MCELIECE'OV KRYPTOSYSTÉM

Václav Šrenkel, Mgr. Zuzana Ševčíková *
KAIVT, FEI STU
vaclav.srenkel@gmail.com

Abstrakt

Táto práca sa zaoberá McEliece'ovým kryptosystémom. Úvod práce obsahuje krátky prehľad o prínosoch kryptografie a tiež popis asymetrickej kryptografie. V prvej časti práce je opis kryptosystému obsahujúci proces šifrovania a dešifrovania a tiež pôvodné a upravené hodnoty parametrov kryptosystému. Ďalšia časť sa zaoberá známymi útokmi na tento kryptosystém, ktoré sa delia na kritické a nekritické. Záver práce obsahuje porovnanie McEliece'ovho kryptosystému s Niederreiterovým kryptosystémom a systémom RSA.

1. Úvod

V dnešnej dobe sa čoraz častejšie stretávame s pojmami kryptológia, kryptografia či bezpečnosť dát. Kryptografia je veda, ktorá sa zaoberá vytváraním šifier. Používa metódy šifrovania na ukrytie citlivých údajov a informácií pred nepovoleným prístupom. Šifrovanie je proces, v ktorom vybraná kryptografická metóda transformuje otvorený text pomocou kryptografického algoritmu a šifrovacieho kľúča do zašifrovaného textu. Dešifrovanie je proces opačný k šifrovaniu. Ak sa pri dešifrovaní použije ten istý kľúč ako pri šifrovaní, ide o symetrickú šifru. Pri asymetrických šifrách sa používajú dva kľúče.

Poznáme tri najvýznamnejšie kryptografické služby:

-utajenie údajov - vo všetkých formách jej existencie (prenos, uloženie na médiách, ...)

-autentizácia - potvrdenie o totožnosti subjektu alebo objektu

-integrita informácií - potvrdenie o správnosti obsahu prenesenej správy.

Ak chceme preniesť tajnú informáciu po verejne prístupnom kanáli, musíme ju zašifrovať. Tu nastáva potreba zaistenia, že počas prenosu nebola správa zmenená. Taktiež je potrebné vedieť, od koho sme správu dostali a či je skutočne on pôvodným odosielateľom.

Prvé asymetrické šifrovacie algoritmy sa objavili v 70. rokoch minulého storočia. Sú založené na jednoduchej myšlienke, že správa sa nedešifruje tým istým kľúčom, ktorým bola šifrovaná. Asymetrická kryptografia je forma kryptografie, v ktorej má používateľ dva kľúče: verejný a súkromný. Verejný kľúč je voľne šíriteľný a používa sa iba na zašifrovanie dát. Súkromný kľúč je tajný a vlastník ho vo vlastnom záujme nesmie nikomu prezradiť. Tento kľúč sa používa iba na dešifrovanie dát. Tieto kľúče spolu matematicky súvisia, ale súkromný kľúč nemôže byť prakticky odvodený od verejného kľúča. Správa zakódovaná s verejným kľúčom môže byť dekódovaná len s príslušným súkromným kľúčom.

Robert McEliece predstavil v roku 1978 kryptosystém s verejným kľúčom, ktorý je založený na matematickej teórii kódovania. Tento algoritmus sa však kvôli určitým nedostatkom nezačal využívať v praxi. Kryptosystém je logický systém, ktorý umožňuje šifrovanie a dešifrovanie údajov. V McEliece'ovom kryptosystéme príjemca najprv skonštruuje ľahko riešiteľný lineárny korekčný Goppa kód C s generujúcou maticou G a potom transformuje túto maticu na maticu G', teda na maticu generujúcu zdanlivo ťažko riešiteľný kód C'. Matica G' a počet chýb t sú verejným kľúčom tohto systému, pretože správa je zakódovaná násobením touto maticou G' a pridaním chýb do výsledných zakódovaných slov. Legitímny príjemca vie rozlúštiť správu s použitím dešifrovacieho algoritmu pre východzí kód C.

Tento kryptosystém využíva triedu samoopravných kódov, tzv. Goppa kódy (sú to lineárne kódy, pre ktoré existujú rýchle dekódovacie algoritmy). Tieto samoopravné kódy sa používajú na prenos dát cez rušený kanál. Fungujú tak, že zakódujú dáta (pridaním redundancie) a takto potom môže byť správa obnovená aj ak sa vyskytne niekoľko chýb.

