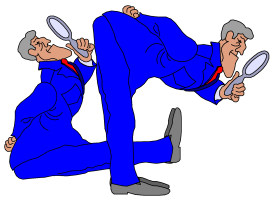
1. Caesar neboli Každý začátek je lehký!

Jan Paseka

Ústav matematiky a statistiky
Masarykova univerzita

23. listopadu 2023

O čem to bude



Opakování I

Nun-e-zuz-o-fuf-e-juj! Bub-u-dud-e-šuš o-sus-vuv-o-bub-o-zuz-e-nun!
Pup-o-zuz-o-rur! Tut-o juj-e tut-e-nun vuv-rur-a-huh!
Astrid Lingrenová

FI Při každém zakódování musí být příjemce vždy o něco před útočníkem. S pomocí této informace může příjemce zprávu rozšifrovat; tato informace nesmí být žádnému útočníkovi k dispozici, neboť by útočník byl schopen rozluštit zprávu zrovna tak lehce jako příjemce. O této exkluzivní informaci mluvíme jako o **klíči**.

Klasické šifrovací metody jsou založeny na tom základě, že odesilatel a příjemce mají společný šifrovací klíč, se kterým odesilatel zprávu zašifruje a příjemce ji rozšifruje; takováto metoda se nazývá **symetrická**.

Opakování II

V dalším uvidíme, že existují také asymetrické šifrovací algoritmy: v těchto systémech potřebuje **pouze příjemce tajný klíč**.

V této kapitole se budeme zabývat v jistém slova smyslu pouze těmi nejjednoduššími šifrovacími algoritmy a to takovými, že v nich je jedno a totéž písmeno nahrazeno jedním a tímtéž symbolem. Například písmeno **e** obsažené v textu by se zašifrovalo pomocí písmena **K**.

*

Terminologie I

Nejdříve několik slov k terminologii. Pojmy **kryptologie** a **kryptografie** pochází z řeckých slov $\kappa\rho\upsilon\pi\tau o\sigma$ (tajný) a $\lambda o\gamma o\sigma$ (slovo, smysl) a $\gamma\rho\alpha\varphi\varepsilon iv$ (psát).Obě slova označují umění a vědu, která se zabývá rozvojem metod k utajení zpráv. (Mnozí autoři rozlišují mezi **kryptografií**, tj. vědou o vývoji kryptosystémů a **kryptoanalýzou**, uměním tyto kryptosystémy prolomit a označují slovem **kryptologie**, spojení těchto věd.)

Text, řetězec znaků nebo písmen, který chceme zprostředkovat se nazývá **zpráva**; obvykle budeme zprávu reprezentovat pomocí malých písmen **a, b, c,**. . . . Zašifrovanou zprávu budeme nazývat **kryptogram**; tento pak budeme psát pomocí velkých písmen **A, B, C,**. . . . Šifrovací postup nazýváme **šifrování**, odšifrovací postup **dešifrování**. Odesilatel tedy šifruje, zatímco příjemce musí dešifrovat, aby si mohl přečíst zprávu.

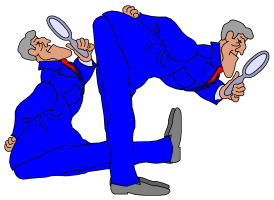
Terminologie II

Texty, které budeme šifrovat, se skládají z **písmen**; tato písmena jsou prvky nějaké **abecedy**. V prvních dvou kapitolách bude obvykle naše abeceda přirozená **abeceda** {**a, b, c,**. . . **}**.

Budeme také vybírat za abecedu např. množinu $\{1, \ldots, 26\}$, množinu $\{0, 1\}$ nebo také množinu $\{(a_1, \ldots, a_{64}) : a_i \in \{0, 1\}\}$ všech binárních posloupností délky 64 v případě, že to bude pro naše úvahy vhodné.

V rozporu s nadpisem kapitoly začínají dějiny kryptografie *před* Caesarem.

O čem to bude



Skytála

Historie začíná asi před 2500 lety. Jak dobře víme z díla řeckého dějepisce Plutarcha, používala vláda ve Spartě následující lstivou metodu pro přenos tajné zprávy pro své generály.

Odesilatel a příjemce museli mít oba tzv. **skytálu**: byly to dva válce o přesně stejném průměru. Odesilatel navinul úzkou pergamenovou pásku spirálovitě okolo své skytály a napsal pak podle délky svou zprávu na pásku.

