Teorija kodiranja in kriptografija 2013/2014

Načini uporabe bločnih šifer in varnost

Arjana Žitnik

Univerza v Ljubljani, Fakulteta za matematiko in fiziko

Ljubljana, 15. 4. 2014

Vsebina

- Načini uporabe bločnih šifer
- Psevdonaključne permutacije
- Semantična varnost

Načini uporabe bločnih šifer

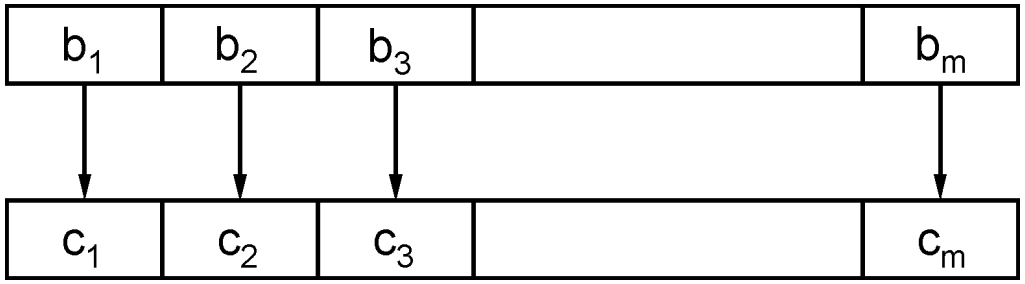
V praksi se uporablja več načinov, kako besedilo poljubne dolžine kodirati z bločno šifro dolžine n:

- **elektronska kodna knjiga** (electronic codebook (ECB))
- **veriženje kodnih blokov** (cipher block chaining (CBC))
- **Način s števcem** (counter mode (CM))
- **izhodna povratna zveza** (output feedback mode (OFB))
- **kodna povratna zveza** (cipher feedback mode (CFB))

Besedilo $b = b_1 b_2 \ldots b_m$ razdelimo na bloke dolžine n, kodiramo vsak blok posebej.

Elektronska kodna knjiga (ECB)

"Naivni" način uporabe bločnih šifer. Z istim ključem kodiramo zaporedoma blok po blok:



Kodiranje: $c_i = E_k(b_i)$, $i = 1, 2, \ldots, m$.
Dekodiranje: $b_i = D_k(c_i)$, $i = 1, 2, \ldots, m$.

Prednosti: napaka pri prenosu enega bloka ne vpliva na dekodiranje naslednjih blokov.
Slabost: identični bloki besedila se (pri istem ključu) kodirajo v identične bloke kriptograma, kar napadalec lahko izkoristi.

Veriženje kodnih blokov (CBC)

Izberemo inicializacijski vektor IV dolžine n (lahko je javno dostopen).

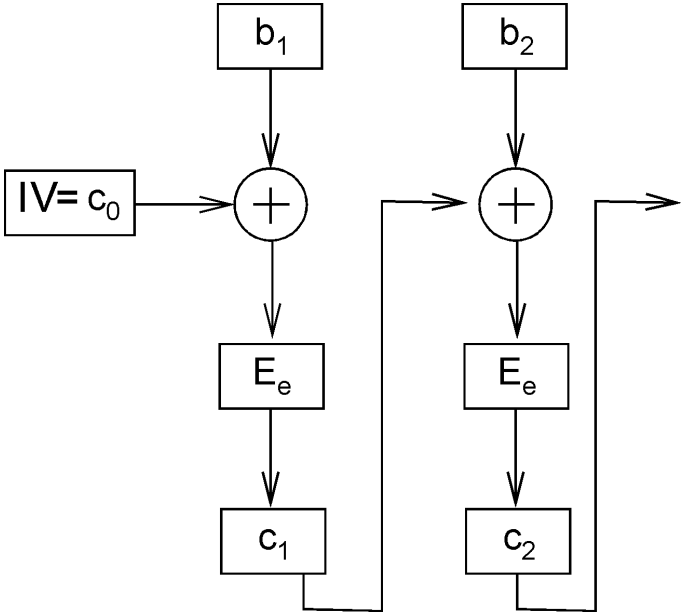
Kodiranje:

