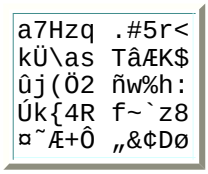
Kryptologie

Kryptoanalyse der Enigma I nach REJEWSKI

Spruchschlüssel-Analyse



Hintergrund

Die Wehrmachts-Enigma I wurde ab 1930 in Betrieb genommen. Sie hatte anders verdrahtete Rotoren als die kommerzielle Enigma - zunächst auch nur drei - und dazu das vorgeschaltete Steckerbrett.

Entscheidend für den ersten Einbruch war eine Schwachstelle in der Handhabung: Der Schlüssel wurde in einem Grundschlüssel und einen Spruchschlüssel aufgeteilt.

- Der Grundschlüssel bestand aus Walzenlage, Ringstellung und Steckerverbindungen (zunächst maximal 6 Stecker) sowie einer Anfangsstellung. Dieser Grundschlüssel wurde längere Zeit, zunächst sogar mehrere Tage, unverändert beibehalten, und war allen möglichen Kommunikationspartnern bekannt.
- Der Spruchschlüssel bestand aus der Anfangsstellung der drei Walzen, die ja leicht mit den Einstellrädern jederzeit zu ändern war, und vom Operator willkürlich festgelegt wurde. Der Sinn des Spruchschlüssels war, zu verhindern, dass alle Nachrichten während der Gültigkeitsdauer des Grundschlüssels phasengleich verschlüsselt wurden und somit durch Untereinanderschreiben reichlich Material für eine Kolonnen-Analyse lieferten.
- Um dem Empfänger den Spruchschlüssel mitzuteilen, wurden die entsprechenden drei Buchstaben zunächst mit dem Grundschlüssel verschlüsselt.
- Da der Funkverkehr störanfällig war und ein falsch übertragener Schlüssel die Nachricht unlesbar gemacht hätte, verschlüsselte man den Spruchschlüssel zweimal hintereinander.
- Es wurden also insgesamt sechs Buchstaben in Grundstellung verschlüsselt, danach wurde der Spruchschlüssel eingestellt und die eigentliche Nachricht verschlüsselt.

Die Polen hörten die verschlüsselten Funksprüche der Deutschen mit konnten aber zunächst damit nichts anfangen. Bis der Mathematiker REJEWSKI zusammen mit seinen Kollegen ROZICKI und ZYGALSKI 1932 an das Problem gesetzt wurde.

Die folgende Darstellung folgt zunächst dem Buch von Bauer, geht aber auf REJEWSKI selbst zurück. Wir vernachlässigen im folgenden (bis auf weiteres) die Störung der Analyse, die durch die (unbekannte) Ringstellung, d. h., durch das unbekannte Weiterschalten der zweiten (und eventuell sogar dritten) Walze verursacht wird.

Aufgefangene Nachrichten

65 an einem Tag aufgefangene Nachrichten begannen etwa so:

```
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
```

Man erkennt daran zweierlei:

1. Es werden öfter die gleichen Spruchschlüssel (auch von verschiedenen Chiffrierern) verwendet. Das lässt gewisse Stereotypien vermuten. Außerdem können verschiedene Nachrichten mit dem gleichen sechsbuchstabigen Beginn einer Koinzidenz-Bestimmung unterzogen werden, wobei sich bestätigt, dass sie mit dem gleichen Schlüssel chiffriert wurden (Zeichenkoinzidenz ungefähr gleich dem Koinzidenzindex der deutschen Sprache).
2. Die Wiederholung der drei Buchstaben des Spruchschlüssels zeichnet sich deutlich ab: Ist der erste Buchstabe zweier Nachrichten gleich, so auch der vierte, wie z. B. ein Z an Stelle 1 ein Y an Stelle 4 nach sich zieht. Gleiches gilt für die Positionen 2 und 5 (U zieht M nach sich) sowie 3 und 6 (W wird von P gefolgt).

Die Spruchschlüssel-Handhabung könnte also durchaus - wenn sie nicht als bekannt angenommen wird - aus dem reinen Geheimtext erschlossen werden. Auf jeden Fall liefert sie reichlich phasengleichen Geheimtext, nämlich jeweils die ersten sechs Buchstaben jeder Nachricht.

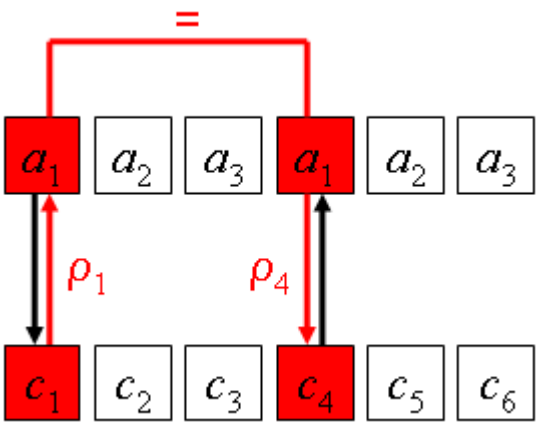
REJEWSKIs Ansatz

Beim wiederholten Spruchschlüssel setzte REJEWSKI an. Seien

- $a_1a_2a_3$ der Spruchschlüssel, also $a_1a_2a_3\ a_1a_2a_3$ die ersten 6 Buchstaben des Klartexts,
- $c_1c_2c_3\ c_4c_5c_6$ der entsprechende Geheimtext,
- $\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6$ die ersten Substitutionen der Enigma, wenn man in der Grundstellung startet.

