Statistički testovi za polialfabetske šifre

Mišić, Fran

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Fran Mišić

STATISTIČKI TESTOVI ZA POLIALFABETSKE ŠIFRE

Diplomski rad

Voditelj rada:
prof. dr. sc. Andrej Dujella

Zagreb, rujan, 2023.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Uvod

Znanost koja se bavi načinima i metodama tajne komunikacije zove se kriptologija. Riječ kriptologija dolazi od grčkih riječi *kryptos*-tajno i *logos*-riječ. Dijeli se na dvije discipline; kriptografiju i kriptoanalizu, od kojih svaka proučava jedan od dva osnovna dijela tajne komunikacije.

Kriptografija je znanstvena disciplina koja se bavi proučavanjem metoda za slanje poruka u takvom obliku da ih samo onaj kome su namijenjene može pročitati. Poruke koji pošiljatelj želi poslati zovu se *otvoren tekst* (engl. *plain text*). Takve poruke se, u postupku koji se naziva šifriranje, transformiraju u *kriptogram* ili *šifrat* (engl. *ciphertext*). Šifrat dalje putuje nekim komunikacijskim kanalom (npr. telefonska linija, računalna mreža i dr.) gdje je izložen trećoj osobi, protivniku ili neprijatelju (engl. *enemy*), koji može saznati sadržaj šifrata, ali ne može odrediti otvoren tekst. Primalac poruke može na temelju dobivenog šifrata i znanja o načinu šifriranja saznati otvoreni tekst odgovarajuće poruke.

Kriptoanaliza je znanstvena disciplina koja se bavi svim principima, metodama i načinima koji se upotrebljavaju u rješavanju, odnosno analizi kriptograma.

Klasična kriptografija se fokusirala na skrivanje sadržaja poslane poruke. Tajnost otvorenog teskta je bio prioritet svih metoda šifriranja. Takve vrste kriptiranja su imale posebno veliku važnost u vojnim operacijama. Povjesničari smatraju da su egipatski hijeroglifi, koji datiraju oko 1900. godine prije Krista, prvi primjeri šifriranja. Francois Champollin, koristeći kamen iz Rosette, uspio je dešifrirati hijeroglife 1822. godine. U antičkoj Grčkoj i Rimskom Carstvu također postoje dokazi ranog šifriranja poruka s namjerom skrivanja njegovog sadržaja. Najpoznatiji primjer je Cezarova šifra koja predstavlja primjer monoalfabetskog kriptiranja. Julije Cezar ju je koristio kako bi komunicirao sa svojim generalima. Veliki doprinos u područjima kriptgrafije i kriptoanalize su imali Arapi koji su prvi sistematski dokumentirali načine dešifriranja tajnih poruka, uključujući analizu frekvencija slova što je jedan od najznačajnijih principa kriptoanalize. Razvoj polialfabetskih šifri se pripisuje talijanu Leonu Battistu Albertiju, ocu zapadne kriptografije.

U novijoj povijesti, najznačajniji primjeri važnosti kriptologije bili su u 20. stoljeću, posebno za vrijeme dva svjetska rata, te razdoblju nakon. Ovaj period pratio je razvoj mehaničkih naprava za šifriranje te prvih računala kao što je britanski Colossus, prvo elektroničko, digitalno računalo sposobno izvoditi programe pomoću kojeg su britanci us-

pjeli dešifrirati njegmačke kriptograme generirane njemačkim Lorenz SZ40/42 strojem. Njemački ministar vanjskih poslova Arthur Zimmerman poslao je telegram 1917. godine u kojem je ponudio tajni vojni pakt Meksiku. Telegram su dešifrirali Britanci i objavili ga američkoj javnosti. Time su pomogli generirati potporu za ulazak Sjedinjenih Američkih Država u Prvi svjetski rat. Jedan od najpoznatijih primjera kriptografskih strojeva bila je njemačka Enigma, koja se koristila za vrijeme Drugog svjetskog rata. U ranim tridesetim godinama 20. stoljeća, poljski matematičar i kriptolog Marian Rejewski prvi je dešifrirao kod Enigme. Britanci su pod vodstvom matematičara A. M. Turinga koristili tehnike i opremu Poljaka te nastavili raditi na razbijanju kodova Enigme što im je pomoglo u pobjedi protiv Nijemaca na Atlantiku. U poslijeratnom periodu kriptogrami su također imali veliki značaj zbog Hladnog rata. Sovjetski i američki špijuni su međusobno komunicirali koristeći šifrate. Razbijanje sovjetskih one-time sustava šifriranja pomoglo je SAD-u otkriti špijune Rosenbergsa i Alger Hissa.

U modernoj kriptografiji tajnost poruke više nije prioritet. Naglasak je na tehnikama provjere integriteta poruka, autentikacije identiteta primatelja/pošiljatelja, digitalnih potpisa, sigurnih izračuna i slično.

U ovome radu proučavat će se polialfabetske šifre klasične kriptografije pri čemu je naglasak na metodama njihove analize i dešifriranja. Jedan od najutjecajnijih kriptologa bio je William Friedman, rođen 1891. godine u Rusiji, iz koje je emigrirao u Sjedinjene Američke Države. Smatra se ocem američke National Security Agency, kraće NSA. Objavio je brojna djela iz područja kriptografije i kriptoanalize te je imao veliki utjecaj u opisivanju metoda dešifriranja neprijateljskih kriptograma u oba svjetska rata. Gotovo sve opisane metode i primjeri u ovome radu temelje se na Friedmanove tri knjige *Military Cryptoanalysis Part I* [2], *Military Cryptoanalysis Part II* [3], *Military Cryptoanalysis Part III* [4].

Diplomski rad napravljen je u sklopu aktivnosti Projekta KK.01.1.1.01.0004 - Znanstveni centar izvrsnosti za kvantne i kompleksne sustave te reprezentacije Liejevih algebri.

Slika 0.1: William F. Friedman

Poglavlje 1

Supstitucijski sustavi

Započinjemo s nekoliko osnovnih pojmova i definicija, koji su potrebni za daljnju analizu. Izlaganje ovog dijela slijedi knjigu *Kriptografija* [1], gdje se mogu pronaći svi definirani pojmovi. *Abeceda A* je konačan skup znakova. Neki od primjera su:

1. abeceda binarnih brojeva,
2. abeceda brojeva,
3. klasična abeceda nekog jezika.

Rječnik nad abecedom A su svi konačni nizovi znakova te abecede. Označavamo ga s A^*. Elemente rječnika zovemo *riječi*, koje predstavljaju konačne nizove znakova abecede. Ne postoje zahtjevi da riječi budu razumljive, odnosno da pripadaju nekom prepoznatljivom jeziku. Na primjer, **DobarDan** i **xeDSmm** su riječi nad abecedom engleskog jezika.

Šifriranje ili *enkripcija* je familija funkcija koje preslikavaju elemente otvorenog teksta u elemente šifrata pri čemu abecede otvorenog teksta i šifrata ne moraju biti jednake. Permutacije abecede otvorenog teksta zovemo *komponentama otvorenog teksta* (engl. *plain component*), dok se permutacije abecede šifrata zovu *komponente šifrata* (engl. *cipher component*). Slovo, broj, riječ, fraza ili rečenica koje usmjeravaju i određuju način šifriranja i tako odabiru jednu funkciju iz familije, zovemo specifičan *ključ* enkripcije. Skup svih ključeva nazivamo *prostor ključeva*. Jedini uvjet za šifriranje jest da je ono u matematičkom smislu injekcija, odnosno postoji i inverzna transformacija koja se naziva *dešifriranje* odnosno *dekripcija* i koja preslikava šifrat u odgovarajući otvoren tekst. Obje matematičke funkcije šifriranja i dešifriranja zovemo *kriptografskim algoritmom*. *Kriptosustav* se sastoji od kriptografskog algoritma te svih mogućih otvorenih tekstova, šifrata i ključeva.

Definicija 1.0.1. *Kriptosustav ili kraće sustav, je uređena petorka (P, C, K, E, D) za koju vrijedi:*

1. *P je konačan skup svih mogućih osnovnih elemenata otvorenog teksta,*
2. *C je konačan skup svih mogućih osnovnih elemenata šifrata,*
3. *K je konačan skup prostora ključeva,*
4. *Za svaki $k \in K$ postoji funkcija šifriranja $T_k \in E$ i odgovarajuća funkcija dešfriranja $D_k \in D$. Pritom su $T_k : P \to C$ i $D_k : C \to P$ funkcije sa svojstvom da je $D_k(T_k(x)) = x$ za svaki otvoren tekst $x \in P$.*

Najvažnije svojstvo u definiciji je $D_k(T_k(x)) = x$. Iz njega slijedi da funkcije T_k moraju biti injekcije. Kada bi vrijedilo

$$T_k(x_1) = T_k(x_2) = y, \tag{1.1}$$

za dva različita otvorena teksta x_1 i x_2, onda primalac ne bi mogao odrediti treba li y dešifrirati u x_1 ili x_2, tj. $D_k(y)$ ne bi bilo definirano.

Možemo postaviti pitanje hoće li neki kriptosustav pružiti sigurnost. Kriptografija je natjecanje između dva protivnika:

- autora sustava (algoritma, prostora ključeva, implementacije),
- neprijatelja, koji pokušava doći do otvorenog teksta x iz šifrata y.

Osnovna pravila ovog natjecanja postavio je Kerckhoffs u knjizi *La Cryptographuc militare*. Naveo je 6 atributa koje sustav mora zadovoljavati kako bi pružio što veću sigurnost.

K1. Sustav bi trebao biti neslomljiv, ako ne u teoriji onda u praksi

Neslomljivost sustava se očituje u nemogućnosti određivanja ključa k te otvorenog teksta x iz šifrata y. Moguće je napraviti sustav koji je u matematičkom smislu neslomljiv, a primjer takvog je jednokratna bilježnica koja je detaljnije opisana u Poglavlju 5. Međutim, takvi sustavi su nepraktični. U praksi je važnije stvarno vrijeme i memorija koji su potrebni računalima da dođu do ključa k i otvorenog teksta x. Claude Shannon je razvio teoriju sigurnosti sustava te funkciju koja kvantificira snagu algoritma. Veće vrijednosti Shannoneve funkcije odgovaraju većoj sigurnosti. Minimalne vrijednosti Shannonove funkcije ovise o stvarnoj aplikaciji algoritma kriptiranja. Osobni medicinski zapisi možda zahtijevaju tajnost na duže vrijeme, od primjerice nekih vojnih planova.

K2. Znanje javnih informacija algoritma ne bi trebalo ugroziti njegovu sigurnost

Kriptografski sustav ima dva tipa informacija:

a) javne informacije, opis algoritama te prostor ključeva K,

b) privatne informacije, specifičan ključ k odabran za šifriranje.

Poznavanje javnih informacija ne smije ugroziti sigurnost sustava koja u potpunosti treba ovisiti o znanju specifičnog privatnog ključa.

K3. Metoda odabira privatnog ključa treba biti lako pamtljiva i promjenjiva

Ključ bi u idealnom sustavu trebao biti odabran potpuno nasumično. Međutim, korisnici trebaju balansirati između opasnosti od gubitka i zaborava ključa i opasnosti da neprijatelj pogodi ključ.

K4. Kriptogram mora biti prenosiv preko telegrafa

Šifrat se mora moći kodirati u niz 0 i 1.

K5. Sva potrebna oprema mora biti prenosiva

Današnji mikroprocesori svakako ispunjavaju ovaj uvjet.

K6. Korištenje sustava ne bi smjelo zahtijevati jako puno pravila i resursa

Jednostavnost, cijena i brzina sustava su glavni problemi šifriranja danas.

Postoje tri načina na koja se otvoreni tekst može pretvoriti u tajni tekst pa imamo podjelu na dvije klase kriptograma. U prvoj klasi, *transpozicije*, nedjeljivi elementi otvorenog teksta, bilo da su to slova, blokovi slova, cijele riječi ili rečenice, zadržavaju svoj identitet, ali mijenjanju svoje relativne pozicije tako da dobiveni šifrat postane nerazumljiv. Neka je m fiksan prirodan broj. Neka je $P = C = (Z_{26})^m$, te neka se K sastoji od svih permutacija skupa $1, 2, ..., m$. Za $\pi \in K$ definiramo

$$T_\pi(x_1, ..., x_m) = (x_{\pi(1)}, ..., x_{\pi(m)}), \tag{1.2}$$

$$D_\pi(y_1, ..., y_m) = (y_{\pi^{-1}(1)}, ..., y_{\pi^{-1}(m)}). \tag{1.3}$$

U drugoj klasi, *supstitucije*, elementi otvorenog teksta zadržavaju relativnu poziciju u tekstu, ali mijenjaju svoj identitet. Postoji i treća vrsta, gdje se otvoreni tekst prvo šifrira supstitucijom pa se primijeni transpozicija, ali su takve vrste ograničene u primjeni.

1.1 Monoalfabetske šifre

Monoalfabetska supstitucija $T : x = (x_1, ..., x_{n-1}) \longrightarrow y = (y_1, ..., y_{n-1})$ je preslikavanje koje svakom znaku abecede otvorenog teksta x_i pridružuje točno jedan znak abecede šifrata $\theta(x_i)$. Funkcija θ je permutacija abecede otvorenog teksta.

$$\theta : x_i \longrightarrow y_i = \theta(x_i), \quad 0 \leq t < n. \tag{1.4}$$

Ako s n označimo veličinu abecede otvorenog teksta, tada postoje točno $n!$ različitih vrsta monoalfabetskih šifrata. Pravilo pridruživanja često se prikazuje pomoću tablice, gdje se svakom znaku abecede otvorenog teksta pridružuje znak abecede šifrata.

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

Tablica 1.1: Jedno moguće šifriranje

U gornjoj tablici prikazano je jedno pravilo pridruživanja. Niz `ABCDEFGHIJKLMNOPQRSTUVWXYZ` je normalna, odnosno standardna komponenta otvorenog teksta, a niz `QWERTZUIOASDFGHJKLYXCVBNM` predstavlja jednu moguću permutaciju abecede.

Standardne abecede šifrata

Postoje dva tipa standardne, odnosno normalne abecede šifrata:

1. **Direktno standardna**

   Komponenta šifrata je jednaka normalnom poretku, ali je pomaknuta u lijevu ili desnu stranu.

   Komponenta teksta $\overrightarrow{\texttt{ABCDEFG}}\texttt{HIJKLMNOPQRSTUVWXYZ}$

   Komponenta šifrata $\texttt{HIJKLMNOPQRSTUVWXYZ}\overrightarrow{\texttt{ABCDEFG}}$

   Očito ima ukupno 25 mogućnosti šifriranja teksta ovom metodom (ako koristimo abecedu engleskog jezika $|A| = 26$). Direktno standardno šifriranje se još naziva i *Cezarovom šifrom* te je određeno jednim od 25 ključeva. Neka je $k \in \{1, 2, 3, ..., 25\}$ jedan ključ. Tada vrijedi:

   $$C_k : x \longrightarrow y = C_k(x) = (x + k) \pmod{26}. \tag{1.5}$$

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

Tablica 1.2: *Cezarova šifra* za ključ $k = 20$

2. **Suprotno standardna**

   Komponenta šifrata je jednaka standardnoj abecedi, ali u suprotnom poretku. Može početi na bilo kojem mjestu abecede pa ih ukupno ima 26.

   Komponenta teksta $\overrightarrow{\texttt{ABCDEFGHIJKLMNOPQRSTUVWXYZ}}$

   Komponenta šifrata $\overleftarrow{\texttt{QPONMLKJIHGFEDCBAZYXWVUTSR}}$

Miješane abecede šifrata

Ako je poredak znakova komponente šifrata različit od direktno ili suprotno standardnog, tada ga nazivamo *miješanim*. Dobiva se pomoću ključa ili nasumično.

