

Podstawy kryptografii

Andrzej M. Borzyszkowski

Instytut Informatyki
Uniwersytet Gdański

sem. letni 2023/2024

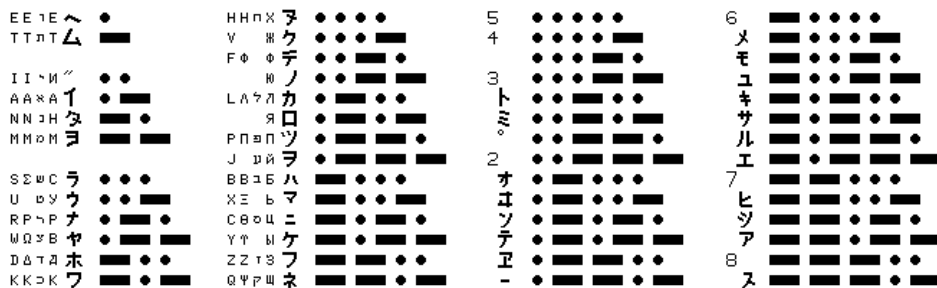
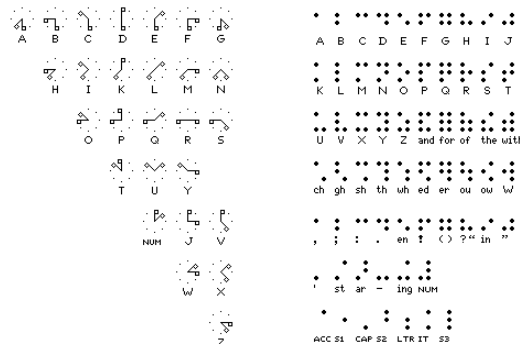
inf.ug.edu.pl/~amb/

- literatura:
 - D. Stinson, M. Paterson, *Kryptografia w teorii i w praktyce*, PWN, 2021
 - J-P. Aumasson, *Nowoczesna kryptografia*, PWN, 2018
 - B. Schneier, *Kryptografia dla praktyków*, WNT, 2002
 - M. Kutyłowski, W-B. Strohmann, *Kryptografia*, Readme, 1999
- materiały są/będą dostępne inf.ug.edu.pl/~amb
- obecność na wykładzie (oczywiście) nieobowiązkowa
- znajomość wykładu (oczywiście) obowiązkowa
- już na ćwiczeniach
- plus podstawowa umiejętność programowania
- program nie przewiduje egzaminu końcowego, wykład zakończy się testem zaliczeniowym

Kodowanie vs. szyfrowanie

- kodowanie: zamiana alfabetu na inny
 - alfabet Braille'a
 - kod ASCII
 - alfabet Morse'a
- kodowanie to nie szyfrowanie

źródło: is.gd/lxNX4A



Działy

- kryptografia: nauka/sztuka szyfrowania (i odszyfrowywania)
- kryptoanaliza: nauka/sztuka łamania szyfrów
- kryptologia: suma powyższych plus całościowe spojrzenie (właściwa nazwa przedmiotu)
jednak powszechne użycie: kryptografia

kodowanie też ma znaczenie

np. kody poprawiające błędy (*error correction codes*)

w dobrym szyfrze zmiana jednego bitu zaszyfrowanej wiadomości może uniemożliwić odszyfrowanie

kod też może być szyfrem jeśli zamiana alfabetów jest tajna

- informacja
dane, możliwość kopiowania, kradzież??
tekst jawny – wiadomość
tekst zaszyfrowany – kryptogram/szyfrogram
- uczestnik (entity)
człowiek, komputer, urządzenie, ...
Alicja, Bolek, Celina, Tadeusz, Pelagia, Wiktor, ...
(*Alice, Bob, Cindy, Trent, Peggy, Victor*)
- przeciwnik, Ewa, Mariola, ... (*Eve, Mallory*)
- klucz
znany nie wszystkim, łatwo zaszyfrować/odszyfrować z kluczem,
trudno bez klucza
uwaga: inne znaczenie niż np. w teorii baz danych

np. n^3 vs. $e^{2\sqrt{n}}$

n	n**3	2*sqrt(n)	exp(2*sqrt(n))
2	8	3	17
4	64	4	55
8	512	6	286
16	4096	8	2981
32	32768	11	81937
64	262144	16	8886111
128	2097152	23	6713706353
256	16777216	32	78962960182681
512	134217728	45	4.507385299E+0019
1024	1073741824	64	6.235149081E+0027
2048	8589934592	91	2.031652223E+0039
4096	68719476736	128	3.887708406E+0055
8192	549755813888	181	4.127610756E+0078
16384	4398046511104	256	1.511427665E+0111