2. Popis McEliece'ovho kryptosystému

Príjemca správy skonštruuje ľahko riešiteľný binárny korekčný Goppa kód C, ktorý má generujúcu maticu G s rozmerom (k x n) a schopnosť opraviť najviac t chýb v jednom kódovom slove. Matica G' je vypočítaná pomocou vzťahu

$$G' = SGP \qquad (1)$$

* Vedúci práce

kde S je regulárna binárna štvorcová matica o rozmere $(k \times k)$ a P je permutačná matica s rozmerom $(n \times n)$. Matica G' s rozmerom $(k \times n)$ je potom zrejme generujúca matica lineárneho kódu C' (t.j. jediná, pre ktorú je známy rýchly algoritmus pre opravu chýb).
Matica G' a počet chýb t sú zverejnené ako šifrovací (verejný) kľúč. Odosielateľ zašifruje k-bitový vektor správy m do n-bitového šifrovaného textu c takto:

$$c = mG' \oplus e \tag{2}$$

kde e je n-bitový chybový vektor s počtom chýb t, ktorý je ľubovoľne zvolený odosielateľom, alebo je náhodne generovaný.
Príjemca, ktorý samozrejme vie, že

$$c = mG' \oplus e = mSGP \oplus e \tag{3}$$

si vypočíta

$$cP^{-1} = (mS)G \oplus eP^{-1} \tag{4}$$

a použije dekódovací algoritmus pre pôvodný kód c na odstránenie chybového vektora eP^{-1} a obnoví vektor mS. Potom je už pre príjemcu ľahko zistiteľná správa, ktorá bola napísaná odosielateľom, takto:

$$m = (mS)S^{-1} \tag{5}$$

Súkromný kľúč tohto systému preto pozostáva z troch matíc G, S a P.
Originálne hodnoty parametrov, ktoré pre tento kryptosystém odporučil McEliece boli:

$n = 1024$
$t = 50$
$k \geq 524$

Po následnom štúdiu a bezpečnostných analýzach tohto kryptosystému sa ukázalo, že nie je s odporúčanými parametrami dostatočne bezpečný a preto boli zvolené nové optimálne hodnoty pre Goppa kód, ktoré maximalizujú útočníkovu náročnosť útoku takto:

$n = 1024$
$t = 38$
$k \geq 644$

Hoci šifrovacie a dešifrovacie operácie tohto kryptosystému sú pomerne rýchle, McEliece'ova schéma pripúšťa aj určité nedostatky:
1. Veľmi veľký verejný kľúč.
Použitím Goppa kódu pre odporúčané parametre $n = 1024$, $t = 38$ a $k \geq 644$ je verejný kľúč veľký 2^{19} bitov, čo môže spôsobiť implementačné problémy.
2. Veľká dĺžka zašifrovanej správy.
Menej závažným nedostatkom je expanzia správy koeficientom n / k. Pre odporúčané parametre $n = 1024$, $t = 38$ a $k \geq 644$ je veľkosť správy 1,6-násobkom otvoreného textu.
Z týchto dôvodov tento kryptosystém získal nízku pozornosť v praxi.

3. Útoky na kryptosystém

Nasledujúca časť popisuje známe útoky na McEliece'ov kryptosystém.

3.1. Nekritické útoky

Účinnosť nasledujúcich útokov bola znížená jednoduchým zväčšením veľkosti parametra, alebo aplikovaním Loidreau'ovej modifikácie bez zväčšenia veľkosti parametra. Preto sa nazývajú nekritické.

