# Enigma Cryptanalysis

**Gaurish Korpal**[1]
gaurish.korpal@niser.ac.in

Summer Internship Project Report

[1] $1^{st}$ year Int. MSc. Student, National Institute of Science Education and Research, Bhubaneswar (Odisha)

# Certificate

Certified that the summer internship project report "Enigma Cryptanalysis" is the bonafide work of "Gaurish Korpal", $1^{st}$ Year Int. MSc. student at National Institute of Science Education and Research, Bhubaneswar (Odisha) carried out under my supervision during July 6, 2015 to July 26, 2015.

Place: Delhi
Date: July 26, 2015

Prof. Geetha Venkataraman
**Supervisor**
Professor of Mathematics,
School of Liberal Studies
Ambedkar University Delhi,
Kashmere Gate Campus,
Delhi 110006

**Abstract**

For centuries, kings, queens and generals have all been aware of the consequences of their messages falling into the wrong hands, revealing precious secrets to rival nations. It was the threat of enemy interception that motivated the development of codes and ciphers. This lead to development of *cryptography*, the practice and study of techniques for secure communication in the presence of third parties. In this project report I will analyse 1930s model of Enigma, an electromechanical rotor cipher machine, which was used by Nazi Germany Army *before* World War II. Objective of this report is to discuss a simple theorem about *permutation groups*, which was the key step involved in breaking 1930s Enigma ciphers.

# Contents

# Introduction

Enigma was invented by the German engineer *Arthur Scherbius* at the end of World War I. The Enigma machine is a combination of mechanical and electrical subsystems. The mechanical subsystem consists of a *keyboard* with 26 keys; 3 rotating disks called *rotors* arranged adjacently along a spindle; and one of various stepping components to turn at least one rotor with each key press and the last rotor came before a *reflector*, a patented feature unique to Enigma among the period's various rotor machines.

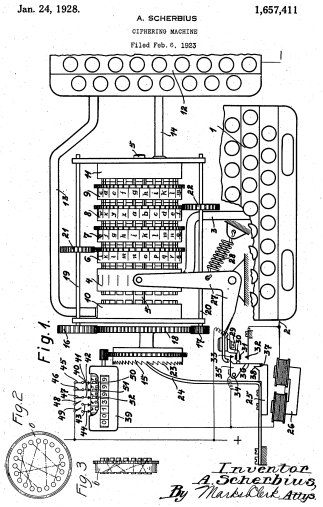


Figure 1: Scherbius's Enigma patent (https://www.google.com/patents/US1657411)

There was a ring around the circumference of each rotor on which the the alphabets A,B, ..., Z or the numbers 01, 02, ..., 26 were engraved. This ring could be rotated around the circumference and then held in place with a pin. The *ring setting* of the key indicated the letter of the alphabet on the ring that corresponded to the position of the pin. The purpose of the ring setting was to set the letters on the ring with respect to the internal wiring of the rotor.

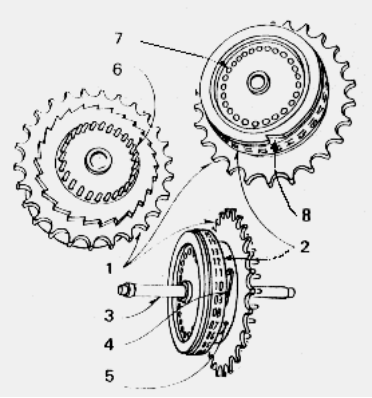


Figure 2: Details of an Enigma rotor: (1) The finger notches used to turn the rotors to a start position; (2) The alphabet RING; (3) The shaft upon which the rotors turn; (4) The catch which locks the alphabet ring to the core (5); (5) The CORE containing the cross-wiring between contacts (6) and discs (7); (6) The spring loaded contacts to make contact with the next rotor; (7) The discs embedded into the core to make contact with the spring loaded contacts in the next rotor. (8) The CARRY notch attached to the alphabet ring; (http://www.codesandciphers.org.uk/virtualbp/enigma/enigma_ files/enigwhls.gif)

The mechanical parts act in such a way as to form a varying *electrical circuit*. When a key is pressed, one or more rotors move to form a new rotor configuration, and a circuit is completed. Current flows through various components in the new configuration, ultimately lighting one display lamp, which shows the output letter.

Early models were used commercially from the early 1920s, and adopted by military and government services of several countries. Scherbius's Enigma is known as *unsteckered Enigma*, Enigma without plugboards.

In 1930 the German military increased security of Scherbius's Enigma by the addition of a plugboard. The plugboard contributed more cryptographic strength than an extra rotor. *Unsteckered Enigma* can be solved relatively straightforwardly using hand methods.

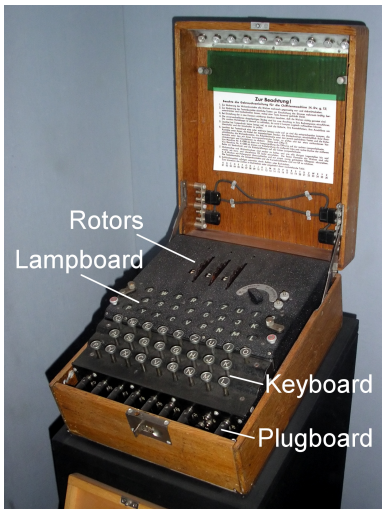


Figure 3: German Army Enigma Machine (https://commons.wikimedia.org/wiki/File:EnigmaMachineLabeled.jpg)

The Polish Cipher Bureau sought to break Enigma ciphers due to the increasing threat that Poland faced from Germany. Near the beginning of 1929, the Polish Cipher Bureau invited math students at Poznań University to take a class on cryptology. On 1 September 1932, 27-year-old Polish mathematician Marian Rejewski and two fellow Poznań University mathematics graduates, Henryk Zygalski and Jerzy Różycki, joined the Bureau full-time and moved to Warsaw.

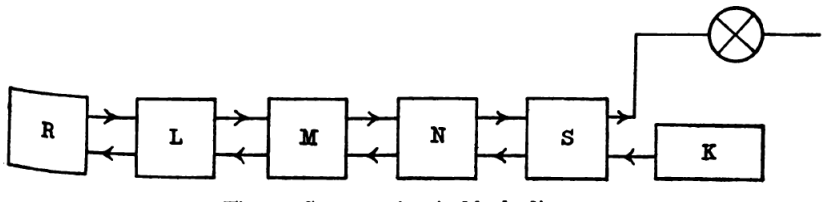


Figure 4: Current circuit diagram [R- reflector; L,M,N - rotors; S - plugboard cables; K - Keyboard; ⊗ - Lampboard] (see [9])

When the Poles began to attack Enigma, 6 plugs were in use. Rejewski reverse-engineered the device, using theoretical mathematics and material supplied by French military intelligence. Subsequently the three mathematicians designed mechanical devices for breaking Enigma ciphers. From 1938 onwards, additional complexity was repeatedly added to the Enigma machines, making decryption more difficult and requiring further equipment and personnel-more than the Poles could readily produce.

# Chapter 1

# Questions on Permutation Groups

These are some warm-up question-answers taken from [1], [2] and [3].

Q1 **Illustrate Cayley's Theorem by calculating the left regular representation for the group $V_4 = \{e, a, b, c\}$ where $a^2 = b^2 = c^2 = e, ab = ba = c, ac = ca = b, bc = cb = a$.**

*Solution.* According to Cayley's Theorem, $V_4$ is isomorphic to some subgroup of $A(V_4)$, let that subgroup be $H$. Hence, there exists a mapping $\Lambda : V_4 \to H$ such the for all $x, y \in V_4$, $\Lambda(x \cdot y) = \Lambda(x) * \Lambda(y)$ where $\cdot$ and $*$ are the binary operations of groups $V_4$ and $H$ respectively and $\Lambda$ is one-to-one. Also, $H$ consists of set of functions, $\lambda_v : V_4 \to V_4$, where, if $v \in V_4$, then $\lambda_v(z) = v \cdot z$ for every $z \in V_4$.

Thus the left regular representation for the group $V_4$ can be derived from Cayley table:

| $\cdot$ | $e$ | $a$ | $b$ | $c$ |
|---|---|---|---|---|
| $e$ | $e$ | $a$ | $b$ | $c$ |
| $a$ | $a$ | $e$ | $c$ | $b$ |
| $b$ | $b$ | $c$ | $e$ | $a$ |
| $c$ | $c$ | $b$ | $a$ | $e$ |

$$\lambda_e : \begin{bmatrix} e & a & b & c \\ e & a & b & c \end{bmatrix} \qquad \lambda_a : \begin{bmatrix} e & a & b & c \\ a & e & c & b \end{bmatrix}$$

$$\lambda_b : \begin{bmatrix} e & a & b & c \\ b & c & e & a \end{bmatrix} \qquad \lambda_c : \begin{bmatrix} e & a & b & c \\ c & b & a & e \end{bmatrix}$$

Now from Cayley table of $H$ we can clearly see the isomorphism:

| $*$ | $\lambda_e$ | $\lambda_a$ | $\lambda_b$ | $\lambda_c$ |
|---|---|---|---|---|
| $\lambda_e$ | $\lambda_e$ | $\lambda_a$ | $\lambda_b$ | $\lambda_c$ |
| $\lambda_a$ | $\lambda_a$ | $\lambda_e$ | $\lambda_c$ | $\lambda_b$ |
| $\lambda_b$ | $\lambda_b$ | $\lambda_c$ | $\lambda_e$ | $\lambda_a$ |
| $\lambda_c$ | $\lambda_c$ | $\lambda_b$ | $\lambda_a$ | $\lambda_e$ |

Q2 **Show that $A_5$ has 24 elements of order 5, 20 elements of order 3, and 15 elements of order 2.**

*Solution.* $A_5$ is the set of permutations of $\{1, 2, 3, 4, 5\}$ that can be expressed as a product of even number of 2-cycles. Also if a permutation $\alpha$ can be expressed as a product of even number of 2-cycles, then every decomposition of $\alpha$ into a product of 2-cycles must have an even number of 2-cycles. Further we can write any cycle of length $m$ as a product of $(m-1)$ 2-cycles. Thus we need to consider only following disjoint cycle structures of the elements of $S_5$:
$(\underline{5})$
$(\underline{3})(\underline{1})(\underline{1})$
$(\underline{2})(\underline{2})(\underline{1})$
$(\underline{1})(\underline{1})(\underline{1})(\underline{1})(\underline{1})$
Now, from simple Combinatorial arguments, we know that

The number of ways to arrange $n$ distinct objects along a fixed circle $= (n-1)!$

Thus using above formula and multiplication principle of counting:
Number of elements of order $5 = \binom{5}{0}(5-1)! = 4! = 24$

Number of elements of order 3 $= \binom{5}{2} \cdot \binom{3}{3}(3-1)! = \frac{5! \cdot 2!}{3! \cdot 2!} = 20$

Number of elements of order 2 $= \binom{5}{1} \cdot \binom{4}{2} \frac{(2-1)!}{2} \cdot \binom{2}{2} = \frac{5 \cdot 4!}{2 \cdot 2! \cdot 2!} = 15$

Number of elements of order 1 $= \binom{5}{5} = 1$

Q3 **Show that if $n \geq m$ then the number of $m$-cycles in $S_n$ is given by $n(n-1)(n-2)...(n-m+1)/m$.**

*Solution.* We can generalize combinatorial argument of previous problem to get[1]

# cycles of order $m$ = (# select $m$ letters from $n$ letters) $\cdot$ (# arrange $m$ distinct objects on a circle)

$$\text{\# cycles of order } m = \binom{n}{m} \cdot (m-1)! = \frac{n!}{(n-m)! \cdot m} = \frac{n(n-1)\dots(n-m+1)}{m}$$

Q4 **Let $\sigma$ be the $m$-cycle $(12\dots m)$. Show that $\sigma^i$ is also an $m$-cycle if and only if $i$ is relatively prime to $m$.**

*Solution.* The permutation $\sigma$ can be visualised by placing the numbers $\{1, 2, \dots, m\}$ clockwise around a circle, then $\sigma$ takes each number to the one next to it clockwise. Since $\sigma$ is a $m-$cycle, $m$ is the smallest number of compositions for which $\sigma^m = \varepsilon$ where $\varepsilon$ is the identity element of the permutation group.