$c_0 = \mathrm{IV}$,

za $j = 1, 2, \ldots m$ **ponovi**

$\qquad c_j := E_e(b_j \oplus c_{i-1})$

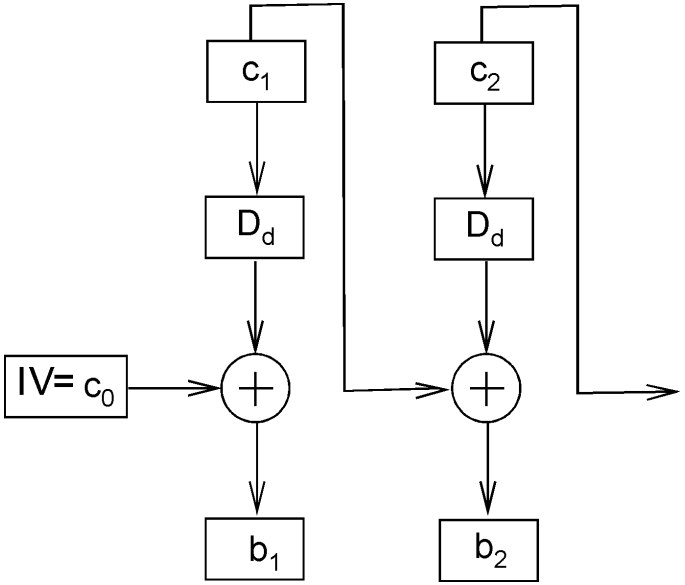
$c = c_1 c_2 \ldots c_m$



CBC - dekodiranje

Dekodiranje:

$c_0 = \mathrm{IV}$,
za $j = 1, 2, \ldots m$
ponovi
$b_j := D_d(c_j \oplus c_{j-1})$
$b = b_1 b_2 \ldots b_m$



Prednosti:

- iz $b_i = b_j$ ne sledi nujno $c_i = c_j$, saj je c_j odvisen od $b_j, b_{j-1}, \ldots, b_1$ in IV.
- zamenjava b_i z b_i' vpliva na $c_i, c_{i+1} \ldots$: zaznamo spremembe sporočil (možnost uporabe za avtentikacijo).
- napake pri prenosu imajo omejen učinek: c_j vpliva le na b_j in b_{j+1}.

Način s števcem (CM)

- Izberemo števec *ctr* dolžine n
- Besedilo razdelimo na bloke dolžine n: $b = b_1 b_2 \ldots b_m$.

Kodiranje:

za $j = 1, 2, \ldots m$ **ponovi**

$$I_j = ctr + j - 1 \pmod{2^n}$$
$$c_j = b_j \oplus E_e(I_j)$$

$c = c_1 c_2 \ldots c_m$

Dekodiranje:

za $j = 1, 2, \ldots m$ **ponovi**

$$I_j = ctr + j - 1 \pmod{2^n}$$
$$b_j = c_j \oplus E_e(I_j)$$

$b = b_1 b_2 \ldots b_m$

Lastnosti načina CM

Prednosti:

- napaka pri prenosu enega bloka ne vpliva na dekodiranje naslednjih blokov,
- hitrost dekodiranja: računanje E_e in D_d je lahko časovno zahtevno, računanje $\oplus$ pa ne; prejemnik lahko zaporeje t_j izračuna vnaprej in tako hitreje dekodira.
- zaporedje ključev $E_e(I_j)$ ni definirano rekurzivno, ampak eksplicitno. Zato lahko CM na vzporedni strojni/programski opremi implementiramo zelo hitro (vse člene zaporedja ključev izračunamo hkrati).

Slabosti:

- c_j odvisen le od števca, b_j in j, kar nasprotniku olajša spreminjanje kriptograma,
- za kodiranje dveh besedil b in b' z istim ključem je potrebno uporabiti drug števec, sicer lahko iz enega besedila izračunamo drugega.

Izhodna povratna zveza (OFB)

- Izberemo inicializacijski vektor IV dolžine n
- Izberemo $r \in \{1, 2, \ldots, n\}$
- Besedilo razdelimo na bloke dolžine r: $b = b_1 b_2 \cdots_t$.

