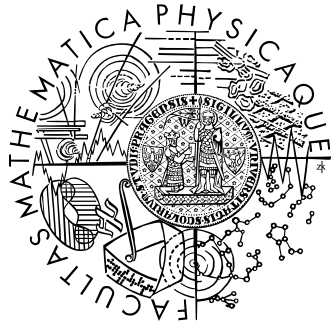
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Petr Veselý

Luštění německého šifrovacího stroje Lorenz

Katedra algebry

Vedoucí bakalářské práce: Mgr. Pavel Vondruška

Studijní program: Matematika
matematické metody informační bezpečnosti

2008

Děkuji vedoucímu práce Mgr. Pavlu Vondruškovi za cenné rady a připomínky a poskytnutí užitečných zdrojů.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 27. května 2008 Petr Veselý

OBSAH

Název práce:	Luštění německého šifrovacího stroje Lorenz
Autor:	Petr Veselý
Katedra (ústav):	Katedra algebry
Vedoucí bakalářské práce:	Mgr. Pavel Vondruška
e-mail vedoucího:	pavel.vondruska@crypto-world.info

Abstrakt: Rotorový šifrovací stroj Lorenz používala za druhé světové války německá armáda k zabezpečení dálnopisného spojení na nejvyšší úrovni velení. Šifra byla založena na Vernamě principu, tedy sčítání otevřeného textu s pseudonáhodným klíčem. Britští kryptoanalytici tento systém rozbili již během jeho zkušebního provozu. Díky chybě německého radisty získali necelých 4000 znaků pseudonáhodného klíče, z nějž odvodili konstrukci šifrovacího stroje. Tato práce stručně popisuje známé historické skutečnosti týkající se používání a luštění šifry Lorenz a obsahuje detailní popis funkce všech používaných verzí šifrátoru. Hlavní část práce prezentuje možný způsob určení vnitřní stavby šifrovacího stroje analýzou jím produkované pseudonáhodné posloupnosti klíče, s využitím pouze těch informací, které měli britští kryptoanalytici k dispozici, včetně stejné sekvence klíče. Součástí práce je softwarový simulátor šifrovacího stroje Lorenz SZ40.

Klíčová slova: kryptoanalýza, Vernamova šifra, Lorenz, Tunny, Bletchley Park

Title:	The German Lorenz Cipher System And How It Was Broken
Author:	Petr Veselý
Department:	Department of Algebra
Supervisor:	Mgr. Pavel Vondruška
Supervisor's e-mail address:	pavel.vondruska@crypto-world.info

Abstract: In the World War 2, the German army used the rotor cipher machine Lorenz to secure their teleprinter communication on the highest level of command. The cipher was based on the Vernam principle, i.e. addition of a pseudorandom key to the plaintext. British cryptanalysts had broken this system already in its experimental period. Due to an error of a German operator they were provided nearly 4000 characters of the pseudorandom key, from which they deduced the operation of the machine. This paper contains a brief review of known historical facts concerning the usage and breaking of the Lorenz cipher and describes in detail the function of all versions of the machine used. The main part of this work presents a possible way of determining the structure of the machine by analysis of the key it produces. Only the information available to the British war cryptanalysts is used in the analysis, including the original key sequence. This work also includes a software simulator of the Lorenz SZ40 machine.

Keywords: cryptanalysis, Vernam cipher, Lorenz, Tunny, Bletchley Park

ÚVOD

Úspěchy britských kryptoanalytiků, kteří v přísně utajeném středisku v Bletchley Parku rozbili za druhé světové války mnohé šifry používané státy Osy, dodnes přitahují pozornost laické i odborné veřejnosti. Jejich práce obsahuje příklady invenčního využití matematiky i názorné ukázky toho, jak katastrofální důsledky pro bezpečnost šifrového systému může mít nedůslednost v dodržování základních kryptografických pravidel, což ji činí stále aktuálním zdrojem inspirace a ponaučení. Kromě toho jsou teprve v posledních letech veřejnosti zpřístupňovány dobové dokumenty, které odhalují dosud neznámé informace a detaily.

Ve stínu patrně nejznámější válečné šifry *Enigma* dlouho zůstávala historie prolomení šifrového systému *Lorenz*, pracujícího na principu Vernamovy šifry, který používala německá armáda k zabezpečení dálnopisné komunikace na nejvyšší úrovni velení. Přitom tento úspěch je hodný pozornosti přinejmenším ze dvou důvodů. Zaprvé, neznámou konstrukci šifrovacího přístroje se pracovníkům Bletchley Parku podařilo odvodit pouhým zkoumáním necelých čtyř tisíc znaků klíče, které získali díky hrubé chybě německého operátora. Dosáhli toho dokonce již během zkušebního provozu šifrátoru a získali tak prakticky až do konce války přístup k informacím o strategických plánech nepřítele. Zadruhé, k luštění zachycených zpráv byla zkonstruována řada pokročilých výpočetních zařízení, včetně elektronkových počítačů *Colossus*, prvních elektronických částečně programovatelných počítačů na světě.

V posledních letech se šifra Lorenz dostává do středu zájmu a věnuje se jí odborná i populární literatura (pravděpodobně zatím žádná však v českém jazyce). Jedním z důvodů je zpřístupnění dobové oficiální zprávy o luštění této šifry, *General Report on Tunny*, v roce 2000. Dalším je nedávné úspěšné dokončení projektu sestavení funkční repliky počítače Colossus. Okolnosti rozbití šifrového systému Lorenz jsou tématem této bakalářské práce.

První kapitola práce stručně seznamuje se známými údaji o používání šifry Lorenz a jejího úspěšného luštění v Bletchley Parku.

Ve druhé kapitole je podrobně vysvětlen princip fungování šifrovacího stroje Lorenz SZ ve všech používaných verzích.

Cílem třetí kapitoly je detailně předvést možný způsob odvození stavby šifrátoru Lorenz SZ40 z části pseudonáhodného klíče, který produkoval. Skutečně použitý postup není možné z kusých informací v dostupných zdrojích přesně zrekonstruovat, analýza proto postupuje po vlastní linii, přičemž se pracuje pouze s informacemi, které podle zdrojů měli nebo mohli mít britští kryptoanalytici k dispozici, včetně stejné sekvence klíče.

Součástí práce je softwarový simulátor šifrovacího stroje Lorenz SZ40.

"They were the geese that laid
the golden eggs and never
cackled…"
Winston Churchill

1. HISTORICKÉ POZADÍ

1.1. DĚNÍ NA KONTINENTU

Během 2. světové války se německé ozbrojené síly na cestě za ovládnutím Evropy setkaly s potřebou bezpečného a spolehlivého spojení na ose mezi hlavním štábem, velitelstvími armádních skupin a jednotlivými bojovými jednotkami. Tomuto účelu sloužila řada šifrovacích přístrojů domácí výroby, které byly často modifikacemi komerčních produktů z předválečné doby. Většina z nich patřila k rotorovým šifrátorům, k jejichž nejvíce ceněným (a přeceňovaným) vlastnostem patřila v té době bezkonkurenční mohutnost klíčového prostoru.

Nejrozšířenějším přístrojem byla známá *Enigma*, která se používala k předběžnému šifrování zpráv před jejich odesláním běžným komunikačním kanálem (off-line šifrování) a díky své přenosnosti patřila mimo jiné do výbavy bojových útvarů nejnižší úrovně. K přenosu zpráv se zpravidla používala Morseova abeceda.

Geheimschreiber T52 firmy *Siemens & Halske AG*, patentovaný ve Spojených státech v roce 1933, byl šifrovací dálnopis umožňující on-line šifrovanou komunikaci po pevné lince, později byla vyvinuta i bezdrátová verze. Toto zařízení používala především *Luftwaffe*.

Lorenz SZ40 (a následné verze *SZ42A*, *SZ42B*) vyráběný společností *C. Lorenz AG* byl přídavným šifrovacím modulem k bezdrátovému dálnopisu a umožňoval rovněž přenos šifrovaných zpráv on-line. Jeho konstrukce pravděpodobně nebyla veřejně známá. Používal se na citlivých linkách mezi nejvyšším velitelstvím pozemní armády v Berlíně a hlavními stany armádních skupin v okupované Evropě a severní Africe. Právě posledně jmenovaný šifrovací stroj je tématem této práce.

Podle [1] byl šifrátor Lorenz SZ40 zprvu nasazen do zkušebního provozu na lince mezi Berlínem, Athénami a Soluní v červnu 1941. Zprávy byly přenášeny ve formátu přístroje *Hellschreiber* (zařízení funkcí podobné faxu) a na přijímající stanici tisknuty na papírovou pásku. Po více než roce zkoušek, během něhož se upravila pravidla používání přístroje, bylo v říjnu 1942 zahájeno ostré vysílání na linkách Berlín – Soluň a Královec – jižní Rusko, nyní již v Baudotově dálnopisném kódu. Komunikační linku vždy tvořil pár bezdrátových dálnopisů s šifrovacím modulem vybavených stejnou sadou klíčů (odlišnou od ostatních linek).

Postupně se otevíraly i další komunikační spoje a od roku 1943 byly stroje SZ40 nahrazovány novějšími modely SZ42A. V době spojenecké invaze do Normandie

v roce 1944 byla síť nejrozsáhlejší, podle [1] ji tvořilo celkem 26 linek s dvěma centrálními ústřednami ve Straußbergu u Berlína a v Královci.

S blížícím se zhroucením německého odporu se postupně hroutila i organizace komunikační sítě, zejména vlivem častých přesunů jednotlivých armád i hlavního velitelství. Zpráva [1] uvádí, že od června 1944 byl postupně zhruba na polovině linek nasazen šifrátor verze SZ42B a zaváděla se další bezpečnostní opatření. Poslední zpráva šifrovaná přístrojem Lorenz SZ byla podle [1] poslána v den německé kapitulace 8. května 1945.

1.2. DĚNÍ NA OSTROVECH

Znalost úmyslů nepřítele může v boji znamenat rozhodující výhodu. V souladu s tímto faktem otevřela vláda Spojeného království roku 1939 ve venkovském sídle v Bletchley Parku, 80 km severozápadně od Londýna, přísně tajné kryptoanalytické středisko (označované *Station X*). Přední matematici, lingvisté a experti v různých oborech zde po celou válku úspěšně pracovali na luštění šifer zemí Osy, zejména Německa. Nepřátelský rádiový provoz byl monitorován systémem odposlouchávacích stanic (tzv. *Stations Y*, zpráva [1] zmiňuje stanici v Knockholtu jižně od Londýna).

Krátce po německé invazi do Ruska v červnu 1941 byla zachycena pravidelná šifrovaná rádiová komunikace mezi Vídní a Athénami, která používala formát přístroje Hellschreiber. Britští kryptoanalytici na základě předcházejících cvičných zpráv na téže lince usoudili, že jde o dálnopisné spojení, a dali mu kódové označení *TUNNY*.

Zkoumáním dalších odposlechnutých zpráv bylo odhaleno, že použitá šifra je Vernamova typu, kde se otevřený text sčítá s pseudonáhodným klíčem. Dlouho se však nedařilo získat dostatečně dlouhý úsek klíče, ze kterého by bylo možné zjistit funkci pseudonáhodného generátoru, a to i přes to, že němečtí operátoři často vysílali různé zprávy zašifrované stejným klíčem, čehož lze v případě Vernamovy šifry snadno využít k luštění.

Dne 30. srpna 1941 vyslal německý radista dvakrát po sobě téměř totožnou zprávu, pokaždé zašifrovanou stejným klíčem. Tyto dvě depeše se dostaly i k pracovníkům Bletchley Parku, kteří nabídnutou příležitost využili.

Zprávy rozluštil (podle [2] během dvou měsíců) plukovník *John H. Tiltman* a získal tak 3976 znaků dlouhou pseudonáhodnou posloupnost. Tuto sekvenci kryptoanalytici zkoumali s cílem určit vnitřní uspořádání přístroje, který ji vygeneroval. Předpokládali, že jde o rotorový šifrovací stroj a částečně mohli vycházet ze známého patentu stroje T52. Přesto konstrukce šifrátoru dlouho odolávala a podle [1, 3] se ji podařilo odhalit až díky téměř náhodnému průlomu mladého matematika *Williama T. Tutteho* v lednu 1942. Možný postup analýzy posloupnosti s cílem zjistit vnitřní konstrukci přístroje je obsahem třetí kapitoly této práce.

Britští kryptoanalytici zjistili, že přístroj zřejmě obsahuje dvanáct rotorů různých velikostí s výklopnými kolíčky po obvodu. Studiem dalších zachycených zpráv zjistili, že počáteční nastavení rotorů se u každé zprávy liší (pokud se operátor nedopustí prohřešku proti kryptografickým pravidlům), zatímco nastavení kolíčků na kolech zůstává stejné po delší časové období. Podrobný popis konstrukce přístroje Lorenz SZ je předmětem druhé kapitoly.

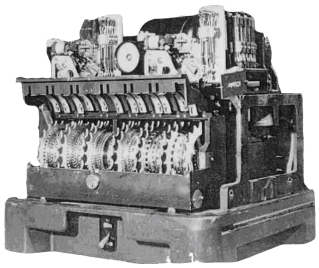
Postupně byla nalezena řada postupů, jak luštit odposlechnuté zprávy. Některé využívaly nedůslednosti německých operátorů, kteří poměrně často vysílali zprávy zašifrované se stejným počátečním nastavením všech nebo téměř všech rotorů. Jiné využívaly naopak jejich přehnané důslednosti: stereotypní hlavičky zpráv (např. *Spruchnummer*, číslo zprávy) umožňovaly útok se znalostí otevřeného textu. Další slabinou byla dvanáctipísmenná indikátorová skupina, která předcházela každé zprávě a určovala počáteční nastavení dvanácti rotorů. Ukázalo se, že s dostatkem zpráv lze z indikátorů určit dokonce nastavení výklopných kolíčků. Díky těmto pokrokům se v červenci 1942 poprvé podařilo vyluštit aktuální zprávy.

Po skončení testovacího provozu němečtí radisté upustili od užívání písmenných indikátorů, různé zprávy zašifrované se stejným nastavením se dařilo získat jen zřídka a byla také zavedena praxe vkládání náhodně zvolených německých výrazů na začátek zprávy, aby se předešlo útokům na obvyklé hlavičky [2]. Pracovníci Bletchley Parku ale v tu dobu již měli k dispozici obecnější algoritmy pro útoky pouze se znalostí šifrového textu, které využívaly nedostatečné náhodnosti generované posloupnosti. Tyto metody jsou spojeny se jmény *Alan Turing*, *Max Newman*, *Donald Michie* aj. a ukázalo se, že jsou snadno adaptovatelné i na pozdější verze šifrovacího zařízení. Podle [1] trvalo Bletchley Parku opětovné rozbití systému po nasazení nového modelu stroje Lorenz SZ vždy přibližně měsíc.

K usnadnění práce a urychlení statistických výpočtů byla zkonstruována řada důmyslných elektromechanických zařízení, například stroj zvaný *Heath Robinson* navržený Maxem Newmanem pro hledání počátečních nastavení rotorů, jehož prototyp byl uveden do provozu v dubnu 1943. Především však jde o elektronkové počítače *Colossus*, vyvinuté *Tommym Flowersem*. První *Colossus* byl připraven k použití v prosinci 1943 a do konce války následovalo dalších devět strojů vylepšené konstrukce. Počítače *Colossus* byly původně používány také pouze k hledání počátečních nastavení rotorů, ukázalo se ale, že díky jejich poměrně velké univerzálnosti je lze použít i k určení nastavení kolíčků kol, což dokázal D. Michie [2, 4].

V průběhu celé války byly vyluštěny zprávy v celkové délce cca. 63 431 000 znaků [1].

Samotná existence kryptoanalytického střediska v Bletchley Parku zůstala před veřejností utajena až do 70. let 20. století. V roce 2000 byla zveřejněna oficiální *Hlavní zpráva o TUNNY* (*General Report on Tunny*), již v roce 1945 napsali pracovníci Bletchley Parku *I. J. Good*, *D. Michie* a *G. A. Timms* [4].



Šifrovací přístroj Lorenz SZ
(obrázek převzat z [1]*)*

2. FUNKCE ŠIFROVACÍHO STROJE LORENZ SZ

2.1. OBECNÉ POZNÁMKY

Přístroj Lorenz SZ je ve všech verzích přídavným modulem bezdrátového dálnopisu (písmena SZ jsou zkratkou německého slova *Schlüsselzusatzgerät*, šifrovací přídavné zařízení).

Dálnopis je telekomunikační zařízení, velmi rozšířené po většinu 20. století, které vzhledem i konstrukcí připomíná elektromechanický psací stroj. Umožňuje elektronicky přenášet psaný text po lince nebo bezdrátově a tisknout zprávy vysílané jinými dálnopisy. Jednotlivé znaky jsou kódovány pětibitovým *Baudotovým kódem*, označovaným také *ITA2* (*International Telegraph Alphabet No. 2*). Většina kódových slov má dva významy (*Letter Shift, Figure Shift*), mezi kterými se přepíná pomocí kontrolních znaků. Významy některých slov v horním registru (*Figures*) nejsou přesně určeny a závisí na zemi použití. Tabulka 2.1 ukazuje význam slov Baudotova kódu podle [1]. Konkrétní signál odpovídající tečce či křížku závisí na podobě komunikační linky.

	Figures	NULL	5	CR	9	SP	£	,	.	LF	)	4	&	8	0	:	=	3	+	□	?	'	6	%	/	-	2	□	FIG	7	1	(	LTR
	Letters		T		O		H	N	M		L	R	G	I	P	C	V	E	Z	D	B	S	Y	F	X	A	W	J		U	Q	K	
impuls	1	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	2	●	●	●	●	●	●	●	●	x	x	x	x	x	x	x	x	●	●	●	●	●	●	●	●	x	x	x	x	x	x	x	x
	3	●	●	●	●	x	x	x	x	●	●	●	●	x	x	x	x	●	●	●	●	x	x	x	x	●	●	●	●	x	x	x	x
	4	●	●	x	x	●	●	x	x	●	●	x	x	●	●	x	x	●	●	x	x	●	●	x	x	●	●	x	x	●	●	x	x
	5	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x	●	x

Tab. 2.1: Baudotův kód (podle [1]). Významy kontrolních znaků: CR – *carriage return*; LF – *line feed*; FIG – přepnout na horní registr (*Figure Shift*); LTR – přepnout na dolní registr (*Letter Shift*). Znaky D a J mají v horním registru po řadě význam *Kdo jsi?* a *zvonek*

Jednotlivé bity slov Baudotova kódu (a posloupnosti těchto bitů) se označují jako impulsy. V dalším textu budou namísto teček a křížků používána po řadě čísla 0 a 1, na které bude pohlíženo jako na prvky tělesa $\mathbb{Z}_2$.

