Literatura

- http://www.obluda.cz/iprednasky/index.html
- PFLEEGER, "*Security in Computing*", Prentice-Hall, 1989, …, 2015
- SCHNEIER, "*Applied Cryptography*", John Wiley & Sons, 1996, … 2015
- MENEZES, OORSCHOT, VANSTONE, *„Handbook of Applied Cryptography“*, CRC Press, 1996, …, 2001
- GOLDREICH, *„The Foundations of Cryptography“*, Cambridge University Press, 2001 (vol. 1), 2004 (vol 2.), http://www.wisdom.weizmann.ac.il/~oded/foc-book.html

Terminologie

Otevřený text [Plaintext] - originální tvar zprávy
Šifrovaný text, šifra [Ciphertext] – "zakódovaný" tvar zprávy po aplikaci šifrování

Šifrování [Encryption, encoding, enciphering] je proces, při kterém je zpráva zakódována tak, že její obsah není zřejmý
Dešifrování [Decryption, decoding, deciphering] je pak proces opačný
Kryptosystém [Cryptosystem] je systém umožňující šifrování a dešifrování zpráv a generování klíčů.

Formální zápis: (C - šifra, P - otevřený text, E,D - (de)šifrovací algoritmus, K - klíč)

šifrování:

$$C = \mathbf{E}(P) \text{ resp. } C = \mathbf{E}(K, P) \ , C = \mathbf{E}(K_E, P)$$

dešifrování:

$$P = \mathbf{D}(C) \text{ resp. } P = \mathbf{D}(K, C), \text{ C} = \mathbf{D}(K_D, P)$$

korektnost:

$$P = \mathbf{D}(\mathbf{E}(P))$$

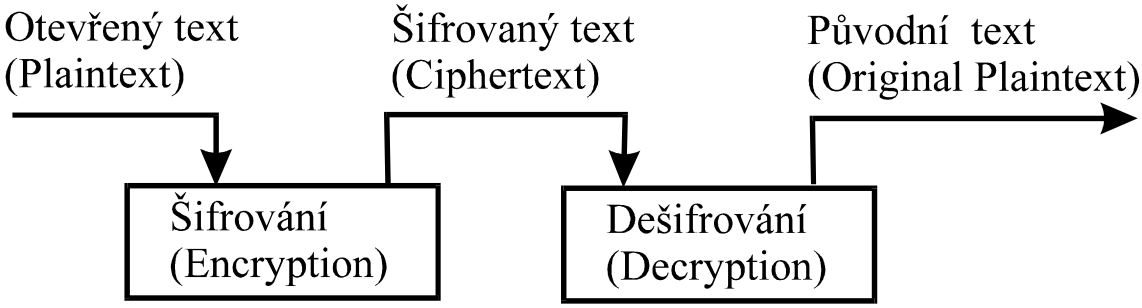


Figure 0.1 Šifrování (Encryption)

