CZU:37.016:512+004.056.55

METODE ALGEBRICE APLICATE ÎN SISTEMELE DE CRIPTARE EL GAMAL ȘI CELE POLIALFABETICE

Dorin AFANAS, doctor, conferențiar universitar

Universitatea de Stat din Tiraspol

Andrei MEDVEȚCHI, Fiodor TIMERCAN,

Academia Militară a Forțelor Armate „Alexandru cel Bun”, mun. Chișinău

Vasile NICHIFOROV, student, Universitatea de Stat din Tiraspol

Rezumat. *Securitatea informațională a căpătat un caracter decisiv în ultimele decenii. În prezentul articol sunt prezentate metode algebrice care pot fi utilizate în sistemul de criptare El Gamal și în criptanaliza sistemelor de criptare polialfabetice. Sunt cercetate avantajele și dezavantajele lor.*

Summary. *Information security has become crucial in recent decades. This article presents algebraic methods that can be used in the El Gamal encryption system and in the cryptanalysis of polyalphabetic encryption systems. Their advantages and disadvantages are investigated.*

Cuvinte cheie: *sistem de criptare, criptanaliza, cheie, logaritmul discret, modul, soluție.*

Keywords: *encryption system, cryptanalysis, key, discrete logarithm, module, solution.*

1. Sistemul de criptare El Gamal

Sistemul de criptare El Gamal, prezentat în anul 1985 [1] de grecul Taher ElGamal, se bazează pe problema logaritmului discret, care poate fi formulată astfel:

Fie p un număr prim și $\alpha, \beta \in Z_p$, $\beta \neq 0$. Determinați $a \in Z_{p-1}$ astfel încât să fie justă relația: $\alpha^a \equiv \beta \pmod p$.

Dacă numărul întreg a există, atunci el este unic și se notează $\log_\alpha \beta$.

Exemplul 1. Dacă $p = 11$ și $\alpha = 6$, atunci elementele din Z_{11} pot fi exprimate ca puteri ale lui α:

a	0	1	2	3	4	5	6	7	8	9
$6^a \pmod{11}$	1	6	3	7	9	10	5	8	4	2

De aici rezultă imediat tabela logaritmilor în baza 6:

β	1	2	3	4	5	6	7	8	9	10
$\log_6 \beta$	0	9	2	8	6	1	3	7	4	5

Însă pentru $\alpha = 3$ nu întotdeauna vom avea soluție, deoarece

a	0	1	2	3	4	5	6	7	8	9
$3^a \pmod{11}$	1	3	9	5	4	1	3	9	5	4

valorile $\beta = \{2, 6, 7, 8, 10\}$ nu pot fi exprimate ca logaritmi în baza 3. Altfel spus, ecuația $\log_3 x = \beta$ nu admite soluție în Z_{11} pentru aceste valori ale lui b.

De asemenea, dacă $p = 7$ și $\alpha = 4$ atunci obținem:

a	0	1	2	3	4	5
$4^a \pmod 7$	1	4	2	1	4	2

și deci pentru $\beta = \{3, 5\}$ ecuația $\log_4 x = \beta$ nu admite soluție în Z_7.

Observația 1. Pentru problema logaritmului discret, nu este obligatoriu ca p să fie un număr prim. Important este ca α să fie rădăcină primitivă de ordinul $p - 1$ a unității, adică pentru orice i, $0 < i < p - 1$, avem α_i nu este congruent cu 1 după modul p. Teorema lui Fermat asigură că $\alpha^{p-1} \equiv 1 \pmod p$. La o alegere convenabilă a lui p, problema este NP – completă. Pentru siguranță, p se alege de minim 512 biți (pentru o securitate pe termen lung se recomandă 1024 biți [2], iar $p - 1$ să aibă cel puțin un divizor prim ”mare”. Pentru un astfel de modulo p, spunem că problema logaritmului discret este dificilă în Z_p. Utilitatea acestei cerințe rezidă în faptul că, deși este foarte dificil de calculat un logaritm discret, operația inversă – de exponențiere este foarte simplă.