- $MMMCDLXXVII * MDCCCXLIV$
było trudne dla Rzymian
ale nie dziś: $3477 * 1844 = 6411588$
- złożoność asymptotyczna, zależy od wielkości zadania, parametr $n \rightarrow \infty$
 - liniowa: n , żadna złożoność
 - wielomianowa, np.: n^2, n^3, n^{100}
 - wykładnicza, np.: $2^n, n!, n^n$
 - podwykładnicza, np.: $e^{\sqrt{n}}, e^{C \cdot \sqrt[3]{n \cdot \ln 2 \cdot \ln(n \cdot \ln 2)}}$
- stała też się liczy, np. $n = 1024$ bity i tylko ta wielkość nas interesuje

Założenia kryptografii

- przestrzeń tekstów jawnych M , kluczy K , kryptogramów C
 - algorytm generowania klucza $G : \rightarrow K$
 - algorytm szyfrowania $E : K \times M \rightarrow C$ (czy deterministyczny ?)
 - algorytm odszyfrowywania $D : K \times C \rightarrow M$
- zasada Kerckhoffs'a (1883):
przeciwnik zna szyfr (tzn. protokół/algorytmy)
przeciwnik ma duże zasoby obliczeniowe i duże umiejętności
przeciwnik NIE ZNA klucza
- dlaczego?
łatwiej utrzymać w tajemnicy klucz niż algorytm
nie da się opracować wielu (tajnych) algorytmów
- JEDYNY BEZPIECZNY szyfr: jednorazowy
w zasadzie nie ma dowodów, że inne szyfry są bezpieczne

- przeciwnik ma tylko tekst zaszyfrowany
- przeciwnik ma przykłady tekstów jawnych plus ich zaszyfrowane wersje
- przeciwnik może żądać zaszyfrowania wiadomości lub odszyfrowania (testowego) kryptogramu
- ataki pasywne vs. aktywne
- ilość: duża liczba tekstów lub par tekstów vs. pojedynczy tekst zaszyfrowany
- atak brutalny: przeszukiwanie całej przestrzeni kluczy K
 - aby zadziałał musi być metoda rozpoznania znalezienia klucza
 - dla obrony przestrzeń kluczy musi być duża, np. $> 2^{80}$ elementów

Kryptografia klasyczna vs. współczesna

- tekst jawny \rightarrow tekst zaszyfrowany \rightarrow tekst jawny
 $M \rightarrow E_K M \rightarrow D_K E_K M$ (zawsze przekształcenie z kluczem)
- klasyczna kryptografia (do lat '70): ten sam klucz
 - obie strony muszą wymienić klucz wspólny klucz
 - jak to zrobić?
- współczesna kryptografia: para kluczy (kryptografia asymetryczna, PKC),
 - idea: Diffie, Hellman (1976)
 - implementacja: RSA (Rivest, Shamir, Adleman) (1977)
 - wada: słaba wydajność
 - zaleta: nie trzeba przedtem przekazywać klucza

- poufność (tajność)
 - tylko uprawnieni uczestnicy mają dostęp do informacji, szyfrowanie
- integralność danych
 - dane są niezmienione (wykrycie zmiany, również/głównie celowej)
- uwierzytelnianie
 - w czasie rzeczywistym: identyfikacja uczestnika
 - odłożone w czasie: identyfikacja źródła dokumentu
- niezaprzeczalność
 - podpis: nie można się wyprzeć
 - niemożliwa w kryptografii klucza symetrycznego

Kryptografia klucza asymetrycznego

- przykład zastosowania
 - 1 Alicja prosi Bolka o przekazanie klucza publicznego, albo odczytuje z ogłoszenia, albo otrzymuje od wspólnego znajomego
 - 2 szyfruje wiadomość kluczem publicznym Bolka
 - 3 przekazuje wiadomość $E_B M$
 - 4 Bolek odszyfrowuje wiadomość swoim kluczem prywatnym
 $D_B E_B M = M$
- NIKT nie przesyła tajnego klucza
- problem: czy to naprawdę Bolek przekazał klucz publiczny?!