Po odmotání pásky mohla zprávu číst jen ta osoba, která měla skytálu stejného rozměru – doufejme, že to byl pouze příjemce.

Uvažujme nyní příklad převedený do moderního jazyka.

Skytála jinak I

Představme si, že jsme zachytili list papíru, na kterém čteme následující řetězec písmen:

UNDTLATEDZEEIOVEMEJKSSMYNZ.EOI
IELAENLTCTENLOIEKRZOAMKKIUENN

Skytála odesílatele má průměr, který můžeme vyjádřit pomocí počtu písmen. Můžeme tedy jednoduše vyzkoušet různé rozsahy u. Zvolíme-li $u = 7$, dostaneme následující nesmysl:

U	E	V	S	O	N	L	O	E
N	D	E	M	I	L	O	A	V
D	Z	M	Y	I	T	I	M	N
T	E	E	N	E	C	E	K	
L	E	J	Z	L	T	K	K	
A	I	K	.	A	E	R	I	
T	O	S	E	E	N	Z	U	,

Skytála jinak II

Zvolíme-li ale uspořádání textu pro $u = 9$, dostaneme:

U	Z	J	E	N	O	M
N	E	K	O	L	I	K
D	E	S	I	T	E	K
T	I	S	I	C	K	I
L	O	M	E	T	R	U
A	V	Y	L	E	Z	E
T	E	N	A	N	O	V
E	M	Z	E	L	A	N
D	E	.				

Skytála je prototyp **transpozičního algoritmu**; přitom písmena zůstávají, ***co*** jsou, nezůstanou však, ***kde*** jsou. Nejedná se o nic jiného, než o permutaci ***míst*** písmen.

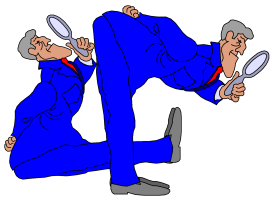
Skytála jinak III

Transpoziční algoritmy jsou důležitým stavebním kamenem pro moderní algoritmy.

Jinou složkou jsou ***substituční algoritmy***; u nich se zpráva stane nečitelnou tím, že každé písmeno se nahradí jiným, ale jeho pozice zůstane zachována.

A nyní nastává doba pro vstup Caesara na scénu.

O čem to bude



Caesarova šifra I

Jeden z prvních, kdo používal kryptologické techniky, byl římský vojevůdce a státník Gaius Julius **Caesar** (100-44 př. n. l.).
U Suetona (Caes. LVI) můžeme číst

> *Exstant et [epistolae] ad Ciceronem, item ad familiares de rebus, in quibus, si qua occultius perferenda erant, per notas scripsit, id est sic structo litterarum ordine, ut nullum verbum effici posset; quae si qui investigare et persequi velit, quartam elementorum litteram, id est D pro A et perinde reliquas commutet.*

Překlad zní přibližně následovně

> *Existují také [Caesarovy dopisy] Cicerovi a známým o věcech, v kterých psal tajným písmem, pokud něco muselo být důvěrně sděleno. Tzn. změnil pořadí písmen tak, že nešlo zjistit jediné slovo. Pokud někdo chtěl toto rozluštit a poznat obsah, musel dosadit čtvrté písmeno abecedy, tedy D, za A, a podobně toto provést se zbývajícími písmeny.*

Caesarova šifra II

Šifru použitou Caesarem obdržíme tím způsobem, že místo abecedy zprávy budeme psát abecedu kryptogramu, ale o 23 míst doprava, což znamená totéž, jako posunutí doleva o 3 místa:

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

Šifruje se tím způsobem, že nahradíme písmeno zprávy pod ním stojícím písmenem kryptogramu. Například ze slova "**zprava**" se stane zdánlivě nesmyslné slovo "**CSUDYD**".
Dešifrování je zrovna tak jednoduché: Každé písmeno kryptogramu se nahradí nad ním stojícím písmenem zprávy.

Caesarova šifra III

Člověk se ovšem může ptát, proč Caesar zvolil právě posunutí o 3 místa.

Odpověď je jednoduchá: nebyl na to vůbec žádný důvod! Samozřejmě můžeme posunout abecedu o libovolný počet míst.