1. **Abeceda određena ključem**

   U ovoj abecedi odredimo neku riječ koja će služiti kao ključ, uklonimo slova koja se ponavljaju te nakon toga navodimo slova abecede koja se ne pojavljuju u ključu. Za ključ `WESTERN FRONT` prvo moramo ukloniti slova koja se ponavljaju pa dobijemo `WESTRNFO`. Enkripcija je sljedeća:

   Komponenta teksta `ABCDEFGHIJKLMNOPQRSTUVWXYZ`

   Komponenta šifrata `WESTRNFOABCDGHIJKLMPQUVXYZ`

   Moguće je nakon odabira ključa primijeniti jednostavnu transpoziciju abecede šifrata pa se tako dobije *transpozicijski-miješana abeceda*.

2. **Abeceda dobivena nasumičnim odabirom**

   Veća sigurnost će se postići ukoliko komponentu šifrata dobijemo nasumičnim odabirom svih slova. U ovom nizu ne možemo biti sigurni u relativan poredak nekih slova, kao što možemo biti ako koristimo ključ. Nedostatak je što se algoritam mora zapisati jer se ne može pouzdano zapamtiti niti reproducirati pamćenjem jedne riječi kao u prethodnom načinu.

Primjer 1.1.1. U ovom primjeru pokazat će se šifriranje otvorenog teksta na engleskom jeziku. Sve prethodno opisano odnosi se na bilo koju abecedu otvorenog jezika. Koristit ćemo miješani način šifriranja s ključem `cipherkey`. Vrijedi sljedeća tablica enkripcije:

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

Otvoreni tekst je sljedeći:

`EXAMPLE OF ENCIPHERING USING MONOALPHABETIC SUPSTITUTION.`

Svakom slovu otvorenog teksta odgovara točno jedan znak komponente šifrata. Na taj ćemo način izvršiti zamjenu i dobiti sljedeći šifrat:

`EWCGMFE LR EJPAMYEOAJK TQAJK GLJLCFMYCIESAP QTMQSASTSALJ.`

Radi preglednosti, običaj je kriptograme pisati u skupinama od pet slova, pa dobijemo:

```
EWCGM FELRE JPAMY EOAJK TQAJK
GLJLC FMYCI ESAPQ TMQSA STSAL J.
```

1.2 Polialfabetske šifre

Kod *monoalfabetskih* šifri koristimo jedno pravilo za šifriranje svih slova

$$y_i = \theta(x_i) \tag{1.6}$$

Kod *polialfabetskih* šifri koristimo više od jednog pravila

$$y_i = \theta_i(x_i), \quad 0 \leq i < n \tag{1.7}$$

Za ključ $k = (k_0, k_1, ..., k_{n-1})$ polialfabetsku supstituciju možemo promatrati kao generalizaciju monoalfabetske Cezarove šifre C_k prema pravilu:

$$x \longrightarrow y = (y_0, y_1, ..., y_{n-1}), \quad y_i = C_{k_i}(x_i), \quad 0 \leq i < n. \tag{1.8}$$

U monoalfabetskom šifratu postoji jedna abeceda šifrata dok je u polialfabetskom broj abeceda jednak duljini ključa. Što je ključ duži to imamo više abeceda te je algoritam generalno sigurniji. Osnovna podjela je na periodičke i aperiodičke sustave. Kada proces šifriranja uključuje metodu koja je repetitivna te koja rezultira pojavom cikličkih fenomena u kriptogramu, sustav nazivamo *periodičnim*. Kada proces nije tog tipa, nazivamo ga *aperiodičnim*. Detaljniju analizu i primjere polialfabetskih šifri obradit ćemo u narednim poglavljima.

Poglavlje 2

Frekvencije

Prije analize kako frekvencije pomažu u dešifriranju kriptograma, navodimo četiri osnovna principa kriptoanalize prema William F. Friedmanu:

1. ***Otkrivanje korištenog jezika***
   Ovaj korak kriptoanalize ne zahtijeva detaljnu analizu. U kriptoanalizi za vojne svrhe, neprijatelj gotovo uvijek koristi materinski jezik. Prije su najčešće korišteni jezici za diplomatske svrhe bili francuski i engleski.
   U specijalnim slučajevima, korišteni jezik može se otkriti iz same prirode i kompozicije šifrata. Primjerice, ako se slova `K` i `W` uopće ne pojavljuju u šifratu, moglo bi se zaključiti da je otvoreni tekst na španjolskom jer se ta slova u španjolskom jeziku koriste samo za pisanje stranih riječi i imena. Određene kombinacije slova mogu biti indikatori jezika. Često pojavljivanje digrafa `CH` može upućivati na njemački jezik.
   Određeni koraci kriptoanalize mogu se izvršiti i prije poznavanja jezika otvorenog teksta, poput analize frekvencija. Nakon analize rezultati se mogu podijeliti s prevoditeljima te se uz njihovu pomoć može odrediti korišten jezik.

2. ***Određivanje korištenog kriptografskog sustava***
   Ovaj korak predstavlja najduži i najteži dio analize. Svaki postupak dešifriranja supstitucijskog kriptograma zasniva se na redukciji na monoalfabetsku supstituciju, ako već nije u tom obliku. Kriptogram se mora reducirati na monoalfabetsku supstituciju koristeći svojstva algoritama i korištenog jezika. Ukoliko taj postupak ne dovodi do rješenja, ne preostaje ništa drugo nego iscrpan proces eliminacije mogućih metoda.
   Nažalost, u kriptoanalizi ne postoje odlučujući procesi ili testovi koje možemo primijeniti na kriptogram kako bi se otkrio korišten algoritam. Zbog toga je analiza izoliranih i kompliciranih kriptograma izuzetno teška. Naglasak je na riječi izoliran kriptogram, zbog toga što se u stvarnim okolnostima zbog nepažnje ili drugih raz-

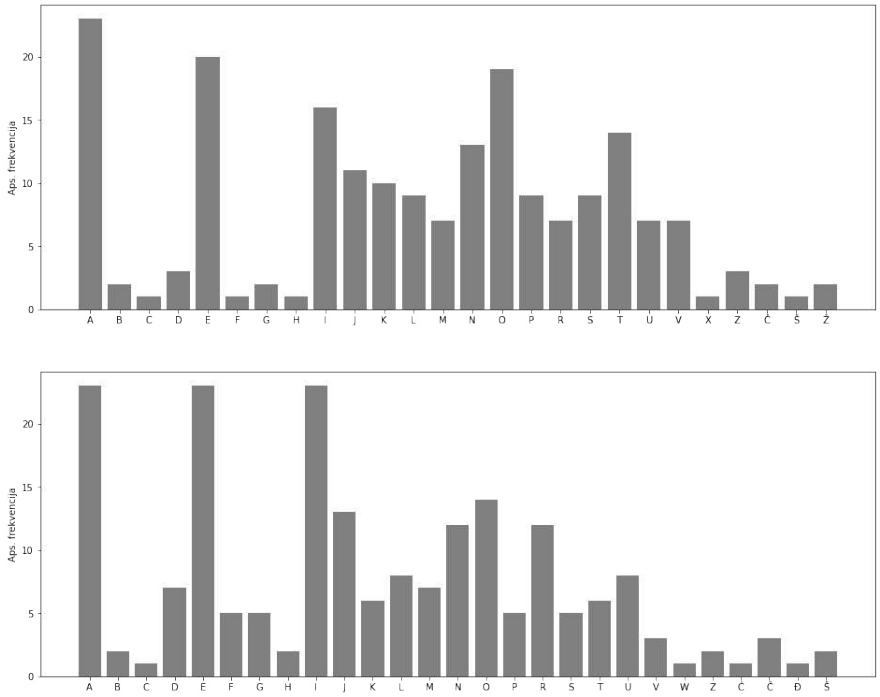
loga događaju greške osoba koje šifriraju poruke. Osim grešaka i curenja informacija neprijatelja, akumulacija dostupnih šifriranih poruka omogućava otkrivanje sustava.

3. ***Rekonstrukcija specifičnog ključa***
   Gotovo sve metode šifriranja koriste neku vrstu ključa za određivanje ili modifikaciju cijelog sustava. U nekim slučajevima nije potrebno odrediti cijeli ključ, nego je dovoljno odrediti neke njegove karakteristike, poput duljine ključa ili broja stupaca korištenih u transpozicijskom sustavu. Ipak, ako se ne radi o izoliranom šifratu, preporuča se potpuna rekonstrukcija jer nam može dati dodatne informacije o prirodi korištenih ključeva. Često se postupak određivanja ključa odvija paralelno s dešifriranjem teksta, a nekada se ključ odredi i nakon što je cijeli tekst dešifriran.

4. ***Rekonstrukcija otvorenog teksta***
   Kao što je prethodno rečeno, ovaj korak se često odvija paralelno s cijelim procesom analize teksta što će se vidjeti u narednim primjerima. U slučaju supstitucijskih šifri znakovi šifrata se zamjenjuju znakovima otvorenog teksta, a u transpozicijskim im se mijenja raspored. Ako je otvoren tekst na stranom jeziku, potreban je prevoditelj.

Početci analize frekvencija javljaju se u 14. stoljeću kada je Qalqashandi objavio veliku enciklopediju koja je uključivala zapažanja Ibn ad-Duraihima. U Europi se analiza frekvencije u kriptologiji pojavila u 15. stoljeću u Italiji.

Dobro je poznato da se sva slova jezika ne pojavljuju u jednakim frekvencijama. Neka slova poput `A, E, I, N` se pojavljuju puno češće od slova `F, Z i B`. Sljedeći grafovi prikazuju apsolutne frekvencije slova za dva odlomka uvodnog dijela teksta ovoga rada. Za prikaz primjera, uklonjena su zadnja slova dužeg odlomka kako bi broj slova bio jednak 200.



Iako se varijabilnost u grafovima očituje u relativnim frekvencijama različitih slova, razlike u apsolutnim frekvencijama pojedinih slova gotovo nema. Ova sličnost u grafovima bila bi još veća kada bi odlomci bili duži. Kada se dva teksta koja sadrže više od 1,000 slova te su sličnog karaktera usporede, pokazuje se da im se grafovi frekvencija gotovo ne razlikuju. Ta činjenica znači da se u normalnom tekstu svako slovo pojavljuje s konstantnom, odnosno *karakterističnom* frekvencijom prema kojoj teži. Što je tekst duži, to je frekvencija bliža karakterističnoj. Iduće tablice prikazuju frekvencije slova u engleskom, hrvatskom i njemačkom jeziku. Podatci za hrvatski jezik su dobiveni analizom tekstova iz dnevnog tiska, dok su podatci za engleski i njemački jezik preuzeti iz F. L. Bauerove knjige *Decrypted Secrets. Methods and Maxims of Cryptology* i W. F. Friedmanove knjige *Military Cryptoanalysis Part I* [2].

Hrvatski		Engleski		Njemački	
A	115	E	130	E	175
I	98	T	92	N	98
O	90	N	79	I	77
E	84	R	76	R	75
N	66	O	75	S	68
S	56	A	74	A	65
R	54	I	74	T	61
J	51	S	61	D	48
T	48	D	42	H	42
U	43	L	36	U	42
D	37	H	34	L	35
K	36	C	31	G	31
V	35	F	28	O	30
L	33	P	27	C	27
M	31	U	26	M	26
P	29	M	25	B	19
C	28	Y	19	F	17
Z	23	G	16	W	15
G	16	W	16	K	15
B	15	V	15	Z	11
H	8	B	10	P	10
F	3	X	5	V	9
		Q	3	J	3
		K	3	Y	1
		J	2	X	0
		Z	1	Q	0

	Frekvencija	Postotak	Zaokružen postotak
6 samoglasnika: A E I O U Y	398	39.8	40
20 suglasnika:			
5 visoke frekvencije: D N R S T	350	35.0	35
10 srednje frekvencije: B C F G H L M P V W	238	23.8	24
5 niske frekvencije: J K Q X Z	14	1.4	1
Ukupno:	1000	100.0	100

U daljnjoj analizi promatrat ćemo englesku abecedu, iako se svi rezultati mogu primijeniti i na hrvatsku, uz neke razlike za pojedina slova.

Primjećujemo da četiri samoglasnika `A, E, I, O` te četiri suglasnika `N, R, S, T` formiraju 661 od svih 1,000 slova otvorenog teksta, odnosno manje od trećine abecede gradi dvije trećine otvorenog teksta. Ova tablica frekvencije slova engleskog jezika dobivena

je iz teksta administrativnog državnog karaktera. Kada bi se uzimale frekvencije slova iz nekog teksta koji je komercijalnog ili nekog drugog tipa, frekvencije slova bi se malo razlikovale. Razlog je u drugačijem načinu pisanja zbog čega se neke riječi, a samim time i neka slova, češće pojavljuju od drugih. Kada bi promatrali telegrame vojnih poruka, riječi poput `STOP`, `PERIOD`, `COMMA`, `PLAN`, `BATTLE` i druge bi se češće pojavljivale te bi za slova koja se u njima pojavljuju frekvencije bile drugačije.

Osim karaktera teksta kojeg promatramo, jako je bitan volumen teksta. Prije je navedeno da frekvencija slova teži prema karakterističnoj frekvenciji što je tekst duži. Pokazalo se da su frekvencije slova značajno blizu karakterističnim kada je njihov broj barem 1,000. Sve više od toga ne pokazuje značajnu praktičnu razliku u frekvencijama. Kao primjer toga navodimo analizu frekvencija koju je proveo Nijemac Kaeding. On je 1898. godine prebrojao pojavljivanje slova u 11,000,000 riječi, što je ukupno oko 62,000,000 slova njemačkog teksta. Kada su se analizirale relativne frekvencije slova, pokazalo se da je razlika praktički neznačajna u odnosu na rezultate koje je dobio Kasiski, njemački kriptograf, koji je prebrojao 1,060 slova.

