Criptosisteme "perfect sigure". Teorema lui Shannon

Până cum am discutat despre câteva criptosisteme cu importanţă mai mult istorică. Am văzut că pentru toate există metode de spargere şi deci sunt nesigure. Există oare criptosisteme sigure (măcar) din punct de vedere probabilistic? În 1948 şi 1949, Claude Shannon a publicat în Bell System Technical Journal două articole ce constituie fundamentul matematic al criptografiei moderne:

• *A mathematical theory of communication*, Bell System Technical Journal, iulie 1948, 28, 379-423: introduce un model matematic pentru comnunicarea prin canale cu "zgomote".

• *Communication theory of secrecy systems*, Bell System Technical Journal, octombrie 1949: folosind ideile din 1948 Shannon pune bazele matematice ale analizei sistemelor de criptare.

Printre aletele, în aceste lucrări el defineşte conceptul de secret perfect (sau necondiţionat), dă primele demonstraţii ale securităţii sistemelor criptografice folosind teoria probabilităţilor şi arată legăturile precise dintre securitatea sistemelor criptografice şi cardinalele mulţimilor cheilor, textelor simple şi ale textelor cifrate pe care sunt date anumite distribuţii ale probabilităţilor. Cea mai importantă realizare a sa a fost demonstrarea existenţei sistemelor criptografice perfect sigure. Intuitiv un sistem este perfect sigur dacă atacatorul (Oscar), care a interceptat un text cifrat, deşi deţine o putere computaţională foarte mare, nu poate obţine informaţii despre textul necifrat. Din păcate astfel de sisteme nu sunt foarte eficiente.

Să vedem mai întâi care este scenariul.

Presupunem că Alice şi Bob comunică printr-un canal nesigur în prezenţa lui Oscar, care ascultă. Ei se hotărăsc să folosească un criptosistem simetric în care cheia de decriptare este aceeaşi cu cheia de criptare. Oscar, atacatorul, ştie mesajul cifrat şi încearcă să obţină din textul criptat, informaţii despre textul simplu necifrat. Conform teoriei lui Shannon un criptosistem este *perfect sigur* dacă atacatorul (Oscar) nu poate obţine informaţii despre mesajul simplu, din mesajele cifrate pe care le interceptează.

Vrem să formalizam matematic aceste lucruri.

• Alice şi Bob folosesc un criptosistem (simetric) (M, C, K, E, D) care satisface condiţia

$$\text{pentru orice cheie } k \in K \text{ şi orice mesaj } m \in M, \text{ avem } d_k(e_k(m)) = m,$$

unde e_k şi d_k, $k \in K$, sunt funcţiile de criptare şi respectiv de decriptare.

• Presupunem că pe spaţiul textelor simple este definită o distribuţie a probabilităţii p_M (depinde de limba folosită dar nu şi de schema de criptare), astfel încât $p_{_M}(m) > 0$, pentru orice $m \in M$. Se presupune că Oscar cunoaşte $p_{_M}$ pentru că ştie limba folosită de Alice şi Bob.

• Pentru fiecare mesaj nou ce va fi criptat Alice alege o nouă cheie din K, independentă de textul simplu ce va fi criptat. Mulţimea K depinde de criptosistem. Presupunem că în K avem o distribuţie a probabilităţii $p_{_K}$ şi că $p_{_K}(k) > 0$, pentru orice $k \in K$.

• Probabilităţile $p_{_M}$ şi $p_{_K}$ induc o distribuţie a probabilităţii $p_{_{M\times K}}$ pe spaţiul produs $M\times K$. Astfel, pentru orice mesaj necifrat şi orice cheie $k\in K$ probabilitatea ca textul simplu m să fie criptat cu cheia k este dată de

$$p_{_{M\times K}}(m,k)=p_{_M}(m)\cdot p_{_K}(k).$$

• Presupunem acum că pe spaţiul textelor criptate avem de asemenea definită o distribuţie a probabilităţii astfel încât $p_{\mathcal{C}}(c)>0$ pentru orice text criptat $c\in\mathcal{C}$.

