Kryptografia i bezpieczeństwo systemów informatycznych

Andrzej M. Borzyszkowski

Instytut Informatyki
Uniwersytet Gdański

sem. letni 2023/2024

inf.ug.edu.pl/~amb/

Organizacja

- literatura:
  - Białas Andrzej, *Bezpieczeństwo informacji i usług w nowoczesnej instytucji i firmie*, WNT, 2007
  - 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

Problemy bezpieczeństwa

Bezpieczeństwo

Bezpieczeństwo dla kogo?

- dla otoczenia – nie ten wykład
- dla siebie: użytkownik ma oczekiwania
  - wygoda
  - szybkość
  - niezawodność
  - poprawność – system jest zgodny ze specyfikacją
    ale problem – kto fatyguje się specyfikacją?
  - ...
  - bezpieczeństwo

Podstawowe definicje

- Dane / informacja
  – dokumenty, pliki,
  – ale również aplikacje, system, konfiguracja systemu.
- Uczestnicy
  – użytkownicy,
  – ale również aplikacje, procesy, serwery, rutery.
- Identyfikacja – rozróżnianie użytkowników (np. login, identyfikator)
- Uwierzytelnianie (*authentication*) – potwierdzenie tożsamości użytkownika
  – użytkownik wie (hasło, protokoły wiedzy zerowej)
  – użytkownik fizycznie posiada (token, karta kryptograficzna, ...)
  – użytkownik ma cechy fizyczne (odcisk palca, tęczówka oka, kształt twarzy/dłoni, charakter pisma, wzorzec zachowań, ...)

Podstawowe definicje c.d.

- Autoryzacja (*authorization*) – przydzielenie praw uczestnikom
- Zarządzanie dostępem (*access control*) – nadzór nad przestrzeganiem praw dostępu
- Poufność (*confidentiality*) – brak dostępu do danych dla użytkowników nieuprawnionych
- Autentyczność – wiedza, kto jest (uprawnionym) autorem danych
- Integralność / spójność (*integrity*)
  – lepiej: niemożność nieuprawnionej modyfikacji informacji
  – a co najmniej wykrycie ew. nieuprawnionej zmiany
- Niezaprzeczalność (*nonrepudiation*) – brak możliwości zaprzeczenia autorstwa dokumentu lub wykonania czynności

Problemy bezpieczeństwa wg PN-I-13335

1. Poufność (tajność)
2. Dostępność (dyspozycyjność) –
   uprawniony użytkownik ma dostęp w określonym czasie
3. Autentyczność (uwierzytelnianie) –
   użytkownik jest naprawdę tym, za kogo się podaje
4. Integralność (spójność) danych i systemu –
   ani dane ani system nie zostały zmienione w sposób nieuprawniony, ani złośliwie, ani przypadkowo
5. Rozliczalność –
   działania użytkownika muszą być jednoznacznie przypisane
   - Niezaprzeczalność – brak możliwości zaprzeczenia autorstwa dokumentu lub wykonania czynności
6. Niezawodność

źródło: `sklep.pkn.pl/pn-i-13335-1-1999p` wikipedia: `is.gd/KFJRB9`

Historia

- 1983 Pomarańczowa Księga
  Trusted Computer System Evaluation Criteria – TCSEC
  pierwsze definicja pojęć, próba standaryzacji
- Kolejne kolorowe księgi
  czerwona – poświęcona bezpieczeństwu sieci
  zielona – poświęcona tematyce haseł
- Zdefiniowanie wspólnych kryteriów (common criteria)
- norma ISO/IEC TR 13335 (polska norma jest kopią części tej normy)

Przeciwnicy

- Hacker – entuzjasta komputerów, ciekawy ich działania, szczególnie w praktyce; niezłośliwy, może szkodliwy
- Cracker, intruz, włamywacz – przestępca, celowo łamie zabezpieczenia by uzyskać korzyść zabronioną
- Wirus komputerowy
- Złodziej fizyczny
- Śmierć/zdrada pracowników odpowiedzialnych
- Zniknięcie firmy serwisującej sprzęt, zarządzającej chmurą, przestarzałe technologie przestają być obsługiwane

