Kriptosistem *je peterica* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ *za katero velja:*

1. *$\mathcal{P}$ je končna množica možnih čistopisov*
2. *$\mathcal{C}$ je končna množica možnih tajnopisov*
3. *$\mathcal{K}$ je končna množica možnih ključev.*
4. *Za vsak ključ $K \in \mathcal{K}$, imamo šifrirni postopek $e_K \in \mathcal{E}$ in ustrezen odšifrirni postopek $d_K \in \mathcal{D}$.*

$$e_K : \mathcal{P} \longrightarrow \mathcal{C} \quad \text{in} \quad d_K : \mathcal{C} \longrightarrow \mathcal{P}$$

sta taki funkciji, da je $d_K(e_K(x)) = x$ za vsak $x \in \mathcal{P}$.

Pomični tajnopis (angl. shift cipher) je poseben primer zamenjalnega tajnopisa.

`wewillmeetatmidnight`

```
22  4 22  8 11 11 12  4  4 19  0 19 12  8  3 13  8  6  7 19
 7 15  7 19 22 22 23 15 15  4 11  4 23 19 14 24 19 17 18  4
```

`HPHTWWXPPELEXTOYTRSE`

Kongruence: naj bosta a in b celi števili in m naravno število.

$$a \equiv b \pmod{m} \iff m | b - a.$$

Afini tajnopis:

$$e(x) = ax + b \pmod{26} \qquad \text{za } a, b \in \mathbb{Z}_{26}$$

Za $a = 1$ dobimo pomični tajnopis.
Funkcija je injektivna, če in samo če je $D(a, 26) = 1$.

Imamo $|\mathcal{K}| = 12 \times 26 = 312$ možnih ključev.

Za pomični tajnopis in afini tajnopis pravimo, da sta **monoabecedna**, ker preslikamo vsako črko v natanko določeno črko.

Vigenerejev tajnopis (1586):

Naj bo m neko naravno število in naj bo

$$\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{26})^m.$$

Za ključ $K = (k_1, k_2, \dots, k_m)$ definiramo

$$e(x_1, \dots, x_m) = (x_1 + k_1, \dots, x_m + k_m)$$
$$d(y_1, \dots, y_m) = (y_1 - k_1, \dots, y_m - k_m)$$

kjer sta operaciji "+" in "−" opravljeni po mo

To ni monoabecedni tajnopis.

Pravimo mu **poliabecedni tajnopis**.

Vigenerejev tajnopis in 26^m možnih ključev.

Za $m = 5$ je število 1.1×10^7 že preveliko, da bi "peš" iskali pravi ključ.

Hillov tajnopis (1929)

Naj bo m neko naravno število in naj bo

$$\mathcal{P} = \mathcal{C} = (\mathbb{Z}_{26})^m.$$

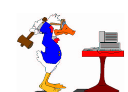
Za K vzemimo obrnljivo $m \times m$ matriko in definirajmo

$$e_K(x) = xK \quad \text{in} \quad d_K(y) = yK^{-1},$$

pri čemer so vse operacije opravljene v $\mathbb{Z}_{26}$.

Ponovimo:

Odšifriranje (razbijanje) klasičnih tajnopisov



Kriptografske sisteme kontroliramo s pomočjo ključev, ki določijo transformacijo podatkov.
Seveda imajo tudi ključi digitalno obliko (binarno zaporedje: 01001101010101...).

Držali se bomo **Kerckhoffovega principa**, ki pravi, da "nasprotnik"

pozna kriptosistem oziroma algoritme,
ki jih uporabljamo, ne pa tudi ključe,
ki nam zagotavljajo varnost.

Ločimo naslednje nivoje napadov na kriptosis

1. **samo tajnopis**: *nasprotnik ima del taj*
2. **poznani čistopis**: *nasprotnik ima del čistopisa ter ustrezen tajnopis,*
3. **izbrani čistopis**: *nasprotnik ima zača voljo šifrirno mašinerijo ter za izbrani x konstruira $e(x)$,*
4. **izbrani tajnopis**: *nasprotnik ima zača voljo odšifrirno mašinerijo ter za izbran konstruira $d(y)$.*

Odšifriranje Vigenerejevega tajnopisa

Test Kasiskega:

poiščemo dele tajnopisa, ki so identični in zabeležimo razdalje $d_1, d_2, \dots$ med njihovimi začetki. Predpostavimo, da m deli največji skupni delitelj.