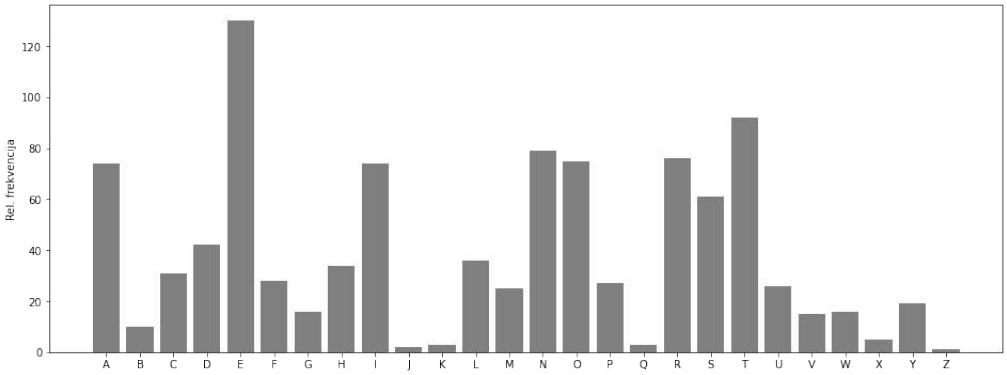
Karakteristične relativne frekvencije koje dobijemo kao rezultat analize velikog volumena teksta možemo smatrati *standardnim*, odnosno *normalnim* frekvencijama slova u pisanom jeziku. Općenito frekvencije konvergiraju prema normalnim frekvencijama. Neka je $f_n(i)$ frekvencija slova rednog broja i u proizvoljnom otvorenom tekstu x duljine n. Zbog zakona velikih brojeva, frekvencija slova će težiti prema standardnoj koju označavamo s $\pi(i)$.

$$\lim_{n\to\infty} f_n(i) = \pi(i). \tag{2.1}$$

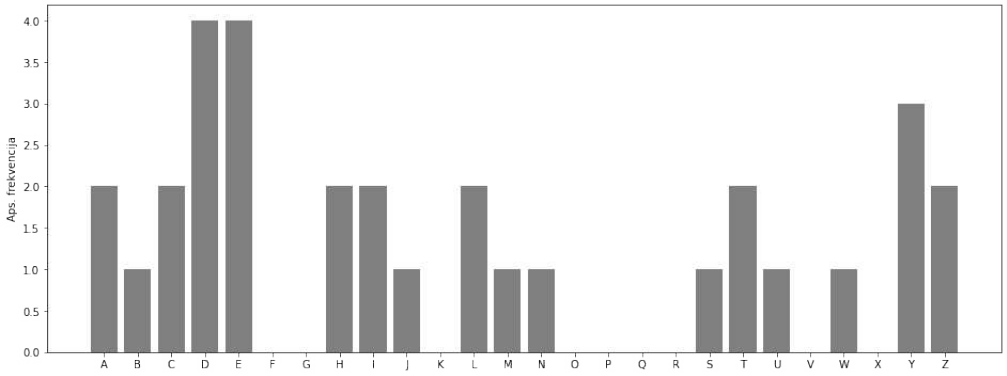
Što je volumen teksta manji, to će aproksimacije normalne frekvencije biti lošije. Zbog toga su kraći šifrati generalno puno teži za dešifriranje jer na temelju analize frekvencija ne možemo doći do pouzdanih zaključaka. Promotrimo izgled grafa normalnih frekvencija graf 1. Zbog velikih razlika u pojavljivanju pojedinih slova, pokazuju se nepravilnosti u izgledu. Kažemo da graf sadrži *brijegove i dolove*, odnosno točke niskih i visokih frekvencija. Relativne pozicije brijegova i dolova, odnosno prostor koji ih dijeli je fiksan zbog konvergencije frekvencija te bi se isti izgled grafa pojavio da analiziramo neki tekst drugog karaktera, ali s istim brojem slova, uz minimalne razlike koje su prije spomenute.

Ako napravimo analizu frekvencija za rečenicu s 40 slova graf 2, vidimo da histogram ne prati normalnu distribuciju. Neka slova se uopće ne pojavljuju, apsolutne i relativne pozicije dolova i brijegova ne odgovaraju normalnoj distribuciji pa ne možemo doći do istih zaključaka.

Osim analize frekvencije zasebnih slova mogu se i analizirati frekvencije digrafa, odnosno brojati pojavljivanje dvaju uzastopnih slova, npr. `AC`, `TH` i drugih. Detaljna analiza frekvencija digrafa, trigrafa i drugih može se pronaći u [2].



Graf 1: Standardna distribucija frekvencija engleskog jezika



Graf 2: Distribucija frekvencija teksta s 40 slova

2.1 Primjene distribucije frekvencije zasebnih slova

Ako se kriptogram sastoji od slova, analizom distribucije frekvencije zasebnih slova možemo saznati četiri činjenice.

Određivanje klase kriptograma
Moguće je relativno jednostavno odrediti pripada li promatrani kriptogram klasi transpozicija ili supstitucija zbog prirode obje metode. U klasi transpozicija, originalna slova otvorenog teksta su samo promijenila svoje pozicije, bez promjene identiteta. Iz toga slijedi da je frekvencija svakog slova otvorenog teksta **jednaka** frekvenciji tog istog slova u šifratu. U klasi supstitucija, slova mijenjaju svoj identitet te se zbog toga frekvencije slova otvorenog teksta i šifrata razlikuju. Zbog prethodno opisanih opažanja, ako su postotci samoglasnika, visoko, srednje i nisko frekventnih suglasnika aproksimativno jednaki onima

u prosječnom normalnom otvorenom tekstu, kriptogram vjerojatno pripada klasi transpozicijskih šifrata. Ako su postotci ovih slova u većoj mjeri različiti od očekivanih, kriptogram vjerojatno pripada klasi supstitucijskih šifrata.

Napomena 2.1.1. Najčešće su postotci u transpozicijskim šifratima blizu normalnih postotaka otvorenog teksta te su najčešće postotci u supstitucijskim šifratima različiti od normalnih postotaka otvorenog teksta. Ponekad je teško klasificirati šifrat s visokim stupnjem sigurnosti na temelju analize frekvencija zbog malog broja slova u šifratu. Kao što je prethodno rečeno, rezultati analize frekvencija često se ne mogu primijeniti ako imamo jako kratak šifrat jer se svojstva jezika ne mogu manifestirati na malom uzorku.

Pokazujemo četiri grafa, preuzeta iz *Military Cryptoanalysis* [2] koji redom prikazuju normalne frekvencije samoglasnika te visokih, srednje i nisko frekventnih suglasnika engleskog jezika na temelju kojih možemo klasificirati šifrat.

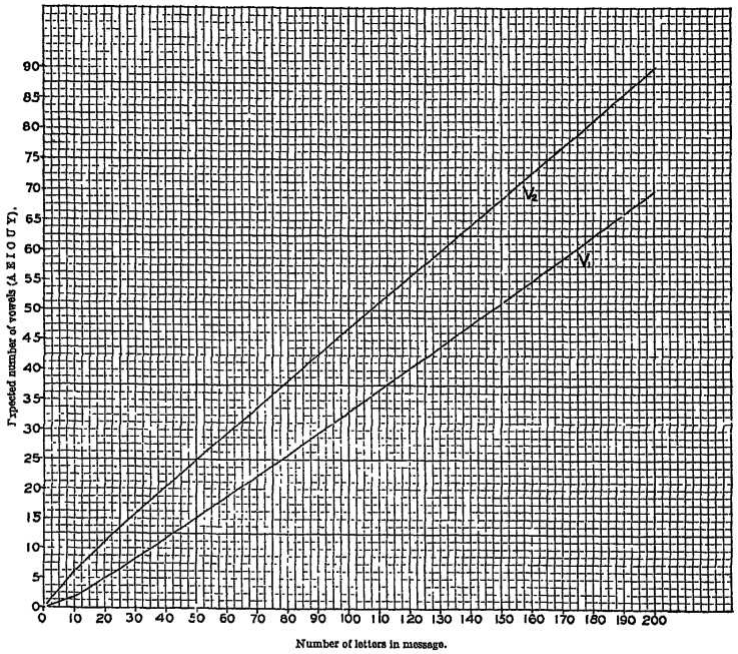


Chart 2. Curves marking the lower and upper limits of the theoretical amount of deviation from the number of vowels theoretically expected in messages of various lengths. (See subpar. 25d.)

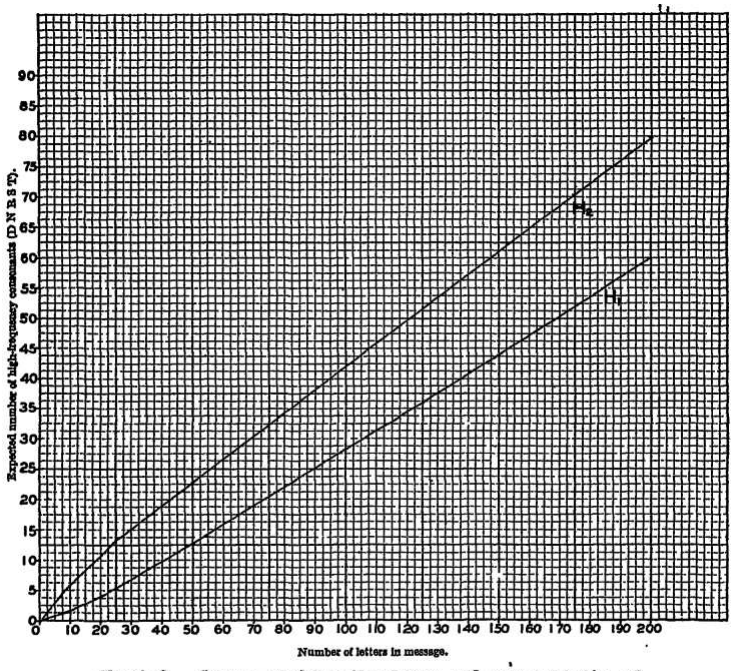


Chart 3. Curves marking the lower and upper limits of the theoretical amount of deviation from the number of high-frequency consonants theoretically expected in messages of various lengths. (See subpar. 25d.)

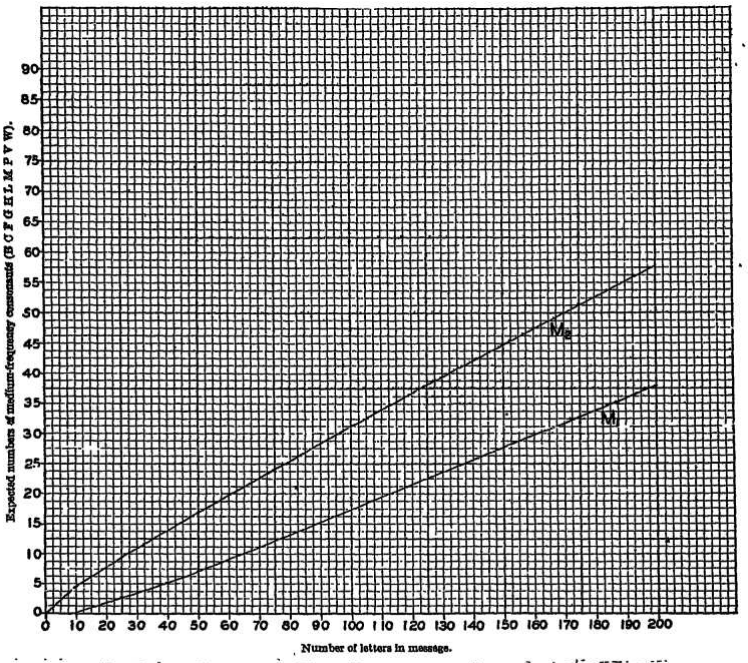


Chart 4. Curves marking the lower and upper limits of the theoretical amount of deviation from the number of medium-frequency consonants theoretically expected in messages of various lengths. (See subpar. 25d.)

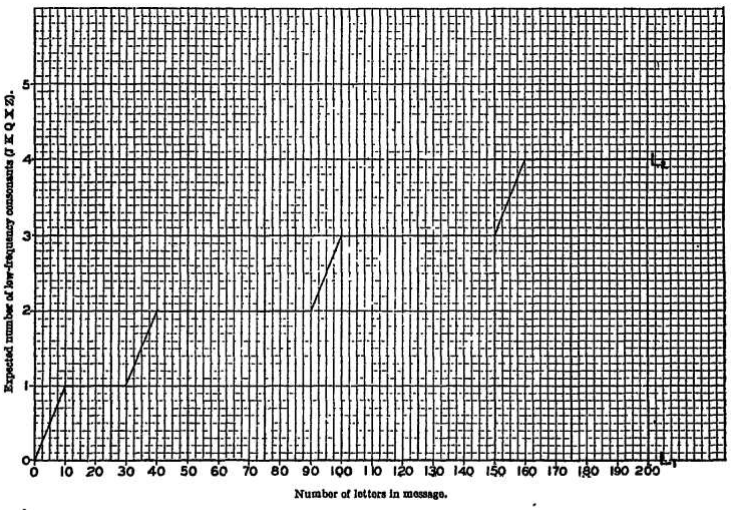


Chart 5. Curves marking the lower and upper limits of the theoretical amount of deviation from the number of low-frequency consonants theoretically expected in messages of various lengths. (See subpar. 25d.)

Za razumijevanje gornjih grafova uzmimo poruku koja ima 100 slova. Ako pronađemo vrijednost 100 na x-osi gornjih grafova, pripadajuće vrijednosti obje krivulje na grafovima određuju donju, odnosno gornju među na očekivan broj samoglasnika te visoko, srednje i nisko frekventnih suglasnika u otvorenom tekstu. Očekivani broj samoglasnika je između 33 i 47, dok je očekivani broj visoko frekventnih suglasnika 28 i 42, srednje frekventnih suglasnika 28 i 42 te nisko frekventnih suglasnika 0 i 3.
Ako frekvencije sve četiri klase slova u kriptogramu odgovaraju očekivanim vrijednostima (nalaze se unutar intervala) kriptogram možemo klasificirati kao transpozicijski. Ako jedna ili više klasa slova kriptograma iskaču iz normalnih ograničenja, možemo reći da je kriptogram supstitucijski. Udaljenost od normalnih vrijednosti pojedine klase može biti gruba mjera vjerojatnosti da šifrat nije transpozicijski.

Još jedna jednostavna metoda za određivanje klase je promatranje ponavljanja uzastopnih grupa slova. Ponavljanja grupa slova, riječi ili cijelih fraza su karakteristike normalnog otvorenog jezika. Kako u transpozicijskim šifratima mijenjamo poziciju slova u poruci, ta svojstva se razbijaju pa šifrat neće sadržavati više ponavljanja grupacija slova. Ako šifrat sadrži ponavljanja grupacija od 4, 5 ili više slova, možemo zaključiti da se radi o supstitucijskom šifratu.

Napomena 2.1.2. Ukoliko se šifrat sastoji od brojeva ili proizvoljnih simbola, očito je riječ o supstitucijskom šifratu.

Određivanje je li šifrat monoalfabetski ili polialfabetski
U normalnoj distribuciji frekvencija slova graf je izgledao neregularno, jer je imao puno brijegova i dolova. Znamo da se u monoalfabetskim kriptogramima svako slovo otvorenog teksta uvijek supstituira s točnom jednim slovom kriptograma. Iz tog bi razloga distribucija frekvencija slova monoalfabetskog šifrata također trebala sadržavati neregularan izgled pun brijegova i dolova, kao i normalna. Međutim, apsolutne pozicije brijegova i dolova neće biti jednake kao u normalnoj, odnosno slova kojima odgovaraju nepravilnosti u grafu neće biti jednaka, ali će opći izgled grafa biti sličan. Stoga, nepravilan izgled grafa distribucije frekvencija pojedinačnih slova koji je pun brijegova i dolova signalizira da postoji jedna abeceda šifrata, odnosno da je riječ o monoalfabetskom šifratu.
Ako je korišteno više abeceda šifrata, odnosno jedno slovo šifrata predstavlja više različitih slova otvorenog teksta, od kojih su neka visoke frekvencije, a druga niske, možemo primijetiti uniformniji izgled distribucije. Nedostatak istaknutih brijegova i dolova u distribuciji frekvencija slova šifrata signalizira da je riječ o polialfabetskom šifratu.

Dodatan test za klasifikaciju u šifratima koji sadrže do 200 slova jest promatranje broja praznina, odnosno broj slova koja se ne pojavljuju u distribuciji. Promatramo sljedeći graf u kojem krivulja P predstavlja prosječan broj praznina koje možemo očekivati u normalnom otvorenom tekstu. Krivulja R predstavlja prosječan broj praznina koje možemo očekivati u tekstu gdje su sva slova odabrana na potpuno nasumičan način. Ukoliko je točka koja predstavlja broj praznina našeg šifrata bliže krivulji P, možemo zaključiti da se radi o monoalfabetskom šifratu. Ako je bliže krivulji R, možemo zaključiti da se radi o polialfabetskom šifratu jer nasumičan odabir slova više nalikuje polialfabetskim šifratima gdje je korišteno više abeceda. Naknadno ćemo definirati Friedmanov indeks koincidencije kojeg također koristimo za razlikovanje monoalfabetskih i polialfabetskih sustava.

