Kryptografia – Laboratorium 1
Szyfrowanie Podstawieniowe

Wstęp

Na początek kilka definicji i pojęć:

Słownik, zbiór znaków

Słownik tekstu jawnego to zbiór **V** znaków używanych do formułowania tekstu jawnego. Słownik tekstu zaszyfrowanego to zbiór **W** znaków służących do zapisu tekstu zaszyfrowanego, czyli kryptogramu (tekstu kodowego). Zbiór **W** może zawierać jakiekolwiek znaki, zbiory **V** oraz **W** mogą nie mieć elementów wspólnych (zbory rozłączne), mogą mieć cześć wspólną lub być identyczne. Z sytuacją ostania mamy do czynienia gdy kodujemy tekst zapisany zwykłym alfabetem łacińskim znakami tego samego alfabetu.

np. **V** = {a,b,c,d,....,x,y,z} oraz **W**=**V**

Słowo to ciąg znaków danego alfabetu. Zbiór słów danej długości n oznaczamy V^n. Zbiór słów długości mniejszej niż n oznaczamy $V^{(n)}$. Zatem $V^{(n)} = \{null\} \cup V \cup V^2 \cup V^3 ... \cup V^n$, gdzie null oznacza znak pusty. V^* oznacza zbiór wszystkich słów utworzonych z liter danego alfabetu.

Szyfrowanie jest definiowane jako relacja X: $V^* \rightarrow W^*$. Deszyfrowanie jest zdefiniowane jako relacja X^{-1}: $V^* \leftarrow W^*$ przy założeniu, że $x \leftarrow y$ wtedy tylko gdy $x \rightarrow y$. Odbiorca wiadomości powinien być w stanie jednoznacznie odszyfrować przesyłaną wiadomość, dlatego też szyfrowanie powinno być różnowartościowe, tzn. jednoznaczne od strony prawej do lewej, tzn. :

$$(x \rightarrow z) \wedge (y \rightarrow z) \Rightarrow (x=y)$$

Oznacza to, że taki sam kryptogram dla dwóch wiadomości można dostać tylko wtedy gdy wiadomości te są identyczne.

System szyfrowania M to niepusty zbiór $\{x_0, x_1, x_2, x_3, ..., x_{k-1}\}$ różnowartościowych relacji $x_i: V^{(ni)} \rightarrow W^{(mi)}$. Każdą z takich relacji nazywamy krokiem szyfrowania. System szyfrowania wraz z odpowiadającym mu systemem deszyfrowania tworzą system kryptograficzny (krypto system).
Szyfrowania określa się jako ciąg kroków szyfrowania z danego systemu szyfrowania.

Przykład: Dany jest system szyfrowania M = $\{ x_0, x_1\}$ gdzie x_0 to cykliczna transpozycja 3 kolejnych elementów tekstu jawnego, a x_1 to podstawienie 2 kolejnych znaków zdefiniowane następującym przyporządkowaniem:

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

Wtedy tekst: *kryptografia* zaszyfrować można szyfrowaniem **X** wygenerowanym z systemu szyfrowania **M** i zdefiniowanym jako ciąg $(x_0, x_1, x_0, x_1, x_0, ...)$

kry pt ogr af ia
ykr qu rog bg ai

Krok szyfrowania nazywamy endomorficznym jeśli **V**=**W**.
Szyfrowanie X nazywamy manoalfabetycznym jeśli wykorzystuje tylko jeden krok szyfrowania danego systemu szyfrowania M. W przeciwnym wypadku jest szyfrowaniem poliafabetycznym.

Szyfrowanie nazywamy monograficznym jeśli wszystkie n_i z użytych kroków szyfrowania są równe 1. W przeciwnym wypadku mówimy o szyfrowaniu poligraficznym. W szczególności przy wykorzystaniu maszyn do szyfrowania korzystnie jest w praktyce zachować stałą szerokość szyfrowania n oraz stałą szerokość szyfru m.

x_i: $V^n \rightarrow W^m$

Dla n = 2,3,4 mamy szyfrowanie odpowiednio dwuznakowe, trójznakowe, oraz trójznakowe, a dla m = 1,2,3 jednodzielne. dwudzielne, trójdzielne. Szyfrowanie takie nazywamy ogólnie szyfrowaniem blokowym.