Thus, $\sigma$ takes the number $k$ to $k+1$, considered modulo $m$.

Then, $\sigma^i$ takes each number $i$ places around the circle in clockwise direction from it, so it takes the number $k$ to $k+i$ modulo $m$. Thus one of the cycles of $\sigma^i$ is $(1, 1+i, 1+2i, \cdots)$ modulo $m$. If this cycle has length $m$, then it must be the only cycle (since all the numbers from 1 to $m$ must be in it), so that $\sigma^i$ is an $m$-cycle. If this cycle has length less than $m$, then there must be more than one cycle an $\sigma^i$ is not an $m$-cycle.

Let $\sigma^i$ be a cycle of length $k$, then $k$ is the smallest positive integer such that $1 + ki \equiv 1 \pmod{m}$. Equivalently, $k$ is the smallest positive integer such that $m|ki$.

If $i$ is relatively prime to $m$, then $m|ik$ implies $m|k$, so that the smallest such $k$ is $k = m$.

If $i$ is not relatively prime to $m$ then let $\gcd(i, m) = d$. So $m = dm'$ and $i = di'$ where $\gcd(m', i') = 1$. Then the smallest positive integer $k$ such that $dm'|di'k$, or equivalently, $m'|i'k$. Since $m', i'$ are relatively prime, the smallest such $k$ is $k = m'$.

Thus, if $i$ is relatively prime to $m$ then $\sigma^i$ is a $m$-cycle. On the other hand, if $i$ is not relatively prime to $m$, then $\sigma^i$ contains an $m'$-cycle, for some $m' < m$, so cannot be a $m$-cycle.

Q5 **Let $n \geq 3$. Prove the following in $S_n$.**

(a) Every permutation of $S_n$ can be written as a product of at most $n-1$ transpositions.

(b) Every permutation of $S_n$ that is not a cycle can be written as a product of at most $n-2$ transpositions.

*Solution.* Here we need to generalize the argument used in Q2.

(a) Since every permutation in $S_n$ is a product of 2-cycles (called transpositions), such that:

$$(a_1 \dots a_p)(b_1 \dots b_q) \dots (c_1 \dots c_r) = (a_1 a_p) \dots (a_1 a_2)(b_1 b_q) \dots (b_1 b_2) \dots (c_1 c_r) \dots (c_1 c_2)$$

Thus:

# transpositions for every permutation of $S_n = (p-1)+(q-1)+\dots+(r-1) = n-$(# disjoint cycles)

Since $p + q + \dots + r = n$, if we explicitly write 1-cycles also and consider them to be disjoint cycles. Thus:

$$\max\{\text{\# transpositions for every permutation of } S_n\} = n - \min\{\text{\# disjoint cycles}\} = n - 1$$

[1] for ease of writing, let # denote "number of ..."

(b) If permutation of $S_n$ is not a cycle then $\min\{\#\text{ disjoint cycles}\} = 2$. Thus,

$$\max\{\#\text{ transpositions for every permutation of } S_n \text{ that is not a cycle}\} = n - 2$$

Q6 **Let $\sigma$ be a permutation of a set $A$. We say that $\sigma$ moves $a \in A$ if $\sigma(a) \neq a$. Let $S_A$ denote the permutations on $A$.**

(a) If $A$ is a finite set then how many elements are moved by a $n$-cycle $\sigma \in S_A$?

(b) Let $A$ be an infinite set and let $H$ be the subset of $S_A$ consisting of all $\sigma \in S_A$ such that $\sigma$ only moves finitely many elements of $A$. Show that $H \leq S_A$.

(c) Let $A$ be an infinite set and let $K$ be the subset of $S_A$ consisting of all $\sigma \in S_A$ such that $\sigma$ moves at most 50 elements of $A$. Is $K \leq S_A$? Why?

*Solution.* (a) $n$ elements

(b) We need to verify two conditions:

i. $\sigma_1, \sigma_2 \in H \Rightarrow \sigma_1 * \sigma_2 \in H$
If a permutation $\sigma_1$ and $\sigma_2$ move a finite number of elements of A, then $\sigma_1 * \sigma_2$, where $*$ is binary operation defined for $S_A$, also moves at most finite number of elements of $A$. Therefore, $H$ is closed under the binary operation.

ii. $\sigma \in H \Rightarrow \sigma^{-1} \in H$
If $\sigma \in H$ then $\sigma$ moves only finite number of elements of A. Notice that $\sigma^{-1}$ moves exactly the same elements as $\sigma$, therefore, $\sigma^{-1}$ also moves only finite number of elements of A, and $\sigma^{-1} \in H$

(c) $K$ is not a subgroup of $S_A$.
Take 100 different elements of $A$, denote them $a_1, a_2, \ldots, a_{100}$. Consider permutations $\sigma_1 = (a_1 a_2 \ldots a_{49} a_{50})$ and $\sigma_2 = (a_{51} a_{52} \ldots a_{99} a_{100})$. Then $\sigma_1, \sigma_2 \in K$, but $\sigma_1 * \sigma_2 \notin K$. Therefore, $K$ is not closed under the binary operation in $S_A$ (composition).

Q7 **Show that if $\sigma$ is a cycle of odd length then $\sigma^2$ is a cycle.**

*Solution.* This is a corollary of Q4. Also explicitly,

$$\sigma = (a_1 a_2 \ldots a_{2m+1})$$

$$\Rightarrow \sigma^2 = (a_1 a_2 \ldots a_{2m+1})(a_1 a_2 \ldots a_{2m+1}) = (a_1 a_3 \ldots a_{2m-1} a_{2m+1} a_2 a_4 \ldots a_{2m-2} a_{2m})$$

Q8 **Let $p$ be a prime. Show that an element has order $p$ in $S_n$ if and only if its cycle decomposition is a product of commuting $p$-cycles. Show by an explicit example that this need not be the case if $p$ is not prime.**

*Solution.* We know following 3 theorems about Permutation Groups (see [3])

(a) Every permutation of a finite set can be written as a cycle or as a product of disjoint cycles.

(b) Disjoint cycles commute

(c) (Ruffini's Theorem) The order of a permutation of a finite set written in disjoint cycle form is the least common multiple of the lengths of the cycle.

Let $\alpha_1, \alpha_2, \ldots \alpha_k$ be the length of $k$ disjoint cycles of a permutation in $S_n$. Now, since order of this permutation is a prime number, $p$.

$$p = \text{lcm}(\alpha_1, \alpha_2, \ldots \alpha_k)$$
$$\Rightarrow \alpha_i | p \qquad \text{for all } 1 \leq i \leq k$$
$$\Rightarrow \alpha_i = 1 \quad \text{or} \quad \alpha_i = p$$

This proves, the given statement, that an element has order $p$ in $S_n$ if and only if its cycle decomposition is a product of commuting $p-$ cycles.

Clearly, from above argument, this need not be the case if $p$ is not prime. Let, $q$ be a composite number, then

*An element has order $q$ in $S_n \not\Rightarrow$ its cycle decomposition is a product of commuting $q-$cycles.*

Consider and element of order 6, in $S_7$, $(123)(45)(67)$, this is not a product of commuting 6-cycles.

Note: *Cycle decomposition of an element in $S_n$ is a product of commuting $q-$cycles $\Rightarrow$ its order is $q$*

Q9 **Show that if $n \geq 4$ then the number of permutations in $S_n$ which are the product of two disjoint 2-cycles is $n(n-1)(n-2)(n-3)/8$.**

*Solution.* Such permutations are of form: $(\underline{2})(\underline{2})\underbrace{(\underline{1})(\underline{1})\dots(\underline{1})}_{(n-4)\text{ 1-cycles}}$, by following Q2, we can say:

$$\#\text{ permutations in } S_n \text{ product of two disjoint 2-cycles} = \binom{n}{n-4}\cdot\binom{4}{2}\frac{(2-1)!}{2} = \frac{n(n-1)(n-2)(n-3)}{8}$$

Q10 **Let $b \in S_7$ and suppose $b^4 = (2143567)$. Find $b$.**

*Solution.* Since $o(b^4) = 7$ it follows that $e = (b^4)^7 = b^{28}$, so $o(b)|28$, hence $o(b) = 1$, 2, 4, 7, 14 or 28.

Suppose that $o(b) = 28$, we know we can decompose $b$ into disjoint cycles, and if $o(b) = 28$, then the least common multiple of these cycles must be 28, but the maximum number of symbols that may appear in any permutation in $S_7$ is 7 so it must be the case that $b$ is a 7-cycle, which cannot be the case since then $b$ would have order 7, or $b$ is a product of 2 and/or 4 cycles, which cannot be the case either, since then the least common multiple of the lengths of cycles would not be 28. So $o(b) \neq 28$.

Suppose that $o(b) = 14$, we know we can decompose $b$ into disjoint cycles, and if $o(b) = 14 = 2 \times 7$, then the least common multiple of these cycles must be 14, but the maximum number of symbols that may appear in any permutation in $S_7$ is 7 so it must be the case that $b$ is a 7-cycle, which cannot be the case since then $b$ would have order 7, then as in previous case, $o(b) \neq 14$.

Since $b^4 \neq e$, clearly $o(b) \neq 1$, 2, or 4.