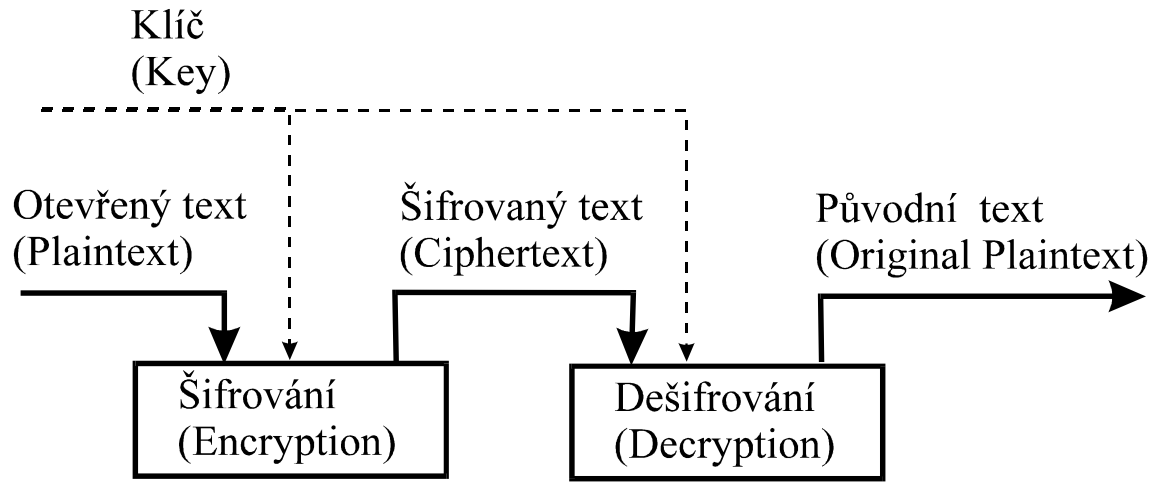


Figure 0.2 Šifrovací systém s jedním klíčem (Single-Key Cryptosystem)

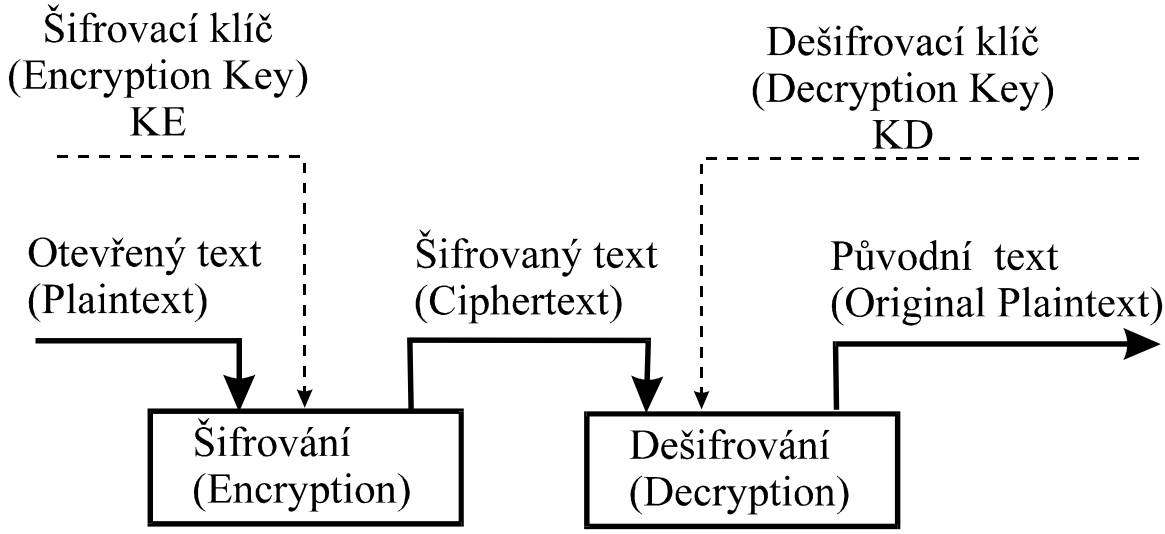


Figure 0.3 Šifrovací systém s dvěma klíči (Two-Key Cryptosystem)

Kryptografie [Cryptography] využívá šifrování k ukrytí dat

Kryptoanalýza [Cryptoanalysis] se zabývá hledáním způsobů jak šifrované zprávy neautorizovaně dešifrovat (encryption break)

Kryptologie [Cryptology] - je věda o šifrování obecně a zahrnuje tedy obě předchozí odvětví

Šifrovací algoritmus může být zlomen - znamená, že s dostatkem času a prostředků může být nalezen způsob jak dešifrovat jím zašifrované zprávy bez znalosti klíče.

Prakticky nezlomitelný - je znám postup jak se domoci otevřeného textu, ale ne v rozumném čase (se současnou technologií a znalostmi!).