Podstawienie proste:

Wśród kroków szyfrowania można wyróżnić dwie duże klasy: podstawienia oraz transpozycje, jak to zostało pokazane w przykładzie powyżej. Są one przypadkami ogólnego kroku szyfrowania

$$V^n \rightarrow W^m$$

W przypadku manoalfabetycznym ustala się jedno konkretne podstawienie (lub transpozycje) i szyfrowanie przebiega cały czas z wykorzystaniem tego samego kroku szyfrowania.

Zajmiemy się teraz dokładniej podstawieniem prostym a dokładniej rzecz biorąc podstawieniem manoalfabetycznym i monograficznym określonym ogólnie:

$$x: V \rightarrow W$$

Dodatkowo założymy ze V=W. Oznacza to ze zarówno tekst jawny jak i tekst zaszyfrowany zapisywane są z wykorzystaniem tego samego alfabetu. Tradycyjnie w szyfrowaniu pomija się znaki odstępu oraz interpunkcyjne. Załóżmy, że naszym alfabetem będzie alfabet łaciński:

abcdefghijklmnopqrstuvwxyz

mający 26 znaków. Przykładem definicji takiego podstawienia może być wypisanie odpowiednich przyporządkowań, np.:

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

Wynika z tego, że literze *a* tekstu jawnego odpowiadać będzie litera *b* w szyfrogramie, literze *b* litera *c* itd. Jak widać podstawienie to powstało po prostu z przesunięcia każdej litery o jedną pozycję w lewo, cyklicznie (*a* pojawiło się na końcu). Jako ze znak kryptogramu nie powinien odpowiadać dwóm rożnym znakom tekstu jawnego tworzony przez nas alfabet jest po prostu permutacją alfabetu podstawowego.

Oprócz notacji podstawień do zapisu permutacji stasuje się również zapis cykliczny. Przykładowo dla

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

Można zastosować zapis cykliczny:

(arktuhfyxzlb)(cw)(den)(g)(imsojpqv)

Podstawienia generować można na wiele sposobów. Znaną metodą mnemotechniczną jest generowanie podstawień na podstawie haseł. Przykładowo, utwórzmy podstawienie na podstawie hasła: SECURITY

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

Zapisujemy najpierw hasło a następnie dopisujemy te litery w kolejności alfabetycznej, które nie wystąpiły a haśle. W przypadku gdy w haśle powtarzają się litery, kolejne wystąpienia tej samej litery są pomijane, gdyż prowadziło by to do sytuacji gdzie dwa (lub więcej) znaki tekstu jawnego będą oznaczone tym samym znakiem.

Powyższy przykład w notacji cyklicznej ma postać:

(asnhyxwvqlfi)(bermgtoj)(c)(dupk)(z)

Jak widać nie jest to do końca dobre podstawienie, gdyż zarówno *c* jak i *z* przechodzą same w siebie, co może być niekorzystne, ze względu na bezpieczeństwo szyfru. Dodatkowe alfabety powyższego typu, zwane alfabetami nieuporządkowanymi, można utworzyć na co najmniej dwa sposoby.

Pierwszy sposób to cykliczne przesuniecie jednego z wierszy, np.

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

czyli

(aeibcuqmh)(drnj)(ftplgyzsok)(v)(w)(x)

Jak widać taka drobna zmiana prowadzi do całkiem innego alfabetu.

Drugi sposób to liczenie potęg danej permutacji. Przykładowo druga potęga dla alfabetu utworzonego z hasła SECURITY ma postać:

(anyqgf)(brgo)(c)(dp)(emtj)(hxvlis)(ku)(z)

Potęgę te tworzy się pisząc co drugą (w tym przypadku) literę z zapisu cyklicznego. Jak widać ta operacja również wnosi pewne urozmaicenie.