The remaining possibility is that $o(b) = 7$, which therefore must be true, and it must be the case that $b$ is 7-cycle, for this is the only choice of lengths of cycles in $S_7$ such that the least common multiple of the lengths of cycles will be 7.

Let, $b = (a_1a_2a_3a_4a_5a_6a_7)$, and as per the question:

$$(a_1a_2a_3a_4a_5a_6a_7)(a_1a_2a_3a_4a_5a_6a_7)(a_1a_2a_3a_4a_5a_6a_7)(a_1a_2a_3a_4a_5a_6a_7) = (2143567)$$

$$\Rightarrow (a_1a_5a_2a_6a_3a_7a_4) = (2143567)$$

$$\Rightarrow a_1 = 2, \quad a_2 = 4, \quad a_3 = 5, \quad a_4 = 7, \quad a_5 = 1, \quad a_6 = 3, \quad a_7 = 6$$

$$\Rightarrow b = (2457136)$$

Q11 **Let $b = (123)(145)$. Write $b^{99}$ in disjoint cycle form.**

*Solution.* Firstly write $b$ in disjoint cycle form to calculate $o(b)$.

$$b = (123)(145) = (14523)$$

$$\Rightarrow o(b) = 5$$

Now, $99 \equiv 4 \pmod 5$, so

$$b^{99} = b^{5\times 19+4} = b^4 = (14523)(14523)(14523)(14523) = (13254)$$

Q12 **Find three elements $\sigma$ in $S_9$ with the property that $\sigma^3 = (157)(283)(469)$.**

*Solution.* Here, $o(\sigma^3) = 3$, thus $\varepsilon = (\sigma^3)^3 = \sigma^9$, hence $o(\sigma) = 1$,3, or 9. Since $\sigma^3 \neq \varepsilon$, we get, $o(\sigma) \neq 1$ or 3.

Thus, $o(\sigma) = 9$ and since $S_9$ has only 9 letters, we get:

$$\sigma = (a_1a_2a_3a_4a_5a_6a_7a_8a_9)$$

And according to question:

$$(a_1a_2a_3a_4a_5a_6a_7a_8a_9)(a_1a_2a_3a_4a_5a_6a_7a_8a_9)(a_1a_2a_3a_4a_5a_6a_7a_8a_9) = (157)(283)(469)$$

$$\Rightarrow (a_1a_4a_7)(a_2a_5a_8)(a_3a_6a_9) = (157)(283)(469)$$

Since disjoint cycles commute we need to consider all $3! = 6$ cases, :

$$\begin{cases} (a_1a_4a_7) = (157), & (a_2a_5a_8) = (283), & (a_3a_6a_9) = (469) \Rightarrow \sigma = (124586739) \\ (a_1a_4a_7) = (157), & (a_3a_6a_9) = (283), & (a_2a_5a_8) = (469) \Rightarrow \sigma = (142568793) \\ (a_2a_5a_8) = (157), & (a_1a_4a_7) = (283), & (a_3a_6a_9) = (469) \Rightarrow \sigma = (214856379) \\ (a_3a_6a_9) = (157), & (a_1a_4a_7) = (283), & (a_2a_5a_8) = (469) \Rightarrow \sigma = (241865397) \\ (a_3a_6a_9) = (157), & (a_2a_5a_8) = (283), & (a_1a_4a_7) = (469) \Rightarrow \sigma = (421685937) \\ (a_2a_5a_8) = (157), & (a_3a_6a_9) = (283), & (a_1a_4a_7) = (469) \Rightarrow \sigma = (412658973) \end{cases}$$

Thus all these 6 elements of $S_9$ satisfy given property.

Q13 **Show that if $H$ is a subgroup of $S_n$, then either every member of $H$ is an even permutation or exactly half of the members are even.**

*Solution.* We are given that $H \leq S_n$, thus:

(a) $\sigma_1, \sigma_2 \in H \Rightarrow \sigma_1\sigma_2 \in H$

(b) $\sigma \in H \Rightarrow \sigma^{-1} \in H$

If $\sigma_1, \sigma_2$ are even permutations then, $\sigma_1\sigma_2$ will also be even permutation, and since inverse of permutation is letters written in reverse order, we conclude that it is possible for this subgroup to have only even permutations as its members. (In other words, $H$ is a subgroup of $A_n$.)

Else, if $\sigma_1, \sigma_2$ are odd permutations then, $\sigma_1\sigma_2$ will be even permutation, and since inverse of permutation is letters written in reverse order, we conclude that $H$ will have some even and some odd permutations as members. Same argument is applicable when only one of $\sigma_1, \sigma_2$ is odd permutation. Now what remains to prove is that, if there are elements of odd permutations, then:

$$\#\text{ elements of } H \text{ with even permutations} = \#\text{ elements of } H \text{ with odd permutations}$$

Let $\sigma$ be an element with odd permutation in $H$.

For each odd permutation $\alpha$, the permutation $\sigma\alpha$ is even and $\sigma\alpha \neq \sigma\beta$ when $\alpha \neq \beta$. Thus:

$$\#\text{ elements of } H \text{ with even permutations} \geq \#\text{ elements of } H \text{ with odd permutations} \tag{1.1}$$

On the other hand, for each even permutation $\alpha$, the permutation $\sigma\alpha$ is odd and $\sigma\alpha \neq \sigma\beta$ when $\alpha \neq \beta$. Thus:

$$\#\text{ elements of } H \text{ with even permutations} \leq \#\text{ elements of } H \text{ with odd permutations} \tag{1.2}$$

Now combining the equations (1.1) and (1.2), we complete proof of given statement.

Q14 **Suppose that $H$ is a subgroup of $S_n$ of odd order. Prove that $H$ is a subgroup of $A_n$.**

*Solution.* This result follows quickly from previous problem. If $H$ is a subgroup of $S_n$, and if $H$ contains an odd permutation, then exactly half the elements of $H$ are odd permutations (form Q13). If $H$ has odd order, it's impossible for exactly half of its elements to be odd permutations; therefore $H$ contains no odd permutations at all, i.e., $H$ is a subgroup of $A_n$.

Q15 **Prove that the smallest subgroup of $S_n$ containing $(12)$ and $(12\dots n)$ is $S_n$. In other words, these generate $S_n$.**

*Solution.* Let $H$ be the smallest subgroup of $S_n$ containing $\{(12),(12\dots n)\}$, now as per properties of subgroup, stated in Q13, we can see that:

$$(12)(12\dots n) = (1)(23\dots n) \in H \quad \text{and} \quad (12\dots n)(12) = (13\dots n)(2) \in H$$

Now we have generated new elements of our group, we can apply above process again:

$$(12)(12\dots n)^2 = \text{can't write explicit cycle since it depends upon parity of } n$$

Recall that, raising a cycle to the $k$-th power simply has the effect of sending each listed number to the one $k$ spaces to the right; thus:

$$\begin{aligned}
(12\dots n)^0 &= \{1\mapsto 1, 2\mapsto 2, \dots\}\\
(12\dots n)^1 &= \{1\mapsto 2, 2\mapsto 3, \dots\}\\
(12\dots n)^2 &= \{1\mapsto 3, 2\mapsto 4, \dots\}\\
&\vdots \qquad\qquad\qquad \vdots\\
(12\dots n)^{n-2} &= \{1\mapsto (n-1), 2\mapsto n, \dots\}\\
(12\dots n)^{n-1} &= \{1\mapsto n, 2\mapsto 1, \dots\}\\
(12\dots n)^n &= \{1\mapsto 1, 2\mapsto 2, \dots\}
\end{aligned}$$

We can rather try to compute:

$$(12\dots n)^k(12)(12\dots n)^{n-k} \quad \text{for } k = 0,1,2,\dots n$$

Thus:

$$\begin{aligned}
(12\dots n)^0(12)(12\dots n)^{n-0} &= (12)\\
(12\dots n)^1(12)(12\dots n)^{n-1} &= (23)\\
(12\dots n)^2(12)(12\dots n)^{n-2} &= (34)\\
&\vdots \qquad\qquad\qquad\qquad \vdots\\
(12\dots n)^2(12)(12\dots n)^{n-2} &= ((n-1)\,n)\\
(12\dots n)^{n-1}(12)(12\dots n)^1 &= (n1)\\
(12\dots n)^n(12)(12\dots n)^0 &= (12)
\end{aligned}$$

Thus, we observe that in general the permutations $(12\dots n)^k(12)(12\dots n)^{n-k}$ with $k$ varying will give the transpositions $(12),(23),(34),\dots,((n-1)n),(n1)$. Thus, all of these transpositions are generated by $(12)$ and $(12\dots n)$.

But we know that *every permutation in $S_n$ is a product of two cycles*, thus what remains to prove is: *all transpositions are generated by those listed above.*

Observe that:

$$(13) = (12)(23)(12) \qquad \text{and} \qquad (24) = (23)(34)(23)$$

Now let's generalize this, consider a transposition, $(ab)$, with $a$ is to the left of $b$.

$$(ab) = \Big(a(a+1)\Big)\Big((a+1)(a+2)\Big)\cdots\Big((b-1)b\Big)\Big((b-2)(b-1)\Big)\cdots\Big((a+1)(a+2)\Big)\Big(a(a+1)\Big)$$

This involves transpositions of consecutive numbers, building up from $a$ to $b$ over the first part and then back down from $b$ to $a$ over the second. Thus $(12)$ and $(123\dots n)$ generated all transpositions, since they generated all transpositions involving consecutive letters.

Since $S_n$ is generated by all transpositions, we thus conclude that it is possible to express any permutation as a product involving only $(12)$ and $(123\dots n)$, so that these two permutations generate all of $S_n$. Hence $H$ is actually, whole of $S_n$.

Q16 **Prove that for $n \geq 3$ the subgroup generated by the 3-cycles is $A_n$.**

*Solution.* Consider a 3-cycle: $\sigma' = (a_1a_2a_3) = (a_1a_3)(a_1a_2)$, thus $\sigma' \in A_n$. Now let the group generated by this 3-cycle be $H$, then clearly, $H \leq A_n$. To prove that in fact this subgroup generated is $A_n$, we need to follow the argument used to prove that unit element, $\varepsilon$, is an even permutation.

Let, $\sigma \in A_n$, then, $\sigma = \alpha_1\alpha_2 \ldots \alpha_{2k}$, where $\alpha_i$ is a 2-cycle. Further we know that $\alpha_{2k-1}\alpha_{2k}$ can be expressed in one of the following forms:

$$\begin{aligned}(ab)(ab) &= \varepsilon \\ (ab)(bc) = (ac)(ab) &= (abc) \\ (ac)(cb) = (bc)(ab) &= (acb) \\ (ab)(cd) = (cd)(ab) &= (abc)(bcd)\end{aligned}$$

If the first case occurs, we may delete $\alpha_{2k-1}\alpha_{2k}$, from original product to obtain: $\sigma = \alpha_1\alpha_2 \ldots \alpha_{2k-2}$.