Indeks naključja (Wolfe Friedman, 1920):

Naj bo $x = x_1x_2\dots x_n$ zaporedje n črk. Indeks naključja zaporedja x, označen z $I_c(x)$, je verjetnost, da sta dva naključna elementa zaporedja x enaka. Naj bodo $f_0, f_1, \dots, f_{25}$ frekvence črk $A, B, C, \dots, Z$ v zaporedju x. Potem je

$$I_c(x) = \sum_{i=0}^{25} \frac{f_i(f_i-1)}{n(n-1)}.$$

Če so p_i pričakovane verjetnosti angleških črk, potem je

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2 = 0.065.$$

Za povsem naključno zaporedje velja

$$I_c(x) \approx 26(1/26)^2 = 1/26 = 0.038.$$

Ker sta števili .065 in .038 dovolj narazen, lahko s to metodo najdemo dolžino ključa

(ali pa potrdimo dolžino, ki smo jo uganili s testom Kasiskega).

Naj bosta $x = x_1x_2\dots x_n$ in $y = y_1y_2\dots y_n$ z
n in n' črk. Vzajemen indeks naključja zapore
y, označen z $M_c(x,y)$, je verjetnost, da je n
element v x enak naključnemu elementu v y.
je

$$MI_c(x,y) = \sum_{i=0}^{25} \frac{f_i f_i'}{nn'}.$$

Po drugi strani pa je

$$M_c(x,y) \approx \sum_{h=0}^{25} p_{h-k_i} p_{h-k_j} = \sum_{h=0}^{25} p_h p_{h-s},$$

kjer je s relativen zamik $(k_i - k_j)$.

Izkaže se, da je $M_c(x,y) \approx 0.065$ za $s = 0$ in $MI_c(x,y) \in [0.031, 0.045]$ za $s \neq 0$.

Z računalnikom izračunamo 260 vrednosti $MI_c(y_i, y_j^s)$, kjer je $1 \le i < j \le 5$ in $0 \le s \le 25$, ter dobimo sistem enačb za $k_1, \dots, k_m$.

Odšifriranje Hillovega tajnopisa

Predpostavimo, da je nasprotnik določil m, ki ga uporabljamo, ter se dokopal do m različnih parov m-teric (2. stopnja – poznan čistopis):

$$x_j = (x_{1,j}, x_{2,j}, \dots, x_{m,j}),\ \ y_j = (y_{1,j}, y_{2,j}, \dots, y_{m,j}),$$

tako da je $y_j = e_K(x_j)$ za $1 \le j \le m$.

Za matriki $X = (x_{i,j})$ in $Y = (y_{i,j})$ dobimo matrično enačbo $Y = XK$.

Če je matrika X obrnljiva, je $K = YX^{-1}$.

Za Hillov tajnopis lahko uporabimo tudi 1. stopnjo napada (samo tajnopis), glej nalogo 1.6.

Koliko ključev imamo na voljo v primeru Hillovega tajnopisa? Glej nalogo 1.2.

Za afino-Hillov tajnopis glej nalogo 1.5.

Kriptoanaliza LFSR tokovnega tajnopisa: zopet lahko uporabimo poznan čistopis, glej nalogo 1.9.

Tokovni tajnopisi

Naj bo $x_1x_2\dots$ čistopis.

Doslej smo obravnavali kriptosisteme z enim
ključem in tajnopis je imel naslednjo obliko.

$$\mathbf{y} = y_1y_2 \cdots = e_K(x_1)e_K(x_2)\dots$$

Takemu tajnopisu pravimo **bločni tajnopis**
(block cipher).

