6 Message Key Analysis by Rejewski

The German Army adopted the Enigma in 1930 as Enigma I. In the first years this variant of the Enigma also had three rotors only—as had the commercial Enigma—but had the rotors wired in another way. Furthermore the additional plugboard, sitting between in/output and the rotors, substantially increased the key space, see Section 2.

The crucial point for the first break-in by the Polish cryptanalysts was a weakness in key handling:

- The key consisted of a daily basic setting and an individual message key.
- The daily basic setting consisted of the rotor order, the ring positions, and the plug connections—first at most 6 plugs—as well as an initial position of the rotors. This setting was valid for all messages of the day—in the first years even for several days. It was known to all participants of the communication network.
- The message key consisted of the initial positions of the three rotors. These could be changed quickly and were to be set by the operator in a random way. This key changed with every message and thereby precluded the alignment in depth of all the messages encrypted with the same daily basic setting.
- The receiver of the message knew the basic setting but not the message key. Therefore the operator encrypted the message key, consisting of three letters, with the basic setting and prefixed this three-letter-cryptogram to the message. This is no diminution of security as long as the keys are selected in a purely random way. In practice they were not.
- Because the radiocommunication was interference-prone, and a distorted key would garble the entire message, the message key was encrypted twice. Thus the proper message had a six-letter prefix. Adding redundancy to a message is not good idea in classical cryptography.

The operator hence had to encrypt six letters, a repeated trigram, using the basic setting, then to set the message key—the rotor positions—and then to encrypt the proper message.

The Polish intercepted the encrypted radio messages of the German Army but couldn't read them—until in 1932 they hired the mathematician Rejewski and his colleagues Różicky und Zygalski.

We describe their approach following Bauer's book [1] whose presentation relies on Rejewski's own description. At first we disregard the obstruction of the analysis that is caused by the (unknown) ring setting, that is, by the unknown stepping of the middle and maybe also the slow rotor.

Some Intercepted Messages

Suppose the first six letters of each of 65 intercepted messages from a single day were (in alphabetic order)

```
AUQ AMN | IND JHU | PVJ FEG | SJM SPO | WTM RAO
BNH CHL | JWF MIC | QGA LYB | SJM SPO | WTM RAO
BCT CGJ | JWF MIC | QGA LYB | SLM SPO | WTM RAO
CIK BZT | KHB XJV | RJL WPX | SUG SMF | WKI RKK
DDB VDV | KHB XJV | RJL WPX | SUG SMF | XRS GNM
EJP IPS | LDR HDE | RJL WPX | TMN EBY | XRS GNM
FBR KLE | LDR HDE | RJL WPX | TMN EBY | XOI GUK
GPB ZSV | MAW UXP | RFC WQQ | TAA EXB | XYW GCP
HNO THD | MAW UXP | SYX SCW | USE NWH | YPC OSQ
HNO THD | NXD QTU | SYX SCW | VII PZK | YPC OSQ
HXV TTI | NXD QTU | SYX SCW | VII PZK | ZZY YRA
IKG JKF | NLU QFZ | SYX SCW | VQZ PVR | ZEF YOC
IKG JKF | OBU DLZ | SYX SCW | VQZ PVR | ZSJ YWG
```

Two observations catch the eye:

1. Frequently even different operators use the same message keys. This could hint at certain stereotypes. Looking for different messages with the same six-letter prefix a coincidence calculation shows that they in fact are encrypted with the same key.

2. The repetition of the three letters of the message key is obvious: If two messages coincide in the first letters, then also their fourth letters coincide. For example a `Z` at position 1 implies a `Y` at position 4. The same holds for positions 2 and 5 (`U` implies `M`) and 3 and 6 (`W` implies `P`).

Therefore the handling of the message keys could be detected from the pure ciphertext, if it was not known already. In any case the cryptanalyst has a lot of ciphertext in depth: The first six letters of each message. If according to the operating instructions the message keys were randomly selected, this observation wouldn't be of much use. However, as it turned out, the message keys were non-random!

REJEWSKI's Approach

REJEWSKI started his analysis by looking at the repeated message keys. Suppose

- $a_1a_2a_3$ is the message key, hence the plaintext starts with the six letters $a_1a_2a_3a_1a_2a_3$.
- The ciphertext starts with the six letters $c_1c_2c_3c_4c_5c_6$.
- The first six Enigma substitutions, starting with the basic setting (+ the first rotor stepping before the first letter is encrypted), are $\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6$.