In the other three cases, we replace the 2-cycles on left by corresponding 3-cycle on right.

Now we can repeat the procedure just describe with $\alpha_{2k-1}\alpha_{2k}$, and we obtain a product of 3-cycles. Thus every $\sigma$ is a product of 3-cycles. Hence our subgroup $H$ is essentially the group $A_n$. Thus proving given statement.

Q17 **Prove that if a normal subgroup of $A_n$ contains even a single 3-cycle it must be all of $A_n$.**

*Solution.* Consider, $N \lhd A_n$, thus for every element $x \in A_n$ and $n \in N$, $xnx^{-1} \in N$.

Suppose that $\sigma = (abc) = (ac)(ab) \in N$, be the only 3-cycle in $N$. Then $S = \{a, b, c\}$ is a unique non-trivial orbit of $\sigma$ (as defined on pp. 77, [1]). From Q2, we can say that there are $(3-1)! = 2$, 3-cycles which have $S$ as their non-trivial orbit, namely $(abc)$ and $(abc)^{-1} = (cba)$. Following previous problem, we need to only show that any 3-element subset $S'$ of $\{1, 2, \ldots, n\}$ is the non-trivial orbit of a 3-cycle in $N$ i.e. $S \neq S'$ or equivalently $|S \cap S'| < 3$.

Now we need to show that there exists $x \in A_n$ such that $\sigma' = x\sigma x^{-1} \in N$, where $\sigma'$ has orbit $S'$. So we have three cases to consider:

Case 1: $|S \cap S'| = 0$
Let, $S' = \{d, e, f\}$ and suppose $\sigma' = (def)$, then we need to find $x$, such that:

$$(def) = x(abc)x^{-1} \qquad \Rightarrow (df)(de) = x(ac)(ab)x^{-1}$$

Then after some trial and error: $x = (ae)(ad)(ab)(cf) = (abde)(cf)$, does the trick.

Case 2: $|S \cap S'| = 1$
Let, $S' = \{a, d, e\}$ and suppose $\sigma' = (ade)$, then we need to find $x$, such that:

$$(ade) = x(abc)x^{-1} \qquad \Rightarrow (ae)(ad) = x(ac)(ab)x^{-1}$$

Thus, clearly, $x = (bd)(ce)$ does the trick. (basic idea was to generate a map of form $a \mapsto c \mapsto e$)

Case 3: $|S \cap S'| = 2$
Let, $S' = \{a, b, d\}$ and suppose $\sigma' = (adb)$, then we need to find $x$, such that:

$$(adb) = x(abc)x^{-1} \qquad \Rightarrow (ab)(ad) = x(ac)(ab)x^{-1}$$

Thus, clearly, $x = (ab)(cd)$ does the trick. (basic idea was to preserve the cycle $a \mapsto b \mapsto a$)

Q18 **Prove that $A_5$ has no non-trivial proper normal subgroups. In other words show that $A_5$ is a simple group.**

*Solution.* Suppose that $N \lhd A_5$ and $N \neq \{e\}$. We will show that $N = A_5$, which will prove that $A_5$ is simple, since $N$ is arbitrary.

Using he fact that $N \neq \{e\}$ and since $N$ is a subset of $A_5$ , we conclude that $N$ contains a non-trivial even permutation $\sigma$. From Q2, we can say that $\sigma$ has disjoint cycle decomposition of one of following 3 types:

Case 1: $\sigma = (abcde)$
Consider: $\alpha = (ab)(cd) \in A_5$, then we have $\sigma' \in N$, such that:

$$\sigma' = \alpha\sigma\alpha^{-1} = (ab)(cd)(abcde)(ab)(cd) = (adceb)$$

But since $N$ is a subgroup of $A_5$, it must also contain $\sigma\sigma'$

$$\sigma\sigma' = (abcde)(adceb) = (aec)$$

From Q17, we can say that, since $N$ contains a 3-cycle, and $N \lhd A_5$, $N = A_5$.

Case 2: $\sigma = (ab)(cd)$
Consider: $\beta = (abe)$, then we have $\sigma' \in N$, such that:

$$\sigma' = \beta\sigma\beta^{-1} = (abe)(ab)(cd)(eba) = (be)(cd)$$

Now again since $\sigma\sigma' \in N$, thus:

$$\sigma\sigma' = (ab)(cd)(be)(cd) = (abe)$$

Again from Q17, we can say that, since $N$ contains a 3-cycle, and $N \lhd A_5$, $N = A_5$.

Case 3: $\sigma = (abc)$
Follows directly from Q17.

Q19 **Show that $Z(S_n)$ is trivial for $n \geq 3$.**

*Solution.* Note that, $Z(G) = \{x \in G : gx = xg \quad \text{for all } x \in G\}$ and $Z(S_n)$ is trivial implies that $Z(S_n) = \{\varepsilon\}$, where $\varepsilon$ is the identity element.

To prove that $Z(S_n)$ is trivial is equivalent to prove that for all $\sigma \in S_n$, such that $\sigma \neq \varepsilon$, there exists $\alpha \in S_n$ such that $\sigma\alpha \neq \alpha\sigma$. (contrapositive of given statement)

Let, $\sigma \in S_n$, be a non identity element. Also, let, $\sigma(a) = b$ where $a \neq b$, then we can choose $\alpha = (bc)$ [i.e. $\alpha(b) = c$] where $\sigma(b) \neq c$ (this is clearly possible since $n \geq 3$). Thus,

$$\sigma\alpha(a) = \sigma(a) = b \qquad \text{and} \qquad \alpha\sigma(a) = \alpha(b) = c$$

Since, $\sigma(b) \neq c$, we have shown that $\sigma\alpha$ and $\alpha\sigma$ act differently[2] on $a$, and so are different elements of $S_n$.

Thus, no non-identity element of $S_n$ commutes with all elements of $S_n$, hence $Z(S_n) = \{\varepsilon\}$.

Q20 **Show that two permutations in $S_n$ are conjugate if and only if they have the same cycle structure or decomposition. Given the permutation $\alpha = (12)(34)$, $\alpha' = (56)(13)$, find a permutation $\beta$ such that $\beta\alpha\beta^{-1} = \alpha'$.**

*Solution.* We need to prove two parts :

Part 1: *Two permutations in $S_n$ are conjugate $\Rightarrow$ they have the same cycle structure or decomposition.*
Let $\alpha, \beta$ be any two permutation in $S_n$ and $\beta\alpha\beta^{-1}$, is conjugate of $\alpha$, call it $\alpha'$.
If, $\alpha = (a_1, a_2, \ldots, a_{k_1})(b_1, b_2, \ldots, b_{k_2}) \cdots (c_1, c_2, \ldots, c_{k_3})$, I claim that:

$$\alpha' = (\beta(a_1), \beta(a_2), \ldots, \beta(a_{k_1}))(\beta(b_1), \beta(b_2), \ldots, \beta(b_{k_2})) \cdots (\beta(c_1), \beta(c_2), \ldots, \beta(c_{k_3})) \tag{1.3}$$

Since both sides of above equation are permutations, we just need to check that both sides have same effect on any integer $j \in \{1, 2, \ldots n\}$. Since $\beta$ is surjective[3], $j = \beta(i)$ for some $i$. By symmetry, we can assume that $j = \beta(a_1)$. Also, $\alpha(a_1) = a_2$, thus for LHS:

$$\beta\alpha\beta^{-1}j = \beta\alpha\beta^{-1}\beta(a_1) = \beta(a_2)$$

Since RHS also takes $\beta(a_1)$ to $\beta(a_2)$, thus the LHS and RHS have the same effect on $j$ and so they must be equal. Proving my claim.
Thus, $\alpha$ and $\alpha' = \beta\alpha\beta^{-1}$ have same cycle structure.

[2] For $\gamma, \delta \in S_n$, $\gamma = \delta$ if and only if $\delta(x) = \gamma(x)$ for all $x$ (letters).
[3] permutations are bijective maps

Part 2: *Two permutations in $S_n$ have same cycle structure $\Rightarrow$ they are conjugate*

Now suppose that $\alpha$ and $\alpha'$ have the same cycle structure. We want to find a permutation $\beta$ that sends $\alpha$ to $\alpha'$ . By assumption the cycles in $\alpha$ and $\alpha'$ have the same lengths. Then we can pick a correspondence between the cycles of $\alpha$ and the cycles of $\alpha'$ .

Pick an integer $j$. Then $j$ belongs to a cycle of $\alpha$. Look at the corresponding cycle in $\alpha'$ and look at the corresponding entry, call it $j'$ . Then $\beta$ should send $j$ to $j'$ . From (1.3), we can check that $\beta\alpha\beta^{-1} = \alpha'$.

Also, we are asked to find $\beta$ for $\alpha = (12)(34)$ and $\alpha' = (13)(56)$. Since there are 6 elements involved, we can assume that we are working in $S_6$.

Now write the two permutations in full cycle notation, writing cycles from longest to shortest (cycles of the same length can be ordered arbitrarily, the starting number of cycle can be chosen arbitrarily from within the cycle).

Let, $\alpha = (12)(34)(5)(6)$ and $\alpha' = (13)(56)(2)(4)$,

$$\beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 3 & 5 & 6 & 2 & 4 \end{bmatrix}$$

In other words, $\beta(1) = 1, \beta(2) = 3$, etc. We can convert $\beta$ to cycle notation, $\beta = (235)(46)$.

Another $\beta$ is found by, let, $\alpha = (12)(34)(5)(6)$ and $\alpha' = (56)(13)(4)(2)$,

$$\beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 1 & 3 & 4 & 2 \end{bmatrix}$$

Hence, $\beta = (1543)(26)$.

Since we can reorder the cycles of the same length, and since we can "cycle" a cycle as much as we want, we actually get many different $\beta$.

# Chapter 2

# Introduction to Cryptanalysis

Here I will discuss the basic information needed to start cryptanalysis of Enigma Cipher.

## 2.1 Hiding Messages

Any message may be hidden in two basic ways. The methods of *steganography* conceal the very existence of the message, for example invisible inks and microdots. The methods of *cryptography*, on the other hand, do not conceal the presence of a secret message but render it unintelligible as *ciphertext*, to outsiders by various transformations of the *plaintext*. Breaking ciphers require ingenuity, creativity and a little math. The skill involved in breaking ciphers is called *cryptanalysis*.

## 2.2 Cipher not Code

A **code** is a mapping from some meaningful unit (word, sentence, phrase) into something else (usually a shorter group of symbols). A code requires a codebook, it is simply a list of these mappings. For example we could make up a code where the word Apple is written as 67, historical example of this is "Zimmermann Telegram Code" for more details refer [14].
**Ciphers** do not involve meaning. Instead they are mechanical operations (known as algorithms) which are performed on the individual or small chunks of letters.
A code is stored as a mapping in a codebook, while ciphers transform individual symbols according to an algorithm.