Protože se naše abeceda sestává z 26 písmen, existuje právě 26 takových šifrování; mluvíme o **posouvacích** neboli **aditivních šifrách** a mezi nimi je samozřejmě **triviální šifrování a $\to$ A, b $\to$ B, . . . , z $\to$ Z, které samozřejmě nikdo nebude používat k šifrování tajných zpráv.**

Vyjasnění pojmů I

Vyjasněme si na této nejjednodušší třídě šifer pojmy "**šifrovací algoritmus**" a "**klíč**".

Šifrovací algoritmus je bezprostředně vidět na šifrování slova zprava.

Naproti tomu **klíč** je např. počet míst, o která je nutno posunout abecedu. Klíč nám přesně popisuje, jak se algoritmus použije ve speciální situaci.

Vyjasnění pojmů II

Partneři, kteří spolu komunikují, se musí dohodnout o šifrovacím algoritmu a o způsobu přenosu klíče.

Šifrovací algoritmus a klíč mají dvě zcela rozličné funkce a musí být proto zcela jasně rozlišitelné.

Algoritmus jako takový je zpravidla velmi "velký". Přitom mnoho algoritmů se realizuje pomocí mechanického zařízení nebo spočívá na více či méně veřejně přístupném postupu.

Z toho plyne, že algoritmus nelze v podstatě udržet v tajnosti. To pak znamená, že ***celková bezpečnost kryptosystému leží na utajení klíče.***

Vyjasnění pojmů III

Tento požadavek se zdá být přehnaný, je ale nanejvýš realistický: pro někoho, kdo chce neoprávněně číst naši zprávu, je srovnatelně lehké získat algoritmus (např. odcizit či zkopírovat přístroj). A pak tento zlosyn ví vše o algoritmu; nezná ale (doufejme) současný klíč.

Z toho nutně plyne, že klíč je nutno předat bezpečnou cestou. K čemu pak vůbec šifrování zprávy? V tom případě jsme mohli hned přenést celou zprávu touto bezpečnou cestou! – Tato námitka je plně oprávněná, lze ji však následujícími argumenty podstatně zmírnit.

Vyjasnění pojmů IV

1. Zpráva bývá zpravidla velmi dlouhá, klíč se obvykle volí co nejkratší. Dodatečná námaha pro spolehlivý přenos klíče je pak silně redukovatelná. Proto je pravděpodobnost, že klíč bude odposloucháván, relativně malá.
2. Odesilatel a příjemce si mohou libovolně vybrat dobu předání klíče. Klíč lze například dohodnout dny před přenosem zprávy. Naproti tomu se zpráva musí odeslat v okamžiku, který není ovlivnitelný komunikujícími partnery (uvažme politické události, vývoj na burze, atd.)
3. S pomocí tzv. Public-Key systémů lze klíče bez nebezpečí vyměnit, abychom pak provedli zakódování s pomocí konvenčních postupů.

Vyjasnění pojmů V

Upozorněme ještě na další nebezpečí. Je-li klíč vyměněn, musí být spolehlivě uložen; nesmí nastat případ, že jej bude možno z přístroje zjistit. Experti souhlasí s tím, že klíč je pouze tehdy bezpečně uložen, pokud přístroj nelze najít fyzikálními prostředky.

Nyní ale nastala doba na změnu stran. Kryptologie se nezabývá pouze tím, že navrhuje bezpečnostní systémy pro přenos zpráv; jedna z jejich ústředních úloh je takové systémy "rozluštit" (nebo se o to alespoň pokusit!).

Kryptoanalýza I

Začněme tedy na okamžik hrát roli zlosyna – to lze vyjádřit trochu slušněji následovně: pracujeme jako **kryptoanalytik** a provádíme **kryptoanalýzu** kryptogramu (případně zkoumaného kryptosystému).

Tvůrce kryptosystému musí vždy počítat s možností, že algoritmus je protivníkovi znám (alespoň po delší dobu). Kromě toho se doporučuje protivníka nepodceňovat a přisoudit mu co nejvyšší inteligenci; nazvěme je **Mr. X**.