Kodiranje:

$I_0 = \mathrm{IV}$,

za $j = 1, 2, \ldots t$ **ponovi**

$I_j = E_e(I_{j-1})$

$t_j =$ prvih r bitov I_j

$c_j = b_j \oplus t_j$

$c = c_1 c_2 \ldots c_m$

Izhodna povratna zveza - zgled

$b = 101100010100101$
$n = 4$, $r = 3$ IV $= 1010$
Kodirna funkcija: permutacijska šifra s ključem

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$$

j	I_j	t_j	b_j	c_j
0	1010			
1	0011	001	101	100
2	0101	010	100	110
3	1100	110	010	100
4	1010	101	100	001
5	0011	001	101	100

$$c = 100110100001100$$

Izhodna povratna zveza - dekodiranje

Dekodiranje je enako kodiranju.

Dekodiranje:

$I_0 = \mathrm{IV}$,

za $j = 1, 2, \ldots t$ **ponovi**

$I_j = E_e(I_{j-1})$

$t_j =$ prvih r bitov I_j

$b_j = c_j \oplus t_j$

$b = b_1 b_2 \ldots b_m$

Dekodiranje - zgled

$$c = 100110100001100$$

j	l_j	t_j	c_j	b_j
0	1010			
1	0011	001	100	101
2	0101	010	110	100
3	1100	110	100	010
4	1010	101	001	100
5	0011	001	100	101

$$b = 101100010100101$$

Lastnosti načina OFB

Podobno kot pri načinu CM.

Prednosti:

- napaka pri prenosu enega bloka ne vpliva na dekodiranje naslednjih blokov,
- hitrost dekodiranja: računanje E_e in D_d je lahko časovno zahtevno, računanje $\oplus$ pa ne; prejemnik lahko zaporeje t_j izračuna vnaprej in tako hitreje dekodira (uporaba pri satelistskih prenosih).

Slabosti:

- c_j odvisen le od IV, b_j in j, kar nasprotniku olajša spreminjanje kriptograma,
- za kodiranje dveh besedil b in b' z istim ključem je potrebno uporabiti drug IV, sicer lahko iz enega besedila izračunamo drugega:

$$c_j \oplus c'_j = b_j \oplus t_j \oplus b'_j \oplus t'_j = b_j \oplus b'_j.$$

Torej: $b'_j = c_j \oplus c'_j \oplus b_j$.

Kodna povratna zveza (CFB)

- Izberemo inicializacijski vektor IV dolžine n
- Izberemo $r \in \{1, 2, \ldots, n\}$
- Besedilo razdelimo na bloke dolžine r: $b = b_1 b_2 \ldots b_t$.

Kodiranje:

$I_0 = \mathrm{IV}$,

za $j = 1, 2, \ldots t$ **ponovi**

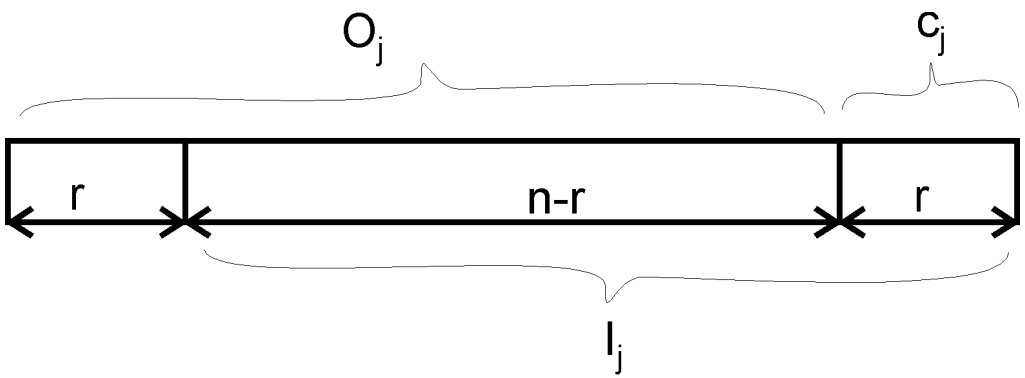
$O_j = E_e(I_{j-1})$

$t_j =$ prvih r bitov O_j

$c_j = b_j \oplus t_j$

$I_j \equiv O_j \cdot 2^r \oplus c_j \pmod{2^n}$

$c = c_1 c_2 \ldots c_m$



Kodna povratna zveza - dekodiranje

Dekodiranje:

$I_0 = \mathrm{IV},$

za $j = 1, 2, \ldots t$ **ponovi**

$O_j = E_e(I_{j-1})$

$t_j =$ prvih r bitov O_j

$b_j = c_j \oplus t_j$

$I_j \equiv O_j \cdot 2^r \oplus c_j \pmod{2^n}$

$b = b_1 b_2 \ldots b_m$

Lastnosti načina CFB

Prednost:

- c_j odvisen od $b_j, b_{j-1}, \ldots, b_1$ in IV.

