Bezpieczeństwo systemów i sieci komputerowych

dr inż. Mirosław Mazurek

Zakład Systemów Złożonych
Bud. F, pok. 305, tel. 17 865 11 04

Podstawowe zasady kryptografii

1. **Nie wolno lekceważyć przeciwnika.**
2. **Nie wolno opierać bezpieczeństwa systemu kryptograficznego na tajności algorytmu.**
3. **Nie należy poprawiać skutecznie działających systemów kryptograficznych.**
4. **Należy zawsze brać pod uwagę możliwość popełnienia błędów kryptograficznych lub naruszenia dyscypliny szyfrowania.**
5. **Prawidłowe stosowanie systemu kryptograficznego nie zwalnia ze stosowania innych zabezpieczeń.**

Podstawowe błędy szyfrowania

1. **Równoczesna transmisja tekstu tajnego i jawnego.**
2. **Użycie tego samego klucza do zaszyfrowania dwóch różnych tekstów jawnych.**
3. **Użycie dwóch różnych kluczy do zaszyfrowania tego samego tekstu jawnego.**
4. **Powtarzanie stereotypowych zwrotów w szyfrowanym tekście.**
5. **Używanie prostych haseł i kluczy.**
6. **Używanie krótkich haseł i kluczy.**
7. **Używanie znaków interpunkcyjnych.**

Częstotliwość spacji jest co najmniej dwukrotnie większa od częstotliwości występowania najczęstszej samogłoski

8. **Poprawność stylistyczna i ortograficzna.**

Kryptografia - metody kryptograficzne

Klasyczne:

– przeszukiwanie całej przestrzeni klucza (ang. brutal force)

- Przeszukiwanie zredukowanej przestrzeni klucza (atak słownikowy)
- Bazujące na statystyce
- Bazujące na negatywnym wzorcu

Nowoczesne:

– metoda różnicowa

- Kluczy powiązanych
- Metoda liniowa

Metoda przeszukiwania całej przestrzeni klucza

Zastosowanie:

- Dla małych przestrzeni klucza
- Skuteczna dla ataku z tekstem jawnym

Długość hasła	Tylko litery [26]	Duże i małe litery [52]	Znaki z klawiatury [95]
1	26	52	95
2	676	2704	9 tys.
3	17,5 tys.	140 tys.	857,4 tys.
4	457 tys.	7,3 mln	81,5 mln
5	11,9 mln	380,2 mln	7,7 mld
6	308 mln	19,8 mld	735,1 mld

Teoria informacji

Długość krytyczna kodu - jest to najmniejsza długość tekstu zaszyfrowanego liczona w znakach, która jest niezbędna do jednoznacznego określenia klucza.

Uwaga: posiadanie tej ilości informacji nie gwarantuje złamanie kodu

Dla dyskretnego źródła informacji miarą ilości informacji w wiadomości jest przeciętna liczba bitów niezbędna do zakodowania wszystkich informacji.

Miarą ilości informacji w wiadomości jest entropia tej wiadomości, mierząca nieokreśloność lub nieprzewidywalność informacji. Im większa entropia, tym większa jest ilość informacji zawarta w wiadomości.

Zerowa entropia oznacza, że wiadomość nie niesie żadnej wiadomości.

Teoria informacji

Entropia – w ramach teorii informacji jest definiowana jako średnia ilość informacji, przypadająca na pojedynczą wiadomość ze źródła informacji. Innymi słowy jest to średnia ważona ilości informacji niesionej przez pojedynczą wiadomość, gdzie wagami są prawdopodobieństwa nadania poszczególnych wiadomości.

Źródło: http://pl.wikipedia.org/wiki/Entropia_%28teoria_informacji%29

$$H(x) = \sum_{i=1}^{n} p(x_i) \log_r \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i) \log_r p(x_i)$$

Gdzie:
p(i) – prawdopodobieństwo zajścia zdarzenia *i*,
n – liczba wszystkich zdarzeń danej przestrzeni.

W przypadku kodowania ciągu znaków jest to prawdopodobieństwo wystąpienia *i*-tego znaku. W teorii informacji najczęściej stosuje się logarytm o podstawie *r*=2, wówczas jednostką entropii jest bit.