Představme si, že Mr. X zachytil následující kryptogram:

BIBV HXZIDI CH VMVQ BIRVI

Úvod Spartská skytála | Posouvací šifry Monoabecední šifry | Další šifry Kryptoanalýza | Caesar Pojmy | Kryptoanalýza Statistika

Kryptoanalýza II

Na základě jistých indicií došel k domněnce, že tento text byl zašifrován pomocí posouvací šifry (např. by mohl "najít" jeden z výše popsaných šifrovacích strojů). Takovýto text pak lze principiálně analyzovat dvěma způsoby.

1. *"Systematické" prozkoušení všech možností*

Protože se jedná pouze o 26 posunutí, není naše námaha příliš velká. Mr. X ale může tuto námahu podstatně zredukovat, omezí-li se pouze na malou část zachyceného textu.

Uvažme např. "slovo" **VMVQ**. Vyzkoušíme-li všechna "posunutí" této posloupnosti písmen, zjistíme snadno, že ze všech možných ekvivalentů pouze slovo **neni** dává smysl. Je tedy více než pravděpodobné, že kryptogram byl získán posunutím o 8 míst. Mr. X pak prověří svou domněnku tím, že dešifruje celkový text.

Kryptoanalýza III

Obdrží pak:

tato zprava uz neni tajna.

Tato metoda pro prolomení posunovací šifry je proto tak dobrá, protože většina kombinací písmen je v češtině zcela bez významu.

Ačkoliv je toto pozorování důležitý základ pro mnoho kryptoanalytických metod, má výše uvedená metoda velkou nevýhodu.

Nelze ji totiž (nebo jen s neúměrně velkou námahou) automatizovat. Pokud by tato metoda měla být provedena počítačem samostatně, pak by bylo nutno uložit všechna (nebo v každém případě velmi mnoho) česká slova.

I když je to v principu možné, používali bychom zbytečně silný nástroj. A to nelze následující metodě vytknout.

Statistická analýza I

V češtině, němčině a angličtině (stejně jako v každém přirozeném jazyku) se nevyskytují všechna písmena se stejnou četností; spíše má každé písmeno svou charakteristickou četnost.

Tyto četnosti jsou uvedeny v následující tabulce:

Statistická analýza II

Písmeno	Četnost v % němčina	Četnost v % angličtina
a	6.51	6.40
b	1.89	1.40
c	3.06	2.70
d	5.08	3.50
e	17.40	10.00
f	1.66	2.00
g	3.01	1.40
h	4.76	4.20
i	7.55	6.30
j	0.27	0.30
k	1.21	0.60
l	3.44	3.50
m	2.53	2.00

Písmeno	Četnost v % němčina	Četnost v % angličtina
n	9.78	5.60
o	2.51	5.60
p	0.79	1.70
q	0.02	0.40
r	7.00	4.90
s	7.27	5.60
t	6.15	7.10
u	4.35	3.10
v	0.67	1.00
w	1.89	1.80
x	0.03	0.03
y	0.04	1.80
z	1.13	0.02

Statistická analýza III

Můžeme pak písmena rozdělit v závislosti na četnosti jejich výskytu do čtyř skupin (např. v němčině). V první skupině budou nejpočetněji se vyskytující **e** a **n**, v druhé skupině budou písmena s ještě relativně velkou četností (cca.**7%**); ve třetí skupině jdou uvedena písmena s malou, ale dosti podstatnou četností zatímco v poslední skupině jsou uvedena zanedbatelná písmena.

Skupina	Počet písmen skupiny v textu
e, n	27.18%
i, s, r, a, t	34.48%
d, h, u, l, c, g, m, o, b, w, f, k z	36.52%
p, v, j, y, x, q	1.82%

Statistická analýza IV

Co se stane, dešifrujeme-li nějakou (německou) zprávu? Pak samozřejmě zůstane četnost písmen zachována; avšak jednotlivé četnosti písmen nemusí být již přiřazeny svým odpovídajícím písmenům. Např. zašifrujeme-li ve zpráve písmeno **e** za **X**, pak bude písmeno **X** nejčetnějším písmenem kryptogramu, zašifrujeme-li **y** za **S**, nebude se **S** v kryptogramu téměř vyskytovat.

Konkrétně Mr. X ve své analýze zprávy

MRNBNA CNGC RBC WRLQC VNQA PNQNRV

vytvoří seznam jednotlivých četností

Písmeno:	**ABCDEFGHIJKLMNOPQRSTUVWXYZ**
Četnost:	**2240001000013601340002 1 000.**

Statistická analýza V

Pozor: Protože celá metoda je založená na statistice, není nic 100% jisté! Mr. X se musí zabývat dalším potvrzením své domněnky. Je-li jeho domněnka správná, jedná se o posunutí o 9 písmen.