Kryptografia klasyczna

- szyfr Cezara
 - przesunięcie liter np. o 3 t.j. $y = x + 3 \pmod{26}$, $x = 0, 1, \dots, 25$
 - kryptoanaliza: wypróbowanie 25 przesunięć
 - jedna litera pary tekst jawny+zaszyfrowany wystarczy!
 $k = y - x \pmod{26}$
- szyfr afiniczny: $y = a \cdot x + b \pmod{26}$
 - odszyfrowywanie: $x = (y - b)/a \pmod{26}$
 - musi być określone dzielenie $1/a = a' \pmod{26}$ t.ż. $a \cdot a' = 1 \pmod{26}$ istnieje w.t.w. gdy $\text{NWD}(a, 26) = 1$
 - dla klucza (13, 4) „input” i „alter” szyfrują się do „ERRER”
 - kryptoanaliza: przestrzeń kluczy ma 312 elementów
 - dwie litery tekstu jawnego+zaszyfrowanego często wystarczają, kilka par prawie na pewno

Szyfr podstawieniowy, książka kodowa

- dwie książki z parami tekst jawny – tekst zaszyfrowany

Februar	13605
fest	13722
finanzielle	13850
folgender	13918
Frieden	17142

 - w jednej kolejność tekstu jawnego, w drugiej zaszyfrowanego
 - kluczem jest para książek! nieślychanie trudno o wymianę
- wersja: dodatkowym kluczem jest przesunięcie
 - $E(M) = \text{Książka}(M) + \text{przesunięcie} \pmod{100.000}$
 - $C(C) = \text{Książka}^{-1}(M - \text{przesunięcie} \pmod{100.000})$

Szyfr monoalfabetyczny

- szyfr monoalfabetyczny (kod), 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
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

 - kryptoanaliza: analiza częstotliwości wystąpień liter
 - ale można mieć kilka odpowiedników dla każdej litery (homofonia)
 - np. $E : \{A..Z\} \rightarrow \{00..99\}$, $|E(e)| = 12$, $|E(z)| = 1$
 - można/trzeba też rozważyć częstotliwości par liter, niektóre nie występują praktycznie wcale, inne b. często

klucz: wektor liczb np. (k_1, k_2, \dots, k_9)

- szyfrowanie: seria szyfrów Cezara: $y_1 = x_1 + k_1, y_2 = x_2 + k_2, \dots, y_9 = x_9 + k_9$, odszyfrowywanie analogicznie
- kryptoanaliza: przeszukiwanie wyczerpujące jest nierealne, liczba kluczy równa 26^n , np. dla $n = 9$ jest ich $5 \cdot 10^{12}$
- para tekst jawny+zaszyfrowany długości klucza definiuje klucz
- gdy znany jest tylko szyfrogram oraz długość klucza, to można/należy przeprowadzić analizę częstotliwości dla fragmentów szyfrogramu, dla zestawów $\{y_1, y_{10}, y_{19}, \dots\}$ itd.
- analiza częstotliwości oznacza przybliżenie wektora częstotliwości wystąpień liter w szyfrogramie z częstotliwościami języka naturalnego

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
82	15	28	43	127	22	20	61	70	2	8	40	24	67	75	19	1	60	63	91	28	1	24	2	20	1

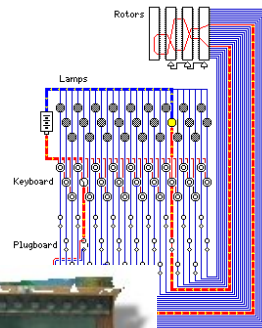
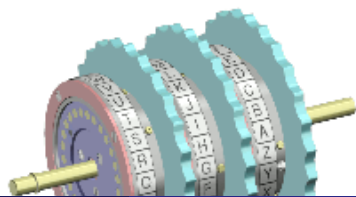
w oparciu o: https://en.wikipedia.org/wiki/Letter_frequency

Enigma

szyfr polialfabetyczny:

- zmiana kodu na bieżąco
- Enigma – rotory obracające się w miarę pisania (17576 kolejnych kodów)
- wielu wynalazców (Scherbius, Hebern, Koch)
- kryptoanaliza: M. Rejewski, Różycki, Zygalski, A. Turing

grafika: <http://www.otr.com/ciphers.shtml>
<http://home.ecn.ab.ca/~jsavard/crypto/intro.htm>
<http://encyclopedia.thefreedictionary.com>



znalezienie długości klucza:

- niech $A_0 = (p_0, p_1, p_2, \dots, p_{25})$ oznacza prawdopodobieństwa występowania liter w tekście
- niech $A_i = (p_i, p_{i+1}, \dots, p_{25}, p_0, \dots, p_{i-1})$ oznacza ten sam wektor z przesuniętymi wielkościami
- testujemy prawdopodobieństwo powtórzenia się litery w szyfrogramie oraz szyfrogramie przesuniętym o n miejsc
- jeśli $n =$ długość klucza, to litery były szyfrowane tym samym przesunięciem, $\text{prawd} = p_0 \cdot p_0 + p_1 \cdot p_1 + \dots + p_{25} \cdot p_{25} = A_0 * A_0$
- jeśli $n \neq$ długość klucza, to koincydencje przypadają na różne przesunięcia, prawdopodobieństwo koincydencji jest uśrednione po różnych iloczynach $A_0 * A_1, A_0 * A_2, \dots$ itd
- iloczyn $A_0 * A_0$ jest znacząco większy od innych (w jęz. ang. $A_0 * A_0 \approx 0.066$, inne iloczyny są w granicach 0.032 do 0.045)