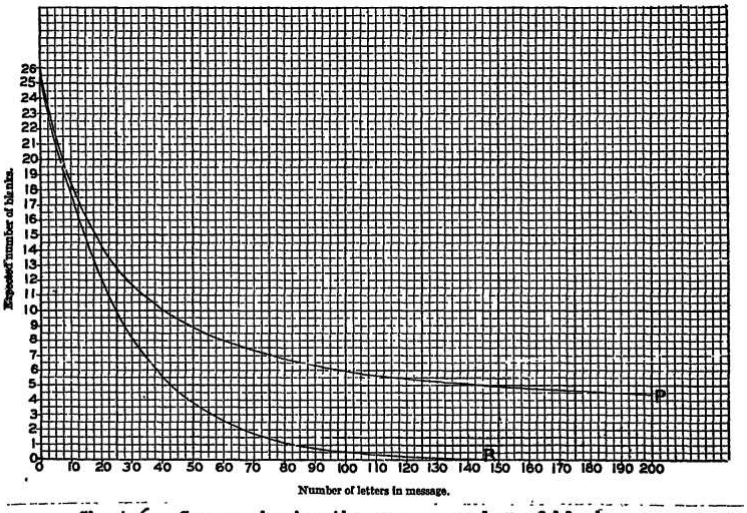


Chart 6. Curves showing the average number of blanks theoretically expected in distributions for plain text (P) and for random text (R) for messages of various lengths. (See subpar. 26f.)

Određivanje je li abeceda šifrata standardna ili miješana

Ako smo pomoću prethodnih metoda odredili radi li se o monoalfabetskom supstitucijskom šifratu, možemo odrediti je li korištena abeceda standardna ili miješana. Ukoliko se pozicije brijegova i dolova distribucije podudaraju s relativnim pozicijama brijegova i dolova normalne distribucije, riječ je o standardnom šifriranju. U suprotnom, velika je vjerojatnost da se radi o miješanom šifratu.

Određivanje je li standardna abeceda direktna ili suprotna

Pregledom poretka, odnosno smjera brijegova i dolova možemo reći radi li se o direktno standardnoj ili suprotno standarndoj abecedi.

Poglavlje 3

Polialfabetski sustavi ponavljajućeg ključa sa standardnom abecedom šifrata

U ovome poglavlju promatramo sustave koji koriste ponavljajući ključ za šifriranje i standardnu abecedu šifrata. Ovakva metoda šifriranja zove se Vigenèreovom šifrom, iako ju je prije izumio G. B. Bellaso. Prije nego krenemo u daljnju analizu, navest ćemo i objasniti tri glavna koraka za dešifriranje polialfabetskih sustava ponavljajućeg ključa:

1. *Određivanje perioda*
   Određivanje duljine specifičnog ključa, što odgovara broju korištenih abeceda šifrata.

2. *Redukcija na monoalfabetske sustave*
   Podjela slova šifrata s obzirom na abecedu kojoj pripadaju. Ovo je korak s kojim polialfabetski šifrat reduciramo na monoalafabetski.

3. *Identifikacija otvorenog teksta*
   Analiza i dešifriranje pojedinačnih monoalfabetskih šifrata.

Prvi korak: Određivanje perioda
Za određivanje perioda koristit ćemo sam kriptogram koji obično pokazuje vanjske cikličke fenomene koji su rezultati korištenja ponavljajućeg ključa. Za prikaz ovog koraka koristimo sljedeći primjer na hrvatskom otvorenom jeziku. Kada koristimo primjere na hrvatskom jeziku, slova Č i Ć, zamijenit ćemo sa slovom C, a slova Đ, DŽ, Lj, Nj, Š, Ž redom s D, DJ, LJ, NJ, S, Z. Pogledajmo sljedeću poruku:

NEPRIJATELJSKE SNAGE NAPREDUJU NA GRANICI RIJEKE DRAVE MOGUC
PRIJEVREMEN NAPAD NA GRAD U PETAK UJUTRO.

Za ključ ćemo odabrati riječ KLIS. Promotrimo abecede otvorenog teksta i 4 abecede šifrata.

Otvoren tekst		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Šifrat	1	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
	2	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
	3	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
	4	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R

S obzirom da ćemo koristiti četiri abecede, podijelit ćemo šifrat u grupe od četiri slova. Ispod svake grupe zapisat ćemo odgovarajuće slovo ključa.

```
NEPR  IJAT  ELJS  KESN  AGEN  APRE  DUJU  NAGR
KLIS  KLIS  KLIS  KLIS  KLIS  KLIS  KLIS  KLIS

ANIC  IRIJ  EKED  RAVE  MOGU  CPRI  JEVR  EMEN
KLIS  KLIS  KLIS  KLIS  KLIS  KLIS  KLIS  KLIS

      NAPA  DNAG  RADU  PETA  KUJU  TRO    .
      KLIS  KLIS  KLIS  KLIS  KLIS  KLIS
```

Sada šifriramo poruku tako što zbrojimo numeričke vrijednosti slova, A = 0, B = 1, C = 2, ..., Z = 25.

```
N + K = 13 + 10 = 23 (mod 26) = 23 = X
E + L = 4 + 11 = 15 (mod 26) = 15 = P
P + I = 15 + 8 = 23 (mod 26) = 23 = X
R + S = 17 + 18 = 35 (mod 26) = 9 = J
I + K = 8 + 10 = 18 (mod 26) = 18 = S
i tako dalje...
```

Dobijemo sljedeću poruku šifrata:

```
XPXJ SUIL OWRK UPAF KRMF KAZW NFRM XLOJ KYQU SCQB OVMV BLDW WZOM MAZA
                TPDJ OXMF XLXS NYIY BLLM ZPBS UFRM DCW.
```

Neovisno o sustavu koji je korišten, slova otvorenog teksta koja su šifrirana istom abecedom uvijek daju jednaka slova u šifratu. Ako je ključ duljine n, tada se otvoreni tekst može podijeliti u n klasa, odnosno stupaca, a svaki od kojih je šifriran jednim od n slova monoalfabetskim šifriranjem.

U slučaju kada je ključ razumne duljine, zbog ograničenosti pozicija slova u odnosu na ključ i zbog toga što se slova ponavljaju u otvorenom tekstu, jednaka slova otvorenog teksta moraju biti šifrirana istom abecedom, što znači da će postati jednako slovo u šifriranom tesktu. Isto vrijedi i za grupacije slova kraće ili jednake duljini ključa, u ovom primjeru za digrafe, trigrafe i tetragrafe. Takva ponavljanja slova ili grupacije slova u šifratu zovemo *uzročnim ponavljanjima* (engl. *casual repetitions*). Ukoliko se ponavljanja slova ili grupacije slova dogode zbog nasumičnosti, bez uočljivog razloga, zovemo ih *slučajnim ponavljanjima*. Primjerice, slovo `N` pripada stupcu koji je šifriran slovom `K`, dok slovo `P` pripada stupcu koji je šifriran slovom `I`, a oba postaju slovo `X`.

Naravno, slučajna ponavljanja će se češće događati s pojedinačnim slovima, rjeđe s digrafima, a još rjeđe s grupacijama od 3 ili više slova. Promatrat ćemo ponavljanja digrafa i trigrafa u primjeru kriptograma te zaključiti da je mala vjerojatnost da su sva ponavljanja slučajna. Udaljenost između dvije uzročne repeticije digrafa, trigrafa ili većih grupacija mora biti višekratnik duljine ključa. Dakle, jedan od faktora te udaljenosti će biti duljina ključa. Postavlja se pitanje kako odrediti točan faktor. Može se pokazati da je vjerojatnost slučajnih ponavljanja trigrafa ili tetragrafa jako niska (*Military Cryptoanalysis Part I* [2]). Zato možemo pretpostaviti da je većina ovih ponavljanja uzročna pa tražimo zajedničke faktore u udaljenostima između ponavljanja.

Ponavljanja	Udaljenost	Faktori
SU	75	3, 5, 15, 25, 75
LO	22	2, 11, 22
FK	4	2, 4
RM (prvo i drugo ponavljanje)	9	3, 9
RM (prvo i treće ponavljanje)	65	5, 13, 65
RM (drugo i treće ponavljanje)	56	2, 4, 7, 8, 14, 28, 56
MF	44	2, 4, 11, 22, 44
AZ	32	2, 4, 8, 16, 32
FR	56	2, 4, 7, 8, 14, 28, 56
XL	36	2, 3, 4, 6, 9, 12, 18, 36
BL	28	2, 4, 7, 14, 28
FRM	56	2, 4, 7, 8, 14, 28, 56

Promatrajući faktore udaljenosti svih ponavljanja, uočavamo da se brojevi 2 i 4 pojavljuju u gotovo svim repeticijama digrafa i u repeticiji trigrafa. Zbog praktičnosti, veća je vjerojatnost da je ključ duljine 4.

Ova promatranja također ovise o duljini poruke. Što je poruka duža, to će statistička svojstva teksta biti vidljivija te ćemo zaključke donositi s većom sigurnošću. U gornjoj tablici vidimo da postoje slučajna ponavljanja kojih bi svakako bilo više da je i poruka duža, ali proporcionalno bi uzročnih ponavljanja bilo još više.

Gore opisana metoda određivanja perioda zove se *faktorizacija intervala između ponavljanja* ili kraće *faktorizacija*.

Napomena 3.0.1. Korištenje šifrata s ponavljajućim kratkim ključem rezultira periodičnim ponavljanjem slova što se zove *ciklički fenomen*. Samo u slučajevima kratkih poruka koje su šifrirane s relativno dugačkim ključem, faktorizacija ne dovodi do rezultata. Prvi slučaj kada faktorizacija ne daje definitne rezultate je kada nismo koristili sustav ponavljajućeg ključa, a drugi slučaj je kada smo koristili monoalfabetski supstitucijski sustav.

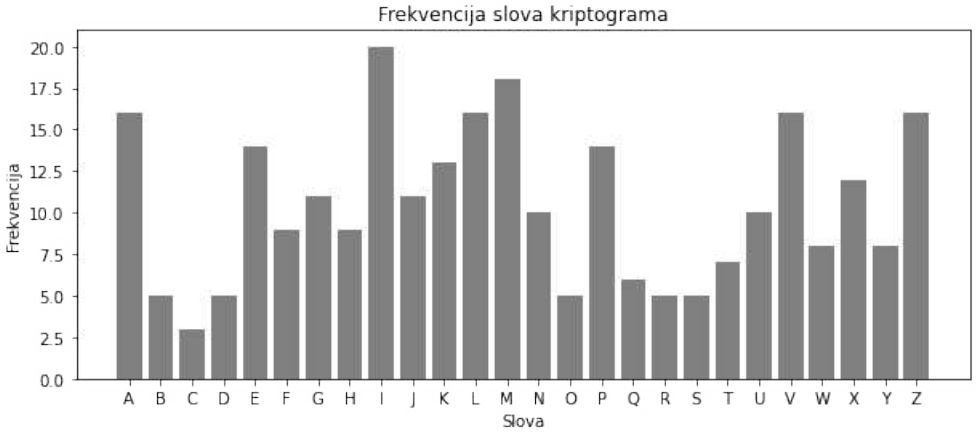
Drugi korak: Redukcija šifrata u monoalfabetske komponente
Nakon što odredimo dužinu k korištenog ključa, podijelimo slova šifrata u k klasa. To možemo vizualno prikazati stupčanom formom, što će se vidjeti u narednim primjerima. Svaka od klasa odgovara jednoj monoalfabetskoj supstituciji. Nakon podjele napravimo distribuciju frekvencija slova. Ako ove distribucije ne posjeduju karakteristične brijegove i dolove koje imaju monoalfabetski sustavi, tada je analiza koja je dovela do zaključka o broju korištenih abeceda kriva.

Treći korak: Rješavanje monoalfabetskih šifrata
Promatranjem distribucije frekvencija slova možemo zaključiti radi li se o standardnoj ili miješanoj abecedi. Analiza je dalje jednaka kao i kod monoalfabetskih šifrata, iako malo teža, jer radimo s manjom količinom teksta nakon što smo ga podijelili na k klasa.

Pogledajmo sljedeći primjer kriptograma:

	1	2	3	4	5
A.	AUKHY	JAMKI	ZYMWM	JMIGX	NFMLX
B.	ETIMI	ZHBHR	AYMZM	ILVME	JKUTG
C.	DPVXK	QUKHQ	LHVRM	JAZNG	GZVXE
D.	NLUFM	PZJNV	CHUAS	HKQGK	IPLWP
E.	AJZXI	GUMTV	DPTEJ	ECMYS	QYBAV
F.	ALAHY	POEXW	PVNYE	EYXEE	UDPXR
G.	BVZVI	ZIIVO	SPTEG	KUBBR	QLLXP
H.	WFQGK	NLLLE	PTIKW	DJZXI	GOIOI
J.	ZLAMV	KFMWF	NPLZI	OVVFM	ZKTXG
K.	NLMDF	AAEXI	JLUFM	PZJNV	CAIGI
L.	UAWPR	NVIWE	JKZAS	ZLAFM	HS

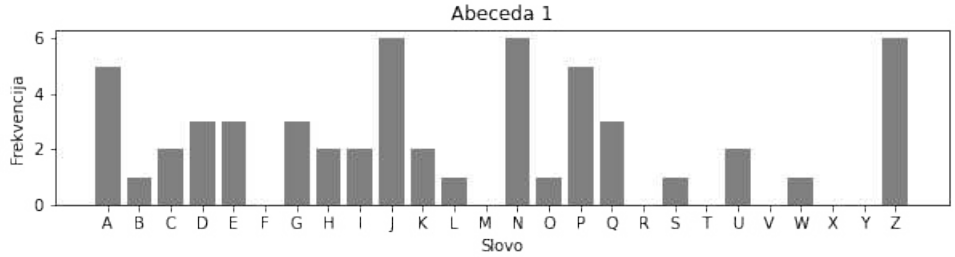
Napravimo distribuciju frekvencije slova. Vidimo da nema karakteristične brijegove i dolove koje asociramo s monoalfabetskim supstitucijama pa zaključujemo da se radi o polialfabetskoj supstituciji.



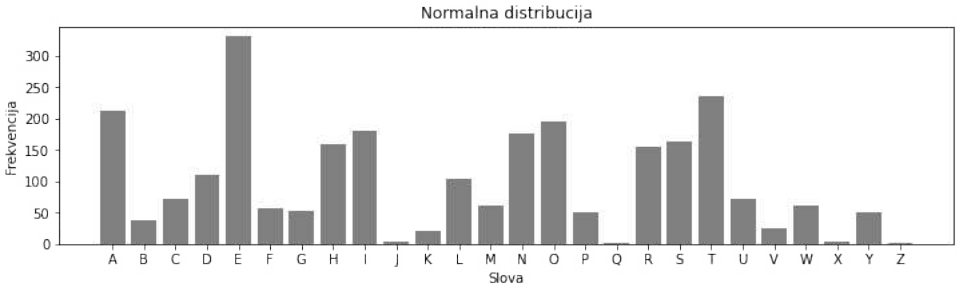
Pretpostavljamo da se radi o sustavu ponavljajućeg ključa. Provest ćemo faktorizaciju i dobiti sljedeću tablicu:

Ponavljanja	Udaljenost	Faktori
LUFMPZJNVC	160	2 4 5 8 10 16 20 32 40 80
JZXIG	90	2 3 5 6 9 10 15 18 30 45
EJK	215	5 43
PTE	50	2 5 10 25 50
QGK	85	5 17
UKH	55	5 11 55
ZLA	65	5 13 65
JA	60	2 3 4 5 6 10 12 15 20 30 60
LL	10	2 5 10
VX	20	2 4 5 10 20

Vidimo da se broj 5 pojavljuje u gotovo svim slučajevima pa zaključujemo da je period jednak 5. Slova šifrata su već grupirana u skupine od 5 slova. Sljedeći korak je analiza frekvencija svake pojedinačne klase.