Pak by muselo **R** odpovídat písmenu **i** (to potvrzuje jeho hypotézu, že se **R** vyskytuje relativně často) a **W** by muselo odpovídat písmenu **n** – to je však v rozporu s tím, že **W** se vyskytlo pouze jednou.

Z druhé strany se vyskytují **A, B, C** relativně často (měly by odpovídat **r, s, t**, a ekvivalenty **x, y, z** (totiž **G, H, I** se prakticky nevyskytují).

Mr. X se tedy pokusí provést posunutí zpět o 9 písmen a obdrží

dieser text ist nicht mehr geheim.

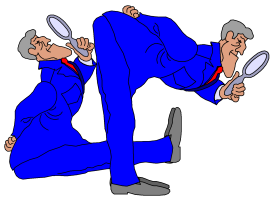
Statistická analýza VI

Výše uvedený text je smysluplný a tedy jeho domněnka je s konečným závěrem potvrzena.

Závěrem několik poznámek. Druhá metoda má bezespornou přednost, že ji počítač může sám lehce provést. Protože zde ale rozhodující roli hrají statistické úvahy, musí být zachována jistá obezřetnost.

Obzvláště při krátkých textech může vést naivní hledání po nejčastěji se vyskytujícím písmenu do slepé uličky. Pokud ale uvážíme ještě četnosti několika jiných písmen, lze použít algoritmy, které jsou i při velmi krátkých textech velmi úspěšné.

O čem to bude



Abecedy I

Šifrování se nazývá **monoabecední**, jestliže každý symbol otevřeného textu nahrazuje odpovídajícím symbolem zašifrovaného textu.

Monoabecední šifrování si můžeme představit tím, že pod abecedu zprávy napíšeme abecedu kryptogramu. Např. následující metody šifrování jsou monoabecední.

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

Ale

Zpráva:	**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**
Kryptogram:	**Q** $\rightarrow$ **E** $\alpha\kappa\beta$ **U** $\perp$ **O P** $\leq$ **S D F G H J** $\geq$ **3 Y 7 9 V 2 4 5.**

Jednoduché substituční šifry se lehce napadají, protože šifra netají frekvenci používání různých symbolů v otevřeném textu.

Abecedy II

Poslední příklad by nám měl připomenout, že zpráva a kryptogram ***nemusí být definovány nad stejnou abecedou***.

Je-li tomu ale tak, pak každému monoabecední šifrování odpovídá permutace písmen abecedy; obráceně lze každé permutaci přiřadit monoabecední šifrování.

Z toho zejména plyne, že máme přesně

$$26! = 26 \cdot 25 \cdot \cdots \cdot 2 \cdot 1 \approx 4 \cdot 10^{26}$$

monoabecedních šifrování nad přirozenou abecedou {**a, b, c, . . . , z**}.

Záměnné šifry I

Chceme-li použít k zakódování písmen počítač, pak identifikujeme obvykle **a** (resp. **A**) s **1**, **b** (resp. **B**) s **2**, atd.; **x** (resp. **X**) s **24**, **y** (resp. **Y**) s **25** a **z** (resp. **Z**) s **0**. Pomocí této reprezentace lze posouvací šifry popsat obzvlášť dobře: posunutí o *s* míst odpovídá přičtení čísla *s* modulo 26.

Konkrétně postupujeme následovně:

- Nejdříve převedeme písmena zprávy do odpovídajících číslic;
- pak připočteme k tomuto číslu číslo *s*;
- z výsledku uvažujeme pouze zbytek, který obdržíme po dělení 26; tento zbytek přeložíme zpátky na odpovídající písmeno.

Tímto způsobem získáme příslušný text kryptogramu.

Záměnné šifry II

Příklad 4.1

Nyní chceme zašifrovat písmeno **a** pomocí posunovací šifry, která posouvá o 3 místa.

- **a** se reprezentuje pomocí číslice **1;**
- **1 + 3 = 4;**
- **4 je reprezentace písmena D** kryptogramu.