Then we have

$$
\begin{aligned}
c_1 &= \rho_1 a_1, \quad c_4 = \rho_4 a_1, \quad a_1 = \rho_1 c_1, \quad c_4 = \rho_4 \rho_1 c_1 \\
c_2 &= \rho_2 a_2, \quad c_5 = \rho_5 a_2, \quad a_2 = \rho_2 c_2, \quad c_5 = \rho_5 \rho_2 c_2 \\
c_3 &= \rho_3 a_3, \quad c_6 = \rho_6 a_3, \quad a_3 = \rho_3 c_3, \quad c_6 = \rho_6 \rho_3 c_3
\end{aligned}
$$

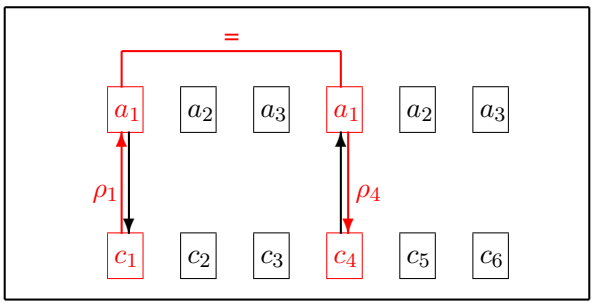

Abbildung 3: *Repeated message key*

Figure 3 illustrates this situation.

The combined permutations $\tau_1 = \rho_4\rho_1$, $\tau_2 = \rho_5\rho_2$, $\tau_3 = \rho_6\rho_3$ are known if we have enough different message keys. In the example the 40 different six-letter groups completely determine τ_1:

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

and τ_2:

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

and τ_3:

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

In REJEWSKI's terminology the triple (τ_1, τ_2, τ_3) was called the **characteristic of the day**.

However we are far from knowing $\rho_1, \ldots, \rho_6$, and far from knowing the basic setting, or even a single message key!

At first sight the plugboard makes trouble. But REJEWSKI as a mathematician knew that the Enigma substitutions with or without plugboard differ only by conjugation with the plugboard substitution η. Therefore there is an *invariant* immune to the effect of the plugboard: the cycle type of the permutations τ_1, τ_2, τ_3, see Appendix A. The cycle decompositions are

τ_1 : `(A)(BC)(DVPFKXGZYO)(EIJMUNQLHT)(RW)(S)` of type $[10, 10, 2, 2, 1, 1]$

τ_2 : `(AXT)(BLFQVEOUM)(CGY)(D)(HJPSWIZRN)(K)` of type $[9, 9, 3, 3, 1, 1]$

τ_3 : `(ABVIKTJGFCQNY)(DUZREHLXWPSMO)` of type $[13, 13]$

From this point the analysis has two possible continuations:

- Assume the rotor wirings are unknown. The cryptanalyst assumes that the message keys are chosen in a stereotypic way—an assumption that in the case of the Wehrmacht-Enigma turned out to be true, see below. This assumption and the material delivered be a German spy and containing the basic settings for a few days including the plug connections enabled RÓŚICKY to derive the wiring of the fast rotor. Since the basic settings changed, each rotor sometimes occupied position 1, so eventually the wirings of all three rotors became known.
- Assume the wirings are known. Then the basic setting can be completely determined and all the messages of the day can be decrypted.

These approaches lead to successes, but not always. REJEWSKI and his colleagues also found some other attack methods, in particular using known plaintext. Here we omit this.

Determining the Enigma Substitution from the Characteristics of the Day

We return to our example and try to determine the first six Enigma substitutions in basic setting, $\rho_1, \ldots, \rho_6$, from the known products $\tau_1 = \rho_4\rho_1$, $\tau_2 = \rho_5\rho_2$, $\tau_3 = \rho_6\rho_3$ whose cycle decomposition is given above. We start with the schema

```
(A) (BC) (DVPFKXGZYO)
(S) (WR) (THLQNUMJIE)

(D) (AXT) (BLFQVEOUM)
(K) (YGC) (NRZIWSPJH)

(ABVIKTJGFCQNY)
(OMSPWXLHERZUD)
```

see Appendix A. We immediately coinclude that ρ_1 and ρ_4 both have the 2-cycle `(AS)`, and ρ_2 and ρ_5 both have the 2-cycle `(DK)`. But even for the 2-cycles of τ_1 we don't get a unique solution: ρ_1 could have the cycles `(BW)(CR)` and ρ_4 the cycles `(BR)(CW)`, or conversely.