Slabost:

- napaka pri prenosu c_j vpliva na $b_j, b_{j+1}, \ldots, b_{j+\lceil n/r \rceil}$

Izbira r:

- majhen r: več klicev kodirne funkcije, a hitrejši prenos posameznega bloka
- velik r: manj klicev kodirne funkcije, a počasnejši prenos posameznega bloka

Psevdo naključna funkcija

Psevdo naključna funkcija nad $(\mathcal{B}, \mathcal{C}, \mathcal{K})$ je funkcija

$$F : \mathcal{K} \times \mathcal{B} \to \mathcal{C},$$

ki jo lahko učinkovito izračunamo.

Psevdo naključna permutacija

Psevdo naključna permutacija nad $(\mathcal{B}, \mathcal{K})$ je funkcija

$$E : \mathcal{K} \times \mathcal{B} \to \mathcal{B},$$

za katero velja

1. obstaja učinkovit determinističen algoritem, ki izračuna $E(k, x)$,
2. funkcija $E(k, \cdot)$ je bijektivna,
3. učinkovito lahko izračunamo inverz $D(k, y)$.

Opomba. Psevdo naključna permutacija je tudi psevdo naključna funkcija. Psevdonaključna funkcija je psevdo naključna permutacija, če velja $\mathcal{B} = \mathcal{C}$ in lahko učinkovito izračunamo njen inverz.

Varne psevdo naključne funkcije

- Vseh funkcij $\mathcal{B} \to \mathcal{C}$ je $|\mathcal{C}|^{|\mathcal{B}|}$
- $S_F = \{F(k, \cdot);\ k \in \mathcal{K}\} \subseteq \mathcal{B}^{\mathcal{C}}$.
- $|S_F| = |\mathcal{K}|$.

Intuitivno: psevdo naključna funkcija je **varna**, če slučajno izbrane funkcije iz $\mathcal{B}^{\mathcal{C}}$ ne moremo razlikovati od slučajno izbrame funkcije iz S_F.

Varne psevdo naključne permutacije

- Vseh permutacij $\mathcal{B} \to \mathcal{B}$ je $|\mathcal{B}|!$
- $S_E = \{E(k, \cdot);\ k \in \mathcal{K}\} \subseteq S(\mathcal{B})$.
- $|S_E| = |\mathcal{K}|$.

Intuitivno: psevdo naključna permutacija je **varna**, če slučajno izbrane permutacije iz $S(\mathcal{B})$ ne moremo razlikovati od slučajno izbrame permutacije iz S_E.

Definicija varnosti kot igra z dvema igralcema

Dva igralca: **izzivalec** in **nasprotnik**.

Definiramo dva poskusa: Poskus(0) in Poskus(1).
Ideja: nasprotnik ne loči med obema poskusoma.

Izzivalec slučajno izbere $b \in \{0, 1\}$.

- Če $b = 0$, izbere $k \in \mathcal{K}$ in $f = F(k, \ )$.
- Če $b = 1$, izbere poljubno funkcijo $f : \mathcal{B} \to \mathcal{B}$.
- Nasprotnik A izbere $x_1, \ldots, x_q \in \mathcal{B}$.
- Izzivalec vrne $f(x_1), \ldots, f(x_q)$.
- Nasprotnik izračuna/ugane $b' \in \{0, 1\}$.

Psevdo naključna funkcija je **varna**, če za vse nasprotnike, ki imajo "polinomsko omejena" računska sredstva velja, da je

$$\text{Prednost}_{PRF}(A, F) = |P(\text{Poskus}(0) = 1) - P(\text{Poskus}(1) = 1)|$$

zanemarljiva (na primer manj kot 2^{-80}).

Zgled

Naj bo $F : \mathcal{K} \times \mathcal{B} \to \mathcal{C}$ varna psevdo naključna funkcija.

Ali je tudi

$$G(k, x) = \begin{cases} 0^{128}, & \text{če je } x = 0, \\ F(k, x), & \text{sicer.} \end{cases}$$

varna psevdo naključna funkcija?