Teoria informacji

Wskaźnik bezwzględny języka R - określa maksymalną liczbę bitów niezbędną do przedstawienia informacji, która mogłaby być zakodowana w dowolnym znaku, przy założeniu, że wszystkie możliwe sekwencje znaków są jednakowo prawdopodobne

$$R = log_2 L$$

gdzie:

L – ilość liter w alfabecie

Wskaźnik względny języka r - określa przeciętną liczbę bitów na jeden znak informacji

$$r = \frac{H(X)}{N log_2 L}$$

gdzie:

N – ilość znaków w wiadomości

Koincydencja znaków

Zastosowanie:

- Dla dwóch dowolnych tekstów jawnych o równej długości można wyznaczyć współczynnik koincydencji

- Wskaźnik koincydencji określa prawdopodobieństwo, że dwie litery wybrane losowo z danego kryptogramu będą identyczne

$$IC = \frac{\sum_{i=0}^{L-1} F_i(F_i - 1)}{N(N-1)}$$

Fi – częstość wystąpienia *i*-tego znaku w szyfrogramie
N – *długość szyfrogramu*
L – długość alfabetu

Koincydencja znaków

Okresy i odpowiadające im wskaźniki koincydencji dla j. angielskiego

d	IC
1	0,0660
2	0,0520
3	0,0473
4	0,0450
5	0,0436
6	0,0427
7	0,0420
8	0,0415
9	0,0411
10	0,0408

d	IC
11	0,0405
12	0,0403
13	0,0402
14	0,0400
15	0,0399
16	0,0397
17	0,0396
18	0,0396
19	0,0395
20	0,0394

Koincydencja znaków

Języki i odpowiadające im wskaźniki koincydencji

Język	IC
Angielski	0,066895
Arabski	0,075889
Duński	0,070731
Fiński	0,073796
Francuski	0,074604
Grecki	0,069165
Hebrajski	0,076844
Hiszpański	0,076613
Holenderski	0,079805
Japoński	0,077236

Język	IC
Malajski	0,085286
Niemiecki	0,076667
Norweski	0,069428
Portugalski	0,074528
Rosyjski	0,056074
Serbsko-chorwacki	0,064363
Szwedzki	0,064489
Włoski	0,073294

Test Kasiskiego

Analiza powtórzeń znaków w kryptogramie sugeruje długość klucza.

Jeśli dwa takie same ciągi znaków znajdują się w tekście jawnym w odstępie wynoszącym wielokrotność okresu klucza, to na odpowiednich miejscach w kryptogramie uzyskuje się identyczne grupy znaków.

Przykłady powtórzeń: …enie, …anie, …ówki,

Metoda Kasiskiego polega na wyszukaniu powtórzeń bloków co najmniej trzech liter w szyfrogramie i obliczeniu odstępu między nimi.