To get on we assume—following REJEWSKI—that `aaa` is the most popular message key with the German operators. (If this would turn out as erroneous we would try some other stereotype.) If we are right, then this corresponds to the encrypted message key `SYX SCW` that occurs five times, and implies the cycles

$$\begin{array}{ll} \texttt{(AS) in } \rho_1, & \texttt{(AS) in } \rho_4, \\ \texttt{(AY) in } \rho_2, & \texttt{(AC) in } \rho_5, \\ \texttt{(AX) in } \rho_3, & \texttt{(AW) in } \rho_6. \end{array}$$

This is nothing new for ρ_1 and ρ_4. But for τ_2 it means that the alignment of the 3-cycles is correct, and we read off the 2-cycles

$$\texttt{(AY)(XG)(TC) in } \rho_2, \quad \texttt{(AC)(GT)(XY) in } \rho_5.$$

For τ_3 the correct alignment is

```
(ABVIKTJGFCQNY)
(XLHERZUDOMSPW)
```

and we find the unique solution

$$\begin{aligned}\rho_3 &= \texttt{(AX)(BL)(CM)(DG)(EI)(FO)(HV)(JU)(KR)(NP)(QS)(TZ)(WY)}\\ \rho_6 &= \texttt{(AW)(BX)(CO)(DF)(EK)(GU)(HI)(JZ)(LV)(MQ)(NS)(PY)(RT)}\end{aligned}$$

Now let's look at other encrypted message keys. The first one in our table is `AUQ AMN`, partially decrypting to the plaintext

```
s?s s?s
```

We suspect the stereotypical message key `sss`. If we are right, then ρ_2 has the 2-cycle `(SU)`, and ρ_5 has the 2-cycle `(MS)`. This gives the correct alignment of the 9-cycles ot τ_2:

```
(D) (AXT) (BLFQVEOUM)
(K) (YGC) (JHNRZIWSP)
```

and completely determines ρ_2 and ρ_5:

$$\begin{aligned}\rho_2 &= \texttt{(AY)(BJ)(CT)(DK)(EI)(FN)(GX)(HL)(MP)(OW)(QR)(SU)(VZ)}\\ \rho_5 &= \texttt{(AC)(BP)(DK)(EZ)(FH)(GT)(IO)(JL)(MS)(NQ)(RV)(UW)(XY)}\end{aligned}$$

The encrypted message key `RJL WPX` occurs four times, and partially decrypts as

```
?bb ?bb
```

Again we are quite sure that this reveals a stereotypical message key: `bbb`. We conclude that ρ_1 has the cycle `(BR)`—hence also the cycle `(CW)`—and ρ_4 has the cycle `(BW)`, hence also the cycle `(CR)`.

For the complete solution the only open problem left is the alignment of the two 10-cycles of τ_1. We look at the group `LDR HDE` and partially decrypt it as

```
?kk ?kk
```

We are quite sure of the message key `kkk`. Then ρ_1 has the 2-cycle `(KL)`, the correct alignment is

```
(A) (BC) (DVPFKXGZYO)
(S) (RW) (IETHLQNUMJ)
```

and the complete solution is

$$\begin{aligned}\rho_1 &= \texttt{(AS)(BR)(CW)(DI)(EV)(FH)(GN)(JO)(KL)(MY)(PT)(QX)(UZ)}\\ \rho_4 &= \texttt{(AS)(BW)(CR)(DJ)(EP)(FT)(GQ)(HK)(IV)(LX)(MO)(NZ)(UY)}\end{aligned}$$

Now we can decrypt all message keys for the actual basic setting. However we do not yet know the basic setting itself, and we cannot decrypt a single message. In particular we do not know the ring setting and the positions of the rotors corresponding to the message keys.

REJEWSKI's Catalogue

In our example the permutations $\tau_1 = \rho_4\rho_1, \tau_2 = \rho_5\rho_2$, and $\tau_3 = \rho_6\rho_3$ are completely determined and their cycle types are the partitions

```
[10 10 2 2 1 1], [9 9 3 3 1 1], [13 13]
```

of the number 26. Now we ask how characteristic is this triple of partitions for the basic setting of the Enigma. The plug connections are irrelevant for this problem. We consider the rotor order as an element of the permutation group $\mathcal{S}_3$, and the initial positions of the three rotors as elements of the cyclic group $\mathbb{Z}/26\mathbb{Z}$. If we disregard the plugboard and the ring settings, the possible basic settings form the set $\mathcal{S}_3 \times (\mathbb{Z}/26\mathbb{Z})^3$. On the other hand we have the set $\mathcal{P}_{13}$ consisting of all the 101 partitions of the number 13 (in bijective correspondence with the partitions of the number 26 in pairwise equal parts), and we have a map

$$\mathcal{S}_3 \times (\mathbb{Z}/26\mathbb{Z})^3 \longrightarrow (\mathcal{P}_{13})^3$$

We would like this map to be injective. This seems not unrealistic in view of the cardinalities: 105,456 different basic settings, $101^3 = 1{,}030{,}301$ different partitions.

To get the complete value table of this map REJEWSKI designed a simple Enigma simulator called Cyclometer that run through all basic settings in about one year. The result, called REJEWSKI's Catalogue, got lost. But there is a recent reconstruction in the paper

Alex KUHL: Rejewski's Catalog. Cryptologia 31 (2007), 326–331.