Často používaná reprezentace znaků (výpočty jsou modulo n=26)

znak	A	B	C	D	E	F	G	H	I	J	K	L	M	N
kód	0	1	2	3	4	5	6	7	8	9	10	11	12	13

znak	O	P	Q	R	S	T	U	V	W	X	Y	Z
kód	14	15	16	17	18	19	20	21	22	23	24	25

Monoalfabetické šifry

- substituční šifry, používající jednu substituční tabulku, která každému znaku abecedy přiřazuje jiný

Caesarova šifra: $c_i=E(p_i)=p_i+3$

Otevřený text	`ABCDEFGHIJKLMNOPQRSTUVWXYZ`
Šifrovaný text	`defghijklmnopqrstuvwxyzabc`

Substituce s klíčem:
Otevřený text `ABCDEFGHIJKLMNOPQRSTUVWXYZ`
Šifrovaný text **`key`**`abcdfghijlmnopqrstuvwxz`

Kryptoanalýza - vyhledávání typických shluků znaků pro daný jazyk, typických prvních/posledních znaků slov a četnost výskytu jednotlivých znaků (**frekvenční analýza**)

Polyalfabetické substituce

nevýhodou monoalfabetických substitucí je, že odrážejí rozložení pravděpodobnosti jednotlivých znaků
použitím více substitučních tabulek polyalf. substituce dosahují rovnoměrného rozložení pravděpodobnosti výskytu jednotlivých znaků v šifrovaném textu

předp. k substituřních tabulek
znak p_{ik+j} je šifrován pomocí j-té tabulky

Vigenêre tableaux

je příkladem polyalf. substituce

klíčové slovo K délky k napíšeme opakovaně nad otevřený text.
znak $K_{i \bmod k}$ určuje řádek tabulky, který bude použit
znak p_i určuje sloupec, v průsečíku se nachází výsledná šifra

	0 1 2 012345678901234567890123 45 abcdefghijklmnopqrstuvwxyz	
A	abcdefghijklmnopqrstuvwxyz	0
B	bcdefghijklmnopqrstuvwxyza	1
C	cdefghijklmnopqrstuvwxyzab	2
D	defghijklmnopqrstuvwxyzabc	3
E	efghijklmnopqrstuvwxyzabcd	4
F	fghijklmnopqrstuvwxyzabcde	5
G	ghijklmnopqrstuvwxyzabcdef	6
H	hijklmnopqrstuvwxyzabcdefg	7
I	ijklmnopqrstuvwxyzabcdefgh	8
J	jklmnopqrstuvwxyzabcdefghi	9
K	klmnopqrstuvwxyzabcdefghij	10
L	lmnopqrstuvwxyzabcdefghijk	11
M	mnopqrstuvwxyzabcdefghijkl	12
N	nopqrstuvwxyzabcdefghijklm	13
O	opqrstuvwxyzabcdefghijklmn	14
P	pqrstuvwxyzabcdefghijklmno	15
Q	qrstuvwxyzabcdefghijklmnop	16
R	rstuvwxyzabcdefghijklmnopq	17
S	stuvwxyzabcdefghijklmnopqr	18
T	tuvwxyzabcdefghijklmnopqrs	19
U	uvwxyzabcdefghijklmnopqrst	20
V	vwxyzabcdefghijklmnopqrstu	21
W	wxyzabcdefghijklmnopqrstuv	22
X	xyzabcdefghijklmnopqrstuvw	23
Y	yzabcdefghijklmnopqrstuvwx	24
Z	zabcdefghijklmnopqrstuvwxy	25

Analýza polyalfabetických substitucí

základem je určení počtu použitých substitucí, dále dokument rozdělíme na části, šifrované stejnou substitucí a na tyto části použijeme postupy analýzy monoalf. šifer

určování počtu použitých substitucí -

Kasiského metoda

pokud se v otevřeném textu vyskytuje *k*-krát stejný řetězec znaků a k šifrování bylo použito *n* substitucí, které se cyklicky střídají, bude daný řetězec zašifrován přibližně *k/n* krát stejně.

