Kasiskiho metóda:

Táto metóda sa spolieha na systematičnosť národných jazykov. V národných jazykoch sa opakujú nielen písmená ale aj skupiny písmen a celé slová. Angličtina napríklad veľmi často používa koncovky: -tk, -ing, -edd, -ion, -tion, -ation; predpony: im- , in- , un- , re- ; a špecifické štruktúry: -eek-, -oot,, -our- , atď.. Často sa v tomto jazyku vyskytujú i krátke slová ako: fo, and, to, with, are, is, a that.

Kasiskiho metóda sa riadi týmito pravidlami:
ak je správa šifrovaná n abecedami s cyklickou rotáciou a ak sa v otvorenom texte tejto správy vyskytuje určité slovo alebo skupina písmen k-krát, potom takáto skupina písmen by mala byť šifrovaná rovnakou abecedou približne k/n krát. Kasiskiho metóda sa venuje duplicitným fragmentom zašifrovaného textu. Aby mohla byť opakovane vyskytujúca skupina znakov otvoreného textu zašifrovaná rovnakým spôsobom dvakrát, tak kľúč by musel mať medzi týmito skupinami celý násobok rotácií a skončiť v rovnakom bode.

Kasiskiho metódu realizujeme v týchto krokoch:

- vyhľadáme opakujúce sa štruktúry o troch a viac znakoch
- pre každú štruktúru určíme "súradnice" jej počiatočného bodu
- zistíme vzdialenosť počiatočných bodov susedných štruktúr (diferencia)
- určíme všetky delitele každej diferencie
- pokiaľ bola pužitá polyalfabetická šifra, potom dĺžku kľúča budú určovať delitele (kroky), ktoré sa vyskytujú najčastejšie

Ku pochopeniu vyššie uvedených faktov nám pomôže nasledujúci príklad šifrovania časti Dickensovej poviedky, v ktorej sa často opakuje skupina: it was the ... Pre výber šifrovanej abecedy použijeme kľúčové slovo dickens:

dicke nsdic kensd icken sdick ensdi ckens dicke
ITWAS THEBE STOFT IMESI TWAST HEWOR STOFT IMESI

nsdic kensd icken sdick ensdi ckens dicke nsdic
TWAST HEAGE OFWIS DOMIT WASTH EAGEO FFOOL ISHNE

kensd icken sdick ensdi ckens dicke nsdic kensd
SSITW ASTHE EPOCH OFBEL IEFIT WASTH EEPOC HOFIN

Zašifrovaný text:

LBYKW GZHJG CXBXW QOOWV LZIUD LRORZ UDSSL LUGCM
GODAV RINYH WHGMF VRUKD ANKWP GKKRG INQYP VKKVG
CWVLZ IUDLR WSWER SSTHT KOJVL ZIUDL RWSWE RSSAQ

Pri Kasiskiho metóde postupujeme tak, že najprv vyhľadáme všetky opakujúce štruktúry zašifrovaného textu. Krátke opakujúce štruktúry k akým patrí napríklad dvojica písmen, sú často náhodné, takže viac problémov by spôsobilo ich akceptovanie ako zanedbanie. Ľubovoľná štruktúra dlhšia než tri znaky takmer určite nebude náhodná.

V každej opakovanej štruktúre (podčiarknutý text) vyznačíme jej počiatočný bod, potom určíme vzdialenosti medzi týmito počiatočnými bodmi:

LBYKW GZHJG CXBXW QOOWV LZIUD LRORZ UDSSL LUGCM
GODAV RINYH WHGMF VRUKD ANKWP GKKRG INQYP VKKVG
CWVLZ IUDLR WSWER SSTHT KOJVL ZIUDL RWSWE RSSAQ

Vzdialenosť počiatočného bodu štruktúry od začiatku šifrovaného textu	Vzdialenosť počiatočného bodu štruktúry od počiatočného bodu predchodzej štruktúry	Delitelia rozdielu vzdialenosti susedných štruktúr
20	–	–
83	63 => (83 - 20)	3, 7, 9, 21, 63
104	21 => (104 - 83)	3, 7, 21

Dĺžka kľúča bude teda pravdepodobne 3 alebo 7. Pokiaľ by bol počet opakovacích štruktúr väčší, bol by aj odhad pravdepodobnej dĺžky kľúča presnejší. V našom prípade budeme preto testovať obidve možné dĺžky kľúča, dĺžku 3 a 7.

Za predpokladu, že dĺžka kľúča môže mať 3 alebo 7 znakov, bude ďalší krok spočívať v rozdelení správy na podmnožiny, ktoré by mohli byť šifrované rovnakou abecedou. Táto miera nám pomôže určiť, či vzorka šifrovaného textu bola šifrovaná len jednou substitúciou (monoalfabetickou), dvoma, alebo viacerými. Takouto mierou je **index koincidencie**. Obecný zápis týchto podmnožín bude vyzerať:

pre dĺžku kľúča 3 :

$S_1 = \{ c_1, c_4, c_7, c_{10}, ... \}, S_2 = \{ c_2, c_5, c_8, c_{11}, ...\}, S_3 = \{ c_3, c_6, c_9, c_{12,} ...\}$

pre dĺžku kľúča 7 :

$S_1 = \{ c_1, c_8, c_{15}, c_{22}, ... \}, S_2 = \{ c_2, c_9, c_{16}, c_{23}, ...\}, ... , S_7 = \{ c_7, c_{14}, c_{21}, c_{28,} ...\}$

Ak všetky znaky v jednej z týchto podmnožín budú šifrované rovnakou abecedou, potom ich frekvenčná distribúcia by sa mala podobať frekvenčnej distribúcii národného jazyka správy otvoreného textu (v našom prípade angličtiny). Predpokladajme, že máme k dispozícii text, o ktorom si myslíme, že bol zašifrofaný monoalfabetickou substitúciou. Ak sa nemílime potom množstvo šifrovaného textu bude rovnaké ako množstvo príslušných písmen národného jazyka. Index koincidencie je potom mierou rozptylu množstva v distribúciách.