2.2. TYP ŠIFRY

Samotná šifra Lorenz je Vernamova typu, šifrový text je tvořen součtem otevřeného textu s pseudonáhodnou posloupností stejné délky, generovanou strojem Lorenz SZ. Sčítání jednotlivých znaků je definováno jako sčítání jejich Baudotových reprezentací po jednotlivých bitech, jak ukazuje následující příklad:

$$A+B=\begin{pmatrix}1\\1\\0\\0\\0\end{pmatrix}+\begin{pmatrix}1\\0\\0\\1\\1\end{pmatrix}=\begin{pmatrix}0\\1\\0\\1\\1\end{pmatrix}=G$$

Podle [1] se v běžném provozu zpráva tiskla na nekonečnou pásku, otevřený text proto nikdy neobsahoval řídicí znaky pro začátek nové řádky (CR, LF), rovněž se v něm nevyskytoval nulový znak. Pseudonáhodný klíč, a tedy i šifrový text, obsahoval všech 32 znaků kódu.

Během zkušebního provozu (červen 1941 až říjen 1942) byl text po zašifrování převeden do formátu používaného přístrojem *Hellschreiber* a přenášen v této formě. Hellschreiber, zařízení vynalezené v roce 1929 *Rudolfem Hellem*, je považován za předchůdce faxu. Přenášené znaky jsou vysílány po jednotlivých pixelech, sloupec tvoří 7 pixelů. Na přijímacím zařízení jsou znaky tisknuty na pásku. Jeho výhodou je především dobrá čitelnost textu i při nekvalitním spojení [5].

V únoru 1942 se přešlo na vysílání přímo v Baudotově kódu [1].

2.3. VNITŘNÍ KONSTRUKCE STROJE LORENZ SZ40

Šifrovací stroj Lorenz SZ40 je zástupcem ve své době velmi oblíbené třídy rotorových šifrovacích zařízení. Jeho hlavní část tvoří 12 rotorů vybavených po obvodu výklopnými kolíčky. Tyto kolíčky mohou být nastaveny do dvou poloh. Svislé aktivní postavení kolíčku (německy označované jako *Nocke*, výstupek, vačka), odpovídá binární hodnotě 1 a šikmé pasivní (*keine*, žádný) hodnotě 0. Nastavení všech kolíčků daného kola bude dále označováno jako *vzorek* tohoto kola (v anglických zdrojích je používán termín *wheel pattern*).

Počty kolíčků na jednotlivých kolech jsou vzájemně nesoudělné. V každém šifrovacím kroku je jeden z kolíčků každého kola *aktivní*, tzn. jeho nastavení ovlivňuje podobu klíče nebo další chování přístroje. Po otočení rotoru se stane aktivním následující kolíček.

Rotory lze podle funkce rozdělit do tří kategorií. Dvě pětičlenné skupiny, v Bletchley Parku označované jako kola ψ (ψ_1, ψ_2, ψ_3, ψ_4, ψ_5) a kola χ (χ_1, χ_2, χ_3, χ_4,

χ_5) vytváří klíč, zbylá dvě tzv. kola μ (μ_1, μ_2) se nazývají *řídicí* (anglicky *motor wheels*), protože řídí otáčení rotorů.

Z typografických důvodů bude v této práci při značení rotorů místo řeckého písmene ψ dále užíváno písmeno S, místo χ písmeno $\mathcal{K}$ a namísto písmene μ písmeno $\mathcal{M}$.

Počty kolíčků jednotlivých kol a jejich pořadí v přístroji zleva doprava shrnuje Tabulka 2.2.

rotor	počet kolíčků	pozice zleva
$\mathcal{K}_1$	41	8 (1)
$\mathcal{K}_2$	31	9 (2)
$\mathcal{K}_3$	29	10 (3)
$\mathcal{K}_4$	26	11 (4)
$\mathcal{K}_5$	23	12 (5)
$\mathcal{M}_1$	61	7
$\mathcal{M}_2$	37	6
S_1	43	1 (8)
S_2	47	2 (9)
S_3	51	3 (10)
S_4	53	4 (11)
S_5	59	5 (12)

Tab. 2.2: Počet kolíčků na jednotlivých rotorech a jejich umístění v přístroji. Pozice je převzata z [1], v závorce pozice uvedená v [2].

V každém kroku šifrování přístroj vygeneruje jeden znak klíče, tedy pětibitové slovo Baudotova kódu, následujícím jednoduchým způsobem: *i*-tý bit je součtem nastavení aktivních kolíčků kola S_i a kola $\mathcal{K}_i$.

Na konci šifrovacího kroku se některé rotory otočí o jednu pozici. Otáčení se řídí následujícími pravidly:

- kola $\mathcal{K}_i$ se otáčí v každém kroku
- kolo $\mathcal{M}_1$ se otáčí v každém kroku
- kolo $\mathcal{M}_2$ se otočí pouze tehdy, je-li hodnota aktivního kolíčku $\mathcal{M}_1$ (před otočením) 1
- kola S_i se všechna otočí pouze tehdy, pokud je hodnota aktivního kolíčku $\mathcal{M}_2$ (před případným otočením) 1; v opačném případě zůstávají všechna stát

Zaveďme následující notaci. Binární posloupnost generovanou během šifrování rotorem $\mathcal{K}_i$, $1 \le i \le 5$, budeme značit rovněž $\mathcal{K}_i$, přičemž z kontextu bude vždy zřejmé, zda jde o rotor, nebo jím tvořenou posloupnost. Posloupnost $\mathcal{K}_i$ je tedy periodickým rozšířením vzorku daného kola. Podobně označíme S_i, $1 \le i \le 5$, binární

posloupnost, kterou by vytvářelo kolo $\mathcal{S}_i$, pokud by se otáčelo v každém kroku. Příslušnou rozšířenou sekvenci, v níž se některé prvky opakují vlivem nepravidelného pohybu kola $\mathcal{S}_i$, budeme značit $\mathcal{S}_i'$. Písmenem $\mathcal{K}$ (respektive $\mathcal{S}$, $\mathcal{S}'$) bez indexu bude označována posloupnost znaků (nebo ekvivalentně příslušných slov Baudotova kódu), jejíchž pět impulsů tvoří posloupnosti $\mathcal{K}_i$ ($\mathcal{S}_i$, $\mathcal{S}_i'$).

S přihlédnutím k zavedenému značení lze šifrovací algoritmus charakterizovat rovnicí

$$\mathbf{C} = \mathbf{P} + \mathcal{K} + \mathcal{S}',$$

kde **C** je posloupnost šifrového textu a **P** posloupnost otevřeného textu.

Nepravidelný pohyb kol $\mathcal{S}_i$ měl zvýšit bezpečnost systému, avšak skutečnost, že se tyto rotory buď otáčely vždy všechny společně, nebo všechny společně stály, se ukázala být jednou z největších slabin systému. Důsledkem této vlastnosti je, že po sobě jdoucí znaky v posloupnosti $\mathcal{S}'$ jsou často shodné, díky čemuž lze tuto sekvenci odlišit od náhodné posloupnosti znaků (v praxi se při luštění využívalo nerovnoměrné distribuce bigramů v sekvenci $\mathbf{P} + \mathcal{S}'$).

2.4. KLÍČE A NASTAVENÍ PŘÍSTROJE

Klíč každé zprávy se skládá z několika částí, které lze rozdělit do dvou skupin. Dlouhodobý klíč tvoří vzorky kol a způsob kódování jejich počátečního nastavení. Klíč zprávy tvoří počáteční nastavení rotorů při šifrování dané zprávy.

Vzorky kol se na každé komunikační lince měnily v pravidelných intervalech. Podle [1] se během zkušebního provozu, tj. od června 1941 do října 1942, měnily vzorky kol $\mathcal{S}_i$ jednou za tři měsíce, vzorky kol $\mathcal{K}_i$ s měsíční periodou a vzorky řídicích kol $\mathcal{M}_i$ každý den. Po nasazení šifrovacího přístroje do ostrého provozu byla platnost vzorků kol $\mathcal{S}_i$ zkrácena na jeden měsíc. Od 1. srpna 1944 se všechny vzorky měnily denně. Tak časté změny klíče však s blížící se německou porážkou narážely na logistické problémy. Autoři zprávy [1] uvádějí, že se dokonce podařilo odposlechnout depeše obsahující nastavení přístroje na další období, což je postup porušující základní kryptografická pravidla.

Klíč zprávy, tzn. počáteční nastavení rotorů, se dohodnutým způsobem zakódoval a přenášel se pomocí indikátorové skupiny v otevřené hlavičce depeše. Během zmíněného zkušebního období měl indikátor podobu dvanácti písmen, přenášených pomocí německé hláskovací tabulky. Každému rotoru byla přiřazena jiná jednoduchá záměna vybraných počátečních pozic za písmena, která platila jeden měsíc.

Tento způsob předávání nastavení přístroje umožňoval kryptoanalytikům rozpoznat zprávy zašifrované s použitím stejného klíče (což je prohřešek proti kryptografickým pravidlům, jehož se němečtí operátoři často dopouštěli), které je možné snadno rozluštit (způsob je blíže popsán v kapitole 3). S přibývajícím počtem rozluštěných zpráv v daném období platnosti substitucí také rostl počet písmen v indikátorech, jejichž význam byl známý, což usnadňovalo luštění dalších depeší. Byla dokonce vyvinuta metoda, jak pomocí indikátorů nalézt vzorky kol.

S přechodem k ostrému provozu byly zavedeny číselné indikátory, přičemž operátoři na obou koncích komunikační linky měli pravděpodobně k dispozici stejnou tabulku s očíslovanými nastaveními. Takový indikátor stále umožňuje rozpoznat zprávy zašifrované stejným klíčem, ale neposkytuje žádné další informace.

Vlastní klíč každé zprávy je tvořen počátečním nastavením rotorů. Podle [1] během zkušebního provozu pravděpodobně radista volil počáteční nastavení kol sám (z těch, které měly substitucí přiřazeno nějaké písmeno). Po zavedení číselných indikátorů zřejmě se zřejmě postupně používala nastavení z předem dohodnutého očíslovaného seznamu.

Stroj Lorenz SZ má, podobně jako jiné rotorové šifrovací stroje, obrovskou mohutnost klíčového prostoru. Možných počátečních nastavení je více než 10^{19} (takový je součin velikostí všech rotorů). Možných vzorků všech kol je teoreticky

$$2^{23+26+29+31+37+41+43+47+51+53+59+61} = 2^{501} > 10^{150}.$$

Vzorky však nelze nastavit libovolně, protože některá nastavení produkují slabou pseudonáhodnou posloupnost. Německá strana užití takových vzorků bránila různými pravidly pro použitelná nastavení. Podle [1] se například vzorky kol $\mathcal{K}_1$ a S_1 sestavovaly s vyrovnaným počtem nul a jedniček a tak, aby neobsahovaly dlouhé posloupnosti stejných hodnot. Další pravidlo, označované jako $ab = ½$, je diskutováno v závěru kapitoly 3.

Zpráva [1] odhaduje počet použitelných nastavení vzorků kol $\mathcal{K}$ na 10^{38}.

Některá nastavení přístroje, přestože se liší vzorky kol nebo počátečními pozicemi rotorů, produkují stejnou pseudonáhodnou posloupnost. Zřejmým příkladem takové ekvivalence je, pokud jsou vzorky kol dvou nastavení šifrátoru vůči sobě cyklicky posunuté a rozdíl odpovídající tomuto posunutí je i mezi počátečními pozicemi rotorů. Méně triviální případ nastává, změníme-li pro nějaké i, $1 \le i \le 5$, nastavení každého kolíčku kola $\mathcal{K}_1$ a S_1 na opačné. Ekvivalence zde vyplývá z vlastností sčítání v tělese $\mathbb{Z}_2$.

2.5. POZDĚJŠÍ VERZE PŘÍSTROJE

V průběhu války byly s cílem zvýšit bezpečnost šifrovacího stroje zavedeny některé úpravy způsobu řízení společného otáčení kol S_1. Všechny fungovaly na stejném principu: v každém kroku se z vnitřního stavu přístroje spočítala jednobitová hodnota, nazývaná *omezení* (v anglických zdrojích *limitation*). Podmínka rotace kol S_1 se pak upravila následujícím způsobem:

- označme binární hodnotu aktivního kolíčku $\mathcal{M}_2$ (před případným otočením) m a aktuální hodnotu omezení ℓ. Kola S_1 se otočí právě tehdy, když platí

$$m \vee \neg \ell = 1.$$

Omezení se v závislosti na verzi přístroje počítalo jako součet některých z následujících hodnot:

- $\overline{\mathcal{K}_2}$, hodnota kolíčku kola $\mathcal{K}_2$ aktivního v předchozím kroku šifrování
- $\overline{\overline{S_1}'}$, hodnota kolíčku kola S_1 aktivního v předminulém kroku šifrování
- $\overline{\overline{\mathbf{P}_5}}$, hodnota 5. impulsu otevřeného textu v předminulém kroku šifrování

Poslední omezení, německy zvané *Klartextfunktion*, efektivně znemožňuje snadné luštění zpráv zašifrovaných stejným nastavením, protože pseudonáhodná posloupnost závisí i na otevřeném textu. Stejně efektivně ale znemožňuje dešifrování zbytku textu v případě, že dojde k chybnému příjmu některého znaku zašifrované zprávy. Podle [1] se tato funkce zkušebně používala v březnu 1943 na lince mezi Římem a armádou maršála Rommela v Tunisku a poté i na linkách v Evropě v prosinci 1943 a v roce 1944, pokaždé se ale od jejího užívání upustilo právě kvůli problémům s dešifrací v případě nedokonalého spojení.

Následující Tabulka 2.3 obsahuje označení jednotlivých verzí šifrovacího stroje Lorenz SZ, datum uvedení do provozu a omezení, která implementovala.

verze přístroje	**uvedení do provozu**	**používaná omezení**
Lorenz SZ40	červen 1941	žádná
Lorenz SZ42A	únor 1943	$\overline{\mathcal{K}_2}$ nebo $\overline{\mathcal{K}_2} + \overline{\overline{\mathbf{P}_5}}$
Lorenz SZ42B	červen 1944	$\overline{\mathcal{K}_2} + \overline{\overline{S_1}'}$ nebo $\overline{\mathcal{K}_2} + \overline{\overline{S_1}'} + \overline{\overline{\mathbf{P}_5}}$

Tab. 2.3: Verze šifrátoru Lorenz SZ, datum uvedení do provozu a používaná omezení (podle [1])

„The construction of long pieces
of key was very difficult…
On 30th August, 1941 the
German cipher operators came
to the rescue."
General Report on Tunny

3. ODVOZENÍ KONSTRUKCE ŠIFROVACÍHO STROJE

3.1. PRVNÍ POZNATKY

Podle [1] byly první zprávy šifrované strojem Lorenz SZ40 zachyceny a zkoumány v červnu roku 1941. Bezdrátový přenos probíhal ve formátu přístroje Hellschreiber, v květnu mu předcházelo zkušební vysílání v Baudotově kódu.

Zprávy měly standardizovanou podobu: nešifrovaná úvodní část obsahovala číslo depeše a sekvenci dvanácti slov německé hláskovací tabulky, zřejmě dvanáctipísmenný indikátor nastavení přístroje. Funkci mezery plnil znak *9*, sekvence *99999* pak uvozovala vlastní šifrový text. Kromě 26 písmen standardní abecedy se v textu vyskytovaly znaky *3*, *4*, *8*, *9*, + a */*.

Prvotní domněnku, že k vysílání zpráv je používán dálnopis a text poté převáděn z Baudotova kódu do formátu Hellschreiberu, potvrdilo 22. července zachycení několika zpráv, které obsahovaly pouze 16 různých znaků, přičemž z písmen abecedy se v nich vyskytovala právě ta, jejichž Baudotova reprezentace začíná nulou. Při jejich vysílání byl zřejmě dálnopis porouchaný a první bit každého znaku původní zprávy změnil na nulu. Z indikátorů zpráv, obsahujících například řetězce H/INRICH nebo TH/O3OR, bylo možné snadno odvodit Baudotovy reprezentace symbolů nepatřících mezi 26 písmen abecedy: / se kupříkladu zjevně liší od (známé) reprezentace *E* v hodnotě prvního bitu. Výsledkem byla následující převodní tabulka používaných znaků a jim odpovídajících slov Baudotova kódu (Tabulka 3.1), přičemž označení kontrolních znaků symboly *3*, *4*, *8*, *9*, + a */* se v Bletchley Parku pravděpodobně používalo jako konvence i poté, co byl Hellschreiber nahrazen v únoru 1942 přímou komunikací pomocí dálnopisů [1]. Některé zdroje používají místo znaku + cifru *5*, v textu [1] jsou na různých místech zmíněny obě varianty.

/	T	3	O	9	H	N	M	4	L	R	G	I	P	C	V	E	Z	D	B	S	Y	F	X	A	W	J	+	U	Q	K	8
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Tab. 3.1: Znaky používané na lince Vídeň-Athény při komunikaci pomocí přístroje Hellschreiber a příslušná slova Baudotova kódu

3.2. VERNAMOVA ŠIFRA

Zpráva [1] uvádí, že během testování nového šifrovacího zařízení byla zachycena řada zpráv se shodným indikátorem. Jejich zkoumání prokázalo, že pro šifrování je používán Vernamův systém, tedy aditivní proudová šifra, používající jako klíč produkci pseudonáhodného generátoru.

3.1 Definice. *Aditivním kryptosystémem* nazveme systém, ve kterém jsou prostor otevřených textů P, prostor šifrových textů C a prostor klíčů K shodné, tedy $P = C = K$, na tomto prostoru je zavedena operace sčítání tak, že $(P,+)$ tvoří aditivní grupu a šifrovací a dešifrovací operace jsou definovány vztahy

$$c = p + k,$$
$$p = c - k,$$

kde p je otevřený text, c šifrový text a k klíč.

Aditivní šifry mají známou slabinu: jsou-li dva texty zašifrovány s použitím stejného klíče, tj. pokud platí

$$c_1 = p_1 + k,$$
$$c_2 = p_2 + k,$$

pak rozdíl šifrových zpráv je rovný rozdílu otevřených zpráv:

$$c_1 - c_2 = (p_1 + k) - (p_2 + k) = p_1 - p_2.$$

Předpokládejme, že otevřené texty p_1, p_2 jsou stejně dlouhé. Zná-li nebo uhodne-li útočník část jedné z otevřených zpráv, snadno pomocí uvedené rovnosti dopočítá odpovídající úsek druhé zprávy. Díky redundanci textu pak pravděpodobně dokáže rozšířit známé části otevřených textů a v ideálním případě tak vyluštit obě zprávy, tj. získat takové texty p_1' a p_2', pro které platí

$$c_1 - c_2 = p_1' - p_2'.$$

Současně s otevřenými zprávami útočník získá i použitý klíč k.

Je-li operace sčítání definována tak, že je ekvivalentní s odčítáním (speciálně sčítání v $\mathbb{Z}_2$), nelze klíč použitý při šifrování určit jednoznačně bez dodatečné informace o tom, který šifrový text odpovídá kterému otevřenému textu. Za daného předpokladu totiž bude platit rovnost

$$c_1 - c_2 = p_1' - p_2' = p_2' - p_1',$$

a útočník při výpočtu klíče podle vztahu $k' = c_1 - p_1' = c_2 - p_2'$ dojde k jednomu ze dvou výsledků:

$$k' = k, \text{ pokud } p_1 = p_1' \text{ a } p_2 = p_2',$$

nebo

$$k' = k + (p_1 + p_2), \text{ pokud } p_1 = p_2' \text{ a } p_2 = p_1'.$$

Dodatečnou informaci vedoucí k vyřešení této nejednoznačnosti lze získat například tehdy, luští-li útočník různě dlouhé zprávy a podaří se mu určit otevřený

text v délce kratší zprávy; potom ke kratší šifrové zprávě bude zřejmě příslušet ten otevřený text, který je ze slohového hlediska rozumně ukončený.