Při dešifrování písmene **B postupujeme následovně:**

- **B** se reprezentuje pomocí číslice **2;**
- **2 – 3 = –1;**
- **Zbytek –1** po dělení**26** je **25** a to odpovídá písmenu **y** zprávy.

S pomocí této metody lze tzv. čísla ***sčítat.***

Záměnné šifry III

Celá věc se stává podstatně zajímavější, pokud budeme písmena ***násobit***. To lze provést následovně:
Abychom mohli násobit písmena číslem *t*, budeme počítat opět modulo 26. Tzn., že vynásobíme číslo odpovídající zadanému písmenu číslem *t* a uvažujeme zbytek po dělení 26. Pak je tomuto zbytku odpovídající písmeno výsledek tohoto "násobení".

Vynásobíme-li hodnotu každého písmena zprávy číslem 2, obdržíme

Zpráva:	**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**
Kryptogram:	**B D F H J L N P R T V X Z B D F H J L N P R T V X Z.**

Vidíme, že vždy dvě písmena (např. **h** a **u**) nám dávají tentýž "součin" (v našem případě **P**). Proto **nemůžeme tuto substituci použít jako šifru**.

Záměnné šifry IV

Pro každé šifrování musí totiž platit doposud nevyslovené ale zcela samozřejmé pravidlo, že *text zprávy musí být s pomocí klíče* **jednoznačně** *rekonstruovatelný z kryptogramu.*

Mnozí považují toto pravidlo za příliš omezující; lze ho však zeslabit a zároveň odůvodnit: Každý kryptogram musí být s pomocí klíče dešifrovatelný **nějakým počítačem.**

Pokusme naše štěstí ještě jednou a vynásobme všechna písmena číslem **3***:*

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

V tomto případě získáme skutečně monoabecední šifrování.

Záměnné šifry V

Lehkým prozkoušením vidíme, že obdržíme monoabecední šifrování právě tehdy, když násobíme jedním z čísel **1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23** nebo **25**; takovéto šifrování nazýváme **multiplikativní šifry**.

Máme tedy (včetně triviální šifry) právě 12 multiplikativních šifrování; jejich počet je tedy ještě menší než počet posouvacích šifer. Proto můžeme od těchto šifer očekávat velmi nepatrnou kryptografickou bezpečnost.

Můžeme ale navzájem kombinovat posouvací a multiplikativní šifry. Za tímto účelem přičteme nejprve k písmenu zprávy číslo s a výsledek vynásobíme dalším číslem t. Podle tohoto předpisu získáme opět šifru, tzv. **záměnnou (afinní)šifru**, kterou budeme označovat $[s, t]$.

Záměnné šifry VI

Klíč záměnné šifry $[s, t]$ sestává z dvojice čísel (s, t) (přirozeně musí být pro každou záměnnou šifru číslo t zvoleno tak, že násobení číslem t je multiplikativní šifra; t tedy musí být jedno z výše uvedených čísel **1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23** nebo **25**).

Počet všech záměnných šifer vypočteme jako počet všech posuvacích šifer vynásobený počtem všech multiplikativních šifer; tedy počet všech záměnných šifer je $26 \cdot 12 = 312$.

Toto číslo je už tak velké, že při ruční kryptoanalýze nám systematické prověření všech možností dá pěkně zabrat.

Komponenty I

Velké množství monoabecedních šifer získáme následujícím způsobem: **klíč** sestává ze dvou komponent – **klíčového slova** a **klíčového písmene**. Nejdříve vytvoříme z klíčového slova posloupnost písmen, ve které se každé písmeno vyskytne pouze jednou. To získáme následujícím způsobem, že každé písmeno se při svém druhém, třetím, . . . výskytu vyškrtne.

Máme-li zvoleno například klíčové slovo

MATEMATIKA,

získáme posloupnost

MATEIK.

Komponenty II

Napišme nyní tuto posloupnost pod abecedu zprávy, a to tím způsobem, že bude začínat právě pod klíčovým písmenem.

Např., zvolili jsme jako klíčové písmeno "j", obdržíme

Zpráva: **abcdefghi j k l mnopqrstuvwxyz**
Kryptogram: **MATE I K .**

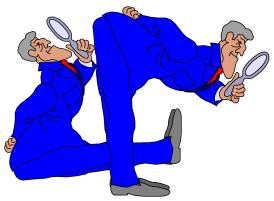
Na závěr napíšeme zbývající písmena kryptogramu v abecedním pořádku, přičemž začneme po posledním písmenu klíčového slova.