3.1.1. Generalized Information-Set-Decoding Attack

Budeme vychádzať z vyššie spomenutej rovnice (2). Keďže m je k-bitový vektor, môžme rovnicu upraviť nasledovne:

$$c_k = mG'_k \oplus e_k \tag{6}$$

kde c_k predstavuje nejakú k prvkovú zložku c ($c_k = c_{i1}, c_{i2}, ..., c_{ik}$), e_k predstavuje prislúchajúcu k zložku e, a G'_k je štvorcová matica pozostávajúca zo stĺpcov $i_1, i_2, ..., i_k$ matice G'. Takto dostaneme

$$c_k \oplus e_k = mG'_k \tag{7}$$

alebo, ak G'_k je inverzná matica, potom dostaneme

$$(c_k \oplus e_k)(G'_k)^{-1} = m \tag{8}$$

Je potrebné dať pozor, aby neboli všetky zložky e_k nulové, pretože tým by sa rovnica zredukovala nasledovne:

$$c_k (G'_k)^{-1} = m \tag{9}$$

a nepriateľ by ľahko vedel rozlúštiť správu bez dekódovania, pretože c a G' sú známe.
Náročnosť tohto útoku sa dá vypočítať nasledovne. Chybový vektor e je n-bitový vektor s t jednotkami a $(n - t)$ nulami. Preto pravdepodobnosť zvolenia k nulových komponentov z e je

$$p = \binom{n-t}{k} / \binom{n}{k} \tag{10}$$

Nepriateľ musí uskutočniť v priemere $1/p$ pokusov, kým bude úspešný, pričom pre každý pokus musí invertovať $(k \times k)$ submaticu G_k. Predpoklad, že na vytvorenie inverznej matice je potrebných k^a krokov, dáva celkovú predpokladanú náročnosť pre tento útok

$$w = k^a \binom{n}{k} / \binom{n-t}{k} \tag{11}$$

3.1.2. Finding-Low-Weight-Codeword Attack

Tento útok využíva algoritmus hľadajúci nízke hodnoty kódových slov spomedzi kódových slov generovaných ľubovoľnou generujúcou maticou použitím databázy získanej výpočtami. Pretože minimálna hodnota kódových slov nasledujúcej generujúcej matice

$$\begin{bmatrix} G' \\ c \end{bmatrix} \tag{12}$$

s rozmerom $(k + 1)\,x\,n$ je chybový vektor e zo zašifrovaného textu c, tento algoritmus môže byť použitý na obnovenie daného zašifrovaného textu c.

3.2. Kritické útoky

Účinnosť nasledujúcich útokov nemôže byť znížená zväčšením veľkosti parametra ani aplikovaním Loidreau'ovej modifikácie. Preto sa tieto útoky nazývajú kritické.

3.2.1. Known-Partial-Plaintext Attack

Čiastočná znalosť otvoreného textu výrazne zredukuje výpočtové náklady útoku na tento kryptosystém. Napríklad nech m_l a m_r označujú ľavé k_l bity a ostávajúce k_r bity v otvorenom texte m, t.j. $k = k_l + k_r$ a $m = (m_l \| m_r)$. Pripusťme, že útočník pozná m_r. Potom problém obnovenia neznámeho otvoreného textu m_l zašifrovaného McEliece'ovým kryptosystémom s parametrami (n, k) je ekvivalentný obnoveniu celého otvoreného textu zašifrovaného týmto kryptosystémom s parametrami (n, k_l). Budeme vychádzať zo vzťahu (2).

$$c = m_l G'_l \oplus m_r G'_r \oplus e \tag{13}$$

$$c \oplus m_r G'_r = m_l G'_l \oplus e \tag{14}$$

$$c' = m_l G'_l \oplus e \tag{15}$$

kde G'_l a G'_r sú horné k_l riadky a ostatné sú spodné k_r riadky v matici G' v tomto poradí. Ak k_l je fixované na malú hodnotu, výpočtové náklady na obnovenie neznámych k_l bitov z c, m_r a G' sú potom rovné polynómu z n ak je použitý nekriticky útok, je to asymptoticky ohraničené vzťahom $k_l^3 C(n, k_l / C(n - t, k_l))$, kde k_l je malá konštanta.