Dann ist

- $c_1 = \rho_1(a_1)$, $c_4 = \rho_4(a_1)$, also $a_1 = \rho_1(c_1)$ und somit $c_4 = \rho_4\rho_1(c_1)$,
- $c_2 = \rho_2(a_2)$, $c_5 = \rho_5(a_2)$, also $a_2 = \rho_2(c_2)$ und somit $c_5 = \rho_5\rho_2(c_2)$,
- $c_3 = \rho_3(a_3)$, $c_6 = \rho_6(a_3)$, also $a_3 = \rho_3(c_3)$ und somit $c_6 = \rho_6\rho_3(c_3)$.



Die Permutationen $\tau_1 = \rho_4\rho_1$, $\tau_2 = \rho_5\rho_2$ und $\tau_3 = \rho_6\rho_3$ sind durch genügend viele Nachrichten also weitgehend bekannt. Im Beispiel können wir aus den 40 verschiedenen Sechsergruppen τ_1 schon vollständig bestimmen:

```
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
```

ebenso 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
```

und 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
```

Das Tripel (τ_1, τ_2, τ_3) nannte REJEWSKI »Tages-Charakteristik«.

Aber damit sind die Substitutionen ρ_1 bis ρ_6 noch lange nicht identifiziert, und auch von der Ermittlung des Grundschlüssels oder der einzelnen Spruchschlüssel sind wir noch weit entfernt!

Zunächst stört das Steckerbrett. REJEWSKI als Mathematiker wusste allerdings, dass sich die Enigma-Substitution mit und ohne Stecker nur durch die Konjugation mit der Steckerbrett-Substitution η unterscheidet. Es gibt also eine *Invariante*, die gegen die Steckerbrett-Substitution unempfindlich ist: den Zykel-Typ der Zykel-Darstellung der Permutationen τ_1, τ_2 und τ_3 [siehe PDF]. Diese Zykel sind:

> τ_1: `(A) (BC) (DVPFKXGZYO) (EIJMUNQLHT) (RW) (S)` vom Typ [10 10 2 2 1 1],
> τ_2: `(AXT) (BLFQVEOUM) (CGY) (D) (HJPSWIZRN) (K)` vom Typ [9 9 3 3 1 1],
> τ_3: `(ABVIKTJGFCQNY) (DUZREHLXWPSMO)` vom Typ [13 13].

Von diesem Punkt aus gibt es zwei Fortsetzungen:

- Bei der ersten benutzt man die Kenntnis der Walzen-Verdrahtung nicht, nimmt aber an, dass stereotype Spruchschlüssel verwendet wurden -- Annahmen, die sich unten bestätigen. Mit diesem Ansatz und einigen Tagesschlüsseln, die ja auch die Steckerstellungen enthielten - und aus französicher Spionage-Arbeit zu den Polen gelangt waren - gelang es ROZICKI sogar, die Verdrahtung der langsamen Walze zu ermitteln, und damit in kurzer Zeit aller Walzen, da wegen des Wechsel jede mal als langsame dran war. [---> Fortsetzung 1]
- Bei der zweiten Fortsetzung nimmt man statt dessen an, dass die Walzen-Verdrahtung schon

bekannt ist, und kommt zu einer vollständigen Bestimmung der Tagesschlüssel und somit zu einer vollständigen Entschlüsselung aller Nachrichten. [---> Fortsetzung 2]

Da diese Ansätze auch zusammengenommen nicht in allen Fällen zum Erfolg führen, entwickelten REJEWSKI und seine Kollegen noch weitere Methoden, insbesondere zur Ausnutzung von bekanntem Klartext, die hier nicht weiter ausgeführt werden.

Nachlese

Auch die Einführung des Steckerbretts erweist sich so als *illusorische Komplikation*: Der Kryptoanalytiker wird dadurch zwar etwas aufgehalten, aber längst nicht im erhofften Umfang, wie es die Erhöhung der Schlüssellänge - in heutigen Begriffen ausgedrückt - von 31 auf 81 Bit glauben ließ. Der Aufwand besteht immer noch »nur« aus der Exhaustion der Walzenlage und -stellung der drei Walzen - und die konnte durch eine umfangreiche universelle Vorberechnung erledigt werden.

Die entschlüsselten 40 verschiedenen der obigen 65 Spruchschlüssel sind übrigens:

```
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
```

Die erschreckend naiven Angewohnheiten der deutschen Chiffrierer werden beim Blick auf die Tastatur-Anordnung der Enigma enthüllt:

```
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
```

Alle Spruchschlüssel gehören zu einer der drei Gruppen

- wiederholter Buchstabe: `sss, fff, ddd, ooo, ...`
- drei nebeneinander liegende Tasten: `rfv, rtz, wer, ikl, ...`
- drei Buchstaben in alphabetischer Folge (fast schon originell): `abc, uvw`

Die Engländer waren übrigens an der Kryptoanalyse der Enigma gescheitert, weil sie versuchten die Eingangsverdrahtung zwischen Tastatur bzw. Lampenfeld und erstem Rotor zu bestimmen. Sie wich nämlich von der Verdrahtung der kommerziellen Enigma D ab, bei der Q mit A, W mit B, E mit C usw. in der Reihenfolge der Tastatur verdrahtet waren. Rejewski, der die Deutschen aus seinem Studium in Göttingen gut kannte, nahm einfach an, dass alles sehr ordentlich und systematisch gemacht worden war und hatte mit der Vermutung "A ist mit A verdrahtet, B mit B usw." Erfolg.

Die Pointe: Diese Verkabelung war auch schon bei der Enigma C verwendet worden und in der Patentschrift im britischen Patent Office beschrieben ...

Autor: Klaus Pommerening, 6. Januar 2008; letzte Änderung: 13. Januar 2008.