Podle [1] útočili popsaným způsobem britští kryptologové na zprávy se shodným indikátorem a jejich úspěchy potvrdily domněnku, že je pro šifrování používána aditivní šifra a že operace sčítání je na kódových slovech Baudotova kódu definována jako operace XOR na odpovídajících si bitech. Zatím se však nedařilo získat souvislou sekvenci klíče o délce dostačující k tomu, aby bylo možné podrobit zkoumání algoritmus, podle kterého je klíč generován.

3.3. HQIBPEXEZMUG

Dne 30. srpna 1941 byly zachyceny dvě zprávy se shodným indikátorem HQIBPEXEZMUG (přezdívané podle něj „ZMUG“), kratší z nich měla délku 3976 znaků [1]. Tyto zprávy se shodovaly v prvních 7 znacích. Byl spočten rozdíl šifrových textů a jako začátek otevřeného textu jedné ze zpráv bylo vyzkoušeno v německé komunikaci často používané slovo SPRUCHNUMMER („číslo zprávy“), výsledkem byl řetězec SPRUCHNR9++U na začátku druhé zprávy.

Tyto dvě zprávy byly vyluštěny v celé délce kratší zprávy, o což se dvouměsíční prací zasloužil plukovník John H. Tiltman [1-3]. Ukázalo se, že jde obsahově o dvě verze téže zprávy, lišící se pouze v použití zkratek, interpunkce, v překlepech atd., což výrazně usnadnilo luštění.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
p_1	S	P	R	U	C	H	N	U	M	M	E	R	9	+	+	U	P	W	U	9
c_1	J	S	H	4	N	Z	Y	M	F	S	/	8	8	4	I	V	K	U	8	Y
c_1-c_2	0	0	0	0	0	0	0	F	O	U	G	F	L	3	M	A	Q	S	G	4
c_2	J	S	H	4	N	Z	Y	Z	Y	4	G	L	F	R	G	X	O	4	S	Q
p_2	S	P	R	U	C	H	N	R	9	+	+	U	P	W	U	9	E	P	X	I

Tab. 3.2: Prvních 20 znaků šifrových zpráv zachycených 30. 8. 1941, jejich rozdílu a příslušných otevřených textů. Převzato z [2] s úpravou označení kontrolních znaků

Díky tomuto hrubému porušení bezpečnostních pravidel ze strany německého operátora se podařilo britským analytikům získat téměř souvislou sekvenci klíče o délce 3976 znaků, která stačila k rozbití celého šifrovacího algoritmu.

Podle [2] byly důvodem opakovaného poslání zprávy atmosférické poruchy, které poškodily první zprávu. Pokud by radista poslal zprávu znovu v přesně stejném znění, ke kompromitaci by nedošlo. Při psaní zprávy na klávesnici ale byly odchylky v interpunkci nebo mezerách velmi pravděpodobné, navíc operátor při druhém vysílání zřejmě kvůli únavě častěji používal zkratky.

Zpráva [1] uvádí, že německá strana si možná toto ohrožení bezpečnosti uvědomila, protože „rádiový provoz na lince na několik dní téměř utichl a do konce roku 1941 už nebyly zachyceny žádné další zprávy šifrované stejným klíčem.“

3.4. ÚVODNÍ POZNÁMKY K ANALÝZE KLÍČE

Použité zdroje neuvádějí plné znění zpráv z 30. srpna 1941, ani celou sekvenci klíče. Softwarový simulátor šifrovacího stroje Lorenz SZ dostupný na internetu [6] však obsahuje příslušné vzorky všech kol přístroje a kniha [2] pak prvních 120 znaků zpráv a klíče. Tyto informace byly použity k nalezení počátečních nastavení kol (vyzkoušením všech možností s pomocí počítače). Pomocí vlastního softwarového simulátoru, který je součástí této práce, bylo vygenerováno 4000 znaků klíče, které jsou podrobeny následující analýze.

Nutno poznamenat, že prvních 120 znaků vygenerovaného klíče se v některých pozicích neshoduje se sekvencí uvedenou v [2]. Možným vysvětlením může být, že ve zmíněném zdroji jsou uvedeny šifrové zprávy tak, jak byly zachyceny, tj. s případnými chybami v přenosu, a do klíče pak jsou tyto chyby vneseny jeho výpočtem, zatímco klíč generovaný pomocí softwarového simulátoru je přenosových chyb prostý. Další možností jsou případné tiskové chyby či nepřesnosti v knize [2], po kterých však v rámci této práce pátráno nebylo.

Následující analýza klíče je volně inspirována postupem pracovníků Bletchley Parku, mezi které patřil i W. T. Tutte, popsaným v [1] a [3], nicméně striktně se ho nedrží a postupuje po vlastní linii, což je ostatně vzhledem ke stručnosti a obecnosti popisu v obou zdrojích jediná možnost. Některé Tutteho objevy navíc byly podmíněny velkou dávkou náhody a štěstí (viz [3]), což jsou prvky, jejichž vliv je jen těžko možné experimentálně zopakovat.

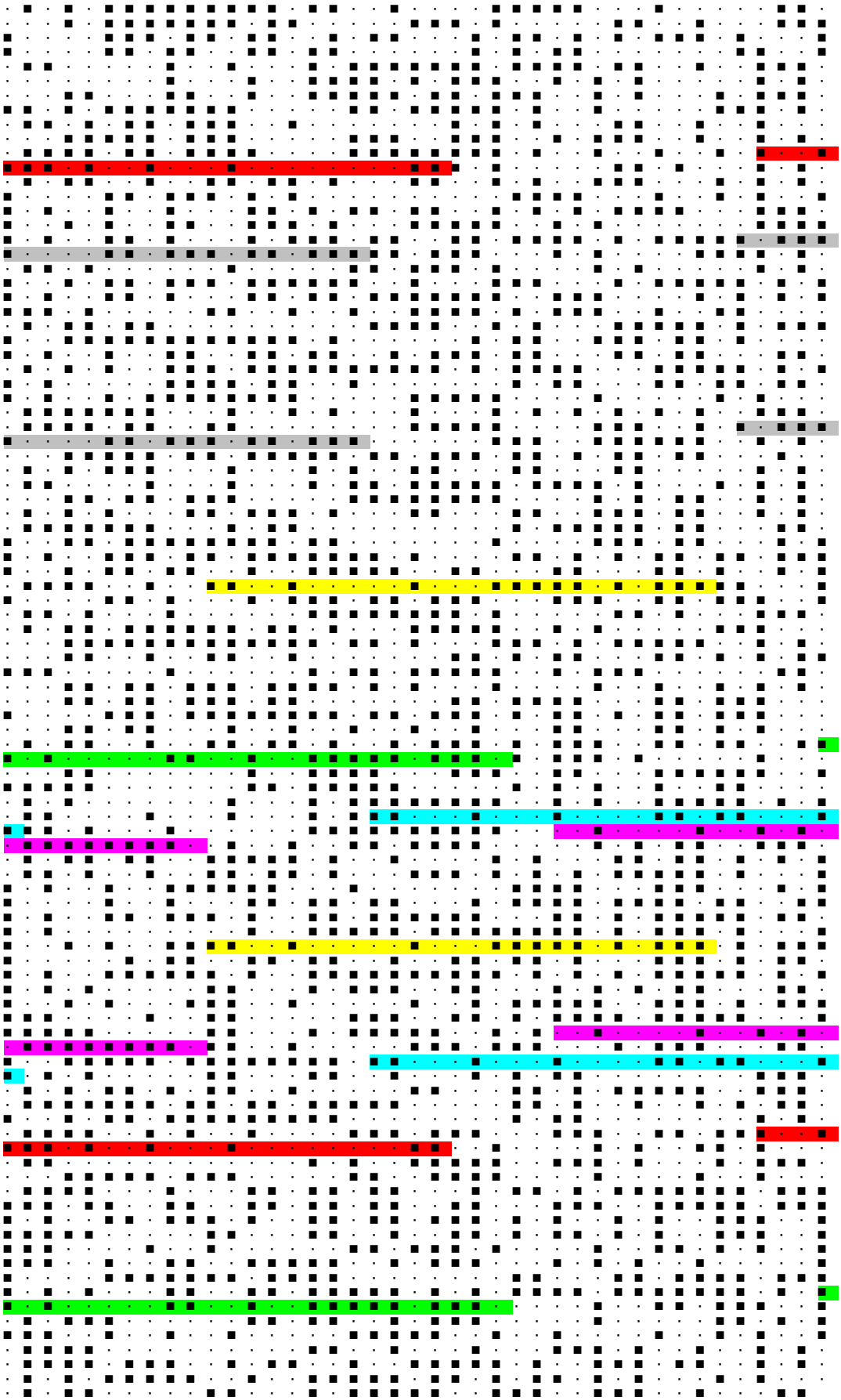
Na závěr jednotlivých fází analýzy je pro porovnání s použitými metodami vždy uveden postup, kterým k výsledku dospěli britští analytici.

3.5. ANALÝZA KLÍČE: PRVNÍ POHLED

Vzhledem k tomu, že sčítání znaků klíče a otevřeného textu se provádí po jednotlivých bitech, je přirozené dívat se na sekvenci klíče nikoli jako na sekvenci znaků, ale jako na pět binárních impulsů, a zkoumat je jednotlivě.

Tomu nahrává i informace, kterou podle [1] Britové získali: úspěšně luštili fragmenty zpráv, jejichž indikátory se shodovaly až na první písmeno, s využitím předpokladu, že se klíče použité k zašifrování těchto zpráv liší pouze v prvním impulsu. Tím dokázali, že první písmeno indikátoru ovlivňuje pouze první impuls klíče, a tedy že impulsy klíče jsou alespoň do jisté míry generovány nezávisle.

Prozkoumejme tedy první impuls klíče, znázorněný na následujícím Obrázku 3.3.



Obr. 3.3: První impuls klíče zprávy „ZMUG“ v délce 4000 znaků s barevně zvýrazněnými shodnými částmi vypsaný do řádků o 41 znacích. Tečky odpovídají nulám, čtverečky jedničkám.

Budeme-li v prvním impulsu klíče hledat opakující se posloupnosti (což je v literatuře někdy označováno jako *Kasiského test*), zjistíme, že v něm je možné najít hned několik vícekrát se vyskytujících sekvencí, nejdelší (na Obrázku 3.3 červeně a zeleně označené) mají délku 26 znaků. Lze spočítat, že tato opakování jsou od sebe vzdálena o násobky čísla 41.

Touto vlastností připomíná zkoumaný klíč text zašifrovaný pomocí šifry s periodickým klíčem (např. Vigenèrovy šifry), kde se shodné části otevřeného textu vzdálené od sebe o násobek délky periody klíče zašifrují na shodné části šifrového textu. Můžeme tedy vyslovit hypotézu, že první impuls klíče je bit po bitu součtem dvou sekvencí: první z nich je periodická s periodou 41 (označíme ji $\mathcal{K}_1$, což, jak vyjde najevo, bude ve shodě s označením osmého kola šifrovacího stroje Lorenz SZ), o druhé (kterou ze stejných důvodů označíme S_1') zatím nemáme žádné informace, kromě toho, že obsahuje dlouhé opakující se části.

Vyslovenou hypotézu lze ověřit tím, se pokusíme získat posloupnosti $\mathcal{K}_1$ a S_1' s využitím postupů od 19. století používaných k luštění Vigenèrovy šifry.

3.6. HLEDÁNÍ PERIODY

K určení periody posloupnosti $\mathcal{K}_1$ použijeme index koincidence, který nyní zavedeme.

3.2 Definice. Buď $T = t_1t_2t_3\ldots t_m$ text délky m. *Index koincidence IC(T)* je pravděpodobnost, že pro náhodně vybrané indexy i, j, $1 \le i < j \le m$, platí $t_i = t_j$.

Výpočet indexu koincidence

Buď $A = \{a_1, a_2,\ldots, a_n\}$ množina všech znaků textu T, tzn. $t_i \in A \ \forall i = 1,\ldots,m$. Buď f_i, $i = 1,\ldots,n$, počet výskytů znaku a_i v textu T. Pak

$$IC(T) = \frac{\sum_{i=1}^{n}\binom{f_i}{2}}{\binom{m}{2}} = \frac{\sum_{i=1}^{n} f_i(f_i-1)}{m_i(m_i-1)}.$$

Index koincidence vyjadřuje míru rozdílnosti frekvencí jednotlivých znaků v textu. Minimální hodnoty dosahuje u náhodného textu, ve kterém je všech n znaků použité abecedy stejně četných. Jeho limita s rostoucí délkou textu je v tomto případě $1/n$.

Index koincidence textu se nezmění, jsou-li znaky textu změněny jednoduchou substitucí, protože jejich relativní četnost zůstává stejná. Toho lze využít při hledání periody klíče polyalfabetické šifry s periodickým klíčem, což je i náš cíl: vypíše-li se šifrový text znak po znaku a řádek po řádku do tabulky o l sloupcích, a spočte-li se *IC* každého z těchto sloupců zvlášť, pak aritmetický průměr získaných indexů koincidence bude výrazně vyšší (a blízký *IC* otevřeného textu) pro l rovné násobku periody klíče, protože v tom případě jsou znaky v jednom každém sloupci šifrovány

jednoduchou záměnou. Pro jiný počet sloupců bude průměrný index koincidence nižší, protože pak jsou znaky v jednom sloupci obecně šifrovány různými substitucemi, což rozdíly v relativní četnosti znaků stírá.

Tento pokus provedeme s prvním impulsem zkoumaného klíče. Protože ale abeceda tohoto textu obsahuje jen dva znaky, jejichž počty jsou v klíči vyrovnané, budeme pro účel výpočtu *IC* za „znak“ považovat bigram, tj. dvojici sousedních znaků, s nadějí, že rozložení bigramů v posloupnosti S_1' je dostatečně nerovnoměrné a rozdíly mezi průměry indexů koincidence budou signifikantní. Výsledky tento předpoklad potvrdí; kdyby tomu tak nebylo, mohli bychom se pokusit uspět s *n*-gramy větší délky.

V tabulce o *l* sloupcích tedy budeme počítat index koincidence bigramů ve dvojicích sloupců 1; 2, 3; 4 atd., až po dvojici $l-1$; *l* pro *l* sudá, nebo dvojici $l-2$; l - 1 pro liché hodnoty *l*, jak je znázorněno v Tabulkách 3.4 a 3.5:

1	2	3	4	5	6	7	8	9	10
0	1	0	1	0	1	1	1	1	1
1	1	1	1	0	1	1	0	0	1
0	0	0	0	1	1	1	1	1	0
0	0	1	0	0	0	0	0	1	1
				…					

Tab. 3.4: První impuls klíče vypsaný do tabulky o 10 sloupcích se zvýrazněnými dvojicemi sloupců pro výpočet *IC*

1	2	3	4	5	6	7	8	9	10	11
0	1	0	1	0	1	1	1	1	1	1
1	1	1	0	1	1	0	0	1	0	0
0	0	1	1	1	1	1	0	0	0	1
0	0	0	0	0	1	1	0	0	0	0
					…					

Tab. 3.5: První impuls klíče vypsaný do tabulky o 11 sloupcích se zvýrazněnými dvojicemi sloupců pro výpočet *IC*

Následující Tabulka 3.6 obsahuje aritmetické průměry indexů koincidence dvojic sloupců pro počet sloupců *l* od 2 do 99.

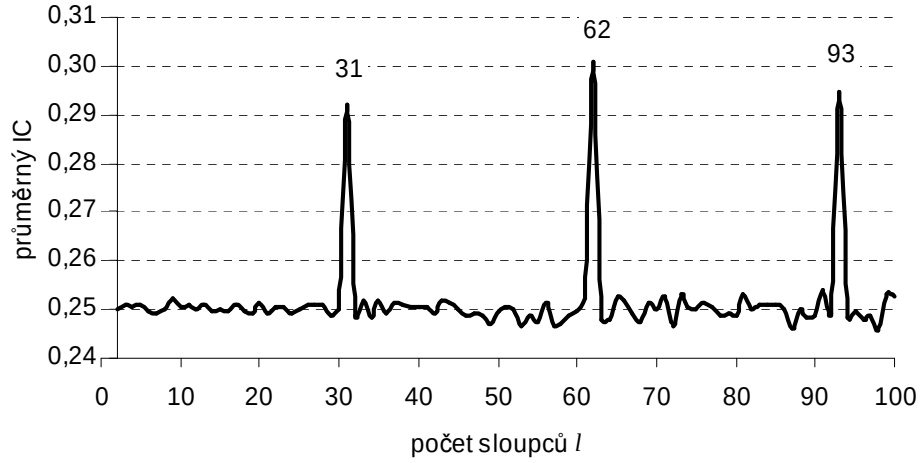
l	pr. IC	l	pr. IC	l	pr. IC	l	pr. IC	l	pr. IC	l	pr. IC	l	pr. IC
2	0,2503	16	0,2501	30	0,2486	44	0,2521	58	0,2481	72	0,2475	86	0,2477
3	0,2500	17	0,2500	31	0,2502	45	0,2498	59	0,2534	73	0,2492	87	0,2533
4	0,2502	18	0,2496	32	0,2494	46	0,2500	60	0,2474	74	0,2535	88	0,2525
5	0,2505	19	0,2484	33	0,2498	47	0,2500	61	0,2506	75	0,2457	89	0,2485
6	0,2498	20	0,2496	34	0,2532	48	0,2494	62	0,2502	76	0,2479	90	0,2502
7	0,2504	21	0,2501	35	0,2498	49	0,2497	63	0,2495	77	0,2495	91	0,2467
8	0,2503	22	0,2516	36	0,2489	50	0,2484	64	0,2530	78	0,2468	92	0,2506
9	0,2506	23	0,2497	37	0,2531	51	0,2503	65	0,2500	79	0,2501	93	0,2479
10	0,2496	24	0,2505	38	0,2486	52	0,2487	66	0,2503	80	0,2468	94	0,2524
11	0,2503	25	0,2490	39	0,2474	53	0,2490	67	0,2490	81	0,2467	95	0,2504
12	0,2496	26	0,2498	40	0,2481	54	0,2495	68	0,2553	**82**	**0,2985**	96	0,2469
13	0,2497	27	0,2499	**41**	**0,2930**	55	0,2483	69	0,2480	83	0,2491	97	0,2483
14	0,2509	28	0,2506	42	0,2496	56	0,2492	70	0,2492	84	0,2492	98	0,2476
15	0,2494	29	0,2486	43	0,2503	57	0,2484	71	0,2494	85	0,2505	99	0,2496

Tab. 3.6: Průměrné indexy koincidence pro jednotlivé počty sloupců *l*.