3.2.2 Related-Message Attack

Tento útok využíva znalosť vzťahov medzi otvorenými textami. Predstavme si, že máme dve správy m_1 a m_2, ktoré sú, každá zvlášť, zašifrované na správy c_1 a c_2, kde

$$c_1 = m_1 G' \oplus e_1 \tag{16}$$

$$c_2 = m_2 G' \oplus e_2 \tag{17}$$

a platí, že $e_1 \neq e_2$. Ak útočník pozná lineárnu závislosť, napr. $\delta m = m_1 \oplus m_2$ potom vie účinne uplatniť tzv. Generalized information-set-decoding attack na c_1 alebo c_2 zvolením k súradníc, ktorých hodnoty sú 0 vo vzťahu $\delta m G' \oplus c_1 \oplus c_2$. Potom

$$e_1 \oplus e_2 = \delta m G' \oplus c_1 \oplus c_2 \tag{18}$$

a Hammingová hodnota t chybového vektora e je oveľa menšia ako $n/2$. Z tohto dôvodu nulové súradnice vo vzťahu $\delta m G' \oplus c_1 \oplus c_2$ musia byť s vysokou pravdepodobnosťou nulové v e_1 a e_2. Keď tú istú správu viackrát zašifrujeme použitím rozdielnych chybových vektorov e_1 a e_2, hodnota $e_1 \oplus e_2$ je daná pomocou vzťahu $c_1 \oplus c_2$. Tento jav je odkázaný na tzv. "Mesage-resend attack".

3.2.3. Reaction Attack

Pri tomto útoku sleduje útočník reakciu príjemcu správy, ktorý ako jediný pozná svoj súkromný kľúč, ale neočakáva prijatie zašifrovanej správy. Myšlienkou tohto typu útoku je, že útočník preklopí jeden alebo niekoľko málo bitov zašifrovaného textu c. Označme zmenený zašifrovaný text ako c'. Útočník pošle tento zmenený zašifrovaný text príjemcovi a čaká na reakciu príjemcu:
Reakcia 1:
Príjemca pošle žiadosť o opakované zaslanie správy z dôvodu indikovanej chyby, alebo z dôvodu nezmyselného rozšifrovaného textu.
Reakcia 2:
Príjemca pošle potvrdenie o prijatej správe, alebo nič nerobí, pretože je dešifrovaný pravý otvorený text m.

Ak celková hodnota chybového vektora nepresiahla hodnotu t, útočník by mal zaznamenať reakciu č. 2, ináč zaznamená reakciu č. 1. Takýmto opakovaním vie útočník určiť chybový vektor a potom už dokáže ľahko rozšifrovať správu použitím Generalized Information-Set-Decoding útoku.

3.2.4 Malleability Attack

Pomocou tohto útoku môže útočník zmeniť časť otvoreného textu v danom zašifrovanom texte bez toho, aby poznal otvorený text, t.j. útočník vytvorí nový zašifrovaný text c', ktorého otvorený text je $m' = m \oplus \delta m$ z daného zašifrovaného textu c bez poznania m. Zašifrovaný text sa potom vypočíta nasledovne:

$$c' = c \oplus G[i] = (m \oplus \delta m)G' \oplus e = m'G' \oplus e \tag{19}$$

Takto vie útočník pomocou dešifrovacieho orákula dešifrovať daný zašifrovaný text. Najprv požiada o dešifrovanie c' a dostane odpoveď $m' = m \oplus \delta m$. Potom už vie obnoviť otvorený text zo zašifrovaného pomocou $m = m' \oplus \delta m$.

4. Porovnanie

4.1. Niederreiter

Niederreiter navrhol verziu kryptosystému, v ktorom verejným kľúčom je matica H' kontrolujúca paritu kódu C' ekvivalentného s tajným kódom. Otvorený text m je n-bitový vektor s hodnotou t a priradený šifrovaný text x je relatívne vyhovujúci so syndrómom správy m k verejnému kódu,

$$x = mH'^{t} \quad (20)$$

Kryptosystémy McEliece'a a Niederreitera sú vlastne rovnaké z pohľadu bezpečnosti, pokiaľ sú nastavené správnymi parametrami. Dané parametre Niederreiterovho kryptosystému majú mnoho výhod.

1. Veľkosť verejného kľúča je $(n - k) / n$ -krát menšia ako v McEliece'ovom systéme.
2. Systematická forma matice H' a nízke hodnoty vektora m významne zmenšujú výpočtové náklady.
3. Na rozdiel od McEliece'ovho kryptosystému, ak Niederreiterovým systémom zašifrujeme otvorený text dvakrát tým istým verejným kľúčom, dostaneme dva rovnaké výsledky. U McEliece'a každým zašifrovaním toho istého otvoreného textu rovnakým verejným kľúčom dostaneme vždy iný zašifrovaný text, čo je zapríčinené použitím náhodného chybového vektora.