Szczególne znaczenie historyczne mają potęgi permutacji monocyklicznych. Przykładowo podstawienie:

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

w zapisie cyklicznym ma postać:

(abcdefghilmnopqrstvx)

Jest to permutacja monocykliczna gdyż zawiera tylko jeden cykl. Jej trzecia potęga to:

(adglorvbehmpsxcfinqt)

Według Swetoniusza szyfrem takim posługiwał się Juliusz Cezar, który szukając potrzebnej litery przesuwał się po prostu w alfabecie o trzy pozycje do przodu. Alfabet taki nazywamy jest alfabetem

CAESAR. Tak nazywane jest również to szyfrowanie. Szyfrowanie takie stosowane było jeszcze w 1915 roku w armii rosyjskiej, kiedy okazało się ze sztaby nie są w stanie przyswoić sobie nic bardzie skomplikowanego.

Przedstawiony powyżej sposób tworzenia alfabetu na podstawie hasła nie jest jedynym. Przykładowo na podstawie tego samego hasła SECURITY można utworzyć alfabet w następujący sposób:

S E C U R I T Y	a e i l o r u x
A B D F G H J K	b f j m p s v y
L M N O P QVW	c g k n q t w z
X Z	d h

Polega on a zapisie hasła w wierszach a odczycie w kolumnach. Otrzymamy zatem:

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

Metody te można oczywiście modyfikować w dowolny sposób, jeśli tylko jest to korzystne.

Korzystanie z tylko jednego alfabetu podczas szyfrowania całego tekstu nie jest dobrym rozwiązaniem. Jest tak dlatego, że szyfrowanie takie zachowuje informacje o częstości występowania danych znaków. Jest to niebezpieczne, gdyż częstości występowania danych znaków są charakterystyczne dla danego języka – niektóre znaki występują częściej niż inne. Może to prowadzić do złamania szyfru i odkrycia alfabetu. Można temu zaradzić na wiele sposobów, oto dwa z nich:

Po pierwsze, zamienić szyfrowanie monoalfabetyczne na polialfabetyczne. Po zakodowaniu jednego znaku (lub jakiejś określonej ich liczby) zmienia się alfabet na inny. Można to zrobić w sposób przedstawiony już wcześniej, mianowicie wykorzystując cykliczne przesunięcia oraz potęgowanie permutacji. Podejście to jest cykliczne – po pewnym czasie dostaniemy początkowy alfabet. Szyfrowanie polialfabetyczne (niekoniecznie w takiej postaci jak opisana w tym akapicie) wykorzystywane było w maszynach szyfrujących, np. w Enigmie.

Drugim sposobem ochrony przed atakiem przez badanie częstości występowania danych znaków jest zastosowanie szyfrowania np. dwudzielnego. Spójrzmy na poniższa tabelkę:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
4,5,6,7,8,9,0	e	t	a	o	n	i	r	s	h	71,09%
2,3	b	c	d	f	g	j	k	l	m	19,46%
1	p	q	u	v	w	x	y	z		9,45%

W tym szyfrowaniu każda litera jest przedstawiana jako para cyfr (numer wiersza i numer kolumny). Jednak niektóre litery (te, które w języku występują częściej) mogą być przedstawione na więcej niż jeden sposobów co powoduje, ze częstość występowania znaków w szyfrogramie wyrównuje się.

Zadanie do zrealizowania:

Napisz program przeprowadzający szyfrowanie/deszyfrowanie wiadomości zapisanej w pliku za pomocą szyfrowania monograficznego jednodzielnego. Alfabet powinien być tworzony na podstawie podanego hasła na jeden ze sposobów omówionych powyżej lub w sposób podobny. Zaszyfrowana/odszyfrowana wiadomość zapisywana jest w osobnym pliku. Za alfabet tekstu jawnego można przyjąć dla uproszczenia cały zbiór znaków ASCII (0-255).