• Pentru un text simplu $m_o\in M$, definim evenimentul $\overline{m}_o$ prin

$$\overline{m}_o=\{(m_o,k)|k\in K\}\subset M\times K.$$

Observaţia 1. *Este clar că avem*

$$p_{_{M\times K}}(\overline{m}_o)=\sum_{(m_o,k)\in\overline{m}_o}p_{_{M\times K}}((m_o,k))=\sum_{k\in K}p_{_M}(m_o)\;p_{_K}(k)=p_{_M}(m_o)\sum_{k\in K}p_{_K}(k)=p_{_M}(m_o),$$

adică probabilitatea ca să fie criptat textul simplu m_o.

• Pentru orice cheie $k_o\in K$, definim evenimentul $\overline{k}_o$ prin

$$\overline{k}_o=\{(m,k_o)|m\in M\}\subset M\times K.$$

Observaţia 2. *Exact ca mai sus, se demonstrază că $p_{_{M\times K}}(\overline{k}_o)=p_{_K}(k_o)$, adică probabilitatea ca să fie aleasă cheia k_o pentru criptare.*

Observaţia 3. *Evenimentele $\overline{m}_o$ şi $\overline{k}_o$ sunt independente deoarece*

$$p(\overline{m}_o\cap\overline{k}_o)=p((m_o,k_o))=p(m_o)p(k_o)=p(\overline{m}_o)p(\overline{k}_o).$$

• Pentru un text cifrat $c_o\in C$ definim evenimentul $\overline{c}_o$ prin

$$\overline{c}_o=\{(m,k)\in M\times K|e_k(m)=c_o\}\subset M\times K,$$

adică evenimentul când rezultatul cifrării este c_o.

Aşa cum am văzut, Oscar cunoaşte probabilitatea $p_{_M}$ pe spaţiul textelor simple, pentru că ştie limba folosită de Alice şi Bob. Presupunem că Oscar interceptează un mesaj cifrat c. Dacă apariţia lui c nu face ca, ţinând cond şi de probabilitatea $p_{_M}$, unele texte simple să fie privilegiate faţă de altele, înseamnă că Oscar nu poate învăţa nimic din c. Acest lucru ne conduce la următoarea definiţie

Definiţia 1. *Un criptosistem (M,C,K,E,E) se numeşte perfect secret (sau perfect sigur) dacă pentru orice text simplu $m\in M$ şi orice text cifrat $c\in C$ avem*

$$p(\overline{m}|\overline{c})=p(\overline{m}).$$

Observaţia 4. *Este clar că un sistem criptografic este perfect secret dacă şi numai dacă pentru orice text simplu $m\in M$ şi orice text cifrat $c\in C$, evenimentele $\overline{m}$ şi $\overline{c}$ sunt independente, adică $p(\overline{m}\cap\overline{c})=p(\overline{m})\cdot p(\overline{c})$.*

Exemplul 1. *(un cifru de jucărie) Fie $M=\{0,1\}$ spaţiul textelor simple, $K=\{A,B\}$ spaţiul cheilor, peste care sunt date probabilităţile: p_M definită prin $p_M(0)=1/4$, $p_M(1)=3/4$ şi p_K dată prin $p_K(A)=1/4$, $p_K(B)=3/4$. Fie $C=\{a,b\}$ spaţiul textelor cifrate. Funcţiile de cifrare sunt date prin*

$$e_{_A}(0)=a,\;\; e_{_A}(1)=b,\;\; e_{_B}(0)=b,\;\; e_{_B}(1)=a.$$

Vrem să vedem dacă sistemul este perfect sigur.

Calculăm $p(\overline{1}|\overline{a}) = \dfrac{p(\overline{1} \cap \overline{a})}{p(\overline{a})}$.

Probabilitatea ca textul cifrat să fie a este

$$\begin{aligned} p(\overline{a}) &= p(\{(m,k) \in M \times K | e_k(m) = a\}) = \\ &= p(0,A) + p(1,B) = \\ &= p(0)p(A) + p(1)p(B) = \\ &= \frac{1}{4} \times \frac{1}{4} + \frac{3}{4} \times \frac{3}{4} = \frac{5}{8}. \end{aligned}$$

Avem

$$\overline{1} \cap \overline{a} = \{(1,A),(1,B)\} \cap \{(0,A),(1,B)\} = \{(1,B)\}$$