4.2. RSA

Kryptosystém RSA zverejnili v roku 1977 Rivest, Shamir a Adleman. Bezpečnosť RSA je postavená na predpoklade, že rozložiť číslo na súčin prvočísel (faktorizácia) je ťažký problém. Verejným kľúčom tohto kryptosystému je dvojica (n, e), pričom platí $1 < e < n$ a tiež $n = pq$ (p a q sú veľké prvočísla),

$$\varphi(n) = (p-1)(q-1) \quad (21)$$

a platí $\gcd(e, \varphi(n))=1$. Vypočítaný súkromný kľúč (n, d) musí spĺňať rovnosť

$$ed \equiv 1 (\mathrm{mod}\, \varphi(n)) \quad (22)$$

Ak označíme otvorený text m a zašifrovaný c, potom pre šifrovanie a dešifrovanie platia tieto vzťahy:
-šifrovanie:

$$c \equiv m^e (\mathrm{mod}\, n) \quad (23)$$

-dešifrovanie:

$$m \equiv c^d (\mathrm{mod}\, n) \quad (24)$$

Fakt, že dešifrovaním získame pôvodnú správu je dôsledkom nasledujúcej rovnosti:

$$c^d \equiv (m^e)^d \equiv m^e d (\mathrm{mod}\, n) \quad (25)$$

Kým McEliece'ov kryptosystém je založený na probléme dekódovania pomocou syndrómov, RSA využíva problém faktorizácie. Výhodou RSA je to, že nemá tak veľký verejný kľúč ako McEliece'ov kryptosystém. Nevýhodou RSA je rýchlosť šifrovania a dešifrovania, kde McEliece'ov kryptosystém je oveľa rýchlejší (50x pri šifrovaní a 100x pri dešifrovaní).
V nasledujúcej tabuľke je porovnanie jednotlivých parametrov popísaných kryptosystémov. Pre RSA bol použitý systém využívajúci Karatsuba'ovu metódu veľkého celočíselného násobenia.

Tab. 1. Porovnanie jednotlivých kryptosystémov

	McEliece [1024,524,101] binárný kód	Niederreiter [1024,524,101] binárný kód	RSA 1024-bit. modul, e=17
Veľkosť verejného kľúča	67,072 bajtov	32,750 bajtov	256 bajtov
Počet bitov prenašaných pri šifrovaní	512	276	1024
Prenášané množstvo	51,17%	56,81%	100%
Počet operácií vykonaných na zašifrovanie 1 bitu	514	50	2,402
Počet operácií vykonaných na dešifrovanie 1 bitu	5,140	7,863	738,112

5. Záver

Ako už bolo spomenuté v texte vyššie, McEliece'ov kryptosystém má napriek pomerne rýchlym šifrovacím a dešifrovacím operáciám dva zásadné nedostatky. Prvým nedostatkom je veľká veľkosť verejného kľúča a druhým je veľkosť zašifrovanej správy, ktorá sa zväčšuje koeficientom n / k. Kvôli spomenutým nedostatkom sa tento kryptosystém neujal v praxi.

Použitá literatúra

[1] Menezes, A., et. al., “Handbook of applied cryptography”,
[2] Canteaut, A., Sendrier, N., “Cryptanalysis of the Original McEliece Cryptosystem”, http://www-rocq.inria.fr/codes/Anne.Canteaut/Publications/Canteaut_Sendrier98.pdf
[3] Adams, C., et. al., ”Security-Related Comments Regarding McEliece’s Cryptosystem“, http://dsns.csie.nctu.edu.tw/research/crypto/HTML/PDF/C87/224.PDF#search=%22McEliece's%20Crypto%20System%22
[4] Kobara, K., et. al., “Semantically Secure McEliece Public-Key Cryptosystems -Conversions for McEliece PKC-“ [citované 01, 2007] http://maths.utime.cn:81/Crypt1998-2003/papers/1992/19920019.pdf
[5] http://en.wikipedia.org/wiki/RSA