Posplošitev: iz enega ključa $K \in \mathcal{K}$ napravimo zaporedje (tok) ključev. Naj bo f_i funkcija, ki generira i-ti ključ:

$$z_i = f_i(K, x_1, \dots, x_{i-1}).$$

Z njim izračunamo:

$$y_i = e_{z_i}(x_i) \qquad \text{in} \qquad x_i = d_{z_i}(y_i).$$

Bločni tajnopis je poseben primer tokovnega tajnopisa (kjer je $z_i = K$ za vse $i \geq 1$).

Tokovni tajnopis *je sedmerica* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{L}, \mathcal{F}, \mathcal{E}, \mathcal{D})$ *za katero velja:*

1. *$\mathcal{P}$ je končna množica možnih* **čistopisov**,
2. *$\mathcal{C}$ je končna množica možnih* **tajnopisov**,
3. *$\mathcal{K}$ je končna množica možnih* **ključev**,
4. *$\mathcal{L}$ je končna množica tokovne* **abecede**,
5. *$\mathcal{F} = (f_1, f_2, \dots)$ je generator toka ključev:*

$$f_i : \mathcal{K} \times \mathcal{P}^{i-1} \longrightarrow \mathcal{L} \quad \text{za } i \geq 1$$

6. *Za vsak ključ $z \in \mathcal{L}$ imamo šifrirni ($e_z \in \mathcal{E}$) in odšifrirni ($d_z \in \mathcal{D}$) postopek, tako da je $d_z(e_z(x)) = x$ za vsak $x \in \mathcal{P}$.*

Za šifriranje čistopisa $x_1 x_2 \dots$ zaporedno računamo

$$z_1, y_1, z_2, y_2, \dots,$$

za odšifriranje tajnopisa $y_1 y_2 \dots$ pa zaporedno računamo

$$z_1, x_1, z_2, x_2, \dots.$$

Tokovni tajnopis je **periodičen** s periodo d kadar, je $z_{i+d} = z_i$ za vsak $i \geq 1$

(poseben primer: Viginerejev tajnopis).

Začnimo s ključi $(k_1, \dots, k_m)$ in naj bo z_i = $i = 1, \dots, m$.
Odfiniramo linearno rekurzijo stopnje m:

$$z_{i+m} = z_i + \sum_{j=1}^{m-1} c_j \, z_{i+j} \mod 2,$$

kjer so $c_1, \dots, c_{m-1} \in \mathbb{Z}_2$ vnaprej določene ko

Za ustrezno izbiro konstant $c_1, \dots, c_{m-1} \in$ neničelen vektor $(k_1, \dots, k_m)$ lahko dobimo tajnopis s periodo $2^m - 1$.

Hitro lahko generiramo tok ključev z uporabo **LFSR** (**Linear Feedback Shift Register**).
V pomičnem registru začnemo z vektorjem

$$(k_1, \dots, k_m).$$

Nato na vsakem koraku naredimo naslednje:

1. k_1 dodamo toku ključev (za XOR),
2. $k_2, \dots, k_m$ pomaknemo za eno v levo,
3. 'nov' ključ k_m izračunamo z

$$\sum_{j=0}^{m-1} c_j \, k_{j+1} \quad \text{(to je “linear feedback”)}.$$

Primer:

$c_0 = 1$, $c_1 = 1$, $c_2 = 0$, $c_3 = 0$,

torej je $k_{i+4} = k_i + k_{i+1}$.

Izberimo $k_0 = 1$, $k_1 = 0$, $k_2 = 1$, $k_3 = 0$.

Potem je $k_4 = 1$, $k_5 = 1$, $k_6 = 0, \dots$.

Naj bo $\boldsymbol{k} = (k_0, k_1, k_2, k_3)^t$ in

$$A := \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Torej je $A(\boldsymbol{k}) = (k_1, k_2, k_3, k_4)^t$,
$A^2(\boldsymbol{k}) = A(k_1, k_2, k_3, k_4)^t = (k_2, k_3, k_4, k_5)^t$
...
$A^i(\boldsymbol{k}) = (k_i, k_{i+1}, k_{i+2}, k_{i+3})^t$.

Najdaljša možna perioda je 15.

Enkrat dobimo:

$$A^i(\boldsymbol{k}) = A^j(\boldsymbol{k})$$

in ker je A obrnljiva

$$A^{i-j}(\boldsymbol{k}) = \boldsymbol{k}$$

Karakteristični polinom matrike A je

$$f(x) = 1 + x + x^4.$$

Ker je $f(x)$ nerazcepen, je $f(x)$ tudi m polinom matrike A.