Semantična varnost - uporaba bločne šifre enkrat

Dana je bločna šifra $(\mathcal{B}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$.

- Napadalec je prestregel en kriptogram (lahko več blokov).
- Njegov cilj je iz kriptograma ugotoviti neko informacijo v zvezi z besedilom.

Semantična varnost kot igra z dvema igralcema

Dva igralca: izzivalec in nasprotnik A.

Definiramo dva poskusa: Poskus(0) in Poskus(1).
Ideja: nasprotnik ne loči med obema poskusoma.

- Izzivalec izbere $k \in \mathcal{K}$ in $b \in \{0, 1\}$.
- Nasprotnik izbere dve besedili $x_0, x_1 \in \mathcal{B}$, $|x_0| = |x_1|$.
- Če $b = 0$, izzivalec vrne $E_k(x_0)$.
- Če $b = 1$, izzivalec vrne $E_k(x_1)$.
- Nasprotnik izračuna/ugane $b' \in \{0, 1\}$.

Kriptosistem $\mathcal{S} = (\mathcal{B}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ je **semantično varen**, če za vse nasprotnike, ki imajo "polinomsko omejena" računska sredstva velja, da je

$$\text{Prednost}_{\mathcal{S}}(A, \mathcal{S}) = |P(\text{Poskus}(0) = 1) - P(\text{Poskus}(1) = 1)|$$

zanemarljiva (na primer manj kot 2^{-80}).

Zgled

- ECB ni semantično varen, če ima sporočilo več kot en blok.
- Deterministični način s števcem je semantično varen način uporabe bločne šifre, če je bločna šifra varna psevdo naključna permutacija.

Semantična varnost - uporaba bločne šifre večkrat

Dana je bločna šifra $(\mathcal{B}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$.

Primeri uporabe istega ključa večkrat:

- kriptosistemi z javnim ključem,
- šifriranje datotek na disku,
- IPSec - nabor protokolov za zaščito internetnih komunikacij; več paketov je zašifriranih z istim ključem...

Potrebujemo torej varnost pred napadom z izbranim besedilom (chosen-plaintext attack CPA).

Semantična varnost kot igra z dvema igralcema

Dva igralca: izzivalec in nasprotnik A.

Definiramo dva poskusa: Poskus(0) in Poskus(1).
Ideja: nasprotnik ne loči med obema poskusoma.

- Izzivalec izbere $k \in \mathcal{K}$ in $b \in \{0, 1\}$.
- Nasprotnik izbere dva nabora besedil $x_{0,1} \ldots, x_{0,q}$ in $x_{1,1} \ldots, x_{1,q}$ iz $\mathcal{B}$, $|x_{0,i}| = |x_{1,i}|$.
- Če $b = 0$, izzivalec vrne $E_k(x_{0,i}), i = 1, \ldots, q$.
- Če $b = 1$, izzivalec vrne $E_k(x_{1,i}), i = 1, \ldots, q$.
- Nasprotnik izračuna/ugane $b' \in \{0, 1\}$.

Kriptosistem $\mathcal{S} = (\mathcal{B}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ je **semantično varen pred napadom z izbranim besedilom**, če za vse nasprotnike, ki imajo "polinomsko omejena" računska sredstva velja, da je

$$\text{Prednost}_{CPA}(A, \mathcal{S}) = |P(\text{Poskus}(0) = 1) - P(\text{Poskus}(1) = 1)|$$

zanemarljiva (na primer manj kot 2^{-80}).

Zgledi

- Deterministična šifra ni semantično varna pri večkratni uporabi ključa.
- Način uporabe bločne šifre CBC z naključno izbranim inicializacijskim vektorjem je semantično varen pred CPA, če je bločna šifra varna psevdo naključna permutacija in q ni prevelik v primerjavi z $|\mathcal{B}|$.

Zgled

OTP ni varen pred napadom aktivnega napadalca, sporočilo lahko spremenimo, a prejemnik ne bo opazil.

Zaključek

Semantična varnost pri CPA zagotavlja varnost pred pasivnim napadalcem, ne pa tudi

- celovitosti podatkov,
- avtentikacije (podatki so prišli od prave osebe).

Aktivni napadalec lahko sporočila spreminja, dodaja svoja.

Rešitev: šifriranje + avtentikacija (authenticated encryption).