şi deci

$$p(\overline{1} \cap \overline{a}) = p(1) \cdot p(B) = \frac{9}{16}.$$

Deci

$$p(\overline{1}|\overline{a}) = \frac{9}{16} \cdot \frac{8}{5} = \frac{9}{10}$$

de unde avem

$$p(\overline{1}|\overline{a}) = \frac{9}{10} \neq \frac{3}{4} = p_M(\overline{1}),$$

adică sistemul nu este perfect sigur. Dacă Oscar interceptează textul cifrat a atunci e aproape sigur că el corespunde textului simplu 1*. Prin calcule asemanătoare se obţine* $p(\overline{0}|\overline{a}) = \dfrac{1}{10}$.

Observaţia 5. *Oare ce ar trebui să modificăm în exemplul precedent pentru a obţine un sistem perfect sigur? (*$p(A) = p(B) = \frac{1}{2}$*, adică să avem o distribuţie uniformă pe spaţiul cheilor)*

Exemplul 2. (Criptosostemul Vernam One-Time Pad)

Cel mai faimos criptosistem perfect secret este One-Time Pad (cheia se foloseşte doar o singură dată) care a fost inventat şi patentat de Gilbert Vernam în 1917 dar abia în 1949 Shannon a demonstrat că acesta este perfect secret. Deşi acest sistem este perfect sigur, el nu este foarte eficient din punct de vedere practic pentru că lungimea cheii este egală cu lungimea mesajului iar o cheie trebuie folosită doar o singură dată şi deci Alice şi Bob ar fi nevoiţi să transfere o cantitate mare de informaţie (cheia) printr-un canal nesigur, sau să se întâlnească periodic pentru a hotărîcheile ce vor fi folosite în viitor. Ne propunem în continuare să dscriem acest sistem de criptare şi să dovedim că este perfect secret.

Fie n *un număr întreg. Definim* $M = C = K = (\mathbb{Z}/2\mathbb{Z})^n$*. Pentru* $k \in (\mathbb{Z}/2\mathbb{Z})^n$*, funcţiile de cifrare sunt definite prin*

$$e_k : (\mathbb{Z}/2\mathbb{Z})^n \to (\mathbb{Z}/2\mathbb{Z})^n,\ e_k(m) = m + k\,(mod\ 2).$$

iar funcţiile de decifrare prin

$$d_k : (\mathbb{Z}/2\mathbb{Z})^n \to (\mathbb{Z}/2\mathbb{Z})^n,\ d_k(c) = c + k\,(mod\ 2).$$

Presupunem că pe spaţiul cheilor avem distribuţia uniformă a probabilităţii, adică $p(k) = \frac{1}{2^n}$ *pentru orice* $k \in K$*.*

Pentru a cifra un text simplu $m \in (\mathbb{Z}/2\mathbb{Z})^n$, Alice alege o cheie la întâmplare (cu distribuţie uniformă) în K şi calculează textul cifrat prin $c = m + k$ (mod 2).

Pentru a arăta că sistemul este perfect sigur trebuie să verificăm condiţia

$$\forall m \in M \text{ şi } \forall c \in C \text{ avem } p(\overline{m}|\overline{c}) = p(\overline{m}),$$

adică evenimentele $\overline{m}$ şi $\overline{c}$ sunt independente.

Din definiţie $p(\overline{m}|\overline{c}) = p(\overline{m} \cap \overline{c})/p(\overline{c})$. Pe de altă parte $p(\overline{m} \cap \overline{c}) = p(\overline{m})p(\overline{c}|\overline{m})$.

Calculăm acum $p(\overline{c}|\overline{m})$. Dacă vrem să criptăm mesajul m, singurul mod de a obţine textul cifrat c este să folosim exact cheia $k = c+m$ (mod 2). Cum $p(k) = \frac{1}{2^n}$ pentru orice $k \in (\mathbb{Z}/2\mathbb{Z})^n$, rezultă $p(\overline{c}|\overline{m}) = \frac{1}{2^n}$ şi deci $p(\overline{m} \cap \overline{c}) = p(\overline{m})\frac{1}{2^n}$.