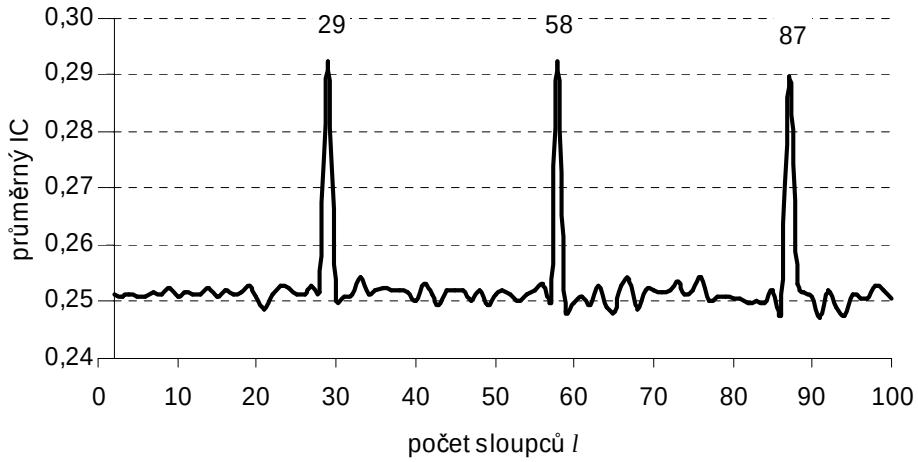
Z Tabulky 3.6 je patrné, že pro téměř všechny hodnoty *l* je průměrná hodnota *IC* blízká 0,25, což je ve zkoumaném případě teoretický index koincidence náhodného textu (množina všech možných bigramů je čtyřprvková). Výrazně vyšší hodnotu má index koincidence pouze pro hodnoty 41 a 82. Přitom 41 je délka periody očekávaná na základě výsledků Kasiského testu.

V Grafech 3.7 až 3.11 jsou vyneseny hodnoty průměrných indexů koincidence v závislosti na hodnotě *l* pro všech pět impulsů zkoumaného klíče. Příslušné tabulky hodnot pro impulsy 2 až 5 jsou uvedeny na přiloženém CR-ROM.

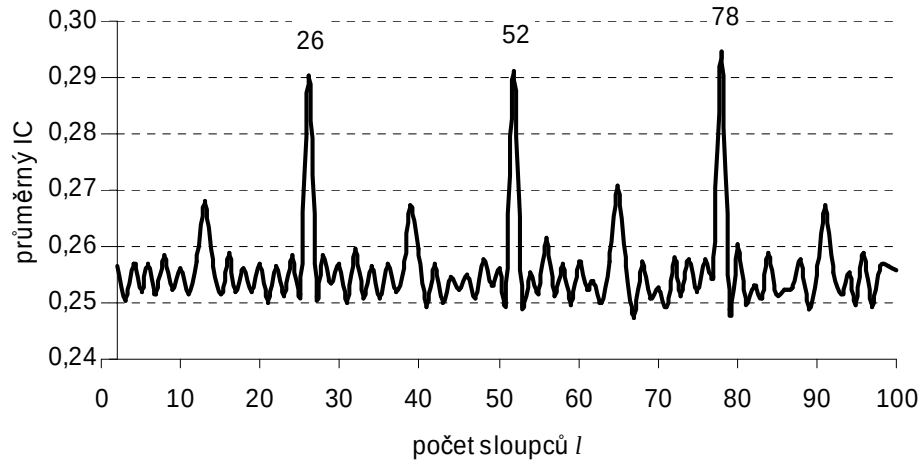
Graf 3.7: Průměrný index koincidence v závislosti na počtu sloupců l pro 1. impuls klíče



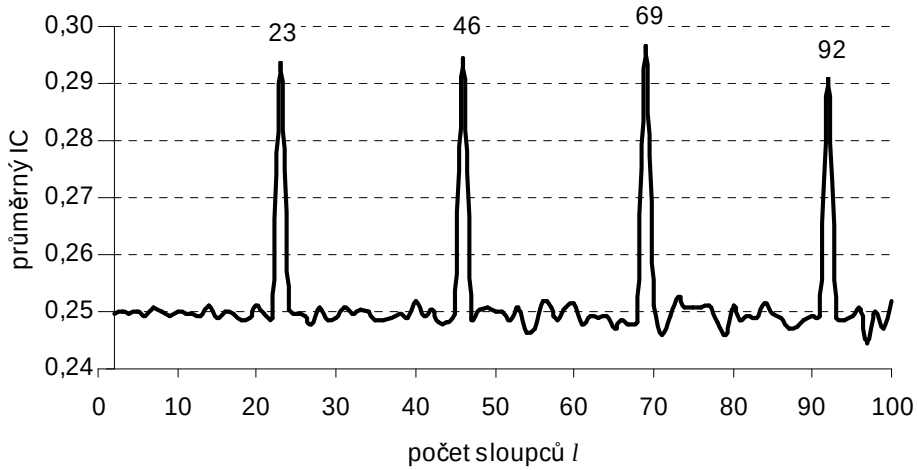
Graf 3.8: Průměrný index koincidence v závislosti na počtu sloupců l pro 2. impuls klíče



Graf 3.9: Průměrný index koincidence v závislosti na počtu sloupců l pro 3. impuls klíče



Graf 3.10: Průměrný index koincidence v závislosti na počtu sloupců l pro 4. impuls klíče



Graf 3.11: Průměrný index koincidence v závislosti na počtu sloupců l pro 5. impuls klíče

Grafy potvrzují hypotézu, že *i*-tý impuls, $1 \le i \le 5$, je bit po bitu součtem periodické posloupnosti $\mathcal{K}_i$ a zatím neznámé posloupnosti S_i'. Jsou z nich patrné i délky period: $\mathcal{K}_1$ má periodu 41, $\mathcal{K}_2$ periodu 31, $\mathcal{K}_3$ periodu 29, $\mathcal{K}_4$ periodu 26 a $\mathcal{K}_5$ má periodu 23. Zajímavý je Graf 3.10, v němž je hodnota indexu koincidence významně větší než 0,25 pro všechna *l* soudělná s 26 (tj. sudá čísla a násobky 13).

Postup britských kryptoanalytiků.

Časté opakující se sekvence vzdálené o násobky čísla 41 byly v prvním impulsu objeveny v podstatě náhodně W. T. Tuttem. Podle [3] pracoval Tutte s informací, že na poslední pozici indikátoru se každý měsíc u všech zpráv vyskytovalo pouze 23 různých písmen, zatímco na ostatních pozicích se mohlo objevit kterékoli z 25 používaných písmen (v indikátoru se nikdy nepoužíval znak J). Očekával tedy periodu 23 nebo 25 a proto vypsal první impuls do tabulky o $23 \cdot 25 = 575$ sloupcích. Všiml si častých diagonálních opakování, tzn. shodných sekvencí vzdálených o násobky 574. Dále už pracoval s číslem 41, které je dělitelem 574.

U dalších impulsů byly podle [1] periody získány metodou hledání opakujících se sekvencí délky 7 a největších společných dělitelů jejich vzdáleností v textu.

3.7. HLEDÁNÍ POSLOUPNOSTÍ $\mathcal{K}_i$

Způsob rozložení impulsu klíče na součet periodické posloupnosti $\mathcal{K}_i$ a posloupnosti S_i' bude detailně předveden pouze pro pátý impuls, protože $\mathcal{K}_5$ má nejkratší periodu. U zbývajících budou uvedeny pouze výsledky, které byly získány stejnou metodou, s malou modifikací v případě $\mathcal{K}_4$, která bude vysvětlena.

Vypočtené hodnoty indexu koincidence naznačují, že frekvence výskytu bigramů v posloupnosti S_5' není vyrovnaná. Vepišme tedy pátý impuls klíče do tabulky s 23 sloupci a spočtěme výskyty jednotlivých bigramů pro každou dvojici sousedních sloupců. Následující Tabulka 3.12 obsahuje získané hodnoty pro jednotlivé dvojice sloupců.

1; 2		**2; 3**		**3; 4**		**4; 5**		**5; 6**		**6; 7**		**7; 8**		**8; 9**		**9; 10**		**10; 11**		**11; 12**	
00	20	00	20	00	23	00	45	00	55	**00**	**67**	00	26	00	59	**00**	**61**	00	35	00	22
01	**62**	01	62	**01**	**67**	01	32	01	14	01	22	**01**	**71**	01	22	01	29	01	56	**01**	**76**
10	**62**	**10**	**70**	10	54	10	24	10	34	10	30	10	55	10	31	10	30	**10**	**63**	10	58
11	30	11	22	11	30	**11**	**73**	**11**	**71**	11	55	11	22	**11**	**62**	11	54	11	20	11	18

12; 13		**13; 14**		**14; 15**		**15; 16**		**16; 17**		**17; 18**		**18; 19**		**19; 20**		**20; 21**		**21; 22**		**22; 23**	
00	32	**00**	**71**	**00**	**74**	00	31	**00**	**63**	**00**	**63**	**00**	**62**	00	23	00	57	00	23	00	28
01	48	01	25	01	24	**01**	**63**	01	27	01	25	01	30	**01**	**66**	01	25	**01**	**62**	01	54
10	**64**	10	27	10	20	10	59	10	25	10	29	10	27	10	59	10	29	10	59	**10**	**65**
11	30	11	51	11	56	11	21	11	59	11	57	11	55	11	26	**11**	**63**	11	29	11	26

Tab. 3.12: Celkové počty jednotlivých bigramů v jednotlivých dvojicích sloupců. Tučným písmem je vyznačen nejčastější bigram v každé dvojici sloupců.

V každé zkoumané dvojici sloupců jsou dva páry přibližně stejně četných bigramů. Přitom bigramy v takovém páru vždy, nahlížíme-li na ně jako na dvouprvkové binární vektory (což budeme v následujícím textu dělat často), mají součet rovný vektoru (1, 1). Četnosti početnějšího páru bigramů jsou přibližně dvojnásobné v porovnání s četnostmi bigramů méně častého páru.

Naším cílem je najít čísla $\boldsymbol{k_i} \in \{0, 1\}$, $1 \leq i \leq 23$, která tvoří první periodu posloupnosti $\mathcal{K}_5$. To znamená, že přičteme-li pro všechna uvažovaná i číslo $\boldsymbol{k_i}$ ke každému číslu v i-tém sloupci (modulo 2), získáme v tabulce posloupnost $\boldsymbol{S}_5'$.

V této fázi budeme opět postupovat analogicky s luštěním Vigenèrovy šifry: byla-li již odhalena délka periody klíče a šifrový text vypsán do tabulky s počtem sloupců rovným této periodě, následuje převod polyalfabetické substituce na substituci jednoduchou, konkrétně Caesarovu. Pro druhý, třetí a každý další sloupec se hledá takový cyklický posun abecedy, po němž bude rozložení konkrétních znaků daného sloupce odpovídat rozložení znaků ve sloupci prvním. Je-li na znacích abecedy obvyklým způsobem zavedeno sčítání, je tento cyklický posun roven rozdílu prvního a druhého (resp. třetího atd.) znaku klíče šifry. (Shodu mezi rozloženími znaků ve dvou sloupcích lze měřit vzájemným indexem koincidence daných sloupců, ten ale v této práci zaváděn nebude, protože jeho použití není vzhledem k malému počtu různých cyklických posunů binárních bigramů nutné.) Jakmile jsou všechny sloupce počínaje druhým převedeny cyklickými posuny tak, že rozložení znaků ve sloupcích si maximálně odpovídá, získáme text zašifrovaný Caesarovou šifrou s klíčem rovným prvnímu znaku klíče původní polyalfabetické záměny. Klíč polyalfabetické substituce pak lze odvodit z použitých cyklických posunů.

V případě zkoumaného impulsu je cyklický posun bigramů ve dvojici sloupců i; $i + 1$ realizován přičtením čísla $a \in \{0, 1\}$ k hodnotám ve sloupci i a čísla $b \in \{0, 1\}$ k hodnotám ve sloupci $i + 1$, neboli vektoru (a, b) ke každému bigramu. Existují tedy právě čtyři různé posuny, které ztotožníme s odpovídajícími vektory (a, b).

Definujme pojem, který budeme často používat.

3.3 Definice. Řekneme, že rozložení bigramů ve dvojici sloupců i; $i + 1$ a ve dvojici sloupců j; $j + 1$ *si odpovídají*, jestliže bigramy tvořící četnější pár v první dvojici sloupců tvoří četnější pár i ve druhé dvojici, a bigramy, které tvoří méně četný pár v první dvojici sloupců, tvoří méně četný pár i ve druhé dvojici.

3.4 Poznámka. Je zřejmé, že vždy existují dva cyklické posuny, kterými lze rozložení bigramů v jedné dvojici sloupců dovést do stavu odpovídajícího rozložení bigramů v jiné dvojici sloupců. Tyto cyklické posuny, vyjádřené jako dvouprvkové binární vektory, mají součet (1, 1), neboť takový součet mají i bigramy v každém páru podobně četných bigramů.

Přitom se nedá jednoznačně rozhodnout, pro který z těchto cyklických posunů si rozložení odpovídají „lépe“. To v důsledku povede k nejednoznačnému určení posloupnosti $\mathcal{K}_5$.

Nyní hledejme takové cyklické posuny, po kterých bude rozložení bigramů ve dvojici sloupců 2; 3, odpovídat rozdělení bigramů ve dvojici sloupců 1; 2. Z Tabulky 3.13 je patrné, že jde o posuny (0, 0) a (1, 1):

1; 2		**2; 3 + (0, 0)**		**2; 3 + (0, 1)**		**2; 3 + (1, 0)**		**2; 3 + (1, 1)**	
00	20	00	20	00	62	00	70	00	22
01	62	01	62	01	20	01	22	01	70
10	62	10	70	10	22	10	20	10	62
11	30	11	22	11	70	11	62	11	20

Tab. 3.13: Četnosti bigramů ve dvojici sloupců 1; 2 a četnosti bigramů ve dvojici sloupců 2; 3 po jednotlivých cyklických posunech

Využijeme skutečnosti, že označíme-li cyklický posun, který dovede rozložení bigramů ve dvou dvojicích sloupců i; $i + 1$ a j; $j + 1$ do odpovídajícího si stavu, jako (a, b), platí (s přihlédnutím k Poznámce 3.4)

$$(\boldsymbol{k}_i, \boldsymbol{k}_{i+1}) + (\boldsymbol{k}_j, \boldsymbol{k}_{j+1}) = (a, b) + \delta\,(1, 1), \text{ kde } \delta \in \{0, 1\}.$$

To je fakt analogický luštění Vigenèrovy šifry, kde, jak bylo zmíněno, je cyklický posun dvou sloupců, který maximalizuje shodu v rozložení znaků v těchto sloupcích, roven rozdílu příslušných znaků klíče.

Pokud tedy budeme s ohledem na Poznámku 3.4 vybírat cyklický posun vždy ve tvaru $(0, b)$, pro dvojice sloupců 1; 2 a 2; 3 platí

$$(\boldsymbol{k}_1, \boldsymbol{k}_2) + (\boldsymbol{k}_2, \boldsymbol{k}_3) = (0, 0) + \delta\,(1, 1),$$

a tedy nezávisle na hodnotě δ je

$$\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}_2 + \boldsymbol{k}_3$$
$$\boldsymbol{k}_3 = \boldsymbol{k}_1.$$

Nyní hledejme cyklický posun, který uvede dvojici sloupců 3; 4 do odpovídajícího stavu s dvojicí 1; 2. Pohled do Tabulky 3.12 řekne, že je to opět posun (0, 0). Platí tedy (hodnota δ může být obecně různá od hodnoty v předchozí rovnosti)

$$(\boldsymbol{k}_1, \boldsymbol{k}_2) + (\boldsymbol{k}_3, \boldsymbol{k}_4) = (0, 0) + \delta\,(1, 1)$$
$$\boldsymbol{k}_1 + \boldsymbol{k}_3 = \boldsymbol{k}_2 + \boldsymbol{k}_4$$

a s ohledem na předchozí výsledek

$$\boldsymbol{k}_4 = \boldsymbol{k}_2.$$

Pokračujme dvojicí sloupců 4; 5. Zde je posun roven (0, 1) a platí

$$(\boldsymbol{k}_1, \boldsymbol{k}_2) + (\boldsymbol{k}_4, \boldsymbol{k}_5) = (0,1) + \delta\,(1,1)$$
$$\boldsymbol{k}_1 + \boldsymbol{k}_4 = \boldsymbol{k}_2 + \boldsymbol{k}_5 + 1$$
$$\boldsymbol{k}_5 = \boldsymbol{k}_1 + 1.$$

Tímto způsobem můžeme pokračovat a získat tak všechna k_i vyjádřená střídavě pomocí k_1, nebo k_2. Následující Tabulka 3.14 obsahuje ke každé dvojici sloupců i; $i + 1$ příslušný cyklický posun vůči dvojici 1; 2 a výsledek pro k_{i+1}.

Dvojice sloupců	Posun vůči 1; 2	Výsledek
1; 2		k_1, k_2
2; 3	(0, 0)	$k_3 = k_1$
3; 4	(0, 0)	$k_4 = k_2$
4; 5	(0, 1)	$k_5 = k_1 + 1$
5; 6	(0, 1)	$k_6 = k_2$
6; 7	(0, 1)	$k_7 = k_1 + 1$
7; 8	(0, 0)	$k_8 = k_2 + 1$
8; 9	(0, 1)	$k_9 = k_1$
9; 10	(0, 1)	$k_{10} = k_2 + 1$
10; 11	(0, 0)	$k_{11} = k_1 + 1$
11; 12	(0, 0)	$k_{12} = k_2 + 1$
12; 13	(0, 0)	$k_{13} = k_1 + 1$
13; 14	(0, 1)	$k_{14} = k_2$
14; 15	(0, 1)	$k_{15} = k_1 + 1$
15; 16	(0, 0)	$k_{16} = k_2 + 1$
16; 17	(0, 1)	$k_{17} = k_1$
17; 18	(0, 1)	$k_{18} = k_2 + 1$
18; 19	(0, 1)	$k_{19} = k_1$
19; 20	(0, 0)	$k_{20} = k_2$
20; 21	(0, 1)	$k_{21} = k_1 + 1$
21; 22	(0, 0)	$k_{22} = k_2 + 1$
22; 23	(0, 0)	$k_{23} = k_1 + 1$

Tab. 3.14: Cyklický posun jednotlivých dvojic sousedních sloupců vzhledem k dvojici 1; 2 a získané výsledky pro hodnoty k_i

Vztah mezi k_1 a k_2 získáme stejným postupem srovnáním rozložení bigramů ve dvojicích sloupců 1; 2 a 23; 1. Bigramy ve dvojici sloupců 23; 1 jsou tvořeny znakem ve sloupci 23 a znakem ve sloupci 1 na následujícím řádku, tedy dvojicemi po sobě jdoucích znaků začínajících ve zkoumaném impulsu na pozicích, které jsou násobky 23.