1. prohledáváme zašifrovaný text na výskyt opakujících se řetězců (délky aspoň 3)
2. zjistíme vzdálenosti začátků jednotlivých řetězců
3. ke každé vzdálenosti získané v předchozím bodě vytvoříme seznam všech dělitelů tohoto čísla
4. počet použitých substitucí by měl odpovídat některému z často se vyskytujících dělitelů

Index koincidence

označme $Freq_i$ počet výskytů symbolu *i* ve zprávě. Index koincidence IC definujeme

$$IC = \sum_{i=a}^{i=z} \frac{Freq_i * \left(Freq_i - 1\right)}{n * \left(n - 1\right)}$$

Pokud má odpovídající otevřený text rozložení znaků blízké normálu, lze z *IC* usuzovat na počet použitých substitucí.

# substitucí	1	2	3	4	5	10	>10
IC	.068	.052	.047	.044	.044	.041	<.038

"Perfektní" substituční šifry

složitost analýzy polyalfabetických šifer roste s počtem použitých substitucí. -> co třeba použít každou substituci jen jednou

One-time pad

- Předpokládáme že máme k dispozici klíče v celkové délce větší, než přenášená zpráva.
- Každý tento klíč použijeme jen jednou, t.j. k zašifrování tolika znaků, jakou má délku. Šifrovat můžeme např. pomocí Vigenêrovy tabulky.
- K dešifrování je třeba mít stejnou sadu klíčů.

Dlouhé sekvence náhodných čísel

… mohou být použity namísto klíčů pro one-time pad systémy

Generátory náhodných čísel

je nutné používat vhodné generátory pracující na základě měření skutečně náhodných veličin
běžné počítačové generátory jsou nevhodné

Kongruenční generátor pseudonáhodných čísel

$$r_{i+1} = \left(a * r_i + b\right) \bmod n$$

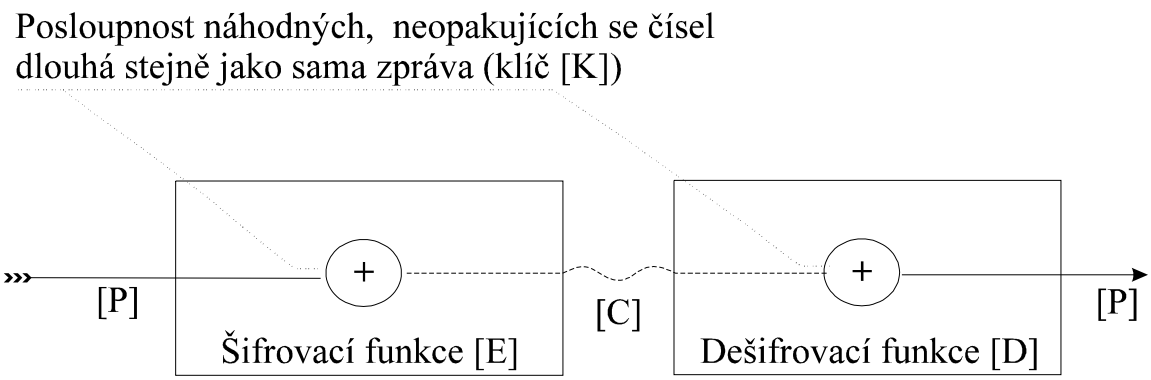
bohužel, pokud útočník získá $r_i, \ldots, r_{i+3}$, může dopočítat *a, b* a *n*.

Sekvence textů z knih

mohou být použity namísto náhodných čísel.
otevřený text a takto vytvořený klíč mají charakteristické rozložení pravděpodobnosti výskytu znaků.
až 25% znaků šifrovaného textu vzniká kombinací několika nejčastějších symbolů.
-> lze rozkrýt část zprávy, což nám umožní efektivně odhadnout zbytek.