Sistemul de criptare El Gamal este următorul:

Fie p număr prim pentru care problema logaritmului discret în Z_p este dificilă și $\alpha \in Z_p$ primitiv.

Fie $P = Z_p$, $C = Z_p \times Z_p$ și $K = \{(p, \alpha, a, \beta): \beta \equiv \alpha^a \pmod p\}$. Valorile p, α și β sunt publice, iar a este secret. Pentru $K = (p, \alpha, a, \beta)$ și $k \in Z_{p-1}$ aleator (secret) se definește $e_K(x, k) = (y_1, y_2)$, unde $y_1 \equiv \alpha^k \pmod p$, $y_2 \equiv x \cdot \beta^k \pmod p$.

Pentru $y_1, y_2 \in Z_p$ se definește $d_K(y_1, y_2) = y_2 \cdot \left(y_1^a\right)^{-1} \pmod p$.

Verificarea este imediată: $y_2 \cdot \left(y_1^a\right)^{-1} = x \cdot \beta^k \cdot \alpha^{-ka} = x \cdot \beta^k \cdot \beta^{-k} = x \pmod p$.

Sistemul este nedeterminist: criptarea depinde de x și de o valoare aleatoare aleasă de X. Există deci mai multe texte criptate corespunzătoare unui anumit text clar.

Exemplul 1. Să considerăm $p = 2579$, $\alpha = 2$ și $a = 765$. Prin calcul obținem $\beta = 2765 \pmod{2579} = 949$. Admitem că X dorește să expedieze mesajul $x = 1299$. El alege aleator k (fie de exemplu, $k = 853$) și calculează $y_1 = 2853 = 435 \pmod{2579}$, apoi $y_2 = 1299 \cdot 949853 = 2396 \pmod{2579}$. Când Y primește mesajul criptat $y = (435, 2396)$, el va determina $x = 2396 \cdot 435^{-765} = 1299 \pmod{2579}$.

Observația 2.

1. Un dezavantaj al sistemului El Gamal constă în dublarea lungimii textului criptat (comparativ cu lungimea textului clar).

2. Dacă (y_1, y_2), (z_1, z_2) sunt textele criptate ale mesajelor m_1, m_2, atunci se poate deduce imediat un text criptat pentru m_1m_2: (y_1z_1, y_2z_2). Similar poate fi dedusă o criptare pentru $2m_1$ (sau $2m_2$). Acest lucru face sistemul El Gamal sensibil la un atac cu text clar ales.

3. Indicația ca pentru criptarea a două texte diferite să se folosească valori diferite ale parametrului k este esențială: astfel, să presupunem că mesajele m_1, m_2 au fost criptate în (y_1, y_2) respectiv (z_1, z_2) folosind același k. Atunci $y_2/z_2 = m_1/m_2$ și cunoașterea unuia din mesaje îl determină imediat pe celălalt.

Sistemul de criptare El Gamal se poate construi pe orice grup (în loc de Z_n) în care problema logaritmului, definită corespunzător este dificilă. Prin urmare, sistemul de criptare El Gamal poate fi generalizat.

Fie $(G, \circ)$ un grup finit. Problema logaritmului discret se definește în G în modul următor:

Fie $\alpha \in G$ și $H = \{\alpha^i : i \geq 0\}$ subgrupul generat de α. Dacă $\beta \in H$, să se determine un a (unic) $(0 \leq a \leq card(H) - 1)$ cu $\alpha^a = \beta$, unde $\alpha^a = \underbrace{\alpha \circ \alpha \circ \alpha \circ \ldots \circ \alpha}_{a \text{ ori}}$.