Index koincidencie " IC " pre daný text sa vypočíta nasledujúcim postupom:

- po rozdelení textu do skupín si v každej skupine spočítame pre každé písmeno zvlášť jeho množstvo výskytu "**Freq**" v danej skupine
- určíme dĺžku "**n**" každej skupiny

Potom IC sa vypočíta podľa vsťahu:

$$IC = \sum_{i=a}^{i=z} \frac{Freq_i * (Freq_i - 1)}{n * (n - 1)}$$

Naše skupiny a ich IC:

pre kľúč dĺžky 3 :

S_1 = [LKZGBQWZDOUSUMDRYHFUAWKGQVVWZDWESTJZDWES] $IC_1 = 0,0423$

S_2 = [BWHCXOVILRDLGGAIHGVKNPKIYKGVILSRTKVILSRA] $IC_1 = 0,0512$

S_3 = [YGJXWOLURZSLCOVNWMRDKGRNPKCLURWSHOLURWSQ] $IC_1 = 0,0449$

Vzorovo:

$$IC_1 = 2.\frac{4.(4-1)}{18.(18-1)} + 5.\frac{2.(2-1)}{18.(18-1)} + 2.\frac{3.(3-1)}{18.(18-1)} + \frac{5.(5-1)}{18.(18-1)} = 0,0423$$

pre kľúč dĺžky 7 :

S_1 = [LHWZRLDHRWIKZSHZSQ] $IC_1 = 0,0654$

S_2 = [BJQIZUAWUPNVIWTIW] $IC_2 = 0,0515$

S_3 = [YGOUUGVHKGQGUEKUE] IC_3 = 0,1029
S_4 = [KCODDCRGDKYCDRODR] IC_4 = 0,1324
S_5 = [WXWLSMIMAKPWLSJLS] IC_5 = 0,0735
S_6 = [GBVRSGNFNRVVRSVRS] IC_6 = 0,125
S_7 = [ZXLOLOYVKGKLWTLWA] IC_7 = 0,0662

Vzorovo:

$$IC_1 = \frac{3.(3-1)}{18.(18-1)} + \frac{2.(2-1)}{18.(18-1)} + \frac{2.(2-1)}{18.(18-1)} + \frac{2.(2-1)}{18.(18-1)} + \frac{2.(2-1)}{18.(18-1)} + \frac{3.(3-1)}{18.(18-1)} = 0{,}0654$$

Pokiaľ zvolíme správnu dĺžku kľúča, tak všetky hodnoty IC sa budú blížiť k číslu 0,068.
V našom príklade aj pri určení správnej dĺžky kľúča 7 sa nám IC nerovnajú danému číslu 0,068 z dôvodu, že text na šifrovanie je veľmi krátky. Ale už aj z tohto krátkeho textu vidieť, že IC správnej dĺžky je v priemeru rádovo vyššie ako IC pri nesprávnej dĺžke kľúča. Naš príklad slúžil iba ako ukážka postupu dešifrovania zašifrovaného textu.

Záverečná poznámka k polyalfabetickým šifrám:

Postup analýzy polyalfabetických šifier je tento:

- Použitím Kasiskiho metódy odhadneme pravdepodobný počet šifrovacích abecied. Ak analýza neposkytne výrazne určité číslo, potom šifrovanie nebude pravdepodobne realizované jednoduchou polyalfabetickou substitúciou.
- Ku overeniu odhadu získaného prvým krokom vypočítame index koincidencie IC.
- Pre sľubnú hodnotu získanú prvým a druhým krokom rozdelíme šifrovaný text na príslušné podmnožiny a pre každú podmnožinu nezávisle vypočítame index koincidencie.

Tieto podmienky platia pre dostatočne dlhý zašifrovaný text.

Kasiskiho metóda analýzy polyalfabetických šifier predpokladá, že v aplikácií abecied existuje určitá periodická zákonitosť. Táto metóda sa pokúša periodickú zákonitosť odhaliť analýzou opakovane sa vyskytujúcich štruktúr zašifrovaného textu, ktorých vzdialenosti nie sú v rozpore s celistvým násobkom možnej dĺžky použitého kľúča.

Index koincidencie sa dá použiť dvoma spôsobmi. Poprvé môže poslúžiť k potvrdeniu odhadu, že daná vzorka šifrovaného textu bola zašifrovaná polyalfabetickou šifrou a za druhé pre daný odhad počtu použitých abecied a zistenia množstva výskytu v častiach textu pravdepodobne šifrovaných týmito abecedami môže potvrdiť súhlas frekvenčnej distribúcii s distribúciou štandardného národného jazyka (v uvažovanom prípade angličtina).

Úspešnosť Kasiskiho metódy a indexu koincidencie závisí na množstve šifrovaného textu, ktorý je k dispozícii. Oba spôsoby sa dobre uplatňujú vtedy, pokiaľ opakované používanie šifrovacích abecied bude prebiehať v periodických intervaloch.