Vernamova šifra

možná implementace one-time pad-u



Transpoziční šifry

Sloupcové transpozice

otevřený text zapíšeme do po řádcích matice
šifrovaný text vznikne přečtením této matice po sloupcích

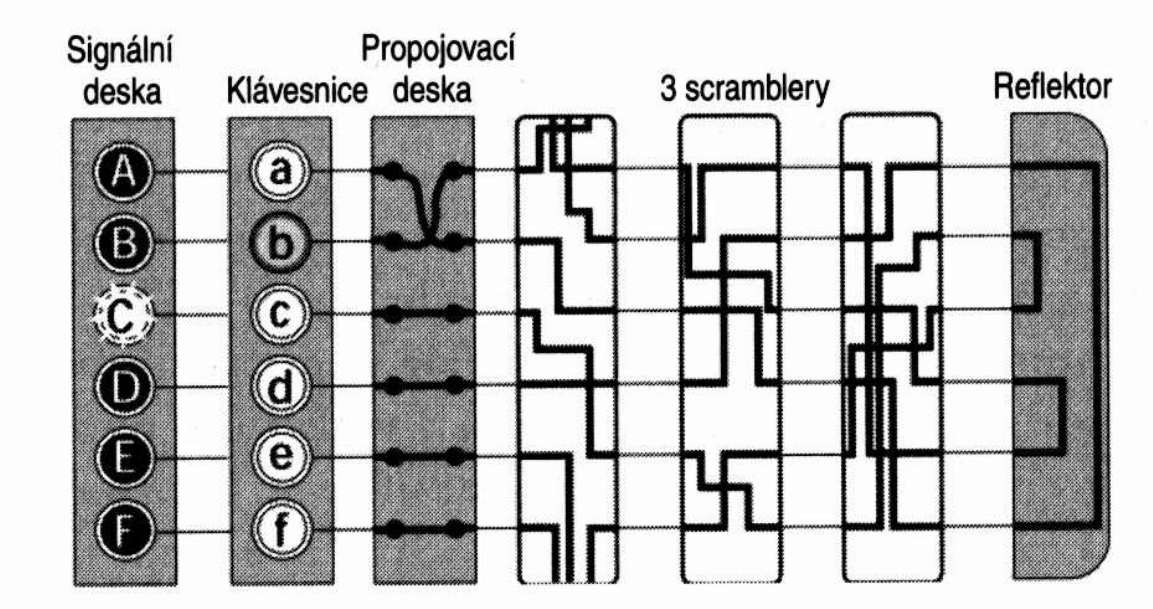
Základem je zjistit jak vypadala šifrovací matice. Analýza se provádí hledáním digramů, trigramů, případně delších sekvencí a jejich frekvenční analýzou.

tssohoa
niwhaasolrsto

tssohoa
niwhaasolrsto

Rotorové stroje

Enigma



Na podobném principu fungovala celá řada dalších strojů Hagellin, ...

Šifrovací algoritmy

kódování – způsob zápisu informace pomocí znaků zvolené abecedy – kódu
šifrování – podtřída kódů, k jejichž interpretaci je nutné znát dodatečnou informaci (***klíč***)

Klasifikace šifrovacích algoritmů

- podle způsobu práce – blokové, proudové
- podle klíčů s tajným klíčem (symetrické), s veřejným klíčem (asymetrické)

Kerckhoffův předpoklad:
Útočník zná všechny aspekty šifrovacího algoritmu s výjimkou použitého klíče.

Naivní charakteristika dobré šifry

Množství práce vynaložené na šifrování a dešifrování by mělo být úměrné požadovanému stupni utajení.
Šifrovací algoritmus by neměl obsahovat zbytečná omezení.
Implementace algoritmu by měla být co nejjednodušší.
Chyby při šifrování by se neměly příliš šířit a ovlivňovat následující komunikaci.
Zprávy by se zašifrováním neměly zvětšovat.

Security through obscurity NEFUNGUJE!

Zmatení (confusion) - nelze predikovat, jakou změnu zašifrovaného textu vyvolá byť jen malá změna otevřeného textu $\Leftrightarrow$ složitá funkční závislost mezi zašifrovaným textem a párem klíč - otevřený text.