Enigma, kryptoanaliza

- Protokół: wybrać klucz sesyjny (3 litery), powtórzyć go dwukrotnie i zaszyfrować kluczem dziennym
 - dane: setki szyfrogramów o powyższym początku
 - klucz dzienny wyznacza permutacje P_1, P_2, \dots, P_6 (i pozostałych 17570), złożenia $P_4 \circ P_1^{-1}$ i pozostałe dwa są podane implícite w danych
 - np. $ABCABC \rightarrow ENIGMA$, więc $P_4 \circ P_1^{-1}(E) = G$ itd.
 - dodatkowe podstawienie λ nieważna dokładną znajomość permutacji $\lambda \circ P_4 \circ P_1^{-1} \circ \lambda^{-1}$
 - ale „kształt” (rozkład na cykle) permutacji jest ustalony
 - opracowano enumeratywną listę rozkładów dla 100.000 kluczy

Kryptografia nowoczesna

Zasady nowoczesnej kryptografii

- Precyzyjne definicje
- Jawne założenia systemu
- Dowody poprawności
- (Kontr)przykład: definicja wymagań dla bezpiecznego systemu kryptograficznego
 - nie da się odtworzyć klucza kryptograficznego
 - nie da się odtworzyć tekstu jawnego
 - nie da się odtworzyć żadnego fragmentu tekstu jawnego
 - nie da się uzyskać żadnej sensownej informacji
 - nie da się obliczyć żadnej funkcji zależącej od tekstu jawnego
- dopiero ostatnia definicja w miarę precyzyjnie opisuje wymogi dla systemu kryptograficznego

Założenia/dowody?

- Model matematyczny a świat fizyczny
 - przykład: złamanie szyfru implementowanego na karcie kryptograficznej w oparciu o analizę zużycia energii
- Jawne założenia + analiza spełnienia założeń
- Różnica pomiędzy oprogramowaniem ogólnego użytku a kryptografią
 - kto zauważa błędy systemu/implementacji?
 - kto celowo szuka błędów w systemie/implementacji?

Bezpieczeństwo szyfrowania

- Wersja z grą:
 - Ewa wybiera dwa teksty jawne: m_0, m_1
 - Tadeusz losuje klucz k , losuje bit $b \in \{0, 1\}$, ogłasza $Enc(m_b, k)$
 - Ewa stara się odgadnąć bit b , wygrywa jeśli odgadnie
 - bezpieczny system znaczy, że Ewa odnosi sukces z prawdopodobieństwem $\frac{1}{2}$
- czyli znajomość kryptogramu nie daje nawet jednego bitu informacji – który z dwóch tekstów jest zaszyfrowany

- Bezpieczeństwo szyfrowania wielokrotnego
 - Ewa nie potrafi odgadnąć, który zestaw wiadomości jest szyfrowany
- Twierdzenie: szyfrowanie musi być niedeterministyczne
 - dw.: Ewa przygotowuje dwa zestawy, jeden identycznych wiadomości, drugi różnych. I sprawdzi, czy zestaw kryptogramów ma identyczne elementy czy różne.
- Atak z wybranym tekstem jawnym (CPA)
 - założenie: Ewa ma dostęp do maszyny szyfrującej
 - i nie potrafi odgadnąć, która wiadomość jest zaszyfrowana
- Twierdzenie: szyfr bezpieczny przeciwko atakowi z wybranym tekstem jawnym musi być niedeterministyczny
 - dw.: Ewa żąda od wyroczeni zaszyfrowania obu wiadomości i sprawdza, która została podana jej jako wyzwanie
- Twierdzenie: atak z wybranym tekstem jawnym jest łatwiejszy niż z zestawem tekstów jawnych, system odporny na atak z wypranym tekstem jawnym jest odporny na atak z zestawem tekstów jawnych.

- Klasyczne szyfry są prawie zawsze nieodporne na atak z tekstem jawnym
 - szyfr Cezara, Vigenere'a, Hilla
 - wielokrotne przykłady łamania szyfrów w praktyce (atak na Midway, „żądanie” zaszyfrowania słowa 'Midway')
- Odporność na atak z wybranym tekstem jawnym jest ważna
 - serwery odpowiadają na żądania użytkowników
 - być może są to ataki
- Tw.: odporność na atak z wybranym tekstem jawnym = odporność CPA przy wielokrotnym szyfrowaniu