Calculăm acum $p(\overline{c})$. Reamintim că $\overline{c} = \{(m, k) \in M \times K | E_k(m) = c\}$ şi deci

$$p(\overline{c}) = \sum_{m \in (\mathbb{Z}/2\mathbb{Z})^n, m+k=c} p(k)p(m) = \frac{1}{2^n} \sum_{m \in (\mathbb{Z}/2\mathbb{Z})^n} p(m) = \frac{1}{2^n}.$$

În sfârşit din cele de mai sus obţinem $p(\overline{m}|\overline{c}) = p(\overline{m})$, adică sistemul satisface condiţia din definiţia sistemelor perfect sigure. Din păcate, dacă Alice şi Bob vor să folosească sistemul Vernam "one-time pad" pentru a schimba n biţi de informaţie, ei trebuie să ştie deja n biţi de informaţie secretă, adică cheia. Acest lucru face sistemul Vernam foarte ineficient pentru comunicaţii în reţele foarte mari. Totuşi, există situaţii când sistemul a fost folosit. De exemplu în comunicaţii strict secrete la nivel diplomatic (pe linia roşie Washington-Kremlin în timpul războiului rece) sau pentru mesaje foarte scurte trimise între spioni şi cartierul lor general (spionul sovietic R.I.Abel folosea în mesajele sale cifrate de la Washington la Moscova această cifrare iar mesajele sale n-au putut fi descifrate de americani). Este evident că sistemul rămâne sigur atâta timp cât cheile nu se repetă. O cheie trebuie sa fie folosită o singura dată (de aceea se mai numeşte "one-time pad")! Să observăm că dacă cheia ar fi folosită de mai multe ori şi Oscar ştie un text simplu m şi corespondentul său cifrat c, atunci el deduce cheia folosită: $m + c = m + m + k = k$ (mod 2). O astfel de greşeală a fost făcută de Uniunea Sovietică în timpul celui de-al doilea Război Mondial.

Din definiţia sistemelor criptografice ştim că funcţiile de criptare $e_k : M \to C$ sunt injective şi deci numărul textelor cifrate trebuie să fie cel puţin la fel de mare ca numărul textelor simple (card $(M) \leq$ card (C)), pentru că altfel decriptarea nu ar fi posibilă. Vom vedea acum că o consecinţă a siguranţei perfecte este faptul că numărul cheilor trebuie să fie de asemenea cel puţin la fel de mare ca numărul textelor simple.

Propoziţia 1. *Fie (M, C, K, E, D) un criptosistem perfect secret. Atunci card$(K) \geq$ card(M).*

Demonstraţie: Fie c_o un text cifrat fixat. Arătăm că pentru orice text simplu $m \in M$, există o cheie $k \in K$ astfel încât $e_k(m) = c_o$.

Presupunem prin absurd că acest fapt nu este adevărat şi că există un text simplu m_o astfel încât pentru orice $k \in K$, $e_k(m_o) \neq c_o$. Atunci $\overline{m}_o \cap \overline{c}_o = \{(m_o, k)|e_k(m_o) = c_o\} = \emptyset$ şi deci $p(\overline{m}_o \cap \overline{c}_o) = 0$, adică $p(\overline{m}_o|\overline{c}_o) = p(\overline{m}_o \cap \overline{c}_o)/p(\overline{c}_o) = 0$. Dar $p(m_o) > 0$ ceea ce contrazice faptul că sistemul este perfect secret.

Fie acum $K = \{k_1, \ldots k_s\}$ şi $M = \{m_1, \ldots, m_t\}$ şi presupunem că $s < t$. Atunci există textele simple $m_i \neq m_j$ astfel încât $e_k(m_i) = e_k(m_j) = c_o$, pentru o anumită cheie $k \in K$. Aplicând acum funcţia d_k ambilor membrii obţinem $m_i = m_j$, contradicţie. Rezultă că $s \geq t$, adică exact ceea ce trebuia să demonstrăm. ■

Având restricţii asupra cardinalelor spaţiilor textelor simple şi cifrate precum şi asupra spaţiului cheilor, adică card$(K)\geq$card(P) şi card$(C)\geq$card(P), nu este nenatural să presupunem că cele trei spaţii au acelasi număr de elemente. Impunând această condiţie, Shannon a demonstrat o teoremă de caracterizare a criptosistemelor perfect secrete.