Difuze (diffusion) - změna otevřeného textu se promítá do mnoha míst zašifrovaného textu.

Bezpečný systém - nelze získat otevřený text na základě znalosti odpovídající šifry

kryptoanalytik hledá transformaci $\mathbf{h} : C \rightarrow P$, $\mathbf{h}$ nebývá jednoznačná

Efektivně bezpečný systém - Prob($\mathbf{h}$(C) = P) < ε.

Ideální stav

Perfektní utajení (perfect secrecy) - mějme *n* možných otevřených textů, stejné množství klíčů a možných šifer.

$$Prob_{C_1}\left(\mathbf{h}\left(C_1\right)=P\right)=Prob\left(\mathbf{h}\left(C_1\right)=P\right)=Prob\left(P\right)$$

Redundance

počet bitů nutný k reprezentaci všech znaků abecedy $A=\left\lceil \mathbf{log}_2\left(k\right)\right\rceil$

počet všech možných zpráv délky $n = 2^{An}$, z toho 2^{Rn} smysluplných. *R* nazýváme *rate* jazyka. Redundance je definována

$$D = A - R$$

Pokud algoritmus šifruje několik různých zpráv, z nichž jedna je smysluplná, do stejné šifry, systém může být bezpečný.

Formálnější charakteristika dobré šifry

Kryptosystém je trojice (G, E, D) pravděpodobnostních p-time algoritmů splňující následující kritéria:

- Algoritmus G (generátor klíčů) nad vstupen 1^n vytvoří dvojici bitových řetězců
- Pro každý pár (*e, d*) z oboru hodnot G(1^n) a pro každé $\alpha \in \{1, 0\}^*$ algoritmus E (šifrování) a D (dešifrování) splňují
  $\Pr(D(d, E(e, \alpha)) = \alpha) = 1$
  kde pravděpodobnost se bere přes interní náhodná rozhodnutí algoritmů E a D.

Kryptosystém (G, E, D) je ***sémanticky bezpečný*** pokud existuje p-time transformace *T* taková že pro každý polynomiálně velký obvod $\{C_n\}$, každou posloupnost $\{X_n\}_{n\in N}$ kde $|X_n|$ je polynomiálně omezeno, každý pár polynomiálních funkcí *f* a *h:* $\{1, 0\}^* \rightarrow \{1, 0\}^*$, každý polynom *p* a všechna dostatečně velká *n*

$$Pr\left(C_n\left(E_{G_1(1^n)}(X_n), 1^{|X_n|}, h(X_n)\right) = f(X_n)\right) < Pr\left(C'_n\left(1^{|X_n|}, h(X_n)\right) = f(X_n)\right) + \frac{1}{p(n)}$$

, kde $C'_n = T(C_n)$ je obvod vytvořený transformací T nad vstupen C_n. Funkce h poskytuje parciální informaci o plaintextu X_n.

Kryptosystém (G, E, D) poskytuje ***nerozlišitelné šifrování***, pokud pro každý polynomiálně velký obvod $\{C_n\}$, každý polynom p, všechna dostatečné velká n, každé $x, y \in \{0,1\}^{poly(n)}$ (tj. "stejně dlouhá") a $z \in \{0,1\}^{poly(n)}$

$$\left|Pr\left(C_n\left(z, E_{G_1(1^n)}(x)\right) = 1\right) - Pr\left(C_n\left(z, E_{G_1(1^n)}(y)\right) = 1\right)\right| < \frac{1}{p(n)}$$

Pravděpodobnost se bere přes náhodná rozhodnutí algoritmů G a E.
Lze ukázat, že obě definice jsou ekvivalentní.

Šifrování obecně:

- *neskrývá samu existenci informace*
- *nezaručuje integritu*
- *nezaručuje autenticitu (původ)*
- *nebrání proti fabrikaci*
- *neskrývá všechny vlastnosti plaintextu*
- *důvěrnost zachovává pouze za určitých podmínek*
- *… a zbyde vám klíč*