Sistemului de criptare El Gamal în subgrupul H în loc de Z_n îl definim astfel:

Fie $(G, \circ)$ un grup și $\alpha \in G$ pentru care Problema Logaritmului Discret în $H = \{\alpha^i : i \geq 0\}$ este dificilă. Fie $P = G$, $C = G \times G$ și $K = \{(G, \alpha, a, \beta) : \beta = \alpha^a\}$. Valorile α și β sunt publice, iar a este secret. Pentru $K = (G, \alpha, a, \beta)$ și un $k \in Z_{card(H)}$ aleator (secret), se definește $e_K(x, k) = (y_1, y_2)$, unde $y_1 = \alpha^k$, $y_2 = x \circ \beta^k$. Pentru $y = (y_1, y_2)$, decriptarea este $d_K(y) = y_2 \circ \left(y_1^a\right)^{-1}$.

De remarcat faptul, că pentru criptare/decriptare nu este necesară cunoașterea ordinului $card(H)$ de mărime al subgrupului. X poate alege aleator un k, $(0 \leq k \leq card(G) - 1)$ cu care cele două procese funcționează fără probleme. Se poate observa de asemenea că G nu este neapărat abelian (H în schimb este, fiind subgrup ciclic).

Să studiem acum problema logaritmului discret "generalizat". Deoarece H este subgrup ciclic, orice versiune a problemei este echivalentă cu Problema Logaritmului Discret într-un grup ciclic. În schimb, se pare că dificultatea problemei depinde mult de reprezentarea grupului utilizat.

Astfel în grupul aditiv Z_n, problema este simplă: aici exponențierea α^a este de fapt înmulțirea cu a modulo n. Deci, Problema Logaritmului Discret constă în aflarea unui număr întreg a astfel încât $a\alpha \equiv \beta \pmod{n}$.

Dacă se alege α astfel ca $cmmdc(\alpha, n) = 1$ (α este generator al grupului), α are un invers multiplicativ modulo n, care se determină ușor cu algoritmul lui Euclid extins. Atunci,

$$a = \log_\alpha \beta = \beta\alpha^{-1} \pmod{n}.$$

Să vedem cum se reprezintă Problema Logaritmului Discret în grupul multiplicativ Z_p cu p prim. Acest grup este ciclic de ordin $p - 1$, deci izomorf cu grupul aditiv Z_{p-1}. Deoarece Problema

Logaritmului Discret se poate rezolva ușor într-un grup aditiv, apare întrebarea dacă putem rezolva această problemă în Z_p reducând-o la Z_{p-1}. Se cunoaște că există un izomorfism φ: $Z_p \to Z_{p-1}$, deci pentru care

$$\varphi(xy \bmod p) = (\varphi(x) + \varphi(y)) \pmod{(p-1)}.)$$

În particular, $\varphi(\alpha^a \bmod p) = a\varphi(\alpha) \pmod{(p-1)}$, adică

$$\beta \equiv \alpha^a \pmod{p} \Leftrightarrow a\varphi(a) \equiv \varphi(\beta) \pmod{(p-1)}.$$

Acum, determinarea lui a se realizează cu $\log_\alpha \beta = \varphi(\beta)(\varphi(\alpha))^{-1} \pmod{(p-1)}$.

Deci, dacă se găsește o metodă eficace pentru calculul izomorfismului φ, atunci se obține un algoritm eficace pentru calculul logaritmului discret în Z_p. Problema este că nu se cunoaște nici o metodă generală de construcție a lui φ pentru un număr prim p oarecare. Deși se știe că cele două grupuri sunt izomorfe, nu există încă un algoritm eficient pentru construcția explicită a unui astfel de izomorfism. Această metodă se poate aplica problemei logaritmului discret într-un grup finit arbitrar. Implementările au fost realizate în general pentru Z_p, (unde Problema Logaritmului Discret este dificilă) sau curbe eliptice.

2. Criptanaliza sistemelor de criptare polialfabetice

Atacul sistemelor polialfabetice este similar cu atacul a n sisteme de substituție monoalfabetică.