1; 2		23; 1	
00	20	00	26
01	62	01	67
10	62	10	55
11	30	11	25

Tab. 3.15: Počty bigramů ve dvojicích sloupců 1; 2 a 23; 1

Z Tabulky 3.15 je patrné, že cyklický posun dvojice sloupců 23; 1 vůči 1; 2 je (0, 0). Platí tedy

$$(\boldsymbol{k_1}, \boldsymbol{k_2}) + (\boldsymbol{k_{23}}, \boldsymbol{k_1}) = (0,0) + \delta(1,1)$$
$$\boldsymbol{k_1} + \boldsymbol{k_{23}} = \boldsymbol{k_2} + \boldsymbol{k_1}$$
$$\boldsymbol{k_1} + \boldsymbol{k_1} + 1 = \boldsymbol{k_2} + \boldsymbol{k_1}$$
$$\boldsymbol{k_2} = \boldsymbol{k_1} + 1.$$

Zjistili jsme, jak vypadá posloupnost $\mathcal{K}_5$ v závislosti na volbě $\boldsymbol{k_1}$, tedy až na záměnu nul a jedniček. To je nejednoznačnost zmíněná v Poznámce 3.4. Je z principu neodstranitelná, odpovídá totiž vlastnosti šifrovacího stroje Lorenz SZ40 zmíněné ve druhé kapitole: změníme-li pro některé i, $1 \le i \le 5$, nastavení každého kolíčku kola $\mathcal{K}_i$ a každého kolíčku kola S_i na opačné, stroj bude produkovat stejný klíč.

Můžeme zavést konvenci, že každá posloupnost $\mathcal{K}_i$ začíná nulou. Tabulka 3.16 obsahuje první periody těchto posloupností.

$\mathcal{K}_1$	0 1 1 0 0 1 1 1 0 0 0 0 1 1 1 0 0 1 1 1 0 0 0 0 1 0 0 1 1 0 1 1 0 0 0 1 1 0 1 1 0
$\mathcal{K}_2$	0 0 0 1 1 0 0 1 1 0 0 0 1 0 1 1 1 1 0 1 1 1 0 0 0 0 1 1 0 1 1
$\mathcal{K}_3$	0 1 1 1 1 0 1 1 0 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 0 0 1 0 0
$\mathcal{K}_4$	0 1 1 0 0 0 1 0 1 1 0 0 1 1 1 0 0 1 1 0 1 0 1 0 0 1
$\mathcal{K}_5$	0 1 0 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 1 1 0 1

Tab. 3.16: První periody sekvencí $\mathcal{K}_i$ klíče zpráv HQIBPEXEZMUG

Tyto výsledky byly získány předvedeným postupem, potřebné tabulky s počty bigramů pro všechny impulsy jsou uvedeny na přiloženém CD-ROM. Postup vyžaduje malou modifikaci v případě posloupnosti $\mathcal{K}_4$. Tato posloupnost má totiž sudou periodu 26 a porovnání rozložení bigramů ve sloupcích 1; 2 a 26; 1 proto informaci o vztahu prvních dvou bitů posloupnosti $\mathcal{K}_4$ nepřinese. Můžeme ji však odvodit jiným způsobem. Nyní totiž známe posloupnosti S_i' pro $i \in \{1, 2, 3, 5\}$. Výpočet rozložení bigramů v nich ukazuje, že po sobě jdoucí znaky jsou ve dvou třetinách až třech čtvrtinách případů stejné (viz Tabulka 3.17). Dá se předpokládat, že tuto vlastnost bude mít i posloupnost S_4'. Protože nejčastější bigramy ve dvojici sloupců 1; 2 čtvrtého impulsu rozepsaného do 26 sloupců jsou 01 a 10, první dva bity posloupnosti $\mathcal{K}_4$ budou různé.

S_1'			S_2'			S_3'			S_4'			S_5'		
00	1358	34%	**00**	1399	35%	**00**	1476	36%	**00**	1360	35%	**00**	1424	35%
01	565	14%	**01**	574	14%	**01**	581	15%	**01**	611	15%	**01**	594	15%
10	565	14%	**10**	574	14%	**10**	582	15%	**10**	610	15%	**10**	595	15%
11	1511	38%	**11**	1452	37%	**11**	1360	34%	**11**	1418	35%	**11**	1386	35%

Tab. 3.17: Četnosti bigramů a jejich percentuální zastoupení v posloupnostech S_i'

Postup britských kryptoanalytiků.

Zpráva [1] krátce uvádí, že pro určení sekvence $\mathcal{K}_1$ byl první impuls vepsán do tabulky o 41 sloupcích a byly počítány výskyty pětic znaků v každých pěti po sobě jdoucích sloupcích. Když byly spočítány četnosti takových *n*-gramů pro dvě různé pětice sloupců, byl hledán takový vektor délky 5, jehož přičtením k jedné z nich by se četnosti dostaly do co největší shody. Z takto získaných vektorů pak byla neuvedeným způsobem rekonstruována posloupnost $\mathcal{K}_1$.

Šlo tedy zřejmě o způsob principiálně podobný metodě použité v této práci. Zdroj [1] se nezmiňuje o tom, jak byl vyřešen problém nejednoznačnosti přičítaných vektorů, ke které dochází i v případě pětic.

Po prozkoumání vlastností posloupnosti S_1' pracovali Britové při analýze dalších impulsů pouze s trigramy.

3.8. ANALÝZA S_i'

Nyní je třeba zjistit, jakým způsobem vznikají sekvence S_i'. Již bylo zmíněno, že dvojice sousedních bitů v těchto posloupnostech jsou v cca. 70 % případů shodné (viz Tabulka 3.17).

Tabulka 3.18 obsahuje prvních dvacet bitů posloupností S_i' a v Baudotově kódu jim odpovídající znaky. Zajímavé je, že stejné znaky se často opakují dvakrát i vícekrát po sobě. Pohled na tuto sekvenci může inspirovat k hypotéze, že každá z posloupností S_i', $1 \le i \le 5$, je tvořena kolem, které se během šifrování pohybuje nepravidelně: v některých krocích se otočí o jednu pozici, v některých krocích zůstane stát. Tato hypotetická kola (a v souladu s dříve zavedenou notací i periodická rozšíření jejich vzorků) označíme S_i. Zmíněná opakování znaků rovněž svádí k domněnce, že se tato kola neotáčí nezávisle na sobě, ale všechna společně.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
S_1'	0	0	1	1	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	0
S_2'	1	1	1	1	0	0	0	0	1	0	0	0	0	0	1	1	1	1	0	1
S_3'	1	1	0	0	1	1	1	1	0	0	0	0	1	1	0	0	0	0	1	0
S_4'	0	0	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0	0	1	1
S_5'	1	1	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	1
S'	P	P	J	J	N	N	M	9	A	D	D	D	M	M	W	A	A	A	F	G

Tab. 3.18: Prvních 20 členů posloupností S_i' a příslušné znaky podle Baudotova kódu

Vyslovenou hypotézu o existenci kol S_i a jejich nepravidelném (zatím nikoli nezbytně společném) otáčení nejlépe potvrdíme tak, že úspěšně určíme velikosti kol S_i a jejich vzorky (což je náplní této podkapitoly) a rovněž odhalíme způsob řízení jejich pohybu (čímž se zabývá podkapitola 3.9).

Nejprve zavedeme následující často používaný pojem.

3.5 Definice. Buď $\mathcal{P}$ posloupnost znaků. *Během* v posloupnosti $\mathcal{P}$ nazveme každý její maximální úsek nenulové délky složený ze shodných znaků.

3.6 Příklad. Posloupnost 00111000010 se skládá z těchto pěti běhů: 00, 111, 0000, 1 a 0.

Podle hypotézy vznikla každá posloupnost S_i' z příslušné sekvence S_i opakováním některých znaků (když v daném kroku příslušné kolo zůstalo stát), jinými slovy prodloužením některých běhů o jeden či více znaků.

3.7 Poznámka. Pro jednoduchost zvolme očíslování kolíčků kol S_i tak, aby byl číslem 1 označen vždy počáteční kolíček toho běhu ve vzorku kola, v němž se nachází kolíček aktivní v prvním kroku šifrování. Je zřejmé, že díky takto zvolenému očíslování kolíčků je počet běhů ve vzorku každého kola sudý, protože běhy jsou tvořeny střídavě nulami a jedničkami a první běh je tvořen jinými znaky než poslední běh. Později v další podkapitole ale budeme muset určit počáteční pozice kol před vlastním šifrováním.

Problém nalezení počtu a nastavení kolíčků kol S_i rozdělíme na dvě dílčí úlohy:

- určení počtu běhů ve vzorcích těchto kol
- zjištění délky každého z těchto běhů

Řešení bude detailně předvedeno na příkladu kola S_1. Pro zbývající kola budou uvedeny pouze výsledky získané stejným postupem.

Odhlédneme od faktu, že otáčení kola S_1 (i ostatních kol) je deterministické a pouze zatím nejsou známa jeho pravidla, a budeme otočení kola považovat za náhodný jev. Prodloužení běhu ve vzorku kola vlivem jeho nepravidelného pohybu pak lze pravděpodobnostně popsat. Označme A_j náhodný jev, že se v j-tém kroku šifrování kolo otočí o jednu pozici. Učiňme následující předpoklady:

i. Jevy A_j mají stejné pravděpodobnostní rozdělení: v každém kroku šifrování je $P(A_j) = a$ a tato pravděpodobnost je stejná pro všechny kroky,
ii. jevy A_j jsou nezávislé.

Je třeba poznamenat, že předpoklad nezávislosti jevů zjevně nemůže být splněn. Nepravidelný pohyb kola je jistě řízen nějakým pseudonáhodným generátorem a výstup takového generátoru závisí na jeho předchozí produkci. Ze stejného důvodu se může i první předpoklad ukázat příliš silným. Přesto při dalším zkoumání budeme z těchto předpokladů vycházet s tím, že ve výsledcích budeme počítat s případnou chybou.

Uvažujme běh v posloupnosti S_1 délky l. Délku jemu odpovídajícího prodlouženého běhu v posloupnosti S_1' označíme l'. Rozdíl jejich délek, $l' - l$, je

náhodná veličina, která má za výše uvedených předpokladů Pascalovo rozdělení Pascal(l, a).

3.8 Definice. Diskrétní náhodná veličina $\boldsymbol{X}$ má *Pascalovo rozdělení* Pascal(r, p), $r \in \mathbb{N}$, $0 < p < 1$, jestliže její distribuční funkce je pro $k \in \mathbb{N}_0$

$$f(k; r, p) = \binom{r+k-1}{k-1} p^r (1-p)^k.$$

Pascalovo rozdělení Pascal(r, p) má náhodná veličina vyjadřující počet neúspěchů před dosažením r-tého úspěchu v sérii nezávislých Bernoulliho pokusů (tj. pokusů s dvěma možnými výsledky – úspěch a neúspěch) se stejnou pravděpodobností úspěchu p.

Rozptyl náhodné veličiny $\boldsymbol{X}$ s rozdělením Pascal(r, p) je

$$\operatorname{var} \boldsymbol{X} = r \frac{1-p}{p^2}.$$

Počet běhů ve vzorku kola $\boldsymbol{S_1}$ označíme n_0 a jejich délky l_i, $1 \le i \le n_0$. Pro $1 \le i \le n_0$ dále zavedeme tyto náhodné veličiny:

- $\boldsymbol{X_i}$, vyjadřující rozdíl $l_i' - l_i$, kde l_i' je délka i-tého běhu po prodloužení vlivem nepravidelného pohybu kola,
- $\boldsymbol{Y_i} = \boldsymbol{X_i} + l_i$ vyjadřující délku i-tého běhu po prodloužení vlivem nepravidelného pohybu kola.

Náhodné veličiny $\boldsymbol{X_i}$ mají, jak již bylo zmíněno, rozdělení Pascal(l_i, a) a platí

$$\operatorname{var} \boldsymbol{X_i} = l_i \frac{1-a}{a^2}.$$

Vzhledem k vlastnostem rozptylu náhodné veličiny platí rovněž

$$\operatorname{var} \boldsymbol{Y_i} = l_i \frac{1-a}{a^2}.$$

Hodnotu n_0 nyní určíme způsobem trochu připomínajícím určení délek period kol $\mathcal{K}_i$ pomocí průměrného indexu koincidence.

Nejprve vytvoříme sekvenci délek běhů v posloupnosti $\boldsymbol{S_1}'$. Z Tabulky 3.18 je patrné, že tato sekvence bude začínat čísly 2, 2, 4, 4, 2, 5, atd. Celkem je ve 4000 znacích posloupnosti $\boldsymbol{S_1}'$ 1129 běhů. Tuto posloupnost délek budeme postupně po řádcích vepisovat do tabulky o n sloupcích pro každý potenciální (sudý) počet běhů n, například $2 \le n \le 100$. Naším cílem je najít nějakou statistiku hodnot ve sloupcích této tabulky, kterou spolehlivě určíme správný počet běhů ve vzorku kola. Touto statistikou bude aritmetický průměr výběrových rozptylů hodnot v jednotlivých sloupcích.

Jestliže $n = n_0$, délky běhů v i-tém sloupci odpovídají různým prodloužením i-tého běhu ve vzorku kola při jeho otáčení, tedy výsledkům různých měření veličiny $\boldsymbol{Y_i}$. Rozptyl veličiny $\boldsymbol{Y_i}$ můžeme odhadnout pomocí výběrového rozptylu s_i^2 hodnot v i-tém sloupci tabulky, který vypočteme podle vztahu

$$s_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(x_j - \overline{x}_i)^2, \quad \overline{x}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_j,$$

kde x_j, $1 \le j \le N_i$, jsou všechny hodnoty v i-tém sloupci (do výpočtu nebudeme zahrnovat délky prvního a posledního běhu v posloupnosti $\boldsymbol{S_1'}$, protože tyto běhy mohou být neúplné).

Pro aritmetický průměr rozptylů veličin $\boldsymbol{Y_i}$, $1 \le i \le n_0$, který aproximujeme právě aritmetickým průměrem výběrových rozptylů s_i^2, $1 \le i \le n_0$, platí

$$\frac{1}{n_0}\sum_{i=1}^{n_0}\operatorname{var}\boldsymbol{Y_i} = \frac{1}{n_0}\sum_{i=1}^{n_0}\left(l_i\frac{1-a}{a^2}\right) = \frac{1-a}{a^2}\frac{1}{n_0}\sum_{i=1}^{n_0} l_i.$$

Pravděpodobnost a můžeme odhadnout zdola. Známe 4000 znaků posloupnosti $\boldsymbol{S'}$ (viz přiložené CD-ROM) a spočítáme hranice mezi běhy v této posloupnosti. Dojdeme k číslu 2357. V těchto místech se určitě otočilo alespoň jedno z kol $\boldsymbol{S_i}$ a budeme předpokládat, že to bylo kolo $\boldsymbol{S_1}$. (Už ostatně bylo zmíněno, že časté opakování znaků v posloupnosti $\boldsymbol{S}$ naznačuje, že se všechna kola otáčejí společně.)

Protože je samozřejmě možné, aby se některé z kol otočilo, a přitom se znak koly tvořený nezměnil, získáváme pro a odhad

$$a \ge \frac{2357}{3999} \doteq 0,63.$$

Jelikož $f(x) = (1-x)/x^2$ je na intervalu (0;1) klesající funkce, platí

$$\frac{1}{n_0}\sum_{i=1}^{n_0}\operatorname{var}\boldsymbol{Y_i} = \frac{1-a}{a^2}\frac{1}{n_0}\sum_{i=1}^{n_0} l_i \le \frac{1-0,63}{0,63^2}\frac{1}{n_0}\sum_{i=1}^{n_0} l_i.$$

Samozřejmě neznáme aritmetický průměr délek běhů. Můžeme zkusit dosadit průměrnou délku běhů ve vzorcích kol $\boldsymbol{\mathcal{K}_i}$ která je rovna téměř přesně 2 (viz Tabulka 3.16).

Budeme tedy očekávat, že pokud je počet sloupců n tabulky s délkami běhů roven skutečnému počtu běhů ve vzorku kola $\boldsymbol{S_1}$, bude platit

$$\frac{1}{n}\sum_{i=1}^{n} s_i^2 \le \frac{1-0,63}{0,63^2}\cdot 2 \doteq 1,86.$$

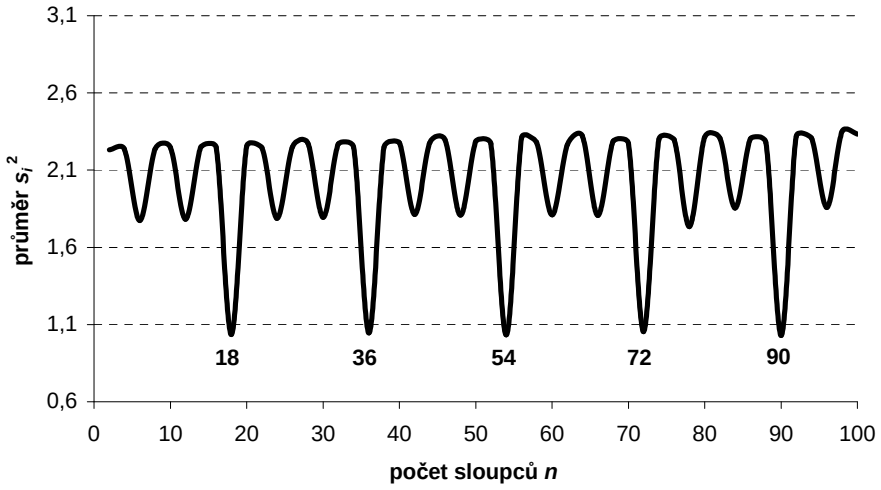
Pokud je n násobkem počtu běhů ve vzorku, lze očekávat přibližně stejnou hodnotu průměru výběrových rozptylů. Naopak pokud bude NSD(n, n_0) = 2, dá se předpokládat, že průměr výběrových rozptylů jednotlivých sloupců bude blízký výběrovému rozptylu celé posloupnosti délek běhů v sekvenci $\boldsymbol{S_1'}$ (bez délek prvního a posledního běhu, které mohou být neúplné). Tento rozptyl je přibližně 2,25.

Následující Tabulka 3.19 obsahuje získané aritmetické průměry výběrových rozptylů pro jednotlivé počty sloupců.

n	$\overline{s}^2$	n	$\overline{s}^2$	n	$\overline{s}^2$	n	$\overline{s}^2$
2	2,2323	28	2,2740	**54**	**1,0335**	78	1,7341
4	2,2362	30	1,7947	56	2,3095	80	2,3142
6	1,7723	32	2,2682	58	2,2785	82	2,3065
8	2,2371	34	2,2488	60	1,8113	84	1,8548
10	2,2453	**36**	**1,0454**	62	2,2638	86	2,3048
12	1,7812	38	2,2479	64	2,3196	88	2,2800
14	2,2498	40	2,2761	66	1,8071	**90**	**1,0294**
16	2,2467	42	1,8129	68	2,2879	92	2,3179
18	**1,0363**	44	2,2765	70	2,2706	94	2,3066
20	2,2541	46	2,2951	**72**	**1,0550**	96	1,8584
22	2,2437	48	1,8081	74	2,2998	98	2,3503
24	1,7877	50	2,2881	76	2,2947	100	2,3321
26	2,2507	52	2,2699				

Tab. 3.19: Průměry výběrových rozptylů pro jednotlivé počty sloupců *n*

Z tabulky a rovněž z následujícího Grafu 3.20, v němž jsou vyneseny tytéž hodnoty, je patrné, že ve vzorku kola S_1 je s největší pravděpodobností 18 běhů.