24, 54, 18, 29, 66

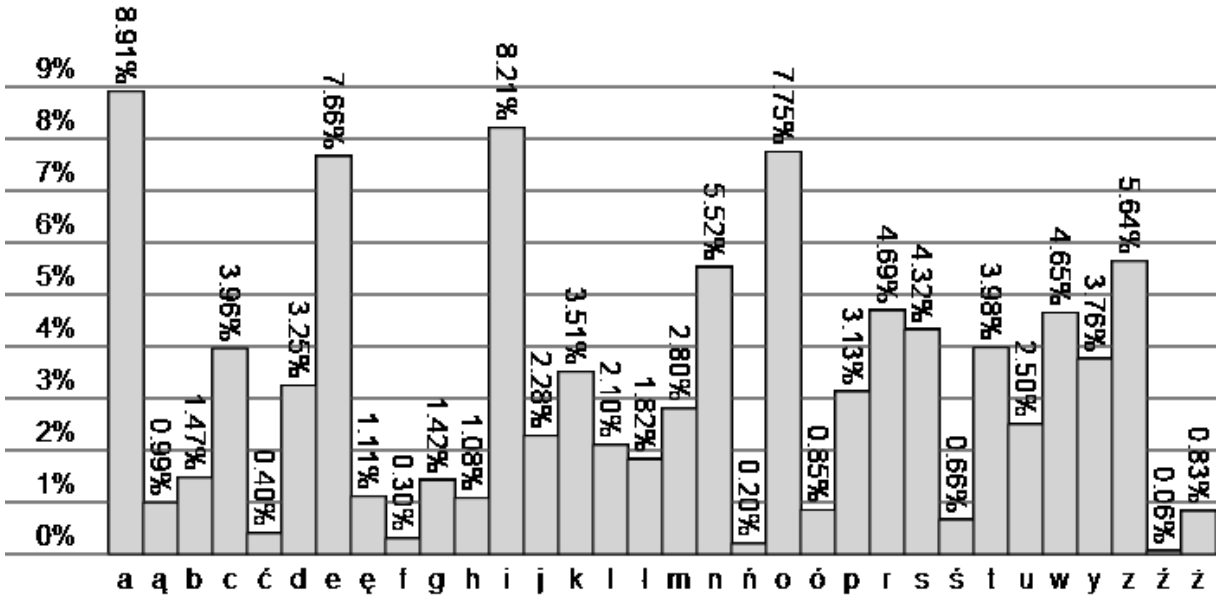
Możliwy okres klucza: **.....**

Test Kasiskiego

Analiza powtórzeń znaków w kryptogramie sugeruje długość klucza.

Jeśli dwa takie same ciągi znaków znajdują się w tekście jawnym w odstępie wynoszącym wielokrotność okresu klucza, to na odpowiednich miejscach w kryptogramie uzyskuje się identyczne grupy znaków.

Metoda Kasiskiego polega na wyszukaniu powtórzeń bloków co najmniej trzech liter w szyfrogramie i obliczeniu odstępu między nimi.



Zbitki literowe

Zbitki dwuliterowe – j.angielski

TH	10,00
HE	9,50
IN	7,17
ER	6,65
RE	5,92
ON	5,70
AN	5,63
EN	4,76
AT	4,72
ES	4,24

ED	4,12
TE	4,12
TI	4,00
OR	3,98
ST	3,81
AR	3,54
ND	3,52
TO	3,50
NT	3,44
IS	3,43

OF	3,38
IT	3,26
AL	3,15
AS	3,00
HA	3,00
NG	2,92
CO	2,80
SE	2,75
ME	2,65
DE	2,65

Zbitki literowe

Zbitki trzyliterowe

THE	10,00
AND	2,81
TIO	2,24
ATI	1,67
FOR	1,65
THA	1,49
TER	1,35
RES	1,26
ERE	1,24
CON	1,20
TED	1,09
COM	1,08

Teoria informacji

Analiza powtórzeń znaków w kryptogramie sugeruje długość klucza.

Jeśli dwa takie same ciągi znaków znajdują się w tekście jawnym w odstępie wynoszącym wielokrotność okresu klucza, to na odpowiednich miejscach w kryptogramie uzyskuje się identyczne grupy znaków.

Przykłady powtórzeń: …enie, …anie, …ówki,

Metoda Kasiskiego polega na wyszukaniu powtórzeń bloków co najmniej trzech liter w szyfrogramie i obliczeniu odstępu między nimi.

24, 54, 18, 29, 66

Możliwy okres klucza: **.....**

Klasyfikacja algorytmów

O(1) – algorytm stały o złożoności niezależnej od długości słowa wejściowego

O(n) – algorytm liniowy o złożoności wprost proporcjonalnej do długości słowa wejściowego

$O(n^t)$ – algorytm wielomianowy - złożoność zależy od n^t przy stałym t

$O(t^{f(n)})$ – algorytm wykładniczy – t – stałe, a funkcja $f(n)$ jest wielomianem zmiennej n

Złożoność obliczeniowa algorytmów

Klasa	Złożoność	Liczba operacji dla $n=10^6$	Czas wykonania przy 10^6 operacji/s
Stały	$O(1)$	1	1 μs
Liniowy	$O(n)$	10^6	1 s
Kwadratowy	$O(n^2)$	10^{12}	11,6 dnia
Sześcienny	$O(n^3)$	10^{18}	32000 lat
Wykładniczy	$O(2^n)$	10^{301030}	10^{301030}x wiek istnienia wszechświata