Red matrike A je najmanjše naravno število s, tako da je $A^s = I$. Naj bo e najmanjše naravno število, tako da $f(x) \,|\, (x^e - 1)$. Potem je $e = s$.

$$1 + x^{15} = (x+1)(x^2+x+1)(x^4+x+1)$$
$$(x^4+x^3+1)(x^4+x^3+x^2+x+1).$$

Splošno: če hočemo, da nam rekurzija stopnje m da periodo $2^m - 1$, potem si izberemo nerazcepen f.

Analiza je neodvisna od začetnega neničelnega vektorja.

2. poglavje

Shannonova teorija

- Popolna varnost
- Entropija
- Lastnosti entropije
- Ponarejeni ključi in enotska razdalja
- Produktni kriptosistemi

Popolna varnost

Omenimo nekaj osnovnih principov za študij varnosti nekega kriptosistema:

- **računska varnost**,
- **brezpogojna varnost**,
- **dokazljiva varnost**.

Kriptosistem je **računsko varen**, če tudi najboljši algoritem za njegovo razbitje potrebuje vsaj N operacij, kjer je N neko konkretno in zelo veliko število.

To je zelo podobno dokazovanju, da je problem NP-poln (dokažemo, da je izbrani problem vsaj tako zahteven kot kakšen drug NP-poln problem, to pa ne pomeni, da smo pokazali, da je absolutno računsko zahteven).

Kriptosistem je **dokazljivo varen** (angl. p
secure), če lahko pokažemo, da se njegova
zreducira na varnost kriptosistema, ki je zasn
dobro preštudiranem problemu.

Ne gre torej za absolutno varnost temveč *r*
varnost.

Gre za podobno strategijo kot pri dokazovanj
določen problem *NP-poln* (v tem primeru do
da je dani problem vsaj tako težak kot nekdru
NP-poln problem, ne podamo pa absolutnega
da gre za računsko težak problem).

Kriptosistem je **brezpogojno varen**, kadar ga napadalec ne more razbiti, tudi če ima na voljo neomejeno računsko moč.

Seveda je potrebno povedati tudi, kakšne vrste napad imamo v mislih. Spomnimo se, da zamične, substitucijske in Vigenere šifre niso varne pred napadom s poznanim tajnopisom (če imamo na voljo dovolj tajnopisa).

Razvili bomo teorijo kriptosistemov, ki so brezpogojno varni pri napadu s poznanim tajnopisom. Izkaže se, da so vse tri šifre brezpogojno varne, kadar zašifriramo le en sam element čistopisa.

Glede na to, da imamo pri brezpogojni varnosti na voljo neomejeno računsko moč, je ne moremo študirati s pomočjo teorije kompleksnosti, temveč s teorijo verjetnosti.

Naj bosta X in Y slučajni spremenljivki,
naj bo $p(x) := P(X = x)$, $p(y) := P(Y = y)$ in
$p(x \cap y) := P((X = x) \cap (Y = y))$ produkt dogodkov.

Slučajni spremenljivki X in Y sta **neodvisni**, če in samo, če je $p(x \cap y) = p(x)p(y)$ za vsak $x \in X$ in $y \in Y$.

Omenimo še zvezo med pogojno verjetnostjo in pa verjetnostjo produkta dveh dogodkov oziroma **Bayesov izrek o pogojni verjetnosti**:

$$p(x \cap y) = p(x/y)p(y) = p(y/x)p(x),$$

iz katerega sledi, da sta slučajni spremenljivki X in Y neodvisni, če in samo, če je $p(x/y) = p(x)$ za vsak x in y.

Privzemimo, da vsak ključ uporabimo za največ eno enkripcijo, da si Anita in Bojan izbereta ključ K z neko fiksno verjetnostno porazdelitvijo $p_{\mathcal{K}}(K)$ (pogosto enakomerno porazdelitvijo, ni pa ta nujna) in naj bo $p_{\mathcal{P}}(x)$ verjetnost čistopisa x.