Graf 3.20: Průměr výběrových rozptylů v závislosti na počtu sloupců *n* pro posloupnost S_1'

Určení délek těchto 18 běhů je nyní již snadné. Podle výše vytvořeného pravděpodobnostního modelu je pravděpodobnost, že se běh délky l neprodlouží, roven a^l (což je pravděpodobnost otočení kola l-krát po sobě). 1129 délek běhů vypsaných do 18 sloupců tvoří necelých 63 řádků, tedy v i-tém sloupci můžeme očekávat přibližně $62\,a^{l_i}$ výskytů hodnoty l_i. Hodnoty výrazů a^l a $62\,a^l$ pro různá l s odhadem $a = 0{,}63$ jsou shrnuty v Tabulce 3.21. Výsledky ukazují, že pro běhy do délky 8 včetně je tento očekávaný počet větší než 1.

l	$0{,}63^l$	$62 \cdot 0{,}63^l$
1	0,63	39,06
2	0,3969	24,61
3	0,2500	15,50
4	0,1575	9,77
5	0,0992	6,15
6	0,0625	3,88
7	0,0394	2,44
8	0,0248	1,54
9	0,0156	0,97
10	0,0098	0,61

Tab. 3.21: Pravděpodobnosti výskytu neprodlouženého běhu pro různé délky běhu l a jejich očekávaný počet mezi 62 hodnotami

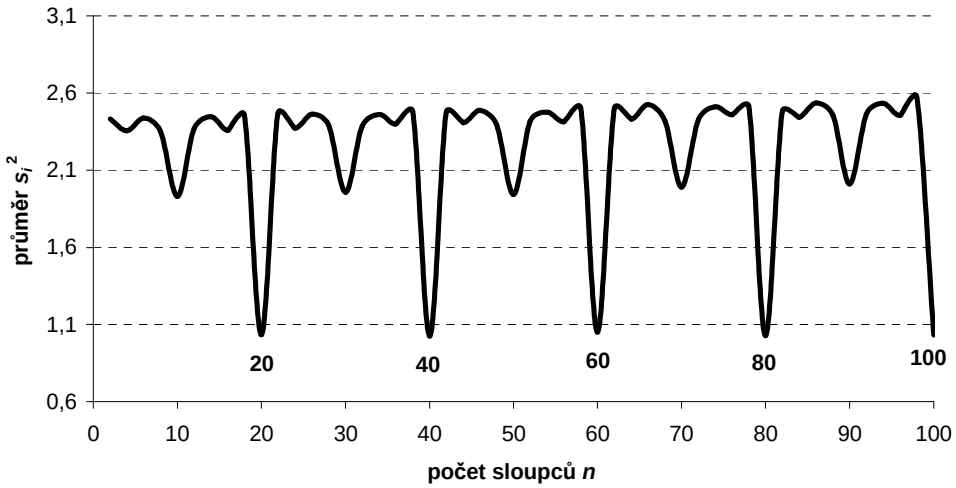
Délky všech 18 běhů ve vzorku kola S_1 tedy určíme prostým nalezením nejmenší hodnoty v každém ze sloupců tabulky s délkami běhů v posloupnosti S_1'. Prvních dvacet pět řádků této tabulky s 18 sloupci je uvedeno jako Tabulka 3.22.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
	2	2	4	4	2	5	6	2	3	3	2	2	5	3	4	5	4	6
	4	2	5	3	3	4	6	2	3	1	1	2	5	5	3	5	2	6
	4	3	4	6	3	3	4	2	3	1	1	4	5	4	2	5	4	7
	7	4	4	2	2	4	4	4	2	2	1	2	4	3	4	4	3	9
	3	4	4	2	2	6	5	2	3	1	1	2	4	7	4	5	2	6
	5	2	6	2	3	5	3	3	2	1	3	3	5	6	2	4	4	6
	7	2	4	2	4	5	5	3	3	3	1	2	5	4	2	7	3	6
	3	3	4	4	2	6	4	2	4	1	1	4	4	5	3	3	3	6
	4	4	6	3	2	3	4	4	2	1	3	3	4	3	3	3	3	7
	4	3	5	2	3	3	4	2	4	2	1	3	3	4	2	4	6	8
	4	4	3	3	2	3	5	2	4	1	1	3	4	5	4	4	5	4
	4	4	3	3	4	4	3	4	2	1	2	2	6	4	3	4	4	4
	4	4	4	4	2	4	3	4	4	2	1	3	4	4	2	4	4	4
	5	2	5	4	2	4	5	3	2	2	3	3	3	4	2	4	4	8
	3	4	5	2	3	4	6	3	2	1	2	2	4	3	3	4	4	6
	3	4	4	4	3	7	3	3	2	2	1	2	4	4	4	5	2	8
	4	4	6	2	2	5	4	2	3	2	1	4	3	5	4	4	3	8
	5	2	4	2	4	4	7	2	4	1	1	4	5	7	3	5	4	4
	4	2	6	3	2	5	5	2	3	2	1	4	6	3	4	5	2	7
	3	3	4	2	4	4	4	4	3	3	1	2	5	5	2	7	3	4
	4	2	4	3	3	7	5	4	2	1	3	2	7	5	2	4	2	5
	6	4	7	2	4	5	3	5	2	1	2	2	5	3	3	5	3	6
	5	4	3	4	2	6	4	2	2	2	1	2	5	6	2	6	3	6
	5	2	6	2	4	3	4	2	3	1	3	4	4	6	2	4	2	6
	7	3	3	4	2	4	7	2	4	2	2	2	3	4	3	3	6	5
MIN	**3**	**2**	**3**	**2**	**2**	**3**	**3**	**2**	**2**	**1**	**1**	**2**	**3**	**3**	**2**	**3**	**2**	**4**

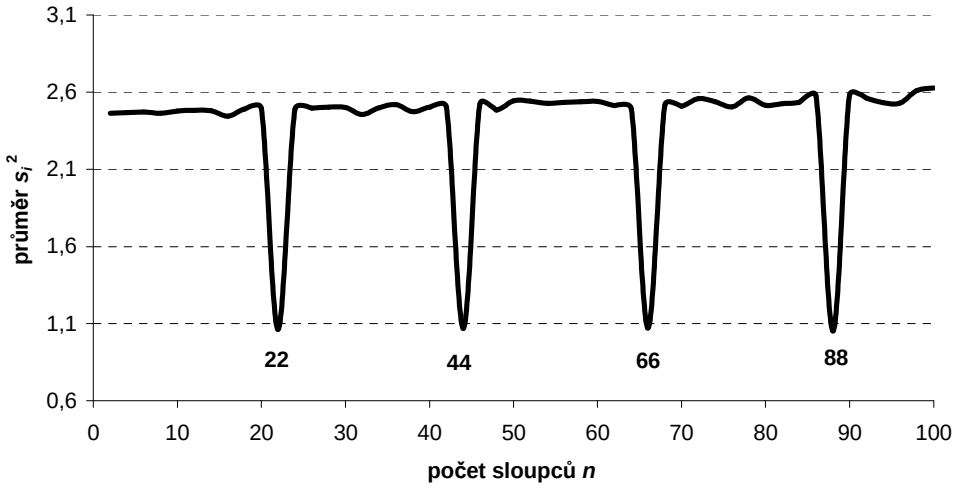
Tab. 3.22: Prvních 25 řádků tabulky délek běhů v S_1' a nejmenší hodnoty každého sloupce

Když sečteme délky běhů, zjistíme, že velikost kola vychází 43. To lze vzhledem ke zvyku volit velikosti kol prvočíselné chápat jako potvrzení správnosti výsledku.

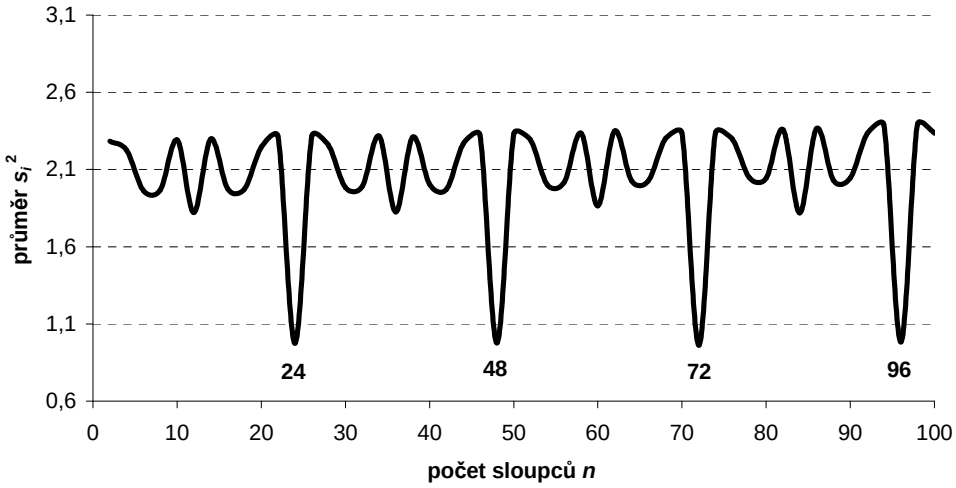
Stejným postupem lze zjistit velikosti a vzorky zbývajících kol. Následující Grafy 3.23 až 3.26 jsou obdobami Grafu 3.20 pro posloupnosti S_2', S_3', S_4' a S_5'.



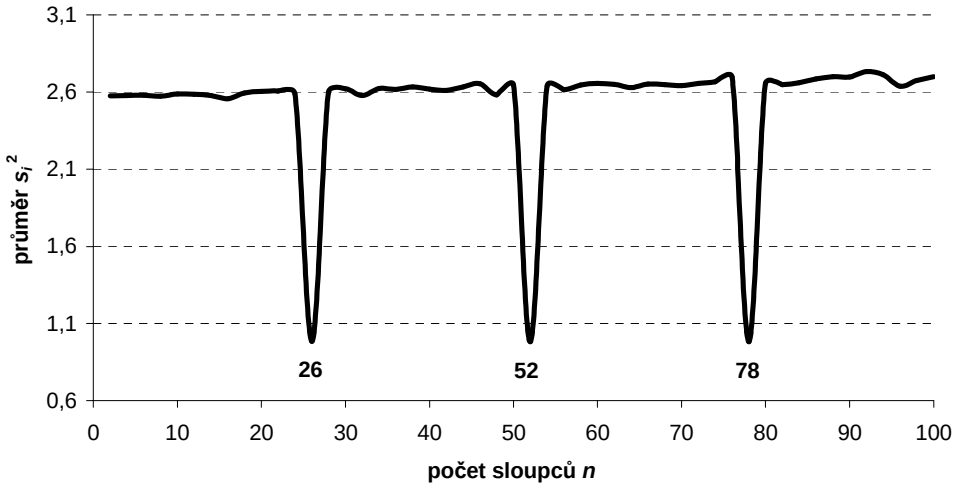
Graf 3.23: Průměr výběrových rozptylů v závislosti na počtu sloupců *n* pro posloupnost S_2'



Graf 3.24: Průměr výběrových rozptylů v závislosti na počtu sloupců *n* pro posloupnost S_3'



Graf 3.25: Průměr výběrových rozptylů v závislosti na počtu sloupců *n* pro posloupnost S_4'



Graf 3.26: Průměr výběrových rozptylů v závislosti na počtu sloupců *n* pro posloupnost S_5'

Z grafů jsou jasně patrné počty běhů ve vzorcích příslušných kol. Následující Tabulka 3.27 shrnuje velikosti, počty běhů a vzorky všech kol S_i, $1 \leq i \leq 5$. Bity, kterými je tvořen první běh každého vzorku, lze snadno určit ze známých posloupností S_i', tvoří totiž rovněž první běhy těchto posloupností.

kolo	vel.	běhů	vzorek
S_1	43	18	00011000110011100011001011000111001110011111
S_2	47	20	11100010001101001110011100110001111000111001100
S_3	51	22	111001110011001011100110001100011100010001111000100
S_4	53	24	11100001100011001100011010001100111001101100011001110
S_5	59	26	00001110111000110100111011000110011000110011100001101100011

Tab. 3.27: Velikosti, počty běhů a vzorky kol S_i

Velikost kola S_3 není prvočíselná, nicméně je, podobně jako v případě kola $\mathcal{K}_5$, součinem dvou malých prvočísel, a tedy nesoudělná s velikostmi ostatních kol.

Podle délek prvních běhů posloupností S_i' můžeme ještě určit přípustná počáteční nastavení jednotlivých kol, jsou-li jejich kolíčky očíslovány ve smyslu Poznámky 3.7. Počáteční nastavení je přípustné, jestliže první běh v posloupnosti S_i není delší než první běh v S_i'. Přípustná počáteční nastavení jednotlivých kol jsou obsahem Tabulky 3.28. Správné nastavení bude určeno, až bude znám způsob řízení pohybů těchto kol.

i	délka 1. běhu S_i'	přípustná počáteční nastavení kola
1	2	2, 3
2	4	1, 2, 3
3	2	2, 3
4	2	2, 3
5	2	3, 4

Tab. 3.28: Přípustná počáteční nastavení kol S_i

Postup britských kryptoanalytiků.

Zpráva [1] neobsahuje žádné podrobnější informace o tom, co pracovníky Bletchley Parku přivedlo na myšlenku, že posloupnosti S_i' jsou tvořeny nepravidelně se otáčejícími koly, ani jak jimi byly odhaleny jejich velikosti a vzorky. Pouze zmiňuje pozorování, že tyto posloupnosti jsou „zhruba periodické“, tzn. odvozené od periodických posloupností prodloužením některých běhů. Vzorky kol S_i pak dle stejného zdroje z posloupností S_i' „přímo plynou“, přičemž nejsou uvedeny žádné detaily o způsobu jejich odvození.

Je proto pravděpodobné, že k určení počtu běhů nebyl použit žádný matematický postup, ale řešení „tužkou a papírem“. Možným způsobem je například počítání běhů mezi dvojicemi sousedních běhů délky 1 v posloupnostech S_i', které se v těchto posloupnostech vyskytují vzácně a ve skutečnosti vždy odpovídají stejné dvojici sousedních jednoprvkových běhů v příslušných sekvencích S_i (neboť taková dvojice je ve vzorku každého kola jen jedna).

Vypsání posloupnosti S_i' po bězích do tabulky o správném počtu sloupců tak, aby běhy v každém sloupci odpovídaly prodloužením téhož běhu ve vzorku kola S_i, a následné sestavení tohoto vzorku z nejkratších běhů v každém sloupci (stejně jako v této podkapitole) se zdá být nejpřímočařejším způsobem, jak určit velikost kola a nastavení jeho kolíčků.

3.9. ODVOZENÍ ŘÍDICÍCH KOL

Nyní již zbývá jen určit pravidla, kterými se řídí pohyb kol S_i. Nejprve definujeme řídicí posloupnosti $\mathcal{R}_i$, pomocí nichž otáčení kol S_i popíšeme. Bez újmy na obecnosti budeme předpokládat, že k otáčení rotorů (které se mají otočit) dochází až na samém

konci každého šifrovacího kroku, tzn. poté, co se z nastavení aktivních kolíčků kol $\mathcal{K}_i$ a S_i spočítá znak klíče.

3.9 Definice. Pro každé $1 \le i \le 5$ definujme řídicí posloupnosti $\mathcal{R}_i = \{r_{i,j}\}_{j=1}^{3999}$ následujícím způsobem:

$r_{i,j} = 1$, pokud se na konci j-tého kroku šifrování kolo S_i otočilo o jednu pozici,

$r_{i,j} = 0$, pokud na konci j-tého kroku šifrování kolo S_i zůstalo stát.

3.10 Poznámka. Definice přiřazuje prvek řídicí posloupnosti $\mathcal{R}_i$ každé dvojici po sobě jdoucích členů posloupnosti S_i'. Je-li $r_{i,j} = 0$, znamená to, že j-tý a po něm následující znak posloupnosti S_i' odpovídají oba stejnému kolíčku kola S_i.

Posloupnosti $\mathcal{R}_i$ lze ze znalosti sekvencí S_i a S_i' rekonstruovat pouze částečně. Je zřejmé, že jsou-li j-tý a po něm následující znak S_i' různé, je $r_{i,j} = 1$. Rovněž pokud běhu v posloupnosti S_i odpovídá běh v S_i' stejné délky, kolo se jistě otočilo v každém kroku, tedy všechny příslušné členy řídicí posloupnosti jsou rovny jedné. Na druhou stranu, víme-li například, že běhu délky 5 v posloupnosti S_i' odpovídá běh délky 2 v S_i, můžeme odvodit, že v příslušných pěti šifrovacích krocích se kolo S_i otočilo právě dvakrát a podruhé se tak stalo v posledním z těchto kroků, ale určit, ve kterém kroku došlo k prvnímu otočení, možné není. Jinými slovy, pokud zmíněný běh délky 5 začíná k-tým znakem posloupnosti S_i', pak víme, že platí

$$\sum_{j=k}^{k+4} r_{i,j} = 2,$$

a rovněž víme, že $r_{i,k+4} = 1$, ale které z dalších $r_{i,j}$, $k \le j \le k + 3$, je rovno jedné, určit nelze.

Nulové členy řídicích posloupností je možné pevně určit jedině v případě, že odpovídají prodloužení běhu délky 1.

Takto částečně odvozené řídicí posloupnosti tedy obsahují izolované běhy jedniček a běhy nul ohraničené z obou stran jedničkami. Mezi těmito skupinami jsou intervaly, v nichž je známý počet jedniček (tzn. součet prvků), ale nikoliv jejich rozmístění.

Následující Tabulka 3.29 obsahuje prvních 33 členů posloupností S_i', příslušné běhy sekvencí S_i a odvozené členy řídicích posloupností $\mathcal{R}_i$ (toto číslo bylo zvoleno s ohledem na hranice běhů v S_i'). Hodnoty ve sloučených buňkách jsou rovny součtům prvků řídicích posloupností na příslušných pozicích. Protože neznáme počáteční pozice kol S_i, nemůžeme první členy řídicích posloupností vůbec určit, protože nevíme, jak dlouhé části prvního běhu ve vzorku kola S_i odpovídá první běh v S_i'. Z tohoto důvodu začíná každá řídicí posloupnost otazníkem.