Teorema 1. *(Shannon) Fie (M, C, K, E, D) un criptosistem cu card(K) = card(M) = card(C). Presupunem că pe fiecare din cele trei spaţii avem distribuţii nenule ale probabilităţilor. Atunci sistemul este perfect secret dacă şi numai dacă următoarele două condiţii sunt satisfăcute:*

1. *Orice cheie $k \in K$ este aleasă cu aceeaşi probabilitate, adică distribuţia probabilităţilor pe spaţiul cheilor este cea uniformă.*
2. *Pentru orice text simplu $m \in M$ şi orice text cifrat $c \in C$ există o unică cheie $k \in K$ astfel ca $e_k(m) = c$.*

Demonstraţie:

• Presupunem mai întâi că schema de criptare este perfect secretă.

Demonstrăm afirmaţia a doua.

Aşa cum am văzut în demonstraţia Propoziţiei 1, pentru fiecare text cifrat c şi pentru orice text simplu m, există o cheie $k \in K$ astfel ca $e_k(m) = c$, adică am probat existenţa. Să arătăm şi unicitatea cheii. Fixăm un text simplu necifrat m_o şi definim funcţia $f : K \to C$ prin $f(k) = e_k(m_o)$. Funcţia f este evident surjectivă. Dar, din ipoteză, numărul de chei este egal cu numărul de texte cifrate, adică cardinalele domeniului de definiţie şi al codomeniului sunt egale. Acest fapt implică şi injectivitatea funcţiei, aşa încât avem şi unicitatea.

Acum probăm prima afirmaţie. Fixăm din nou un text cifrat c_o. Fie $n =$ card (K) şi $M = \{m_1, \ldots, m_n\}$. Reamintim că card(K) = card(M). Etichetăm cheile $\{k_1, \ldots, k_n\}$ astfel încât pentru orice $i = \overline{1,n}$ avem $e_{k_i}(m_i) = c_o$. Indexarea poate fi făcută datorită punctului (2), adică prin fixarea lui c_o, avem că pentru orice m există o unică cheie k astfel încât $e_k(m) = c_o$, şi deci pentru fiecare m_i asociem cheia k_i. Să remarcăm că această indexare acoperă toate cheile. Intr-adevăr, dacă pentru două mesaje diferite $m_i \neq m_j$, ar exista o aceeaşi cheie k astfel încât $e_k(m_i) = e_k(m_j) = c_o$, decriptarea nu s-ar mai putea face (funcţia e_k este injectivă). Deoarece sistemul este perfect secret, folosind Teorema lui Bayes precum şi corespondenţa de mai sus dintre texte simple şi chei, vom avea:

$$p(\overline{m}_i) = p(\overline{m}_i|\overline{c}_0) = \frac{p(\overline{c}_0|\overline{m}_i)p(\overline{m}_i)}{p(\overline{c}_o)} = \frac{p(\overline{k}_i)p(\overline{m}_i)}{p(\overline{c}_o)}.$$

Cum pentru orice text simplu m avem $p(m) > 0$, rezultă că $p(k_i) = p(c_0)$ pentru orice i şi deci distribuţia probabilităţii pe spaţiul cheilor este uniformă.

• Reciproc, presupunem că sunt adevărate condiţiile (1) şi (2) şi arătăm că sistemul este perfect secret.

Pentru un text simplu m şi un text cifrat c, fie $k = k(m, c)$ unica cheie astfel încât $e_k(m) = c$. Folosind din nou teorema lui Bayes avem

$$p(\overline{m}|\overline{c}) = \frac{p(\overline{m})p(\overline{c}|\overline{m})}{p(\overline{c})} = \frac{p(\overline{m})p(\overline{k}(m,c))}{\sum\limits_{q\in M} p(\overline{q})p(\overline{k}(q,c)).}$$

Pe de altă parte, cheile sunt uniform distribuite în K şi deci $p(\overline{k}(m, c)) = 1/\text{card}(K)$. În plus avem

$$\sum_{q\in M} p(\overline{q})p(\overline{k}(q, c)) = \frac{1}{\text{card } K} \sum_{q\in M} p(\overline{q}) = \frac{1}{\text{card } M}.$$

Obţinem astfel că $p(\overline{m}|\overline{c}) = p(\overline{m})$ pentru orice m şi c, adică sistemul este perfect secret. ■