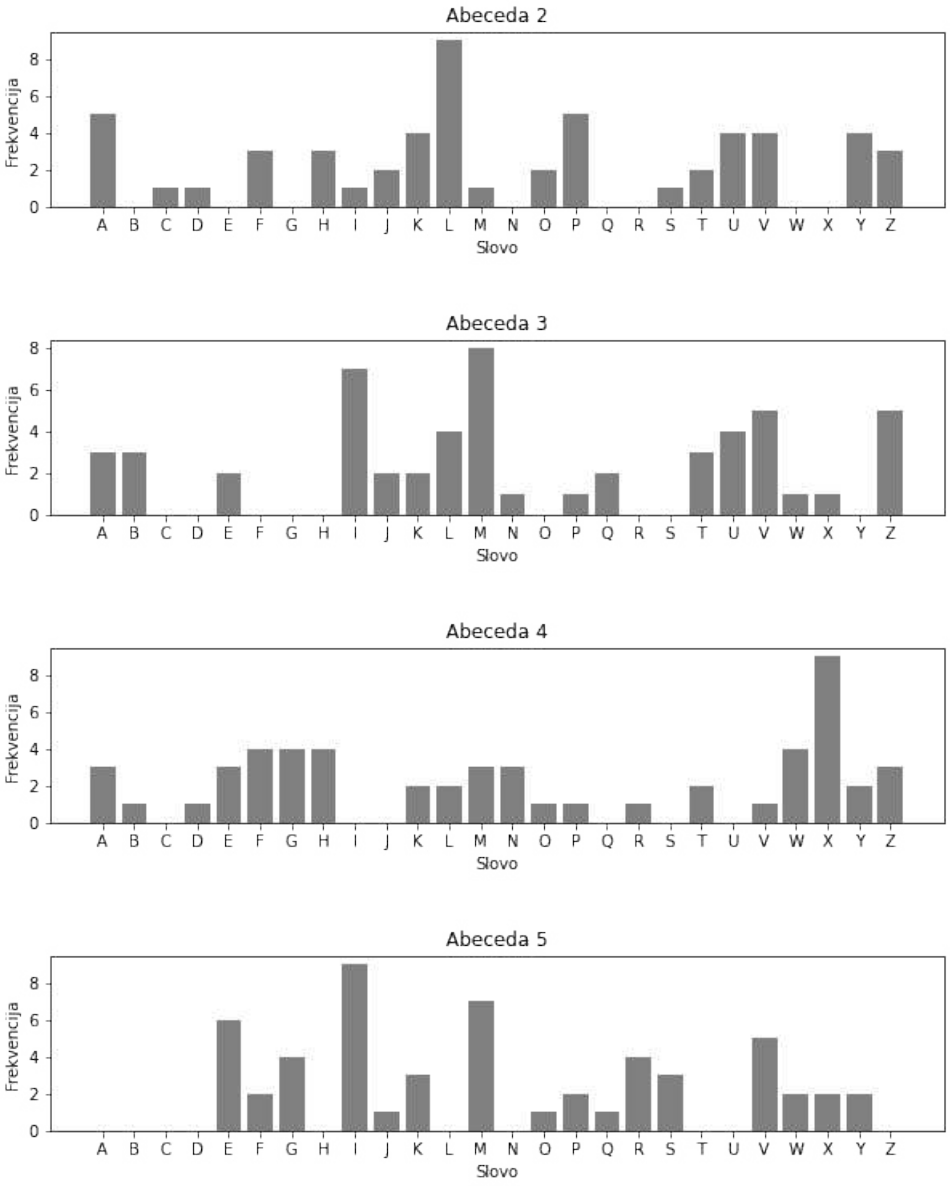


Promatranjem distribucije primjećujemo karakterističke brijegove i dolove monoalfabetske supstitucije u svim abecedama pa zaključujemo da je odabir perioda dobar.
Izgledno je da se radi o standardnim abecedama s obzirom na relativne pozicije brijegova i dolova pa nastavljamo s tom pretpostavkom. Promatranjem svake distribucije te pozicije brijegova i dolova, možemo nakon eksperimentiranja zaključiti koliki je pomak u svakoj abecedi, odnosno od kojeg slova počinje.
Primjerice, za prvu abecedu vidimo da su lokacije brijegova kod slova `A, J, N, P i Z`, a dolovi kod slova `B, F, L, M, O, R, S, T, V, W, X i Y`.



Uspoređujući s normalnom distribucijom možemo zaključiti, s obzirom na relativne pozicije brijegova i dolova, da vrijedi $\mathtt{A_p} = \mathtt{W_c}$, što znači da se slovo `A` u otvorenom tekstu šifrira u slovo `W`, te je korištena sljedeća pomaknuta standardna abeceda šifrata za prvu klasu slova:

Otvoren tekst	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Šifrat 1	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V



Analogno postupamo za ostale 4 abecede i dolazimo do ključa `WHITE`, te sljedeće tablice s 5 standardnih abeceda šifrata.

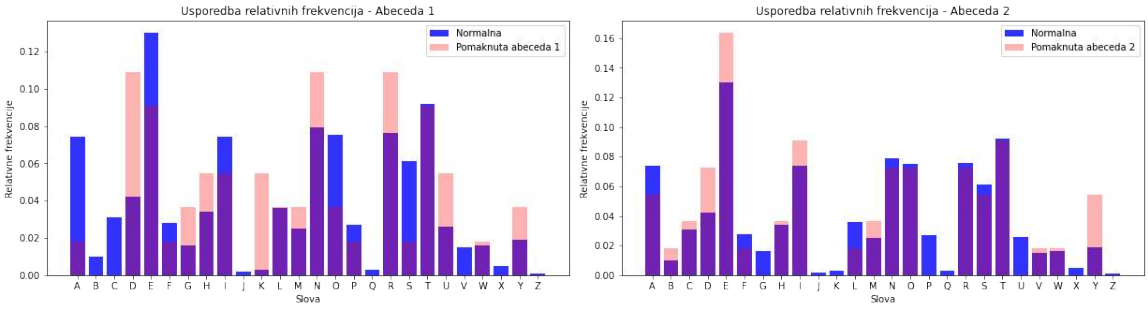
Otvoren tekst		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	1	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Šifrat	2	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
	3	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
	4	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	5	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

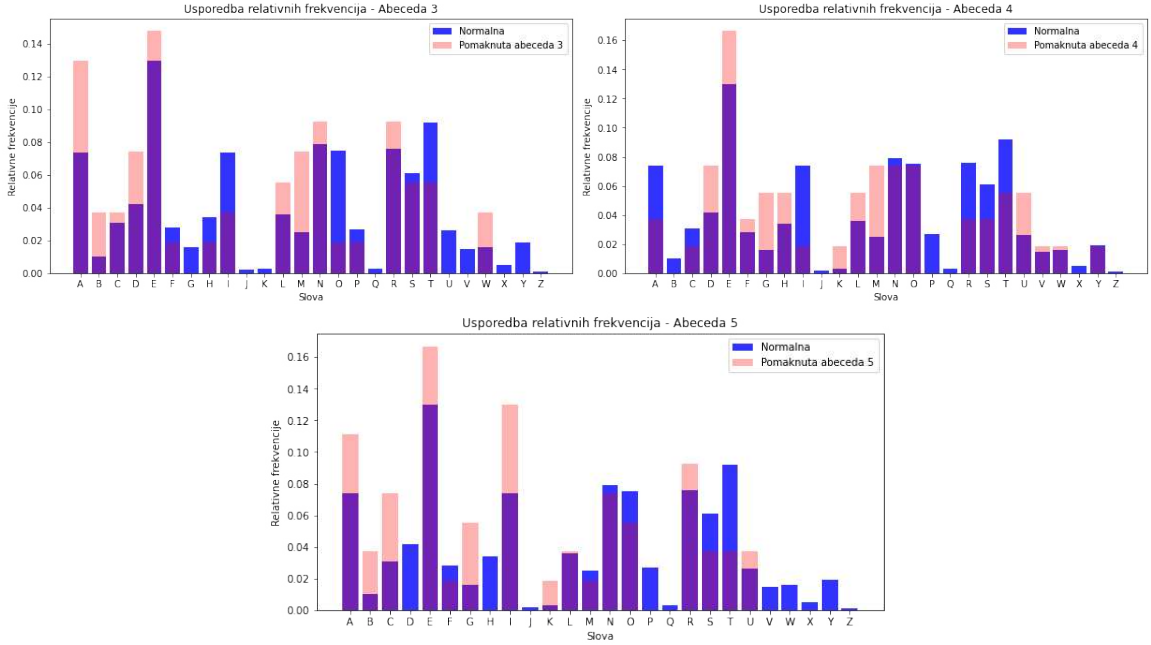
Tablica 3.1: Pet abeceda šifrata

Na kraju je jednostavno doći do otvorenog teksta, svako slovo preslikamo nazad u originalnu vrijednost, ovisno o abecedi koja je korištena za njegovo šifriranje. Dolazimo do sljedećeg teksta.

ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MACHINE GUN COMPANY IN TRUCKS NEAR EMMITSBURG AM HOLDING MIDDLE CREEK NEAR HILL FIVE FOUR THEE SOUTHWEST OF FAIRPLAY WHEN FORGED BACK WILL CONTINUE DELAYING REDS AT MARSH CREEK HAVE DESTROYED BRIDGES ON MIDDLE CREEK BETWEEN EMMITSBURG TANEYTOWN ROAD AND RHODESMILL.

Jedini pravi način da verificiramo naš postupak je konačan rezultat iz kojeg vidimo jesmo li došli do razumljivog teksta.
Navodimo 5 grafova na kojima su uspoređene relativne frekvencije normalne, odnosno standardne distribucije sa svakom od 5 distribucija abeceda šifrata, nakon što je napravljen odgovarajući pomak prikazan u tablici iznad 3.1.





3.1 Indeks koincidencije

Prethodni postupak koristi faktorizaciju za određivanje perioda sustava te analizu frekvencije slova kako bi eksperimentalno i vizualno pokušali odrediti vrijednost ključa. Metodu je uveo Friedrich Kaskiki 1863. godine. Sljedeća metoda je sistematičnija te se manje oslanja na intuiciju kriptoanalitičara. Koristi teoriju koincidencije koju je objavio William Friedman 1920. godine u djelu *Indeks koincidencije i njegove primjene kriptografiji*[5].

Definicija 3.1.1. *Neka je* $x = (x_0, ..., x_{n-1})$ *proizvoljan tekst. Indeks koincidencije od x, označavamo* $I_c(x)$*, definiramo kao vjerojatnost da su dva slučajno odabrana slova iz x jednaka. Ako s* $f_0, f_1, ..., f_{25}$ *označimo apsolutne frekvencije slova* $A, B, ..., Z$ *u tekstu x, indeks koincidencije je jednak:*

$$I_c(x) = \sum_{i=0}^{25} \frac{f_i(f_i - 1)}{n(n-1)}. \tag{3.1}$$

Dva elementa iz teksta x možemo odabrati na $\frac{n(n-1)}{2}$ *načina, a za svaki i postoji točno* $\frac{f_i(f_i-1)}{2}$ *načina odabira dva puta i-tog slova.*

Prethodna definicija vrijedi za proizvoljan tekst. U slučaju kada je tekst nasumično generiran, apsolutne frekvencije svih slova bi trebale biti otprilike jednake pa je indeks koincidencije jednak $I_c(x_{rand}) = 26 \cdot (\frac{1}{26})^2 = \frac{1}{26} = 0.038$. Ovu vrijednost označavamo s

κ_r. Ako računamo indeks koincidencije otvorenog teksta nekog jezika, njegova vrijednost ovisi o njegovoj veličini te o standardoj distribuciji frekvencija tog jezika. U slučaju hrvatskog jezika, za dovoljno velike tekstove, možemo očekivati da će indeks koincidencije biti jednak $I_c(x) \approx \sum_{i=0}^{25} \pi(i)^2 \approx 0.064$, označavamo κ_p. Za druge jezike pogledajmo sljedeću tablicu, preuzetu iz Konheimove knjige *Computer Security and Cryptography*.

Jezik	$I_c(x)$
Engleski	0.0688
Francuski	0.0778
Njemački	0.0762
Talijanski	0.0738
Španjolski	0.0775
Ruski	0.0529

Zaključujemo kako možemo koristiti indeks koincidencije kako bi odredili pripada li šifrat monoalfabetskom/transpozicijskom ili polialfabetskom sustavu. S obzirom da u transpozicijskom sustavu frekvencije svih slova ostaju jednake kao što su i u otvorenom tekstu, a u monoalfabetskom samo promijene svoj identitet, možemo očekivati, uz pretpostavku dovoljne veličine teksta, da će vrijednost indeksa koincidencije biti blizu očekivane za tražen jezik. Distribucija frekvencija kod polialfabetskih sustava su uniformnije pa možemo očekivati da će indeks koincidencije biti bliže $\kappa_r = 0.038$, kao kod nasumično generiranog teksta.

Isti primjer kojeg smo riješili Kaskikijevom metodom možemo riješiti i pomoću Friedmanovog indeksa koincidencije.

	1	2	3	4	5
A.	AUKHY	JAMKI	ZYMWM	JMIGX	NFMLX
B.	ETIMI	ZHBHR	AYMZM	ILVME	JKUTG
C.	DPVXK	QUKHQ	LHVRM	JAZNG	GZVXE
D.	NLUFM	PZJNV	CHUAS	HKQGK	IPLWP
E.	AJZXI	GUMTV	DPTEJ	ECMYS	QYBAV
F.	ALAHY	POEXW	PVNYE	EYXEE	UDPXR
G.	BVZVI	ZIIVO	SPTEG	KUBBR	QLLXP
H.	WFQGK	NLLLE	PTIKW	DJZXI	GOIOI
J.	ZLAMV	KFMWF	NPLZI	OVVFM	ZKTXG
K.	NLMDF	AAEXI	JLUFM	PZJNV	CAIGI
L.	UAWPR	NVIWE	JKZAS	ZLAFM	HS

Umjesto faktorizacije, izračunamo indeks koincidencije šifrata pa dobijemo $I_c = 0.0425$. Vidimo da je broj blizu 0.038 pa možemo pretpostaviti da se radi o polialfabetskom sus-

tavu, uz dodatnu provjeru distribucije frekvencija.
Sljedeći korak je određivanje perioda sustava. Princip je sljedeći; pretpostavimo da je p vrijednost perioda i podijelimo šifrat po stupcima u p klasa, svaka od kojih odgovara jednoj od p abeceda koje su korištene za šifriranje. Ako je odabir p bio dobar, svaka od p klasa trebala bi odgovarati monoalfabetskom sustavu, što provjeravamo izračunom indeksa. Ovaj korak obično se koristi s faktorizacijom kako bi bili sigurniji u naš odabir perioda.

Indeks koincidencije

period	Grupa 1	Grupa 2	Grupa 3	Grupa 4	Grupa 5	Grupa 6	Grupa 7
1	0.0425						
2	0.0401	0.0468					
3	0.0413	0.0474	0.0365				
4	0.0435	0.0435	0.0360	0.0558			
5	0.0552	0.0620	0.0657	0.0552	0.0734		
6	0.0386	0.0473	0.0323	0.0475	0.0374	0.0444	
7	0.0418	0.0324	0.0459	0.0513	0.0432	0.0432	0.0370

Vidimo da su vrijednosti svih klasa za period 5 blizu očekivanom indeksu koincidencije otvorenog teksta κ_p, pa možemo pretpostaviti da je period točan.