S_1'	S_1	$\mathcal{R}_1$	S_2'	S_2	$\mathcal{R}_2$	S_3'	S_3	$\mathcal{R}_3$	S_4'	S_4	$\mathcal{R}_4$	S_5'	S_5	$\mathcal{R}_5$
0	0	?	1	1		1	1	?	0	0	?	1	1	?
0	0	1	1	1	?	1	1	1	0	0	1	1	1	1
1	1	1	1	1		0	0	1	1	1		0	0	
1	1	1	1		1	0	0	1	1	1	3	0	0	2
0	0		0	0		1	1		1	1		0	0	
0	0	2	0	0	2	1	1	2	1	1		0		1
0	0		0	0		1	1		1		1	1	1	1
0		1	0		1	1		1	0	0	1	0	0	
1	1		1	1	1	0	0		0	0	1	0	0	2
1	1	1	0	0		0	0	1	1	1		0	0	
1			0	0	2	0			1	1	2	0		
1		1	0	0		0		1	1	1		0		1
0	0	1	0			1	1	1	1			1	1	1
0	0	1	0		1	1	1	1	1		1	1	1	1
1	1		1	1		0	0		0	0		1	1	1
1	1	2	1	1	1	0	0	1	0	0	1	0	0	
1	1		1			0			0			0	0	1
1			1		1	0		1	0		1	0		
1		1	0	0	1	1	1	1	1	1	1	0		1
0	0		1	1	1	0	0	1	1	1	1	1	1	1
0	0		0	0		1	1		0	0		0	0	0
0	0	2	0	0	1	1	1		0	0	1	0		1
0			0			1	1	2	0			1	1	
0			0			1			0			1	1	1
0		1	0		1	1			0		1	1		
1	1	1	1	1	1	1		1	1	1	1	1		1
1	1	1	1	1	1	0	0	1	1	1	1	0	0	
0	0	1	1	1	1	0	0	1	1	1	1	0	0	2
0	0		0	0		1	1		0	0		0	0	
0		1	0	0	1	1	1	1	0	0	1	0		1
1	1	0	0			1			0			1	1	0
1		0	0			1			0			1		0
1		1	0		1	1		1	0		1	1		1

Tab. 3.29: Odvozené řídicí posloupnosti $\mathcal{R}_i$ pro každou dvojici sekvencí S_i', S_i

Nyní je možné ověřit hypotézu, že kola S_i se otáčejí všechna společně, která byla vyslovena v úvodu předchozí podkapitoly na základě údajů v Tabulce 3.18. Dalším argumentem ve prospěch této domněnky je fakt, že indikátorová skupina má pouze 12 znaků. Již bylo zmíněno, že první znak indikátoru ovlivňuje pouze první impuls klíče (pracovníci Bletchley Parku tento poznatek získali luštěním zpráv, jejichž indikátory se shodovaly až na první znak). Rovněž bylo uvedeno, že na poslední pozici indikátorové skupiny se v daném období vyskytovalo pouze 23 různých znaků, lze se tedy domnívat, že tato pozice udávala nastavení kola $\mathcal{K}_5$ s periodou 23. Je tedy pravděpodobné, že šifrovací stroj má 12 rotorů a každý znak indikátoru udává nastavení jednoho z nich. K řízení pohybu všech kol S_i tím pádem zbývají jen

dva rotory. Pokud by se měla kola S_i otáčet nezávisle na sobě, bylo by patrně potřeba alespoň pět řídicích kol.

Porovnáme-li odvozené řídicí posloupnosti jednotlivých kol S_i, zjistíme, že mezi nimi nejsou žádné rozpory, skutečně tedy mohou být částečnými popisy společné řídicí posloupnosti $\mathcal{R}$. Prvních 33 členů takto získané posloupnosti je uvedeno v Tabulce 3.30.

$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$	$\mathcal{R}_5$	$\mathcal{R}$
?	?	?	?	?	**?**
1		1	1	1	**1**
1		1	3	2	**1**
1	1	1			**1**
2	2	2			**0**
				1	**1**
			1	1	**1**
1	1	1	1	2	**1**
1	1	1	1		**1**
	2		2		**0**
					0
1		1		1	**1**
1		1		1	**1**
1	1	1	1	1	**1**
2	1	1	1	1	**1**
				1	**0**
					0
	1	1	1		**1**
1	1	1	1	1	**1**
2	1	2	1	0	**0**
				1	**1**
				1	**0**
					0
1	1		1		**1**
1	1	1	1	1	**1**
1	1	1	1	2	**1**
	1	1	1		**1**
					0
1				1	**1**
0	1	1	1	0	**0**
0				0	**0**
1	1	1	1	1	**1**

Tab. 3.30: Řídicí sekvence $\mathcal{R}_i$ a společná řídicí posloupnost $\mathcal{R}$ vytvořená jejich kombinací

I posloupnost $\mathcal{R}$, jejíchž všech 3999 členů je uvedeno v přílohách (CD-ROM), obsahuje některé neznámé úseky, ale v mnohem menším počtu. Nejdelší souvislý úsek známých hodnot má délku 200 bitů. Je proto možné ji podrobit běžné analýze.

Provedeme-li na posloupnosti Kasiského test s tím, že se omezíme pouze na známé členy sekvence, nalezneme řadu shodných úseků maximální možné délky (vzhledem k hranicím známých částí posloupnosti), jejichž vzdálenost je 2257. To napovídá, že posloupnost $\mathcal{R}$ je periodická s touto periodou. Skutečně, předpoklad

$$\boldsymbol{r}_j = \boldsymbol{r}_{j+2257}, \text{ kde } \mathcal{R} = \left\{\boldsymbol{r}_j\right\}_{j=1}^{3999},$$

nevede k žádným sporům a navíc umožňuje přesně určit další neznámé prvky řídící posloupnosti. Po doplnění má nejdelší souvislý známý úsek posloupnosti $\mathcal{R}$ délku 658 členů a začíná prvkem $\boldsymbol{r}_{674}$.

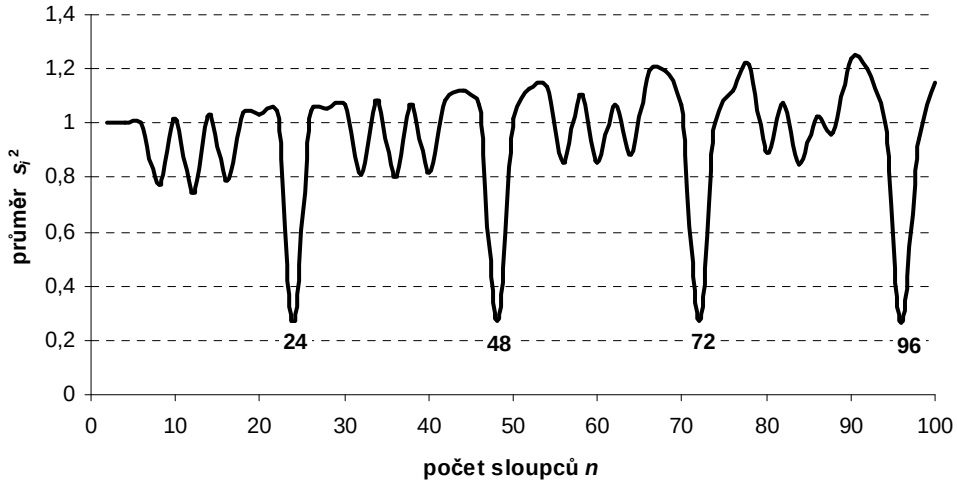
Pro další postup je důležitý poznatek, že $2257 = 37 \cdot 61$. Tato prvočísla jsou velmi věrohodnými kandidáty na velikosti kol šifrovacího stroje: seřadíme-li periody dosud známých kol podle velikosti, zjistíme, že spolu s čísly 37 a 61 tvoří souvislou řadu prvočísel (samozřejmě s výjimkou čísel 26 a 51). Lze se tedy domnívat, že posloupnost $\mathcal{R}$ je nějakým způsobem tvořena součinností kol právě těchto velikostí, a proto budeme tato kola rovněž nazývat řídicími koly.

Jednou z možností je, že se tato kola otáčejí v každém kroku a řídicí posloupnost je tvořena součtem nastavení jejich aktivních kolíčků. Taková sekvence by měla vzhledem k nesoudělnosti čísel 37 a 61 periodu 2257. Rozklad posloupnosti na součet dvou periodických posloupností, jejichž periody p_1 a p_2 známe, lze řešit jako soustavu lineárních rovnic o $p_1 + p_2$ neznámých. To je pro periody 37 a 61 s pouhou tužkou a papírem, které měli k dispozici britští kryptoanalytici, sice pracné, ale snadné, a periodické posloupnosti tak lze určit jednoznačně až na negaci jejich členů. Jak se však ukáže, tato cesta by ve skutečnosti byla slepou uličkou, což by se projevilo neřešitelností soustavy rovnic.

Další možností, jak by mohla být řídicí posloupnost $\mathcal{R}$ tvořena dvěma rotory tak, aby měla periodu rovnou součinu jejich velikostí, je podobná otáčení samotných kol $\mathcal{S}_i$. Jedno z kol se otáčí v každém šifrovacím kroku, zatímco druhé, které tvoří samotnou posloupnost $\mathcal{R}$, se otočí jen tehdy, je-li aktivní kolíček prvního kola nastaven na 1. Jinými slovy, periodická posloupnost tvořená pravidelným otáčením prvního z kol je řídicí posloupností pohybu druhého rotoru.

V takovém případě by měla být posloupnost $\mathcal{R}$, řečeno slovy pracovníků Bletchley Parku, „hrubě periodická“, tzn. délky běhů v posloupnosti R vzdálené o počet běhů ve vzorku nepravidelně se pohybujícího řídicího kola by měly mít malý rozptyl. Můžeme se pokusit určit počet běhů v běhu tohoto kola stejně, jako jsme postupovali v případě rotorů $\mathcal{S}_i$. Je třeba se však omezit na nějaký souvislý úsek známých členů posloupnosti $\mathcal{R}$, protože rozmístění hodnot v neznámých intervalech ovlivňuje jak délky, tak počty běhů.

V následujícím Grafu 3.31 jsou vyneseny průměrné výběrové rozptyly délek běhů ve zmíněném nejdelším souvislém úseku řídicí posloupnosti $\mathcal{R}$ v závislosti na předpokládaném počtu běhů nepravidelně se pohybujícího řídicího kola. Hodnoty byly získány stejným způsobem, jako pro obdobné grafy v předchozí podkapitole.



Graf 3.31: Průměr výběrových rozptylů v závislosti na počtu sloupců *n* pro nejdelší souvislý úsek posloupnosti $\mathcal{R}$

Z grafu vyplývá, že řídicí posloupnost $\mathcal{R}$ patrně skutečně vznikla rozšířením periodické posloupnosti s 24 běhy. To lze považovat za důkaz zkoumané hypotézy. To z řídicích kol, které se otáčí v každém šifrovacím kroku, budeme nadále označovat $\mathcal{M}_1$. Nepravidelně se otáčející kolo budeme značit $\mathcal{M}_2$. Prodloužením některých běhů posloupnosti $\mathcal{M}_2$ pak vzniká posloupnost $\mathcal{R}$.

Délky běhů ve vzorku kola $\mathcal{M}_2$ můžeme opět určit jako minima délek příslušných prodloužených běhů. Tímto způsobem získáme posloupnost

1 1 1 1 0 1 1 0 1 0 1 1 0 1 0 1 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1 1 1 0,

jejíž délka je 42. Očekávaná délka posloupnosti však je 37, nebo 61. Vzhledem k tomu, že zkoumaný úsek posloupnosti $\mathcal{R}$ je krátký (má 285 běhů), je třeba počítat s tím, že některé delší běhy se v tomto úseku neprodloužené nemusejí vyskytovat. Získaná posloupnost obsahuje pět běhů délky 4, ale žádný běh délky 3. Můžeme zkusit zkrátit všechny běhy délky 4 o jeden znak, čímž získáme následující posloupnost požadované délky 37:

1 1 1 0 1 1 0 1 0 1 1 0 1 0 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1 0 1 1 0 1 1 1 0.

Další postup je již zcela analogický odvozování řídicích posloupností kol S_i v první části této podkapitoly. Budeme opět pracovat na tomtéž dlouhém souvislém úseku známých prvků posloupnosti $\mathcal{R}$. Protože nyní známe délku každého běhu před jeho prodloužením, můžeme částečně odvodit posloupnost $\mathcal{M}_1$ s tím, že tato posloupnost bude opět obsahovat intervaly, v nichž budeme znát počty výskytů jednotlivých hodnot, avšak nikoliv jejich přesné umístění.

Posloupnost $\mathcal{M}_1$ má podle předpokladů periodu 61. Vzhledem k délce zkoumaného úseku, která činí 658 znaků, je k dispozici necelých 11 period, tedy deset alternativních popisů první periody. Ukazuje se, že nejen že mezi těmito popisy nejsou žádné kontradikce, ale navíc z nich lze jednoznačně rekonstruovat celý vzorek kola $\mathcal{M}_1$, jak ukazuje následující Tabulka 3.32.

per. 1	per. 2	per. 3	per. 4	per. 5	per. 6	per. 7	per. 8	per. 9	per. 10	per. 11	$\mathcal{M}_1$
?	1	1	0	2	2	2	0	2	2	1	**0**
	1		1				1				**1**
1	1	1	1	1	1	1	1			1	**1**
1	1	1	1	1	1	1	1	1	1	1	**1**
1	0	0	2	2	2	0	0	0	0	0	**0**
	1	1				1	1	1	1	1	**1**
1	1	1				1	2	2	1	1	**1**
0		0				1				1	**0**
1	1	1	1	1	1				1		**1**
1	1	1	1	1	1	1	1	1	1	1	**1**
0	0	2	2	2	0	0	0	0	0	0	**0**
1	1				1	1	1	1	1	1	**1**
1	1				1	2	2	1	1	1	**1**
	0				1				1	0	**0**
1	1	1	1	1				1		1	**1**
1	1	1	1	1	1	1	1	1	1	1	**1**
0	2	2	2	0	0	0	0	0	0	0	**0**
1				1	1	1	1	1	1	1	**1**
1				1	2	2	1	1	1	2	**1**
1	1	1	1	1			1		1		**1**
0	0	0	0				0		0		**0**
1	1	1	1	1	1	1	1	1	1	1	**1**
2	2	2	0	0	0	0	0	0	0	0	**0**
			1	1	1	1	1	1	1	1	**1**
			0	2	2	1	0	0	2	2	**0**
			1				1	1			**1**
1	1	1	1			1	1	1			**1**
1	1	1	1	1	1	1	1	1	1	1	**1**
2	2	0	0	0	0	0	0	0	0	0	**0**
		1	1	1	1	1	1	1	1	1	**1**
		1	2	2	1	1	1	2	2	2	**1**
1	1	1			1	1	1				**1**
0	0				0		0				**0**
1	1	1	1	1	1	1	1	1	1	1	**1**
2	0	0	0	0	0	0	0	0	0	0	**0**
	1	1	1	1	1	1	1	1	1	1	**1**
	1	2	2	1	1	1	2	2	2	1	**1**
	1				1	0				0	**0**
1				1		1				1	**1**
1	1	1	1	1	1	1	1	1	1	1	**1**
0	0	0	0	0	0	0	0	0	0		**0**
1	1	1	1	1	1	1	1	1	1	1	**1**
0	2	2	1	0	0	2	2	2	0	0	**0**
1				1	1				1	1	**1**
1				1	0				0	2	**0**
			1		1				1		**1**
1	1	1	1	1	1	1	1	1	1		**1**
1	1	1	1	1	1	1	1	1	1	1	**1**
2	2	1	1	1	2	2	2	1	1		**1**
			1	0				0	2		**0**
		1		1				1			**1**
		0		0				1			**0**
1	1	1	1	1	1	1	1				**1**
0	0	0	0	0	0	0	0				**0**
1	1	1	1	1	1	1	1	1	1		**1**
2	1	1	1	2	2	2	1	1	1		**1**
	1	1	1				1	2	2		**1**
	0		0								**0**
1	1	1	1	1	1	1					**1**
1	1	1	1	1	1	1	1	1	1		**1**
1	1	1	2	2	2	1	1	1	1		**1**

Tab. 3.32: Odvození první periody posloupnosti $\mathcal{M}_1$ z částečných popisů necelých 11 jejích period

Tento výsledek potvrzuje, že domněnka o způsobu tvorby posloupnosti $\mathcal{R}$ i odhad podoby vzorku kola $\mathcal{M}_2$ jsou správné. Periodickým rozšířením posloupnosti $\mathcal{M}_1$ oběma směry po celé délce řídicí sekvence $\mathcal{R}$ a simulací odpovídajícího nepravidelného pohybu kola $\mathcal{M}_2$ je možné určit jejich nastavení v prvním kroku šifrování, stejně jako zatím neznámé počáteční nastavení rotorů S_i. Výsledky, spolu s dalšími údaji o všech rotorech, jsou obsaženy v Tabulce 3.33.

Kolo	Vel.	Počáteč. nast.	Vzorek
$\mathcal{K}_1$	41	1	01100111000011100111000010011011000110110
$\mathcal{K}_2$	31	1	00011001100010111110111000011011
$\mathcal{K}_3$	29	1	01111011000111000110001100100
$\mathcal{K}_4$	26	1	01100010110011100110101001
$\mathcal{K}_5$	23	1	01011110001011100001101
S_1	43	3	0001100011001110001100101100011100111001111
S_2	47	1	11100010001101001110011100110001111000111001100
S_3	51	3	111001110011001011100110001100011100010001111000100
S_4	53	3	11100001100011001100011010001100111001101100011001110
S_5	59	4	00001110111000110100111011000110011000110011100001101100011
$\mathcal{M}_1$	61	60	01110110110110110111010101110111010110110101011110101011101111
$\mathcal{M}_2$	37	30	11101101011010101110111011101011011010

Tab. 3.33: Velikosti, počáteční nastavení a vzorky všech kol

Postup britských kryptoanalytiků.

Pracovníci Bletchley Parku postupovali podle zprávy [1] podobným způsobem, jako je výše popsaný. Definovali a částečně odvodili řídicí posloupnosti jednotlivých kol S_i a brzy si všimli, že jde o různé popisy téže posloupnosti (což předpokládali na základě zmíněného argumentu o omezeném počtu řídicích kol). Způsob generování posloupnosti $\mathcal{R}$ jim však údajně dlouho unikal. O periodičnosti této posloupnosti s periodou 2257, která se zdá být významnou indicií, se zdroj [1] v souvislosti s odvozováním konstrukce šifrovacího stroje nezmiňuje. Jakmile si kryptoanalytici všimli, že $\mathcal{R}$ je podobně jako sekvence S_i' „zhruba periodická“, „snadno“ odvodili velikosti a vzorky řídicích kol opět blíže nepopsaným způsobem.

Nyní je odhalena celá konstrukce šifrovacího stroje. V další fázi rozbíjení systému je třeba určit, jak často se mění jednotlivé části klíče, tzn. vzorky kol, jejich počáteční nastavení a převodní tabulky písmen indikátoru a počátečních nastavení rotorů, zda se mění pořadí rotorů atd. Britští kryptoanalytici tyto údaje zjistili luštěním depeší z období zachycení zpráv ZMUG.

3.10. PRAVIDLO *ab*=½

Úspěšná analýza klíče popsanými metodami byla proveditelná především díky velmi nerovnoměrnému rozdělení bigramů v posloupnostech $\boldsymbol{S}_\mathbf{i}'$. Na této jejich vlastnosti záviselo jak určení délek period sekvencí $\mathcal{K}_\mathbf{i}$ pomocí indexu koincidence, tak následné nalezení těchto posloupností samotných. Tuto zranitelnost si uvědomili i Němci a zavedli proto (podle [1] nejpozději v březnu 1942) pravidlo pro používaná nastavení kolíčků rotorů $\boldsymbol{M}_\mathbf{i}$ a $\boldsymbol{S}_\mathbf{i}$, které zajišťovalo rovnoměrné rozdělení bigramů v posloupnostech $\boldsymbol{S}_\mathbf{i}'$. Nejprve však definice důležitého pojmu.