Ataki

- pasywne vs. aktywne – jedynie podsłuch (*eavesdropping*), zamach na poufność
  albo również możliwość wykonywania operacji, w tym
  - podszywanie się pod innych uczestników (*masquerading*)
  - powtórne użycie zdobytych informacji (*replaying*)
  - manipulowanie danymi i oprogramowaniem (*tampering*), zamach na integralność
  - DOS (*denial of service*) – zarzucenie systemu nadmiarem żądań i w efekcie uniemożliwienie działania, zamach na dostępność
- lokalne vs. zdalne – atakujący już ma konto w systemie i próbuje nieuprawnionych operacji albo w ogóle nie ma konta

Dylematy bezpieczeństwa

- Bezpieczeństwo kosztuje
  również finansowo – czy równoważy ew. straty?
  zmniejsza efektywność programów – j.w.
  może zagrozić dyspozycyjności
- Atak DOS (denial of service)
  zarzucenie systemu z funkcją bezpieczeństwa taką ilością wymagań, że nie może ich wszystkich spełnić
  albo system jest bezpieczny i przestaje działać
  albo wyłączane są zabezpieczenia i system działa ryzykownie
- Audyt bezpieczeństwa
  wadliwie działający program raportuje użytkownik
  wadliwie działające zabezpieczenia pozostają tajemnicą
  zabezpieczyć należy wszystkie słabe punkty
  do ataku wystarczy jedna słabość

Cechy konieczne bezpieczeństwa

- Prostota
  system złożony może mieć więcej luk
- Jawność zasad
  w kryptografii „zasada Kerckhoffsa" – algorytmy są znane powszechnie, tajny jest klucz
  umożliwia to publiczną ocenę zasad, zmniejsza szansę na błędy
- Polityka bezpieczeństwa
  zdefiniowanie co jest chronione, jakie są zagrożenia, jak im przeciwdziałać
  określenie procedur i osób odpowiedzialnych
  okresowy/stały przegląd i nadzór

Rozwiązania

- Ochrona fizyczna – zabezpiecza przed kradzieżą, nieuprawnionym dostępem, zniszczeniem
- Odpowiednie oprogramowanie i sprzęt
- Procedury bezpieczeństwa
  - również szkolenia pracowników
  - zasada „need to know" i minimalnego przywileju
  - podział obowiązków, konieczność współdziałania (podział sekretu)

Kryptografia

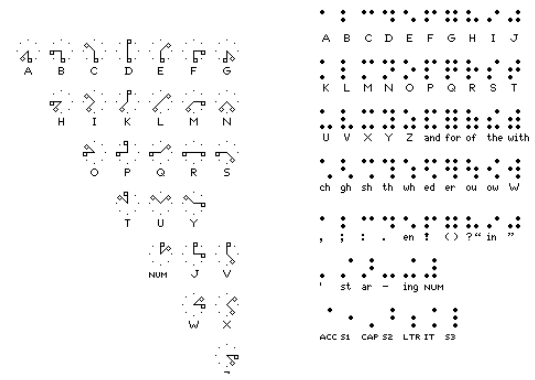
Kryptografia

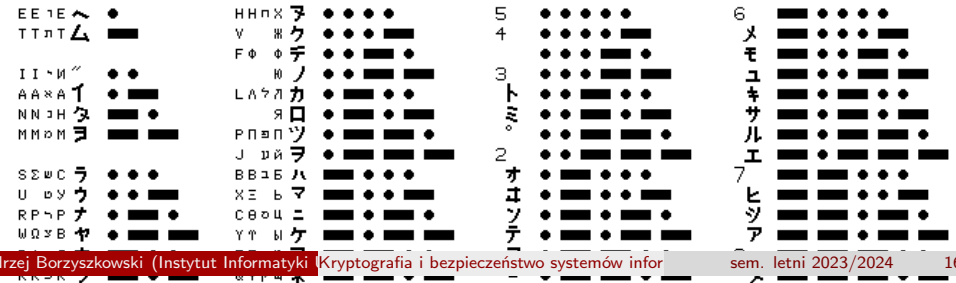
1. Poufność (tajność) – cała kryptografia klasyczna i dużą część współczesnej
2. Dostępność – w dużej mierze problematyka niezależna od kryptografii
   ale atak DOS (*denial of service*)
   ale *ransomware* – złośliwe zaszyfrowanie i żądanie okupu
3. Autentyczność (uwierzytelnianie) – hasła, podpis cyfrowy, Kerberos, algorytmy wiedzy zerowej, ...
4. Integralność (spójność) danych i systemu – MAC, funkcje skrótu
5. Rozliczalność –
   Niezaprzeczalność – podpis cyfrowy
6. Niezawodność – poza kryptografią