Končno, predpostavimo, da sta izbira čisto
ključa neodvisna dogodka.

Porazdelitvi $\mathcal{P}$ in $\mathcal{K}$ inducirata verjetnostno
porazdelitev na $\mathcal{C}$. Za množico vseh tajnopiso
za ključ K

$$C(K) = \{e_K(x) \,|\, x \in \mathcal{P}\}$$

velja

$$p_{\mathcal{C}}(y) = \sum_{\{K \,|\, y \in C(K)\}} p_{\mathcal{K}}(K) p_{\mathcal{P}}(d_k(y))$$

in

$$P(Y = y/X = x) = \sum_{\{K \,|\, x = d_K(y)\}} p_{\mathcal{K}}(K$$

Sedaj lahko izračunamo pogojno verjetnost $p_{\mathcal{P}}(x/y)$, tj. verjetnost, da je x čistopis, če je y tajnopis

$$P(X=x/Y=y) = \frac{p_{\mathcal{P}}(x) \times \displaystyle\sum_{\{K \mid x=d_K(y)\}} p_{\mathcal{K}}(K)}{\displaystyle\sum_{\{K \mid y \in C(K)\}} p_{\mathcal{K}}(K)\, p_{\mathcal{P}}(d_k(y))}$$

in opozorimo, da jo lahko izračuna vsakdo, ki pozna verjetnostni porazdelitvi $\mathcal{P}$ in $\mathcal{K}$.

Primer: $\mathcal{P}=\{a,b\}$ in $\mathcal{K}=\{K_1,K_2,K_3\}$:

$$p_{\mathcal{P}}(a)=1/4 \quad \text{in} \quad p_{\mathcal{P}}(b)=3/4.$$

$$p_{\mathcal{K}}(K_1)=1/2 \quad \text{in} \quad p_{\mathcal{K}}(K_2)=p_{\mathcal{K}}(K_3)=1/4.$$

Enkripcija pa je definirana z $e_{K_1}(a)=1$, $e_{K_1}(b)=2$; $e_{K_2}(a)=2$, $e_{K_2}(b)=3$; $e_{K_3}(a)=3$, $e_{K_3}(b)=4$.

Potem velja

$$p_{\mathcal{C}}(1)=\frac{1}{8},\quad p_{\mathcal{C}}(2)=\frac{7}{16},\quad p_{\mathcal{C}}(3)=\frac{1}{4},\quad p_{\mathcal{C}}(3)=\frac{3}{16}.$$

$$p_{\mathcal{P}}(a/1)=1,\; p_{\mathcal{P}}(a/2)=\frac{1}{7},\; p_{\mathcal{P}}(a/3)=\frac{1}{4},\; p_{\mathcal{P}}(a/4)=0.$$

Kriptosistem $(\mathcal{P},\mathcal{K},\mathcal{C})$ ima **popolno varnost**, če je

$$P(X=x/Y=y)=p_{\mathcal{P}}(x) \quad \text{za vse} \quad x\in\mathcal{P} \quad \text{in} \quad y\in\mathcal{C},$$

tj. "končna" verjetnost, da smo začeli s tajnopisom x pri danem čistopisu y, je identična z "začetno" verjetnostjo čistopisa x.

V prejšnjem primeru je ta pogoj zadoščen samo v primeru $y=3$, ne pa tudi v preostalih treh.

Izrek 1. *Če ima vseh 26 ključev pri šifri enako verjetnost $1/26$, potem ima z verjetnostno porazdelitev čistopisa zamic popolno varnost.*

Dokaz: $\mathcal{P}=\mathcal{C}=\mathcal{K}=\mathbb{Z}_{26}$, $e_K(x)=x+K$

$$p_{\mathcal{C}}(y)=\frac{1}{26}\sum_{K\in\mathbb{Z}_{26}} p_{\mathcal{P}}(y-K)=\frac{1}{26},$$

$$P(Y=y/X=x)=p_{\mathcal{K}}(y-x \bmod 26)=$$

Torej lahko zaključimo, da zamične šifre ne razbiti, če za vsak znak čistopisa uporabi naključno izbran ključ.