V našem případě pak získáme

Zpráva: **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**
Kryptogram: **QRSUVWXYZMATE I KBCDFGHJLNOP.**

Zřejmě lze ***každé monoabecední šifrování získat pomocí vhodného klíčového slova.***

O čem to bude



Hillova šifra I

Hillova šifra lineárně transformuje d znaků otevřeného textu na d znaků šifrového textu.

Bude-li $d = 2$, pak

$$\mathbf{C} = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}, \mathbf{M} = \begin{pmatrix} m_0 \\ m_1 \end{pmatrix}, \mathbf{K} = \begin{pmatrix} k_0 & k_2 \\ k_1 & k_3 \end{pmatrix}.$$

Šifrování zahrnuje násobení regulární matice $\mathbf{K}$ blokem otevřeného textu $\mathbf{M}$, t.j $\mathbf{C} = \mathbf{KM}$.

Dešifrování zahrnuje násobení matice $\mathbf{K}^{-1}$ blokem šifrového textu $\mathbf{C}$, tj $\mathbf{M} = \mathbf{K}^{-1}\mathbf{C}$.

Hillova šifra II

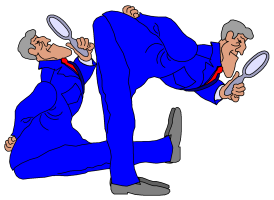
Příklad 5.1

$d = 2$ a budeme pracovat nad abecedou s 27 znaky (26 písmen a mezera), tj. nad $\mathbf{Z}_{27}$. Zvolme

$$\mathbf{K} = \begin{pmatrix} 4 & 6 \\ 1 & 7 \end{pmatrix}, \text{ pak } \mathbf{K}^{-1} = \begin{pmatrix} 4 & 12 \\ 11 & 10 \end{pmatrix}.$$

V tomto případě je determinant $\mathbf{K}$ nesoudělný s 27, tedy inverzní matice $\mathbf{K}^{-1}$ opravdu existuje.

O čem to bude



Kerckhoff I

"Filozofii" moderní kryptoanalýzy lze popsat Kerckhoffovým principem; tento princip byl poprvé formulován v knize *La cryptographie militaire* (1883) holandským jazykovědcem Jeanem Guillaumem Hubertem Victorem Francoisem Alexandrem Augustem **Kerckhoffem von Nieuwenhofem** (1835–1903).

Kerckhoffův princip. ***Spolehlivost kryptosystému nesmí záviset na utajení algoritmu. Spolehlivost je založena pouze na utajení klíče.***

To je zásadní varování pro tvůrce kryptosystémů. Nesmíme být tak naivní a přepokládat, že Mr. X nemá možnost získat znalost algoritmu. Dějiny kryptografie jsou plné příkladů, kdy objevitel kryptosystému založil důvěru na něm tím, že jeho algoritmus nikdy nemohl být znám.

Kerckhoff II

Naopak: Cílem moderní kryptografie musí být vývoj systémů, které zůstanou bezpečné i v tom případě, že o algoritmu bylo dlouhou dobu veřejně diskutováno.
"Příkladem" je AES-algoritmus.

Útoky na šifru I

Kryptoanalytik rozlišuje následující případy útoku na šifru:

1. Known ciphertext attack:
Kryptoanalytik zná relativně velkou část kryptogramu. To je opravdu reálný předpoklad, protože je zpravidla celkem jednoduché zajistit si (libovolně dlouhé) části kryptogramu.

2. Known plaintext attack:
Kryptoanalytik zná relativně malou část související zprávy/kryptogramu. Tato hypotéza je reálnější, než se na první pohled zdá. Totož často "ví" Mr. X, o co se jedná, a může proto uhádnout několik hlavních slov. Mimoto lze nalézt zpravidla standardní úvodní a závěrečné fráze atd.

Úvod Spartská skytála | Posouvací šifry Monoabecední šifry | Další šifry Kryptoanalýza | Kerckhoff | Útoky na šifru

Útoky na šifru II

3. Chosen plaintext attack:
Má-li kryptoanalytik přístup k šifrovacímu algoritmu (s aktuálním klíčem), může pak za účelem zjištění klíče kódovat vybrané části zprávy a pokusit se udělat z obdrženého kryptogramu závěry o struktuře klíče.