It was definitively stated in 1883 by the Dutch linguist Auguste Kerckhoffs von Nieuwenhof in his book La Cryptographie militaire:

Kerckhoffs' Principle: The security of a cryptosystem must not depend on keeping secret the crypto-algorithm. The security depends only on keeping secret the key.

The process of converting a message from plane language into secret language by systematic treatment of its letters is called *enciphering*. The inverse process of restoring the original message from the cipher text by reversing the steps of encipherment with full knowledge of the details is called *deciphering*.

## 2.3 Substitution Cipher

**Substitution** is a function which uses a set of rules to transform elements of a sequence into a new sequence using a set of rules which "translate" from the original sequence to its transformation.

In the encryption methods which uses **substitution**, every character of the *plaintext* is replaced by another character in the *ciphertext*. For the decryption, the reverse substitution has to be performed. The easiest substitution is given when each character is replaced by exactly one other character. This encryption can be broken with statistical methods because in every language characters appear with a particular probability.
Examples: Caesar Cipher (a Monoalphabetic Cipher), Vigenère cipher (a Polyalphabetic Cipher), Dancing

Men Cipher [*The Adventure of Dancing Men* by Arthur Conan Doyle] (a symbol substitution cipher), Pig-Pen cipher (a symbol substitution cipher), Hill Cipher (a polygraphic cipher[1])

### 2.3.1 Caesar Cipher

Consider the 26 alphabets of English Language and associate each letter with the number representing its position in normal sequence:

| Plain Alphabet | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Corresponding Number | 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 |

One of the earliest cryptographic systems known was used by Roman dictator Gaius Julius Caesar and is referred to by his namethe *Caesar Cipher*. It consisted of a replacement of each letter of the message by the letter three places beyond it in the normal alphabet.

Now, to encipher the message, `BEWARE OF ZOMBIES`, after deleting spaces, we proceed as follows (each step is illustrated below):

1. Replace each letter by the number to which it corresponds
2. Add 3 to each of these numbers and reduce each mod26.
3. Replace the resulting numbers by their letter equivalents in the letter-number correspondences

| Plaintext | B | E | W | A | R | E | O | F | Z | O | M | B | I | E | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Corresponding Numbers | 1 | 4 | 22 | 0 | 17 | 4 | 14 | 5 | 25 | 14 | 12 | 1 | 8 | 4 | 18 |
| Adding 3 (mod 26) | 4 | 7 | 25 | 3 | 20 | 7 | 17 | 8 | 2 | 17 | 15 | 4 | 11 | 7 | 21 |
| Ciphertext | E | H | Z | D | U | H | R | I | C | R | P | E | L | H | V |

There is nothing special about the number 3 as the amount of shift between the cipher sequence and the plain sequence. We can choose any number at all, so long as an understanding has been made with our correspondent about how the enciphering was to be accomplished. Given the number of shifts (can call it *key*), the substitution alphabet can be constructed and used for encipherment or for decipherment. A substitution alphabet in which both the plain and the cipher sequences are the normal alphabet (with the cipher sequence shifted a specific number of places) is called a *direct standard alphabet.*

In the equivalent numerical process, expressible as $\mathcal{C} = \mathcal{P} \oplus \mathcal{K}$, the number of places of shift ($\mathcal{K}$) is the number to be added modulo 26 to the numerical equivalent of each plain language character ($\mathcal{P}$) to determine its cipher replacement ($\mathcal{C}$). To decipher the ciphertext simply perform the inverse process, i.e. subtract $\mathcal{K}$ from $\mathcal{C}$ modulo 26.

**Cryptanalysis** (Primary Frequency Analysis)**.** .
Suppose we are not the intended receiver, but we somehow intercept the message:

`EHZDUHRICRPELHV`

Now we don't know the *key* needed to decipher this message. If we would know that this is Caesar Cipher then, we can easily check the 25 possible keys and decipher the message.
We now examine the possibility of a different method of procedure which does not assume, but instead will prove, that the system of encipherment used a shift of the normal alphabet. This method is based on a fundamental property of language, the relative frequencies of occurrence of the different letters of the alphabet.

| Alphabet | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Percentage | 8.2 | 1.5 | 2.8 | 4.3 | 12.7 | 2.2 | 2.0 | 6.1 | 7.0 | 0.2 | 0.8 | 4.0 | 2.4 | 6.7 | 7.5 | 1.9 | 0.1 | 6.0 | 6.3 | 9.1 | 2.8 | 1.0 | 2.4 | 0.2 | 2.0 | 0.1 |

Table 2.1: A standard table of relative frequencies. The table was compiled by H. Beker and F. Piper, and originally published in *Cipher Systems: The Protection of Communication*

In general, short texts are likely to deviate significantly from the standard frequencies, and if there are fewer than a hundred letters, then decipherment will be very difficult. On the other hand, longer texts are

[1] A system of cryptography in which a group of $n$ plain text letters is replaced as a unit by a group of $n$ cipher letters is called a *polygraphic system.*

more likely to follow the standard frequencies, although this is not always the case.

For given cipher text we calculate relative frequency: Let us assume that the commonest alphabets in

| Alphabet | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occurrences | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| Percentage | 0 | 0 | 6.7 | 6.7 | 13.3 | 0 | 0 | 20 | 6.7 | 0 | 0 | 6.7 | 0 | 0 | 0 | 6.7 | 0 | 13.4 | 0 | 0 | 6.7 | 6.7 | 0 | 0 | 0 | 6.7 |

Table 2.2: Table of relative frequencies for `EHZDUHRICRPELHV`

the ciphertext probably represent the commonest letters in the English alphabet, but not necessarily in the right order. Nevertheless we can assume `H` = `E`, since both are alphabets with highest frequency in both Table 2.1 and Table 2.2. Thus, we get $\mathcal{K} = 3$ and we broke the cipher!

### 2.3.2 Vigenère Cipher

The first step in encipherment is to draw up a so-called Vigenère[2] square, as shown in Table 2.3, a plain alphabet followed by twenty-six cipher alphabets, each shifted by one letter with respect to the previous alphabet. To unscramble the message, the intended receiver needs to know which row of the Vigenère square

| Plain Alphabet | 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 | 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 | P | 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 |
| C | 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 |
| I | 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 |
| P | 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 |
| H | 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 |
| E | 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 |
| R | 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 |

Table 2.3: Vigenère square,

has been used to encipher each letter, so there must be an agreed system of switching between rows. This is achieved by using a *keyword*.

Now, let's again encipher the previous message, `BEWARE OF ZOMBIES`, after deleting spaces, with keyword `ELEPHANT` we proceed as follows (each step is illustrated below):

1. Select a keyword.

[2]Pronounced "vidjenair"

2. The keyword is spelled out above the message and plaintext is written in columns of width equal to length of keyword. In the language of modular arithmetic every letter whose position number in the original message is congruent to $a \pmod{b}$, where $b$ is length of keyword, would be enciphered by the alphabet designated by the $a$-th letter of the keyword.

3. To encipher each letter begin by identifying the key letter above it, which in turn defines a particular row in the Vigenère square.

| Keyword | E | L | E | P | H | A | N | T |
|---|---|---|---|---|---|---|---|---|
| Plaintext | B | E | W | A | R | E | O | F |
| Ciphertext | F | P | A | P | Y | E | B | Y |
| Plaintext | Z | O | M | B | I | E | S | |
| Ciphertext | D | Z | Q | Q | P | E | F | |

**Cryptanalysis** (Secondary Frequency Analysis)**.** .
The fact that a letter that appears several times in the ciphertext can represent a different plaintext letter on each occasion generates tremendous ambiguity for the cryptanalyst. Equally confusing is the fact that a letter that appears several times in the plaintext can be represented by different letters in the ciphertext. Thus Vigenère cipher is invulnerable to primary frequency analysis (try yourself).

Consider the ciphertext:

FPAPYEBYDZQQPEF

Suppose we know that it was enciphered using the Vigenère cipher, but we know nothing about the original message, and the keyword is a mystery. The first stage in cryptanalysis is to look for sequences of letters that appear more than once in the ciphertext.

There are two ways that such repetitions could arise. The most likely is that the same sequence of letters in the plaintext has been enciphered using the same part of the key. Alternatively, there is a slight possibility that two different sequences of letters in the plaintext have been enciphered using different parts of the key, coincidentally leading to the identical sequence in the ciphertext. If we restrict ourselves to long sequences, then we largely discount the second possibility, and in this case we shall consider repeated sequences only if they consist of four letters or more.

After listing which sequences repeat themselves and the spacing between these repetitions we identify the factors of the spacing and these factors are possible lengths of the keyword.

Unfortunately, our intercepted message has no repeated sequence[3], but it at least has repeated alphabets, Table 2.4 is a log of such repetitions, along with the spacing between the repetition and possible length of key.

| Repeated Sequence | Repeat Spacing | Possible length of key (or factors) |
|---|---|---|
| F | 14 | 2,7,14 |
| P | 2 | 2 |
| | 9 | 3,9 |
| | 11 | 11 |
| E | 8 | 2,4,8 |
| Y | 3 | 3 |

Table 2.4: Repetitions and spacings in FPAPYEBYDZQQPEF

Now, 2,3,4,7,8,9,11 and 14 are possible lengths of keyword (ignore consecutive repetition). From infinitely possible lengths of keywords we have identified the finitely many cases which we need to check.

Now what remains is to identify the exact keyword. If $k$ is length of the keyword then let, $L_1L_2\cdots L_k$ represents the keyword. What we need to do is to write cipher text in column of length $k$ and write keyword above the column. Then apply primary frequency analysis to alphabets below each alphabet of keyword and check for existence of keyword.

As an illustration observe the case when $k = 2$

[3]This is worst case scenario, when repeated sequences are of length 1

| Keyword | $L_1$ | $L_2$ |
|---|---|---|
| | F | P |
| | A | P |
| | Y | E |
| Ciphertext | B | Y |
| | D | Z |
| | Q | Q |
| | P | E |
| | F | |

Now we have two independent cipher texts, FAYBDQPF and PPEYZQE to be checked by *primary frequency analysis*. We won't be able to find $\mathcal{K}$ for both of these!

So on check for $k = 3, 4, 7$, for $k = 8$ we get:

| Keyword | $L_1$ | $L_2$ | $L_3$ | $L_4$ | $L_5$ | $L_6$ | $L_7$ | $L_8$ |
|---|---|---|---|---|---|---|---|---|
| Ciphertext | F | P | A | P | Y | E | B | Y |
| | D | Z | Q | Q | P | E | F | |

Then the frequency distribution for each alphabet of keyword is:

| Alphabet | | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Percentage | $L_1$ | 0 | 0 | 0 | 50 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $L_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 |
| | $L_3$ | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $L_4$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $L_5$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 0 |
| | $L_6$ | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $L_7$ | 0 | 50 | 0 | 0 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $L_8$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |

Now compare above results with Table 2.1.