3.11 Definice. Buď $\mathcal{P} = p_1p_2p_3\ldots p_n$ binární posloupnost. Pak *diferencí* $\mathcal{P}$ neboli $\Delta\mathcal{P}$ nazveme binární posloupnost $q_1q_2q_3\ldots q_{n-1}$ definovanou vztahem

$$q_i = p_i + p_{i+1}, 1 \le i < n.$$

Například $\Delta\boldsymbol{S}_\mathbf{i}'$ zkoumaného klíče zpráv ZMUG obsahovaly významnou většinu nul, protože po sobě jdoucí znaky byly většinou stejné. Účelem později zavedeného pravidla bylo vyrovnat počty nul a jedniček v posloupnostech $\Delta\boldsymbol{S}_\mathbf{i}'$.

3.12 Tvrzení. Označme *a* podíl jedniček v řídicí posloupnosti $\mathcal{R}$ definované v předchozí podkapitole. Dále označme, pro nějaké zvolené *i*, *b* podíl jedniček v posloupnosti $\Delta\boldsymbol{S}_\mathbf{i}$. Potom podíl jedniček v posloupnosti $\Delta\boldsymbol{S}_\mathbf{i}'$ je roven *ab*.

Důkaz. Hodnotu *a* lze chápat jako pravděpodobnost, že se kola $\boldsymbol{S}_\mathbf{i}$ při přechodu stroje Lorenz SZ40 z jednoho stavu do následujícího pohnou. Hodnotu *b* pak jako pravděpodobnost, že aktivní kolíček kola $\boldsymbol{S}_\mathbf{i}$ je nastaven na opačnou hodnotu, než kolíček následující. Jednička v posloupnosti $\Delta\boldsymbol{S}_\mathbf{i}'$ se objeví právě tehdy, když se při přechodu mezi stavy kolo $\boldsymbol{S}_\mathbf{i}$ pohne a nový aktivní kolíček má opačné nastavení, než předchozí. Budeme-li tyto dva jevy považovat za nezávislé, je pravděpodobnost jejich společného výskytu rovna *ab*.

Německá strana nastavovala kolíčky kol přístroje Lorenz SZ tak, aby pro každé kolo $\boldsymbol{S}_\mathbf{i}$ platilo $ab = ½$. Konkrétní podmínka nastavení kolíčků na kole $\boldsymbol{S}_\mathbf{i}$ se počítala následujícím způsobem:

Označme *d* počet nul v nastavení řídicího kola $\boldsymbol{M}_\mathbf{2}$. Podíl jedniček v periodickém rozšíření vzorku tohoto kola je tedy

$$a = 1 - \frac{d}{37} = \frac{37-d}{37}.$$

Stejný podíl jedniček je vzhledem k velikostem kol $\boldsymbol{M}_\mathbf{1}$ a $\boldsymbol{M}_\mathbf{2}$ i v každém intervalu řídicí posloupnosti $\mathcal{R}$ délky 2257.. Hodnota *b* se vypočte podle vztahu:

$$b = \frac{1}{2a} = \frac{1}{2 \cdot \frac{37-d}{37}} = \frac{37}{74-2d}.$$

Buď p_i velikost kola S_i. Počet jedniček v každém ΔS_i tedy musí být roven

$$z_i = bp_i = \frac{37\,p_i}{74-2d}.$$

Vytvořme pro ilustraci nastavení kola S_1, které splňuje podmínku $ab = ½$, ostatní kola ponechme v nastavení stejném jako u zpráv ZMUG a podrobme první impuls získaného klíče analýze dle předchozích částí této kapitoly.

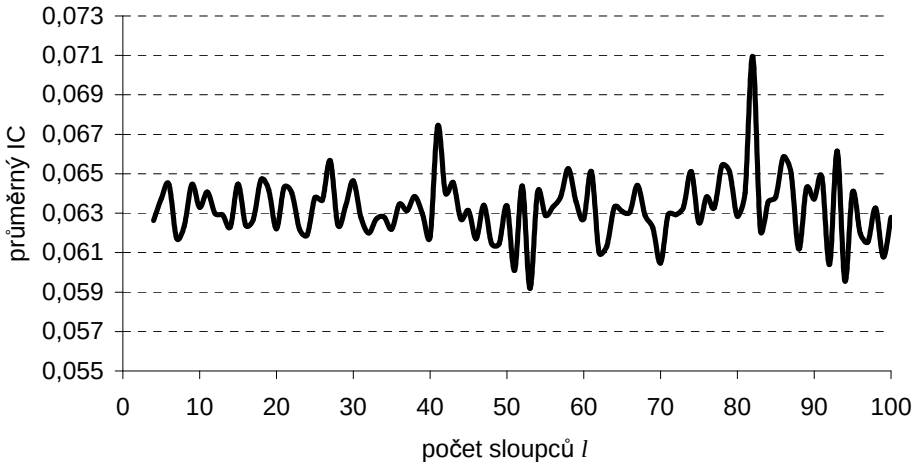
U zpráv ZMUG je $d = 12$, proto musí být $z_1 = 31{,}82 \doteq 32$. To splňuje například takovéto nastavení:

S_1	0 1 1 0 1 1 0 1 0 1 0 1 0 0 1 0 0 1 1 0 1 0 0 1 1 0 1 0 0 1 0 1 0 0 1 1 0 1 0 1 0 1 0
ΔS_1	1 0 1 1 0 1 1 1 1 1 1 1 0 1 1 0 1 0 1 1 1 0 1 0 1 1 1 0 1 1 1 1 0 1 0 1 1 1 1 1 1 1 0

Následující Graf 3.34 závislosti průměrného indexu koincidence na počtu sloupců, vytvořený stejným způsobem jako grafy v podkapitole 3.6, ukazuje, že při použití pravidla $ab = ½$ je metoda hledání periody posloupnosti $\mathcal{K}_1$ popsaná ve zmíněné podkapitole nepoužitelná. Tím je efektivně znemožněna i jakákoli následná analýza.

Graf 3.34: Průměrný index koincidence v závislosti na počtu sloupců *l* pro 1. impuls klíče dodržujícího pravidlo *ab* = ½

O něco lepší výsledky lze dosáhnout počítáním indexu koincidence *n*-gramů pro $n > 2$. Graf 3.35 ukazuje výsledky získané počítáním indexu koincidence pro tetragramy.



Graf 3.35: Průměrný index koincidence tetragramů v závislosti na počtu sloupců *l* pro 1. impuls klíče dodržujícího pravidlo *ab* = ½

Průměrný index koincidence je zde nejvyšší pro správné hodnoty 41 a 82, ale rozptyl jeho hodnot je velký a hypotézy o délce periody by byly méně věrohodné. Nejednoznačnosti by vyvstaly i při pokusech zrekonstruovat sekvenci $\mathcal{K}_1$.

Sami britští kryptoanalytici v [1] přiznávají, že by šifru Lorenz SZ40 pravděpodobně nikdy neprolomili, nemít k analýze dostatečně dlouhou sekvenci klíče, která pravidlo *ab* = ½ nesplňovala. Souhra dvou velkých nedůsledností ze strany Němců v již prvních měsících zkušebního provozu stroje Lorenz SZ40 tedy dokázala kompromitovat celý šifrový systém.

ZÁVĚR

Rozbití systému Lorenz bylo možné především díky vážným prohřeškům proti kryptografickým pravidlům na německé straně. Nedostatečná opatření proti posílání zpráv šifrovaných stejným klíčem při používání Vernamovy šifry a užívání slabých klíčů (nesplňujících pravidlo $ab=½$) ve svém důsledku vedly k prozrazení šifrovacího stroje.

Samotný algoritmus vytváření pseudonáhodné posloupnosti se rovněž ukázal jako chatrný a jeho bezpečnost do velké míry závisela na jeho utajení. Společný nepravidelný pohyb kol S_i, která se otáčela pouze v některých krocích, měl zkomplikovat luštění šifry, ale ukázal se být naopak její největší slabinou, kterou neodstranilo ani zavedení omezení ve verzích SZ42A a SZ42B. Díky tomuto konstrukčnímu řešení bylo možné najít správná počáteční nastavení kol $\mathcal{K}_i$ a dokonce i jejich vzorky pouze ze znalosti šifrového textu. Princip využití této slabiny je stručně vysvětlen v následujícím odstavci.

Z rovnice

$$\mathbf{C} = \mathbf{P} + \mathcal{K} + S'$$

popisující šifrování, kde **C** je šifrový text, **P** otevřený text a $\mathcal{K}$ a S' posloupnosti znaků tvořených po řadě koly $\mathcal{K}_i$ a S_i vyplývá

$$\mathbf{C} + \mathcal{K} = \mathbf{P} + S'.$$

Je zřejmé, že platí rovněž

$$\Delta\mathbf{C} + \Delta\mathcal{K} = \Delta\mathbf{P} + \Delta S',$$

kde Δ značí diferenci zavedenou v Definici 3.11. Z nerovnoměrnosti rozdělení bigramů v otevřeném textu vyplývá nerovnoměrnost rozdělení znaků v posloupnosti $\Delta\mathbf{P}$. Nepravidelný pohyb kol S_i způsobuje velký podíl bigramů tvořených stejnými znaky v posloupnosti S', což je příčinou stejně velkého podílu znaku / (s Baudotovou reprezentací (0, 0, 0, 0, 0)) v posloupnosti $\Delta S'$ (přestože počty nul a jedniček jsou v každém impulsu posloupnosti $\Delta S'$ vyrovnané díky pravidlu $ab = ½$). Protože symbol / je při zavedeném sčítání znaků neutrální prvek, rozdělení znaků v součtu posloupností $\Delta\mathbf{P} + \Delta S'$ je rovněž nerovnoměrné a koreluje s rozdělením znaků v diferenci otevřeného textu. Při hledání počátečních nastavení kol $\mathcal{K}$ pak útočník může počítat $\Delta\mathbf{C} + \Delta\mathcal{K}$ pro všechna nastavení kol a najít díky popsané vlastnosti to správné. Použití pravidla $ab = ½$ efektivně brání provedení popsaného útoku na jediném impulsu, ale důležitým poznatkem britských kryptoanalytiků je fakt, že není nutné ani hledat nastavení všech kol $\mathcal{K}$ současně, stačí pracovat s jejich dvojicemi. Například možných nastavení kol $\mathcal{K}_1$ a $\mathcal{K}_2$ je pouze $41 \cdot 31 = 1271$, útok tedy byl s pomocí výpočetních strojů proveditelný.

Dalšími příklady počínání, které výrazně usnadňovalo Bletchley Parku práci, byly používání písmenných indikátorů (z nichž bylo možné odvodit řadu dodatečných informací, včetně vzorků kol), stereotypní začátky zpráv a ustálené výrazy

(umožňující útoky se znalostí otevřeného textu) a především hrubé podcenění analytických a výpočetních schopností a možností protivníka.

Prolomení šifry Lorenz je dalším z mnoha příkladů v dějinách utajované komunikace, který potvrzuje, že nedostatky v návrhu, spoléhání se na utajení algoritmu, nedostatek disciplíny při používání a podcenění nepřítele jsou zaručené cesty ke zkáze každého šifrovacího systému.

LITERATURA

[1] Good, J., Michie, D., Timms, G. A. *General Report on Tunny*. [online] URL:< http://www.alanturing.net/tunny_report/>
<http://www.ellsbury.com/tunny/tunny-000.htm>
<http://www.codesandciphers.org.uk/documents/newman/newman.pdf>

[2] Bauer, F. L. *Decrypted Secrets: Methods and maxims of Cryptology*. Springer, Berlin, 2007

[3] Tutte, W. T. Fish and I. *Coding Theory and Cryptography*. Springer, Berlin, 2000

[4] Michie, D. Colossus and the Breaking of the Wartime „Fish“ Codes. *Cryptologia*, 2002, roč. 26, č. 1, s. 17-58.

[5] Hanuš, R. *Dálnopisná technika systému HELL*. Nakladatelství dopravy a spojů, Praha, 1974

[6] Sale, T. *The Updated Virtual Tunny machine*. [online] URL:<http://www.codesandciphers.org.uk/tunny/tunny.htm>

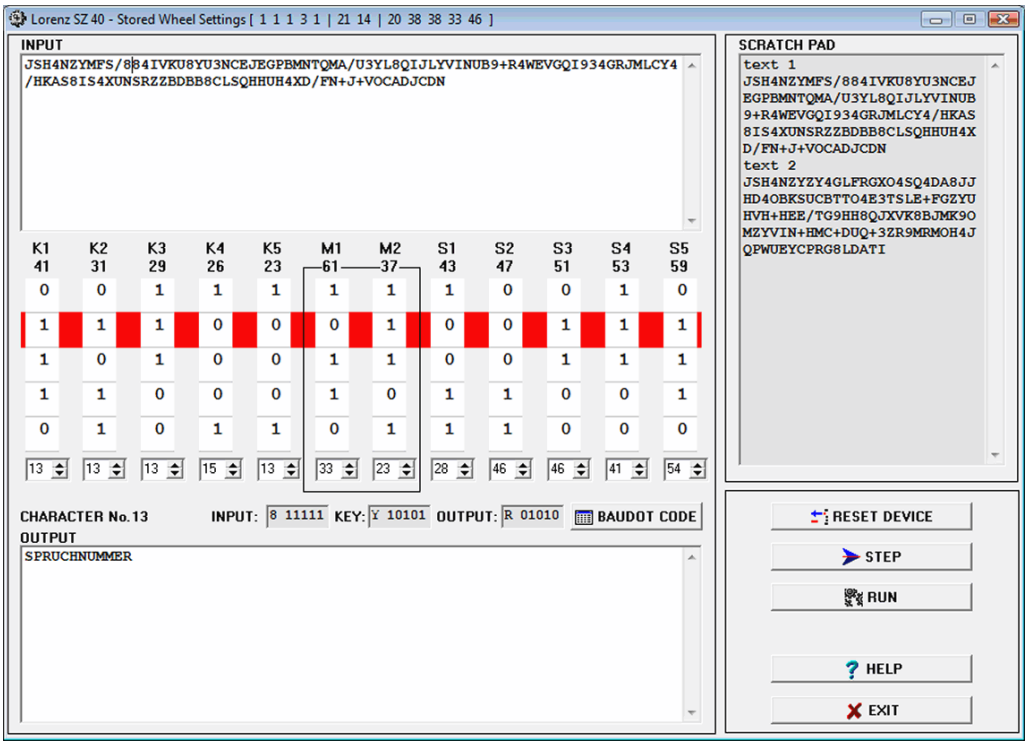
PŘÍLOHY

OBSAH PŘILOŽENÉHO CD-ROM

\

klic_zmug	
key.txt	posloupnost znaků klíče zpráv ZMUG
key_impuls*i*.txt	*i*-tý impuls této posloupnosti, $1 \leq i \leq 5$
s_prime.txt	posloupnost S' z klíče zpráv ZMUG
s_prime_impuls*i*.txt	*i*-tý impuls této posloupnosti, $1 \leq i \leq 5$
simulator	
src	
.	zdrojové soubory simulátoru v jazyce Delphi
lorenz.exe	softwarový simulátor stroje Lorenz SZ40
msg1.txt	začátek jedné ze zpráv ZMUG
msg2.txt	začátek druhé ze zpráv ZMUG
zmug_settings.txt	vzorky a nastavení odvozené v kapitole 3
tabulky	
grafy_ic.xls	tabulky hodnot ke grafům 3.7-11, 3.34, 3.35
grafy_pocetbehu.xls	tabulky hodnot ke grafům 3.20, 3.23-26, 3.31
ridici_posloupnost.xls	řídicí posloupnost $\mathcal{R}$ odvozená na str. 40
tabulky_bigramy.xls	tabulky počtů bigramů ve dvojicích sloupců pro odvození vzorků kol $\mathcal{K}_i$
tabulky_vzorky_s.xls	tabulky s délkami běhů v posloupnostech S_i' pro určení vzorků kol S_i
prace.pdf	text této práce ve formátu PDF

UŽIVATELSKÁ DOKUMENTACE SIMULÁTORU LORENZ SZ40



OVLÁDACÍ PRVKY

Simulátor má snadné a intuitivní ovládání, většina funkcí je přístupná přes kontextová menu příslušných prvků aplikace. Rozhraní programu je pro univerzálnější použití v anglickém jazyce.

POLE INPUT

Toto pole je určeno pro text k zašifrování nebo dešifrování. Text je možné vložit pomocí klávesnice, schránky nebo načíst z textového souboru. Tato možnost je přístupná přes kontextové menu pole.

Text je při šifrování zpracováván od pozice kurzoru v tomto poli. Nezáleží na velikosti písmen. Znaky, které nemají přiřazenu Baudotovu reprezentaci, jsou do výstupu zkopírovány beze změny.

POLE OUTPUT

V tomto poli se objevuje výstup stroje. Je možné jej uložit do textového souboru (kontextové menu).

POLE SCRATCH PAD

Toto je pomocné textové pole pro uživatele, mimo jiné lze jeho obsah načíst z či uložit do textového souboru.

ROTORY

Nastavení kolíčků rotorů jsou znázorněny hodnotami 0 a 1. Aktivní kolíčky jsou vyznačeny červeným pruhem.

Vzorky jednotlivých kol je možné nastavit manuálně (po kliknutí na kolo se zobrazí příslušné dialogové okno), k manuálnímu nastavení počátečních pozic kol slouží pole s šipkami pod každým rotorem.

Kontextové menu oblasti rotorů umožňuje:

- uložit aktuální nastavení kol do paměti pro pozdější reset do těchto pozic (uložené nastavení je zobrazeno v titulku okna)
- načíst vzorky a nastavení kol z textového souboru (formát souboru je blíže popsán v nápovědě programu)
- uložit vzorky a nastavení kol do textového souboru
- zvolit, zda se mají kola otáčet po každém kroku šifrování
- vložit do výstupního pole samotný pseudonáhodný klíč (všechny impulsy nebo vybraný impuls)

TLAČÍTKA

- **Baudot Code** zobrazí tabulku Baudotova dálnopisného kódu podle konvencí Bletchley parku
- **Reset Device** nastaví kola do dříve uložených pozic, smaže obsah pole Output a nastaví kurzor v poli Input na začátek
- **Step** provede jeden krok šifrování
- **Run** zpracuje zbylý text v poli Input od pozice kurzoru
- **Help** zobrazí nápovědu programu
- **Exit** ukončí aplikaci

UKÁZKOVÉ SOUBORY

Na přiloženém CD-ROM jsou ve stejném adresáři, jako vlastní aplikace, ukázkové textové soubory:

- *msg1.txt*, *msg2.txt* jsou začátky zpráv ZMUG podle [Bauer], kontrolní znaky jsou upraveny podle konvencí Bletchley Parku
- *zmug_settings.txt* obsahuje nastavení a vzorky kol odvozené ve třetí kapitole ve formátu pro jejich přímé načtení do aplikace