Koincidencije nam mogu pomoći odrediti i vrijednost ključa. Tu nam pomaže *međusobni indeks koincidencije dvaju nizova.*

Definicija 3.1.2. *Neka su $x = (x_0, ..., x_{n-1})$ i $y = (y_0, ..., y_{m-1})$ dva proizvoljna teksta. Međusobni indeks koincidencije od x i y, u oznaci $MI_c(x, y)$, definiramo kao vjerojatnost da je slučajno odabran element od x jednak slučajno odabranom elementu od y. Ako frekvencije slova abecede u x i y označimo s $f_0, ..., f_{25}$, odnosno $g_0, ...g_{25}$, vrijedi*

$$MI_c(x, y) = \sum_{i=0}^{25} \frac{f_i g_i}{nm}. \tag{3.2}$$

Neka su $c_1, ..., c_p$ klase slova, svaka od njih odgovara jednoj od p abeceda korištenih za šifriranje. Pretpostavimo da znamo i ključ $k = (k_1, ..., k_p)$. Pokušajmo procijeniti vrijednost $MI_c(c_i, c_j)$, za dvije proizvoljne klase c_i i c_j. Vjerojatnost da su dva proizvoljna slova u klasama c_i i c_j jednaka slovu `A` bit će jednaka vjerojatnosti da je na tom mjestu u otvorenom tekstu bilo slovo čiji je numerički ekvivalent $-k_i \pmod{26}$ ($-k_i + k_i = 0 \pmod{26} =$ `A`). Vjerojatnost da su dva proizvoljna slova jednaka `B` analogno će biti jednaka vjerojatnosti da je na tom mjestu slovo s numeričkim ekvivalentom $(1 - k_i) \pmod{26}$. Zaključujemo:

$$MI_c(c_i, c_j) \approx \sum_{l=0}^{25} \pi_{l-k_i} \pi_{l-k_j} = \sum_{l=0}^{25} \pi_l \pi_{l+k_i-k_j}. \tag{3.3}$$

Vrijednost $k_i - k_j$ nazivamo relativnim pomakom c_i i c_j. U sljedećoj tablici prikazane su vrijednosti očekivanog međusobnog indeksa koincidencije za različite vrijednosti po-

maka u hrvatskom jeziku. S obzirom da vrijedi $\sum_{l=0}^{25} \pi_l \pi_{l+k_i-k_j} = \sum_{l=0}^{25} \pi_l \pi_{l-(k_i+k_j)}$ dovoljno je promatrati prvih 13 vrijednosti.

relativni pomak	očekivana vrijednost od MI_c
0	0.064
1	0.039
2	0.031
3	0.031
4	0.044
5	0.040
6	0.039
7	0.033
8	0.040
9	0.042
10	0.036
11	0.036
12	0.036
13	0.039

Vidimo da postoji razlika u vrijednosti MI_c ako je pomak jednak 0 i za sve druge vrijednosti pomaka. To možemo iskoristiti za računanje relativnih pomaka klasa slova. Fiksiramo jednu klasu c_i te promatramo šifriranje c_j sa svim pomacima $r = 0, 1, 2, ..., 25$ i za svaku dobivenu klasu c_j^r izračunamo indeks $MI_c(c_i, c_j^r)$. Za $r = k_i - k_j$ bi vrijednost indeksa trebala biti oko 0.064, a inače između 0.031 i 0.044. Tako izračunamo sve relativne pomake te nam onda ostaje 26 mogućnosti za ključ koje provjerimo jedan po jedan.
Ukoliko znamo na kojem je jeziku otvoreni tekst, možemo prilagoditi metodu kako bi bila efikasnija. Ako x odgovara tipičnom tekstu hrvatskog jezika vrijednosti $\frac{f_i}{n}$ su približno jednake $\pi(i)$. Za svako slovo k_j ključa k izračunamo indeks između standardnog otvorenog teksta jezika x i pripadajuće klase slova c_j^r duljine n' po formuli

$$MI_c(x, c_j^r) \approx \sum_{i=0}^{25} \frac{\pi(i) f_{i-r}}{n'} \tag{3.4}$$

Ako je $r \equiv -k_j \pmod{26}$, vrijednost bi trebala biti blizu 0.064, a inače bi trebala biti manja od 0.045. Odredimo maksimalnu vrijednost indeksa r_{max} po svim slovima te stavimo $k_j \equiv -r_{max} \pmod{26}$. Izračunajmo sve vrijednosti međusobnog indeksa za 5 klasa u prethodnom primjeru.

Za svaku su grupu podebljane maksimalne vrijednosti te zbog prethodno opisanog, odabir numeričke vrijednosti slova s kojim je šifrirana ta abeceda je jednak $-r_{max} \pmod{26}$.

Abeceda	Vrijednosti međusobnog indeks koincidencije						
	0.03823	0.04259	0.03692	0.03456	0.05905	0.04859	0.03605
1	0.02691	0.03948	0.03573	0.03808	0.03848	0.02991	0.04001
	0.04261	0.04513	0.03032	0.04417	0.03745	0.04029	0.03873
	0.03922	0.02754	0.03293	0.04026	0.03663		
	0.03394	0.02757	0.04035	0.04236	0.04457	0.03634	0.04446
2	0.03679	0.04673	0.04209	0.03737	0.03201	0.02995	0.03253
	0.03516	0.04754	0.03106	0.02950	0.03656	0.06890	0.04232
	0.03377	0.03036	0.04191	0.03760	0.03817		
	0.03662	0.03420	0.03335	0.04028	0.03379	0.05036	0.04132
3	0.04472	0.03475	0.04818	0.03941	0.03559	0.02527	0.03166
	0.04300	0.03080	0.02648	0.03244	0.06577	0.04610	0.04097
	0.03325	0.04988	0.03136	0.03369	0.03674		
	0.03508	0.03530	0.02990	0.03903	0.03806	0.03649	0.04044
4	0.06064	0.04534	0.03418	0.02964	0.04055	0.03433	0.03841
	0.03984	0.03706	0.03770	0.04291	0.03729	0.03482	0.04304
	0.04145	0.04328	0.03843	0.03526	0.03153		
	0.04688	0.03274	0.03854	0.03365	0.03059	0.04317	0.03669
5	0.04366	0.03486	0.05552	0.03592	0.03870	0.03438	0.04233
	0.03559	0.03023	0.03585	0.02620	0.04899	0.02661	0.03590
	0.03480	0.06849	0.03277	0.04106	0.03587		

Dobijemo:

Abeceda 1:	$-4 \pmod{26} = 22 \pmod{26}$	Broj 22 odgovara slovu W.
Abeceda 2:	$-19 \pmod{26} = 7 \pmod{26}$	Broj 7 odgovara slovu H.
Abeceda 3:	$-18 \pmod{26} = 8 \pmod{26}$	Broj 8 odgovara slovu I.
Abeceda 4:	$-7 \pmod{26} = 19 \pmod{26}$	Broj 19 odgovara slovu T.
Abeceda 5:	$-22 \pmod{26} = 4 \pmod{26}$	Broj 4 odgovara slovu E.

Kada znamo ključ postupamo, jednako kao u prethodnoj metodi. Uvijek je dobro koristiti obje metode za rješavanje kriptograma kako bi mogli doći do zaključaka s većim stupnjem sigurnosti.

Poglavlje 4

Polialfabetski sustavi ponavljajućeg ključa s miješanom abecedom šifrata

Kod sustava koji koriste standardnu abecedu šifrata, vidjeli smo da povećavanje broja abeceda skriva neke frekvencijske fenomene koji su bili očiti kod sustava koji koriste jednu abecedu. Ta adaptacija je otežala dešifriranje kriptograma, ali zbog korištenja standardnih abeceda bila su vidljiva druga svojstva. Pomoću cikličkih fenomena koji su se manifestirali zbog ponavljanja ključa za šifriranje opisali smo dvije pouzdane metode za rješavanje takvih sustava. Korištenjem miješanih abeceda otežat ćemo dešifriranje kriptograma u polialfabetskim sustavima.
Miješana abeceda koristi ključ kako bi se odredila komponenta šifrata, odnosno niz u kojem će se koristiti slova abecede.
Za ključ `LEAVENWORTH` dobijemo sljedeću komponentu i pripadnu matricu:

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

Jedan ključ se koristi za definiciju gornje matrice, odnosno za odabir mogućih miješanih abeceda, a drugi ključ se koristi za odabir onih abeceda koje će biti korištene. Neka je BLUE ključ koji odabire koje abecede će se koristiti. Dobijemo sljedeću tablicu.

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

Princip direktne simetrije pozicije

Vidimo da su sve, od 26 mogućih miješanih abeceda prikazanih u tablici, zapravo jedan niz slova, određen ključnom riječi LEAVENWORTH, koji je pomaknut u intervalima od 1, 2, 3, ..., 25. Za simetriju koja se uočava kažemo da je *vidljiva*, odnosno *direktna*.

Primjećujemo, ako se u prvoj abecedi slovo V nalazi 15 mjesta udesno od slova G, tada će se

slovo V nalaziti 15 mjesta udesno od slova G u sve četiri korištene abecede. Zaključujemo da su im relativne pozicije jednake u svim mogućim abecedama matrice.
Dakle, ako su relativne pozicije dva slova θ_1 i θ_2 u nekoj abecedi C_1 poznate i ako znamo poziciju θ_1 u abecedi C_2, tada znamo i poziciju slova θ_2 u abecedi C_2. Ovaj postupak ćemo koristiti paralelno s analizom frekvencija digrafa i trigrafa kako bi dešifrirali neka slova i došli do "kostura" teksta na temelju kojeg ćemo zaključiti neke riječi i ponovno nastaviti ovaj postupak.

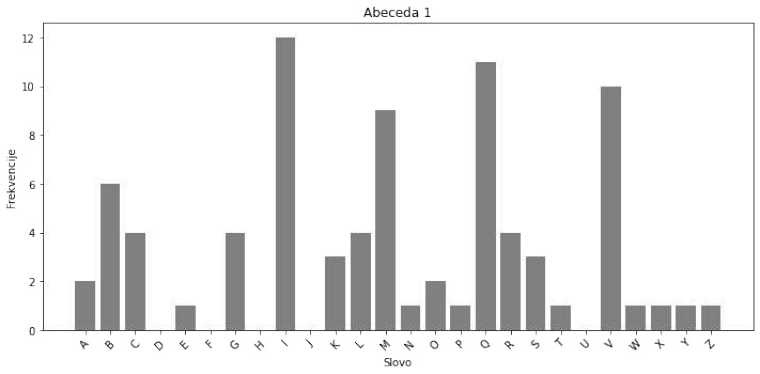
Pogledajmo sljedeći kriptogram.

	1	2	3	4	5
A	QWBRI	VWYCA	ISPJL	RBZEY	QWYEU
B	LWMGW	ICJCI	MTZEI	MIBKN	QWBRI
C	VWYIG	BWNBQ	QCGQH	IWJKA	GEGXN
D	IDMRU	VEZYG	QIGVN	CTGYO	BPDBL
E	VCGXG	BKZZG	IVXCU	NTZAO	BWFEQ
F	QLFCO	MTYZT	CCBYQ	OPDKA	GDGIG
G	VPWMR	QIIEW	ICGXG	BLGQQ	VBGRS
H	MYJJY	QVFWY	RWNFL	GXNFW	MCJKX
I	IDDRU	OPJQQ	ZRHCN	VWDYQ	RDGDG
J	BXDBN	PXFPU	YXNFG	MPJEL	SANCD
K	SEZZG	IBEYU	KDHCA	MBJJF	KILCJ
L	MFDZT	CTJRD	MIYZQ	ACJRR	SBGZN
M	QYAHQ	VEDCQ	LXNCL	LVVCS	QWBII
N	IVJRN	WNBRI	VPJEL	TAGDN	IRGQP
O	ATYEW	CBYZT	EVGQU	VPYHL	LRZNQ
P	XINBA	IKWJQ	RDZYF	KWFZL	GWFJQ
Q	QWJYQ	IBWRX			

Provest ćemo faktorizaciju i odrediti Friedmanov indeks koincidencije kako bismo odredili period.

Multigrafi	Faktorizacija udaljenosti
QWBRIVWY	45 = 3, 5, 9, 15
CGXGB	60 = 2, 3, 4, 5, 6, 10, 12, 15, 20
PJEL	95 = 5, 19
ZZGI	145 = 5
BRIV	285 = 3, 5, 15, 19
BRI	45 = 3, 5, 9, 15
KAG	75 = 3, 5, 15
QRD	165 = 3, 5, 15
QWB	45 = 3, 5, 9, 15
QWB	275 = 5, 11
WIC	130 = 2, 5, 10, 13
ZTC	145 = 3, 5

Vidimo da je faktor 5 čest u svim ponavljanjima pa pretpostavljamo da je period jednak 5. Pretpostavku provjerimo izračunom indeksa koincidencije. Dobijemo vrijednosti 0.075, 0.0741, 0.0873, 0.0662, 0.0711 pa zaključujemo da je pretpostavka dobra.



Pokušaji usporedbe distribucije frekvencija s normalnom ili suprotno normalnom ne dovode do rezultata. Možemo pretpostaviti da se radi o miješanoj abecedi što nam sugeriraju relativne pozicije brijegova i dolova grafa.
Prelazimo na analizu frekvencija trigrafa. S obzirom da pretpostavljamo da se radi o sustavu s periodom 5, dakle imamo 5 abeceda, trigrafi su razdvojeni po klasama. Ako imamo trigraf `QAC` koji će se pojaviti u distribuciji abecede 1, znači da je znak `Q` iz abecede prije nje, dakle abecede 5, znak `A` iz abecede 1, a znak `C` iz abecede poslije nje, abecede 2.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
QC	GW	NT		TV		AE		AS		UD	UW	IT	UT	QP	NX	-W	LB	LA	LA		IW	NN	QI	UX	QR
PT	OP	TC				AD		WC		FI	QX	II		UP		YW	YW	DE			IW				
	GK	TT				LX		HW		FW	LV	OT				NW	QD	RB			UE				
	OW	WB				LW		ND			LR	SY				QC	QD				LC				
	GL							GV				WC				GI					GP				
	GX							WC				GP				QL					QB				
								XD				AB				RI					NW				
								GB				JF				YV					QE				
								IV				DI				NY					IP				
								NR								SW					UP				
								AK								QW									
								QB																	

Tablica 4.1: Trigrafi centrirani u prvoj abecedi

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
SN	RZ	IJ	IM	GG	MD			MB		IW	QF		WB		BD		ZH	IP	MZ		IX	QB	GN	MJ	
TG	VG	QG	GG	VZ				QG		BZ	BG				OD		IG		CG		QF	VY	BD	QA	
	IE	VG	ID	SZ				QI							VW		LZ		NZ		LV	QY	PF		
	MJ	CB	RG	VD				KL							OJ				MY		IJ	LM	YN		
	SG	IG	KH					MY							MJ				CJ		EG	QB	LN		
	CY	MJ	RZ					XN							VJ				AY			VY			
	IW	AJ													VY							BN			
																						IJ			
																						BF			
																						RN			
																						VD			
																						QB			
																						KF			
																						GF			
																						QJ			