În criptanaliză, pentru determinarea unui cifru de substituție, se parcurg următoarele etape:

1) *analiza criptogramelor:*

1.1) pregătirea unui tabel de frecvențe;

1.2) căutarea repetițiilor;

1.3) determinarea tipului de sistem utilizat;

1.4) pregătirea unei foi de lucru;

1.5) pregătirea unui alfabet individual;

1.6) tabelarea repetițiilor lungi.

2) *Clasificarea vocalelor și consoanelor prin studierea:*

2.1) frecvențelor;

2.2) spațiilor;

2.3) combinațiilor de litere;

2.4) repetițiilor.

3) *Identificarea literelor:*

3.1) partiționarea literelor în clase de probabilitate;

3.2) verificarea presupunerilor;

3.3) înlocuirea valorilor corecte în criptogramă;

3.4) descoperirea altor valori pentru a avea soluția completă.

4) *Reconstrucția sistemului:*

4.1) reconstrucția tabelei de cifrare;

4.2) reconstrucția cheilor folosite în operația de cifrare;

4.3) reconstrucția cheilor sau a cuvintelor cheie ce au fost utilizate pentru construcția șirurilor de alfabete.

De exemplu, criptanaliza sistemului Vigenere constă în următoarele: fie $c = c_0\, c_1 \ldots c_{n\text{-}1}$ textul criptat cu cheia $k = k_0\, k_1 \ldots k_{p-1}$. Putem aranja acest text sub forma unui tabel cu p linii şi $[n/p]$ coloane, astfel:

c_0	c_p	c_{2p}	…
c_1	c_{p+1}	c_{2p+1}	…
…	…	…	…
c_{p-1}	c_{2p-1}	c_{3p-1}	…

Elementele de pe prima linie au fost criptate după formula:

$$c_{pr} = a_{pr} + k_0 \pmod{26},\ k \geq 0,$$

adică cu un sistem Cezar (k_0 fiind o valoare fixată din Z_{26}). În mod similar şi celelalte linii.

Prin urmare, dacă s-ar cunoaşte lungimea p a cheii, problema s-ar reduce la criptanaliza a p texte criptate cu Cezar – sistem de criptare monoalfabetic. Sânt cunoscute următoarele metode pentru aflarea lungimii cheii: testul lui Kasiski și indexul de coincidenţă. Prima metodă constă în studiul textului criptat și aflarea de perechi de segmente de cel puţin 3 caractere identice (această lungime este propusă de Kasiski). Pentru fiecare astfel de pereche, se determină distanta dintre segmente. După ce s-au găsit mai multe astfel de distanţe, valoarea lui p va fi cel mai mare divizor comun al lor (sau – eventual un divizor al acestuia).

A doua metodă de aflare a lungimii cheii de criptare într-un sistem Vigenere se bazează pe un concept definit în anul 1920 de Wolfe Friedman [3] – *indexul de coincidenţă*. Dacă $c = c_1c_2...c_n$ este o secvenţă de n caractere alfabetice, probabilitatea ca două caractere din c, alese aleator, să fie identice se numeşte ”*index de coincidenţă*” $I_c(x)$ al lui c.

Exemplul 2. Admitem ca s-a interceptat următorul text criptat, despre care se face presupunerea că s-a folosit sistemul Vigenere:

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

Textul este destul de scurt (146 litere) și nu se mai cunoaște nici un text trimis anterior. Folosind metoda Kasiski, se găsește secvența QLNCJ care apare pe rândul al doilea. Distanța dintre

cele două apariții este 27. De asemenea, apar două cuvinte foarte asemănătoare: AQLN și AOLN, având între ele distanța 57. Deci putem presupune că avem de-a face cu un cuvânt cheie de lungime cmmdc(27, 57) = 3. Rescriem textul pe coloane, fiecare coloană având trei elemente:

D O O C W F C R S A S H Q C S J W W H I A Q C I Q G S S W G C O S C W S P R W W W G

V E G G W R K V R K T D L J R V E K Z E R L J K V K J G D K W L Z V Z C I V Y V E T

L G L I A S A S A R U A N T U C A O T U I N C A A A T R A R A N P Z C E E M A M E U

Numărând frecvența apariției literelor pe fiecare linie, obținem tabelul:

	A	B	C	D	E	F	G	H	I	J	K	L	M
Linia 1	2	0	6	1	0	1	3	2	2	1	0	0	0
Linia 2	0	0	1	2	4	0	3	0	1	3	6	3	0
Linia 3	11	0	3	0	3	0	1	0	2	0	0	2	2

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Linia 1	0	3	1	3	2	7	0	0	1	8	0	0	0
Linia 2	0	0	0	0	4	0	2	0	6	2	0	1	3
Linia 3	3	1	1	0	3	2	3	4	0	0	0	0	1

În limba română, primele litere cu frecventă mare sunt *A*−*E*−*I*, aflate la distanță egală una de alta. Deci vom căuta pe fiecare linie tripletele de litere situate pe pozițiile (k, k + 4, k + 8) având frecvența semnificativ de mare (maximă în cazul unui text lung). Pentru linia a 3-a, alegerea este simplă: ea este chiar *A*−*E*−*I* (16 apariții din 49 posibile), deci o deplasare 0 în codul lui Cezar. Pentru prima linie, sunt două posibilități: *O*−*S*−*W* (deplasare 14) sau *S*−*W*−*A* (deplasare 18), ambele cu câte 18 apariții.

Tot două variante apar și pentru a doua linie: *C*−*G*−*K* (deplasare 2) cu 10 apariții, sau *R*−*V*−*Z* (deplasare 14) cu 13 apariții. Deplasările dau exact codificările cheii. Deci trebuie luate în considerare patru variante de cuvânt cheie: OCA, ORA, SCA sau SRA. Cum de obicei cuvântul cheie are o semnificație semantică (pentru a putea fi reținut mental ușor), putem presupune că el este OCA sau ORA. O simplă verificare reține drept cuvânt cheie ORA, care conduce la decriptarea corectă a textului (spațiile și semnele de punctuație se pun corespunzător):

PE LANGA PLOPII FARA SOT ADESEA AM TRECUT MA CUNOSTEAU VECINII TOTI TU NU MAI CUNOSCUT ACEASTA ESTE PRIMA STROFA A UNEI POEZII CELEBRE DE MIHAI EMINESCU

Concluzii

1. Pentru problema logaritmului discret, nu este obligatoriu ca p să fie un număr prim. Important este ca α să fie rădăcină primitivă de ordinul $p - 1$ a unității, adică pentru orice i, $0 < i < p - 1$, avem α_i nu este congruent cu 1 după modulo p.

2. Un dezavantaj al sistemului El Gamal constă în dublarea lungimii textului criptat (comparativ cu lungimea textului clar).

3. Sistemul de criptare El Gamal se poate construi pe orice grup (în loc de Z_n) în care problema logaritmului, definită corespunzător este dificilă. Prin urmare, sistemul de criptare El Gamal poate fi generalizat.

4. Putem aplica două metode la aflarea lungimii cheii pentru sistemul de criptare polialfabetic: testul lui Kasiski și indexul de coincidență.

Bibliografie

1. EL GAMAL, T. A public key cryptosystem and a signature scheme based on discrete algorithms, IEEE Transactions on Information Theory. 31 (1985), pp. 469 – 472.
2. MENEZES, A., OORSCHOT, P., VANSTOME, S. Handbook of applied cryptography, crc press; Ediția a 2-a, 1997, 250 p.
3. FRIEDMAN, W. Military Cryptanalysis, Aegean Park Press, 1980, 150 p.