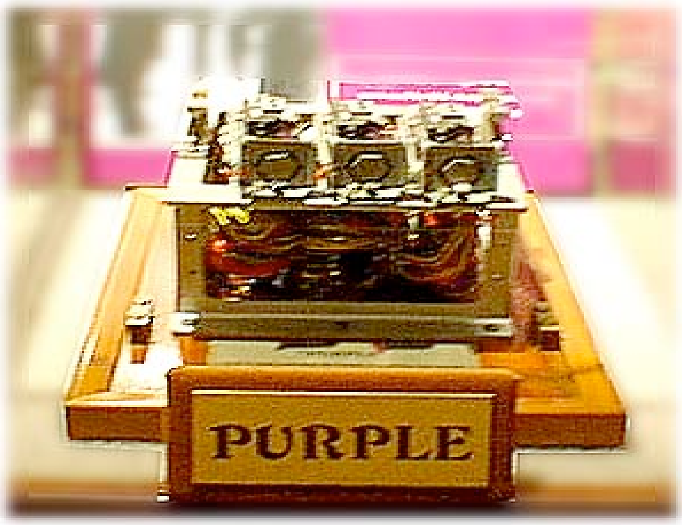
- The whole encryption process for a single letter contains a minimum of 7 remappings (the current passes through the rotors twice) and a maximum of 9 remappings (if the letter has a connection in the plug board).
  - Plug board performs the first remapping, if the letter has a connection in the plug board.
  - Rotors remap letters. Each rotor contains one-to-one mappings of letters but since the rotors rotate on each key press, the mappings of the rotors change on every key press.
  - The reflector does one more remapping, the one-to-one mappings are always the same.

Decryption

- Need the encrypted message, and know which rotors were used, the connections on the plug board and the initial settings of the rotors.
- Without the knowledge of the state of the machine when the original message was typed in, it is extremely difficult to decode a message.

Japanese Purple Machine

- Electromechanical stepping switch machine modelled after Enigma.
- Used telephone stepping switches instead of rotors
- Pearl Harbor attack preparations encoded in Purple, decoded hours before attack.



Summary

- Shift ciphers are easy to break using brute force attacks, they have small key space.
- Substitution ciphers vulnerable to frequency analysis attacks.
- Vigenere cipher is vulnerable: once the key length is found, a cryptanalyst can apply frequency analysis.

Coming Attractions …

- HW1 will be handed in class
- Perfect Secrecy
- Entropy
- Unicity Distance
- Recommended reading:
  Stinson Chapter 2.