Tablica 4.2: Trigrafi centrirani u drugoj abecedi

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
YH	WR		PB	BY	WE	CQ	RC	IE	CC		IC	WG	WB		SJ						VC	PM	VC	WC	BE
	IK		PK		LC	EX	DC		WK			DR	WF									KJ		WE	TE
	WR		DR		VW	IV			YJ				XF									BR		WI	EY
	CY		WY		XP	TY			CK				XF											TZ	KZ
	WI		XB		WZ	CX			PQ				AC											IZ	TA
	NR		FZ		WJ	DI			PE				XC											TE	EZ
			EC			CX			BJ				IB											BZ	RN
						LQ			TR															PH	DY
						BR			CR																
						DD			VR																
						BZ			PE																
						AD			WY																
						RQ																			
						VQ																			

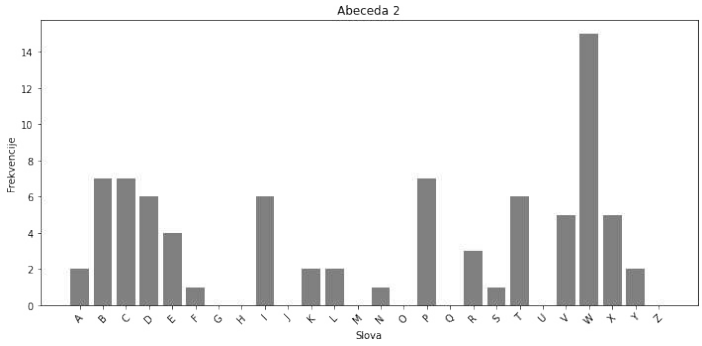
Tablica 4.3: Trigrafi centrirani u trećoj abecedi

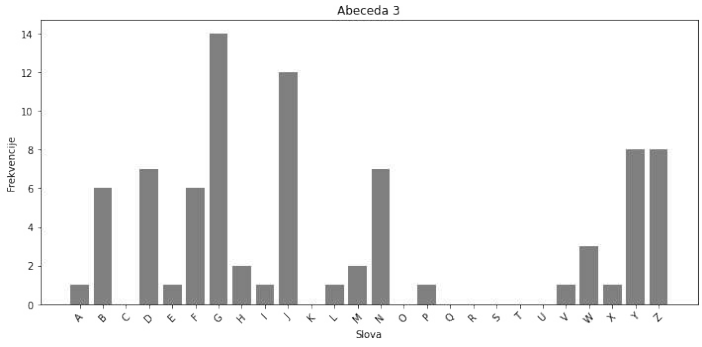
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ZO	NQ	YA	GG	ZY	NL	MW	AQ	YG	PL	BN		WR	ZQ		FU	GH	BI				GN	FY	GN	ZG	ZG
	DL	JI	GN	YU	NW		YL	GG	JY	JA						GQ	BI						GG	GO	YT
	DN	XU		ZI	NG			BI	JF	DA						JQ	MU						GG	BQ	ZG
	NA	FO		FQ					WQ	JX						GP	GS							DQ	DT
		HN		IW					FQ							GU	DU							EU	YQ
		ND		JL													JD							ZF	GN
		HA		JL													JR							JQ	YT
		LJ		YW													JN								FL
		DQ															BI								
		NL															WX								
		VS																							

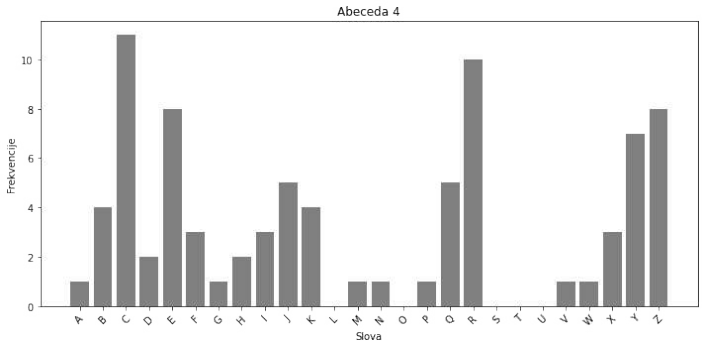
Tablica 4.4: Trigrafi centrirani u četvrtoj abecedi

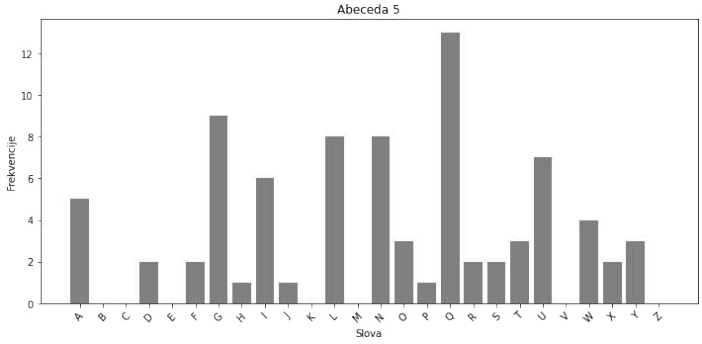
Sada pokušavamo metodom pokušaja i promašaja identificirati slova u abecedama šifrata pomoću analize frekvencija slova i trigrafa. Prisjetimo se da četiri samoglasnika `A, E, I, O` i četiri suglasnika `N, R, S, T` formiraju preko dvije trećine svih riječi engleskog jezika. S obzirom da je slovo `E` najfrekventnije slovo otvorenog teksta, na temelju distribucije frekvencija probamo zaključiti da vrijedi:

$$\mathtt{E_p} = \mathtt{I_c^1}, \mathtt{W_c^2}, \mathtt{G_c^3}, \mathtt{C_c^4}, \mathtt{Q_c^5}$$









Slova najveće frekvencije u abecedi 1 su: I, M, Q, V, B, G, K, R, S i C. Za slovo $\mathtt{I}_\mathtt{c}$ smo već pretpostavili da je $\mathtt{E}_\mathtt{p}$. Kako smo pretpostavili $\mathtt{W}_\mathtt{c}^2$ i $\mathtt{Q}_\mathtt{c}^5$, analizom frekvencija trigrafa možemo pretpostaviti koja od slova M, Q, V, B, G, K, R, S i C su samoglasnici, a koja suglasnici, gledajući prefikse od $\mathtt{W}_\mathtt{c}^2$ i sufikse od $\mathtt{Q}_\mathtt{c}^5$.

Q	V	L	B	I	R	K	G
5	3	1	2	1	1	1	1

Q	O	V	Z	R	A	L	X	I
3	1	2	1	2	1	1	1	1

Pogledajmo sada slovo M iz abecede 1, uočavamo da se ne pojavljuje kao prefiks niti kao sufiks slova Q i W iz abeceda 2 i 5. Stoga, možemo zaključiti da se radi o samoglasniku jer je frekvencija digrafa, samoglasnik + slovo I, jako niska. Kako imamo relativno visoku frekvenciju, možemo pretpostaviti da se radi o slovu $\mathtt{O}_\mathtt{p}$. Isto tako, primjećujemo da se slovo $\mathtt{Q}_\mathtt{c}^1$ pojavljuje pet puta kao prefiks i tri puta kao sufiks pa možemo zaključiti da se radi o frekventnom suglasniku na primjer slovu R. Analogno, slovo $\mathtt{V}_\mathtt{c}^1$ se pojavljuje tri puta kao prefiks i dva puta kao sufiks pa opet možemo pretpostaviti da je jednak frekventnom suglasniku $\mathtt{T}_\mathtt{p}$.
Slova $\mathtt{S}_\mathtt{c}^1$ i $\mathtt{C}_\mathtt{c}^1$ se ne pojavljuju kao prefiksi i sufiksi pa možemo zaključiti da se radi o samoglasnicima. Međutim, nakon promatranja sufiksa i prefiksa ta dva slova, sigurniji smo da je slovo $\mathtt{C}_\mathtt{c}^1$ samoglasnik, a ne slovo $\mathtt{S}_\mathtt{c}^1$. Prefiksi slova C su N, T i W iz abecede 5 za koja smo pretpostavili da su suglasnici, kao i sufiksi T, C i B iz abecede 2. Slovo $\mathtt{S}_\mathtt{c}^1$ za sada smatramo nedefiniranim.

Primijenimo princip direktne simetrije pozicija. Pogledajmo koja smo slova abeceda šifrata do sada pretpostavili:

Otvoren tekst	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Abeceda 1	C?				I				C?						M			Q		V						
Abeceda 2					W																					
Abeceda 3					G																					
Abeceda 4					C																					
Abeceda 5					Q																					

Ako pretpostavimo da je komponenta otvorenog teksta normalna, te da su abecede miješane, možemo koristiti princip direktne simetrije pozicija da dodatno ispunimo tablicu. Vidimo da prva i peta abeceda imaju vrijednosti za $\mathtt{Q}_\mathtt{c}$. To ćemo iskoristiti kako bismo ostala slova abecede 1 premjestili u abecedu 5, poštujući relativne pozicije slova. Tako dobijemo sljedeću tablicu:

Otvoren tekst	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Abeceda 1					I				C						M			Q		V						
Abeceda 2					W																					
Abeceda 3					G																					
Abeceda 4	I				C						M			Q		P										
Abeceda 5		M			Q		V											I				C				

Slova `M`, `V` i `I` smo mogli direktno upisati u abecedu 5. Iz toga smo mogli doći do dodatnih zaključaka. Kako su $\mathtt{B}_\mathtt{p}$ i $\mathtt{G}_\mathtt{p}$ niskofrektventna slova, trebalo bi vrijediti da su frekvencije slova $\mathtt{M}_\mathtt{c}$ i $\mathtt{V}_\mathtt{c}$ u abecedi 5 također niska. Kada pogledamo njihove frekvencije, vidimo da se ta dva slova uopće ne pojavljuju u klasi što odgovara našim dosadašnjim pretpostavkama. Isto tako, slovo $\mathtt{I}_\mathtt{c}^5 = \mathtt{R}_\mathtt{p}$ bi trebalo imati visoku frekvenciju, što je također točno. Do sada nismo bili sigurni na kojem bi mjestu u abecedi 1 trebalo biti slovo C. Poštujući relativnu poziciju slova C u odnosu na slovo `Q` iz prve abecede, mora vrijediti $\mathtt{C}_\mathtt{c}^5 = \mathtt{N}_\mathtt{p}$ ili $\mathtt{C}_\mathtt{c}^5 = \mathtt{V}_\mathtt{p}$. Kada bi `C` bilo jednako slovu `N` u abecedi 5, njegova frekvencija bi trebala biti visoka, ali se slovo `C` uopće ne pojavljuje zbog čega zaključujemo da je jednako slovu `V` u abecedi 5, a time je jednako slovu `I` u abecedi 1.

Nakon što smo pretpostavili vrijednosti nekoliko slova u abecedama, korisno je unijeti vrijednosti u kriptogram kako bismo mogli na temelju nekih pretpostavljenih riječi doći do dodatnih zaključaka. Pogledajmo kriptogram:

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

Poznavanje tipa teksta pomaže u ovom koraku. S obzirom da je ovaj primjer preuzet iz Friedmanove knjige *Military Cryptoanalysis* [3], radi se o vojnoj poruci pa možemo pretpostaviti sljedeće riječi iz kostura dešifriranog kriptograma: REPORT, ATTACK. Prvo uvrstimo slova u odgovarajuće abecede koje smo zaključili uvrštavanjem riječi REPORT i ATTACK. Kada uvrstimo slovo W u abecedu 3, tako da vrijedi $\mathtt{W}_c^3 = \mathtt{C}_p$ (zaključeno iz riječi ATTACK), dolazimo do kontradikcije. U abecedama 1 i 5, drugo mjesto lijevo od slova G je bilo slovo I, dok je sada u abecedi 3 drugo mjesto lijevo od slova G slovo W. Po principu direktne simetrije pozicija, slova moraju biti jednaka. Postavlja se pitanje koje ćemo odabrati.
Do zaključka da vrijedi $\mathtt{G}_c^3 = \mathtt{E}_p$ smo došli samo na temelju analize frekvencija. $\mathtt{G}_c^3$ se svakako ponaša kao samoglasnik, ali možda nije E. Pretpostavimo drugu vrijednost i neka je $\mathtt{E}_c^3 = \mathtt{O}_p$. S novim vrijednostima dobijemo sljedeću tablicu:

Otvoren tekst	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Abeceda 1			S		I		G	B	C						M		P	Q	R	V	W					
Abeceda 2	P	Q	R	V	W								S		I		G	B	C						M	
Abeceda 3	R	V	W								S		I		G	B	C						M		P	Q
Abeceda 4	I		G	B	C						M		P	Q	R	V	W								S	
Abeceda 5		M		P	Q	R	V	W								S		I		G	B	C				

Ponovno ćemo uvrstiti novodobivene vrijednosti u kriptogram koji je prikazan na sljedećoj stranici.

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

Nakon što smo dobili ovakav kostur otvorenog teksta, jednostavno je zaključiti pojedine riječi i time nadopuniti abecede šifrata u potpunosti kao što smo prije napravili.

Miješana abeceda je zadana ključem EXHAUSTING, a abecede su odabrane ključnom riječi APRIL. Slijedi prikaz potpuno ispunjene abecede šifrata i dešifrirani tekst.

Otvoren tekst	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Abeceda 1	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H
Abeceda 2	P	W	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O
Abeceda 3	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q
Abeceda 4	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T
Abeceda 5	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K

REPORTED ENEMY HAS RETIRED TO NEWCHESTER. ONE TROOP IS REPORTED AT HENDERSON MEETING HOUSE: TWO OTHER TROOPS IN ORCHARD AT SOUTHWEST EDGE OF NEWCHESTER. 2D SQ IS PREPARING TO ATTACK FROM THE SOUTH. ONE TROOP OF 3D SQ IS ENGAGING HOSTILE TROOP AT NEWCHESTER. REST OF 3D SQ IS MOVING TO ATTACK NEWCHESTER FROM THE NORTH. MOVE YOUR SQ INTO WOODS EAST OF CROSSROAD 539 AND BE PREPARED TO SUPPORT ATTACK OF 2D AND 3D SQ. DO NOT ADVANCE BEYOND NEWCHESTER MESSAGES HERE.

Poglavlje 5

Poboljšanja prethodnih sustava

U prethodna dva poglavlja opisali smo polialfabetske sustave koji koriste ponavljajući ključ za dobivanje kriptograma. Pokazali smo nekoliko načina za rješavanje takvih sustava, koji su se temeljili na pojavi cikličkih fenomena unutar kriptograma. U ovom poglavlju ćemo kratko opisati metode kojima prikrivamo ili u potpunosti izbjegavamo pojavu cikličkih fenomena te još jednu korisnu metodu za dešifriranje. Jedan od mogućih načina je korištenje aperiodičkih sustava šifriranja. Takvi sustavi se mogu dobiti na sljedeća dva načina:

1. koristeći ključ konstantne duljine za šifriranje grupa otvorenog teksta koje su varijable duljine,

2. koristeći ključeve varijable duljine za šifriranje grupa otvorenog teksta koje su konstantne duljine.

Neka je ključ k jednak UTORAK. U sljedećem primjeru šifriramo poruku prvim načinom, koristeći standardnu abecedu.