We observe that as per standard frequency distribution, C & U; G & Y; J & X; M & W; Q & Z are the ones with equal frequency. Now we need to check, if there is a pair with difference between corresponding numbers same as any of cipher alphabets. Voil! for $L_3$, the gap between A and Q is $16 \ominus 0 = 16$ or $0 \ominus 16 = 10$ and the gap between M and W is $22 \ominus 12 = 10$ or $12 \ominus 22 = 16$. Thus for gap of 10, we get: A= W and Q = M, leading to $L_3 = E$.

Also, since E is the alphabet with highest frequency in Table 2.1, for $L_6$ (not $L_8$ since it has only one alphabet), we put E = E, thus $L_6 = A$.

Further, {F, G, M, P, W, Y}; {H, R, S}; {I, N}; {J, Q, X, Z} ;{K, V}; are set of alphabets with approximately equal frequency. But no pair from these sets will work. (check yourself).

Now consider the pairs with difference in percentage approximately 1. Even then we fail!

Thus by above arguments we can deduce only that: *if the keyword is of 8 letters then it will be of form* $L_1L_2EL_4L_5AL_7L_8$. Then we will have to use brute-force method, by checking all $26^6 = 308915776$ possibilities by computer.

For breaking a Vigenère cipher by frequency analysis the length of the cipher text alone is not the crucial part. What really matters is the proportion $\frac{\text{ciphertext length}}{\text{key length}}$, as this indicates how many characters of the clear text are enciphered by the same character of the key.

As seen in above example, frequency analysis based on monograms (single letters) will definitely fail. We can try to break the cipher by using frequency analysis of bigrams, trigrams or quadgrams instead, see [5] for illustration.

Both, *Frequency Analysis*[4] and *Index of Coincidence*[5], only work if the cipher text is much longer than the key.

[4] for more details refer pp. 64-73 of [5]

[5] this is a statistical way of dealing with the problem, refer pp. 61-74 of [4]

Another approach is using *word dictionaries*[6], this tool can break extremely short Vigenère ciphers. It requires that the clear text as well as the keyword consists of words only which are found in the dictionary.

If the message is shorter than the key, then the Vigenère cipher is essentially the *one-time pad*, which is unbreakable for a random key (proven by *Claude Shannon*). If the key is not random, then you may get some information on the plaintext.

## 2.4 Transposition Cipher

**Permutation** of a set $X$ is a bijective function $\sigma : X \to X$ that for each element $x \in X$ assigns a unique value $\sigma(x) \in X$.
A transposition is a permutation of two elements and any permutation is also a product of transpositions.

In **transposition cipher** the order of the characters is rearranged but the actual characters are not changed. Transposition can be broken by statistical methods because the pairs of successive characters in a normal language have typical likelihood. Other pairs do occur much less often. If the messages is short, some characters may not appear thus it is possible to say which words do not exist in the text.
Examples: Rail Fence Cipher or Scytale Cipher (a matrix transposition cipher)

Note the small difference between transposition ciphers and substitution cipher, substitution ciphers replace each letter with a different letter or symbol to produce the ciphertext, in a transposition cipher, the letters are just moved around.

### 2.4.1 Rail Fence Cipher

The text is written with alternate letters on each of $n$ rows, and then read row by row.

As an example, consider the message `BEWARE OF ZOMBIES`, and $n = 3$, then we get:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Row 1* | B | | | | R | | | | Z | | | | I | |
| *Row 2* | | E | | A | | E | | F | | O | | B | | E |
| *Row 3* | | | W | | | | O | | | | M | | | | S |

Thus the cipher text will be: `BRZIEAEFOBEWOMS`.

Decrypting the message is easy if the number of rows and row boundaries are known. Just write down the rows in order:
`BRZI`
`EAEFOBE`
`WOMS`
and reconstruct the "rails" of the fence:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | | | | R | | | | Z | | | | I | | |
| | E | | A | | E | | F | | O | | B | | E | |
| | | W | | | | O | | | | M | | | | S |

If no row boundaries are present, it is not difficult to reconstruct the fence, as long as you know how many rows there are and in which order they are written[7].

We can assign a number corresponding to position of each alphabet of plaintext to get:

| Plaintext | B | E | W | A | R | E' | O | F | Z | O' | M | B | I | E'' | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Position Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

Now since $n = 3$ we get position numbers for ciphertext as:

| Ciphertext | B | R | Z | I | E | A | E' | F | O' | B | E'' | W | O | M | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Position Number | 1 | 5 | 9 | 13 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 3 | 7 | 11 | 15 |

Writing this under the plain text message we have:

[6] see here: http://www.sichere.it/vigenere_ tool.php?language=EN

[7] For added complexity, a key can be used to indicate the order of reading the rows. For example, the key 213, gives `EAEFOBEBRZIWOMS` as ciphertext.

```
plaintext position numbers  1 2 3  4  5 6 7 8  9  10 11 12 13 14 15
ciphertext position numbers 1 5 9 13 2 4 6 8 10 12 14  3  7  11 15
```

Now apply the cycle procedure. We end up with following disjoint cycles:

```
(2, 5)(3, 9, 10, 12)(4, 13, 7, 6)(11, 14)
```

In most cases of transposition ciphers, the resulting permutation consists of several cycles with no particular relation between their lengths. The total number of letters represented in these cycles is less than the length of the message by the number of letters whose positions are unchanged, i.e. are in cycles consisting of just one number.

The advantages of using the permutation concept for transposition are mainly theoretical. The permutation which expresses the result of a transposition is a function not only of the method but also of the length of the message. Consequently the permutation procedure is not really practical for enciphering and deciphering messages of differing lengths.

In general, it can be shown that $n$ applications of the transposition yield a decimation at interval $n$ of the original permutation. This result now permits us to answer the question: *How many times does a transposition have to be applied before the cipher becomes identical with the original plain text?* Consider a single cycle of the original permutation, say of length $x$. Such a cycle is reduced to cycles of one letter only if decimated at any interval which is a multiple of $x$. In other words, all the letters of that cycle return to their original positions if the transposition is repeated a multiple of $x$ times. Since such a statement is true for every cycle:

**Theorem 2.4.1** (Order of a Permutation)**.** *The number of times a transposition must be applied to return to the original plain language message is the least common multiple of the lengths of all the cycles included in it.*

*Proof.* Refer pp. 102 of [3]. □

Thus our example above, the ciphertext will become a plaintext if transposition is applied 4 times, since $\text{lcm}(4, 2) = 4$.

*Thus the disjoint permutation cycle can also be used as key*, instead of number of rows and order of rows.

**Cryptanalysis** (Brute Force)**.** .
If you know (or suspect) that a message was encrypted with a Rail Fence Cipher, it can easily be deciphered by brute force because the letters break into rows according to certain fixed patterns based on the number of rows in the key.

For example, in `BRZIEAEFOBEWOMS`, we try out row lengths:

1. If $n = 2$, then, then letters $1, 3, 5, \ldots$ of the plaintext are in row one and letters $2, 4, 6, \ldots$ are in row two. Thus possible keys are 12 or 21, with 8 alphabets in row 1 and 7 alphabets in row 2.

```
key: 12
       B   R   Z   I   E   A   E   F
         O   B   E   W   O   M   S
                 OR
key: 21
       F   O   B   E   W   O   M   S
         B   R   Z   I   E   A   E
```

   None of these make any sense!

2. If $n = 3$, then letters $1, 5, 9, \ldots$ are in row one, letters $2, 4, 6, 8, \ldots$ are in row two and letters $3, 7, 11, \ldots$ are in row three, with 4 alphabets in row 1 and 7 alphabets in row 2 and 4 alphabets in row 3.

   Now we will have to decide possible spacings and ordering out of $3! = 6$ possibilities:

Case 1: `BRZIEAE FOBE WOMS`
Since, two of rows are of same length, we have two ways of forming rail fence:

i. row 1 : FOBE and row 3: WOMS

```
F       O       B       E
  B   R   Z   I   E   A   E
    W       O       M       S
```

Meaningless!

ii. row 1 : WOMS and row 3: FOBE

```
W       O       M       S
  B   R   Z   I   E   A   E
    F       O       B       E
```

Meaningless!

Case 2: BRZI EAEF OBEWOMS

Since, two of rows are of same length, we have two ways of forming rail fence:

i. row 1 : BRZI and row 3: EAEF

```
B       R       Z       I
  O   B   E   W   O   M   S
    E       A       E       F
```

Meaningless!

ii. row 1 : EAEF and row 3: BRZI

```
E       A       E       F
  O   B   E   W   O   M   S
    B       R       Z       I
```

Meaningless!

Case 3: BRZI EAEFOBE WOMS

Since, two of rows are of same length, we have two ways of forming rail fence:

i. row 1 : WOMS and row 3: BRZI

```
W       O       M       S
  E   A   E   F   O   B   E
    B       R       Z       I
```

Meaningless!

ii. row 1 : BRZI and row 3: WOMS

```
B       R       Z       I
  E   A   E   F   O   B   E
    W       O       M       S
```

*Meaningful!*

And we are done!

# Chapter 3

# Permutation Groups and Enigma Ciphers

Here I will primarily discuss the work of Marian Rejewski (see Part-I of [9]).[1] Observe that in Enigma, each rotor and reflector performed a *substitution cipher* and plugboard and ring settings[2] performed *transposition cipher*.

## 3.1 Enigma Algorithm

In 1931 and 1932 the French cryptographer *Gustave Bertrand* obtained the algorithm used by the German Enigma from a spy, Hans-Thilo Schmidt, known by the code name Asche. In Nazi Germany military till 1938, the following regulations were obeyed (see [19]):

1. For a message to be correctly encrypted and decrypted, both the sender and receiver needed to set up their Enigma in exactly the same way. These settings were distributed in key sheets. The key sheets were distributed on beforehand, and contained the basic settings for a whole month, per day. In general, the key sheets were in the custody of an officer, responsible for setting up the *rotor order*, *plugboard* and *ring settings*. After setup, he could lock the machine front panel with a key.

2. The operator could only select the *rotor orientation* (or rotor start position). Then he chose the individual key for a message, consisting of three letters which were ciphered twice, thus obtaining six letters placed at the beginning of the message.

In each place of the message they form a one-to-one transformation of the set of letters onto itself and hence they are permutations. These permutations, denoted subsequently by letters $A, B, C, D, E$ and $F$, are unknown to the cryptologist.

As shown in Figure 4, $S$ represents the plugboard, $N$ represents the right-hand, or fast, rotor; $M$ represents the middle rotor; $L$ represents the left-hand, or slow, rotor; and $R$ represents the reflector.