Mohl by například do stroje vkládat pravidelná zdrojová slova, např. posloupnosti stejných písmen (**aaa . . .**) za účelem jejich zakódování. Nebezpečí takovéhoto útoku spočívá v tom, že by se mohlo Mr. X podařit přimět šifrovací stroj k tomu, aby zakódoval ***zdánlivě*** neškodné zprávy, s jejichž pomocí pak Mr. X může zašifrovat zprávu, kterou by odesilatel nikdy nezašifroval.

Jak nebezpečný může být takovýto útok, se obzvláště ukáže, když pomyslíme na to, že mnohé šifrovací přístroje pouze nešifrují, ale také "podpisují".

Útoky na šifru III

Pokud je algoritmus tak slabý, že dovolí tento útok, pak by mohl Mr. X vytvořit z nevinně vyhlížejících podepsaných zpráv brizantní, platně podepsaný dokument.

*

Každé monoabecední šifrování přirozeného jazyka může být dosti lehce prolomeno. Musíme si pouze ujasnit, že každé monoabecední šifrování (přirozeného jazyka) lze prolomit již za vysoce slabého (přičemž opravdu realistického) předpokladu 1.

Předvedeme pouze "princip" algoritmu.

Představme si, že Mr. X zachytil kryptogram o délce asi 500 písmen a že ví, že kryptogram byl zašifrován pomocí monoabecedního šifrování.

Útoky na šifru IV

Krok 1. Nejdříve Mr. X zjistí četnosti písmen kryptogramu. Tím získá ekvivalent **e, n** spolu s množinou písmen **{i, s, r, a, t}**. Jednotlivá písmena z této množiny přitom zpravidla ještě nemůže identifikovat.

Krok 2. Nyní Mr. X spočte bigramy, tzn. páry po sobě sledujících písmen. Nejčastější bigramy německého jazyka jsou uvedeny v následující tabulce:

Bigram	Četnost	Bigram	Četnost
en	3.88%	**nd**	1.99%
er	3.75%	**ei**	1.88%
ch	2.75%	**ie**	1.79%
te	2.26%	**in**	1.67%
de	2.00%	**es**	1.52%

Takto může Mr. X izolovat písmena o největším výskytu.

Útoky na šifru V

Např. dvojice **er** má velmi velkou četnost, zatímco všechny jiné kritické kombinace s **e** se vyskytují dosti zřídka (**ea** a **et** jsou opravdu velmi řídké – pod 0.5%) a také **es** se vyskytuje se signifikantně malou četností.

Konkurentem by mohla být dvojice **ei**; tu však lze vyřadit tím způsobem, že testujeme inverzní dvojice: jen u těchto dvojic je tomu tak, že jak v původní posloupnosti, tak i v obráceném pořadí se vyskytují s prakticky stejnou četností.

Tímto způsobem může Mr. X nejprve izolovat rozlišitelná písmena skupiny { **i, s, r, a, t**} s druhou největší četností. Dále může rozpoznat písmena **c** a **h** podle toho, že se jako dvojice vyskytují relativně často ale samostatně velmi zřídka.

Tímto způsobem může prakticky bezchybně identifikovat nejčastěji se vyskytující písmena; tj. písmena **e, n, i, s, r, a, t, h, c**, která tvoří více než dvě třetiny textu.

Útoky na šifru VI

Krok 3 Pak nechá Mr. X dosadit rozpoznaná písmena do celého textu. Jinak řečeno: počítač rozšifruje známé díly textu. Ten se objeví na obrazovce, příčemž nerozluštěná písmena se nahradí účelně prázdnými znaky.
Zpravidla tento text není rozšifrován, nebo se jeho šifrování provádí ještě stále náročně. Další písmena však může inteligentní Mr. X na základě kontextu lehce *hádat*! To provede a nechá si vždy po každém kroku ukázat změněný text. Po dvou nebo třech krocích dospěje k velmi dobře čitelnému textu.

Shrnutí: Monoabecední šifrování nad přirozeným jazykem jsou pozoruhodně nejistá (**přirozený jazyk** má málo písmen, jež jsou dost nerovnoměrně rozdělena). V současné době proto používáme buď monoabecední šifrování nad *umělým jazykem* nebo *poly*abecední šifrování.