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

Cztery główne pojęcia

- informacja
  dane, możliwość kopiowania, kradzież??
  tekst jawny – wiadomość
  tekst zaszyfrowany – kryptogram
- 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 w teorii baz danych

Złożoność

- *MMMCDLXXVII* $*$ *MDCCCXLIV*
  było trudne dla Rzymian
  ale nie dziś: $3477 * 1844 = 6411588$
- złożoność asymptotyczna, zależy od wielkości zadania, parametr $n \longrightarrow \infty$
  - 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

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

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

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 Kerckhoffsa (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

Scenariusze ataków

- przeciwnik ma tylko tekst zaszyfrowany
- przeciwnik ma przykłady tekstów jawnych plus ich zaszyfrowane wersje
- przeciwnik może zaszyfrować żądaną wiadomość lub odszyfrować żądany tekst
- 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
  - 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

Kryptografia klucza asymetrycznego

- przykład zastosowania
  1. Alicja prosi Bolka o przekazanie klucza publicznego, albo odczytuje z ogłoszenia, albo 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?!

Algorytmy dawnej kryptografii

Kryptografia klasyczna

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

Szyfr monoalfabetyczny

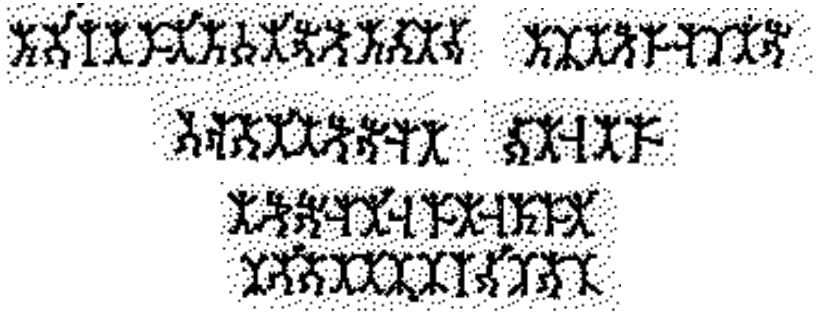
- szyfr monoalfabetyczny (kod), np.

  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

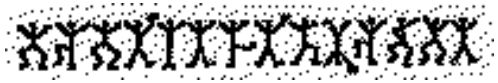
  – kryptoanaliza: analiza częstotliwości wystąpień liter
  – ale można mieć kilka odpowiedników dla każdej litery (homofonia)
  – np. $E : \{A..Z\} \to \{00..99\}$, $|E(e)| = 12$, $|E(z)| = 1$
  – można/trzeba też rozważać częstotliwości par liter, niektóre nie występują praktycznie wcale, inne b. często

Szyfr monoalfabetyczny

- przykład: Sherlock Holmes:



- był to szyfr monoalfabetyczny, S.H. znalazł klucz i przygotował sam wiadomość



grafika: www.trincoll.edu/depts/cpsc/cryptography/holmes/dancingmen

Szyfr Vigenere'a

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

- szyfrowanie: seria szyfrów Cezara: $y_1 = x_1 + k_1$, $y_2 = x_2 + k_2$, ..., $y_9 = x_9 + k_9$, 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 $\{y_1, y_{10}, y_{19}, \ldots\}$ 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`

Szyfr Vigenere'a, c.d.

znalezienie długości klucza:

- niech $A_0 = (p_0, p_1, p_2, \ldots, p_{25})$ oznacza prawdopodobieństwa występowania liter w tekście
- niech $A_i = (p_i, p_{i+1}, \ldots, p_{25}, p_0, \ldots, p_{i-1})$ oznacza ten sam wektor z przesuniętymi wielkościami
- testujemy prawdopodobieństwo powtórzenie się litery w szyfrogramie oraz szyfrogramie przesuniętym o n miejsc
- jeśli $n =$ długość klucza, to litery były szyfrowane tym samym przesunięciem, prawd $= p_0 \cdot p_0 + p_1 \cdot p_1 + ... + p_{25} \cdot p_{25} = A_0 * A_0$
- jeśli $n <>$ długość klucza, to koincydencje przypadają na różne przesunięcia, prawdopodobieństwo koincydencji jest uśrednione po różnych iloczynach $A_0 * A_1$, $A_0 * A_2$, 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

Szyfr Vigenere'a, III

znalezienie klucza:

- dzielimy kryptogram na grupy mod n, gdzie n jest długością klucza
- obliczamy wektor częstotliwości poszczególnych liter w pierwszej grupie $B_0 = (q_0, q_1, q_2, \ldots, q_{25})$
- jest on równy w przybliżeniu $A_i = (p_i, p_{i+1}, \ldots, p_{25}, p_0, \ldots, p_{i-1})$ dla i będącego przesunięciem w danej grupie, czyli pierwszą liczbą klucza
- testujemy iloczyn skalarny $B_0 * A_i$ dla kolejnych przesunięć, będzie on największy, gdy trafimy właściwie i
- kolejne liczby klucza obliczamy podobnie dla kolejnych grup znaków kryptogramu

Proste szyfry, poligramy

- kodować można pary i więcej liter – poligram
- szyfr angielski: Playfair
  `TO JEST PRZYKLAD SZYFRU UZYWANEGO V BAHAMACX`

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

`TF→ EL, NY→ VR, RK→ ZP, RA→ AN`

`pary są niedozwolone`
`Q nie ma, trzeba czymś zastąpić`

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 cała książka ! niesłychanie trudno o wymianę
- wersja: dodatkowym kluczem było przesunięcie
  – $E(M) = \text{Książka}(M) + \text{przesunięcie} \mod 100.000$

Proste szyfry, szyfr przestawieniowy

- szyfr niemiecki: ADFGX, kluczem części 1 jest tabela

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

  – część 1: T→ AA, O→ AD, itd. litery ADFGX mają bardzo różne kody Morse'a
  – część 2: kluczem jest słowo, np. RHEIN, tekst zaszyfrowany jest zapisany wierszami w 5 kolumnach, kolumny są przestawiane alfabetycznie względem klucza, na końcu są odczytywane pionowo

Szyfr przestawieniowy, przykład

- tekst jawny: KRYPTOGRAF

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

- krok 1: DX DD DG DA AA AD GG DD FD FG
- krok 2:

R	H	E	I	N		E	H	I	N	R
D	X	D	D	D		D	X	D	D	D
G	D	A	A	A		A	D	A	A	G
A	D	G	G	D		G	D	G	D	A
D	F	D	F	G		D	F	F	G	D

- kryptogram: DAGD XDDF DAGF DADG DGAD

Szyfry blokowe

- Playfair – blok dwuliterowy
  – obecnie znacznie dłuższe bloki (64, 128, 1024 bitów)
- Szyfr Hilla (ok. 1920, właściwie nieużywany)
  – klucz: liczba naturalna n, macierz kwadratowa wymiaru n
  – szyfrowanie: tekst jawny jest zapisany jako ciąg wektorów długości n, następnie mnożony przez macierz
  – odszyfrowywanie: trzeba znać macierz odwrotną, tzn. $A * B = I \mod 26$, możliwe jeśli $NWD(det(A), 26) = 1$
  – kryptoanaliza: analiza częstotliwości nie ma zastosowania przy kombinacjach wieloliterowych
  – znajomość tekstu jawnego i szyfrogramu pozwala obliczyć klucz (wektory muszą być liniowo niezależne, łatwe gdy można je wybierać)

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

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 $P_1, P_2, \ldots, P_6$, złożenia $P_4 \circ P_1^{-1}$ i pozostałe dwa są podane implicite w danych
  – np. $ABCABC \rightarrow ENIGMA$, więc $P_4 \circ P_1^{-1}(E) = G$ itd.
  – dodatkowe podstawienie λ unieważnia 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