We can think of $S, N, M, L$ and $R$ as permutations. In addition to these permutations, we have a permutation $P$, corresponding to the motion of the fast rotor which moves forward on letter each time a key is pressed.

$$P = \texttt{(abcdefghijklmnopqrstuvwxyz)}$$

Thus, the unknown permutations from $A$ to $F$ can be represented in the form (composing permutations from left to right):

$$A = SPNMLRL^{-1}M^{-1}N^{-1}P^{-1}S^{-1} = (SPNML)R(SPNML)^{-1}$$

$$B = SP^2NMLRL^{-1}M^{-1}N^{-1}P^{-2}S^{-1} = (SP^2NML)R(SP^2NML)^{-1}$$
$$C = SP^3NMLRL^{-1}M^{-1}N^{-1}P^{-3}S^{-1} = (SP^3NML)R(SP^3NML)^{-1}$$
$$D = SP^4NMLRL^{-1}M^{-1}N^{-1}P^{-4}S^{-1} = (SP^4NML)R(SP^4NML)^{-1}$$
$$E = SP^5NMLRL^{-1}M^{-1}N^{-1}P^{-5}S^{-1} = (SP^5NML)R(SP^5NML)^{-1}$$
$$F = SP^6NMLRL^{-1}M^{-1}N^{-1}P^{-6}S^{-1} = (SP^6NML)R(SP^6NML)^{-1}$$

[1] There are excellent articles on this topic by *Jiří Tůma* (see [7]) and *Chris Christensen* (see [8])

[2] Only the position of the notches on the right hand and middle rotors contributed to cryptographic security of Enigma.

Whether the middle and left-hand rotors move or not, an Enigma permutation is always a conjugate of the reflector. So, an Enigma permutation is always a product of 13 disjoint transpositions.

The reflector was "half a rotor." There were only 26 contacts on the right-hand side of the reflector. Internally, the 26 contacts were joined in pairs by wires to create a permutation consisting of 13 disjoint transpositions. The fact that every Enigma permutation is a product of 13 disjoint transpositions is what permits Enigma to encipher and decipher in the same mode. Every Enigma permutation is self-reciprocal.

## 3.2 Number of Possible Keys

We observed in last chapter, that cryptanalysis becomes more and more difficult with increase in number of possible keys. Using the information provided in introduction of this report, for 1930s German Army model of Enigma:

$$\begin{aligned}\#\text{possible keys} = (\#\text{possible plugboard settings}) \times (\#\text{possible rotor orders}) \\ \times (\#\text{possible rotor orientations}) \times (\#\text{possible ring settings})\end{aligned}$$

$$\Rightarrow \#\text{possible keys} = \left(\frac{26!}{(26-(2\times 6))!} \times \frac{1}{6!\times 2^6}\right) \times (3!) \times (26\times 26\times 26) \times (1\times 26\times 26)$$

$$\Rightarrow \#\text{possible keys} = (100391791500) \times (6) \times (17576) \times (676)$$

$$\Rightarrow \#\text{possible keys} = 7,156,755,732,750,624,000$$

## 3.3 Weakness of Enigma

Thus the individual keys for the given day had the following properties:

- All individual message keys were ciphered in the same basic position unknown to the cryptologist;
- Each individual key was ciphered twice, so that the first letter meant the same as the fourth, the second the same as the fifth and so on.

If a sufficient number of messages (approximately 80) of the same day are available, then, in general, all alphabet letters are present in their six initial places.

Being self-reciprocal was also a weakness. The reflector permutation guarantees that every Enigma permutation is self-reciprocal, but it also guarantees that no letter can be enciphered as itself.

But the transitions from the first letter of each message to the fourth one, from the second to the fifth and from the third to the sixth form also permutations which, contrary to the individual ones, are entirely known to the cryptologist since they are the products $AD$, $BE$, $CF$ of the above-mentioned permutations and are also given by the formulas:

$$AD = SPNMLRL^{-1}M^{-1}N^{-1}P^{-1}S^{-1}SP^4NMLRL^{-1}M^{-1}N^{-1}P^{-4}S^{-1}$$
$$BE = SP^2NMLRL^{-1}M^{-1}N^{-1}P^{-2}S^{-1}SP^5NMLRL^{-1}M^{-1}N^{-1}P^{-5}S^{-1}$$
$$CF = SP^3NMLRL^{-1}M^{-1}N^{-1}P^{-3}S^{-1}SP^6NMLRL^{-1}M^{-1}N^{-1}P^{-6}S^{-1}$$

The first part of cryptanalysis is, in principle, to solve this set of equations in which the left-hand sides are known and the permutation $P$ and its powers on the right-hand sides as well, whereas the permutations $S, L, M, N, R$ are unknown. Since in this form the set is certainly unsolvable, we have to simplify it.

Since the cycle structure is not affected by the plugboard, the first step is to let

$$P_\alpha = P^\alpha NMLRL^{-1}M^{-1}N^{-1}P^{-\alpha}$$

where each of $P_\alpha$, $\alpha = \{1,2,3,4,5,6\}$, is determined by only rotor orientation and rotor order. Thus:

$$AD = SP_1S^{-1}SP_4S^{-1} = SP_1P_4S^{-1}$$
$$BE = SP_2S^{-1}SP_5S^{-1} = SP_2P_5S^{-1}$$
$$CF = SP_3S^{-1}SP_6S^{-1} = SP_3P_6S^{-1}$$

Consider following theorem from elementary permutation theory:

**Theorem 3.3.1.** *Two permutations in $S_n$ are conjugate if and only if they have the same cycle structure or decomposition.*

*Proof.* See Q20 in Chapter 1 of this report. □

From this theorem, the disjoint cycle structure of $AD$ is the same as it would be if there were no plugboard. Similarly, the disjoint cycle structure of $BE$ and $CF$ is not affected by the plugboard. Thus we can find rotor orientation and rotor order without considering $100,391,791,500$ possible plugboard connections. Momentarily we can also ignore the 676 ring settings, thus we are only left with $105,456$ rotor orders and rotor orientations.

We can assume that the middle and left-hand rotor did not turn during the six permutation $A, B, \ldots, E, F$. This was a reasonable assumption because the middle rotor turned only once in 26 turns of the right-hand rotor. If a turnover did occur, the above method will not work.

## 3.4 Rejewski's Insight

Now we aim to get disjoint unknown permutations from $A$ to $F$ from known products $AD$, $BE$, $CF$. As seen in previous section, the unknown permutations consists only of transpositions, and the expressions $AD$, $BE$, $CF$ are their products. Rejewski proved following theorems to handle above problem.

**Theorem 3.4.1** (Rejewski's Theorem)**.** *A permutation of even degree includes cycles of same length in even numbers if and only if this permutation is product of two permutations consisting only of disjoint transpositions.*

*Proof.* We need to prove two parts:

Part 1: *Two permutations of same degree consist only disjoint transpositions ⇒ their product consist of an even number of disjoint cycles of the same length.*

Let $X$ and $Y$ stand for the permutations to be multiplied and let their degree be $2n$, since there are only disjoint transpositions.

If in the permutation $X$ a transposition identical with a transposition in $Y$, for example, `(ab)`, incidentally occurs, then in the product $XY$ a pair of single-letter cycles `(a)(b)` will be observed i.e. we will get an identity element. With respect to transpositions, identical in the two permutations, the theorem is thus true.

After rejecting identical transpositions we can assume, without loss of generality, that the follow transpositions occur:

$$X = (a_1a_2)(a_3a_4)(a_5a_6)\cdots(a_{2k-3}a_{2k-2})(a_{2k-1}a_{2k})$$

$$Y = (a_2a_3)(a_4a_5)(a_6a_7)\cdots(a_{2k-2}a_{2k-1})(a_{2k}a_1)$$

Indeed, the initial letter $a_1$ must finally appear in the permutation $Y$. When we perform the operation of multiplying $XY$, we will always get two cycles of the same length $k \leq n$:

$$\Rightarrow XY = (a_1a_3a_5\ldots a_{2k-3}a_{2k-1})(a_{2k}a_{2k-2}\ldots a_6a_4a_2)$$

If in this way not all letters of the permutation are exhausted, we continue our procedure to exhaust all the letters.

Simultaneously we note that the letters of a given transposition are always observed in two different cycles of the same length in the permutation $XY$ and if two letters appearing in two different cycles of the same length in the permutation $XY$ belong to the same transposition, then their neighbouring letters (the left neighbour and the right one) belong to the same transposition. Thus completing proof of this part.

Part 2: *A permutation of even degree includes cycles of same length in even numbers ⇒ this permutation is product of two permutations consisting only of disjoint transpositions*[3]

[3] Recall that each of $AD, BE$ and $CF$ satisfy the conditions of this part of theorem.

Given: $XY = (a_1a_3a_5 \dots a_{2k-3}a_{2k-1})(a_{2k}a_{2k-2} \dots a_6a_4a_2)$
Then we can write:(one obvious factor, though many other possible)

$$X = (a_1a_2)(a_3a_4)(a_5a_6) \cdots (a_{2k-3}a_{2k-2})(a_{2k-1}a_{2k})$$

$$Y = (a_2a_3)(a_4a_5)(a_6a_7) \cdots (a_{2k-2}a_{2k-1})(a_{2k}a_1)$$

Thus proving other part of theorem.

Combining both parts we complete proof of the theorem. □

The above theorem on the product of transpositions does not lead us to the point we are aiming to get at, it brings us, however, to the proximity of it. For an illustration of application of above theorem refer pp. 263-268 of [8].

# Chapter 4

# Finding the Key

According to the work of Marian Rejewski, as seen in last chapter, we now are left with a lesser number of *probable keys*, but still we need to find the *key*, i.e. the exact initial settings[1] of the machine for a given day in a reasonable period of time. Thus following techniques (based on brute-force and frequency analysis) were created by Polish mathematicians to search for initial keys from the probable ones.

## 4.1 Cyclometer

Marian Rejewski invented a machine, which consisted of, in effect, two Enigma machines side by side with their right hand wheels offset by three places. He had worked out his theory of Characteristics and by using the Cyclometer had constructed a lookup table with 105,456 entries, a characteristic for each three wheel start combination for all possible wheel orders for three wheels. $(26 \times 26 \times 26 \times 3!)$. The characteristics were built up by following loops of letters from firstly the first and fourth positions in the double enciphered message key. Then between the second and fifth positions and finally between the third and sixth positions. For more details and examples refer [16].