U	T	O	R	A	K	U	T	O	R	A	K	U
1	12	123	1234	12345	1	12	123	1234	12345	1	12	123
P	RI	PRE	MECE	SEODR	Z	AT	IUV	OJAR	NIUVI	N	KO	VCI
J	KB	DFS	DVTV	SEODR	J	UN	BNO	CXOF	EZLMZ	N	UY	PWC

R	A	K	U	T
1234	12345	1	12	1
MAID	UCITJ	E	DA	N
DRZU	UCITJ	O	NK	G

Od otvorenog teksta

```
PRIPREME CE SE ODRZATI U VOJARNI U VINKOVCIMA IDUCI TJEDAN
```

dobijemo sljedeći šifrat:

```
JKBDFSDV TV SE ODRJUNB N OCXOFEZ L MZNUYPWCDR ZUUCI TJONKG
```

U nastavku prikazujemo drugi način u kojem koristimo varijabilne duljine ključa. Neka se ponavljanje ključa prekine pojavom slova R u otvorenom tekstu. Dobijemo sljedeće šifriranje:

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

Ponavljanje ključa možemo prekinuti pojavljivanjem nekog slova u šifratu. Može se pokazati da asimetrični sustavi ne uspijevaju izbjeći pojave cikličkih fenomena u dovoljnoj mjeri da se oteža dešifriranje. Kako bismo dobili veću sigurnost, preporuča se korištenje alternativne metode koja će uspjeti izbjeći takve fenomene. Opisat ćemo nekoliko metoda kojima ćemo produžiti ključ šifriranja. Kada je dužina ključa dovoljno velika, kriptoanalitičaru ostaje ograničen broj perioda koje može analizirati. U slučajevima kada je ključ dulji od cijelog otvorenog teksta, ključ se koristi samo jednom. Najjednostavnija metoda produženja ključa je korištenje dugih fraza ili čitavih rečenica. U stvarnosti se to gotovo nikada ne koristi zbog nepraktičnosti.

U prvoj metodi ćemo od neke riječi ili kratke fraze dobiti numeričku vrijednost koja će nam poslužiti kao ključ u transformaciji normalne abecede. Rezultat takve transformacije nam može koristiti kao dugačak ključ.

U drugoj metodi opet uzimamo neku riječ. Numeričke vrijednosti slova te riječi će nam odrediti koji dio riječi će se nadodati na ključ.

Riječ	K	R	I	P	T	O	L	O	G
Brojevi	3	8	2	7	9	5	4	6	1

Produženi ključ	K	R	I	P	T	O	L	O	G	K	R	I	K	K	R	I	P	T	O	L	K	R	I	P
	T	O	K	T	I	P	T	O	L	O	K	R	I	P	T	K	R	K	R	I	P	T		

Vidimo da od riječi duljine n dobijemo ključ duljine $1 + 2 + 3 + ... + n = \frac{n(n+1)}{2}$.

Postoje brojne druge metode odabira duljeg ključa. Možemo odabrati recipročnu vrijednost nekog broja. Ako dobijemo dugačak niz znamenaka, one nam mogu poslužiti za prekid ponavljanja ključa. Pogledajmo decimalan zapis recipročnog zapisa broja 49, $\frac{1}{49} = 0.02040815$. Koristimo riječ `CHRISTMAS` za šifriranje. Prva znamenka decimalnog zapisa je 0 pa se abeceda slova `C` neće koristiti. Druga znamenka je 2 pa se abeceda `H` koristi za šifriranje prvih i drugih slova. Treća znamenka je opet 0 pa se abeceda `R` neće koristiti. Četvrta znamenka je 4 pa se abeceda `I` koristi za šifriranje trećih, četvrtih, petih i šestih slova, i tako dalje.

5.1 Metoda vjerojatne riječi

Sustavi u kojima se ključ nikada ne ponavlja zovemo "running-key" sustavi. Najčešći i najpraktičniji način odabira ključeva u takvim sustavima je otvoreni tekst neke unaprijed dogovorene knjige. Takvi sustavi su jako teški za dešifriranje. Metoda koju ćemo ovdje koristiti, a može se koristiti i u prethodnim sustavima za ubrzano rješavanje, je metoda vjerojatne riječi (engl. *probable-word method*). Ako znamo kojeg je tipa otvoreni tekst, možemo pretpostaviti pojavljivanje nekih riječi. Ukoliko je riječ o tekstu vojnog karaktera, riječi poput `VOJSKA, GENERAL, NAPAD, OBRANA, VOJNIK, TENK` i druge, češće će se pojavljivati. Isto tako, možemo pretpostaviti pojavljivanje nekih riječi koje se često pojavljuju u svim tekstovima, neovisno o njihovom karakteru. U engleskom jeziku to su riječi `THE, AND, THIS, THAT, WITH, THEY, HAVE, FROM` i druge. U hrvatskom jeziku neke od najčešćih riječi su `GODINA, HTJETI, IMATI, MOCI, DAN` i druge. Iako su neke kratke riječi poput `I, JE, U, SE` puno češće, koristimo dulje riječi kako bismo dobili veće dijelove kostura ključa i tako zaključili sadržaj teksta.

Pogledajmo sljedeći šifrat:

```
CVLGI RAEFR DNTEC KONMO RVBDC EISGL ZCJVO BGETU UUZPE PLKK
```

Pretpostavimo da je u njemu riječ `VOJARNA`. Redom pokušavamo odgonetnuti dio ključa tako što stavljamo riječ `VOJARNA` uz šifrat pa kada dobijemo razumljiv dio teksta ključa možemo zaključiti da smo napravili dobar odabir pozicije riječi `VOJARNA`.

Počinjemo s prvim slovom šifrata.

```
Pretpostavljeni dio ključa...  H  H  C  G  R  E  A  .  .  .  .  .  .  .  .
Šifrat.......................  C  V  L  G  I  R  A  E  F  R  D  N  T  E  C
Otvoreni tekst...............  V  O  J  A  R  N  A  .  .  .  .  .  .  .  .
```

Vidimo da prvi dio pretpostavljenog ključa HHCGREA nije smislena riječ pa povučemo riječ VOJARNA jedno mjesto u desno i nastavimo postupak. Nakon nekoliko pokušaja dolazimo do slova T.

```
Pretpostavljeni dio ključa... . . . . . T G O D I N I . . . . . .
Šifrat..................... Z K O L M M O U X D Z A I P D L V Y G V
Otvoreni tekst............. . . . . V O J A R N A . . . . . .
```

Pojavljuje se razumljiva riječ GODINI unutar pretpostavljenog ključa. Sljedeći korak u rješavanju je daljnje ispunjavanje kostura ključa. Ispred riječi GODINI nalazi se slovo T. U hrvatskom jeziku slovo T se često pojavljuje u idućim digrafima: ST, AT, ET, UT. Pretpostavimo da je riječ o digrafu ST.

```
Pretpostavljeni dio ključa... . . . . . S T G O D I N I . . . . . .
Šifrat..................... Z K O L M M O U X D Z A I P D L V Y G V
Otvoreni tekst............. . . . . U V O J A R N A . . . . . .
```

U otvorenom tekstu javlja se izraz U VOJARNA što nije po pravilima hrvatskog jezika. Možemo pretpostaviti da je izbor digrafa ST bio kriv pa pokušati s drugim ili da je oblik riječi VOJARNA zapravo VOJARNI.

```
Pretpostavljeni dio ključa... . . . . . S T G O D I N A . . . . . .
Šifrat..................... Z K O L M M O U X D Z A I P D L V Y G V
Otvoreni tekst............. . . . . U V O J A R N I . . . . . .
```

Nadopunimo izraz ...ST GODINA s najizglednijim izborom SEST GODINA pa dobijemo:

```
Pretpostavljeni dio ključa... . . . S E S T G O D I N A . . . . . .
Šifrat..................... Z K O L M M O U X D Z A I P D L V Y G V
Otvoreni tekst............. . . . T I U V O J A R N I . . . . . .
```

Kraj neke nepoznate riječi u otvorenom tekstu je ...TI što može upućivati na infinitiv nekog glagola. Ovaj postupak nastavljamo dok ne rekonstruiramo tekst. Djelujemo u oba smjera u nizu pretpostavljenog teksta i pretpostavljenog ključa. Primjer je šifriran početkom knjige *Mali Princ*, a otvoreni tekst je jednak prijašnjem

```
PRIPREME CE SE ODRZATI U VOJARNI U VINKOVCIMA IDUCI TJEDAN.
```

5.2 Jednokratna bilježnica

Jednokratna bilježnica (engl. *one-time pad*) je jedini poznati algoritam koji daje matematički dokazan neslomljiv šifrat ukoliko poštuje određene uvjete. Uvjeti jednokratne bilježnice su:

- duljina ključa je veća ili jednaka dužini otvorenog teksta,
- ključ je generiran na potpuno nasumičan način,
- operacije ključa i otvorenog teksta se računaju modulo 10 (brojevi), modulo 26 (slova) ili modulo 2 (binarno),
- svaki ključ se koristi točno jednom te se uništi nakon uporabe,
- postoje samo dvije kopije ključa, jedna je kod pošiljateljan i jedna kod primatelja.

Vjerojatnost da se dogodi događaj B, ako znamo da se već dogodio događaj A, zovemo *uvjetna vjerojatnost* i označavamo $P(B \mid A)$. Vrijedi sljedeća formula:

$$P(B \mid A) = \frac{P(A \cap B)}{P(B)}.$$

Neka su M, C, K konačni skupovi mogućih otvorenih tekstova, šifrata i ključeva. Izraz $P(C = c)$ označava vjerojatnost da je odabran šifrat c iz skupa šifrata C (analogno za skupove K i M). Kažemo da je sustav savršeno siguran ako vrijedi:

$$P(M = m \mid C = c) = P(M = m), \qquad \text{za sve } m \in M \text{ i } c \in C. \tag{5.1}$$

Teorem 1. Neka je N duljina otvorenog teksta i k duljina ključa takva da je $k \geq N$. Ako je ključ odabran na potpuno nasumičan način, tada je jednokratna bilježnica savršeno sigurna.

Dokaz. Kako se sav alfanumerički ASCII tekst kodira u niz nula i jedinica, bez smanjenja općenitosti, možemo pretpostaviti da su K, M i C skupovi binarnih nizova.
Šifrat se dobije primjenom XOR bit-operacije nad ključem i otvorenim tekstom. Neka su $m \in M$ i $k \in K$ nizovi nula i jedinica. Za rezultat šifriranja $c \in C$ vrijedi $c = m \otimes k$.

Moramo pokazati da vrijedi (5.1) za sve parove m i c. Ako je $|K| = k$, tada je vjerojatnost odabira nekog ključa jednaka $\frac{1}{k}$. Pokazat ćemo da je vjerojatnost bilo kojeg šifrata iz skupa C također $\frac{1}{k}$. Znamo da je svaki šifrat dobiven od jednog ključa, a odabiri ključeva su disjunktni događaji. Vrijedi:

$$P(C = c) = \sum_i P(C = c \cap K = i).$$

Odabir šifrata i ključa je nezavisan pa vrijedi:

$$\begin{aligned} P(C = c \cap K = i) &= P(M = c \otimes i \cap K = i) \\ &= P(M = c \otimes i) \cdot P(K = i) \\ &= P(M = c \otimes i) \cdot \frac{1}{k} \end{aligned}$$

Kako idemo po svim ključevima iz skupa K dobijemo:

$$\sum_i P(M = c \otimes i) = P(M = \text{neka poruka } m) = 1.$$

$$P(C = c) = \sum_i P(C = c \cap K = i) = \sum_i P(M = c \otimes i)\frac{1}{k} = \frac{1}{k}.$$

Zbog nezavisnosti K i M dobijemo:

$$\begin{aligned} P(M = m \mid C = c)P(C = c) &= P(C = c \cap M = m) \\ &= P(K = c \otimes m \cap M = m) \\ &= P(K = c \otimes m)P(M = m) \end{aligned}$$

Kako su $P(C = c) = P(K = c \otimes m) = \frac{1}{k}$, možemo sve pomnožiti s k tako da dobijemo traženo svojstvo. □

Bibliografija

[1] Andrej Dujella i Marcel Maretić, *Kriptografija*, Element, Zagreb, 2007.

[2] William F. Friedman, *Military Cryptanalysis Part I*, Aegean park press, California, 1980.

[3] _____, *Military Cryptanalysis Part II*, Aegean park press, California, 1984.

[4] _____, *Military Cryptanalysis Part III*, Aegean park press, California, 1992.

[5] _____, *The Index of Coincidence And Its Applications In Cryptoanalysis*, (2014), https://www.nsa.gov/portals/75/documents/news-features/declassified-documents/friedman-documents/publications/folder_233/41761039080018.pdf.

Sažetak

Cilj ovog rada bio je predstaviti glavne metode dešifriranja šifrata polialfabetskih sustava koristeći statistička svojstva jezika.

Na početku rada definirani su osnovni pojmovi s kojima se susrećemo te su opisane glavne metode šifriranja klasične kriptografije.

Prije analize polialfabetskih sustava, provedena je frekvencijska analiza općenitog teksta. Rezultati analize daju nam informacije koje možemo koristiti u svim sustavima šifriranja. Koriste nam u otkrivanju jezika, određivanju sustava, rekonstrukciji ključa i rekonstrukciji otvorenog teksta. Definirana je standardna, odnosno normalna distribucija frekvencije slova. Pokazan je utjecaj veličine teksta na statističku analizu.

Prilikom analize polialfabetskih sustava ponavljajućeg ključa sa standardnom abecedom šifrata, opisane su dvije glavne metode rješavanja. To su Kasiskijeva metoda, koja se oslanja na faktorizaciju i Friedmanov indeks koincidencije. Obje metode koriste ciklički fenomen koji se manifestira u šifratima takvih sustava.

Nadalje, u sustavu s miješanom abecedom šifrata opisana je metoda koja se većinom oslanja na frekvencijsku analizu digrafa, trigrafa i drugih, te na principu direktne simetrije pozicija. Primjene metoda su ilustrirane na primjerima koji su većinom bili vojnog karaktera.

Na kraju su opisani aperiodični sustavi i metode kojima možemo produžiti ključ šifriranja. Kod takvih sustava se izbjegava pojava cikličkih fenomena koji su olakšali rješavanje prethodnih sustava. Pokazano je da korištenje dužih ključeva osigurava sigurniji šifrat. Opisana je i metoda vjerojatne riječi, s kojom smo riješili takav kriptogram.

Za sustav jednokratne bilježnice smo dokazali da je savršeno siguran sustav, ako poštuje određene uvjete.

Summary

The prime objective of this thesis is to give an overview of the main methods used to decipher polyalphabetic ciphers, using statistical language properties.

At the beginning, the basic concepts encountered are defined and the main methods of encryption of classical cryptography are described.

Before analyzing polyalphabetic systems, a frequency analysis of a general text was carried out. The results of the analysis give us information which we can use in every encryption system. They are useful in determining the language used, the general system, in reconstruction of the key and the reconstruction of the plain text. Standard or normal distribution of frequencies of letters was also defined. The impact of text volume on the statistical analysis was shown.

While analysing polyalphabetic systems with repeating keys and standard cipher alphabets two methods were described. The Kasiski method, which uses factorization and Friedman's index of coincidence. Both methods use cyclic phenomena which manifests in these systems.

Furthermore, in systems with mixed cipher alphabets, a method is described which mostly relies on frequency analysis of digraphs, trigraphs and others, and on the principal of direct symmetry of positions. The applications of the methods are illustrated using examples which were mostly of a military nature.

At the end, aperiodic systems and methods with which we can extend encryption keys are described. Such systems avoid the occurrences of cyclical phenomena that facilitated the solutions of previous systems. Using longer keys has been shown to provide a more secure cipher. The probable word method, with which we solved such a cryptogram, is also described.

For the one-time pad system, we proved that it is perfectly secure system, if it respects certain conditions.

Životopis

Rođen sam 23. studenoga 1998. godine u Vinkovcima. Pohađao sam Gimnaziju Matije Antuna Reljkovića. Tijekom školovanja sudjelovao sam na županijskim i državnim natjecanjima iz matematike. Nakon završetka srednje škole, 2017. godine upisujem Preddiplomski sveučilišni studij Matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu. Tijekom studiranja bio sam demonstrator iz kolegija Matematička analiza 1 i 2. Diplomski studij Računarstva i matematike upisujem 2020. godine na istom fakultetu.