It turned out that the above map is *not* injective, but "almost" so: Many triples of partitions have a unique preimage, most have only a few ones. However a few triples occur quite frequently, the top ten being

Triple of partitions			Frequency
[13 13]	[13 13]	[13 13]	1771
[12 12 1 1]	[13 13]	[13 13]	898
[13 13]	[13 13]	[12 12 1 1]	866
[13 13]	[12 12 1 1]	[13 13]	854
[11 11 2 2]	[13 13]	[13 13]	509
[13 13]	[12 12 1 1]	[12 12 1 1]	494
[13 13]	[13 13]	[11 11 2 2]	480
[12 12 1 1]	[13 13]	[12 12 1 1]	479
[13 13]	[11 11 2 2]	[13 13]	469
[12 12 1 1]	[12 12 1 1]	[13 13]	466

All in all there are 21230 different triples in the image of the map. 19604 of these, that is 92%, occur at most ten times, the numbers of these rare triples are

Pre-Im	1	2	3	4	5	6	7	8	9	10
Freq	11466	3381	1658	958	660	456	343	265	234	183

Using the catalogue the Polish cryptanalysts usually found the correct basic setting in at most 20 minutes. It is unknown what they did in the exceptional situations where there are too many false positives. Certainly some other useful details could be used. In any case we may assume that the method was successful for at least 92% of all triples, corresponding to roughly 50% of all cases.

We neglected the effect of the ring setting. This causes a rotor movement because the stepping mechanism is connected with the alphabet ring. Now, what could happen? As long as only the fast rotor moves we are in a situation included in the catalogue. The analysis is hampered if the middle rotor moves between two of the first six letters. The chances are 5 of 26 ring settings, that is about 19%. This lowers the total probability of success from 50% to about 40%.

There is even more potential for drawing conclusions from the collected message keys. For example the moving of the middle rotor gives information about the ring setting of the first rotor. An approach to determining the plugboard connections uses the fact that in the first years at most six letter pairs were interchanged. If the cryptanalysts assume that there are no plugs at all, then some true plaintext shows through the tentatively decrypted text. This enables them to reconstruct the plugboard connections.

Epilogue

The plugboard turns out to be an illusory complication: It slows the cryptanalyst down a bit, but not as much as the increase in keylength from 29 to 76 bits—expressed in terms of today—suggested. The main cost of the cryptanalysis is exhausting the rotor order and positions, and this could be made efficient by compiling lookup tables.

By the way the decrypted 40 different message keys from the list of 65 above are:

```
AUQ AMN : sss | IKG JKF : ddd | QGA LYB : xxx | VQZ PVR : ert
BNH CHL : rfv | IND JHU : dfg | RJL WPX : bbb | WTM RAO : ccc
BCT CGJ : rtz | JWF MIC : ooo | RFC WQQ : bnm | WKI RKK : cde
CIK BZT : wer | KHB XJV : lll | SYX SCW : aaa | XRS GNM : qqq
DDB VDV : ikl | LDR HDE : kkk | SJM SPO : abc | XOI GUK : qwe
EJP IPS : vbn | MAW UXP : yyy | SUG SMF : asd | XYW GCP : qay
FBR KLE : hjk | NXD QTU : ggg | TMN EBY : ppp | YPC OSQ : mmm
GPB ZSV : nml | NLU QFZ : ghj | TAA EXB : pyx | ZZY YRA : uvw
HNO THD : fff | OBU DLZ : jjj | USE NWH : zui | ZEF YOC : uio
HXV TTI : fgh | PVJ FEG : tzu | VII PZK : eee | ZSJ YWG : uuu
```

The astonishingly naive habits of the German cipher operators become obvious by looking at the keyboard layout of Enigma:

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

All message keys belong to one of three groups of stereotypes

- iterated letters: `sss`, `fff`, `ddd`, `ooo`, ...

- three consecutive keys: `rfv`, `rtz`, `wer`, `ikl`, ...
- three letters in alphabetic order: `abc`, `uvw`

Before World War II the British cryptanalysts failed with the cryptanalysis of Enigma because they tried to determine the wiring between in-/output and first rotor. The commercial Enigma D connected `Q` with `A`, `W` with `B`, `E` with `C` and so on in the order of the keyboard. Assuming this for Enigma I didn't work. REJEWSKI who knew the Germans since he was a student at Göttingen simply assumed that the wiring in any case should follow a simple scheme, and succeeded with the assumption "`A` is connected to `A`, `B` to `B` etc."

The point: Enigma C also had had this simple wiring, and this information could be found in the patent file in the British Patent Office ...

For later attacks (from 1938 on) of the Polish cryptanalysts against the Enigma, including a complete example, see the paper

> David LINK, Resurrecting *Bomba Kryptologiczna*: Archeology of Algorithmic Artefacts, I. Cryptologia 33 (2009), 166–182.