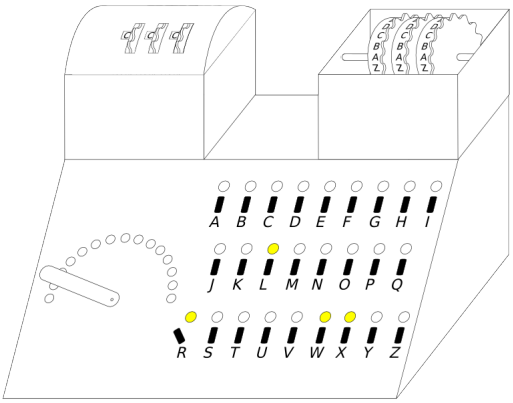


Figure 4.1: Cyclometer [Drawn using Inkscape] (http://commons.wikimedia.org/wiki/File:Cyclometer4.png)

## 4.2 Zygalski Sheets

Henryk Zygalski realised that the analysis of the vast amount of information required could be achieved by a grill method using perforated sheets. The sheet procedure involves working through each of the six possible wheel orderings for three wheels and for each wheel order working through the 26 possible left hand wheel ring letters, 156 tries in all, but on average only half before the answer is found. There were sheets prepared for each left wheel letter for each wheel order. There were sheets prepared for each left wheel letter for each wheel order. Each sheet contained four squares of 26 by 26 i.e. two alphabets along the top and down the side.The $26 \times 26$ matrix represented the 676 possible starting positions of the middle and left rotors and was duplicated horizontally and vertically: **a-z**, **a-y**. One can try overlaying Zygalski sheets at *http://www.codesandciphers.org.uk/virtualbp/poles/zygalski.htm*.

[1] I will primarily discuss the work published by Marian Rejewski (see Part-II of [9])

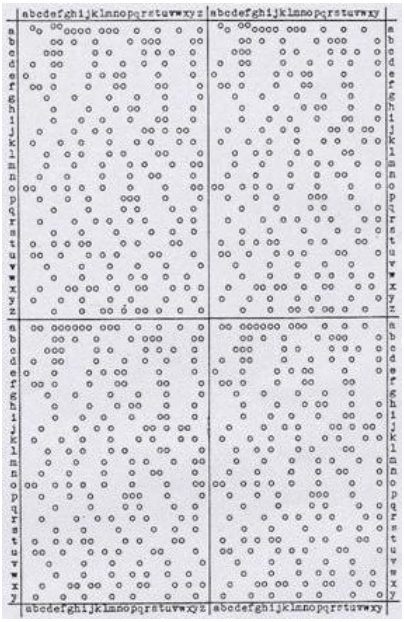

Figure 4.2: A Zygalski Sheet (see [9])

## 4.3 Bomba Kryptologiczna (Cryptologic Bomb)

When Marian Rejewski, Jerzy Różycki[2] and Henryk Zygalski had been studying the double enciphered message settings in order to construct Rejewski's characteristics, cases had been noticed where the same enciphered letter occurred in either the 1st and 4th, or 2nd and 5th, or 3rd and 6th positions in the enciphered message settings.

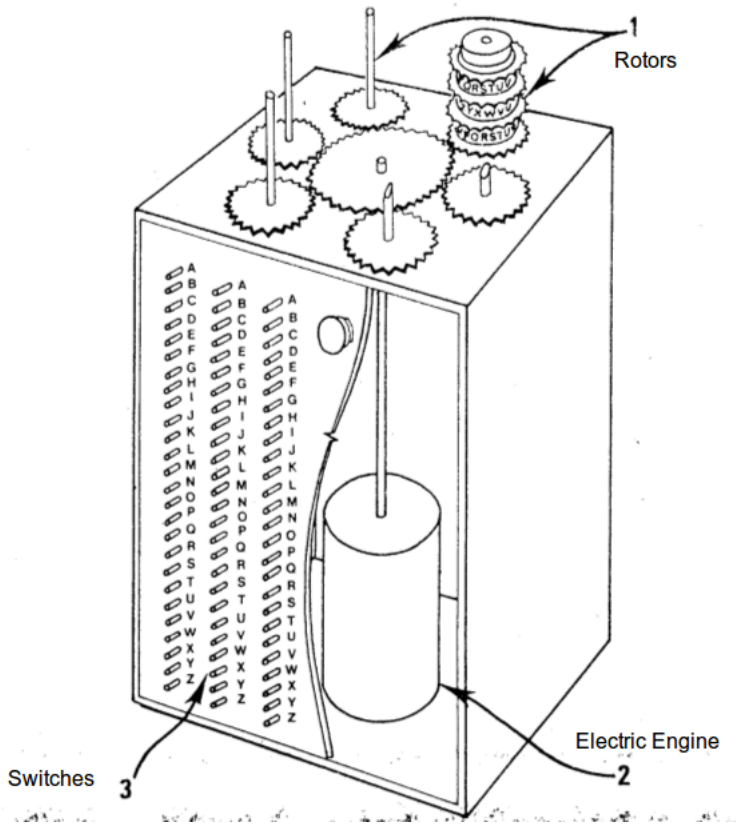


Figure 4.3: For clarity shown in the upper part of the Bomba Kryptologiczna only one set of rotors encryption (see [10])

[2] Jerzy Różycki invented the "clock" method, which sometimes made it possible to determine which of the machine's rotors was at the far right, that is, in the position where the rotor always revolved at every depression of a key.

These positions, which were called females, corresponded to positions at which the same letter had been keyed by the German operator into the Enigma machine because of the repeat of the three letter message key. Rejewski also had the idea for a mechanical method for finding the Enigma ring settings from the females in the double enciphered message settings. The idea was to rotate six sets of enigma wheels in synchronism with each set being one fast wheel position in advance of the preceding one so that the six positions corresponding to the double encipherment of the message setting could be examined simultaneously looking for repeating enciphered letters. Six of these machines were required, each set with one of the six possible wheel order for the possible three wheels in the Enigma machine. The Bombas were not very reliable and Zygalski's sheets produced better results. For more details refer [17].

# Conclusion

In Chapter-1 we discussed questions from basic concepts regarding permutation groups needed to understand Rejewski's Theorem.

Further in Chapter-2 we saw that, the weakness of Caesar cipher is that it uses one *fixed* shift for whole message. Then Vigenère cipher was created by eliminating that weakness of Caesar cipher, using a fixed keyword so that consecutive alphabets have different shifts. Though for short messages Vigenère cipher turns out to be very secure, but for long messages there were repetitions of shifts after fixed intervals, which turned out to be the weakness this cipher.

Then *Scherbius* invented Enigma machine by eliminating weakness of Vigenère cipher, using rotors to keep changing keywords (but of fixed length) for a given message. But unlike Vigenère cipher, since the key length was fixed, the brute-force method was successful.

Then Nazi Germany Army, added plugboards to Enigma, which, unlike the statistical complications caused by substitution cipher systems, caused complications of form of transposition cipher. These complications were handled in an elegant way by Rejewski as seen in Chapter-3 by using theory of permutation groups to reduce number of probable keys and making addition of plugboard as well as rotations nearly ineffective. Also Polish mathematicians developed novel methods of finding initial key as seen in Chapter-4.

*Reflector was responsible for ease of usage of Enigma by allowing same "key" (initial settings), to be used to encipher and decipher the message. But at same time, reflector ensured that no alphabet is taken to itself (so as to ensure disjoint cycles), which turned out to be biggest weakness of Enigma.*

## Acknowledgement

I would like to show my gratitude to Mitul Verma, Research and Teaching Assistant, Ashoka University (Haryana) for sharing his pearls of wisdom with me during the course of this project.

# Bibliography


[1] I. N. Herstein : *Topics in Algebra*, John Wiley & Sons, Xerox Corporation (1975)

[2] David S. Dummit & Richard M. Foote : *Abstract Algebra*, John Wiley & Sons, Inc., ISBN: 0-471-43334-9 (2004)

[3] Joseph A. Gallian : *Contemporary Abstract Algebra*, Brooks/Cole, Cengage Learning, ISBN: 978-0-547-16509-7 (2010)

[4] Abraham Sinkov : *Elementary Cryptanalysis - A Mathematical Approach*, The Mathematical Association of America, New Mathematical Library Series, Volume 22: ISBN: 0-88385-622-0 (1966)

[5] Simon Singh : *The code book - how to make it, break it, hack it, crack it*, Random House, Inc., eISBN: 0-375-89012-2 (2001)

[6] David Kahn : *The Codebreakers - The story of secret writing*, The Macmillan Company (1973)

[7] Jiří Tůma : *Permutation Groups and the Solution of German Enigma Cipher*, https://cryptocellar.web.cern.ch/cryptocellar/Enigma/tuma2003.pdf

[8] Chris Christensen : *Polish Mathematicians Finding Patterns in Enigma Messages*, Mathematics Magazine, Vol. 80, No. 4 (Oct., 2007), pp. 247-273, http://www.jstor.org/stable/27643040

[9] Marian Rejewski : *An application of the theory of permutations in breaking the Enigma cipher*, Zastosowania Mathematyki (Applicationes Mathematicae), XVI, 4 (1980)

[10] Marian Rejewski : *Matematyczne podstawy rozwazania niemieckiego szyfru maszynowego Enigma (Mathematical foundations of the German Enigma cipher machine solutions)*, http://www.spybooks.pl/en/archiwum_ tekst.html?id=b05aa114591085b7a8012516bc3533958fea

[11] Eric W. Weisstein : *Substitution System*, http://mathworld.wolfram.com/SubstitutionSystem.html

[12] WebLearn Archive : *Substitution and Transposition*, Fachbereich Elektrotechnik & Informatik, Hochschule Bremen (University of Applied Sciences), http://www.weblearn.hs-bremen.de/sonstiges/Diplomarbeiten/Secure_ Mechanisms/Secure_ Mechanisms_ html/node17.html

[13] Khan Academy : *Journey into Cryptography*, https://www.khanacademy.org/computing/computer-science/cryptography/ciphers/a/ciphers-vs-codes

[14] National Archives and Records Administration : *Teaching With Documents - The Zimmermann Telegram*, http://www.archives.gov/education/lessons/zimmermann/decoding-activity.html

[15] The Late Tony Sale's Codes and Ciphers Website : *The Breaking of Enigma by the Polish Mathematicians*, http://www.codesandciphers.org.uk/virtualbp/poles/poles.htm

[16] The Late Tony Sale's Codes and Ciphers Website : *Investigating Rejewski's Characteristics*, http://www.codesandciphers.org.uk/virtualbp/poles/cyclom.htm

[17] Crypto Museum : *The Polish Bomba*, http://www.cryptomuseum.com/crypto/bombe/

[18] Sophia Knight : *The Rail Fence Cipher*, http://www.cs.trincoll.edu/ crypto/historical/railfence.html

[19] Enigma Message Procedures : *The Heer and Luftwaffe Procedures*, http://users.telenet.be/d.rijmenants/en/enigmaproc.htm


Prepared in LaTeX $2_\varepsilon$ by *Gaurish Korpal*